

Next-Generation Sequencing Technologies in Blood Group Typing

Daniel Fürst Chrysanthi Tsamadou Christine Neuchel Hubert Schrezenmeier
Joannis Mytilineos Christof Weinstock

Institute for Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Transfusion Service, Baden Wuerttemberg/Hessen, and University Hospital Ulm, Ulm, Germany; Institute of Transfusion Medicine, University of Ulm, Ulm, Germany

Keywords

Next-generation sequencing · Sequencing · Blood group · DNA typing · Donor typing program

Abstract

Sequencing of the human genome has led to the definition of the genes for most of the relevant blood group systems, and the polymorphisms responsible for most of the clinically relevant blood group antigens are characterized. Molecular blood group typing is used in situations where erythrocytes are not available or where serological testing was inconclusive or not possible due to the lack of antisera. Also, molecular testing may be more cost-effective in certain situations. Molecular typing approaches are mostly based on either PCR with specific primers, DNA hybridization, or DNA sequencing. Particularly the transition of sequencing techniques from Sanger-based sequencing to next-generation sequencing (NGS) technologies has led to exciting new possibilities in blood group genotyping. We describe briefly the currently available NGS platforms and their specifications, depict the genetic background of blood group polymorphisms, and discuss applications for NGS approaches in immunohematology. As an example, we delineate a protocol for large-scale donor blood group screening established and in use at our institution. Furthermore, we discuss technical challenges and limitations as well as the prospect for future developments, including long-read sequencing technologies.

© 2019 S. Karger AG, Basel

Introduction

Next-generation sequencing (NGS) technology refers to the second generation of genomic sequencing technologies which succeeded the first generation based on Sanger dideoxy sequencing [1]. All NGS methods share the features of relatively short reads and the possibility of sequencing many, usually many millions of, templates or fragments concomitantly. Therefore, these sequencing approaches are often described as massively parallel sequencing. Third-generation sequencing technologies, sometimes also called next-next-generation sequencing, are characterized by the direct sequencing of single-sequence templates [2]. These methods allow for much longer sequence reads, albeit the sequence quality is lower compared to short-read second-generation methods. NGS has decreased the costs per megabase and increased the throughput drastically [3]. Knowledge about disease-causing mutations and other genomic variations has increased exponentially [4]. Sophisticated molecular barcoding techniques allow simultaneous sequencing of hundreds or even thousands of samples and inclusion of many different genomic regions [5]. Such approaches are standard now in many laboratories performing donor HLA typing for registry purposes and often donors are upfront more comprehensively typed than required for clinical purposes in histocompatibility assessment [6].

Although serology is still the standard method in immunohematological diagnostics, molecular typing is getting increasingly important in certain instances, particu-

larly when serology is inconclusive or not feasible [7]. As the genetic polymorphisms for almost all blood group antigens have been elucidated, a molecular approach is possible for most diagnostic problems including donor screening [8]. Particularly high-throughput NGS methods offer advantages beyond classic techniques for DNA typing [9]. In the following, we describe the different most widely used NGS platforms with their specifications as well as their advantages and drawbacks for immunohematological diagnostics. We additionally discuss potential applications of these methods as well as their limitations and share our experience with the establishment of a comprehensive amplicon-based NGS donor screening protocol for high-throughput typing of blood group alleles. Finally, we describe the perspective for future developments with an emphasis on long-read technologies.

Short-Read Sequencing Technologies

Sequencing techniques of the second generation are characterized by the introduction of high-throughput massively parallel sequencing. The most common sequencing methods are Thermo Fisher's Ion Torrent semiconductor sequencing and Illumina's sequencing-by-synthesis techniques. Semiconductor sequencing is based on indirect detection of nucleotide incorporation via changes in electric potentials [10]. Sequencing capacity is determined by the number of wells per chip. Ideally, each well is loaded with a single sequencing bead that has been coated with clonal strands of a sequencing template. Upon incorporation of nucleotides, protons are cleaved (H^+), resulting in a change of electric potentials. If more than 1 nucleotide of the same type has been incorporated during a sequencing cycle, the change in electrical potential is correspondingly higher, which makes the system prone to errors in homopolymer stretches with indel error rates as high as 1.5% [11]. The main advantage of this system is its sequencing speed taking only a couple of hours per run. Ion Torrent semiconductor sequencing with the PGM machine offers read numbers from 0.4 to 5.5 million. Read lengths vary from 200 bp up to 400 bp.

Illumina's sequencing-by-synthesis method is based on direct detection of nucleotide incorporation via optical signals [12]. A solid phase is coated with adapter oligonucleotides. The sequencing library is flushed over the flow cell and captured by the adaptors followed by on-board cluster generation and incremental sequencing of each nucleotide with optical detection based on fluorescence of the dye-labelled bases. Illumina's systems offer the possibility of paired-end sequencing, which means that each template is sequenced from both ends. The error patterns consist mostly of substitutions which occur at a rate of approximately 0.1% [11]. Therefore, the system delivers a high raw-

read accuracy. The use of optical detection results in relatively long run times taking up to 56 h depending on the read length as well as the capacity of the flow cell.

The MiSeq's different sequencing kits offer the greatest flexibility with regard to read length and output. Short-read NGS has been successfully used for blood group typing based on whole genome data. Particularly blood group antigens resulting from single nucleotide variations (SNVs) are typed with high accuracy given sufficient coverage. However, more complex situations, such as phase resolution of more distant polymorphisms, mapping problems due to homologous genes, or structural changes in the RHCE and RHD genes, may be problematic due to the short raw sequences and have to be addressed with sophisticated bioinformatics approaches [13]. Whole genome approaches are not limited to antigen typing but may also identify polymorphisms relevant for antigen expression using association studies. This has been demonstrated in a study identifying polymorphisms correlating with expression in P1 and Xg^a antigens [14].

Long-Read Single-Molecule Technologies

These techniques generate sequencing data from single molecules. The single-molecule real-time (SMRT) method from Pacific Biosciences is based on fluorescence detection and sequencing by synthesis [15]. During sequencing, the polymerase incorporates fluorescently labelled nucleotides. The emitted fluorescent signal is detected by a camera in real time and subsequently assigned to the corresponding DNA nucleotide of the sequenced DNA molecule. Multiple sequencing cycles are possible. Thus, high sequencing coverage from a single molecule can be achieved [16].

The nanopore sequencing method from Oxford Nanopore Technologies determines the nucleotide sequence of single-stranded DNA molecules through diminutive changes in electric currents. The sequencing flow cell contains a membrane coated with protein nanopores. Each DNA molecule is coupled to a motor protein which facilitates nanopore passing. The conformation of these nanopores is altered by the passing nucleotide 5- or 6-mer string. Each change in pore conformation results in a change of the electric current. 512 sensors, each connected to 4 nanopores, are continuously measuring the electrical current over the pore [17]. Base calling is performed in real time either locally or through a cloud-based service. Once the DNA molecule has passed the pore, it can no longer be sequenced. Other than that, depth of coverage is highly dependent on the amount of similar DNA molecules that have been sequenced together in the same run.

Both methods allow for exceptionally long read lengths. SMRT sequencing achieves read lengths up to 45

kb. The nanopore technique is theoretically capable of sequencing DNA strands up to 200 kb in length [18].

With reads this long, the problem of sequencing and assembly through repetitive elements can be overcome. Further advantages are better phasing of polymorphisms and the ability for direct characterization of epigenetic modifications [19, 20].

A special feature of Oxford Nanopore's MinION device is the small size, which is comparable to an USB stick. Compared to short-read sequencing technologies, the SMRT methods have a higher error rate (11–15%), mainly consisting of insertions [21]. This error rate, however, can be corrected by using an algorithm for sequencing analysis that has been specifically designed, or it can be reduced by increasing the number of generated sub-reads [22, 23]. This proneness to indel errors ultimately results in the need of higher sequencing depth and also complicates the alignment procedure as analysis software is needed that specifically addresses the indel problem [19].

Both sequencing techniques have been used to generate high-quality HLA sequencing data [24, 25]. Homopolymer stretches seem to be a problem particularly for the nanopore device.

Genetic Variations Creating Blood Group Polymorphisms

Any structure of the outer red cell membrane can become a blood group antigen, when it elicits an antibody response in an individual who does not carry this structure. The expression or lacking expression of blood group antigens by different individuals is in most cases caused by genetic variations.

SNVs in the coding sequence are a major cause of blood group polymorphism. These nucleotide exchanges can cause amino acid changes which may alter protein antigens. For example, the alleles *JK*^{*}*A* (838G) and *JK*^{*}*B* (838A) differ by 1 nucleotide [26]. This nucleotide exchange determines the expression of the antigens *Jk*^a (280Asp) or *Jk*^b (280Asn).

Polysaccharide antigens are built by enzymes which transfer carbohydrates to precursor structures. Variations of the genes coding these enzymes may lead to altered reactivity or substrate specificity and, thereby, indirectly affect antigen expression. A prominent example is the ABO blood group system. The gene encoding the α 1,3-*N*-acetyl-D-galactosaminyltransferase (A-transferase), which converts the H antigen into the A antigen, differs by 7 nucleotide exchanges from the gene encoding the α 1,3-D-galactosyltransferase (B-transferase) [27, 28], which converts the H antigen into the B antigen. Three of these 7 nucleotide exchanges are si-

lent; the other 4 changes cause amino acid changes. The variant enzyme no longer transfers *N*-acetyl-D-galactosamine but now transfers D-galactose. Finally, gene conversion is thought to be the mechanism causing hybrid alleles like the alleles of the *RHD*^{*}*DVI* family, where up to 4 *RHD* exons are replaced by *RHCE* exons [29, 30].

Genetic Variations Causing Null Phenotypes

Cells that lack a particular antigen have a null phenotype for this antigen. Null phenotypes arise from homozygosity for a genetic variation that prevents the expression of the antigen or from heterozygosity for 2 different genetic variants, both preventing the expression of the antigen. One mechanism for a null phenotype is the introduction of a premature stop codon as in the very rare *JK*^{*}*01N.02* allele, where the nucleotide exchange 202C>T alters codon 68 into a stop codon. In some null phenotypes, the coding sequence is not changed, but other mechanisms interfere with the expression of the antigen. For example, the nucleotide exchange -67T>C in the promotor region of the *ACKR1* gene, which encodes the FY protein, prevents its transcription in hematopoietic precursor cells [31]. Nucleotide exchanges within the splice site region may hamper the ligation of the transcripts to form the final mRNA. Incomplete mRNA causes the formation of erroneous proteins which are not integrated into the cell membrane, thereby causing null phenotypes. For example, the exchange of guanosine by adenine at the first position of intron 8 (IVS8+1G>A) of the RhD gene disturbs the splicing in a way that detectable RhD protein is not expressed [32]. The deletion of guanosine at position 261 in the gene of the A-transferase causes a frame shift leading to a premature termination of the peptide. The resulting protein is without function and is unable to convert the H antigen to the A antigen. The *ABO*^{*}*O.01.01* allele is the most frequent cause for the blood group O phenotype. A deletion of 17 nucleotides in exon 3 of the Small Integral Membrane Protein 1 gene (*SMIM1*) also causes a frame shift and a premature stop codon leading to the Vel-negative phenotype [33]. The duplication of 37 base pairs in exon 4 of the *RHD* allele causes a premature stop codon [34]. This allele (*RHD* Ψ) is a common cause for the RhD-negative phenotype in Africans. The most frequent cause of the RhD-negative phenotype is the deletion of the whole *RHD* gene [35]. Other phenotypes arise by the deletion of exons, for example in the Gerbich blood group system (GE): the deletion of exon 2 causes the phenotype GE:-2,3,4 (former: Yus phenotype) and the deletion of exon 3 causes the phenotype GE:-2,-3,4 (former: Gerbich phenotype) [36, 37].

NGS of Blood Groups

Serological typing of blood group antigens is fast and for many antigens inexpensive but has its limitations: It is often impossible to type pre-transfused patients, for many blood group antigens antisera are not available, and weak antigen expression may falsely be typed as negative [38, 39]. In these situations, genotyping is superior. Genetic variants carrying 1 or very few nucleotide exchanges can easily be typed by using sequence-specific primer (SSP) polymerase chain reaction (PCR) [40] or by techniques using sequence-specific oligo probes (SSO) [41, 42]. For genes with many variants, such as *RHD* or *RHCE*, genotyping by SSP-PCR gets elaborate and sequencing can be advantageous [43, 44].

Genotyping results are impatiently awaited when decisions on transfusion practice have to be made. Thus, genotyping of patient samples by NGS is unfavorable, because the generation of results with current short-read technologies requires several days [45]. In addition, the workflows are designed and are cost-effective only for large sample numbers [46].

In contrast to patient typing, NGS is ideal for donor typing: the optimal number of samples can be collected and processed calmly, and a large number of blood group alleles can be tested for each donor simultaneously by NGS. A well-typed inventory is very helpful for blood banks when patients with blood group antibodies have to be transfused [47]. Additional primers to screen for very rare phenotypes, e.g. Vel negative, Lan negative, or In(b-) phenotypes, can be easily included in the panel.

High-Throughput Amplicon-Based Donor Typing with NGS

For the development of a high-throughput blood group donor typing protocol, an amplicon-based approach using Illumina short-read technology was chosen [48]. A total of 32 PCR primer pairs were designed. These primer pairs were distributed into 5 multiplex PCRs to reduce costs of PCR amplification. Iterative evaluation of the primer pairs was performed, changing the composition of each PCR mix and concentration of primers until all PCR reactions performed adequate amplification of the target sequences. Targeted polymorphisms are summarized in Table 1. All primers included the Illumina generic TruSeq (TS) sequences at the 5' end of all forward and reverse primers, respectively. Initial amplification of target regions is followed by a second PCR which adds sample-wise dual indexes as well as Illumina adaptors via the flanking TS sequences [49].

All pipetting is performed by liquid handler workstations. DNA isolation is performed in deep-well plates.

DNA is afterwards transferred to standard PCR plates in 96-well format. One plate includes 94 samples plus 1 positive and 1 negative control. From each sample, 5 multiplex PCRs are generated followed by PCR cleanup and second PCR. After another round of cleanup, plate-wise pooling is performed combining all uniquely barcoded samples for each multiplex PCR leading to 5 pools (Fig. 1). These pools are then quantified and combined in equimolar amounts. The final pool is then prepared for sequencing according to Illumina's library preparation guide. Due to the large number of primers, enough sequence diversity is generated and a PhiX spike-in of 5% was sufficient. The protocol scales to $x \times 94$ samples, and runs with up to $8 \times 94 = 752$ samples are possible using V2 500 cycles chemistry. The number of samples, which may be sequenced in 1 run, can be calculated by the number of total reads divided by the number of amplicons and the number of reads per amplicon.

The dual indexing strategy consisting of individual oligonucleotide sequence (multiplexing indexes [MID]) combinations at the 5' and 3' end of the sequencing templates allows multiplexing large sample numbers with a limited amount of index sequences [50]. In this protocol, we are using 34 individual MID's within read 1 and 48 individual MID's within read 2. Barcode oligo-sequences were designed with a minimum distance of 2 differences between each combination. The first step of bioinformatics is performed by the instrument using a fastq-only workflow resulting in a collection of gzip-compressed fastq files for each sample. Genotyping is performed using custom R/Bioconductor functions and scripts, which perform identification of the amplicon using forward and reverse target-specific primer sequences forming the beginning of each sequence read [51]. The blood group-defining polymorphisms are detected using pattern matching with known sequences. Individual cutoffs for presence of a polymorphism are used, which range from 15 to 20% of the allele counts. Furthermore, an index is computed between the number of the counted primer sequences, i.e., the number of amplicons, and the sum of target polymorphisms. This index is usually close to 1; extreme values are flagged for review as they may represent rare variants or novel alleles. Validation showed that a minimum threshold of 15 reads per amplicon is sufficient for correct genotyping; however, the assay is designed to result in read counts of at least 100 per amplicon on average. In a second step, the genotype counts are converted to blood group phenotypes (positive or negative) according to standard nomenclature [52]. This conversion results in a final csv file, which is ready for review or import into a LIMS (laboratory information system).

Validation of the assay was performed for 570 samples with known genotypes obtained by a TaqMan assay. Complete concordance was achieved for the alleles of the following antigens and phenotypes: K, k, Kp(a,b), Js(a,b),

Table 1. Overview of the most relevant polymorphisms typed

Blood group	Gene	Exon(s)	Polymorphisms	Antigen(s)/ phenotypes/ allele(s)	Nucleotide	Amino acid
MNS	GYP A	2	rs7682260; rs7687256; rs7658293	M, N	c.59C>T; c.71G>A; c.72T>G	p.Ser20Leu; p.Gly24Glu
	GYP B	4	rs7683365	S, s	c.143C>T	p.Thr48Met
	GYP A	3	rs144802902	Vw, Hut	c.140C>T	p.Thr47Lys
Rhesus	RHCE	1	rs586178; rs138268848	Diverse, C ^w	c.48G>C; c.122A>G	p.Trp16Cys; p.Gln41Arg
	RHCE	5	rs147357308; rs144163296; rs1053361; rs1029541644; rs1132763	Diverse	c.667G>T; c.712A>G; c.733C>G; c.748G>A; c.787A>G	p.Val223Phe; p.Met238Val; p.Leu245Val; p.Val250Met; p.Arg263Gly
Kell	KEL	6	rs8176058	K, k	c.578C>T	p.Thr193Met
	KEL	8	rs8176059	Kp(a), Kp(b)	c.841C>T	p.Arg281Trp
	KEL	8	rs61729034	KEL:–11,17	c.905T>C	p.Val302Ala
	KEL	17	rs8176038	Js(a), Js(b)	c.1790T>C	p.Leu597Pro
Duffy	ACKR1	2	rs12075	Fy(a), Fy(b)	c.125A>G	p.Asp42Gly
	ACKR1	2	rs34599082; rs13962	Fy [*]	c.265C>T; c.298G>A	p.Arg89Cys; p.Ala100Thr
	ACKR1	Promoter	rs2814778	Fy(a–, b–)	c.–67T>C	p.0
Kidd	SLC14A1	9	rs1058396	Jk(a), Jk(b)	c.838A>G	p.Asn280Asp
Lutheran	B-CAM	3	rs28399653	Lu(a), Lu(b)	c.230G>A	p.Arg77His
	B-CAM	6	rs28399656	LU:–8,14	c.611T>A	p.Met204Lys
	B-CAM	12	rs1135062	Au(a), Au(b)	c.1615A>G	p.Thr539Ala
Cartwright	ACHE	2	rs1799805	Yt(a), Yt(b)	c.1057C>A	p.His353Asn
Dombrock	ART4	2	rs11276	Do(a), Do(b)	c.793A>G	p.Asn265Asp
	ART4	2	rs28362797	Hy–	c.323G>T	p.Gly108Val
	ART4	2	rs28362798	Jo(a–)	c.350C>T	p.Thr117Ile
Colton	AQP1	1	rs28362692	Co(a), Co(b)	c.134C>T	p.Ala45Val
	AQP1	1	rs17852043	Co(b)	c.133G>A	p.Ala45Thr
	AQP1	1	rs377506522	Co(a–b–)	c.140A>G	p.Gln47Arg
Diego	SLC4A1	19	rs2285644	Di(a), Di(b)	c.2561C>T	p.Pro854Leu
	SLC4A1	16	rs75731670	Wr(a), Wr(b)	c.1972G>A	p.Glu658Lys
Scianna	ERMAP	4	rs56025238	Sc1, Sc2	c.169G>A	p.Gly57Arg
	ERMAP	4	rs56136737	Sc4	c.178C>G	p.Pro60Ala
LW	ICAM-4	1	rs77493670	LW(a), LW(b)	c.299A>G	p.Gln100Arg
Indian	CD44	2	rs369473842	In(a), In(b)	c.137G>C	p.Arg46Pro
	CD44	3	rs1071695	INFI–	c.255C>G	p.His85Gln
Lan	ABCB6	2	rs149202834	Lan–	c.574C>T	p.Arg192Trp
	ABCB6	12	rs145526996	Lan weak	c.1762G>A	p.Gly588Ser
	ABCB6	14	rs13402964	Lan–	c.1942C>T	p.Arg648X
Vel	SMIM1	3	rs566629828	Vel–	c.64_80delAGCCTAGGGGCTGTGTC	p.Ser22Glnfs*?
Jr	ABCG2	4	rs201121511; rs72552713	Jr(a–)	c.337C>T; c.376C>T	p.Arg113X; p.Gln126X
	ABCG2	7	rs140207606; na; rs200190472; rs200473953; rs387906870	Jr(a–)	c.706C>T; c.730C>T; c.736C>T; c.784G>T; c.791_792delTT	p.Arg236X; p.Gln244X; p.Arg246X; p.Gly262X; p.Leu264HisfsX

Co(a,b), Di(a,b), Wr(a,b), Jk(a,b), M, N, S, s, Sc1, Sc2, Lan, KEL:11, KEL:17, and Fy(a–b–). For Fy(a,b), 2 typings (0.4%) were discrepant, one sample was close to the low-read cutoff and resequencing resulted in the correct typ-

ing, the other discrepant typing turned out to be a novel allele [53]. For Do(a,b), 10 typings (1.8%) were discrepant, which led to a redesign of the primers and showed complete concordance afterwards. For Lu(a,b), 1 typing

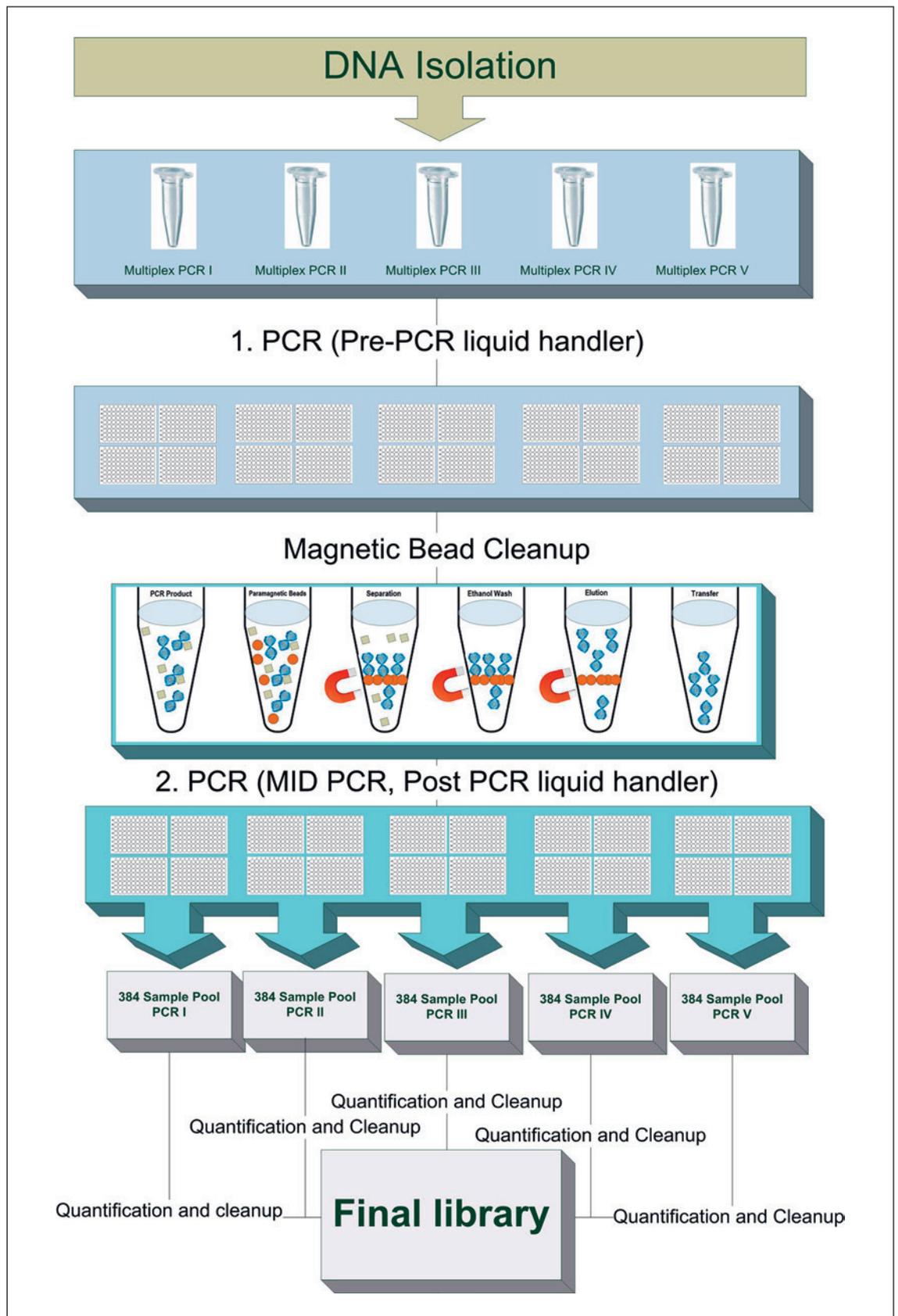


Fig. 1. Workflow of the donor blood group screening protocol based on an Illumina amplicon sequencing protocol. Two libraries with different sets of MID's may be combined allowing simultaneous blood group typing of up to 768 samples with v2 500 cycles chemistry.

was discrepant with a low read count. This sample was typed correctly in retesting. For Vel, 8 samples (1.4%) differed from reference testing. Further evaluation showed that all NGS results were correct, and the reference result had to be corrected. For Yt(a,b), 2 discrepant results (0.4%) were observed. One result in NGS showed low read count and was correct in resequencing. In the other result, NGS turned out to be correct, and the reference had to be corrected. The overall performance of the assay was acceptable. Genotype PCR dropouts within the multiplex reactions were observed occasionally in 1.7% of the amplicons and cluster mostly around some samples with poor DNA amount or purity. Such samples would prompt repeating DNA isolation and retyping.

Challenges and Technical Limitations

The formation of primer dimers is a common problem of PCR-based techniques [54]. They result in nonspecific products and may result in sequencing reads that cannot be mapped or that produce noise signal in the alignments [55]. Complementarity between primers can initiate amplification of primer dimer artifacts. Factors that further promote the formation of primer dimers are high primer concentrations, low template concentration, and less stringent PCR conditions, such as low annealing temperature and high number of PCR cycles. Particularly in multiplex PCR, primer dimers can pose a serious problem as the interactions between all primers are hard or even impossible to predict [56]. A useful tool to analyze primer interactions is the Multiple Primer Analyzer online tool by Thermo Fisher [57]. Careful primer design and evaluation is necessary to handle this problem. Additionally, optimization of PCR conditions (annealing temperature, number of cycles, magnesium chloride concentration, nucleotide concentration, primer concentration, and template concentration) can be helpful as can be the use of Hot Start Taq polymerase. Also, structural modification of primers has been reported as being able to reduce or prevent the formation of primer dimers [58].

Another common problem in NGS amplicon-based sequencing is the formation of hybrid alleles in heterozygous polymorphic alleles or between highly homologous genes [59]. The formation of such hybrid alleles is an artefact introduced by PCR amplification of the target sequences and occurs during the first few cycles if low amounts of template are present or during the final cycles when high amounts of PCR products compete in multiplex PCR. The mechanisms that may account for this effect are the formation of heteroduplexes between incompletely elongated fragments to the competing allele which primes a subsequent elongation [60]. This leads to the production of an artefactual sequence consisting of a 5' part from one allele

and a 3' part from the other allele or gene, respectively. Another mechanism, which may lead to the generation of hybrid alleles, is template switching [61]. This effect is commonly perceived to be a property of reverse transcriptase but has also been described in DNA amplification with Taq polymerase. Formation of such hybrid alleles during the first few cycles leads to exponential amplification in subsequent PCR cycles and may lead to a significant proportion of artificial sequences in data analysis. It is commonly observed in HLA sequencing and in sequencing of the FCGR3B, which forms the HNA-1 neutrophil antigen, but may also occur in sequencing the genes of ABO, Rhesus, Kell, and other highly polymorphic blood group systems. Careful primer design and optimization of PCR protocols may reduce formation of hybrid alleles, but often not completely. Awareness of this effect is necessary in analysis of NGS blood group sequencing data to recognize this problem and to preform adequate correction.

Pseudogenes and homologous genes are a challenge both in primer design as well as in data analysis [62]. The *RHD* and the *RHCE* genes, for example, share 92% sequence similarity [63]; also, the *GYP A* and *GYP B* genes encoding the MNS blood group show more than 95% sequence identity [64]. It may be difficult particularly with short-read techniques to map the reads unambiguously in all situations. Uncommon results or novel alleles should be reviewed critically taking into consideration the problem of primer specificity and co-amplification of homologous genomic sequences. Careful validation is necessary to identify such problems.

Genomic variations within the primer binding sites may lead to allele dropouts leading to incorrect typings [65]. Comprehensive frequency data for SNVs are now available from various sources, and primer binding sites should be checked for common variations upon design of the primers.

Applications for Long-Read Sequencing in Immunohematology

The most obvious application of long-read technologies for molecular typing of blood groups is the resolution of ambiguities which cannot be resolved with conventional methodology [66]. If the distance between 2 or more heterozygous bases is longer than the maximum length of the sequencing template, it is impossible to infer the haplotypic configuration with Sanger or short-read NGS [67]. The phase of 2 SNVs can only be resolved if both are sequenced within the same (hemizygous) read. The long-read technologies thus enable a better resolution as templates with more distant variations may be sequenced. This is particularly important in highly polymorphic genes like the transferases of the ABO system

and the *RH* genes. Only long-read sequencing technologies allow the definition of complete hemizygous sequences in cases involving weak or partial RHD antigens as well as nonexpressed variants without previous cloning in heterozygous samples. Such methods may enrich the diagnostic portfolio with respect to clinical setting as they may well be integrated in a diagnostic workup of questionable cases. Another advantage of the ultralong-read technologies is that sequencing performed in real time allows in principle short turn-around times of less than 24 h [68]. Above this, epigenetic aspects, such as promoter DNA methylation, may be taken into consideration in the future [69]. Particularly the nanopore technology offers exciting possibilities due to the ability to sequence templates up to 100 kb or even more in certain instances. A method to sequence the complete *RHD* gene with overlapping amplicons has already been described [43]. Oligonucleotide bait enrichment may allow sequencing complete large genes, such as the transferases of the ABO system, the *RHD* or the *RHCE* genes, directly from genomic DNA without the need for prior PCR [70].

Direct RNA sequencing with the nanopore device may aid investigation of low or questionable expression of blood group antigens [71]. In the future, nanopore devices may be able to sequence directly from whole blood without prior DNA or RNA isolation, which may simplify the workflow even more [72]. Finally, the size and weight of the nanopore device similar to a USB stick allows usage in environments with limited infrastructure [73].

Summary and Perspective

The introduction of short-read NGS technologies has led to a dramatic reduction in sequencing costs [74]. The principle of massively parallel sequencing allows sequencing of millions of reads in 1 run. Multiplexing strategies using molecular barcodes and automatization protocols allow cost-efficient genotyping of large cohorts with short turn-around time [75]. This allows the application of NGS technology in donor screening programs, which will eventually improve provision of blood products for patients with rare phenotypes or patients immunized for frequent antigens [76]. Inclusion of genetic variants leading to null phenotypes is easy and straightforward in most instances [77]. Powerful bioinformatics tools augment their detection with limited hands-on time. Also, detection of novel or rare variants is possible, where serology is difficult or no antibodies are available. This will eventually boost knowledge about the allelic diversity of the human blood groups systems and will improve safety and efficacy of blood products with low additional costs.

NGS reads also allow quantitative or semi-quantitative assessment of mixed phenotypes or chimeric samples in selected applications [78, 79]. Single, long-read sequencing, particularly nanopore sequencing, will allow resolution of polymorphisms, which were previously difficult or impossible to resolve particularly when allelic ambiguities are present [80]. This approach is especially promising for *RH* variants. Further development of this technique will possibly allow RNA sequencing or direct DNA sequencing from blood without previous library preparation and may provide novel diagnostic tools in clinical and scientific settings.

Statement of Ethics

The authors have no ethical conflicts to disclose.

Disclosure Statement

The authors have no conflicts of interest to declare.

Funding Sources

No funding was received in the preparation of the manuscript.

Author Contributions

D.F., C.T., C.N., H.S., J.M., and C.W. researched and wrote the manuscript.

References

- 1 Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. *Genomics*. 2016 Jan;107(1):1–8.
- 2 van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet*. 2018 Sep;34(9):666–81.
- 3 Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell*. 2013 Sep;155(1):27–38.
- 4 Lohmann K, Klein C. Next generation sequencing and the future of genetic diagnosis. *Neurotherapeutics*. 2014 Oct;11(4):699–707.
- 5 Lyons E, Sheridan P, Tremmel G, Miyano S, Sugano S. Large-scale DNA Barcode Library Generation for Biomolecule Identification in High-throughput Screens. *Sci Rep*. 2017 Oct;7(1):13899.
- 6 Schöfl G, Lang K, Quenzel P, Böhme I, Sauter J, Hofmann JA, et al. 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned. *BMC Genomics*. 2017 Feb;18(1):161.
- 7 Seltsam A, Doescher A. Sequence-Based Typing of Human Blood Groups. *Transfus Med Hemother*. 2009;36(3):204–12.

- 8 Reid ME. Overview of molecular methods in immunohematology. *Transfusion*. 2007 Jul; 47(1 Suppl):10S–6S.
- 9 Liu Z, Liu M, Mercado T, Illoh O, Davey R. Extended blood group molecular typing and next-generation sequencing. *Transfus Med Rev*. 2014 Oct;28(4):177–86.
- 10 Merriman B, Rothberg JM; Ion Torrent R&D Team. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*. 2012 Dec;33(23):3397–417.
- 11 Marinier E, Brown DG, McConkey BJ. Pollux: platform independent error correction of single and mixed genomes. *BMC Bioinformatics*. 2015 Jan;16(1):10.
- 12 Ambardar S, Gupta R, Trakroo D, Lal R, Vakhlu J. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J Microbiol*. 2016 Dec;56(4):394–404.
- 13 Lane WJ, Westhoff CM, Gleadall NS, Aguad M, Smeland-Wagman R, Vege S, et al.; MedSeq Project. Automated typing of red blood cell and platelet antigens: a whole-genome sequencing study. *Lancet Haematol*. 2018 Jun; 5(6):e241–51.
- 14 Lane WJ, Aguad M, Smeland-Wagman R, Vege S, Mah HH, Joseph A, et al.; MedSeq Project. A whole genome approach for discovering the genetic basis of blood group antigens: independent confirmation for P1 and Xga. *Transfusion*. 2019 Mar;59(3):908–15.
- 15 Ardui S, Ameer A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res*. 2018 Mar;46(5):2159–68.
- 16 Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*. 2015 Oct;13(5):278–89.
- 17 Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol*. 2008 Oct;26(10):1146–53.
- 18 Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*. 2018 Apr;36(4):338–45.
- 19 Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016 Nov;17(11):239.
- 20 Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010 Jun; 7(6):461–5.
- 21 Carapito R, Radosavljevic M, Bahram S. Next-Generation Sequencing of the HLA locus: methods and impacts on HLA typing, population genetics and disease association studies. *Hum Immunol*. 2016 Nov;77(11):1016–23.
- 22 Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013 Jun;10(6):563–9.
- 23 Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin Microbiol Infect*. 2018 Apr;24(4):335–41.
- 24 Mayor NP, Robinson J, McWhinnie AJ, Ranade S, Eng K, Midwinter W, et al. HLA Typing for the Next Generation. *PLoS One*. 2015 May;10(5):e0127153.
- 25 Liu C, Xiao F, Hoisington-Lopez J, Lang K, Quenzel P, Duffy B, et al. Accurate Typing of Human Leukocyte Antigen Class I Genes by Oxford Nanopore Sequencing. *J Mol Diagn*. 2018 Jul;20(4):428–35.
- 26 Olivès B, Merriman M, Bailly P, Bain S, Barnett A, Todd J, et al. The molecular basis of the Kidd blood group polymorphism and its lack of association with type 1 diabetes susceptibility. *Hum Mol Genet*. 1997 Jul;6(7):1017–20.
- 27 Yamamoto F, Hakomori S. Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *J Biol Chem*. 1990 Nov;265(31): 19257–62.
- 28 Yamamoto F, Clausen H, White T, Marken J, Hakomori S. Molecular genetic basis of the histo-blood group ABO system. *Nature*. 1990 May;345(6272):229–33.
- 29 Matassi G, Chérif-Zahar B, Mouro I, Cartron JP. Characterization of the recombination hot spot involved in the genomic rearrangement leading to the hybrid D-CE-D gene in the D(VI) phenotype. *Am J Hum Genet*. 1997 Apr;60(4):808–17.
- 30 Wagner FF, Gassner C, Muller TH, Schonitzer D, Schunter F, Flegel WA. Three molecular structures cause rhesus D category VI phenotypes with distinct immunohematologic features. *Blood*. 1998 Mar;91(6):2157–68.
- 31 Tournamille C, Le Van Kim C, Gane P, Cartron JP, Colin Y. Molecular basis and PCR-DNA typing of the Fya/fyb blood group polymorphism. *Hum Genet*. 1995 Apr;95(4):407–10.
- 32 Wagner FF, Frohmajer A, Flegel WA. RHD positive haplotypes in D negative Europeans. *BMC Genet*. 2001;2(1):10.
- 33 Storry JR, Jöud M, Christophersen MK, Thuresson B, Åkerström B, Sojka BN, et al. Homozygosity for a null allele of SMIM1 defines the Vel-negative blood group phenotype. *Nat Genet*. 2013 May;45(5):537–41.
- 34 Singleton BK, Green CA, Avent ND, Martin PG, Smart E, Daka A, et al. The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D-negative blood group phenotype. *Blood*. 2000 Jan;95(1):12–8.
- 35 Wagner FF, Flegel WA. RHD gene deletion occurred in the Rhesus box. *Blood*. 2000 Jun; 95(12):3662–8.
- 36 Daniels G. Gerbich blood group system. In: Daniels G. *Human Blood Groups*. 3rd ed. Chichester, UK: Wiley-Blackwell; 2013. pp. 410–26.
- 37 Gourri E, Denomme GA, Merki Y, Scharberg EA, Vrignaud C, Frey BM, et al. Genetic background of the rare Yus and Gerbich blood group phenotypes: homologous regions of the GYPC gene contribute to deletion alleles. *Br J Haematol*. 2017 May;177(4):630–40.
- 38 Flegel WA, von Zabern I, Wagner FF. Six years' experience performing RHD genotyping to confirm D- red blood cell units in Germany for preventing anti-D immunizations. *Transfusion*. 2009 Mar;49(3):465–71.
- 39 Murphy MT, Templeton LJ, Fleming J, Ferguson M, Peterkin M, Fraser RH. Comparison of Fy(b) status as determined serologically and genetically. *Transfus Med*. 1997 Jun;7(2):135–41.
- 40 Prager M. Molecular genetic blood group typing by the use of PCR-SSP technique. *Transfusion*. 2007 Jul;47(1 Suppl):54S–9S.
- 41 Finning K, Bhandari R, Sellers F, Revelli N, Villa MA, Muñoz-Díaz E, et al. Evaluation of red blood cell and platelet antigen genotyping platforms (ID CORE XT/ID HPA XT) in routine clinical practice. *Blood Transfus*. 2016 Mar;14(2):160–7.
- 42 Drago F, Karpasitou K, Poli F. Microarray Beads for Identifying Blood Group Single Nucleotide Polymorphisms. *Transfus Med Hemother*. 2009;36(3):157–60.
- 43 Tounsi WA, Madgett TE, Avent ND. Complete RHD next-generation sequencing: establishment of reference RHD alleles. *Blood Adv*. 2018 Oct;2(20):2713–23.
- 44 Dezan MR, Ribeiro IH, Oliveira VB, Vieira JB, Gomes FC, Franco LA, et al. RHD and RHCE genotyping by next-generation sequencing is an effective strategy to identify molecular variants within sickle cell disease patients. *Blood Cells Mol Dis*. 2017 Jun;65:8–15.
- 45 Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*. 2014 Feb;56(2):61–4, 66, 68, passim.
- 46 Lange V, Böhme I, Hofmann J, Lang K, Sauter J, Schöne B, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics*. 2014 Jan;15(1):63.
- 47 Meny GM, Flickinger C, Marcucci C. The American Rare Donor Program. *J Crit Care*. 2013 Feb;28(1):110.e9–18.
- 48 Fichou Y, Audrézet MP, Guéguen P, Le Maréchal C, Férec C. Next-generation sequencing is a credible strategy for blood group genotyping. *Br J Haematol*. 2014 Nov;167(4):554–62.
- 49 Boccoz SA, Fouret J, Roche M, Lachuer J, Legras-Lachuer C, Corgier BP, et al. Massively parallel and multiplex blood group genotyping using next-generation-sequencing. *Clin Biochem*. 2018 Sep;60:71–6.
- 50 MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, et al. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics*. 2018 Jan;19(1):30.
- 51 Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*. 2015 Feb;12(2): 115–21.
- 52 International Society of Blood Transfusion [Internet]. Red Cell Immunogenetics and Blood Group Terminology. [cited 2019 Oct 30]. Available from: <http://www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology/>
- 53 Weinstock C, Mytilineos J, Bugert P, Sitzmann N, Pensel E, Schrezenmeier H, et al. A novel allele of the atypical chemokine receptor 1 (ACKR1) gene containing the nucleotide change c.126 T[GT]G (p.42Glu) encodes a third Duffy blood group protein sequence antithetical to that encoding Fya and Fyb antigens. *Transfusion*. 2019 Jun;59(6):2158–9.

- 54 Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, et al. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res.* 1997 Aug;25(16):3235–41.
- 55 Peng Q, Vijaya Satya R, Lewis M, Randad P, Wang Y. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics.* 2015 Aug;16(1):589.
- 56 Markoulatos P, Siafakas N, Moncany M. Multiplex polymerase chain reaction: a practical approach. *J Clin Lab Anal.* 2002;16(1):47–51.
- 57 Breslauer KJ, Frank R, Blöcker H, Marky LA. Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci USA.* 1986 Jun;83(11):3746–50.
- 58 Poritz MA, Ririe KM. Getting things backwards to prevent primer dimers. *J Mol Diagn.* 2014 Mar;16(2):159–62.
- 59 McDewitt SL, Bredeson JV, Roy SW, Lane JA, Noble JA. HAPCAD: an open-source tool to detect PCR crossovers in next-generation sequencing generated HLA data. *Hum Immunol.* 2016 Mar;77(3):257–63.
- 60 Thompson JR, Marcelino LA, Polz MF. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* 2002 May;30(9):2083–8.
- 61 Odelberg SJ, Weiss RB, Hata A, White R. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Res.* 1995 Jun;23(11):2049–57.
- 62 Chen SM, Ma KY, Zeng J. Pseudogene: lessons from PCR bias, identification and resurrection. *Mol Biol Rep.* 2011 Aug;38(6):3709–15.
- 63 Le van Kim C, Mouro I, Chérif-Zahar B, Raynal V, Cherrier C, Cartron JP, et al. Molecular cloning and primary structure of the human blood group RhD polypeptide. *Proc Natl Acad Sci USA.* 1992 Nov;89(22):10925–9.
- 64 Kudo S, Fukuda M. Structural organization of glycoprotein A and B genes: glycoprotein B gene evolved by homologous recombination at Alu repeat sequences. *Proc Natl Acad Sci USA.* 1989 Jun;86(12):4619–23.
- 65 Silva FC, Torrezan GT, Brianese RC, Stabelini R, Carraro DM. Pitfalls in genetic testing: a case of a SNP in primer-annealing region leading to allele dropout in BRCA1. *Mol Genet Genomic Med.* 2017 May;5(4):443–7.
- 66 Duke JL, Mosbrugger TL, Ferriola D, Chitnis N, Hu T, Tairis N, et al. Resolving MiSeq-generated ambiguities in HLA-DPB1 typing by using the Oxford Nanopore technology. *J Mol Diagn.* 2019 Sep;21(5):852–61.
- 67 Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. *Hum Mol Genet.* 2018 Aug;27 R2:R234–41.
- 68 Jeck WR, Lee J, Robinson H, Le LP, Iafrate AJ, Nardi V. A Nanopore Sequencing-Based Assay for Rapid Detection of Gene Fusions. *J Mol Diagn.* 2019 Jan;21(1):58–69.
- 69 Nakano K, Shiroma A, Shimoji M, Tamotsu H, Ashimine N, Ohki S, et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Hum Cell.* 2017 Jul;30(3):149–61.
- 70 Karamitros T, Magiorkinis G. A novel method for the multiplexed target enrichment of MinION next generation sequencing libraries using PCR-generated baits. *Nucleic Acids Res.* 2015 Dec;43(22):e152.
- 71 Henley RY, Carson S, Wanunu M. Studies of RNA Sequence and Structure Using Nanopores. *Prog Mol Biol Transl Sci.* 2016;139:73–99.
- 72 Mueller A, Fischer K, Suluku R, Hoenen T. Sequencing of mRNA from Whole Blood using Nanopore Sequencing. *J Vis Exp.* 2019 Jun;(148).
- 73 Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A, Judson SD, et al. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg Infect Dis.* 2016 Feb;22(2):331–4.
- 74 Park ST, Kim J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. *Int Neurolog J.* 2016 Nov; 20 Suppl 2:S76–83.
- 75 Orzińska A, Guz K, Mikula M, Kulecka M, Kluska A, Balabas A, et al. A preliminary evaluation of next-generation sequencing as a screening tool for targeted genotyping of erythrocyte and platelet antigens in blood donors. *Blood Transfus.* 2018 May;16(3):285–92.
- 76 Flegel WA, Gottschall JL, Denomme GA. Integration of red cell genotyping into the blood supply chain: a population-based study. *Lancet Haematol.* 2015 Jul;2(7):e282–9.
- 77 Jungbauer C. Molecular Bases and Genotyping for Rare Blood Types. *Transfus Med Hemother.* 2009;36(3):213–8.
- 78 Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet.* 2018 May;19(5):269–85.
- 79 Aloisio M, Licastro D, Caenazzo L, Torboli V, D'Eustacchio A, Severini GM, et al. A technical application of quantitative next generation sequencing for chimerism evaluation. *Mol Med Rep.* 2016 Oct;14(4):2967–74.
- 80 Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. *Front Genet.* 2019 May;10:426.