A preliminary evaluation of next-generation sequencing as a screening tool for targeted genotyping of erythrocyte and platelet antigens in blood donors

Agnieszka Orzińska[1], Katarzyna Guz[1], Michał Mikula[2], Maria Kulecka[3], Anna Kluska[2], Aneta Balabas[2], Monika Pelc-Kłopotowska[1], Jerzy Ostrowski[2,3], Ewa Brojer[1]

[1]Department of Immunohaematology and Immunology of Transfusion Medicine, Institute of Haematology and Blood Transfusion; [2]Department of Genetics, "Maria Sklodowska-Curie" Memorial Cancer Centre and Institute of Oncology; [3]Medical Centre of Postgraduate Education, Department of Gastroenterology, Hepatology and Clinical Oncology, Warsaw, Poland

Background. Matching the compatibility of donor blood with the recipient's antigens prevents alloimmunisation. Next-generation sequencing (NGS) technology is a promising method for extensive blood group and platelet antigen genotyping of blood donors. It circumvents the limitations of detecting known alleles based on predefined polymorphisms and enables targeted sequencing on a massive scale. The aim of this study was to evaluate the NGS AmpliSeq application on the Ion Torrent platform as a screening tool for genotyping blood donors' erythrocyte/platelet antigens.

Materials and methods. Primers for regions encoding antigens RhD (exons 5, 7), Rhc, RhE/e, Fya/b, Jka/b, M/N, S/s, HPA-1, 2, 3, 5, 15 were designed with Ion AmpliSeq Designer with manual inclusion of *RHCE*C* primers. DNA libraries of 57 regular blood donors with determined phenotype/genotype (prepared using the Ion AmpliSeq Library Kit and 14 primer pairs) were sequenced on the Ion Torrent PGM using 316v2 chips and 200 bp chemistry.

Results. Sequencing was successful in all but the MN and HPA-5 regions. Mean sequencing coverage in one experiment was 4,606 reads, except for the *RHCE*C* region (mean 568 reads). NGS results agreed with the known phenotype/genotype of donors except in one phenotypically Fy(a+b−) case in whom *FY*A/FY*B* alleles were found. Reading rates for homozygotes were 97-100%, while they were around 50% for heterozygotes. NGS of *RHD* regions led to identification of mutations in two RhD negative donors.

Discussion. NGS can be performed as a screening test to determine erythrocyte/platelet antigens in blood donors. This method allowed testing of 48 donors for 14 features (200 bp long) with the depth of a few thousand reads simultaneously, and the estimation of natural chimerism or hemi/homozygotic status. NGS screening can be adjusted to the genetic background of a given tested population.

Keywords: next-generation sequencing, blood group, platelet antigens, blood donors.

Introduction

Immunisation to red blood cell and platelet antigens in transfusion recipients can cause haemolytic transfusion reactions or lead to alloimmune thrombocytopenia in the foetus and refractoriness to transfused platelets. It is often a challenge to find blood products for transfusion in alloimmunised patients because of the insufficient number of blood donors with compatible phenotypes[1]. This may cause delays in transfusion and adversely affect patients' health. Faced with such challenges, a significant limitation of current serological techniques is their relatively low testing throughput, since high throughput is required for massive testing of donor phenotypes.

Molecular biology techniques have opened a new era in immunohaematology. Since the 1990s, when the basics of blood cell antigen genotyping emerged, several molecular methods, such as single-strand conformation polymorphism analysis, restriction fragment length polymorphism analysis and real-time polymerase chain reactions (PCR) have been introduced for blood cell antigen genotyping. These were especially useful in reference laboratories when faced with a lack of specific antisera, weakly expressed antigens in test samples or a low quantity/quality of material from the transfusion recipient. It has been shown that DNA-based methods are very accurate and that genotyping precisely predicts the blood group phenotype. Many studies have proven that genotyping can be easier, faster and

probably cheaper than phenotyping for establishing huge registries of typed blood donors[2-5]. It is considered that the introduction of molecular methods for blood group typing in both patients and blood donors, by enabling selection of compatible donors for specific groups of recipients (e.g., transfusion-dependent patients), would prevent alloimmunisation[6].

In recent years, blood group testing platforms adjusted to a high throughput scale have become used for mass screening of donors in blood transfusion centres[7-9]. However, commercially available beads or microarray-based genotyping platforms only detect single nucleotide changes based on the assumption that the sequence next to or around the target single nucleotide polymorphism (SNP)/reference SNP (rs) is known and complementary to the primer/probe used. Thus, a lack of complementarity in primers or probes for cases of novel variants may yield false results. Since new mutations are continually being identified in the regions amplified, testing must be regularly updated to minimise the risk of technical errors[9,10].

Next-generation sequencing (NGS) is a new technology that circumvents these limitations, combining complex sequencing with an enormous capacity for sequencing entire genomes[10,11]. It should be stressed that this technology provides detailed insight into the real nucleotide sequence of the tested gene panel together with a quantitative estimate of reads for a large group of individuals. At present, the platforms used are those which support huge capacity genotyping of antigens at low cost. However, the problems of platforms based on SNP detection include the lack of specificity of both primers and probes for mutations in the amplified region along as well as their inability to detect mutations. Sequencing those regions containing the SNP allows the large-scale throughput to be maintained, while overcoming the lack of specificity in defined primers/probes and enabling the region studied to be viewed. Undoubtedly, the advantages of NGS are its ability to determine the sequence of a person's entire gene encoding red blood cell antigens or human platelet antigens (HPA) and its potential to identify novel variants[12-14]. Nevertheless, such an approach only allows a few people to be studied at any one time in a given experiment, and is thus an ineffective means for screening. From an immunohaematological point of view, NGS of SNP regions only enables the screening of clinically important antigens for many individuals in one test and is a promising method for massive blood group genotyping in blood donors[11,15-17].

In Poland as well as in other countries, NGS technology is now being introduced into laboratories for routine genotyping of human leucocyte antigens of donors for transplantation and is also widely used in screening studies for foetal aneuploidy in pregnant women[18-20]. NGS enables the identification not only of weak phenotypes, variants and new alleles, but also the detection of naturally occurring chimerism as well as providing an estimation of zygosity in only one sequencing run. The HPA genotype can be obtained during the same run as the red cell genotypes, which is important in light of an increasing need for blood products from HPA-typed donors.

The aim of this study was to evaluate the NGS AmpliSeq application on an Ion Torrent platform as a screening tool for simultaneous genotyping of erythrocyte and platelet antigens in blood donors with known phenotypes/genotypes. Such an approach resolves the problem of platforms routinely used for antigen genotyping while maintaining their large-scale throughput.

Materials and methods

Samples

Blood samples from 57 regular blood donors with a determined phenotype/genotype were collected into vacutainer tubes containing EDTA. These donors were from the Institute of Haematology and Transfusion Medicine (IHTM) and Regional Blood Transfusion Centres in Radom, Slupsk, Walbrzych, Gdansk, Poznan and Warsaw. DNA was isolated using the innuPREP Kit (Analytik Jena, Überlingen, Germany). In addition, we tested: (i) an artificial mixture of blood from selected donors imitating chimerism; (ii) unique blood samples from blood donors with natural chimerism previously detected by serological tests; and (iii) unique blood samples from two donors with *RHD* variants found in Polish blood donors[21,22].

Reference genotyping of blood group or platelet antigens for all samples tested was performed using in-house Taqman assays or commercially available tests (RBC-FluoGene vERYfy, HPA-FluoGene, Inno-train, Kronberg, Germany)[5].

Targeted next generation sequencing

Primer design

Primers, complementary to regions encoding antigens RhD (exons 5, 7), Rhc, RhE/e, Fya/b, Jka/b, M/N, S/s, and HPA-1, 2, 3, 5, 15, were designed using the Ion AmpliSeq Designer (Thermo Fisher, Waltham, MA, USA), with manual inclusion of RhC primers (see Table SI in the Online Supplementary Content). Primers for the *RHD* gene originally suggested by the software were modified, in order to avoid amplification of the *RHCE* gene, which is highly homologous to the *RHD* gene. Analogous multiple modifications were made to the originally proposed primers for the identification of the *RHCE*c* allele. To determine the presence of the

*RHCE*C* allele, primers were designed to detect the gene fragment in intron 2 of *RHCE*, which was absent in the reference sequence. The primers were designed based on the sequence AJ633650 from the National Center for Biotechnology Information (NCBI data) containing the insertion in intron 2 of *RHCE*C*.

Library preparation

The library was generated using Ion AmpliSeq Library Kit 2.0 (Thermo Fisher) and 28 designed primers (see Table SI in the Online Supplementary Content) in the multiplex reaction of the amplification donor's DNA. Fourteen amplicons were generated in 18 PCR cycles. Following purification from the primers, adapters and barcodes were added using Ion Xpress Barcode Adapters (Thermo Fisher).

Template preparation

Barcoded libraries underwent automated template preparation with the Ion PGM IC 200 Kit (Thermo Fisher) on the Ion Chef Instrument (Thermo Fisher). DNA library samples were sequenced on a 316 chip v2 using the sequencing reagents provided as part of the Ion PGM IC 200 Kit, according to the manufacturer's instructions. Targeted sequencing for 200 bp was performed using the PGM Ion Torrent (Thermo Fisher).

Data analysis

Sequencing results were automatically processed by Ion Torrent Suite Process software (Thermo Fisher) and aligned to the hg19 reference genome and human *RHCE* gene, intron 2 (sequence AJ633650). Variants were indicated by the Torrent Variant Caller 4.4 with low-stringency parameters for germline variants. In addition, results were visually inspected with the browser for genomic IGV to confirm the presence of variants for the amplicons in the samples. The proportion of each base for a given position was confirmed by bam-readcount (GitHub, Inc., San Francisco, CA, USA)[23]. The presence of an insert was determined by computing the proportion of reads with mapping quality (to sequence AJ633650) of 20 or higher which corresponded to the insert region (between bases 625 and 730 of the reference sequence). Detailed NGS results obtained for each donor include the number of sequence reads corresponding to a given polymorphism in relation to the total number of detected sequences in the tested region.

Results

Using the designed primer set, we sequenced the polymorphic regions of alleles determining phenotypes RhD, RhE/e, Rhc, RhC, Kk, Jk, Fy, Ss, HPA-1a, HPA-2, HPA-3, HPA-5, and HPA-15 in the donors. The region encoding antigens of the MN blood group were not sequenced, probably because of the poor specificity of the primers used for that genomic region.

We initially assessed the performance of our 14 fragments by using the AmpliSeq assay on nine donors. The median coverage was 20,788 reads (range 450 to 51,000) and the frequency of non-specific reads in the tested rs position ranged from 0 to 0.32% for individual nucleotides.

Sequencing results were fully consistent with the known phenotype/genotype of the donors. For two RhD negative donors with silent *RHD* gene variants (S256X and Y343X), which had previously been found in the Polish population, the sequencing indicated a mutation in the polymorphic position of the *RHD* gene (in exon 5 rs780356470 c.767C>G and in exon 7 rs771554591 c.1029C>A)[22].

The results showed a clear discrimination of homozygotes and heterozygotes for the analysed blood group systems and platelet antigens. Read rates for homozygotes ranged from 99.6 to 100%, while those for heterozygotes were approximately 50% (range, 44.21 to 55.66%, depending on the tested region). Table I details exemplary sequencing of two donors with homozygous and heterozygous genotypes. The results of the NGS chimerism analysis of an artificial mixture of these donors' DNA are also shown; chimerism ranged from 1 to 8%, depending on the region analysed.

Knowing that our assay could amplify 14 examined regions and call the variants, we next simultaneously sequenced 48 samples in a single PGM run on a 316 v2 chip. The depth of sequencing ranged from 152 to 11,692 reads. For the HPA-5 region, the coverage was too low for the genotype to be correctly interpreted (Table II). The contribution of non-specific reads ranged from 0 to 0.82% for the individual nucleotides.

The sequencing results agreed with the known phenotype/genotype of the donors in all cases except one; a donor phenotyped as Fy(a+b−) had *FY*A* and *FY*B* alleles. Further investigations of this donor's DNA confirmed the presence of *FY*X*, a weak variant of the Fyb antigen.

The level of coverage achieved unambiguously allowed discrimination between homozygotes and heterozygotes for the analysed blood group or platelet systems. The read rates for homozygotes ranged from 97 to 100%, while those for heterozygotes were about 50% (range, 30.59 to 69.25%) depending on the tested region, excluding the HPA-5 data (Table SII in the Online Supplementary Content).

In one of the male donors, naturally occurring chimerism was detected using serological tests, with about 20% of his erythrocytes containing RhD protein[21]. In this case, the number of sequence reads for the *RHD* gene, when compared with the mean number of

Table I - Examples of NGS results for two donors and an artificial mixture of their DNA (3% chimerism).

Antigen	rs number	SNP	Donor n. 1			Donor n. 2			Mixture of donor n. 1 with 3% donor n. 2		
			Phenotype/ genotype	*Number of reads*	*Nucleotide: quantity (%) detected in rs position*	*Phenotype/ genotype*	*Number of reads*	*Nucleotide: quantity (%) detected in rs position*	*Phenotype/ genotype*	*Number of reads*	*Nucleotide: quantity (%) detected in rs position*
RhD−/+	rs146093871	_/G	D+	32,304	A: 0.07; C:0.01; G:99.91; T: 0.01	D−	0		D+	23,839	A: 0.13; G: 99.86; T: 0.01
RhD−/+	rs3118454	_/G	D+	38,138	A: 0.04; G: 99.92; T: 0.00	D−	0		D+	25,785	A: 0.03; C: 0.00; G: 99.95
Rhc−/+	rs1053343	_/T	C+c−	0		C−-c+	30,798	C: 0.06; G: 0.04; T: 99.90	C+c− with 3% c+	1,515	C: 0.33; T: 99.67
RhE/e	rs609320	G/C	E−e+	31,557	C: 99.94; T: 0.06	E−e+	45,775	A: 0.00; C: 99.93; G: 0.00; T: 0.07	E−e+	22,771	C: 99.98; T: 0.02
Kk	rs8176058	A/G	K−k+	32,658	A: 0.04; C: 0.00; G: 99.96; T: 0.00	K+k+	46,916	A: 49.48; C: 0.01; G: 50.45; T: 0.05	K−k+ with 3% K+	23,716	A: 3.03; G: 96.96; T: 0.00
JK a/b	rs1058396	G/A	Jk(a−b+)	39,051	A: 99.72; C: 0.01; G: 0.26; T: 0.01	Jk(a+b+)	49,743	A: 48.34; C: 0.00; G: 51.65; T: 0.01	Jk(a−b+) with 3% Jka+	34,591	A: 91.79; C: 0.00; G: 8.20; T: 0.00
FY a/b	rs12075	G/A	Fy(a+b+)	34,504	A: 53.02; G: 46.97; T: 0.01	Fy(a+b−)	48,345	A: 0.05; G: 99.95	Fy(a+b+)	27,404	A: 47.35; C: 0.00; G: 52.63; T: 0.02
MN	rs7682260	C/T	M+N−	0		M−N+	0		M+N−	0	
Ss	rs7683365	A/G	S+s+	31,031	A: 55.66; C: 0.01; G: 43.33; T: 0.05	S−s+	46,673	A: 0.07; C: 0.02; G: 99.88; T: 0.03	S+s+	26,399	A: 44.21; C: 0.03; G: 55.01; T: 0.05
HPA-1 a/b	rs5918	T/C	HPA-1 b/b	50,748	C: 99.68; G: 0.17; T: 0.07	HPA-1 a/b	47,181	A: 0.01; C: 47.56; G: 0.08; T: 52.27	HPA-1 b/b with 3% HPA1a+	31,223	C: 98.17; G: 0.21; T: 1.54
HPA-2 a/b	rs6065	C/T	HPA-2 a/a	39,876	C: 99.97; G: 0.01; T: 0.02	HPA-2 a/a	44,251	C: 99.97; G: 0.01; T: 0.02	HPA-2 a/a	29,000	C: 99.94; G: 0.00; T: 0.05
HPA-3 a/b	rs5911	A/C	HPA-3 a/b	17,502	A: 55.19; C: 44.67; G: 0.05; T: 0.09	HPA-3 a/b	16,723	A: 58.39; C: 41.45; G: 0.07; T: 0.09	HPA-3 a/b	12,445	A: 57.97; C: 41.83; G: 0.08; T: 0.12
HPA-5 a/b	rs1801106	G/A	HPA-5 a/a	5,683	A: 0.30; G: 99.70	HPA-5 a/a	7,594	A: 0.05; G: 99.95	HPA-5 a/a	6,194	A: 0.02; G: 99.97
HPA-15a/b	rs10455097	C/A	HPA-15a/a	17,690	A: 0.01; C: 99.95; G: 0.01; T: 0.02	HPA-15a/a	21,884	A: 0.01; C: 99.96; T: 0.02	HPA-15a/a	10,501	A: 0.04; C: 99.89; T: 0.05

NGS: next-generation sequencing; SNP: single nucleotide polymorphism.

Table II - Number of sequenced reads for homozygotes and heterozygotes and nucleotide percentage detected in rs positions for each antigen tested in 48 donors.

Antigen	Phenotype/ genotype	N. of donors tested	Range of number of reads (median; mean; SD)	Range of nucleotide percentages detected in rs position (%)			
				G	*A*	*T*	*C*
RhD exon 5	D+	29	225-7,783 (3,246; 3,418; 1,644)	99.64-100.00	0.00-0.28	0.00-0.44	0.00-0.02
	D−	16	0-1 (0;0;0)	0.00-100.00	0.00	0.00	0.00
RhD exon 7	D+	29	338-8,188 (3,712; 3,946; 1,882)	99.78-100.00	0.00-0.11	0.00-0.13	0.00-0.02
	D−	16	0-4 (0,0,0)	0.00	0.00	0.00	0.00
RhC	CC	14	174-1,577 (689; 723; 349)	Whole sequence mapped on reference AJ633650			
	Cc	10	191-779 (447; 424; 202)				
	cc	21	0-12 (0; 1; 3)				
Rhc	CC	13	0-5 (1;1;1)	0.00	0.00	0.00	0.00
	Cc	10	1,524-5,694 (2,228; 2,591; 729)	0.00-0.07	0.00-0.05	99.48-99.95	0.05-0.49
	cc	22	255-7,651 (4,994; 4,709; 1,419)	0.00-0.39	0.00-0.04	99.61-99.94	0.00-0.29
K/k	KK	4	865-10,120 (5,606; 5,549; 2,937)	0.02-0.15	99.64-99.78	0.06-0.30	0.00
	Kk	5	3,955-6,391 (5,528; 5,357; 726)	45.91-49.77	50.14-54.07	0.02-0.11	0.00
	kk	36	520-8,784 (4,852; 5,042; 1,507)	99.80-100.00	0.00-0.19	0.00-0.10	0.00-0.03
RhE/e	EE	5	3,929-6,611 (5,392; 5,338; 600)	99.85-99.92	0.00-0.06	0.00-0.03	0.02-0.06
	Ee	4	3,958-6,127 (4,691; 4,866; 630)	46.99-52.46	0.00-0.04	0.03-0.15	47.46-52.78
	ee	36	675-8,781 (5,457; 5,332; 1,719)	0.00-0.14	0.00-0.05	0.00-0.16	99.84-100.00
Jk	aa	14	699-7,528 (5,875; 5,005; 1,846)	99.58-100.00	0.00-0.42	0.00-0.08	0.00
	ab	18	1,320-11,692 (6,742; 7,219; 1,757)	48.07-54.49	45.51-57.00	0.00-0.08	0.00
	bb	13	2,731-10,572 (7,636; 7,023; 2,068)	0.11-0.53	99.42-99.89	0.00-0.08	0.00-0.04
Fy	aa	11	656-7,915 (5,397; 5,135; 1,898)	99.84-99.95	0.03-0.16	0.00-0.03	0.00-0.01
	ab	16	4,044-11,675 (6,561; 6,761; 1,616)	45.99-53.61	46.40-53.98	0.00-0.03	0.00-0.05
	bb	18	1,017-9,348 (5,116; 5,368; 1,868)	0.16-0.35	99.59-99.82	0.00-0.08	0.00-0.05
Ss	SS	12	384-6,005 (2,868; 2,615; 1,237)	0.00-0.82	95.17-99.00	0.00-0.52	0.00-0.20
	Ss	16	782-8,303 (3,880; 4,049; 1,174)	49.87-64.59	35.79-49.52	0.00-0.15	0.00-0.08
	ss	17	1,796-7,490 (3,525; 3,900; 1,069)	99.77-99.97	0.17-0.17	0.00-0.17	0.00-0.12
HPA-1	aa	25	955-10,859 (6,296; 6,406; 1,880)	0.00-0.07	0.00-0.05	99.67-99.93	0.07-0.25
	ab	13	1,231-8,406 (5,180; 5,207, 1,399)	0.12-0.51	0.00-0.05	48.59-53.13	46.54-51.20
	bb	7	599-8,610 (6,001; 5,587; 240)	0.00-0.73	0.00-0.02	0.00-0.19	98.97-99.83
HPA-2	aa	33	563-10,036 (6,312: 5,949; 1,815)	0.00-0.05	0.00-0.09	0.00-0.36	99.64-100
	ab	10	1,611-9,477 (6,290; 6,729; 1,958)	0.00-0.05	0.00-0.02	48.48-53.49	46.51-51.51
	bb	2	5,528-6,694 (6,111; 6,111; 583)	0.04-0.06	0.00-0.03	99.72-99.78	0.14-0.19
HPA-3	aa	14	1,592-3,734 (2,517; 2,679; 483)	0.00-:0.16	99.76-100	0.00-0.05	0.00-0.05
	ab	20	433-4,749 (1,980; 2,425; 1,025)	0.00-0.24	48.62-61.42	0.00-0.32	38.39-51.21
	bb	11	236-5,075 (2,438; 2,592; 1,018)	0.00-0.38	0.00-0.04	0.00-0.47	99.38-100.00
HPA-5	aa	36	11-792 (99; 167; 127)	99.85-100.00	0.00-0.15	0.00	0.00
	ab	7	56-183 (78; 103; 44)	48.21-67.12	31.51-48.21	0.00	0.00
	bb	2	99-105 (102; 102; 3)	0.95-1.01	96.97-99.05	0.00	0.00
HPA-15	aa	10	853-6,114 (2,636; 2,747; 1,151)	0.00-0.04	0.00-0.07	0.00-0.11	99.82-100.00
	ab	25	152-5,223 (2,240; 2,243; 1,086)	0.00-0.30	30.59-47.16	0.00-0.11	52.78-69.25
	bb	10	447-2,936 (1,483; 1,634; 820)	0.00-0.22	99.35-100.00	0.00-0.04	0.00-0.43

SD: standard deviation.

sequence reads from other regions, showed that the *RHD* sequence was present in approximately 24%, confirming serological testing results previously measured by flow cytometry.

Discussion

The aim of our study was to develop a protocol which can be used as a screening test for reliable determination of clinically important blood group or platelet antigens in blood donors. We chose a target enrichment NGS protocol, as the best method, since this enables genotyping of a large number of donors simultaneously. We designed an assay for sequencing regions with SNP encoding 14 clinically important blood group or platelet antigens using the Ion Torrent PGM Sequencer and AmpliSeq technology. We managed to detected the rs positions responsible for forming all but one blood group system (M/N antigens) simultaneously in all samples investigated. Our results were in full accordance with the genotypes determined using allelic discrimination PCR but they also gave the full sequence of nucleotides. In a single run, 48 donors were tested for the 14 antigen features, meaning that 700 genotyping results were obtained with a full 200 nucleotide sequence for each gene and its quantitative estimation. A significant amount of data was generated from the designed experiments, providing full information on the sequenced region and any polymorphic variations. The number of mapped reads varied between samples and between sequenced regions, from a 100 reads for the *ITGA2* region to 50,000 for *ITGAB3*. For some regions (Table II) the depth of sequencing was not sufficient to obtain reliable results. Some unique cases from the archived material also showed a lower number of reads. Further refinements of primer sequences in poor quality amplicons could overcome these shortcomings.

Obviously, antigen genotyping platforms currently used in immunohaematological laboratories can achieve similarly high throughputs as those possible with NGS, but they lack the ability to look into tested fragments, which is especially important in genotyping specific ethnic groups[7-9,24,25]. This may lead to false-positive or false-negative results, which are unacceptable when typing blood donors. Using TaqMan chemistry and an OpenArray real-time PCR system (Thermo Fisher), Flegel *et al*. obtained 32/89,596 discrepant results, caused by the presence of variant alleles[7]. Boccoz *et al*. found a range of discrepancy of 0.17 to 1.7% in their blood group typing when using HIFI technology[26]. Platforms require continuous modifications of the versions of tests used in order to provide broader coverage. Delaney *et al.* showed that 4.3 to 4.7% of donors tested using the HEA BeadChip Kit (Bioarray Solution Ltd., Warren, NJ, USA) had a discrepancy between SNP and serology findings, depending on the version used of the test for certain ethnic groups with unknown genetic variants[24]. The advantage of NGS is that it goes beyond the detection of SNP encoding known alleles. Our study was, however, focused on identifying nucleotide changes only in the region amplified around the antigen SNP to circumvent the limits of the tests currently used while maintaining the high throughput scale of genotyping.

The NGS assay we developed identified clinically important HPA regions. Similarly, Avent *et al*., in their experiments of sequencing whole blood group genes, included all HPA alleles in the tested panel for blood group sequencing[17]. Given the increasing requests for blood products from HPA-typed donors, the HPA genotype should be obtained together with blood group genotypes for regular blood donors. The IHTM has been creating a registry of available blood donors with determined platelet antigens. The increasing rates of foetal/neonatal alloimmune thrombocytopenia identified in Poland, in connection with the PREVFNAIT HPA-1a screening programme, have caused a higher demand for HPA-typed products[27,28]. In our work, the range of sequencing coverage was correct for HPA-1, 2, 3 and 15, whereas the number of reads for the HPA-5 region was low using low depth sequencing. This problem is probably related to the fact that the sequence of the HPA-5 region has repeats of 13 adenines that made the correct sequencing and interpretation of the data produced by the Ion Torrent PGM Sequencer difficult.

We, like others, encountered difficulties in designing the pool of primers for the NGS experiments[14,16,29,30]. We initially used the Ion AmpliSeq Designer to design the primers for all regions. However, since this tool had not checked the primers' specificity, it was proposed that non-specific primers be used for highly homologous genes such as *RHD* and *RHCE*. We, therefore, prepared gene-specific primers to generate libraries including regions from exons 5 and 7 of the *RHD* gene. Similar observations have been reported regarding NGS of *RH* genes. Fichou *et al.* applied specific primers and generated libraries separately for *RHD* and *CE* genes and Avent *et al.* applied long-range PCR amplification of *D* and *CE* genes with specific primers before using NGS as a standard procedure to circumvent non-specific reactions[15,17]. Additionally, our primers were designed to cover particular regions of *RHD* thus allowing identification of non-expressed D variants, found in our previous study of 32,000 Rh-negative Polish donors[22]. The *RHD* regions flanked by the designed primers contain positions 767 in exon 5 and 1029 in exon 7 which were analysed in order to exclude a silent *D* gene. In our experiment we included samples from those *RHD*-positive but RhD-negative donors and confirmed the coverage of these positions. From the screening point of

view it was important to select *RHD* regions matching the molecular background of our population. Obviously, the *RHD* genotyping based on these two fragments enables detection of the majority of *RHD* variants known to express D antigens but it also produces amplicons from non-functional *RHD* alleles, falsely classifying such donors as RhD-positive.

For Rhc, RhE and Rhe and other tested antigens we used primers proposed by the software, but we were unable to sequence the rs positions responsible for forming M/N antigens. After the failure to detect MN, we checked the complementarity of our primers using software on the NCBI website and found that the region with the MN polymorphism in the *GYPA* gene is identical to highly homologous genes: *GYPB* and pseudogene *GYPE*. So, using the proposed MN primers, complementary to highly duplicated regions, probably made the sequencing and analysis impossible.

For *RHCE*C*, we based our analysis on the insertion in intron 2 of the gene. The design was complicated, since the reference genome hg19 comes from a Rh(C−c+) individual and does not include the insertion, thus primers were designed for the non-reference target AJ633650 sequence. In the NGS analysis presented here, we succeeded in specific detection of *RHD vs RHCE* genes by testing sequences of *RHD* exon 5 and 7. The assay detected the *RHD* gene specifically, regardless of the presence of other RhCE antigens.

In this study, we tested both randomly selected and unique samples. The former cases included a donor with a known Fy(a+b−) serotype, in whom the *FY*B* allele was found. The presence of the weak Fy^b variant was then confirmed through additional phenotyping and genotyping. The result indicates that in future we should include this SNP region (*FY*X*) in our panel for antigen screening, all the more so as the weak Fy^b antigen occurs at quite high frequency in our population[31]. Laboratories that use genetic methods report a high incidence of allelic variations expressing weak phenotypes which are not detected when using routine serological tests[12-14]. Avent *et al.* established a NGS protocol on the Ion Torrent PGM for 11 blood group genes where, in a relatively small cohort, they found several new alleles causing weakened or silenced antigen expression (Jk^a or Fy^b)[17].

Since NGS technology is widely applied for measuring the fraction of foetal material in maternal plasma DNA during pregnancy, we decided to use the technology to determine the level of chimerism which may occur naturally at the blood cell level[21,32]. For this NGS experiment, we analysed a donor sample that had been sent to our reference laboratory for further advanced studies because of difficulties in Rh phenotyping. Using cytometric and molecular methods, the donor was found to have naturally occurring chimerism in RhD and RhC antigens[21]. The NGS test provided the same results in the blood group screening mode and for quantifying the level of chimerism. To our knowledge this is the first case of estimating natural chimerism in blood groups using NGS technology.

Conclusions

In summary, NGS can be performed as a screening test for determining erythrocyte and platelet antigens in blood donors. The targeted multiplex amplification method allowed testing of 48 donors for about 14 features (200 bp long) with a depth of a few thousand reads simultaneously, as well as estimates of natural chimerism or hemi/homozygotic status. The screening protocol can be adapted to the genetic background of the tested population. Specially dedicated software does, however, need to be developed to analyse blood gene sequences before routine use in the diagnostic laboratory.

Acknowledgements

We would like to express our sincere thanks to the Directors of the Regional Blood Transfusion Centres in Gdansk, Poznan, Radom, Slupsk, Walbrzych and Warsaw and their staff for providing blood samples for our study, as well as for their excellent and fruitful cooperation.

Funding and resources

This study was supported by a Ministry of Health grant.

Authorship contributions

AO and KG designed the panel of rs and of blood donors for sequencing, analysed the data and wrote the manuscript. MM supervised the design of the primer pool and performance of the study and participated in data discussion. MK analysed the data. AK and AB performed the study and analyses. MP-K provided serological data. JO and EB revised the manuscript.

The Authors declare no conflicts of interest.

References

1) Evers D, Middelburg RA, de Haas M, et al. Red-blood-cell alloimmunisation in relation to antigens' exposure and their immunogenicity: a cohort study. Lancet Haematol 2016; **3**: 284-92.
2) Peyrard T. Molecular tools for investigating immunohematology problems. ISBT Science 2015; **10**: 31-8.
3) Wieckhusen C, Rink G, Scharberg EA, et al. Molecular screening for Vel- blood donors in Southwestern Germany. Transfus Med Hemother 2015; **42**: 356-60.
4) Orzińska A, Guz K, Michalewska B, Brojer E. Molecular screening of the C^w antigen for typing donors compatible with patients with anti-MAR-like antibodies. Blood Transfus 2015; **14**: 573-6.
5) Ficko T, Galvani V, Rupreht R, et al. Real-time PCR genotyping of human platelet alloantigens HPA-1, HPA-2,

HPA-3 and HPA-5 is superior to the standard PCR-SSP method. Transfus Med 2004; **14**: 425-32.
6) Zimring JC, Welniak L, Semple JW, et al. Current problems and future directions of transfusion-induced alloimmunization: summary of an NHLBI working group. Transfusion 2011; **51**: 435-41.
7) Flegel WA, Gottschall JL, Denomme GA. Implementing mass-scale red cell genotyping at a blood center. Transfusion 2015; **55**: 2610-5.
8) McBean RS, Hyland CA, Flower RL. Blood group genotyping: the power and limitations of the Hemo ID Panel and MassARRAY platform. Immunohematology 2015; **31**: 75-80.
9) Finning K, Bhandari R, Sellers F, 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; **14**: 160-7.
10) Liu Z, Liu M, Mercado T, et al. Extended blood group molecular typing and next-generation sequencing. Transfus Med Rev 2014; **28**: 177-86.
11) Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 2012; **13**: 341.
12) Tilley L, Grimsley S. Is next generation sequencing the future of blood group testing? Transfus Apher Sci 2014; **50**: 183-8.
13) Mori A, Deola S, Xumerle L, et al. Next generation sequencing: new tools in immunology and hematology. Blood Res 2013; **48**: 242-9.
14) Lane WJ, Westhoff CM, Uy JM, et al. Comprehensive red blood cell and platelet antigen prediction from whole genome sequencing: proof of principle. Transfusion 2016; **56**: 743-54.
15) Fichou Y, Audrézet MP, Guéguen P, et al. Next-generation sequencing is a credible strategy for blood group genotyping. Br J Haematol 2014; **167**: 554-62.
16) Altayar A, Kiernan M, Reynolds AJ, Li X. Extensive genotyping of blood groups Duffy Kidd and ABO by next generation sequencing. Vox Sang 2014; **107**: 187.
17) Avent ND, Madgett TE, Halawani AJ. Next-generation sequencing: academic overkill or high-resolution routine blood group genotyping? ISBT Science Series 2015; **10**: 250-6.
18) Erlich H. HLA DNA typing: past, present, and future. Tissue Antigens 2012; **80**: 1-11.
19) Chitty LS, Bianchi DW. Next generation sequencing and the next generation: how genomics is revolutionizing reproduction. Prenat Diagn 2015; **35**: 929-30.
20) Taylor-Phillips S, Freeman K, Geppert J, et al. Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis. BMJ Open 2016; **6**: e010002.
21) Guz K, Orzińska A, Walenciak J, et al. Evaluation of RHD and RHCE*C mosaicism/chimerism by real-time PCR in a blood donor with mixed-field agglutination in Rh phenotyping. Vox Sang 2013; **105**: 235.
22) Orzińska A, Guz K, Polin H, et al. RHD variants in Polish blood donors routinely typed as D−. Transfusion 2013; **53**: 2945-53.
23) GitHub. BAM read count. Available at: https://github.com/genome/bam-readcount. Accessed on 28/04/2016.
24) Delaney M, Harris S, Haile A, et al. Red blood cell antigen genotype analysis for 9087 Asian, Asian American, and Native American blood donors. Transfusion 2015; **55**: 2369-75.
25) Fichou Y, Mariez M, Le Maréchal C, Férec C. The experience of extended blood group genotyping by next-generation sequencing (NGS): investigation of patients with sickle-cell disease. Vox Sang 2016; **111**: 418-24.
26) Boccoz S, Le Goff G, Mandon C, et al. Development and validation of a fully automated platform for extended blood group genotyping. J Mol Diagn 2016; **18**: 144-52.
27) Uhrynowska M, Dębska M, Guz K, et al. [PREVFNAIT prevention of foetal/neonatal alloimmune thrombocytopenia (FNAIT) in Polish foetuses and newborns--the PREVFNAIT program]. Ginekol Pol 2015; **86**: 62-6. [In Polish.]
28) Brojer E, Husebekk A, Dębska M, et al. Fetal/neonatal alloimmune thrombocytopenia: pathogenesis, diagnostics and prevention. Arch Immunol Ther Exp 2016; **64**: 279-90.
29) Stabentheiner S, Danzer M, Niklas N, et al. Overcoming methodical limits of standard RHD genotyping by next-generation sequencing. Vox Sang 2011; **100**: 381-8.
30) Wienzek-Lischka S, Krautwurst A, Fröhner V, et al. Noninvasive fetal genotyping of human platelet antigen-1a using targeted massively parallel sequencing. Transfusion 2015; **55**: 1538-4.
31) Daniels G. *Human Blood Groups*. 3rd ed. Oxford: Wiley-Blackwell; 2013.
32) Eikmans M, van Halteren AG, van Besien K, et al. Naturally acquired microchimerism: implications for transplantation outcome and novel methodologies for detection. Chimerism 2014; **5**: 24-39.

Arrived: 8 September 2016 - Revision accepted: 16 November 2016
Correspondence: Agnieszka Orzińska
Institute of Haematology and Blood Transfusion
5st Chocimska Street
00-791 Warsaw, Poland
e-mail: aorzinska@ihit.waw.pl