

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Advanced Blood Group Typing: A Long-Read Sequencing
Approach for Comprehensive Blood Group Identification

verfasst von | submitted by

Gergő Marton-Szücs BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 830

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Molecular Microbiology, Microbial
Ecology and Immunobiology

Betreut von | Supervisor:

ao. Univ.-Prof. i.R. Dipl.-Biol. Dr. Angela Witte

Abstract

Blood group typing is essential for safe transfusion medicine, transplantation, and the prevention of alloimmunization¹. Although serological methods remain the routine standard, they can be limited in cases involving rare variants, hybrid alleles, recently transfused patients, or loci with high sequence similarity between related genes^{1,2}. Molecular genotyping can overcome several of these limitations, but conventional targeted assays and short-read sequencing may not fully resolve structurally complex loci or determine which variants occur on the same allele^{1,3}. Long-read sequencing preserves linkage between variants across extended genomic regions, enabling haplotype phasing and the detection of novel variants, hybrid alleles, and structural rearrangements^{3,4}.

This thesis aimed to establish and optimize a long-range PCR-based enrichment workflow for clinically relevant blood group loci, including ABO, KEL, MNS-associated regions, and RHD/RHCE-associated regions, compatible with downstream Oxford Nanopore sequencing. Several commercially available DNA polymerases and PCR master mixes were compared for large-fragment amplification. Takara PrimeSTAR® GXL Fast DNA Polymerase provided the best balance of amplification reliability, specificity, reproducibility, and PCR runtime and was therefore selected for further workflow development.

Long-range primer pairs were designed de novo or adapted from published approaches and experimentally tested for the selected loci. Agarose gel electrophoresis showed distinct high-molecular-weight products at the expected approximate positions for ABO, KEL, MNS-associated, and RHD/RHCE-associated regions. Three overlapping ABO amplicons and two amplicons covering the selected KEL regions were generated. Selected ABO and KEL products were subsequently analyzed by Oxford Nanopore sequencing. Sequencing confirmed their origin from the intended genomic regions, supporting primer specificity and the suitability of the optimized PCR workflow for downstream long-read sequencing.

MNS and RHD/RHCE-associated PCR products were not sequenced because of time and cost constraints and were therefore evaluated only as gel bands with approximate sizes consistent with the expected products. As agarose gel electrophoresis cannot establish genomic identity, these products cannot yet be considered validated target-specific amplicons and require sequencing-based confirmation to exclude off-target amplification from homologous regions.

Overall, this study establishes a foundation for targeted long-read blood group analysis and demonstrates the feasibility of combining optimized long-range PCR with Oxford Nanopore sequencing for selected ABO and KEL loci. The workflow provides a basis for future validation of MNS and RHD/RHCE-associated products and subsequent haplotype-resolved characterization of blood group alleles.

Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die mich während meines Studiums und bei der Entstehung dieser Masterarbeit begleitet und unterstützt haben.

Mein besonderer Dank gilt meiner Familie und insbesondere meinen Eltern, die mich während meines gesamten akademischen Werdegangs stets unterstützt, motiviert und mir den notwendigen Rückhalt gegeben haben. Ebenso möchte ich Kathi von Herzen danken, die mir sowohl im Privatleben als auch während herausfordernder Phasen meines Studiums stets zur Seite stand, viel Geduld und Verständnis zeigte und mich immer wieder ermutigte.

Ein besonderer Dank gilt außerdem Ivan, der für mich nicht nur ein herausragender Professor, sondern auch eine wichtige Inspirationsquelle während meines Studiums war. Seine Begeisterung für die Wissenschaft, seine Art zu lehren und sein persönlicher Umgang mit den Studierenden haben meinen akademischen Weg nachhaltig geprägt.

Darüber hinaus möchte ich mich herzlich bei meinen Arbeitskollegen Filip und Andrea bedanken. Mit ihrem Fachwissen, ihren wertvollen Ratschlägen und ihrer kontinuierlichen Unterstützung haben sie wesentlich zum Gelingen dieser Arbeit beigetragen. Ihre Hilfsbereitschaft und die zahlreichen fachlichen Gespräche waren für mich sowohl während der praktischen Arbeit als auch beim Verfassen dieser Masterarbeit von großer Bedeutung.

Abschließend danke ich allen weiteren Personen, die mich auf meinem bisherigen Weg begleitet, unterstützt und motiviert haben.

Abstract.....	2
Danksagung.....	3
1 Introduction and Background.....	5
1.1 Research Problem and Motivation.....	6
2 Objective and aim of the study.....	8
2.1 Specific Objectives.....	8
2.2 Target Blood Group Systems.....	9
2.2.1 ABO Blood Group System.....	9
2.2.2 Rh Blood Group System: RHD and RHCE.....	10
2.2.3 MNS Blood Group System.....	10
2.2.4 Kell Blood Group System.....	11
3 Materials and Methods.....	12
3.1 Study Design and General Workflow.....	12
3.2 Selection of Blood Group Target Regions.....	12
3.3 Primer Design Strategy.....	13
3.4 Primer Synthesis and Preparation.....	14
3.5 Primer Tails for Oxford Nanopore PCR Barcoding.....	14
3.6 Genomic DNA Templates.....	15
3.7 Polymerase Selection for Long-Range PCR.....	16
3.8 PCR Reaction Setup.....	17
3.9 PCR Thermocycling Conditions.....	18
3.10 Agarose Gel Electrophoresis and Visualization.....	19
3.11 PCR Product Cleanup and Size Selection.....	20
3.12 Oxford Nanopore Amplicon Sequencing.....	20
3.13 Sequencing Data Processing and Amplicon Confirmation.....	22
3.14 Documentation and Evaluation Criteria.....	22
4 Results.....	24
4.1 Polymerase Comparison and PCR Optimization.....	24
4.2 Gel-Based Evaluation of Primer Performance.....	25
4.2.1 ABO.....	25
4.2.2 KEL.....	26
4.2.3 RHD/RHCE-associated regions.....	27
4.2.4 MNS-associated regions.....	28
4.3 Confirmation of Amplicon Identity by Oxford Nanopore Sequencing.....	29
4.3.1 ABO sequencing results.....	29
4.3.2 KEL sequencing results.....	30
5 Discussion.....	32
5.1 Principal Findings and Scope of the Study.....	32
5.2 Long-Range PCR Optimization and Polymerase Selection.....	33
5.3 Primer Design and the Overlapping-Amplicon Strategy.....	34
5.4 Interpretation of the ABO and KEL Sequencing Results.....	34
5.5 Interpretation of the RHD/RHCE and MNS-Associated Results.....	36
5.6 Scientific and Potential Clinical Relevance.....	37
5.7 Limitations.....	38
5.8 Future Development and Conclusion.....	39
6 References.....	40
7 List of abbreviations.....	44
8 List of figures and tables.....	46
9 Appendix.....	47
9.1 Primer sequences used for long-range PCR amplification.....	47
9.2 Deutsche Zusammenfassung.....	50

1 Introduction and Background

Blood group typing is a central component of transfusion medicine, transplantation, pregnancy management, and the prevention of immune-mediated transfusion complications¹. In routine transfusion practice, serological testing remains the established standard for determining clinically relevant red blood cell antigen expression and supporting compatibility testing^{1,5}. Molecular genotyping is used to complement serology, particularly in recently transfused patients, patients with autoantibodies or complex antibody histories, cases involving weak or unavailable serological reagents, and extended donor typing^{1,5}. High-throughput molecular typing also enables large-scale testing of blood donors for numerous clinically relevant red blood cell antigens and the expansion of donor databases with extended antigen profiles⁶. Such databases support the identification of antigen-negative or extensively matched blood products for alloimmunized and chronically transfused patients^{1,6}.

Routine serological and molecular methods are highly effective for their intended clinical purposes, which include determining or predicting antigen profiles, compatibility testing, and donor-recipient matching^{1,5}. However, many routine molecular assays, including allele-specific PCR and single nucleotide polymorphism-based genotyping platforms, analyze predefined antigen-associated variants and use these variants to predict the expected phenotype^{1,5}. This targeted strategy is efficient for clinical blood group typing, but it may not detect previously undescribed variants or determine the complete sequence of a blood group gene or locus^{1,7}. Furthermore, conventional targeted assays generally cannot determine whether multiple variants are located together on the same allele or distributed between the two homologous chromosomes⁷.

Genetic variation underlying blood group alleles is not limited to exonic single nucleotide variants^{2,7}. It can also include small insertions and deletions, intronic and regulatory variants, gene conversions, deletions, duplications, hybrid alleles, and other structural rearrangements^{2,7}. At loci containing highly homologous genes, particularly RHD/RHCE and the glycophorin genes associated with the MNS blood group system, recombination and gene conversion can produce hybrid alleles that are difficult to resolve using assays restricted to individual variants or short genomic regions⁷⁻⁹. The identification of novel variants and reconstruction of complete allele sequences can therefore help explain unusual serological findings, improve blood group allele definitions, and expand reference databases^{7,10}.

Long-read sequencing technologies, including Oxford Nanopore sequencing, address several of these limitations by preserving the linkage between variants across extended genomic regions^{3,7}. Their principal advantage in blood group genetics is haplotype phasing, because variants detected on the same read or connected by overlapping reads can be assigned to the same allele and distinguished from variants located on the homologous chromosome⁷. Long-read sequencing can also support the detection of novel variants, characterization of intronic and regulatory variation, and resolution of structural variants and hybrid genes that are difficult to reconstruct from short sequencing reads^{3,7}. Long-read sequencing of the

complete ABO gene has already enabled the generation of fully resolved ABO haplotype sequences and the identification of variants with potential relevance for allele discrimination¹⁰. Long-read sequencing is therefore not intended to replace established serological testing for routine compatibility decisions, but to complement it by providing more comprehensive and haplotype-resolved genetic characterization of blood group loci^{1,7}.

1.1 Research Problem and Motivation

The research problem addressed in this thesis is the need for a reliable targeted enrichment strategy capable of generating large, locus-specific DNA fragments for long-read sequencing of selected blood group genes and genomic regions. Whole-genome long-read sequencing can provide comprehensive genomic information, but it remains comparatively resource-intensive for focused blood group analysis and directs most sequencing output toward genomic regions unrelated to the specific research question⁷. Targeted enrichment can improve the efficiency of such analyses by restricting sequencing to selected blood group loci⁷. One suitable enrichment strategy is end-point long-range PCR, in which defined genomic regions are amplified before sequencing-library preparation⁷⁻⁹.

Long-range PCR is technically demanding because reliable amplification of fragments exceeding 10 kb depends on template integrity, primer specificity, polymerase performance, reaction composition, and optimized thermocycling conditions^{8,9}. Primer specificity is particularly important for homologous loci such as RHD/RHCE and MNS, where primers may also bind related genomic regions and generate cross-amplified or off-target products^{7,8}. Consequently, the presence of a gel band at the expected approximate size provides evidence that a high-molecular-weight PCR product was generated, but does not independently confirm that the product originated from the intended genomic locus^{7,8}. Sequencing-based confirmation is therefore required to establish the genomic identity of long-range PCR products, especially when homologous genes or repetitive sequences are present^{7,8}.

The selected targets represented different levels of genomic complexity. ABO and KEL were included as clinically relevant loci for establishing and sequencing-confirming the workflow, whereas MNS-associated and RHD/RHCE-associated regions were included because their homologous gene structures and susceptibility to recombination, gene conversion, and hybrid allele formation create additional challenges for locus-specific amplification and sequence analysis^{2,7}.

The aim of this thesis was therefore to establish and evaluate a long-range PCR-based enrichment workflow for ABO, KEL, MNS-associated, and RHD/RHCE-associated regions, with downstream compatibility for Oxford Nanopore long-read sequencing. In this thesis, full-gene characterization refers to sequencing an entire gene, including its coding and intronic regions, whereas extended-locus characterization refers to sequencing a defined target interval that also includes relevant flanking or adjacent genomic regions. The long-range PCR step was used for targeted enrichment, and selected resulting PCR products were subsequently analyzed by Oxford Nanopore sequencing. PCR products from all selected target regions were

assessed by agarose gel electrophoresis, while selected ABO and KEL products were additionally confirmed by sequencing. MNS-associated and RHD/RHCE-associated PCR products were not sequenced because of time and cost constraints and therefore remained candidates for future sequence-based validation. The long-term purpose of the established workflow is to support haplotype phasing, discovery of novel variants, full-gene or extended-locus characterization, and a better understanding of the genetic mechanisms underlying blood group diversity.

2 Objective and aim of the study

The aim of this thesis was to establish and optimize a long-range PCR-based enrichment workflow for selected clinically relevant blood group loci, with downstream compatibility for Oxford Nanopore long-read sequencing. The investigated targets comprised the ABO and KEL genes as well as selected regions associated with the MNS and Rh blood group systems.

The study was designed as a method-development project focusing on target selection, primer design, long-range PCR optimization, polymerase comparison, gel-based assessment of PCR products, purification of selected products, and sequencing-based confirmation of selected amplicons. The project was not intended to constitute a complete clinical validation of molecular blood group typing or to replace established serological methods.

The selected loci represented different levels of genomic complexity. ABO and KEL were included as clinically relevant genes for the establishment and sequencing-based confirmation of the workflow. MNS-associated and RHD/RHCE-associated regions were included as technically more challenging targets because of the high sequence similarity between related genes and the potential occurrence of recombination, gene conversion, and hybrid alleles^{2,8,9}.

Long-range PCR products were initially evaluated by agarose gel electrophoresis. The presence of a distinct high-molecular-weight band at the expected approximate position was considered evidence that a PCR product of the anticipated size range had been generated. However, gel electrophoresis alone was not considered sufficient to confirm the genomic identity of a PCR product, particularly for homologous regions such as RHD/RHCE and the glycoporphin gene cluster associated with the MNS system.

Selected ABO and KEL PCR products were therefore analyzed by Oxford Nanopore sequencing and aligned to the human GRCh38 reference genome to confirm their genomic origin. MNS-associated and RHD/RHCE-associated PCR products were not sequenced because of time and cost constraints. These products were consequently assessed only at the gel level and cannot be regarded as validated target-specific amplicons until sequencing-based confirmation has been performed.

The established workflow is intended to provide a methodological basis for future sequencing of additional blood group loci, haplotype phasing, detection of novel variants, and full-gene or extended-locus characterization.

2.1 Specific Objectives

The specific objectives of this study were:

1. To identify clinically relevant genomic target regions within the ABO, KEL, MNS, and Rh blood group systems.
2. To design new long-range PCR primer pairs or adapt published primer sequences for the amplification of large genomic fragments from the selected loci.

3. To design overlapping amplicons where necessary to enable continuous coverage of the selected genes or genomic regions.
4. To evaluate the predicted specificity of candidate primer pairs in silico using the human GRCh38 reference genome and appropriate genome visualization and in silico PCR tools.
5. To compare commercially available DNA polymerases and PCR master mixes for their suitability for long-range amplification of the selected blood group targets.
6. To optimize long-range PCR conditions, including polymerase selection, reaction composition, additive concentrations, primer concentrations, genomic DNA input, thermocycling temperatures, extension times, and ramping conditions.
7. To assess the generation of high-molecular-weight PCR products by agarose gel electrophoresis based on band presence, approximate size, intensity, reproducibility, and the occurrence of visible non-specific products.
8. To purify and size-select selected PCR products for downstream Oxford Nanopore library preparation.
9. To sequence selected ABO and KEL PCR products using Oxford Nanopore technology and confirm their genomic identity by alignment to the human GRCh38 reference genome.
10. To establish a reproducible workflow that can be expanded in future studies to include sequencing-based validation of MNS-associated and RHD/RHCE-associated products, additional blood group loci, haplotype-resolved allele characterization, and novel-variant detection.

2.2 Target Blood Group Systems

The blood group loci investigated in this thesis were selected based on their clinical relevance, genomic structure, and suitability for evaluating different challenges associated with long-range PCR enrichment. ABO and KEL were included as clinically important genes with defined target regions suitable for workflow establishment and sequencing-based confirmation. The MNS and Rh systems were included because their genomic organization contains highly homologous genes and is associated with recombination, gene conversion, and hybrid allele formation^{2,8,9}.

Together, the selected loci provided targets of differing technical complexity for evaluating primer design, long-range amplification, and downstream Oxford Nanopore sequencing. Sequencing-based validation was completed for selected ABO and KEL products, whereas MNS-associated and RHD/RHCE-associated products remained at the gel-evaluation stage because of time and cost constraints.

2.2.1 ABO Blood Group System

The ABO blood group system is one of the most clinically important blood group systems in transfusion medicine and transplantation¹¹. The major ABO phenotypes are determined by

glycosyltransferase activity encoded by the ABO gene, which is located on chromosome 9q34.2^{11,12}. The A and B glycosyltransferases modify the H antigen by adding different terminal monosaccharides, whereas common O alleles generally encode an inactive or non-functional enzyme^{11,12}.

The ABO gene consists of seven exons, with exons 6 and 7 containing most of the coding sequence^{11,13}. Numerous variants affecting enzyme activity or substrate specificity are located within these exons and can result in altered, weak, or subgroup phenotypes^{13,14}. Additional variants located outside the routinely investigated coding positions may contribute to allele differentiation and haplotype characterization¹⁰.

ABO was selected as a representative target for the design of overlapping long-range PCR products and subsequent sequencing-based confirmation. Three overlapping amplicons were designed to cover the selected ABO target region and to provide overlap between adjacent fragments.

2.2.2 Rh Blood Group System: RHD and RHCE

The Rh blood group system is one of the most clinically significant blood group systems and includes the highly immunogenic D antigen as well as the C, c, E, and e antigens^{2,15}. The D antigen is encoded by RHD, whereas the C/c and E/e antigens are encoded by RHCE^{2,15}. RHD and RHCE are located in close proximity on chromosome 1 and share a high degree of sequence similarity^{8,16}.

The homologous structure of RHD and RHCE facilitates recombination, gene conversion, and the formation of hybrid alleles^{8,16}. These genetic mechanisms contribute to the formation of weak D, partial D, DEL, and other Rh variant alleles^{16,17}. Correct characterization of these variants is clinically relevant because individuals carrying certain partial or weak antigen variants may remain at risk of alloimmunization^{16,17}.

The high sequence similarity between RHD and RHCE complicates the identification of gene-specific primer-binding sites and increases the risk of cross-amplification⁸. RHD/RHCE-associated regions were therefore included as technically challenging targets for evaluating long-range PCR primer design. Six overlapping PCR products were designed to cover the selected Rh-associated regions. Because these products were not sequenced owing to time and cost constraints, their genomic identity remains to be confirmed in future work.

2.2.3 MNS Blood Group System

The MNS blood group system is primarily encoded by the glycophorin genes GYPA and GYPB, while the closely related GYPE gene also contributes to the genomic organization and evolution of the glycophorin gene cluster^{2,15}. These genes are located in a homologous cluster on chromosome 4 and share substantial sequence similarity^{2,18}. The system includes the M, N, S, and s antigens as well as numerous low-prevalence and hybrid antigens^{2,15}.

The homologous organization of the glyophorin genes promotes unequal recombination, gene conversion, and hybrid gene formation^{9,18}. Hybrid structures such as GYP(A-B-A) and GYP(B-A-B) can encode variant antigen profiles that may be difficult to resolve using serological methods or molecular assays targeting only individual polymorphic positions^{9,18}. Some glyophorin hybrid alleles are clinically relevant because the resulting antibodies can be associated with hemolytic transfusion reactions or hemolytic disease of the fetus and newborn^{19,20}.

The MNS-associated regions were included as technically challenging targets for long-range PCR enrichment because of their high degree of homology and potential structural complexity. Eight overlapping PCR products were designed across selected regions of the glyophorin gene cluster. These products were evaluated by agarose gel electrophoresis but were not sequenced because of time and cost constraints. Their genomic identity and specificity therefore require confirmation in future sequencing experiments.

2.2.4 Kell Blood Group System

The Kell blood group system is clinically important because several Kell antigens are highly immunogenic, particularly the K antigen^{2,15}. Antibodies directed against Kell antigens can cause hemolytic transfusion reactions and hemolytic disease of the fetus and newborn^{2,15}. The system is encoded by the KEL gene on chromosome 7, which produces a red blood cell membrane glycoprotein with zinc endopeptidase activity²¹.

The K/k antigen polymorphism is caused by a single nucleotide substitution that results in an amino acid change in the Kell glycoprotein^{2,21}. The KEL gene also contains numerous additional clinically relevant variants associated with low-prevalence antigens, weakened antigen expression, null phenotypes, and rare alleles². Molecular analysis beyond the common K/k polymorphism may therefore be required for comprehensive characterization of uncommon KEL alleles^{1,2}.

KEL was selected as a clinically relevant target for evaluating the ability of the optimized workflow to produce sequencing-compatible long-range PCR products. Two overlapping KEL target regions were amplified, and selected products were subsequently confirmed by Oxford Nanopore sequencing.

3 Materials and Methods

3.1 Study Design and General Workflow

This study was conducted as a method-development project to establish a long-range PCR-based enrichment workflow for selected blood group loci and to assess the compatibility of selected PCR products with Oxford Nanopore long-read sequencing.

The experimental workflow comprised target-region selection, primer design and *in silico* specificity assessment, primer synthesis and preparation, comparison and optimization of long-range PCR enzymes and reaction conditions, agarose gel electrophoretic evaluation of PCR products, bead-based cleanup and size selection, Oxford Nanopore library preparation and sequencing, and alignment-based confirmation of selected PCR products.

Primer pairs targeting ABO, KEL, MNS-associated, and RHD/RHCE-associated regions were tested using human genomic DNA as the template. PCR products were initially evaluated by agarose gel electrophoresis based on band presence, approximate size, intensity, and the occurrence of visible non-specific products. Selected ABO and KEL products were subsequently purified, size-selected, prepared for Oxford Nanopore sequencing, sequenced, and analyzed by alignment to the human reference genome.

MNS-associated and RHD/RHCE-associated products were evaluated by agarose gel electrophoresis but were not sequenced because of time and cost constraints. Consequently, gel electrophoresis was used only to document the generation of high-molecular-weight PCR products within the expected approximate size ranges and not to confirm their genomic identity.

3.2 Selection of Blood Group Target Regions

Genomic target regions were defined for the ABO and KEL genes and for selected MNS-associated and RHD/RHCE-associated regions using the Genome Reference Consortium Human Build 38 reference assembly (GRCh38)²². Gene annotations and genomic coordinates were inspected using the UCSC Genome Browser and the NCBI Genome Data Viewer^{23,24}.

For each locus, exon–intron boundaries, adjacent intronic and flanking sequences, and the positions of known clinically relevant regions were inspected. Target intervals were selected to cover the intended gene or genomic region while allowing long-range PCR primers to be placed in suitable flanking sequences.

The ABO and KEL target intervals were defined *de novo* using the corresponding GRCh38 gene annotations. The selected regions were divided into overlapping long-range PCR fragments to provide continuous coverage of the intended genomic intervals.

For the MNS-associated targets, the amplicon arrangement and target coverage described for long-read analysis of the glycoporphin gene cluster were considered when defining regions within GYPA, GYPB, and GYPE⁹. The published study was used as a reference for target-region selection, while the primer sequences used in the present study were evaluated and adjusted for the intended GRCh38 target intervals.⁹

For the RHD/RHCE-associated targets, the overlapping long-range PCR design and primer sequences described by Tounsi et al.⁸ were reviewed. Published primer sequences were retained where suitable or modified and redesigned where necessary to obtain the intended target coverage and improve compatibility with the experimental conditions used in this study⁸.

The detailed primer-design procedure is described in Section 3.3. Final primer sequences and expected amplicon sizes are provided in Appendix Section 9.1.

3.3 Primer Design Strategy

Primer design was performed manually using the GRCh38 human reference assembly and the target intervals defined in Section 3.2²². Candidate primer-binding regions were inspected using the UCSC Genome Browser, and predicted amplification products were evaluated using the UCSC In-Silico PCR tool²³.

The design process focused on obtaining overlapping long-range PCR products that covered the selected genomic intervals without leaving gaps between adjacent fragments. Forward and reverse primers were selected with similar calculated melting temperatures, preferably differing by no more than approximately 2–3 °C. Primer GC content was maintained within a moderate range where possible, and sequences with pronounced self-complementarity, complementarity to the paired primer, or predicted stable secondary structures were avoided. For primers containing universal 5' tails, these parameters were evaluated using only the gene-specific target-binding sequence; the tailing strategy is described separately in Section 3.5.

For ABO, three primer pairs were designed de novo to generate overlapping long-range PCR products across the selected target interval. The overlaps were included to provide continuous coverage of the region and redundant sequence information at the boundaries between adjacent products.

For KEL, two genomic target intervals were selected. Two closely related primer variants were tested for each interval, resulting in two alternative primer pairs for KEL Fragment 1 and two alternative primer pairs for KEL Fragment 2.

For the MNS-associated regions, eight primer pairs were designed to target selected intervals within the glycoporphin gene cluster. The target arrangement was informed by the long-read MNS approach described by Xu et al.⁹. Because GYPA, GYPB, and GYPE contain extensive

homologous sequence, candidate primers were evaluated for predicted binding to more than one glycoporphin gene and for the generation of unintended products⁹.

For the RHD/RHCE-associated regions, six primer pairs were selected to generate overlapping long-range PCR products. Primer sequences described by Tounsi et al.⁸ were retained where they were compatible with the selected target intervals, whereas other primers were modified or redesigned for the workflow established in this study. Because RHD and RHCE share a high degree of sequence similarity, predicted binding to both genes and potential cross-amplification were specifically assessed during primer selection⁸.

All candidate primer pairs were evaluated by *in silico* PCR against GRCh38. Primer pairs that produced predicted off-target products, multiple genomic products, or unintended amplification of homologous regions were excluded or redesigned. Only primer pairs with an acceptable predicted amplification profile were ordered for experimental testing. *In silico* analysis was used to assess predicted primer specificity; definitive confirmation of the genomic origin of the resulting PCR products required sequencing.

The final primer sequences and expected product sizes are provided in Appendix Section 9.1.

3.4 Primer Synthesis and Preparation

The final oligonucleotide primers were synthesized by Microsynth AG (Balgach, Switzerland) and supplied in lyophilized form. Each primer was reconstituted in nuclease-free water to a stock concentration of 100 μ M using the oligonucleotide amount and resuspension volume specified in the corresponding manufacturer's technical datasheet²⁵.

Following the addition of nuclease-free water, the primer solutions were incubated briefly at 65 °C, vortexed, and centrifuged to ensure complete dissolution and collect the solution at the bottom of the tube.

A 10 μ M working solution was prepared from each 100 μ M stock solution using nuclease-free water. The concentrated stock solutions were stored at –20 °C, while the 10 μ M working solutions were stored at 4 °C for short-term use. The working solutions were used for PCR preparation to limit repeated handling and freeze–thawing of the concentrated stock solutions.

The complete primer sequences, primer labels, and expected amplicon sizes are provided in Tables A1–A4 in Appendix Section 9.1. Simplified fragment labels are used in these tables instead of the original internal laboratory designations.

3.5 Primer Tails for Oxford Nanopore PCR Barcoding

Primer pairs intended for Oxford Nanopore PCR barcoding were synthesized with universal sequences at their 5' ends, as specified in the Oxford Nanopore Technologies dual-barcoding protocol²⁶. The universal forward-primer tail was 5'-TTTCTGTTGGTGCTGATATTGC-3',

and the universal reverse-primer tail was 5'-ACTTGCCTGTCGCTCTATCTTC-3'²⁶. These sequences are required for amplicon input in the dual-barcoding workflow using the PCR Barcoding Expansion 1–96 (EXP-PBC096) and Native Barcoding Kit 24 V14 (SQK-NBD114.24)²⁶.

Each tailed primer consisted of the universal 5' sequence followed by the locus-specific primer sequence at the 3' end. During the first amplification cycles, the locus-specific 3' portion annealed to the genomic target, whereas the universal 5' tail did not participate in genomic target recognition. Primer melting temperature, GC content, predicted secondary structure, and genomic specificity were therefore assessed using the locus-specific portion of each primer.

After first-round long-range PCR and product cleanup, the incorporated universal tails served as binding sites for the barcode primers supplied with EXP-PBC096. This PCR-barcoding step incorporated a sample-specific PCR barcode, referred to as the inner barcode, into each amplicon²⁶. During subsequent library preparation, a native barcode from SQK-NBD114.24 was ligated to each pool as a second, outer barcode²⁶. The sequential incorporation of PCR and native barcodes enabled dual-barcoded sample identification following sequencing²⁶.

The complete primer sequences are provided in Tables A1–A4 in Appendix Section 9.1. For tailed primers, the universal 5' sequence and the locus-specific 3' sequence should be displayed separately within the primer tables to make their respective functions unambiguous.

3.6 Genomic DNA Templates

Two sources of human genomic DNA were used in this study: genomic DNA extracted from a buccal swab collected from the author and commercially available Human Genomic DNA (Cat. No. G3041; Promega Corporation, Madison, WI, USA). No biological material from third-party donors was included. The DNA templates were used for method development, primer testing, and PCR optimization and were not selected based on a known blood group phenotype or genotype.

Genomic DNA from the buccal swab was extracted using the Maxwell® CSC Instrument together with the Maxwell® CSC Genomic DNA Kit (Cat. No. AS1850; Promega Corporation, Madison, WI, USA). Extraction was performed according to the manufacturer's technical manual²⁷.

DNA concentration was measured using a DeNovix DS-11 spectrophotometer (DeNovix Inc., Wilmington, DE, USA). Before PCR preparation, the DNA samples were diluted with nuclease-free water to a working concentration of 100 ng/μL. For the standard 15 μL primer-screening reactions, 1 μL of genomic DNA was added, corresponding to a total DNA input of 100 ng per reaction.

Commercial Human Genomic DNA (Cat. No. G3041; Promega Corporation, Madison, WI, USA) was handled according to the manufacturer's product information²⁸. The commercially

available DNA was used as an additional standardized human genomic DNA template during primer testing and PCR optimization.

3.7 Polymerase Selection for Long-Range PCR

Several commercially available DNA polymerases and PCR master mixes were evaluated for long-range amplification of the selected blood group target regions. The tested products were GoTaq® Long PCR Master Mix (Promega Corporation, Madison, WI, USA), Q5® High-Fidelity DNA Polymerase and LongAmp® Taq DNA Polymerase (New England Biolabs, Ipswich, MA, USA), and PrimeSTAR® GXL DNA Polymerase, PrimeSTAR® GXL Premix Fast, and PrimeSTAR® LongSeq DNA Polymerase (Takara Bio Inc., Shiga, Japan).

GoTaq® Long PCR Master Mix was included because it had previously been used for in-house long-range PCR experiments. The master mix contains an optimized hot-start polymerase blend, reaction buffer, magnesium chloride, and deoxynucleotide triphosphates and is intended for amplification of long genomic DNA targets²⁹.

Q5® High-Fidelity DNA Polymerase was evaluated because of its high replication fidelity and its suitability for amplification of long or difficult DNA targets³⁰. LongAmp® Taq DNA Polymerase was included because it is formulated for long-range PCR and combines Taq DNA polymerase with a proofreading polymerase³¹. Selected Q5® and LongAmp® reactions were additionally tested with Q5® High GC Enhancer during optimization.

PrimeSTAR® GXL DNA Polymerase was evaluated as a high-fidelity enzyme developed for long and difficult genomic templates³². PrimeSTAR® GXL Premix Fast was tested as a separate ready-to-use 2× formulation containing PrimeSTAR® GXL DNA Polymerase, reaction components, and a density reagent for accelerated PCR setup and processing³³. PrimeSTAR® LongSeq DNA Polymerase was included because it was developed for amplification of long and difficult genomic targets intended for downstream long-read sequencing³⁴.

Each enzyme or master mix was initially used according to the corresponding manufacturer's recommended reaction composition and thermocycling conditions. Reaction conditions were subsequently adjusted during optimization, including changes to annealing temperature, extension time, additives, and temperature-ramping conditions.

Polymerase performance was evaluated by agarose gel electrophoresis. The evaluation criteria were the presence of a high-molecular-weight product at the expected approximate position, band intensity, reproducibility between repeated reactions, smearing, the presence of additional visible products, and total thermocycling time. Agarose gel electrophoresis was used to compare product formation but was not considered sufficient to confirm the genomic identity or locus specificity of the resulting PCR products.

Following the comparative experiments, PrimeSTAR® GXL Premix Fast provided the most suitable combination of high-molecular-weight product formation, reproducibility, and PCR

runtime under the conditions tested in this study. It was therefore selected for the standard primer-screening reactions and for the preparation of PCR products intended for downstream Oxford Nanopore sequencing.

3.8 PCR Reaction Setup

Initial primer-screening reactions were performed in a total volume of 15 μL using PrimeSTAR® GXL Premix Fast (Takara Bio Inc., Shiga, Japan). The standard reaction composition used in this study is shown in Table 3.1. PrimeSTAR® GXL Premix Fast is supplied as a 2 $\times$ PCR master mix and was used at a final concentration of 1 $\times$ ³³.

Table 3.1. Standard reaction composition for long-range PCR using PrimeSTAR® GXL Premix Fast.

Component	Volume per reaction
PrimeSTAR® GXL Premix Fast	7.5 μL
Betaine, 5 M	1.0 μL
Nuclease-free water	4.0 μL
Forward primer, 10 μM	0.75 μL
Reverse primer, 10 μM	0.75 μL
Genomic DNA, 100 ng/ μL	1.0 μL
Total volume	15.0 μL

Each standard screening reaction contained 0.5 μM of each primer, 0.33 M betaine, and 100 ng of genomic DNA. Betaine was included because it has been shown to improve PCR amplification of GC-rich DNA sequences³⁵. Betaine can reduce sequence-dependent differences in DNA melting behavior and thereby facilitate amplification of templates containing GC-rich regions or stable secondary structures³⁵.

For reactions intended for downstream Oxford Nanopore sequencing, the total reaction volume was increased proportionally to generate sufficient PCR product for bead-based cleanup, PCR barcoding, and library preparation. The relative proportions and final concentrations of the reaction components were maintained unless otherwise stated.

PCR mixtures were assembled on ice before thermocycling. PrimeSTAR® GXL Premix Fast uses an antibody-mediated hot-start mechanism that suppresses polymerase activity during reaction preparation and becomes active during the initial high-temperature incubation³³. No-template controls were prepared by replacing the genomic DNA with nuclease-free water and were included to monitor possible contamination.

3.9 PCR Thermocycling Conditions

PCR amplification was performed using an Eppendorf Mastercycler® X50 thermal cycler (Eppendorf SE, Hamburg, Germany). The final PrimeSTAR® GXL Premix Fast workflow used a two-step PCR protocol in which primer annealing and extension were combined within a single temperature step. PrimeSTAR® GXL Premix Fast is compatible with two-step PCR using 68 °C as the combined annealing and extension temperature³³. The thermocycling profile used in this study is shown in Table 3.2.

Table 3.2. Thermocycling profile used for long-range PCR with PrimeSTAR® GXL Premix Fast.

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	1 min	1
Denaturation	98 °C	10 sec	30
Annealing/extension	68 °C	4–10 min	30
Final extension	68 °C	5–10 min	1
Hold	4 °C	∞	—

The annealing and extension time was adjusted between 4 and 10 min according to the expected amplicon length and the purpose of the reaction. Shorter extension times were used during initial primer screening, whereas longer extension times were used for the largest target regions and for reactions intended for downstream sequencing preparation. The duration of the final extension step was adjusted correspondingly between 5 and 10 min.

During preliminary PCR optimization, temperature-gradient reactions were performed to evaluate the effect of the combined annealing and extension temperature on product formation. Reduced ramping during the transition from denaturation to the annealing and extension step was also evaluated. The resulting PCR products were compared by agarose gel electrophoresis based on band presence, approximate size, intensity, and the occurrence of additional visible products. Following optimization, 68 °C was retained as the combined annealing and extension temperature, and the standard ramp settings of the Mastercycler® X50 were used for the routine workflow.

The use of 68 °C for the extension step is specified in the official PrimeSTAR® GXL Premix Fast documentation³³.

3.10 Agarose Gel Electrophoresis and Visualization

PCR products generated during polymerase comparison, primer screening, and PCR optimization were analyzed by agarose gel electrophoresis. Gel images were evaluated for the presence and migration of DNA products, approximate product size relative to a DNA marker, band intensity, smearing, and the occurrence of additional visible bands. Agarose gel electrophoresis was not considered sufficient to establish the genomic identity or locus specificity of a PCR product.

Agarose gels were prepared at a concentration of 1.0% (w/v) by dissolving 4.0 g of Biozym LE Agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany; Lot No. 0001273523; EEO: –mr 0.09–0.13) in 400 mL of 1× TBE buffer. The running buffer was prepared by diluting Invitrogen™ UltraPure™ 10× TBE Buffer (Cat. No. 15581-044; Thermo Fisher Scientific, Waltham, MA, USA) with distilled water to a final concentration of 1×.

After the molten agarose had cooled to approximately 60 °C, 20 µL of SYBR™ Safe DNA Gel Stain (Cat. No. S33102; Invitrogen™, Thermo Fisher Scientific, Waltham, MA, USA) was added directly to the 400 mL gel solution. The solution was mixed, poured into a horizontal gel-casting tray containing the appropriate combs, and allowed to solidify at room temperature. Electrophoresis was performed using 1× TBE as the running buffer.

PCR products were loaded together with the Quantitas Pro DNA Marker 100 bp–10 kb (Art. No. 617121; Biozym Scientific GmbH, Hessisch Oldendorf, Germany). The marker contains 13 DNA fragments ranging from 100 bp to 10 kb, with enhanced reference bands at 1 and 3 kb³⁶.

Electrophoresis was performed at 100 V for 90–120 min. The run duration was adjusted according to the expected product size and the degree of separation required between high-molecular-weight products.

Following electrophoresis, the gels were visualized under ultraviolet illumination, and digital images were recorded. Gel images were evaluated visually for band presence, migration relative to the DNA marker, band intensity, smearing, and additional visible products.

Because most expected PCR products were larger than the 10 kb upper limit of the DNA marker, products migrating above the highest marker band could only be classified as high-molecular-weight products consistent with the expected approximate size range. Exact fragment sizes and genomic identities could not be determined by agarose gel electrophoresis alone.

Representative gel images for the ABO, KEL, RHD/RHCE-associated, and MNS-associated reactions are presented in Figures 4.1–4.4 in Section 4.2.

3.11 PCR Product Cleanup and Size Selection

PCR products selected for downstream Oxford Nanopore sequencing were purified and size-selected using MAGFLO™ NGS magnetic beads (INTEGRA Biosciences AG, Zizers, Switzerland) in a 96-well plate format. MAGFLO™ NGS magnetic beads are intended for PCR-product purification and single- or double-sided nucleic-acid size selection, with fragment retention influenced by the bead-to-sample ratio³⁷.

A single-sided cleanup was performed using a bead-to-sample ratio of 0.4×. The required bead volume was calculated by multiplying the PCR-product volume by 0.4. This comparatively low bead ratio was selected experimentally to favor recovery of the long-range PCR products while reducing retention of shorter DNA fragments. Lower bead-to-sample ratios preferentially enrich larger DNA fragments, whereas increasing the bead ratio increases the retention of shorter fragments³⁷.

Before use, the magnetic beads were brought to room temperature and mixed thoroughly to resuspend the particles. The calculated bead volume was added to each PCR product, and the samples were mixed until homogeneous. The mixtures were incubated at room temperature for 5 min and subsequently placed on a magnetic stand until the solution had cleared and a bead pellet had formed.

The supernatant was removed without disturbing the bead pellet. While the plate remained on the magnetic stand, the bead-bound DNA was washed twice with freshly prepared 80% ethanol. Residual ethanol was removed, and the beads were briefly air-dried while avoiding excessive drying. The DNA was subsequently eluted in nuclease-free water, and the cleared eluate was transferred to a new plate after magnetic separation.

The cleanup step was performed after the initial long-range PCR and before PCR barcoding and library preparation. Purification at this stage removes residual primers, deoxynucleotide triphosphates, salts, enzymes, and other reaction components that may interfere with subsequent barcoding and library-preparation reactions²⁶. The Oxford Nanopore dual-barcoding protocol specifically requires purification after the initial PCR used to incorporate the universal primer tails and before amplification with the PCR barcodes²⁶.

The 0.4× cleanup was used as an enrichment step for long DNA products rather than as a method for determining their exact fragment sizes. Purified products were subsequently processed using the dual-barcoding and ligation-based Oxford Nanopore library-preparation workflow described in Section 3.12.

3.12 Oxford Nanopore Amplicon Sequencing

Selected purified long-range PCR products from the ABO and KEL loci were prepared for Oxford Nanopore sequencing using the Ligation Sequencing DNA V14 – dual barcoding protocol with the PCR Barcoding Expansion 1–96 (EXP-PBC096) and Native Barcoding Kit 24 V14 (SQK-NBD114.24)²⁶. The dual-barcoding protocol is intended for genomic or

amplicon DNA, requires PCR-based and native barcoding steps, and is compatible with R10.4.1 MinION/GridION flow cells (FLO-MIN114)²⁶.

Purified first-round PCR products containing the universal 5' primer tails described in Section 3.5 were used as input for library preparation. The protocol specifies an input of 100 ng of each tailed first-round PCR product in a volume of 45 μ L²⁶. The purified PCR products were adjusted to the required DNA input before the barcoding procedure.

The purified products were subjected to the initial end-preparation and barcode-adaptor ligation steps specified in the protocol²⁶. Sample-specific PCR barcodes from EXP-PBC096 were subsequently incorporated by amplification using the universal primer-tail sequences as binding sites²⁶. The PCR-barcoded products were purified, quantified, and pooled as specified by the manufacturer²⁶.

Following pooling, the DNA underwent a second end-preparation step before ligation of the native barcodes supplied with SQK-NBD114.24²⁶. The native-barcoded pools were combined, sequencing adapters were ligated, and the final library was purified and eluted according to the manufacturer's protocol²⁶. The combination of an internal PCR barcode and an external native barcode provided two levels of sample identification within the multiplexed sequencing library²⁶.

The final library was primed and loaded onto an R10.4.1 MinION flow cell (FLO-MIN114) according to the manufacturer's loading instructions²⁶. Sequencing was performed using an Oxford Nanopore Technologies MinION Mk1B device. The MinION Mk1B is compatible with R10.4.1 MinION flow cells and is operated using the MinKNOW software³⁸.

Sequencing setup, device control, data acquisition, and real-time run monitoring were performed using MinKNOW³⁹. MinKNOW was used to monitor the number of available pores, sequencing activity, accumulated output, and read-quality metrics during the sequencing run³⁹. Basecalling, demultiplexing, and downstream alignment are described separately in Section 3.13.

Within the scope of this study, Oxford Nanopore sequencing was performed for selected ABO and KEL long-range PCR products. MNS-associated and RHD/RHCE-associated PCR products were not sequenced because of time and cost constraints.

The official dual-barcoding protocol confirms the combination of EXP-PBC096, SQK-NBD114.24, and FLO-MIN114, as well as the sequential PCR-barcoding, native-barcoding, adapter-ligation, and flow-cell-loading steps.²⁶ MinKNOW controls sequencing acquisition and provides real-time sequencing and run information.³⁹ The MinION Mk1B documentation confirms compatibility with MinION flow cells, including FLO-MIN114.³⁸

3.13 Sequencing Data Processing and Amplicon Confirmation

Sequencing device control, nanopore signal acquisition, basecalling, and barcode demultiplexing were performed using MinKNOW software provided by Oxford Nanopore Technologies³⁹. MinKNOW converts the recorded nanopore signal into nucleotide sequences and supports barcode demultiplexing during or after the sequencing run³⁹. The resulting basecalled and demultiplexed reads were exported in FASTQ format for subsequent analysis.

The demultiplexed reads were imported into the EPI2ME analysis platform provided by Oxford Nanopore Technologies. Reference-based alignment and visualization were performed in EPI2ME using the Genome Reference Consortium Human Build 38 assembly as the reference sequence^{22,40}. EPI2ME workflows support the alignment of Oxford Nanopore reads to a supplied reference sequence and the generation of mapping and coverage information⁴⁰.

The analysis was restricted to confirming the genomic origin and coverage of the selected ABO and KEL long-range PCR products. Amplicon identity was assessed by determining whether reads assigned to each PCR product aligned to the intended genomic locus and covered the expected primer-defined interval.

For ABO, reads obtained from the three long-range PCR products were inspected for alignment to their intended ABO target regions. Coverage across the three individual fragments and the overlaps between adjacent fragments were evaluated to determine whether the selected ABO region was represented continuously.

For KEL, reads obtained from the selected long-range PCR products were inspected for alignment to the intended KEL target regions. The two fragments were assessed independently based on their genomic alignment positions and the correspondence between the observed read coverage and the expected primer-defined intervals.

A PCR product was considered sequence-confirmed when reads assigned to that product mapped predominantly to the intended genomic locus and extended across the expected target interval. This alignment-based assessment was used to confirm the identity of the ABO and KEL PCR products but was not used for clinical blood group genotype interpretation, formal variant calling, allele assignment, or haplotype phasing.

MNS-associated and RHD/RHCE-associated PCR products were not sequenced because of time and cost constraints. Their genomic identity therefore remained unresolved.

3.14 Documentation and Evaluation Criteria

For each PCR experiment, the primer pair, target region, expected product size, DNA polymerase or PCR master mix, reaction composition, thermocycling conditions, genomic DNA input, gel-electrophoresis result, and subsequent processing steps were documented.

PCR-product formation was initially evaluated by agarose gel electrophoresis. Gel-based evaluation included the presence or absence of a visible DNA band, migration relative to the

DNA marker, correspondence with the expected approximate size range, band intensity, smearing, additional visible products, and reproducibility between repeated reactions. Agarose gel electrophoresis was not used to establish the genomic identity or locus specificity of a PCR product.

A reaction was classified as producing an expected-size high-molecular-weight PCR product when a distinct band migrated at a position consistent with the expected approximate product size. This classification indicated successful formation of a DNA product of the anticipated size range but did not demonstrate that the product originated from the intended genomic locus.

A PCR product was classified as sequence-confirmed when demultiplexed sequencing reads aligned predominantly to the intended GRCh38 target region and provided coverage across the expected primer-defined interval in the EPI2ME visualization. For overlapping PCR products, coverage across the expected overlap between adjacent target intervals was additionally evaluated.

The distinction between gel-detected high-molecular-weight PCR products and sequence-confirmed PCR products was maintained throughout the Results and Discussion. This distinction was particularly important for the homologous RHD/RHCE and glycophorin gene regions, where primers may bind to closely related genomic sequences and generate unintended products^{8,9}.

ABO and KEL products were evaluated using both gel electrophoresis and sequencing-based confirmation. MNS-associated and RHD/RHCE-associated products were evaluated only by agarose gel electrophoresis because they were not sequenced owing to time and cost constraints. Their genomic identity therefore remained unresolved.

4 Results

4.1 Polymerase Comparison and PCR Optimization

The DNA polymerases and PCR master mixes evaluated during workflow optimization differed in their ability to generate high-molecular-weight PCR products and in the total time required for amplification. The complete experimental comparison is summarized in Table 4.1.

Initial reactions using Q5® High-Fidelity DNA Polymerase and LongAmp® Taq DNA Polymerase did not produce distinct bands at the expected approximate high-molecular-weight positions. Supplementation with Q5® High GC Enhancer and adjustment of annealing temperature, extension time, and additive conditions did not result in reproducible product formation. These enzymes were therefore not pursued further in the established workflow (Table 4.1).

GoTaq® Long PCR Master Mix generated visible high-molecular-weight PCR products for some of the tested targets. However, the long extension periods required for these reactions resulted in complete thermocycling times of up to approximately 11 h. This runtime limited the suitability of the GoTaq-based protocol for repeated primer screening and optimization experiments (Table 4.1).

The PrimeSTAR®-based formulations produced the most consistent high-molecular-weight PCR products under the conditions tested. PrimeSTAR® LongSeq DNA Polymerase generated long-range products but did not provide a clear improvement over the PrimeSTAR® GXL-based reactions and required longer thermocycling times. PrimeSTAR® GXL Premix Fast produced reproducible bands at the expected approximate high-molecular-weight positions while reducing the total PCR runtime to approximately 4 h (Table 4.1).

Based on high-molecular-weight product formation, reproducibility, the occurrence of only limited additional visible products, and total PCR runtime, PrimeSTAR® GXL Premix Fast was selected for subsequent primer screening and preparation of PCR products for Oxford Nanopore sequencing (Table 4.1). Gel electrophoresis was used for this comparative assessment but did not establish the genomic identity of the observed products.

Table 4.1. Comparison of DNA polymerases and PCR master mixes evaluated during long-range PCR optimization. Manufacturer recommendations were obtained from the corresponding product documentation^{29–34}.

Polymerase / system	Manufacturer recommendation	Estimated extension per cycle (14 kb)	Practical result in this study
NEB Q5 High-Fidelity DNA Polymerase + Q5 High GC Enhancer	Usually 20–30 s/kb; for >6 kb often 40–50 s/kb	~4.7–11.7 min	No reliable target band
NEB LongAmp Taq DNA Polymerase + Q5 High GC Enhancer	50 s/kb	~11.7 min	No reliable target band
Promega GoTaq Long PCR Master Mix	~1 min/kb	~14 min	Worked, but up to ~11 h total runtime
Takara PrimeSTAR GXL DNA Polymerase	Standard 1 min/kb; for 10–30 kb: 10 min at 68 °C in 2-step PCR	~10–14 min	Best overall balance
Takara PrimeSTAR LongSeq DNA Polymerase	20–30 s/kb	~4.7–7 min	Worked, but not clearly better and longer in practice

4.2 Gel-Based Evaluation of Primer Performance

Following PCR optimization, the primer pairs targeting ABO, KEL, RHD/RHCE-associated regions, and MNS-associated regions were evaluated using PrimeSTAR® GXL Premix Fast. Product formation was assessed by agarose gel electrophoresis based on the presence of high-molecular-weight bands, migration relative to the DNA marker, band intensity, visible smearing, additional products, and reproducibility between repeated reactions. Because the DNA marker extended only to 10 kb, products larger than 10 kb could not be sized precisely and were evaluated only according to whether they migrated within the expected high-molecular-weight range.

4.2.1 ABO

The three ABO primer pairs produced distinct high-molecular-weight PCR products. The expected product sizes were approximately 14.1 kb for ABO Fragment 1, 13.2 kb for ABO Fragment 2, and 13.4 kb for ABO Fragment 3. All three products migrated above the 10 kb upper limit of the DNA marker, consistent with their expected approximate size range (Figure 4.1).

The gel result demonstrated the formation of three long PCR products suitable for further processing but did not independently establish their genomic identity. The ABO products were therefore subjected to Oxford Nanopore sequencing, and their alignment-based confirmation is presented in Section 4.3.

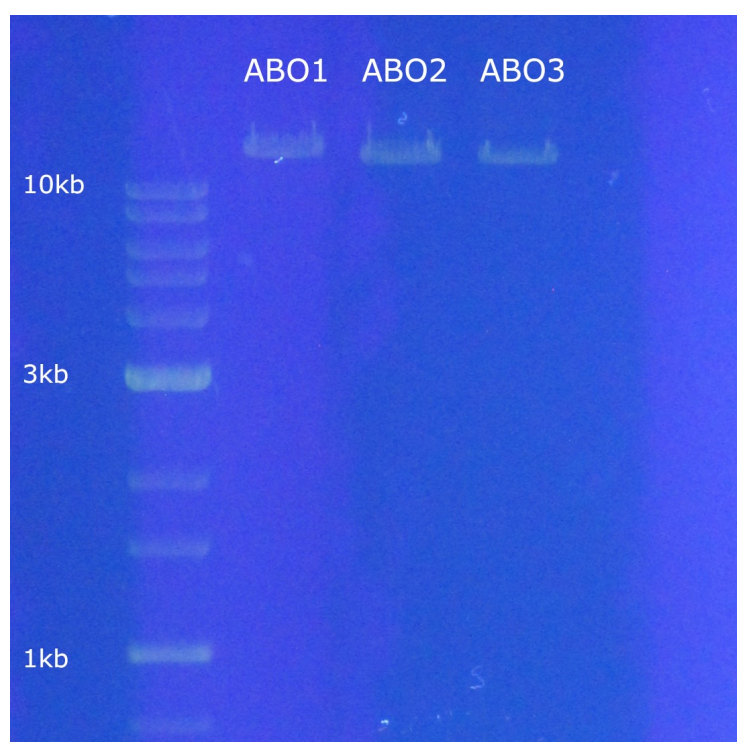


Figure 4.1. Agarose gel electrophoresis of ABO long-range PCR products. Products generated for ABO Fragment 1, Fragment 2, and Fragment 3 were separated on a 1.0% agarose gel. The expected product sizes were approximately 14.1, 13.2, and 13.4 kb, respectively. M, Quantitas Pro DNA Marker 100 bp–10 kb.

4.2.2 KEL

The primer pairs targeting the two KEL intervals also produced distinct high-molecular-weight PCR products. The expected product sizes were approximately 14.9 kb for KEL Fragment 1 and 13.9 kb for KEL Fragment 2. The observed bands migrated above the 10 kb marker band, consistent with the expected approximate size ranges of the two KEL products (Figure 4.2).

The tested primer variants generated products under the optimized PrimeSTAR® GXL Premix Fast conditions. As with the ABO products, gel electrophoresis demonstrated high-molecular-weight product formation but did not by itself confirm that the products originated from the intended KEL intervals. Selected KEL products were therefore analyzed by Oxford Nanopore sequencing, as described in Section 4.3.

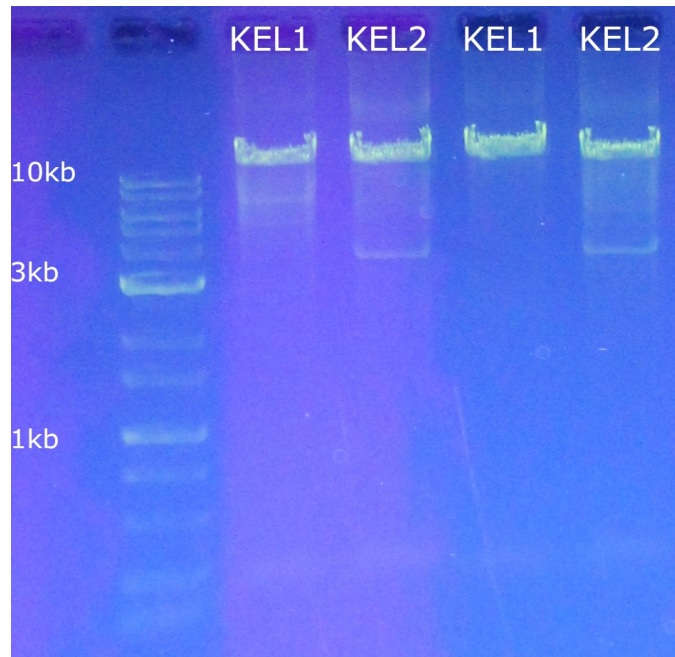


Figure 4.2. Agarose gel electrophoresis of KEL long-range PCR products. Products generated for the two KEL target intervals were separated on a 1.0% agarose gel. The expected product sizes were approximately 14.9 kb for Fragment 1 and 13.9 kb for Fragment 2. M, Quantitas Pro DNA Marker 100 bp–10 kb.

4.2.3 RHD/RHCE-associated regions

Six primer pairs targeting RHD/RHCE-associated regions produced visible high-molecular-weight PCR products. The expected product sizes were approximately 16.6 kb for Fragment 1, 16.9 kb for Fragment 2, 10.8 kb for Fragment 3, 9.9 kb for Fragment 4, 11.6 kb for Fragment 5, and 11.3 kb for Fragment 6. The observed products migrated near or above the upper region of the DNA marker, consistent with the expected approximate size ranges (Figure 4.3).

Because most of the expected products exceeded the 10 kb upper limit of the DNA marker, their exact sizes could not be determined by gel electrophoresis. Furthermore, because of the high sequence similarity between RHD and RHCE, the observed bands could not be assigned definitively to the intended genomic intervals without sequencing. The RHD/RHCE-associated products were not sequenced because of time and cost constraints, and their genomic identity therefore remained unresolved.

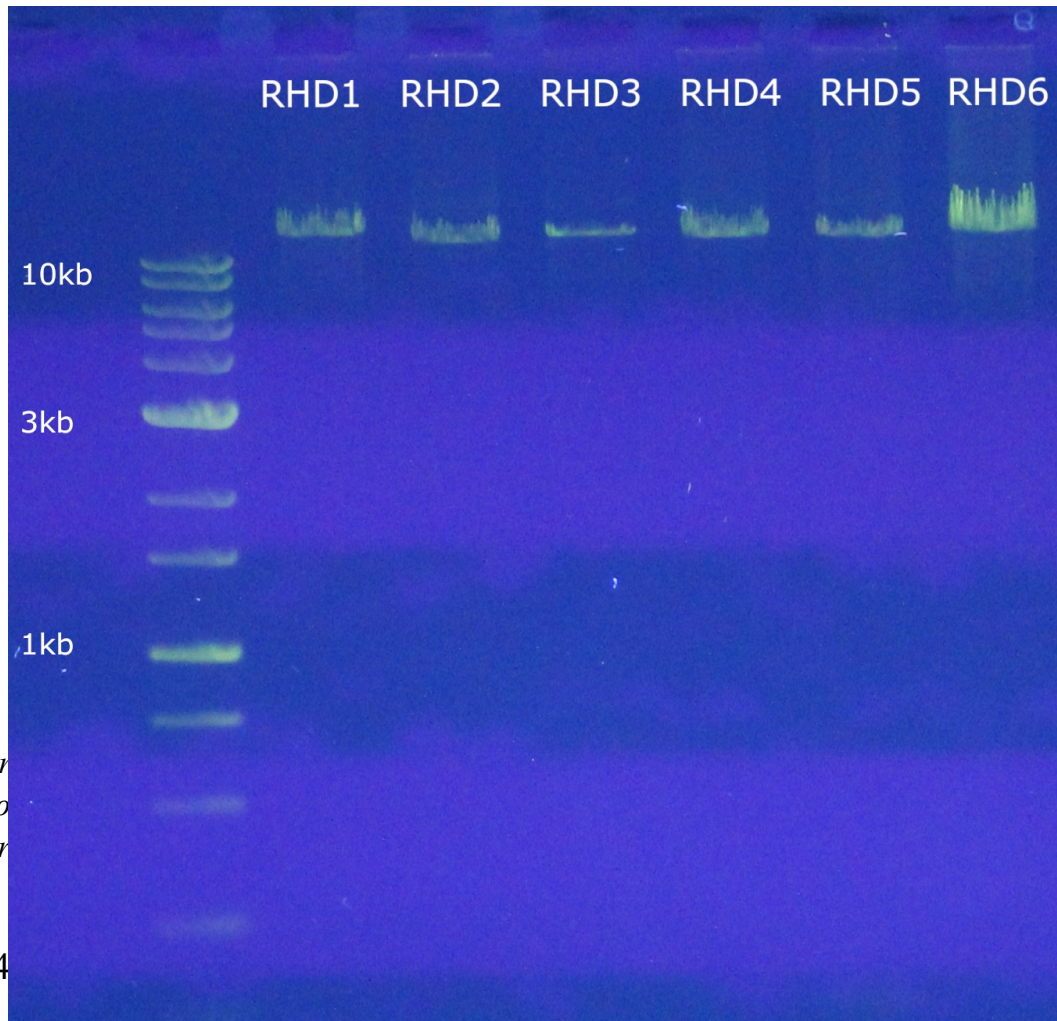
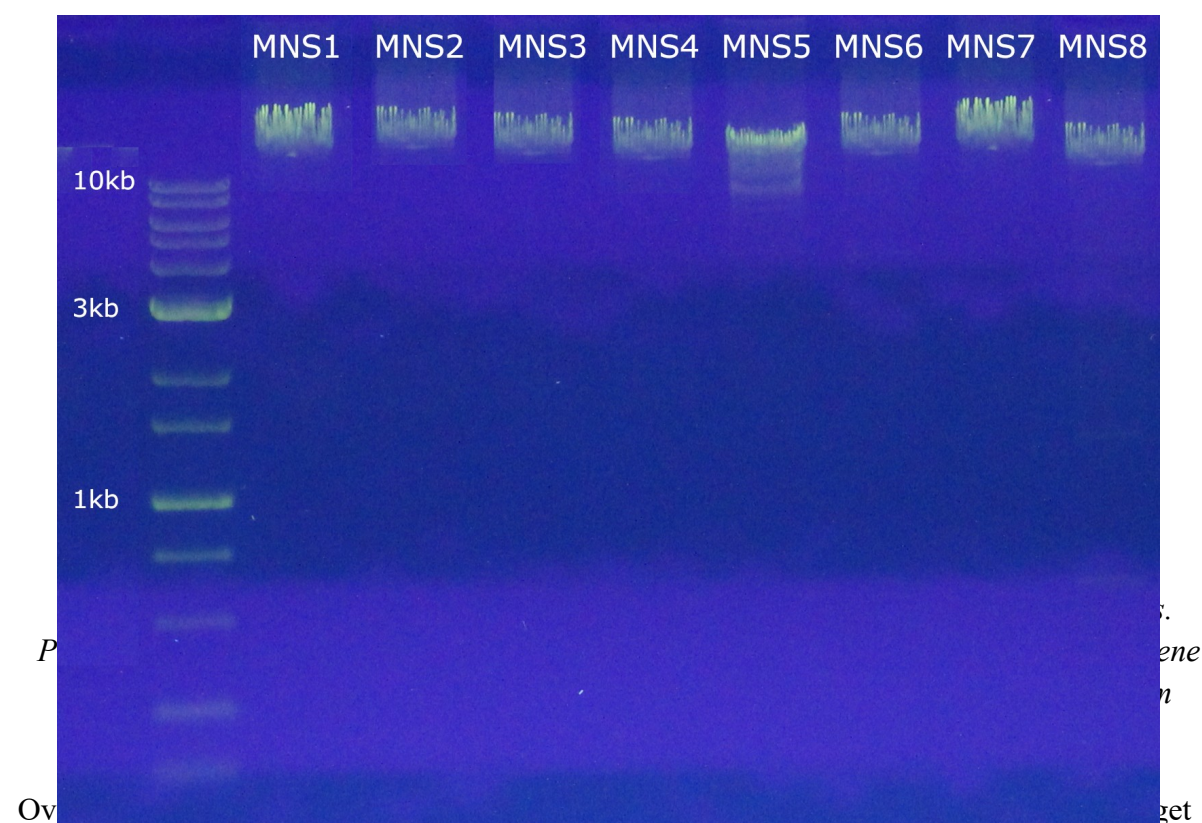


Figure 4.4. PCR products of the glycoprotein gene cluster. The products were separated by 9.9

4.2.4

Eight primer pairs targeting regions within the glycoprotein gene cluster produced visible high-molecular-weight PCR products. The expected product sizes were approximately 15.8 kb for Fragment 1, 14.1 kb for Fragment 2, 15.8 kb for Fragment 3, 11.4 kb for Fragment 4, 12.8 kb for Fragment 5, 13.2 kb for Fragment 6, 14.0 kb for Fragment 7, and 11.7 kb for Fragment 8. The products migrated near or above the upper region of the DNA marker, consistent with the expected approximate size ranges of the designed MNS-associated products (Figure 4.4).

Because the glycoprotein gene cluster contains highly homologous sequences within GYPA, GYPB, and GYPE, gel electrophoresis could not determine whether the observed products originated exclusively from the intended target intervals. The MNS-associated products were not sequenced because of time and cost constraints, and their genomic identity therefore remained unresolved.



groups. The ABO and KEL products selected for sequencing were subsequently confirmed by alignment to the intended genomic regions. In contrast, the gel-detected MNS-associated and RHD/RHCE-associated products were not sequence-confirmed and should therefore not be described as definitively validated target amplicons.

4.3 Confirmation of Amplicon Identity by Oxford Nanopore Sequencing

Selected ABO and KEL long-range PCR products were analyzed by Oxford Nanopore sequencing. Following basecalling and demultiplexing in MinKNOW, reads were aligned to GRCh38 and visualized using EPI2ME. Amplicon identity was evaluated by comparing the observed alignment positions with the expected primer-defined target intervals.

4.3.1 ABO sequencing results

For ABO, three distinct sequencing coverage profiles corresponded to Fragment 1, Fragment 2, and Fragment 3 (Figure 4.5). Each coverage profile aligned to its expected position within the selected ABO target region, confirming the genomic origin of the three PCR products (Figure 4.5).

Coverage from adjacent fragments extended across the expected overlapping intervals. No gap was visible between the three primer-defined target regions in the EPI2ME visualization, resulting in continuous coverage of the selected ABO region (Figure 4.5). The sequencing results therefore confirmed the three overlapping ABO long-range PCR products.

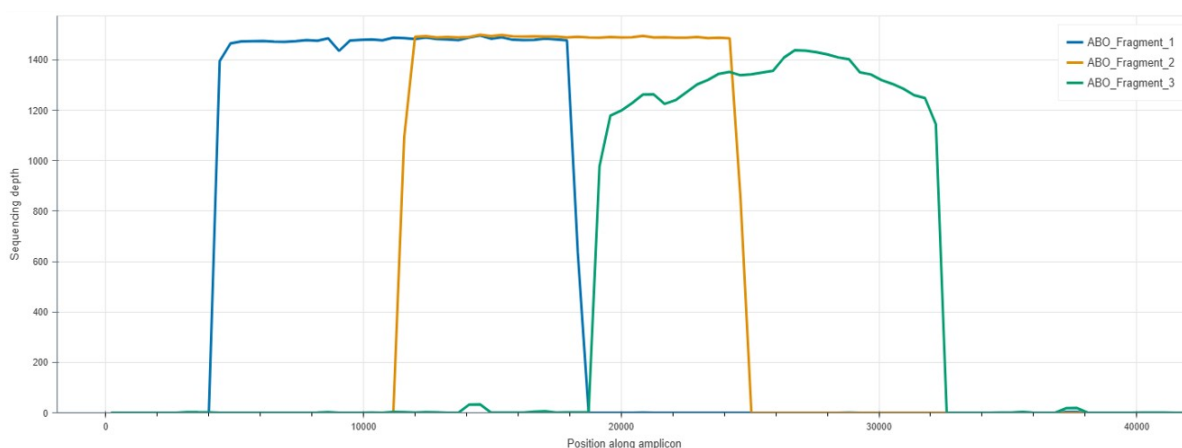


Figure 4.5. EPI2ME alignment and coverage visualization of the three overlapping ABO long-range PCR products.

Oxford Nanopore reads were aligned to the GRCh38 reference genome. The coverage profiles corresponding to ABO Fragment 1, Fragment 2, and Fragment 3 span the selected ABO target region and overlap at the boundaries between adjacent fragments.

4.3.2 KEL sequencing results

For KEL, two principal sequencing coverage profiles aligned to the expected target intervals for Fragment 1 and Fragment 2 (Figure 4.6). In the visualization, the coverage profile corresponding to KEL Fragment 1 was manually highlighted in yellow, while the profile corresponding to KEL Fragment 2 was manually highlighted in orange (Figure 4.6). The positions of both principal profiles agreed with the expected primer-defined KEL intervals, confirming the genomic origin of the selected KEL PCR products (Figure 4.6).

Additional smaller coverage signals were visible outside the two principal highlighted profiles (Figure 4.6). The origin of these additional signals could not be determined from the alignment visualization alone. They were therefore documented as additional sequencing signals but were not assigned to a specific PCR product or experimental cause.

Depth of coverage

Coverage along the individual amplicon, (use the dropdown menu to view the plots for the individual amplicons).

NG_007492_3_Homo_sapiens_Kell_metallo_endopeptidase_Kell_blood_group_KEL_RefSeqGene_LRG_799_on_chromosome_7 ▾

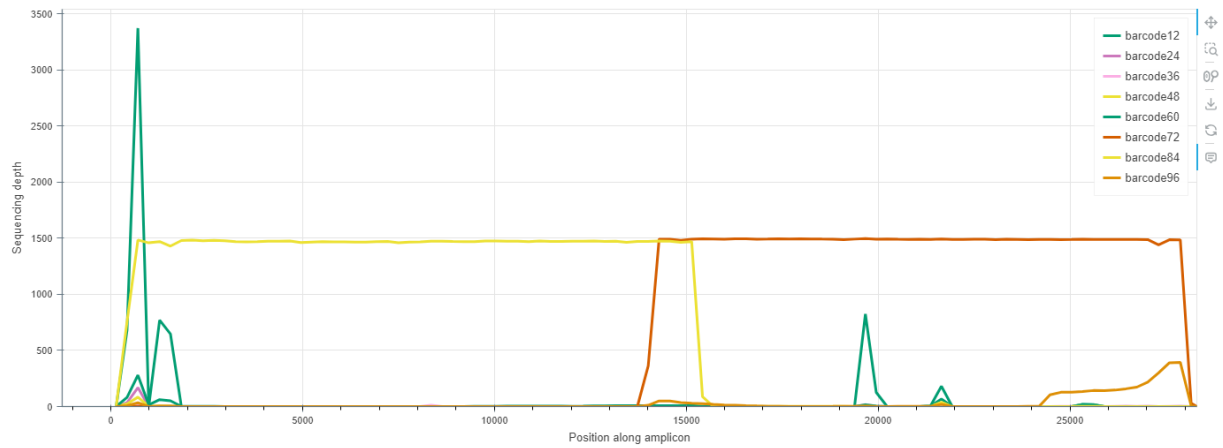


Figure 4.6. EPI2ME alignment and coverage visualization of the KEL long-range PCR products.

Oxford Nanopore reads were aligned to the GRCh38 reference genome. KEL Fragment 1 is highlighted in yellow and KEL Fragment 2 in orange. Additional smaller coverage signals are visible outside the two principal highlighted target intervals.

5 Discussion

5.1 Principal Findings and Scope of the Study

The aim of this thesis was to establish a long-range PCR-based enrichment workflow for selected clinically relevant blood group loci and to determine, through Oxford Nanopore sequencing and reference-based alignment, whether selected PCR products originated from the intended genomic target regions.

Agarose gel electrophoresis demonstrated the generation of high-molecular-weight PCR products for the ABO, KEL, RHD/RHCE-associated, and MNS-associated reactions. The bands were observed at positions consistent with the expected approximate size ranges of the respective PCR products, although most products exceeded the 10 kb upper limit of the DNA marker and could therefore not be sized precisely (Figures 4.1–4.4). Oxford Nanopore sequencing and subsequent reference-based alignment confirmed that the selected ABO and KEL products originated from the intended genomic intervals (Figures 4.5 and 4.6).

The MNS-associated and RHD/RHCE-associated products were not sequenced because of time and cost constraints. The formation of high-molecular-weight products in these reactions is therefore established by the gel results, but their genomic identity and target specificity remain unresolved (Figures 4.3 and 4.4).

The computational analysis comprised basecalling and barcode demultiplexing using MinKNOW, followed by alignment of the demultiplexed reads to GRCh38 and visualization of the resulting mapping and coverage profiles using EPI2ME. The analysis was sufficient to assess whether the principal ABO and KEL sequencing signals aligned to the expected genomic intervals, but it did not include consensus-sequence generation, systematic variant calling, haplotype phasing, allele assignment, or clinical genotype interpretation.

Consequently, the conclusions of this thesis are restricted to PCR-product formation, confirmation of the genomic origin of selected ABO and KEL products, and the technical establishment of the upstream enrichment and sequencing workflow. No conclusions can be drawn regarding the complete sequence, variant composition, haplotype structure, or blood group allele of the analyzed DNA templates.

This distinction is important because the principal advantage of long-read sequencing in blood group genetics is not merely the generation of long reads but the preservation of linkage between distant variants and the resolution of complex haplotypes, hybrid genes, and structural variants⁷. These downstream applications require validated target enrichment together with appropriate consensus generation, variant calling, phasing, and allele-assignment procedures⁷.

5.2 Long-Range PCR Optimization and Polymerase Selection

PrimeSTAR® GXL Premix Fast provided the most suitable combination of reproducible high-molecular-weight product formation, limited additional visible products, and a comparatively short thermocycling time under the conditions tested in this study (Table 4.1). The resulting protocol reduced the total PCR runtime to approximately 4 h, whereas the evaluated GoTaq® Long PCR workflow required up to approximately 11 h (Table 4.1). This reduction was practically important because numerous primer pairs and optimization conditions had to be evaluated.

The performance differences observed between the tested DNA polymerases and PCR master mixes are consistent with published comparisons showing that long-range PCR enzymes can differ substantially in their ability to amplify large genomic targets⁴¹. Jia et al. found that PrimeSTAR GXL amplified targets of different lengths and primer melting temperatures using a unified two-step protocol, whereas several alternative enzymes required additional target-specific optimization or did not generate the largest investigated products⁴¹. These findings support empirical polymerase selection using the relevant target regions rather than relying exclusively on the maximum product length stated by the manufacturer⁴¹.

Q5® High-Fidelity DNA Polymerase and LongAmp® Taq DNA Polymerase did not produce reproducible high-molecular-weight products under the conditions evaluated in this thesis (Table 4.1). This result does not demonstrate that the enzymes are generally unsuitable for long-range PCR, but it also does not establish that additional optimization would necessarily have produced the intended products. It only demonstrates that the tested conditions did not provide a sufficiently reliable workflow for the primer pairs and templates used here.

Betaine was included in the compared reactions and therefore cannot independently explain the superior performance of PrimeSTAR® GXL Premix Fast. Betaine can facilitate amplification of GC-rich templates by reducing sequence-dependent differences in DNA melting and the formation of stable secondary structures³⁵. However, because betaine was not evaluated in a separate controlled comparison, its individual contribution to product formation could not be determined.

Similarly, the effects of the polymerase formulation, reaction buffer, primer concentration, DNA input, extension time, and cycling conditions were not separated in a factorial experimental design. The selected protocol should therefore be interpreted as a functional combination of reaction components rather than as evidence that a single component was solely responsible for the observed improvement.

The PrimeSTAR® GXL Premix Fast protocol provides a practical starting point for further work with the investigated primer pairs. Nevertheless, transfer of the workflow to additional templates, alleles, instruments, or laboratories would require renewed verification of product formation, reproducibility, and target specificity.

5.3 Primer Design and the Overlapping-Amplicon Strategy

The three-overlapping-amplicon design developed for ABO generated three high-molecular-weight PCR products whose genomic origins were subsequently confirmed by Oxford Nanopore sequencing (Figures 4.1 and 4.5). The mapped-read profiles occupied the expected ABO target intervals and overlapped at the boundaries between adjacent fragments, providing continuous representation of the selected genomic region in the EPI2ME visualization (Figure 4.5).

The use of overlapping long-range PCR products is consistent with published approaches for extended blood group gene sequencing^{7,10}. Overlapping products provide redundant sequence information at fragment boundaries and can facilitate the reconstruction of larger genomic regions from multiple separately amplified products^{7,10}.

The three-fragment ABO arrangement is particularly comparable to published long-read approaches in which overlapping long-range PCR products were used to generate extended or full-length ABO sequences¹⁰. Extended ABO analysis can detect intronic and regulatory variation in addition to the major coding variants investigated by many targeted genotyping assays^{10,42}.

Nevertheless, overlapping PCR products do not automatically produce a completely phased haplotype. Connecting phase information across separately amplified fragments requires informative heterozygous variants within the overlapping regions, sufficient sequence quality, adequate read depth, and representation of both parental alleles^{7,43}. Because consensus generation and haplotype phasing were not performed in this thesis, the ABO result demonstrates overlapping locus coverage but not the reconstruction of complete ABO haplotypes.

The design of primers for RHD/RHCE and the glycophorin gene cluster is substantially more challenging because the respective loci contain highly homologous genes and recurrent structural variation⁷⁻⁹. Primers predicted to be specific *in silico* may still amplify a homologous genomic region or may fail to amplify alleles containing variation at a primer-binding site⁷. PCR-based enrichment may consequently produce cross-amplification, allelic imbalance, or allelic dropout in genetically diverse or structurally altered samples⁷.

The RHD/RHCE-associated and MNS-associated primer pairs therefore require sequencing-based evaluation using multiple genetically characterized samples. Only such an analysis can determine whether the primer pairs consistently amplify the intended locus, whether closely related paralogous regions are co-amplified, and whether relevant allele classes are represented.

5.4 Interpretation of the ABO and KEL Sequencing Results

Oxford Nanopore sequencing provided evidence of genomic origin that could not be obtained from agarose gel electrophoresis alone. While the gel results demonstrated that

high-molecular-weight DNA products had been generated, reference-based alignment was required to determine whether the sequenced products corresponded to the intended ABO and KEL target regions.

For ABO, three principal read-alignment profiles were observed at the expected genomic positions for Fragment 1, Fragment 2, and Fragment 3 (Figure 4.5). The profiles extended across the expected primer-defined intervals and overlapped at the boundaries between adjacent fragments, as displayed along the genomic-coordinate axis in Figure 4.5. These results confirm the genomic origin of the three principal ABO PCR products and demonstrate that the primer design, PCR conditions, clean-up procedure, barcoding workflow, sequencing, and alignment analysis collectively generated reads from the intended ABO region.

Published long-read ABO studies have shown that extended sequence information can support the characterization of complete allelic backgrounds and variants outside commonly examined coding positions¹⁰. Regulatory variation can also contribute to altered ABO phenotypes, as demonstrated by an intronic RUNX1-binding-site variant associated with an A₃ phenotype⁴². These observations illustrate the potential value of the extended target region generated in the present study^{10,42}.

However, the present analysis did not generate a polished consensus sequence or identify and phase sequence variants. The ABO results therefore confirm target identity and overlapping genomic representation but do not establish the complete sequence or allele composition of the analyzed DNA template.

For KEL, the two principal read-alignment profiles corresponded to the expected target intervals for Fragment 1 and Fragment 2 (Figure 4.6). The principal profile for Fragment 1 was highlighted in yellow and the profile for Fragment 2 in orange in the EPI2ME visualization (Figure 4.6). Their genomic positions support the conclusion that the selected KEL PCR products originated from the intended KEL regions (Figure 4.6).

Extended KEL sequencing may be valuable because clinically relevant variation is distributed throughout the gene and includes rare, weakened, and null alleles that may not be represented in predefined genotyping panels^{1,15}. A recently described Oxford Nanopore workflow used overlapping long-range PCR products to investigate KEL together with additional blood group genes and demonstrated reliable variant detection in characterized samples⁴³. The KEL findings in the present thesis are consistent with the technical feasibility of this general strategy, although the current analysis was restricted to alignment and target-region visualization rather than variant-level validation.

Additional smaller sequencing signals were visible outside the two principal KEL profiles (Figure 4.6). Possible explanations include shorter non-specific PCR products, residual DNA fragments after insufficient clean-up, barcode misassignment, or alignment artefacts. The available data do not distinguish among these possibilities, and none of them can therefore be presented as a confirmed cause.

Future evaluation should include fragment-size assessment after clean-up, per-barcode read counts, mapping-quality filtering, quantitative on-target read assessment, and inspection of the genomic positions represented by the minor signals. Comparable quality-control measures are necessary when developing long-read workflows for reproducible blood group variant analysis⁴³.

Overall, the ABO and KEL sequencing results confirm the genomic origin of the principal selected products and demonstrate that they could be processed through the established Oxford Nanopore workflow. They do not establish analytical validity for variant calling, allele assignment, or blood group phenotype prediction.

5.5 Interpretation of the RHD/RHCE and MNS-Associated Results

The RHD/RHCE-associated and MNS-associated reactions generated high-molecular-weight PCR products at positions consistent with their expected approximate size ranges (Figures 4.3 and 4.4). The generation of these products is an experimental finding rather than preliminary evidence. What remains unresolved is whether the products originated specifically and exclusively from the intended genomic target intervals.

The products were not sequenced because of time and cost constraints. Consequently, their genomic identity, locus specificity, and complete target coverage could not be assessed. This distinction is particularly important for RHD/RHCE and the glycophorin gene cluster because both regions contain highly homologous genes and frequent structural rearrangements⁷.

RHD and RHCE are closely related duplicated genes in which recombination, gene conversion, deletion, duplication, and hybrid-allele formation contribute to clinically relevant diversity^{8,16,44}. Oxford Nanopore-based RHD genotyping has previously been performed using six overlapping long-range PCR products⁴⁵. High-accuracy long-read sequencing has also been applied to reconstruct RHD and RHCE haplotypes and characterize structural variants within the locus⁴⁴.

In comparison with these published studies, the present work established reaction conditions that generated high-molecular-weight products in six RHD/RHCE-associated reactions (Figure 4.3). It did not determine whether each product was derived from RHD, RHCE, both genes, or another homologous sequence. The products must therefore not be described as sequence-confirmed or target-specific RHD/RHCE amplicons.

A similar limitation applies to the MNS-associated reactions. GYPA, GYPB, and GYPE form a homologous genomic cluster in which gene conversion, hybrid genes, and deletions contribute to blood group diversity^{7,9}. Long-read sequencing has been used to resolve GYPB deletions and characterize extended GYPA and GYPB regions, including complex MNS haplotypes^{9,46}.

The eight MNS-associated reactions generated high-molecular-weight products at the expected approximate gel positions (Figure 4.4). Sequencing would be required to determine

whether these products represent the intended target intervals, homologous glycophorin regions, or mixtures of related products. Future analysis should specifically evaluate multi-mapping reads, paralog-specific positions, allele balance, and possible hybrid breakpoints^{7,44}.

Until this sequencing-based evaluation has been completed, the MNS-associated and RHD/RHCE-associated products should be reported as gel-detected high-molecular-weight PCR products rather than validated target-specific amplicons.

5.6 Scientific and Potential Clinical Relevance

The workflow developed in this thesis is not intended to replace serological blood group typing. Serology remains the routine method for assessing red blood cell antigen expression and supporting transfusion-compatibility decisions¹. Established molecular assays complement serology when known antigen-associated variants are investigated or when serological interpretation is limited¹.

The principal added value of long-read sequencing lies in its ability to characterize extended allelic backgrounds, link distant variants, and resolve non-coding changes, hybrid genes, deletions, duplications, and other structural rearrangements that may not be adequately captured by predefined genotyping assays⁷. These capabilities are among the central motivations for applying long-read sequencing to complex blood group loci⁷.

The value of extended sequencing has already been demonstrated in published blood group studies. In the Kidd blood group system, overlapping long-range PCR products followed by Oxford Nanopore sequencing resolved genotype–phenotype discrepancies and identified novel variants and an approximately 5 kb deletion that had not been detected by the preceding routine methods⁴⁷. This study illustrates how extended gene sequencing can detect variation outside predefined target positions and place individual variants within a broader allelic background⁴⁷.

The present thesis contributes to this field by establishing the upstream PCR-enrichment, clean-up, barcoding, sequencing, and reference-alignment steps for selected blood group loci. However, the current workflow remains a research-stage method rather than a validated diagnostic assay.

Diagnostic implementation would require predefined acceptance criteria for DNA quality, PCR yield, target identity, read quality, mapping quality, coverage, allele balance, and variant-calling confidence^{7,43}. The workflow would also require evaluation using a sufficiently large and genetically diverse sample set with established serological and molecular reference results^{7,43}.

5.7 Limitations

The first major limitation is the small number of genomic DNA templates and the absence of a larger cohort of phenotyped and molecularly characterized donors or patients. Analytical sensitivity, specificity, repeatability, reproducibility, allele-detection rate, and concordance with serological typing could therefore not be determined. Successful product formation using a limited number of templates does not demonstrate equivalent amplification across genetically diverse alleles, particularly when sequence variants occur within primer-binding sites⁷.

A second limitation is the incomplete sequencing validation. Only selected ABO and KEL products were sequence-confirmed, whereas the RHD/RHCE-associated and MNS-associated products were evaluated only by agarose gel electrophoresis because sequencing was not performed owing to time and cost constraints. Their genomic origin and target specificity therefore remain unresolved.

Most expected PCR products were also larger than the 10 kb upper limit of the DNA marker. The gel images consequently demonstrated high-molecular-weight product formation but could not provide precise size measurements or establish genomic identity (Figures 4.1–4.4).

A third limitation arises from PCR-based target enrichment. Primer-dependent amplification may introduce allelic imbalance or allelic dropout when sequence variants alter primer-binding sites or structural changes remove one of the binding regions⁷. Overlapping products reduce the risk of unrepresented intervals between adjacent fragments but do not eliminate amplification bias⁷. PCR-free or capture-based long-read enrichment may provide complementary approaches for structurally complex alleles, although these methods differ in DNA-input requirements, sequencing output, cost, and analytical complexity⁷.

A fourth limitation concerns the extent of the bioinformatic analysis. The computational work performed in this thesis consisted of basecalling, barcode demultiplexing, reference-based alignment to GRCh38, and mapping and coverage visualization using MinKNOW and EPI2ME. Within the available project timeframe, no systematic read-quality analysis, consensus-sequence generation, variant calling, haplotype phasing, allele assignment, or orthogonal sequence confirmation was performed.

The sequencing results can therefore support conclusions regarding the genomic origin and approximate target coverage of the principal ABO and KEL products, but not regarding their complete sequence or allele composition. Reproducible variant detection would require defined thresholds for read quality, mapping quality, sequencing depth, strand balance, and allele balance⁴³. Variants and structural alleles identified in future analyses should also be confirmed using an independent method appropriate to the respective finding before clinical interpretation^{43,47}.

5.8 Future Development and Conclusion

The immediate experimental priority is sequencing-based validation of the RHD/RHCE-associated and MNS-associated products. Each PCR product should be assigned an unambiguous barcode and evaluated individually for locus specificity, off-target amplification, and representation of the expected primer-defined interval. Primer pairs that generate substantial off-target sequencing signals should be redesigned using paralog-specific sequence positions.

The workflow should subsequently be evaluated using a larger cohort containing common alleles and, where available, weak, partial, null, and hybrid alleles. Comparison with established serological and molecular results would permit evaluation of concordance and identify primer pairs susceptible to allelic bias or dropout^{7,43}.

Further development must include a defined bioinformatic workflow for read-quality assessment, high-accuracy basecalling, demultiplexing, reference alignment, consensus generation, small- and structural-variant calling, haplotype phasing, allele assignment, and standardized report generation^{7,43}. Highly homologous regions such as RHD/RHCE and the glycophorin gene cluster may require locus-specific reference sequences, alternative haplotypes, graph-based references, or de novo assembly when conventional linear-reference mapping is ambiguous^{7,44}.

Higher throughput through automation, balanced pooling, or multiplexing should be considered only after the performance of the individual primer pairs has been established. Multiplex long-range PCR may introduce primer interactions and unequal amplification between target regions^{7,41}.

In conclusion, this thesis established a long-range PCR-based enrichment workflow for selected blood group loci. PrimeSTAR® GXL Premix Fast provided the most suitable balance of high-molecular-weight product formation, limited additional visible products, reproducibility, and runtime under the tested conditions (Table 4.1).

The selected ABO and KEL products were confirmed by Oxford Nanopore sequencing and reference-based alignment to originate from the intended genomic regions (Figures 4.5 and 4.6). High-molecular-weight products were also generated in the MNS-associated and RHD/RHCE-associated reactions, but their genomic identity remains unresolved because they were not sequenced owing to time and cost constraints (Figures 4.3 and 4.4).

The present work therefore provides a methodological foundation for targeted long-read analysis rather than a validated diagnostic assay. With systematic sequence confirmation, quantitative quality assessment, variant calling, haplotype phasing, allele assignment, and evaluation in clinically characterized samples, the workflow could contribute to the comprehensive molecular characterization of blood group alleles and complex genotype–phenotype discrepancies⁷.

6 References

1. Westhoff CM. Blood group genotyping. *Blood*. 2019;133(17):1814-1820. doi:10.1182/blood-2018-11-833954.
2. Reid ME, Lomas-Francis C, Olsson ML. *The Blood Group Antigen FactsBook*. 3rd ed. Academic Press; 2012.
3. Ameer A, Kloosterman WP, Hestand MS. Single-molecule sequencing: towards clinical applications. *Trends in Biotechnology*. 2019;37(1):72-85. doi:10.1016/j.tibtech.2018.07.013.
4. Slatko BE, Gardner AF, Ausubel FM. Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology*. 2018;122(1):e59. doi:10.1002/cpmb.59.
5. Veldhuisen B, van der Schoot CE, de Haas M. Blood group genotyping: from patient to high-throughput donor screening. *Vox Sanguinis*. 2009;97(3):198-206. doi:10.1111/j.1423-0410.2009.01209.x.
6. Jungbauer C, Hobel CM, Schwartz DWM, Mayr WR. High-throughput multiplex PCR genotyping for 35 red blood cell antigens in blood donors. *Vox Sanguinis*. 2012;102(3):234-242. doi:10.1111/j.1423-0410.2011.01542.x.
7. Thun GA, Gueuning M, Mattle-Greminger MP. Long-read sequencing in blood group genetics. *Transfusion Medicine and Hemotherapy*. 2023;50(3):184-197. doi:10.1159/000530652.
8. Tounsi WA, Madgett TE, Avent ND. Complete RHD next-generation sequencing: establishment of reference RHD alleles. *Blood Advances*. 2018;2(20):2713-2723. doi:10.1182/bloodadvances.2018017871.
9. Xu H, Su X, Zuo Q, Zhang L, Chu X. The research of a large-scale analysis platform for MNS blood group identification based on long-read sequencing. *Transfusion Medicine Reviews*. 2024;38(4):150836. doi:10.1016/j.tmr.2024.150836.
10. 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 Advances*. 2023;7(6):878-892. doi:10.1182/bloodadvances.2022007133.
11. Storry JR, Olsson ML. The ABO blood group system revisited: a review and update. *Immunohematology*. 2009;25(2):48-59.
12. Yamamoto F, Marken J, Tsuji T, White T, Clausen H, Hakomori S. Cloning and characterization of DNA complementary to human UDP-GalNAc:Fuc alpha 1-2Gal alpha 1-3GalNAc transferase (histo-blood group A transferase) mRNA. *Journal of Biological Chemistry*. 1990;265(2):1146-1151.
13. Olsson ML, Chester MA. Polymorphisms at the ABO locus in subgroup A individuals. *Transfusion*. 1996;36(4):309-313. doi:10.1046/j.1537-2995.1996.36496226142.x.
14. Seltsam A, Hallensleben M, Kollmann A, Blasczyk R. The nature of diversity and diversification at the ABO locus. *Blood*. 2003;102(8):3035-3042. doi:10.1182/blood-2003-03-0955.
15. Daniels G. *Human Blood Groups*. 3rd ed. Wiley-Blackwell; 2013.
16. Flegel WA. The genetics of the Rhesus blood group system. *Blood Transfusion*. 2007;5(2):50-57. doi:10.2450/2007.0011-07.

17. Westhoff CM. The Rh blood group system in review: a new face for the next decade. *Transfusion*. 2004;44(11):1663-1673. doi:10.1111/j.0041-1132.2004.04237.x.
18. Blumenfeld OO, Huang CH. Molecular genetics of glycophorin MNS variants. *Transfusion Clinique et Biologique*. 1997;4(4):357-365. doi:10.1016/S1246-7820(97)80041-9.
19. Wei L, Shan ZG, Flower RL, Wang Z, Wen JZ, Luo GP, et al. The distribution of MNS hybrid glycoproteins with Mur antigen expression in Chinese donors, including identification of a novel GYP.Bun allele. *Vox Sanguinis*. 2016;111(3):308-314. doi:10.1111/vox.12421.
20. Mallari RA, Chan A, Powers RJ, Pandipati S, Bensing KM, Biese D, et al. Fetal inheritance of GP*^{Mur} causing severe HDFN in an unrecognized case of maternal alloimmunization. *Transfusion*. 2020;60(4):870-874. doi:10.1111/trf.15709.
21. Lee S, Russo D, Redman CM. The Kell blood group system: Kell and XK membrane proteins. *Seminars in Hematology*. 2000;37(2):113-121. doi:10.1016/S0037-1963(00)90036-2.
22. Schneider VA, Graves-Lindsay T, Howe K, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Research*. 2017;27(5):849-864. doi:10.1101/gr.213611.116.
23. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Research*. 2002;12(6):996-1006. doi:10.1101/gr.229102.
24. National Center for Biotechnology Information. Genome Data Viewer [Internet]. Bethesda (MD): National Library of Medicine; [cited 2026 Jul 20]. Available from: <https://www.ncbi.nlm.nih.gov/gdv/>
25. Microsynth AG. Oligonucleotide technical datasheets supplied with the synthesized primers. Balgach, Switzerland: Microsynth AG; 2026. Unpublished manufacturer documentation.
26. Oxford Nanopore Technologies. Ligation sequencing DNA V14 - dual barcoding (SQK-NBD114.24 with EXP-PBC096) [Internet]. Oxford: Oxford Nanopore Technologies; [cited 2026 Jul 20]. Available from: <https://nanoporetech.com/document/ligation-sequencing-dual-barcoding-v14>
27. Promega Corporation. Maxwell CSC Genomic DNA Kit Technical Manual, Cat. No. AS1850, Technical Manual TM677 [Internet]. Madison (WI): Promega Corporation; [cited 2026 Jul 20]. Available from: <https://www.promega.com/resources/protocols/technical-manuals/500/maxwell-csc-genomic-dna-kit-protocol/>
28. Promega Corporation. Human Genomic DNA, Cat. No. G3041: product information [Internet]. Madison (WI): Promega Corporation; [cited 2026 Jul 20]. Available from: <https://www.promega.com/products/biochemicals-and-labware/nucleic-acids/human-genomic-dna/>
29. Promega Corporation. GoTaq Long PCR Master Mix Technical Manual, Cat. Nos. M4020 and M4021, Technical Manual TM359 [Internet]. Madison (WI): Promega Corporation; revised 2017 May [cited 2026 Jul 20]. Available from: <https://www.promega.com/resources/protocols/technical-manuals/101/gotaq-long-pcr-master-mix-protocol/>

- 30.** New England Biolabs. PCR using Q5 High-Fidelity DNA Polymerase (NEB #M0491) [Internet]. Ipswich (MA): New England Biolabs; [cited 2026 Jul 20]. Available from: <https://www.neb.com/en/protocols/pcr-using-q5-high-fidelity-dna-polymerase-m0491>
- 31.** New England Biolabs. PCR protocol for LongAmp Taq DNA Polymerase (M0323) [Internet]. Ipswich (MA): New England Biolabs; [cited 2026 Jul 20]. Available from: <https://www.neb.com/en/protocols/m0323-longamp-taq-dna-polymerase-protocol>
- 32.** Takara Bio Inc. PrimeSTAR GXL DNA Polymerase, Cat. No. R050A: product manual, version 201906Da [Internet]. Shiga, Japan: Takara Bio Inc.; 2019 [cited 2026 Jul 20]. Available from: https://takara.co.kr/file/manual/pdf/R050A_e.v1906Da.pdf
- 33.** Takara Bio Inc. PrimeSTAR GXL Premix Fast, Cat. Nos. R053S/R053A: data sheet [Internet]. Shiga, Japan: Takara Bio Inc.; [cited 2026 Jul 20]. Available from: https://www.takarabio.com/documents/User%20Manual/R053S/R053S_R053A_DS.pdf
- 34.** Takara Bio Inc. PrimeSTAR LongSeq DNA Polymerase, Cat. Nos. R055S/R055A: data sheet [Internet]. Shiga, Japan: Takara Bio Inc.; [cited 2026 Jul 20]. Available from: https://www.takarabio.com/documents/User%20Manual/R055S/R055S_R055A_DS.pdf
- 35.** Henke W, Herdel K, Jung K, Schnorr D, Loening SA. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Research*. 1997;25(19):3957-3958. doi:10.1093/nar/25.19.3957.
- 36.** Biozym Scientific GmbH. Quantitas Pro DNA Marker 100 bp-10 kb, Art. No. 617121: product information [Internet]. Hessisch Oldendorf, Germany: Biozym Scientific GmbH; [cited 2026 Jul 20]. Available from: <https://www.biozym.com/biochemikalien/molekularbiologie/dna-und-rna-marker/quantitas-pro-dna-marker-100-bp-10-kb/617121>
- 37.** INTEGRA Biosciences AG. MAGFLO NGS Instruction Manual, document 137960, version 01 [Internet]. Zizers, Switzerland: INTEGRA Biosciences AG; 2025 [cited 2026 Jul 20]. Available from: https://www.integra-biosciences.com/sites/default/files/documents/137960_v01_im_magflo_ngs_en.pdf
- 38.** Oxford Nanopore Technologies. MinION Mk1B user manual [Internet]. Oxford: Oxford Nanopore Technologies; [cited 2026 Jul 20]. Available from: <https://nanoporetech.com/document/minion-user-manual>
- 39.** Oxford Nanopore Technologies. MinKNOW user guide [Internet]. Oxford: Oxford Nanopore Technologies; [cited 2026 Jul 20]. Available from: <https://nanoporetech.com/document/experiment-companion-minknow>
- 40.** Oxford Nanopore Technologies. EPI2ME workflow: wf-alignment [Internet]. Oxford: Oxford Nanopore Technologies; [cited 2026 Jul 20]. Available from: <https://nanoporetech.com/document/epi2me-workflows/wf-alignment>
- 41.** Jia H, Guo Y, Zhao W, Wang K. Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. *Scientific Reports*. 2014;4:5737. doi:10.1038/srep05737.
- 42.** Thun GA, Gueuning M, Sigurdardottir S, et al. Novel regulatory variant in ABO intronic RUNX1 binding site inducing A3 phenotype. *Vox Sanguinis*. 2024;119(4):377-382. doi:10.1111/vox.13580.
- 43.** Worner L, Rink G, Merkle X, Bugert P. Introduction of an optimized protocol for long-read nanopore sequencing of blood group genes in immunohematology case studies.

Transfusion Medicine and Hemotherapy. 2026. Advance online publication. doi:10.1159/000550249.

44. Zhang Z, An HH, Vege S, Hu T, Zhang S, Mosbrugger T, et al. Accurate long-read sequencing allows assembly of the duplicated RHD and RHCE genes harboring variants relevant to blood transfusion. *American Journal of Human Genetics*. 2022;109(1):180-191. doi:10.1016/j.ajhg.2021.12.003.

45. Tounsi WA, Lenis VP, Tammi SM, et al. Rh blood group D antigen genotyping using a portable nanopore-based sequencing device: proof of principle. *Clinical Chemistry*. 2022;68(9):1196-1201. doi:10.1093/clinchem/hvac075.

46. Lane WJ, Gleadall NS, Aeschlimann J, Vege S, Sanchis-Juan A, Stephens J, et al. Multiple GYPB gene deletions associated with the U-negative phenotype in those of African ancestry. *Transfusion*. 2020;60(6):1294-1307. doi:10.1111/trf.15839.

47. 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. doi:10.3390/biomedicines12010225.

Ich habe mich bemüht, sämtliche Inhaber*innen der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung der Bilder in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um Meldung bei mir.

7 List of abbreviations

Abbreviation	Full term
ABO	ABO blood group system; <i>ABO</i> glycosyltransferase gene
bp	base pair(s)
DEL	DEL phenotype; an RhD phenotype with extremely low D-antigen expression
DNA	deoxyribonucleic acid
EEO	electroendosmosis
FASTQ	text-based sequence format containing nucleotide sequences and associated quality scores
GC	guanine–cytosine
GRCh38	Genome Reference Consortium Human Build 38
GYPA	glycophorin A gene
GYPB	glycophorin B gene
GYPE	glycophorin E gene
hg38	UCSC designation for the GRCh38 human reference genome assembly
kb	kilobase(s)
KEL	Kell blood group system; <i>KEL</i> gene
LE	low electroendosmosis
LR-PCR	long-range polymerase chain reaction
MNS	MNS blood group system
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NGS	next-generation sequencing
PCR	polymerase chain reaction
Rh	Rhesus blood group system
RHD	gene encoding the RhD blood group antigen
RHCE	gene encoding the RhCE blood group antigens

RUNX1	RUNX family transcription factor 1
Taq	<i>Thermus aquaticus</i> DNA polymerase
TBE	Tris–borate–EDTA
UCSC	University of California, Santa Cruz
w/v	weight per volume

8 List of figures and tables

Figure 4.1. Agarose gel electrophoresis of ABO long-range PCR products	26
Figure 4.2. Agarose gel electrophoresis of KEL long-range PCR products	27
Figure 4.3. Agarose gel electrophoresis of RHD/RHCE-associated long-range PCR products.....	28
Figure 4.4. Agarose gel electrophoresis of MNS-associated long-range PCR products	29
Figure 4.5. EPI2ME alignment and coverage visualization of the three overlapping ABO long-range PCR products	30
Figure 4.6. EPI2ME alignment and coverage visualization of the KEL long-range PCR products	31
Table 3.1. Standard reaction composition for long-range PCR using PrimeSTAR® GXL Premix Fast	17
Table 3.2. Thermocycling profile used for long-range PCR with PrimeSTAR® GXL Premix Fast	18
Table 4.1. Comparison of DNA polymerases and PCR master mixes evaluated during long-range PCR optimization	25
Table A1. ABO primer sequences	47
Table A2. KEL primer sequences	47
Table A3. RHD/RHCE-associated primer sequences.....	48
Table A4. MNS-associated primer sequences.....	49

9 Appendix

9.1 Primer sequences used for long-range PCR amplification

Table A1. ABO primer sequences

Blood group	Primer label	Primer sequence (5'–3')	Amplicon size (bp)
ABO	Fragment 1_F	AAGAAGCCTCTGGGACCCTT	14058
ABO	Fragment 1_R	GGAGCAAGTAAGCACACCCT	
ABO	Fragment 2_F	AGTCTGACGTTAGCATTTCTCCTCAAG	13201
ABO	Fragment 2_R	CTAGGCTTCAGTTACTCACAACAG	
ABO	Fragment 3_F	AGCCTTGACCTCCCTTCTGA	13388
ABO	Fragment 3_R	TCTCACGCCTGGTAAACCAAA	

Table A2. KEL primer sequences

Blood group	Primer label	Primer sequence (5'–3')	Amplicon size (bp)
KEL	Fragment 1A_F	TGCTAAGAGGGAGACCACACA	14891
KEL	Fragment 1A_R	ATAAGGAAGGCCACTGCGTA	
KEL	Fragment 2A_F	TTCCTCAGGTGTGGTGTGAAG	13945
KEL	Fragment 2A_R	CTCCATTCTTGTGCCTCGCC	
KEL	Fragment 1B_F	TTGCTAAGAGGGAGACCACACA	14889
KEL	Fragment 1B_R	AGGAAGGCCACTGCGTAACA	
KEL	Fragment 2B_F	TTCCTCAGGTGTGGTGTGAA	13944
KEL	Fragment 2B_R	TCCATTCTTGTGCCTCGCC	

Table A3. RHD/RHCE-associated primer sequences

Blood group	Primer label	Primer sequence (5'–3')	Amplicon size (bp)
RHD/RHCE	Fragment 1_F	ATCCACTTTCCACCTCCCTGC	16559
RHD/RHCE	Fragment 1_R	GTGGATCACTTGAGGTCGGGAGTTCC	
RHD/RHCE	Fragment 2_F	TATTGAGCGTTCTCCATGTGCCAGTC	16872
RHD/RHCE	Fragment 2_R	CACCCTCTTGATAAGATGAAGCTGCCC	
RHD/RHCE	Fragment 3_F	GTTTGAGCCCAGGAGTTAGGGACCGAG	10789
RHD/RHCE	Fragment 3_R	CCCACTGTGACCACCCAGCATTCTA	
RHD/RHCE	Fragment 4_F	CATACCTTTGAATTAAGCACTTCAC	9895
RHD/RHCE	Fragment 4_R	CAGAATGGCCTTTACCAGCCAT	
RHD/RHCE	Fragment 5_F	GTTCAAGCTGTCAAGGAGACATCACTATACA	11628
RHD/RHCE	Fragment 5_R	CCAGTTTTAAGAATTTGTCGGCCGGTCTG	
RHD/RHCE	Fragment 6_F	ATACATTCCATCCAGAACTGTTACCC	11284
RHD/RHCE	Fragment 6_R	AGGCCAAGAGATCCTGGTGAAACTATCC	

Table A4. MNS-associated primer sequences

Blood group	Primer label	Primer sequence (5'–3')	Amplicon size (bp)
MNS	Fragment 1_F	GGGAAGGCTTTGTGGAGGAA	15807bp
MNS	Fragment 1_R	AGAAGGCACAGAGGCTTGTT	
MNS	Fragment 2_F	GGAGAATGGAGTGAACCCGG	14082 bp
MNS	Fragment 2_R	TTTTACCCTGTGAGCCGTGG	
MNS	Fragment 3_F	GGGAGGCCTCAGGAAACAAT	15807 bp
MNS	Fragment 3_R	GCAAACAACCAGGGCTGAAG	
MNS	Fragment 4_F	ATGGCTCCACAACAGCTACC	11357bp
MNS	Fragment 4_R	ACCAACTCTCAAGCAAGCCT	
MNS	Fragment 5_F	GCAGGTCACCCTACAATGCT	12812 bp
MNS	Fragment 5_R	TCTCTACGTGCTCCTACCCC	
MNS	Fragment 6_F	CAGTTGGGTGGCTCCTGTC	13170 bp
MNS	Fragment 6_R	GCACTGTGGTCTGAGAGACA	
MNS	Fragment 7_F	TGGAAAGGCACATCCGGTAA	14000bp
MNS	Fragment 7_R	GTTGTGGGTGGTATATGGGCT	
MNS	Fragment 8_F	CTGGAGTGTGCAGTGGTGT	11650bp
MNS	Fragment 8_R	GTTTCTATGAGTATCTTGGAGGTGC	

9.2 Deutsche Zusammenfassung

Die Bestimmung von Blutgruppen ist für eine sichere Transfusionsmedizin, für Transplantationen sowie zur Vermeidung von Alloimmunisierungen von zentraler Bedeutung¹. Obwohl serologische Methoden weiterhin den diagnostischen Standard darstellen, können sie bei seltenen Varianten, Hybridallelen, kürzlich transfundierten Patient:innen oder bei Blutgruppenloci mit hoher Sequenzähnlichkeit zwischen verwandten Genen an ihre Grenzen stoßen^{1,2}. Molekulargenetische Verfahren können einige dieser Einschränkungen überwinden. Konventionelle zielgerichtete Analysen und Short-Read-Sequenzierungen ermöglichen jedoch häufig keine vollständige Charakterisierung strukturell komplexer Loci und keine eindeutige Zuordnung gemeinsam auftretender Varianten zu demselben Allel^{1,3}. Long-Read-Sequenzierung erhält die Verknüpfung von Varianten über größere genomische Bereiche und ermöglicht dadurch Haplotypisierung sowie den Nachweis neuer Varianten, Hybridallele und struktureller Veränderungen^{3,4}.

Ziel dieser Masterarbeit war die Etablierung und Optimierung eines auf Long-Range-PCR basierenden Anreicherungsverfahrens für klinisch relevante Blutgruppenloci. Untersucht wurden ABO, KEL, MNS-assoziierte Regionen sowie RHD/RHCE-assoziierte Regionen. Der entwickelte Workflow sollte mit einer nachfolgenden Oxford-Nanopore-Sequenzierung kompatibel sein. Mehrere kommerziell verfügbare DNA-Polymerasen und PCR-Mastermixe wurden hinsichtlich ihrer Eignung zur Amplifikation großer genomischer Fragmente verglichen. Die Takara PrimeSTAR® GXL Fast DNA Polymerase zeigte die beste Kombination aus zuverlässiger Amplifikation, Spezifität, Reproduzierbarkeit und PCR-Laufzeit und wurde daher für die weitere Methodenentwicklung ausgewählt.

Long-Range-Primerpaare wurden neu entworfen oder aus publizierten Ansätzen adaptiert und experimentell getestet. Die Agarosegelelektrophorese zeigte für ABO, KEL sowie die MNS- und RHD/RHCE-assoziierten Regionen deutliche hochmolekulare PCR-Produkte an den erwarteten ungefähren Positionen. Für ABO wurden drei überlappende Amplicons erzeugt, während für die ausgewählten KEL-Regionen zwei Amplicons erhalten wurden. Ausgewählte ABO- und KEL-Produkte wurden anschließend mittels Oxford-Nanopore-Sequenzierung analysiert. Die Sequenzierung bestätigte, dass die untersuchten Produkte aus den vorgesehenen genomischen Regionen stammten. Dies unterstützte die Spezifität der entsprechenden Primerpaare sowie die Eignung des optimierten PCR-Workflows für die nachfolgende Long-Read-Sequenzierung.

Die MNS- und RHD/RHCE-assoziierten PCR-Produkte wurden aus Zeit- und Kostengründen nicht sequenziert und daher ausschließlich anhand ihrer Bandenpositionen im Agarosegel beurteilt. Da die Agarosegelelektrophorese keine eindeutige Bestimmung der genomischen Herkunft erlaubt, können diese Produkte noch nicht als validierte, zielregionsspezifische Amplicons eingestuft werden. Eine sequenzbasierte Bestätigung ist erforderlich, um Off-Target-Amplifikationen aus homologen genomischen Regionen auszuschließen.

Insgesamt schafft diese Arbeit eine Grundlage für die gezielte Long-Read-basierte Analyse von Blutgruppengenen und zeigt die grundsätzliche Eignung einer Kombination aus optimierter Long-Range-PCR und Oxford-Nanopore-Sequenzierung für ausgewählte ABO- und KEL-Loci. Der etablierte Workflow bildet zudem die Grundlage für die zukünftige Validierung der MNS- und RHD/RHCE-assoziierten Produkte sowie für eine haplotypaufgelöste Charakterisierung von Blutgruppenallelen.