SDU
University of Southern Denmark

The use of next-generation sequencing for the determination of rare blood group genotypes

Jakobsen, M. A.; Dellgren, C.; Sheppard, C.; Yazer, M.; Sprogøe, U.

Published in:
Transfusion Medicine

DOI:
10.1111/tme.12496

Publication date:
2019

Document version:
Accepted manuscript

Citation for pulished version (APA):
Jakobsen, M. A., Dellgren, C., Sheppard, C., Yazer, M., & Sprogøe, U. (2019). The use of next-generation sequencing for the determination of rare blood group genotypes. *Transfusion Medicine*, *29*(3), 162-168. https://doi.org/10.1111/tme.12496

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 23. May. 2022

Jakobsen Marianne (Orcid ID: 0000-0003-2352-9276)
Yazer Mark (Orcid ID: 0000-0001-5937-3301)

The Use of Next Generation Sequencing for the Determination of Rare Blood Group genotypes.

Marianne Antonius Jakobsen[1,2], Christoffer Dellgren[1], Chelsea Sheppard[3], Mark Yazer,[2,4] and Ulrik Sprogøe[1]

1. **Department of Clinical Immunology, Odense University Hospital.**
2. **University of Southern Denmark, Denmark.**
3. **Virginia Blood Services, Richmond, US.**
4. **Department of Pathology, University of Pittsburgh, Pittsburgh, US.**

Corresponding author:

Marianne A. Jakobsen
Department of Clinical Immunology
Odense University Hospital
J.B. Winsløws Vej 4
5000 Odense C
Denmark
Mail: marianne.jakobsen@rsyd.dk

Running title: NGS used for Genotyping of Blood Group Antigens

Author Manuscript

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tme.12496

This article is protected by copyright. All rights reserved.

Author Manuscript

Abstract

Objectives Next generation sequencing (NGS) for the determination of **rare** blood group genotypes was tested in 72 individuals from different ethnicities.

Background Traditional serological-based antigen detection methods, as well as genotyping based on specific single nucleotide polymorphisms (SNP) or single nucleotide variants (SNV), are limited to detecting only a limited number of known antigens or alleles. NGS methods do not have this limitation.

Methods/materials NGS using Ion torrent PGM was performed with a customized Ampliseq panel targeting 15 different blood group systems on 72 blood donors of various ethnicities **(Caucasian, Hispanic, Asian, Middle Eastern, and Black).**

Results Blood group genotypes for 70 out of 72 samples could be obtained for 15 blood group systems in one step using the NGS assay and for common SNPs are consistent with previously determined genotypes using commercial SNP assays. However, particularly for the Kidd, Duffy and Lutheran blood group systems several SNV were detected by the NGS assay that revealed additional coding information compared to other methods. Furthermore, the NGS assay allowed for the detection of genotypes related to VEL, Knops, Gerbich, Globoside, P1PK and Landsteiner-Wiener blood group systems.

Conclusion The NGS assay enables a comprehensive genotype analysis of **many** blood group systems and is capable of detecting common and rare alleles, including alleles not currently detected by commercial assays.

Introduction

Blood group antigens are variations on red blood cell (RBC) membrane surface structures including proteins, glycoproteins and glycolipids. Alloimmunization due to transfusion of red cells carrying

This article is protected by copyright. All rights reserved.

Author Manuscript

mismatched blood group antigens between donor and recipient can impede the rapid provision of compatible RBCs and can sometimes lead to life-threatening events including hemolytic disease of the fetus and newborn, and hemolytic transfusion reactions. Therefore, in highly transfused patients, ensuring antigen fidelity between donor and recipient can be an important means of enhancing the safety of transfusions. Currently, 36 blood groups systems have been described (Keller 2015; Reid 2012). Similarly, the genes and the genetic basis for the expression of most blood group antigens are known, with the most common motif of genetic variation within a system being single nucleotide polymorphisms (SNPs) in the coding region (Reid 2012). The most common method for the detection of blood group antigens is hemagglutination, which is an inexpensive and widely used technique. However, this technique is limited by the availability of reagents and can be problematic to perform in recently transfused patients. These limitations can be overcome by blood group antigen genotyping techniques. The majority of the routinely employed genetic methods detects only the most common SNPs, and are not designed to detect novel polymorphisms. As a consequence, new or uncommon single nucleotide variants (SNV) are left undetected, which can lead either to incorrect phenotype interpretation or confounding of the antibody identification process (Goldman *et al.* 2015; Keller 2015; Paccapelo *et al.* 2015; Reid 2009). By contrast, Next Generation Sequencing (NGS) enables simultaneous sequencing of all relevant genes in one or more blood group systems in one step, without the limitation of only detecting previously identified SNPs, thereby facilitating comprehensive blood group genotyping including rare polymorphisms.

In this study, an NGS assay that analyzes the genes underlying 15 blood group systems, excluding ABO, MNS and Rh, was utilized to determine the genotypes of blood donors from different ethnicities.

Materials and methods

DNA Samples

Deoxyribonucleic acid (DNA) samples from 46 American blood donors, 18 Danish blood donors or patients, and from 8 External Quality Assessment (EQA)(Instand e.V.; Germany, 5 samples; NEQAS, 3 samples) samples were studied. The samples were selected from EQA distributions (n=8) or persons of different ethnic descent (Caucasian (19), Hispanic (9), Asian (4), Middle Eastern (6) or Black (26)) without taking any previously determined genotypes into consideration except the sample that was VEL-. The NGS

This article is protected by copyright. All rights reserved.

Author Manuscript

investigators were blinded to the donor's ethnicity and previously determined genotypes until all of the NGS sequencing and analysis had been performed.

Preparing NGS libraries and Sequencing

A customized Ampliseq panel based on multiplex PCR for use with Ion Torrent PGM was designed to cover the complete coding regions and intron splice sites of 13 blood group system genes and to partially cover the coding regions of the CR1 and C4 system genes (table 1). The design also covers part of the promoter region of FY defining the GATA site. *RHD*, *RHCE*, *GYPA* and *GYPB* genes were intentionally omitted from this analysis due to the high level of homology within their respective systems (RH and MNS). With Sanger sequencing it is possible to generate sequence analysis that is specific for each gene by placing primers at non-homology intronic sites. AmpliSeq does not take the presence of homologous genes into account during primer design and therefore RH and MNS genotyping would be based on an estimation of how the detected SNV and SNPs should group together rather than by a clear resolution of the correct genotype. DNA samples were diluted 1:10 in molecular grade water and the DNA content was quantified by spectrophotometer (Qubit Fluorometer, Life Technologies, Denmark) using the Qubit dsDNA HS Assay kit (Life technologies, Denmark) followed by dilution to a final DNA concentration of approximately 5 ng/µl. For each sample, a library based on the AmpliSeq panel was generated with the Ion AmpliSeq Library Kit 2.0 (Life Technologies) according to the manufacturer's instructions. The contents of DNA of libraries were quantified using either the Bioanalyzer 2000 (Agilent, Denmark) or Qubit dsDNA HS Assay kit (Life technologies, Denmark) and diluted to 100 pM before pooling. Emulsion PCR and enrichment were carried out on Ion OneTouch 2 System (Life Technologies) using Hi-Q OT2 Kit (Life Technologies) set at 400 bp mode. After collection, beads were loaded onto an Ion 314 Chip or 316 Chip (Life Technologies) and sequenced on an Ion PGM using the Ion PGM Hi-Q Sequencing kit (Life Technologies).

Next generation sequencing data analysis

The NGS data analysis was performed by a bioinformatician and was done as follows: an initial selection of passed samples was based on the general number of reads per amplicon (called coverage). Only samples with general coverage >40 were used for further analysis. Targeted regions were extracted from the designed Browser Extensible Data (BED) file generated in the Ion Ampliseq™ Designer system. Signal processing and sequence generation including adaptor trimming were performed with the Torrent Suite

This article is protected by copyright. All rights reserved.

version 5.05 (Life Technologies) software and exported as unaligned Fastq files. Alignment to the reference genome (hg19), quality control, variant calling and variant annotation was performed in CLC Biomedical Workbench software version 3.5.2 (QIAGEN Aarhus A/S, Denmark). All alleles are named according to the nomenclature used in The Blood Group Antigen Facts Book, 3rd edition (Reid 2012).

ID Core XT.

ID Core XT (Progenika, Spain) is a microarray, Luminex-based assay designed to detect common SNPs in the blood group genes (number of alleles assayed by ID Core XT), *KEL* (6), *JK* (4), *FY* (4), *DI* (2), *DO* (4), *AQP1* (2), *YT* (2) and *LU* (2). The *MNS* and *RHCE* genes are also assayed by ID Core XT, but are not shown, since, as mentioned elsewhere, *MNS* and *RHCE* are not covered by the NGS assay. The ID Core XT assay and genotyping was performed in accordance with the manufacturer's instructions.

Sanger sequencing

The CDS regions and part of relevant non-coding regions of *FY*, *JK* and *CR1 (only exon 26 and 29)* were amplified by 1X PCR buffer I, 0,2 mM dNTP mix, 0,25 pmol/μl of each primer in a total volume of 25 μl using 2 ng/μl template and 0,05 U/μl AmpliTaqGold (all reagents are from Life Technologies.). The PCR program used was as follows: an initial denaturation at 96°C for 12 min, 40 cycles of 96°C for 30 sec, 56 or 60°C (depending on the product) for 30 sec and 72°C for 60 sec with a final 7-min extension at 72°C (A full list of primers can obtained on request). Sequencing was performed using BigDye® Terminator v3.1 Cycling Sequencing kit. The sequence reactions were run on an ABI3500 Dx and analyzed with SeqScapev.3 (Life Technologies)

Results

A total of 72 samples from persons with different ethnic descent (for details see Materials and Methods) were examined by the NGS assay. Two samples were omitted from NGS data analysis because their general coverage (reads per amplicon) was insufficient. Sixty-four (due to an error in data retrieval, KEL genotypes were only available for 62 samples) of the remaining 70 samples were genotyped using ID Core XT. They comprised samples from 45/46 American blood donors, 11/18 Danish blood donors or patients and from 8/8 EQA distributions.

This article is protected by copyright. All rights reserved.

Author Manuscript

Author Manuscript

For brevity, the NGS assay and ID Core XT genotyping results are presented as names of blood group system reference alleles or the name(s) of the allele(s) deviating from the reference allele (Reid 2012). For example, for the KEL system, the reference allele, KEL*02 encodes k (KEL2), KEL4, KEL5, KEL7, KEL11, KEL12, KEL 13, KEL14, KEL15, KEL16, KEL17, KEL18, KEL19, KEL20, KEL22, KEL26, KEL27, KEL29, KEL30, KEL32, KEL33, KEL35 and KEL36, of which all are assayed by the NGS assay, and of which KEL2, KEL4 and KEL7 are assayed by ID Core XT, the notation in table 2 only states, e.g. KEL*02/KEL*02 for the homozygous presence of the KEL reference allele. As a consequence, the *Concordance* column in table 2 only specifies the alleles that are assayed by both tests.

Kell

The genotypes determined by NGS were consistent with the genotypes obtained by ID Core XT for all 62 samples tested by both methods. Though many different SNPs are described for the *KEL* gene, only the common SNPs encoding the K/k, Kp^a/Kp^b and Js^a/Js^b blood group antigens were detected.

Kidd

There was concordance between the results of the ID Core XT and the NGS assay for the SNP determining JK*01/02 for the majority (61/64) samples. For 2 samples, the NGS assay failed to detect JK*01/02 (Jk^a/Jk^b heterozygosity) due to low coverage in the region of the JK*01/02 polymorphism. The true JK*01/02 genotype of the 2 samples was subsequently confirmed by Sanger sequencing. Furthermore, for 3 of 8 EQA samples and for 20 of 56 donor or patient samples, the NGS assay yielded a more comprehensive *JK* genotype, since the assay was able to detect the presence of 4 different JK*01W variants encoding weak or partial expression of the Jk^a antigen in these 23 samples. In 7 samples the JK*01W variant was present either as homozygous or compound heterozygous. Detection of JK*01W.01 and JK*01W.01/JK*01W.02 variants were confirmed by Sanger sequencing.

Duffy

One important SNP encoding a null variant (previously described as FY_GATA) is situated in the promoter region at -67 bp. The NGS assay detected both the coding region and the Null polymorphism in the promoter. There was complete accord between the NGS data and ID Core XT data for the common SNP encoding the Fy^a and Fy^b antigens and for the FY_GATA SNP in the promoter. The linkage of the FY*01/02

This article is protected by copyright. All rights reserved.

SNP with the FY_GATA SNP is not elucidated by either ID Core XT nor by the NGS assay which makes it difficult to determine the precise genotype (FY*01N.01 or FY*02N.01) in samples which are FY*01/02 heterozygous and carry the FY_GATA Null variant. Two of the donors were shown by the NGS assay to be FY*02M.01 encoding a weak Fy^b antigen, this was confirmed by Sanger sequencing. The 2 donors did not have their genotypes determined by ID Core XT. However, one of the 2 had previously been interpreted as Fy(a+b-) by serological testing.

Diego

There was full concordance between the NGS and ID Core XT with regard to the detection of *DI*B* (unfortunately, no patients or donors in this study were *DI*A*). However, the NGS assay did not include exon 16 of *SLC4A1* thereby missing the SNP (DI*02.03) encoding the Wr^a antigen and other SNPs located in exon 16. Other variants in *SLC4A1* that are currently of unknown clinical significance were detected. Two donors (Hispanic) were heterozygous for c.166G (p.56Glu), and one donor (Middle Eastern) was c. 2716 C (p.906Gln) (data not shown).

Dombrock

The common *DO* variants, DO*01, DO*02, DO*HY and DO*JO alleles encoding Do^a, Do^b, Hy and Jo^a antigens, respectively, were detected by both assays. There was full concordance between the NGS and ID Core XT results.

Colton

For some samples the NGS assays missed exon 4. However, no SNPs have been described in this region. The NGS results were consistent with the ID core results such that that all samples were either CO*01, 01 or CO*01, 02.01. This was to be expected as other known mutations are rarely described (Halverson *et al.* 2010) (Reid 2012).

Cartwright

For all 70 samples, the designed AmpliSeq panel failed to amplify exon 4. This is likely caused by the high (>60%) GC base pair content in this exon. However, none of the previously described mutations are located in this region. The NGS analysis distinguished between YT*02.01 and YT*02.02, whereas ID Core XT only

This article is protected by copyright. All rights reserved.

Author Manuscript

detected the c.1057C>A SNP that is shared by the two genotypes associated with YT*B. Similarly, six samples that were homozygous for both c.1057C (p.353His) and c.1775G (p.592Arg) and were probably a variant of the YT*01 genotype, were detected by the NGS method. Also, a previously undescribed mutation c.752A (p.251Asp) was detected in one sample (Hispanic donor) using the NGS method (data not shown). The detected YT*02.01 and YT*02.02 variants and SNVs were not confirmed by other methods.

Lutheran

The AmpliSeq panel failed to amplify exons 1, 14 and 15, probably because these exons have a GC base pair content of >60 %. However, no variants have been previously reported in these exons. The NGS results of the SNP determining the LU*01/02 genotype was in accordance with the ID Core XT results for all 64 samples typed with both methods. However, for 38 out of 64 samples other SNPs than LU*01/02 are also detected by NGS leading to a more specific genotype. Nine samples were homozygous and 26 were heterozygous for LU*02.19, eight samples were heterozygous for LU*01.02 and one sample was heterozygous for LU*02.-13. The rare mutation c.1351C (p.451Gln) was also detected in one sample (Hispanic) using the NGS method (data not shown).

Vel

One sample from a Caucasian person was homozygous for the deletion c.64_80del encoding the Vel- phenotype. This was in accordance with the serologically obtained phenotype. One black donor was heterozygous and 68 samples were all typed as homozygous VEL+ by NGS. This finding could not be confirmed by other methods as this SNP is not included in the ID Core XT assay.

Knops

Only exons 26 and 29 of the CR1 gene were detected in the NGS assay. Commercial kits only detect the SNP that determines the Kn(a+b-) and Kn(a-b+) phenotypes, whereas other SNPs within the CR1 gene including the genotypes KN*01.-05 (YK(a-)), KN*01.07(Sl(a-)Vil+) and KN*01.-09(KCAM-) also are detected by the NGS assay.

Gerbich

This article is protected by copyright. All rights reserved.

Author Manuscript

Part of exon 1 was not covered by the NGS method, and the rare genotypes GE*01.05 and GE*01.08 would therefore have gone undetected. All 70 samples were found to be homozygous for GE*01, the reference allele for the Gerbich system encoding GE2, GE3, GE4, GE10, GE11 and GE12.

Globoside

Variants in the Globoside system are very rare, and as expected, all of the tested samples were found to be GLOB*01 homozygous.

Landsteiner-Wiener

Only the LW*05 genotype, encoding the common LW(a+b-) phenotype was detected by NGS in all 70 samples.

P1Pk

The NGS assay only covers exon 3 which contains all of the known SNPs that give rise to Null phenotypes. Therefore, the SNP in exon 2a encoding the P_1 and P_2 phenotypes was not detected. Only the A4GALT4*01 and 02 genotypes were observed in all 70 samples.

Chido/Rodgers

The NGS assay was designed to cover exons 25-28 of *CH*(C4B) and *RG*(C4A) that contain the SNPs encoding the different Ch/Rg phenotypes. Even though all exons were fully covered by the NGS assay, it was not possible to discriminate between the different genotypes, mainly because of the high degree of homology between *CH* and *RG* genes. .

Discussion

These data demonstrate that NGS genotyping based on massive parallel sequencing can be used for the detection of common and rare alleles in 14 blood groups systems in one step. The results obtained with NGS for common SNPs are concordant with those obtained using the ID Core XT when the NGS assay coverage of a particular region of the gene is satisfactory. However, in addition to common SNPs the NGS

This article is protected by copyright. All rights reserved.

Author Manuscript

Author Manuscript

analysis also detects rare SNVs, which may potentially lead to a more comprehensive genotype profile. In some of the cases, the extended genotypes corresponded to variant or weakly expressed antigens. This may have clinical implications, since information on weak or variant antigen expression may be very important in the context of antibody identification in clinical samples or in blood donor phenotyping. For example, the presence of a *JK* variant (encoding a Jk^a variant) as detected by NGS in some of the samples would be very helpful information when determining the nature of an apparent anti-Jk^a in a patient typing Jk^a positive by serology. A more comprehensive genotype profile was revealed by the NGS testing, particularly for the Kidd, Duffy and Lutheran blood group systems. Furthermore, the NGS assay enabled genotyping for certain blood group systems such as VEL, Knops, Gerbich, Globoside, P1PK and Landsteiner-Wiener that are not currently addressed by commercial SNP assays. Although, a number of red cell antibodies with specificity for antigens in these systems are clinically benign, e.g. anti-Kn^a, it is nevertheless important to be able to identify these antibodies to high frequency antigens in order to rule out or detect clinically important antibodies, and NGS facilitates this investigation.

Since it was not possible to obtain samples in which confirmatory genomic testing for these blood group systems with PCR or serology had been carried out, the NGS genotypes presented here cannot be fully validated. A 17 bp-deletion in *SMIM1* resulting in a Vel- phenotype in two samples was detected by NGS and this result was confirmed by serology. However, it was not possible to determine the genotype corresponding to the Chido/Rodgers phenotypes by the NGS assay because clear discrimination between *CH* and *RG* was impossible. This is due to the fact that *CH* and *RG* are 99 % homologous and that variation in the *CH* gene is similar to the wildtype *RG* and vice versa (Daniels 2013; Mougey 2010). Thus, the NGS assay needs further refinement in order to be able to resolve the *CH* and *RG* variants from wild types.

Not all coding regions are fully covered by the NGS assay. This is probably due to difficulties in amplifying target regions with high GC base pair content by multiplex PCR. The missed regions were scanned for known SNPs and SNVs compared to the listings in Reid et al.{Reid, 2012 43 /id} and Erytohrogene.com. Accordingly, it is only for the Diego and P1PK systems that common alleles will be missed. Low coverage may also be caused by regions with repeats or homopolymers. Whether the problem with missed regions only applies to this AmpliSeq panel or also applies to other NGS methods is unclear. It is however, clear, that regions with low coverage will have to be investigated by other methods.

RHD, *RHCE*, *GYPA* and *GYPB* genes were intentionally omitted from this AmpliSeq panel due to the high level of homology within their respective systems, which was expected to constitute a methodological

This article is protected by copyright. All rights reserved.

problem. However, other researchers have successfully used NGS for genotyping of blood group alleles, including those of the RH and MNS systems, using AmpliSeq panels and relevant analysis software; simply by assisting their AmpliSeq panel with a simple method for testing of exon 1 and 2 of *RH* and *GYPA/GYPB* (Fichou *et al.* 2014; Fichou *et al.* 2016). Similarly, several groups have shown that data from whole exome sequencing can be used for blood group genotyping including *RHD/RHCE* and *GYPA/GYPB*; simply by including a test for copy number variation (CNV) (Lane *et al.* 2016; Möller *et al.* 2016). We have used the CLC Biomedicals program for detecting variations from reference sequences, and exported the data into an Excel worksheet that was designed to "call" common and uncommon SNPs defined by Reid et al, 2012 (Reid 2012). However, interpretation of sequence data and determining the blood group alleles is complicated by the fact that some of the available reference sequences in Ensemble.org are a combination of different genotypes (Lane *et al.* 2016; Möller *et al.* 2016). Furthermore, the available databases for blood group alleles primarily contain only the most clinically important SNPs and not complete reference sequences for each allele (isbt-web.org,)(Reid 2012). As a consequence, the data obtained from NGS analysis often contains a combination of SNPs that have not been described in the available databases, such as the findings in the YT blood group in this study. In contrast, NGS of HLA-loci is performed routinely and is based on the comparison of sequencing results with a library containing complete reference sequences for most HLA alleles obtained from the IPD-IMGT/HLA database (Barone *et al.* 2015). Recently, Erythrogene.com was described, and it contains the complete coding region sequence of many different alleles obtained from the 1000genome project (Möller et al. 2016). This may simplify NGS Blood group genotyping by providing the sequence of most reference alleles at the coding region level and eliminate the need for confirming NGS results using Sanger sequencing. However, studies have shown that SNPs or SNVs in non-coding regions like promoter regions and splice sites also affect the expression of blood group antigens and must be considered for a complete blood group antigen profile (Christophersen *et al.* 2017; Hoher *et al.* 2017; Lawicki *et al.* 2016; Lomas-Francis *et al.* 2010; Lucien *et al.* 1998; Meny 2010; Reid 2012; Tournamille *et al.* 1995). The effect of splice site mutations may be examined at the mRNA level and similarly, a quantitative real-time PCR protocol would reveal the effect on transcription from nucleotide variations in the promoter region. Furthermore, when previously undescribed genetic variations are discovered, the clinical significance of these new alleles are difficult to predict without clinical correlation. In this study a number of rare SNVs were detected, which are described at Erythrogene.com as 1000Genome alleles with no assigned phenotype. Further studies are needed to determine whether these

This article is protected by copyright. All rights reserved.

Author Manuscript

missense mutations actually encode hitherto undescribed blood group antigens and whether they are clinically significant. Currently, the lack of a comprehensive reference database hampers genetic blood grouping by NGS, and turnaround times may be several weeks or longer if other testing modalities must be performed to confirm the NGS results. Still, this NGS assay can be used as a point of orientation in more complicated serology cases where time is not a critical issue.

Acknowledgement: A special thanks to Pernille Stentoft Rasmussen for technical assistance.

Marianne A. Jakobsen, Mark Yazer and Ulrik Sprogøe designed the research project and wrote the manuscript. Christoffer Dellgren analyzed the data. Chelsea Sheppard contributed the samples and ID Core XT data from the US.

Reference List

1. Barone,J.C., Saito,K., Beutner,K., Campo,M., Dong,W., Goswami,C.P., Johnson,E.S., Wang,Z.X., & Hsu,S. (2015) HLA-genotyping of clinical specimens using Ion Torrent-based NGS. *Hum.Immunol.*, **76**, 903-909.

2. Christophersen,M.K., Joud,M., Ajore,R., Vege,S., Ljungdahl,K.W., Westhoff,C.M., Olsson,M.L., Storry,J.R., & Nilsson,B. (2017) SMIM1 variants rs1175550 and rs143702418 independently modulate Vel blood group antigen expression. *Sci.Rep.*, **7:40451. doi: 10.1038/srep40451.**, 40451.

3. Daniels, G. Human Blood groups. 2013.

4. Fichou,Y., Audrezet,M.P., Gueguen,P., Le,M.C., & Ferec,C. (2014) Next-generation sequencing is a credible strategy for blood group genotyping. *Br.J.Haematol.*, **167**, 554-562.

5. Fichou,Y., Mariez,M., Le,M.C., & Ferec,C. (2016) The experience of extended blood group genotyping by next-generation sequencing (NGS): investigation of patients with sickle-cell disease. *Vox Sang.*, **111**, 418-424.

6. Goldman,M., Nuria,N., & Castilho,L.M. (2015) An overview of the Progenika ID CORE XT: an automated genotyping platform based on a fluidic microarray system. *Immunohematology.*, **31**, 62-68.

This article is protected by copyright. All rights reserved.

Author Manuscript

Author Manuscript

7. Halverson,G.R. & Peyrard,T. (2010) A review of the Colton blood group system. *Immunohematology.*, **26**, 22-26.

8. Hoher,G., Fiegenbaum,M., & Almeida,S. (2017) Molecular basis of the Duffy blood group system. *Blood Transfus.*, 1-8.

9. Keller,M.A. (2015) The role of red cell genotyping in transfusion medicine. *Immunohematology.*, **31**, 49-52.

10. Lane,W.J., Westhoff,C.M., Uy,J.M., Aguad,M., Smeland-Wagman,R., Kaufman,R.M., Rehm,H.L., Green,R.C., & Silberstein,L.E. (2016) Comprehensive red blood cell and platelet antigen prediction from whole genome sequencing: proof of principle. *Transfusion.*, **56**, 743-754.

11. Lawicki,S., Covin,R.B., & Powers,A.A. (2016) The Kidd (JK) Blood Group System. *Transfus.Med.Rev.*, 10.

12. Lomas-Francis,C. & Reid,M.E. (2010) The Dombrock blood group system: a review. *Immunohematology.*, **26**, 71-78.

13. Lucien,N., Sidoux-Walter,F., Olives,B., Moulds,J., Le Pennec,P.Y., Cartron,J.P., & Bailly,P. (1998) Characterization of the gene encoding the human Kidd blood group/urea transporter protein. Evidence for splice site mutations in Jknull individuals. *J.Biol.Chem.*, **273**, 12973-12980.

14. Meny,G.M. (2010) The Duffy blood group system: a review. *Immunohematology.*, **26**, 51-56.

15. Möller, M., Jöud, M., Storry J.R., and Olsson M.L. Erythrogene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. Blood advances 1, 240-249. 2016.

16. Mougey,R. (2010) A review of the Chido/Rodgers blood group. *Immunohematology.*, **26**, 30-38.

17. Paccapelo,C., Truglio,F., Antonietta,V.M., Revelli,N., & Marconi,M. (2015) HEA BeadChip technology in immunohematology. *Immunohematology.*, **31**, 81-90.

18. Reid,M.E. (2009) Transfusion in the age of molecular diagnostics. *Hematology.Am.Soc.Hematol.Educ.Program.*, 171-177.

19. Reid, M. E. Lomas Francis C. and Olsson M. L. The Blood Group Antigen FactsBook. 2012.

20. Tournamille,C., Colin,Y., Cartron,J.P., & Le Van,K.C. (1995) Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat.Genet.*, **10**, 224-228.

This article is protected by copyright. All rights reserved.

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript

Table 1: Blood group genes targeted by the NGS assay.

					CDS experimental coverage		
Blood Group system	**Gene**	**Number of exons in Gene**	**CDS accession number**	**CDS length (bp)**	**bp**	**%**	**Missed exons**
Gerbich	GYPC	4	NM_001256584.1	427	401	93.91	1
Yt	YT	6	NM_015831.2	2026	1836	90.62	4
Dombrock	DO	3	NM_021071.2	975	975	100	
Colton	AQP1	4	NM_198098.3	1113	1113	100	
Kell	KEL	19	NM_000420.2	2389	2389	100	
Kidd	JK/SLC14A1	11	NM_001308278.1	1428	1428	100	
P1Pk	A4GALT	4	NM_001318038.1	1072	1072	100	
Globoside	GLOB	5	NM_033169.2	1006	1006	100	
Duffy	FY	2	NM_002036.3	1058	1058	100	
Lutheran	LU	15	NM_001013257.2	2041	1839	90.1	1, 14, 15
Landsteiner-Wiener	LW	3	NM_001039132.2	943	943	100	
Diego	SLC4A1	20	NM_0006342.3	2926	2749	93.95	16
Vel	VEL	4	NM_001288583.1	257	257	100	
Knops	CR1 (exons 26, 29)	47	NM_000651.4	750*	750	100	
Chido/Rodgers	C4 (exons 25-28)	41	NM_001242823.2	800*	0	0	25-28
Total					17,816	91.24	

AmpliSeq use GRC37/hg19 as reference. CDS stands for coding region.

* Basepairs refers to targeted region and includes more than the coding region.

This article is protected by copyright. All rights reserved.

Table 2: Detected blood group genotypes

			Genotype by ID core		
Blood group	**Occurence**	**NGS genotype**	**Concordance**	**Discordance**	**Comprehensive[1]**
Kell	62	KEL*02/KEL*02	57		
	4	KEL*01.01/KEL*02	3		
	2	KEL*02.06/KEL*02	2		
	1	KEL*02.06/KEL*02.06	1		
	1	KEL*01.01/KEL*02.03	1		
Kidd[4]	17	JK*01/JK*01	13		
	11	JK*02/JK*02	8		
	19	JK*01/JK*02	16	2	
	7	JK*01/JK*01W.01	NA[2]		7
	9	JK*02/JK*01W.01	NA		9
	2	JK*01W.01/JK*01W.01	NA		2
	1	JK*01W.01/JK*01W.02	NA		1
	4	JK*01W.03/JK*01W.04	NA		4
Duffy[4]	13	FY*01/FY*01	13		
	11	FY*02/FY*02	8		
	13	FY*01/FY*02	12		
	6	FY*02N.01/FY*02	6		
	17	FY*02N.01/FY*02N.01	16		
	1	FY*02N.02/FY*02	1		
	5	FY*01N.01 or FY*02N.01	5		
	2	FY*01/FY*02M.01	ND[3]		
	1	FY*02/FY*02M.01	ND		
	1	FY*01N.01/FY*01N.01	1		
Diego	70	DI*02/DI*02	64		
Dombrock	4	DO*01/DO*01	4		
	29	DO*02/DO*02	25		
	31	DO*01/DO*02	29		
	1	DO*01.-05/DO*01.-05	1		
	2	DO*02/DO*02.-04.01	2		
	2	DO*01/DO*01.-05	2		
	1	DO*02/DO*01.-05	1		
Colton	66	CO*01/CO*01	60		
	4	CO*01/CO*02.01	4		
Yt	58	YT*01/YT*01	52		
	1	YT*02.01/YT*02.01	NA		1
	3	YT*01/YT*02.01	NA		3

Author Manuscript

This article is protected by copyright. All rights reserved.

	2	YT*01/YT*02.02		2
	6	YT*01variant		6
Lutheran	4	LU*01.02/LU*02		4
	4	LU*01.02/LU*02.19		4
	29	LU*02/LU*02	26	
	22	LU*02/LU*02.19	NA	21
	9	LU*02.19/LU*02.19	NA	8
	1	LU*02/LU*02.-13	NA	1
Vel	68	Vel+/Vel+	NA	
	1	Vel+/Vel-	NA	
	1	Vel-/Vel-	NA	
Knops	9	KN*01/KN*01	NA	
	4	KN*01.-05/KN*01.-05	NA	
	4	KN*01.-09/KN*01.-09	NA	
	4	KN*01.07;01.-09/KN*01.07;01.-09	NA	
	1	KN*01.06; 1.07; 1.-09/KN*01.06; 1.07; 1.-09	NA	
	48	Various heterozygous	NA	
Gerbich	70	GE*01/GE*01	NA	
Globoside	70	GLOB*01/GLOB*01	NA	
Landsteiner-Wiener	70	LW*05/LW*05	NA	
P1PK	26	A4GALT*01/*01	NA	
	6	A4GALT*02/*02	NA	
	38	A4GALT*01/*02	NA	

The alleles are named according to the nomenclature in The Blood Group antigen Factsbook (Reid et al, 2012) .Although comparing SNPs, all coding regions have been examined by the NGS assay and only SNPs deviating from the wildtype are reported.

1. Comprehensive indicates areas of the genotype where NGS detects additional SNPS compared to ID Core XT.
2. NA; not measurable by ID Core XT
3. ND, not evaluated by ID Core XT
4. For Kidd and Duffy, two ID Core XT results are missing hence the sum is 62.

Author Manuscript

This article is protected by copyright. All rights reserved.