2021

HIGH THROUGHPUT SEQUENCING TECHNOLOGIES IN BLOOD GROUP GENOTYPING AND APPLICATIONS IN TRANSFUSION MEDICINE: ESTABLISHMENT OF RH GENES REFERENCE ALLELE SEQUENCES

Tounsi, Wajnat Ali

http://hdl.handle.net/10026.1/17241

http://dx.doi.org/10.24382/795
University of Plymouth

HIGH THROUGHPUT SEQUENCING TECHNOLOGIES IN BLOOD GROUP GENOTYPING AND APPLICATIONS IN TRANSFUSION MEDICINE: ESTABLISHMENT OF RH GENES REFERENCE ALLELE SEQUENCES

by:

Wajnat A. Tounsi

A thesis submitted in partial fulfilment for the degree of

Doctor of Philosophy

School of Biomedical Sciences, Faculty of Health

June, 2021

To Raed and my children

Yousef, Sulaf and Sufanah

Acknowledgements

This thesis has come together as a results of many peoples' knowledge, time, patience, dedication and faith, to whom I express my deepest gratitude. I express my sincere appreciation to my director of study Dr. Tracey E. Madgett for her guidance, patience and kindness. I would like also to extend my sincere thanks to my supervisors Prof. Neil D. Avent and Dr. Robert Belshaw for their invaluable critiques of this work.

I would also like to thank Dr. Vasileios P. Lenis for his willingness to give his time so generously and his contribution to MinION™ data analysis. I am also thankful to Dr. Garry S. Farnham for his practical support with MinION™ sequencing and data analysis. Thanks should also go to Dr. Michele Kiernan for carrying out sequencing run on the Ion PGM™ and Dr. Paul Waines for his help and support around the lab.

I am also thankful for Dr. Katri Haimila and Silja Tammi for the opportunity to sequence and analyse the Finnish samples and for their invaluable contribution to the analysis and discussion of these samples results.

I am also forever in debt for my husband, Raed, for being my rock through the years and never let me down.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

Work submitted for this research degree at University of Plymouth has not formed part of any other degree either at University of Plymouth or at another establishment. This study was financed with a funded scholarship from King Abdulaziz University, Jeddah, Saudi Arabia.

Name: Wajnat A. Tounsi

Signature:

June 8, 2021

Word count of the main body of thesis: 35 512 words.

Publications

- Silja M. Tammi*, Wajnat A. Tounsi*, Susanna Sainio, Michele Kiernan, Neil D. Avent, Tracey E. Madgett, Katri Haimila. Next-generation sequencing of 35 *RHD* variants in 16,253 serologically D negative pregnant women in Finnish population. *Blood Advances* 4(20), 4994-5001. * Joint first authors

- Wajnat A. Tounsi, Tracey E. Madgett, and Neil D. Avent. Complete *RHD* next generation sequencing: establishment of reference *RHD* alleles (2018). *Blood Advances* 2(20), 2713-2723.

- Kelly A. Sillence, Amr J. Halawani, Wajnat A. Tounsi, Kirsty A. Clarke, Michele Kiernan, Tracey E. Madgett, and Neil D. Avent (2017). Rapid *RHD* zygosity determination using digital PCR. *Clinical Chemistry* 63(7), 1388-1397.

Abstracts from these publications are included in Appendices.

Abstract

High throughput sequencing technologies in blood group genotyping and applications in transfusion medicine: establishment of Rh genes reference allele sequences

Wajnat A. Tounsi

The Rh blood group system (ISBT004) is the second most important blood group after ABO and the most polymorphic one, with 55 antigens encoded by two genes, *RHD* and *RHCE*. The two genes are homologous and located on chromosome 1 with 10 coding exons.

This research uses next generation sequencing (NGS) on Ion Torrent Personal Genome Machine™ (Ion PGM™) to sequence the Rh genes (*RHD*, *RHCE*, *RHAG*) using overlapping Long Range-PCR (LR-PCR) amplicons. The aim was to study different *RHD* and *RHCE* alleles present in the population to establish reference *RHD* and *RHCE* allele sequences by utilising the analysis of intronic single nucleotide polymorphisms (SNPs) and their correlation to a specific Rh haplotype. We also aimed to use single molecule sequencing (SMS), MinION™, to test the feasibility of applying SMS in blood group genotyping (BGG).

Genomic DNA (gDNA) samples (n=218) were sequenced for the *RHD* gene and 100 samples were sequenced for the *RHCE* gene using Ion PGM™. The *RHAG* gene was also sequenced from samples where no mutation was detected in the *RHD* gene that would explain weak D reactivity in serological testing. The *RHD* gene was sequenced from samples (n=13) using MinION™. Data generated was mapped to the human genome reference sequence hg38 and variants were called. Variants detected by the Min-

ION™ were compared to the ones detected from the Ion PGM™.

The *RHD* gene data analysis lead to the detection of different exonic SNPs that correlate to known variants and the identification of 10 novel *RHD* variant alleles. Multiple *RHD* intronic SNPs were found in all samples: 21 intronic SNPs were present in all samples indicating their specificity to the *RHD*DAU0 (RHD*10.00)* allele which the hg38 reference sequence encodes. Twenty-three intronic SNPs were found to be R_2 haplotype specific, and 15 were linked to $\mathrm{R}_1, \mathrm{R}_0, \mathrm{R}_\mathrm{Z}$ haplotypes which lead to the establishment of two *RHD* reference sequences one for R_2 and the other for $\mathrm{R}_1, \mathrm{R}_0, \mathrm{R}_\mathrm{z}$. Intronic SNPs were also detected in the *RHCE* gene in which 89 SNPs were specific to samples with C+/c- and 22 SNPs were found in C-/c+ samples which lead to the conclusion that these SNPs are Rhc specific. This work resulted in the establishment of three *RHCE* allele specific reference sequences in which one is C+ specific, one is c+ specific and the other is for e+ allele. MinION™ was successful for determining the *RHD* allele in all 13 samples sequenced.

NGS is a high throughput technique that enables the sequencing of polymorphic blood group genes and identification of novel variants present in the population which cannot be detected using the available BGG molecular platforms. Intronic SNPs may represent a novel diagnostic approach to investigate known and novel variants of the *RHD* and *RHCE* genes, whilst being a useful approach to establish reference *RHD* and *RHCE* allele sequences. MinION™ is a powerful tool and was successful here in BGG. However, challenges remain with developing a straightforward user-friendly data analysis pipeline for translating this work into transfusion facilities.

Contents

List of Figures

List of Tables

Abbreviation List

ADCC	Antibody-dependent cell-mediated cytotoxicity	ISPs	ion sphere particles
AGO1	*Argonaute RISC component 1*	ITAM	immunoreceptor tyrosine based activatory motif
AHG	Anti-human globulin	ITIM	immunoreceptor tyrosine based inhibitory motif
AMIS	Antibody-mediated immune suppression	kb	kilobases
Amt	Ammonia transporter	kDa	kilodalton
anti-Id	Anti-idiotypic antibodies	LISS	low ionic strength solution
BCR	B-cell receptor	LR-PCR	long range-polymerase chain reaction
BGG	blood group genotyping	LW	Landsteiner and Wiener glycoprotein
BHQ-1	Black Hole Quencher 1	Mb	megabytes
bp	base pair	MHC	major histocompatibility complex
cffDNA	cell free fetal DNA	MPS	mononuclear phagocyte system
DAT	Direct antiglobulin test	NAT	nucleic acid testing
ddNTPs	dideoxynucleotide triphosphates	NCBI	National Centre for Biotechnology Information
ddPCR	droplet digital polymerase chain reaction	NGS	Next-generation sequencing
DNA	deoxyribonucleic acid	NHSBT	National Health Service Blood and Transplant
dNTPs	deoxynucleotide triphosphates	PCR	polymerase chain reaction
EDTA	ethylenediaminetetraacetic acid	PEG	polyethylene glycol
FAM	6-carboxyfluorescein	RBCs	red blood cells
FcRn	neonatal Fc receptor	RFLP-PCR	restriction fragment length polymorphism-polymerase chain reaction
FNHTRs	Febrile non-haemolytic transfusion reactions	RHAG	Rh-associated glycoprotein
GB	gigabytes	RhBG	Rh type B glycoprotein
gDNA	Genomic DNA	RhCG	Rh type C glycoprotein
GYP	Glycophorin	*RHD* ψ	*RHD* pseudogene
HDFN	haemolytic disease of the fetus and newborn	SCD	sickle cell disease
HEX	Hexachlorofluorescein	SMRT	single-molecule real time
HGP	Human Genome Project	SNPs	single nucleotide polymorphisms
HIV	human immunodeficiency virus	SMS	single molecule sequencing
HLA	human leukocyte antigens	*SMP1*	*small membrane protein 1*
HPLC	high performance liquid chromatography	SNV	single nucleotide variant
HTRs	haemolytic transfusion reactions	SSP-PCR	sequence-specific primers-polymerase chain reaction
ICAM-4	intercellular cell adhesion molecule 4	T_a	Annealing temperature
IDAT	indirect antiglobulin test	TAE	Tris-acetate-EDTA
IgG	immunoglobulin G	TGS	Third generation sequencing
IgM	immunoglobulin M	T_m	melting temperature
IGV	Integrated Genome Viewer	*TMEM50A*	*Transmembrane 50A*
Ion PGM™	Ion Torrent Personal Genome Machine™	TPH	transplacental haemorrhage
ISBT	International Society of Blood Transfusion	WES	whole exome sequencing
ISFET	ion-sensitive field-effect transistor	WGS	whole genome sequencing

1. Introduction

1.1 Introduction to Blood Groups

Blood group antigens are genetically inherited mostly protein or carbohydrate structures present on the outer membrane of red blood cells (RBCs) that can be detected by their corresponding antibody (Reid and Mohandas, 2004; Daniels, 2005; Reid et al., 2012; Daniels, 2013c). The most clinically significant blood group, ABO, was discovered in 1901 by Karl Landsteiner, after he observed that some patients' plasma agglutinated others' RBCs. His discovery explained the issue of inconsistent outcome from blood transfusion at that time (Landsteiner, 1961; Yamamoto, 2000; Poole and Daniels, 2007). After the introduction of the indirect antiglobulin test (IDAT) in 1945, scientists were able to identify more blood group antigens (Coombs et al., 1945). As serological tests became more sensitive and specific, and further techniques like molecular and genetic tests became available, more blood group antigen polymorphisms were revealed creating further diversity (Daniels, 2005). Nowadays, there are 43 blood groups registered with the ISBT which are listed in Table 1.1 ISBT.

The expression of most blood group antigens is controlled by one gene, whilst a few are encoded by closely related multiple genes, such as Rh and MNS (Daniels, 2005). Polymorphism of blood group antigens are most often caused by one or more single nucleotide polymorphisms (SNPs), resulting in different forms of an antigen in a population. The Rh blood group system is the most polymorphic blood group with 55 different antigens, and over 300 alleles (Storry et al., 2019). Blood group antigens have been linked to variant functions. These functions include membrane physical support, molecules transportation through the cell membrane, receptors for extracellular ligands, adhesion molecules, complement components and regulator (Reid and Mohandas, 2004).

1.1.1 Blood Antigen Identification and Classification

Blood groups verified by serology are categorized into either systems, collections, low or high incidence antigens. Systems contain one or more antigens encoded by a single gene, such as Kel blood group system, or two genes that are closely linked, such as Rh blood group system (Daniels et al., 2004), see Table 1.1 for a list of blood group systems. Collections are antigens that might be linked serologically, biochemically or genetically, yet lack some of the criteria to fit into a blood group system. Antigens that cannot be assigned to a system due to the lack of carrier protein or carbohydrate on the RBCs are grouped together into collections which consist of serologically, biochemically, or genetically related antigens. Col-

lections are classified into potentially, sometimes, and not clinically significant. Low and high incidence antigens include low and high incidence blood antigens that do not meet the other two classification criteria.

Since blood groups were discovered over 100 years ago, they were denoted in different styles. In the 1980s, the ISBT initiated the Working Party on Terminology for Red Cell Surface Antigens to establish a universal process for describing blood antigens to eliminate confusion from using different terminologies and create a genetically based database for all blood group antigens.

The ISBT blood antigen numerical terminology consist of 6 digits; the first 3 represent the blood group system (e.g 006 for the Kell) and the second 3 digits represent the antigen (e.g. 006003 for $\mathrm{Kp^a}$) (Table 1.1). Also, the system symbol can be used followed by the antigen number, and zeros on the left may be deleted (e.g. KEL003 or KEL3) (Daniels et al., 2004). Phenotype is denoted by using the system symbol followed by a semicolon followed by the antigen number (e.g. KEL;3); following antigen numbers are separated by commas (e.g. KEL:3,4). In case of a missing antigen, an antigen number is preceded by a minus sign (e.g. KEL:1,-2,3).

To represent the blood group gene, allele, or a haplotype, the ISBT system symbol is used instead of the gene name (italicized) followed with an asterisk and then the antigen number is used, subsequent antigens are sep-

arated by commas (e.g. *DCe* would be represented by *RH*1,2,5*) (Daniels et al., 2004; Harmening, 2005).

Although ISBT terminology is useful for tracking and keeping up-to-date with newly identified antigens , it is hard to use in a clinical setting as numbers can be hard to remember and distinguish between different blood groups. Therefore, alternative terminology is used, for example, Fisher-Race and Wiener terminologies are still in use to report Rh haplotypes, sections 1.3.2.1 and 1.3.2.2.

1.2 Rh Blood Group System

The Rh blood group system (ISBT004) is the second most important blood group system after ABO (Avent and Reid, 2000; Avent et al., 2006) and one of the most polymorphic blood group systems. The *RHD* and the *RHCE* genes located on chromosome 1 (1p33.1-1p36), encode the RhD protein and RhCcEe protein, respectively (Colin et al., 1991; Wagner and Flegel, 2000). The two closely related genes encode 55 different antigens in the Rh system; five of them are considered the leading antigens (RhD, RhC, Rhc, RhE, and Rhe). Rh antigens are expressed on the surface of the RBCs on the Rh protein complex that is integrated within the cell membrane (Iwamoto, 2005). The RhD antigen is the most clinically significant antigen in the Rh system due to its high immunogenicity and being the main cause of haemolytic disease of the fetus and newborn (HDFN), section 1.8.2.

Table 1.1: Blood Group Systems, genes encoding the blood groups, number of antigens, and chromosomal location. Table adapted from ISBT (2021).

No.	System name	System symbol	Gene name(s)	No. of antigens	Chromosomal location
001	ABO	ABO	*ABO*	4	9q34.2
002	MNS	MNS	*GYPA, GYPB, GYPE*	50	4q31.21
003	P1PK	P1Pk	*A4GALT*	3	22q13.2
004	Rh	RH	*RHD, RHCE*	55	1p36.11
005	Lutheran	LU	*BCAM*	27	19q13.2
006	Kell	KEL	KEL	*36*	7q33
007	Lewis	LE	*FUT3*	6	19q13.3
008	Duffy	FY	*ACKR1*	5	1q21-q22
009	Kidd	JK	*SLC14A1*	3	18q11-q12
010	Diego	DI	*SLC4A1*	22	17q21.31
011	Yt	YT	*ACHE*	5	7q22
012	Xg	XG	*XG, MIC2*	2	Xp22.32
013	Scianna	SC	*ERMAP*	9	1p34.2
014	Dombrock	DO	*ART4*	10	12p13-p12
015	Colton	CO	*AQP1*	4	7p14
016	Landsteiner-Wiener	LW	*ICAM4*	3	19p13.2
017	Chido/Rodgers	CD/RG	*C4A, C4B*	9	6p21.3
018	H	H	*FUT1*	1	19q13.33
019	Kx	XK	*XK*	1	Xp21.1
020	Gerbich	GE	*GYPC*	13	2q14-q21
021	Cromer	CROM	*CD55*	20	1q32
022	Knops	KN	*CD1*	12	1q32.2
023	Indian	IN	*CD44*	6	11q13
024	Ok	OK	*BSG*	3	19p13.3
025	Raph	RAPH	*CD151*	1	11p15.5
026	John Milton Hagen	JMH	*SEMA7A*	8	15q22.3-q23
027	I	I	*GCNT2*	1	6p24.2
028	Globoside	GLOB	*B3GALNT1*	2	3q25
029	Gill	GIL	*AQP3*	1	9p13
030	Rh-associated glycoprotein	RHAG	*RHAG*	4	6p12.3
031	FORS	FORS	*GBGT1*	1	9q34.13-q34.3
032	JR	JR	*ABCG2*	1	4q22.1
033	LAN	LAN	*ABCB6*	1	2q36
034	Vel	VEL	*SMIM1*	1	1p36.32
035	CD59	CD59	*CD59*	1	11p13
036	Augustine	AUG	*SLC29A1*	4	6p21.1
037	KANNO	KANNO	*PRNP*	1	20p13
038	Sid	SID	*B4GLANT2*	1	17q21.32
039	CTL2	CTL2	*SLC44A2*	2	19p13.2

continued ...

No.	System name	System symbol	Gene name(s)	No. of antigens	Chromosomal location
040	PEL	PEL	*ABCC4*	1	13q32.1
041	MAM	MAM	*EMP3*	1	19q13.33
042	EMM	EMM	*PIGG*	1	4p16.3
043	ABCC1	ABCC1	*ABCC1*	1	16p13.11

1.3 Rh Blood Group System

The Rh blood group system (ISBT004) is the second most important blood group system after ABO (Avent and Reid, 2000; Avent et al., 2006) and one of the most polymorphic blood group systems. The *RHD* and the *RHCE* genes located on chromosome 1 (1p33.1-1p36), encode the RhD protein and RhCcEe protein, respectively (Colin et al., 1991; Wagner and Flegel, 2000). The two closely related genes encode 55 different antigens in the Rh system; five of them are considered the leading antigens (RhD, RhC, Rhc, RhE, and Rhe). Rh antigens are expressed on the surface of the RBCs on the Rh protein complex that is integrated within the cell membrane (Iwamoto, 2005). The RhD antigen is the most clinically significant antigen in the Rh system due to its high immunogenicity and being the main cause of HDFN, which is discussed in section 1.8.2.

1.3.1 Discovery

The Rh blood group system was first reported in the literature in 1939 by Levine and Stetson (1939). They reported a case of a female, with blood group O, who gave birth to a stillborn infant and suffered severe

haemolytic transfusion reactions (HTRs) after being transfused with her husband's blood, blood group O. After more investigations, serological tests revealed that her blood agglutinated about 80% of the blood units of the same blood group. They assumed that an agglutinating factor, caused by the dead infant, was the main reason for the alloimmunisation (Levine and Stetson, 1939). In 1941, Landsteiner and Wiener used rhesus blood to immunise rabbits and isolated an agglutinating factor that agglutinated human blood, which they called Rhesus factor. They came to a conclusion that the agglutinating factor responsible for the unexplained transfusion reactions were the Rh antibodies (Levine et al., 1941). It was later found that the antibody they identified was not the Rh antibody but rather an antibody now known as anti-LW (Avent and Reid, 2000).

1.3.2 Rh Terminology

There are multiple nomenclature systems in use to denote Rh blood group haplotype and phenotype. Three Rh terminologies including Fisher-Race (Fisher and Race, 1946), Wiener (Wiener, 1969), and Rosenfield (Rosenfield et al., 1962) are used to donate Rh antigens. Fisher-Race and Wiener terminologies were based on the researchers postulation of the Rh genetic mechanism. Fisher-Race and Wiener terminologies are used in this work therefore it is important to discuss them in detail in the following section.

Table 1.2: The Rh most common haplotypes in Wiener, Fisher-Race, and ISBT numerical terminology, and haplotype frequency in Caucasians.

Wiener	Fisher-Race	ISBT terminology	Occurrence (%)*
R_0	Dce	RH*1,4,5	4
R_1	DCe	RH*1,2,5	42
R_2	DcE	RH*1,3,4	14
R_z	DCE	RH*1,2,3	0
r	dce	RH*4,5	37
r′	dCe	RH*2,5	2
r''	dcE	RH*3,4	1
r^y	dCE	RH*2,3	<1

*Frequency in Caucasians according to Singleton et al. (2000); Daniels (2013b).

1.3.2.1 Fisher-Race (DCE) Terminology

Fisher and Race, when studying the Rh antigens in the 1940s, hypothesised that the Rh antigens are encoded by three closely linked alleles, in which each allele produces a specific antigen (Fisher and Race, 1946; Watkins et al., 1988; Garratty et al., 2000). They assigned each antigen a letter with a gene name that corresponded to the same letter but italicized. They named the Rh antigens (D, d, C, c, E, and e) and the three genes encoding them *D* gene, *C/c* gene, and *E/e* gene. It is now established that there is no d antigen but it is used to indicate the absence of antigen D. The Rh phenotype in the Fisher-Race terminology is based on the presence of D, C, c, E, and e antigens (Garratty et al., 2000; Harmening, 2005).

Fisher and Race proposed that an individual inherits two set of Rh genes (haplotype), one from each parent, and because the Rh genes are co-dominant,

each gene expresses the related antigen on the RBCs. Inherited haplotypes determine the Rh genotype which controls the Rh phenotype. The most common Rh haplotypes expressed in Fisher-Race terminology are listed in Table 1.2. Rh haplotype in the Fisher-Race terminology is written as CDE as they speculated that the Rh genes are located in *C/c*, *D/d*, *E/e* order; however, the order now has been changed to DCE after the identification of the Rh genetic basis.

Individuals with no Rh antigens expressed on the RBCs have $\mathrm{Rh_{null}}$ phenotype and individuals with weak expression have $\mathrm{Rh_{mod}}$ phenotype, although there is no special way to describe them in the Fisher-Race terminology. Fisher-Race terminology is widely used today to denote Rh haplotype.

1.3.2.2 Wiener (Rh-Hr) Terminology

Wiener hypothesized that one Rh gene produces three antigens able to cause agglutination with the corresponding antibodies (Wiener, 1969; Dunstan, 1986; Harmening, 2005). The Rh gene in Wiener terminology produced 3 antigens, which present the Rh haplotype, for example, $\mathrm{R_1}$ haplotype in Wiener terminology produces D, C, e antigens. See Table 1.2 for the most common Rh haplotypes presented in Wiener terminology. Rh haplotype frequencies and distribution depend on the population ethnicity (Garratty et al., 2000; Harmening, 2005; Daniels, 2013b).

Wiener terminology describes the Rh haplotype by indicating the presence

or the absence of antigen D with (R) or (r), respectively. The presence of antigen C is denoted by (1) or (′). Antigen c is implied in the absence of (1) or (′). The presence of antigen E is denoted by (2) or (″), and the absence of E antigen symbols implies the presence of antigen e. If both C and E antigens are present they are indicated by (z) or (y), (Table 1.2).

To describe Rh antigen phenotypes in Wiener (Rh-Hr) terminology, (′) indicates C or c antigens and (″) indicates E or e antigens. The r preceding h indicates the presence of C (rh′) or E (rh″) antigen; while the h preceding the r indicates the presence of c (hr′) or e antigen (hr″). Interestingly, Wiener terminology for Rh phenotype does not have designation to express the absence of antigen D. Wiener terminology for Rh haplotype is still used today but not phenotype designation as it is difficult to use and did not accommodate other Rh antigens identified later (Wiener, 1969; Garratty et al., 2000; Harmening, 2005).

To summarize, Rh phenotyping by serology assesses Rh antigens present on the outer surface of the RBCs. The results are then expressed as Rh haplotype using either the Fisher-Race or Wiener terminology. The Rh haplotype is determined based on the most common haplotype in a certain population, for example, the most common haplotype in Caucasian is $\mathrm{R_1}$ (DCe) (Daniels, 2013b). If phenotyping results came as D+, C+, c-, E-, e+, then the most common haplotype will be $\mathrm{R_1R_1}$ (DCe/DCe) or $\mathrm{R_1r'}$ (DCe/dCe), depending on the *RHD* gene zygosity which cannot be deter-

mined by serology. Other possible haplotypes ($\mathrm{R_0r'}$) are excluded due to their low frequency in the population.

1.3.2.3 ISBT Terminology

The ISBT Working Party assigned the Rh blood group system number 004, and the most common Rh antigens are numbered as follow: 001 for RhD, 002 for RhC, 003 for RhE, 004 for Rhc, and 005 for Rhe, (Section 1.1.1). Each antigen is represented in a unique number, for example, the RhD antigen is represented by 004001. To shorten, the blood system symbol is used (all capitals) in conjunction with the antigen number and zeros are removed (e.g. RH001 or 004001 would be RH1).

Rh Phenotype is represented by the system symbol followed by the antigen number. A minus sign precedes the antigen number when an antigen is missing (RH:-1). The Rh alleles/ haplotypes are presented by the system symbol (capitalised) followed by the antigen numbers (all italicized) (e.g. *DCe* would be represented by *RH*1,2,5*) (Garratty et al., 2000; Daniels et al., 2004; Harmening, 2005), (Table 1.2).

Although ISBT terminology is useful for tracking and keeping up to date with new antigens and blood groups, it is hard to use in clinical settings as numbers can be hard to remember and distinguish. The Fisher-Race and Wiener terminologies are referred to in this work.

1.3.3 Genetic Structure of Rh Genes

The Rh blood group system is the most polymorphic blood group system. Two genes, *RHD* and *RHCE*, give rise to 55 Rh antigens. The Rh protein is expressed in a complex transmembrane structure, where the RhCcEe protein is encoded by the *RHCE* gene and the RhD protein is encoded by the *RHD* gene (Colin et al., 1991; Carritt et al., 1997).

The *RHD* and *RHCE* genes are located on chromosome 1, short arm, at position 1p36.11, where they lie next to each other in a mirror image format separated by the *small membrane protein 1 (SMP1)* gene (Wagner and Flegel, 2000), which is now called *transmembrane 50A (TMEM50A)* . The two genes are composed of 10 coding exons, where *RHD* is located between two homologous *Rhesus boxes* that are 9 kilobases (kb) each (Figure 1.1 -1) (Avent et al., 2006). The *RHD* and *RHCE* genes are homologous. The evolution of the Rh genes and their different alleles are discussed in more detail in section 1.4.

Recombination, deletion, and point mutations in these two genes generate the eight most common Rh haplotypes (Table 1.2) (Noizat-Pirenne et al., 1998). Different studies were conducted to clarify the molecular structure of the Rh proteins (Section 1.6). This research has improved as the molecular technology became more advanced, accurate and affordable in the last two decades. In 2000, the *RHD* and *RHCE* genes were fully sequenced us-

ing Sanger sequencing; nucleotide substitutions, gaps and recombination events were studied (Okuda et al., 2000).

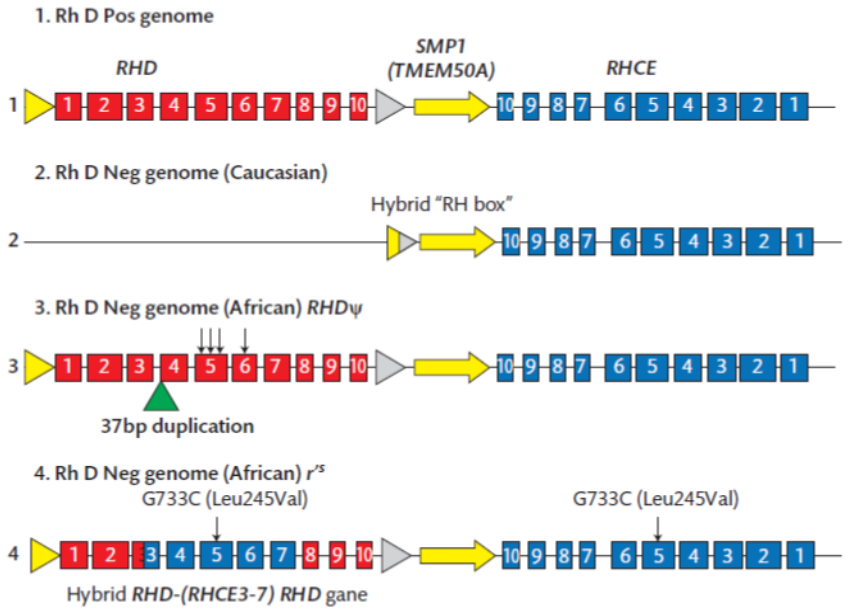


Figure 1.1: The *RHD* and *RHCE* genetic structure. (1) The *RHD* and *RHCE* genes are located on chromosome one and separated by gene segment called *TMEM50A*. The *RHD* gene is located between two homologous structures called *Rhesus boxes*. (2) RhD negative phenotype in Caucasians occurs as a result of a complete deletion of the *RHD* gene by incomplete cross over resulting in a hybrid *Rhesus box*. (3) *RHD* pseudogene (*RHD* ψ) is a common variant in RhD negative Africans caused by an insertion of 37bp at the start of exon 4 resulting in a shift in the reading frame, leading to a stop codon that prevents successful translation of the RhD protein. (4) A common variant allele in RhD negative individuals of Afro/ Caribbean descent known as $\mathrm{r'^{s}}$ haplotype which have mutated Rh genes including a hybrid *RHD-RHCE-RHD* allele on the *RHD* gene, and $\mathrm{ce^{s}}$ on the *RHCE* gene. The combination of the two mutated genes encode for negative D, partial c, weak partial e known as $\mathrm{e^{s}}$, VS, V antigens. Image courtesy of Avent (2018).

1.4 The Evolution of the Rh Genes

According to Okuda et al. (2000) the *RHD* and *RHCE* genes are 93.8% homologous in all introns and coding exons. The similarities between these two genes gives an indication to their evolutionary origin from the same ancestral gene through duplication event that occurred 240-300 million years ago (Carritt et al., 1997; Okuda and Kajii, 2002). The findings of Carritt et al. (1997) following analysis of the Rh genes suggest that the *RHD* gene is a duplicate of the *RHCE*ce* allele, which means that the original Rh lineage is Dce (R_0), similar to findings by Kemp et al. (1999) from the *RHCE* gene microsatellite variation analysis.

The *RHCE*C* allele is suggested to originate from one non-reciprocal genetic recombination event of the *RHD* gene sequence into the *RHCE*ce* allele. This was inferred by the identical sequence of the *RHD* gene and the *RHCE*C* allele in exon 2 and its flanking non-coding regions (Carritt et al., 1997). The *RHCE*Ce* allele was found to include a 109 base pair (bp) insertion in intron 2 that does not exist in either the *RHD* gene or the *RHCE*ce* allele (Kemp et al., 1999; Zhou et al., 2008). This 109 bp insertion is thought to be a remainder of the genetic conversion event from which the *RHCE*C* allele originated (Carritt et al., 1997).

The other Rh haplotypes arose from the Dce (R_0) allele by various types of mutation. For example, DcE (R_2) haplotype emerged from a single muta-

tion in exon 5 c.676C>G predicting amino acid change p.Pro226Ala.

1.5 Rh Genetic Polymorphism

1.5.1 *RHD* Genetic Variation

Although there are many variants of the RhD antigen, they can be placed into five main categories: RhD positive, RhD negative, RhD weak, RhD partial, and very weak D (DEL). RhD variations are encoded by the *RHD* gene, in which single nucleotide variant (SNV), deletion, and/or recombination lead to RhD polymorphism. RhD positive individuals carry normal D antigen on their RBCs. In Caucasians, 82%-85% of individuals are RhD positive (Singleton et al., 2000).

1.5.1.1 RhD Negative Phenotype

The RhD negative phenotype can be either due to complete deletion of the *RHD* gene or associated with *RHD* gene presence without the protein production (Figure 1.12). In Rh-D negative Caucasians, complete deletion of the *RHD* gene is the most common cause for RhD negative phenotype (Flegel and Wagner, 2002). While in Asians and Africans, RhD negativity is associated with an intact or partial *RHD* gene. Translation to viable RhD protein fails because of mutations within the *RHD* gene (Iwamoto, 2005). For example, the *RHD* pseudogene (*RHD* ψ) is a common variant in RhD negative Africans (66%). *RHD* ψ is caused by an insertion of 37 bp (Singleton et al., 2000) in exon 4 that causes a shift in the reading frame, leading

to a stop codon that prevents successful translation to a valid RhD protein (Figure 1.13). Other point mutations found in exons 5 and 6 are also associated with *RHD* ψ (Singleton et al., 2000; Flegel and Wagner, 2002).

1.5.1.2 RhD Weak & Partial Phenotypes

Weak expression of antigen D arises when missense mutations lead to a decreased number of antigen D epitopes expressed on the RBCs. Partial D phenotype can also be caused by the loss of some antigen D epitopes due to the presence of hybrid *RHD-RHCE* genes. Hybrid *RHD-RHCE* genes are caused by partial or full exons changes in the *RHD* gene with ones from the *RHCE*. Partial D amino acid changes generally affect the ones at the extracellular loops of the RhD protein and weak D changes generally affect amino acids located in the transmembrane loop segments (Figure 1.2).

The difference that distinguishes partial and weak D phenotypes is antibody production. Individuals with partial D phenotype are able to produce antibodies against missing epitopes of normal antigen D when exposed to complete RhD through transfusion or pregnancy. Individuals with weak D allele are unable to produce antibodies against the D antigen, as mutations in the *RHD* gene weaken the efficiency of protein anchoring in the cell membrane leading to decreased number of D epitopes expressed but the integrity of epitopes is not affected (Avent et al., 2006).

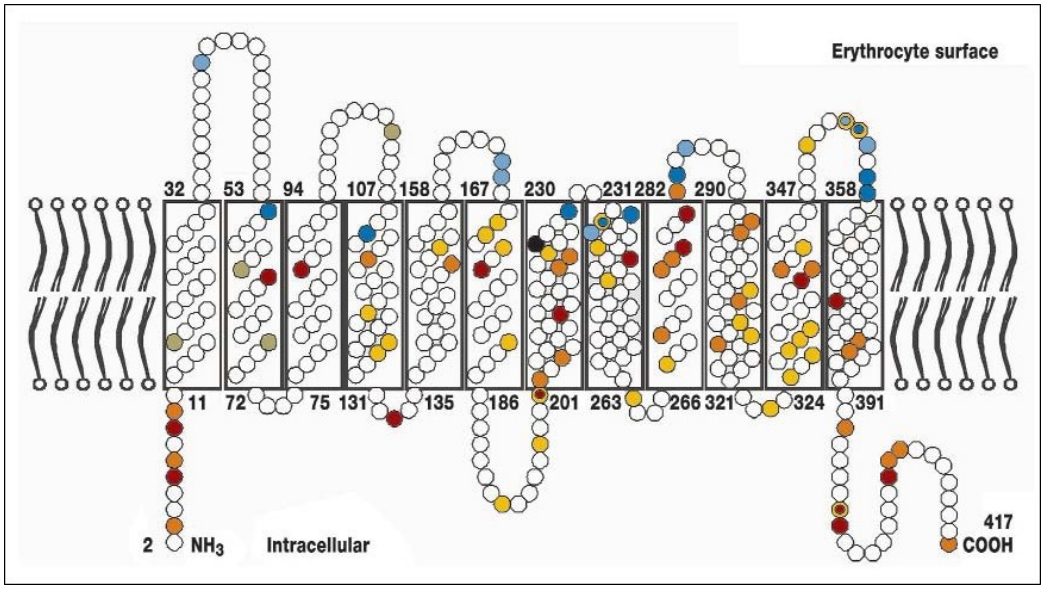


Figure 1.2: The Rh protein integrated in the red cell membrane in 12 spanning domains. Mature Rh protein show 416 amino acids presented as circles, in which yellow ones represent the difference between RhD and RhCE; blue and red ones represent amino acid changes in partial D and weak D, respectively; green represent amino acids encoding C; black represent amino acid that encode E. Image courtesy of Flegel (2007b).

1.5.1.3 Very Weak D (DEL) Phenotype

Individuals with DEL phenotype express very low level of antigen D on the outer membrane of the red cells that cannot be identified using serology agglutination (Westhoff, 2007). Adsorption, elution and molecular genotyping techniques are the only way to detect a DEL allele (Kwon et al., 2017). DEL phenotype is often mis-phenotyped as D negative by serology. Transfusing RhD negative patients with seemingly RhD negative blood (in fact DEL) imposes a risk of sensitisation (Kwon et al., 2017). The risk of immunisation in transfusing D negative patients with DEL phenotype RBCs

used to be considered unlikely. However, cases of D sensitisation in D negative patients have been reported indicating DEL phenotype in donors as the cause of sensitisation (Wagner et al., 2005; Yasuda et al., 2005). Studies have linked the transfusion of DEL red cell, such as *RHD(K409K)* and *RHD(IVS5-38del4)*, to D negative patients to secondary anti-D immunisation (Wagner et al., 2005; Yasuda et al., 2005).

There are over 40 DEL alleles listed in the RhesusBase website (Wagner and Flegel, 2014). DEL phenotype is a common allele among RhD negative Asians (10-30%) (Shao et al., 2002) but significantly lower among D negative Caucasians (0.1%) (Kwon et al., 2017). Molecular basis of DEL allele differs in populations, for example, the most common DEL allele in Asians is *RHD*DEL1* which is caused by c.1227G>A (p.Lys409Lys) in exon 9 predicting splice site disruption (Kwon et al., 2017).

1.5.1.4 $\mathbf{Rh_{null}}$ Phenotype

The $\mathrm{Rh_{null}}$ phenotype is a very rare condition in which an individual lacks the Rh complex; therefore, RBCs show morphological (stomato-spherocytosis) and biochemical abnormalities such as defective cation co-transport and water and phospholipid imbalance. Individuals with $\mathrm{Rh_{null}}$ phenotype are diagnosed with Rh deficiency syndrome. The phenotype is also linked to expression deficiency of intercellular cell adhesion molecule 4 (ICAM-4), previously known as Landsteiner and Wiener glycoprotein (LW), CD47, and Glycophorin (GYP) B. $\mathrm{Rh_{null}}$ phenotype has also been linked to Rh-

associated glycoprotein (*RHAG)* gene mutations which may cause disruption of the Rh complex leading to Rh complex complete absence from the cell membrane (Polin et al., 2016; Mu et al., 2019). $\mathrm{Rh_{null}}$ carriers suffer chronic haemolytic anaemia (Huang, 1998; Avent et al., 2006).

1.5.2 *RHCE* Genetic Variations

Nucleotide substitutions in the *RHCE* gene are the reason behind RhC/c and RhE/e polymorphism. Although there are six point mutations that lead to four amino acid alterations (p.Cys16Trp; p.Ile60Leu; p.Ser68Asn; p.Ser103Pro) which have been linked to the RhC/c, the p.Ser103Pro substitution has been identified as the main mutation that leads to C/c polymorphism (Mouro et al., 1993; Simsek et al., 1994). p.Pro226Ala amino acid substitution is the main cause of Rh E/e polymorphism.

Similarly to RhD, RhE and RhC antigenicity and expression are affected by different mutations that could lead to a weakened or a partial expression of the antigens. For example, missense mutations in the *RHCE* gene cause the expression of the $\mathrm{C_x}$ and $\mathrm{C_w}$ antigens, which are a weakened form of the E and c antigens (Mouro et al., 1993; Noizat-Pirenne et al., 1998; Avent et al., 2006). Rh $\mathrm{E_w}$ is also a form of weakened expression of antigen E, which results from p.Met167Lys amino acid substitution (Strobel et al., 2004). Antigen e expression is altered in people who express the VS antigen which occurs as a result of the amino acid substitution p.Leu245Val

and a hybrid gene (Faas et al., 2001), and it can also be encoded by the *RHCE* gene Westhoff et al. (2001). Complete depletion of the C/c and E/e antigen expression occurs very rarely when the *RHD* gene replaces the *RHCE*. The lack of the RhCcEe protein expression was not linked to any physical abnormalities (Gardner et al., 1991).

1.6 Rh Protein Complex

In the 1980s, researchers began to investigate the molecular basis of the Rh antigens, which led to the purification of the RhcE, RhD and RhAG polypeptides (Avent et al., 1990; Gardner et al., 1991; Mouro et al., 1993; Simsek et al., 1994). Rh protein is about 30 kilodalton (kDa) integrated in the RBCs membrane (Avent et al., 1990; Le van Kim et al., 1992). The mature RhD or RhCE protein is composed of 416 amino acid residues (Figure 1.2), which is found to be limited to the RBCs. Rh polypeptides are integrated in the cell membrane in 12 plasma-membrane-spanning domains (Okuda and Kajii, 2002; Avent et al., 2006).

The core of the Rh-complex is made up of one Rh non-glycosylated (RhD or RhCE) protein and two glycosylated RhAG molecules that create a trimeric structure (Conroy et al., 2005; Avent et al., 2006), where RhAG plays a crucial role in this complex structure (Conroy et al., 2005). Figure 1.3 shows a 3D module of the Rh protein trimeric complex, which is thought to be integrated in the cell membrane with band3. This complex binds to the

RBC membrane by the second repeat of ankyrin and protein 4.2 to the cytoskeleton (Nicolas et al., 2006; Polin et al., 2016); therefore, this complex has a crucial function in sustaining the structural integrity of the RBC membrane. Rh-complex is also integrated in the cell membrane with different associated proteins; which include CD47 glycoprotein, GYP B, ICAM-4, and Duffy glycoprotein. Rh type B glycoprotein (RhBG) and Rh type C glycoprotein (RhCG) are two non-erythroid specific Rh glycosylated proteins found in various tissues (Conroy et al., 2005; Avent et al., 2006), (Figure 1.4).

The specific functions of Rh-proteins are not clearly understood; nevertheless, Rh protein complex is essential for RBC structure and morphology. Oxygen and cation transportation were linked to this protein complex (Bruce et al., 2003; Nicolas et al., 2006; Bruce et al., 2009; Polin et al., 2016). Different studies have linked Rh-proteins to ammonium transporters (Okuda and Kajii, 2002; Avent et al., 2006; Planelles, 2007). RhAG protein in red cells and RhBG and RhCG proteins expressed on different human organs that are associated with ammonium transportation, are linked to ammonium ion transportation (Liu et al., 2000; Conroy et al., 2005; Avent et al., 2006; Planelles, 2007). These proteins have shown similarities to the ammonium channel Ammonia transporter (Amt) proteins, which play an important role in the transportation of ammonium ion (NH_4^+) (Liu et al., 2001; Conroy et al., 2005). However, there is no evidence to link RhD and RhCE

proteins to ammonium transportation (Conroy et al., 2005).

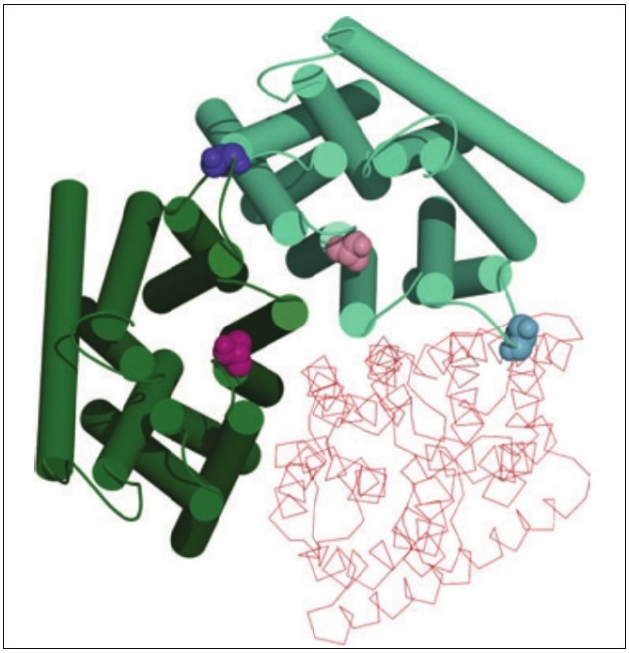


Figure 1.3: 3D module of the trimeric structure of the Rh protein complex ($\mathrm{RhD}-\mathrm{RhAG}_2$). The red backbone structure represents the RhAG molecule and the green structure represents the RhD molecules. $\mathrm{RhD}-\mathrm{RhAG}_2$ is the most probable structure for the Rh protein complex. Image courtesy of Conroy et al. (2005).

1.7 Rh Associated Glyoprotein (RhAG)

The *RHAG* gene is located on chromosome 6 with 10 coding exons. The RhAG protein is composed of 409 amino acids (Ridgwell et al., 1992; Reid et al., 2012; Polin et al., 2016). The RhD, RhCE, and RhAG proteins are expressed in 12 transmembrane spanning domains with both N-terminus and C-terminus in the cytoplasm (Reid et al., 2012; Polin et al., 2016). The

RhAG protein is essential for the expression of Rh proteins and different mutations in the *RHAG* gene have been linked to Rh deficiency. Over 20 *RHAG* alleles were linked to severe decline in Rh expression ($\mathrm{Rh}_{\mathrm{mod}}$) or complete absence of Rh ($\mathrm{Rh}_{\mathrm{null}}$) and related antigens like CD47 and ICAM-4. Different studies have linked *RHAG* specific SNPs to a weakened RhD expression (Tian et al., 2011; Polin et al., 2016; Mu et al., 2019; Wen et al., 2019). For example, c.241G>C mutation in the *RHAG* gene leads to the replacement of non-polar Glycine to polar Arginine at position 81 which is predicted to be at the third transmembrane -spanning domain of the protein. This change alters the hydrophobicity of the protein in that area which weakens the RhAG protein expression (Tian et al., 2011; Polin et al., 2016).

People with $\mathrm{Rh}_{\mathrm{mod}}$ or $\mathrm{Rh}_{\mathrm{null}}$ may suffer haemolytic anaemia and RBCs cation leak (Bruce et al., 2009). Mutations in the *RHAG* gene are also associated with RBC morphological changes, for example, stomatocytosis and spherocytosis Bruce et al. (2009), section 1.5.1.4.

1.8 Blood Group Antigenicity

Among the 55 different antigens in the Rh blood group system, RhD is the most clinically significant one. Antigen D is considered the highest immunogenic antigen in the Rh system (50x more than other Rh antigens) referring to its ability to elicit an immune response (Urbaniak and Greiss,

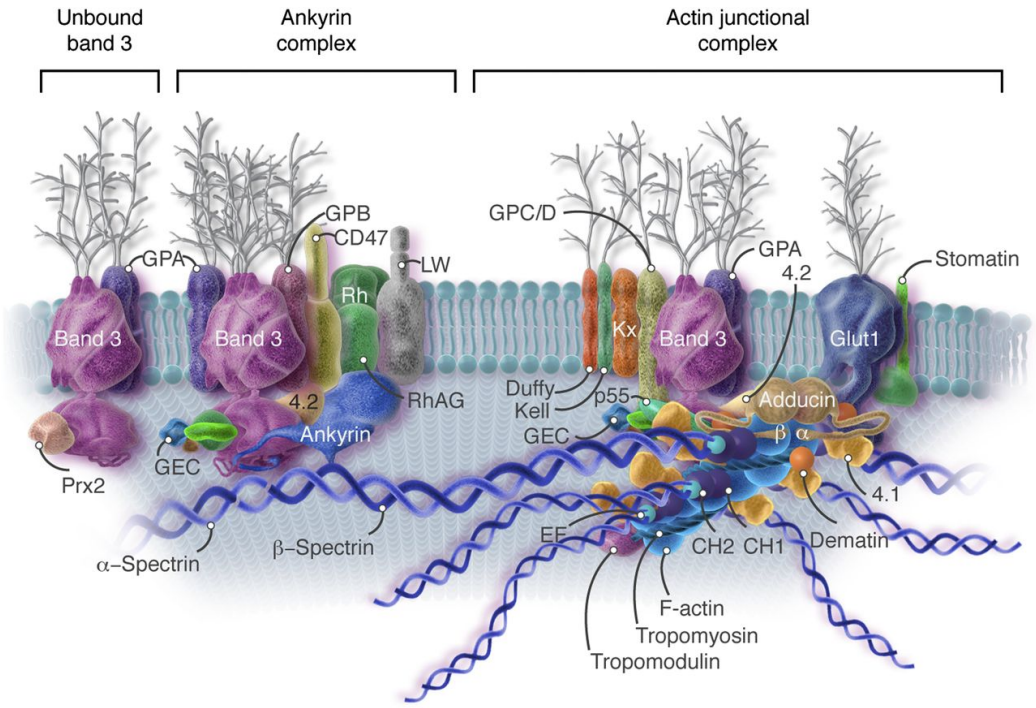


Figure 1.4: The interaction of the Rh protein complex with other proteins in the red cell membrane. Rh protein complex is integrated in the cell membrane by interacting with Band3 and CD47 proteins which connect to ankyrin and Band4.2, respectively. Image courtesy of Iwamoto (2005).

2000). The D antigen immunogenicity might be due to the high number of epitopes expressed on the cell surface and the position they take on the cell surface. , referring to its ability to elicit an immune response (Urbaniak and Greiss, 2000). Antigen D clinical significance comes from the ability of its antibodies to cause HTRs, HDFN, and autoimmune haemolytic anaemia (Le van Kim et al., 1992). The first two will be discussed in detail in the following section.

1.8.1 Transfusion Reactions

Since the discovery of the ABO blood group in the early 1900s, blood transfusion has become a common and a safe practice that saves life. However, adverse immune reactions following a blood transfusion can reduce the quality of life or in severe cases lead to death (Lubart et al., 2014). Transfusion reactions can range from mild to a severe life threatening events. Febrile non-haemolytic transfusion reactions (FNHTRs) are the most common post transfusion reactions, associated with fever, and chill-rigor symptoms without haemolysis, which cause by antibodies directed against donor leucocytes and human leukocyte antigens (HLA) antigens.

HTRs are caused when RBCs are destroyed by an immune response, which occurs as a result of transfusing an incompatible blood product. Haemolysis can be either intravascular, which is caused by immunoglobulin M (IgM), or extravascular, which is caused by immunoglobulin G (IgG). Intravascular HTRs are more common with the ABO incompatibility, where IgM binds to the membrane of the RBC activating the classical complement pathway, leading to direct penetration of the RBC's membrane. Symptoms usually include chills, shocks, low blood pressure, anaemia, and uraemia. Symptoms can be more severe and include disseminated intravascular coagulopathy and kidney failure. Intravascular HTRs fatality rate is about 10% (Klein and Anstee, 2005). Extravascular HTRs are most commonly involved with other blood groups incompatibility, like Kell and

Rh. It is caused when IgG (mainly IgG3 and IgG1) bind to the RBC's membrane targeting them for phagocytosing by splenic macrophages. Symptoms are the same involved with intravascular HTRs but are less severe and associated with a lower mortality rate (Poole and Daniels, 2007).

1.8.2 Haemolytic Disease of the Fetus and Newborn (HDFN)

HDFN occurs when an antigen-negative female has been immunised, whether through blood transfusion or pregnancy, to develop alloantibodies to an RBC's antigen that her baby expresses on the RBCs. Different antibodies are involved with HDFN; however, anti-D is the most common causative agent (Bowman, 1998; Urbaniak and Greiss, 2000; Kumpel, 2008). RhD negative females are susceptible to develop anti-D after delivering the first RhD positive baby or after Rh incompatible transfusion, (Figure 1.5). During labour some of the fetus's blood enters the mother's circulation through transplacental haemorrhage (TPH) which causes an immune response to develop antibodies against non-self-antigens, in this case anti-D. The initial immune response stimulates the production of IgM antibodies, which cannot cross the placenta because of their pentamer structure as they are not bound by the IgG transporter neonatal Fc receptor (FcRn). The secondary immune response causes production of IgG antibodies which are able to cross the placenta and cause HDFN. The passing of the IgG through the placenta is mediated by the FcRn located in the syncytiotrophoblast chorionic villi which bind the IgG Fc region in an acidic endo-

some. The IgG is released in physiological pH of villous stroma, which then bind FcRn on placental macrophages to be carried to fetal capillaries (Simister, 2003; Wilcox et al., 2017). The IgG antibodies cross the placenta and bind to the corresponding antigen on the RBCs of the fetus, which induces destruction of the fetal RBCs by the immune system (Urbaniak and Greiss, 2000; Poole and Daniels, 2007; de Haas et al., 2015).

The mechanism of cell lysis in HDFN is the same as other antibody mediated cell destruction, though Rh antibodies do not bind complement. The severity of cell destruction in HDFN depends on antibody concentration, antigen expression, and the placenta's ability to transfer antibodies. Haemolysis starts when IgG coated cells are removed mostly by splenic macrophages. The opsonised RBCs are either completely engulfed by the macrophages or lose a part of the cell membrane (scission) and go back to the circulation as microspherocytes, which reduces the cells' survival rate. Antibody-dependent cell-mediated cytotoxicity (ADCC) also plays a role in cell destruction (Urbaniak and Greiss, 2000; Kumpel and Elson, 2001).

Rapid fetal RBC destruction through splenic macrophages results in severe anaemia, and a high level of bilirubin in fetal circulation ("hyperbilirubinaemia"). If hyperbilirubinaemia is left untreated, bilirubin accumulation in the brain leads to a disease called "kernicterus" which causes irreversible damage to the nervous system associated with high morbidity rate (Kumpel, 2008; de Haas et al., 2015). In 1940, 1% of newborns suffered

HDFN, with about a 40% mortality rate (Kumpel, 2008).

In the 1960s, anti-D Ig prophylaxis (section 1.8.3) was introduced causing a significant decrease in the incidence of HDFN from 15% to 1.6% (de Haas et al., 2015). Mortality rate from HDFN decreased from 50% to 5-9% in the 1970s (Urbaniak and Greiss, 2000).

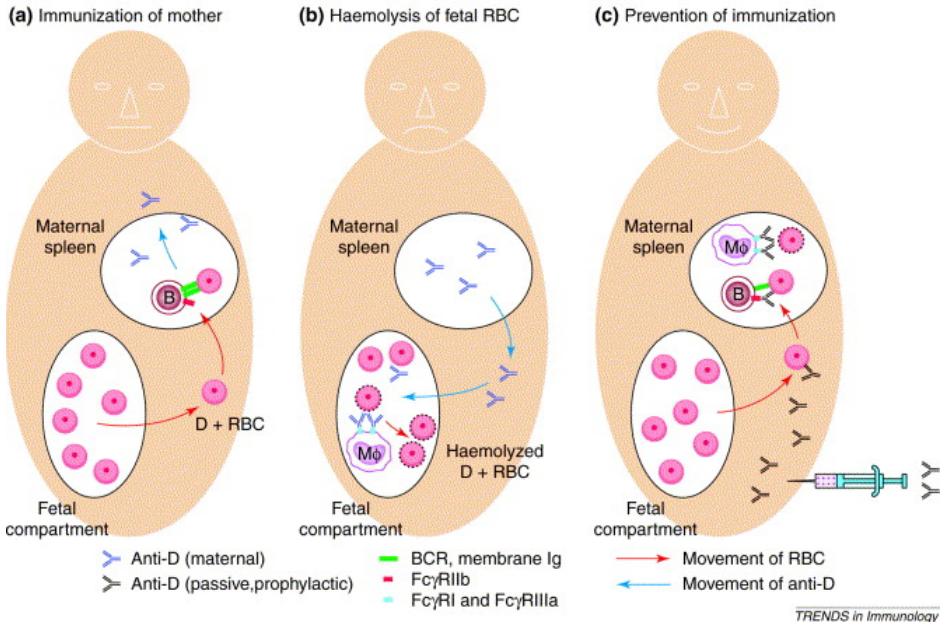


Figure 1.5: RhD incompatibility between RhD positive fetus and RhD negative mother leads to HDFN. (a) Fetal RhD positive RBCs enter the mother's circulation and elicit an immune response and eventually cause the production of anti-D (IgG) antibody by B cells in the secondary immune response. (b) In the second pregnancy, RhD IgG cross the placenta and bind to antigen D inducing haemolysis. (c) Prophylactic anti-D prevent the initiation of the first immune response by blocking B cell activation. Image courtesy of Kumpel and Elson (2001).

1.8.3 Anti-D Prophylaxis

Nowadays, the standard neonatal care includes administration of anti-D Ig (20 μM / 1 mL of packed RBCs) prophylaxis between the 28^{th} and the 34^{th} week of gestation and at birth to prevent sensitisation (NICE, 2007). Prophylactic anti-D, administered to RhD negative women to prevent antigen D immunisation, is the most successful application of antibody-mediated immune suppression (AMIS) (Urbaniak and Greiss, 2000; Kumpel and Elson, 2001; Kumpel, 2008), although the exact mechanism is not fully understood. There are different suggested mechanisms of how antibody administration can prevent an immune activation from the corresponding antigen. It is important to know that anti-D can only suppresses initial immune response. Therefore it is crucial to prevent primary immunisation because once initiated it cannot be reversed.

A suggested mechanism of action for anti-D is through inhibition of B cells activation through antigen-antibody complex. Antibody coated RBCs bind through antibody Fc region to Fcγ RIIb (which has an immunoreceptor tyrosine based inhibitory motif (ITIM) intracellularly) on B cells with concurrent B-cell receptor (BCR) binding to antigen D on the RBCs, (Figure 1.5). This complex receptor binding generates a signal of inhibition and can lead to the lysis of B cells (Kumpel and Elson, 2001; Avent, 2018).

A different mechanism suggests that anti-D completely occupies antigen

D binding sites not allowing BCR interaction, which causes inactivation of B cells. Anti-idiotypic antibodies (anti-Id) are also believed to participate in anti-D mechanism by cross-binding BCR with Fcγ RIIb preventing B cells activation (Kumpel et al., 1995; Anderson and Sinclair, 1998; Avent and Reid, 2000; Kumpel and Elson, 2001).

Anti-D prophylaxis has reduced HDFN occurrence about 50% since the 1970s (Visser et al., 2019) nevertheless, HDFN caused by Rh antibodies is still reported (Bhutani et al., 2013; Hassan et al., 2019). HDFN prevalence among RhD negative females varies, for example, from 1.5% in Chile to 19% in Brazil, 15% in White European (mainly East Europe), and about 0.5% in China but expected to rise after the amendment to the 1 child policy (Bhutani et al., 2013; Visser et al., 2019). HDFN occurs mainly as a result of failure to provide adequate dosage of anti-D prophylaxis to prevent sensitisation after a sensitisation episode such as abortion, amniocentesis, antepartum bleeding. Failure to provide an adequate dosage of anti-D can be as a result of lack of anti-D avilability in some countries like China (Visser et al., 2019).

Anti-D is a biological product and is collected through apheresis from donors with high anti-D concentration which could be a limited source in some countries (Kumpel and Elson, 2001; Visser et al., 2019). The lack of antenatal care in some developing countries in Africa and Asia is another cause of HDFN (Bowman, 1988; Urbaniak and Greiss, 2000; Bhutani et al.,

2013; Visser et al., 2019). According to Bhutani et al. (2013) HDFN results in 50,000 fetal deaths and 114,000 preventable neonatal morbidities. The majority of these cases (75%) are reported from Sub-Saharan Africa and South Asia (Visser et al., 2019).

The limited access to anti-D in some countries and the possibility of unnecessary administration of anti-D, as in the case of RhD negative fetuses, created the need for a test to determine whether anti-D administration was required. Fetal molecular *RHD* genotyping through testing cell free fetal DNA (cffDNA) proved to be effective to determine the fetal RhD status (Finning et al., 2008; Fan et al., 2009; Sillence et al., 2015). Using sequence-specific primers-polymerase chain reaction (SSP-PCR) is one of the molecular testing advancement to prevent HDFN (Clausen et al., 2012; de Haas et al., 2016; Hyland et al., 2017; Clausen, 2018).

1.8.4 Fetal *RHD* Genotyping

Prenatal *RHD* genotyping has been routinely applied in different countries such as Belgium (Minon et al., 2015), Finland (Haimila et al., 2017), the Netherlands (Oepkes et al., 2006; de Haas et al., 2016), and Denmark (Clausen, 2018) to monitor RhD negative pregnant females who either have been immunised or are susceptible to RhD immunisation (Hyland et al., 2017). cffDNA genotyping through targeting one or more of the *RHD* exons (5, 7, 10) by SSP-PCR has proven to be an accurate method for fetal *RHD* genotyping (Finning et al., 2008; Fan et al., 2009; Clausen et al.,

2012; Sillence et al., 2015; de Haas et al., 2016; Clausen, 2018). Reliable genotyping of the fetal *RHD* gene results in a better use of anti-D Ig, in which it is only administered to RhD negative females with RhD positive fetus and eliminates the irrelevant administration of anti-D Ig to RhD negative females with RhD negative fetus, which was about 40% of the RhD negative pregnancies (Blanco et al., 2018; Clausen, 2018).

1.9 Serological Testing

1.9.1 Antigen Detection

Blood group antigen detection depends on the agglutination of RBCs by the corresponding antibody (Landsteiner, 1961). Agglutination occurs in two steps: first, antigen-antibody reaction, in which antigen epitopes bind the Fab section of the antibody; second, multiple antigen-antibody binding occur building a web like structure that can be visualised as an agglutination (Mujahid and Dickert, 2015). There are different methods to carry out agglutination tests such as classical tube testing or more advanced assays such as microplate and gel centrifugation methods (Kim et al., 2006; Malomgré and Neumeister, 2009). The agglutination is influenced by different factors such as antigen:antibody ratio, temperature, and type of immunoglobulin used (Harmening, 2005).

1.9.2 Direct Antiglobulin Test (DAT)

Direct antiglobulin test (DAT) is used to detect in-vivo sensitised RBCs (i.e. antibody/ complement coated RBCs) as in case of HDFN and HTRs. Coombs et al. (1946) first described the use of an antiglobulin test to detect in-vivo sensitised RBCs of newborns diagnosed with HDFN. Anti-human globulin (AHG) could be mono- or poly-specific. Mono-specific anti-globulin reagents contain antibody to one specific component either anti-IgG or anti-complement. Poly-specific anti-globulin reagents contain antibodies against both immunoglobulin and complement. The test is carried out by using patient washed RBCs mixed with antiglobulin reagent, which are then incubated to optimise reaction conditions and then the reaction is examined for agglutination, (Figure 1.6). A positive result might indicate immunological activity against RBCs (Harmening, 2005; Pamphilon and Scott, 2007). Such immunological activity might include HDFN, HTRs, autoimmune haemolytic anaemia, or drug-induced haemolytic anaemia.

1.9.3 Indirect Antiglobulin Test (IDAT)

IDAT is used to detect in-vitro sensitisation of the RBCs, for example, the presence of unexpected antibodies in the serum (for e.g. antibody screening, crossmatch), and detect blood group antigen (for e.g. weak D). Incomplete antibodies (IgG), when bound to red cells, are often unable to produce visible agglutination depending on the number of antigen binding sites and antigen:antibody ratio in a reaction. Therefore, complete an-

tibodies (AHG) is added to cross linking the binding IgG to produce a visible agglutination. Coombs et al. (1945) described this test for detection of incomplete (weak) D expression in 1945.

To carry out the IDAT for antibody detection in patients' serum (Figure 1.6), serum is mixed with known cells (e.g. D+ cells) which are then incubated to allow antigen-antibody binding. After incubation, the mixture is washed to remove unbound antibodies and then AHG is added. If antibodies are present in the serum, the AHG bind to the antigen-antibody structure forming a visible agglutination (Coombs et al., 1945; Coombs and Roberts, 1959; Pamphilon and Scott, 2007). IDAT can also be used for blood antigen phenotyping (e.g. for Du test), where patients' red cells are tested with known antisera (IgG) which blocks antigen binding sites. Complete antibodies are then added to produce visible agglutination (AHG). Different factors affect the reaction such as the ratio of antibody to cells, temperature, and reaction medium (Harmening, 2005). Different reaction media have been reported to enhance antigen-antibody reactivity such as albumin (Stroup and MacIlroy, 1965), low ionic strength solution (LISS) (Löw and Messeter, 1974; Walker, 2018), and polyethylene glycol (PEG) (Shirey et al., 1994; Dinardo et al., 2014).

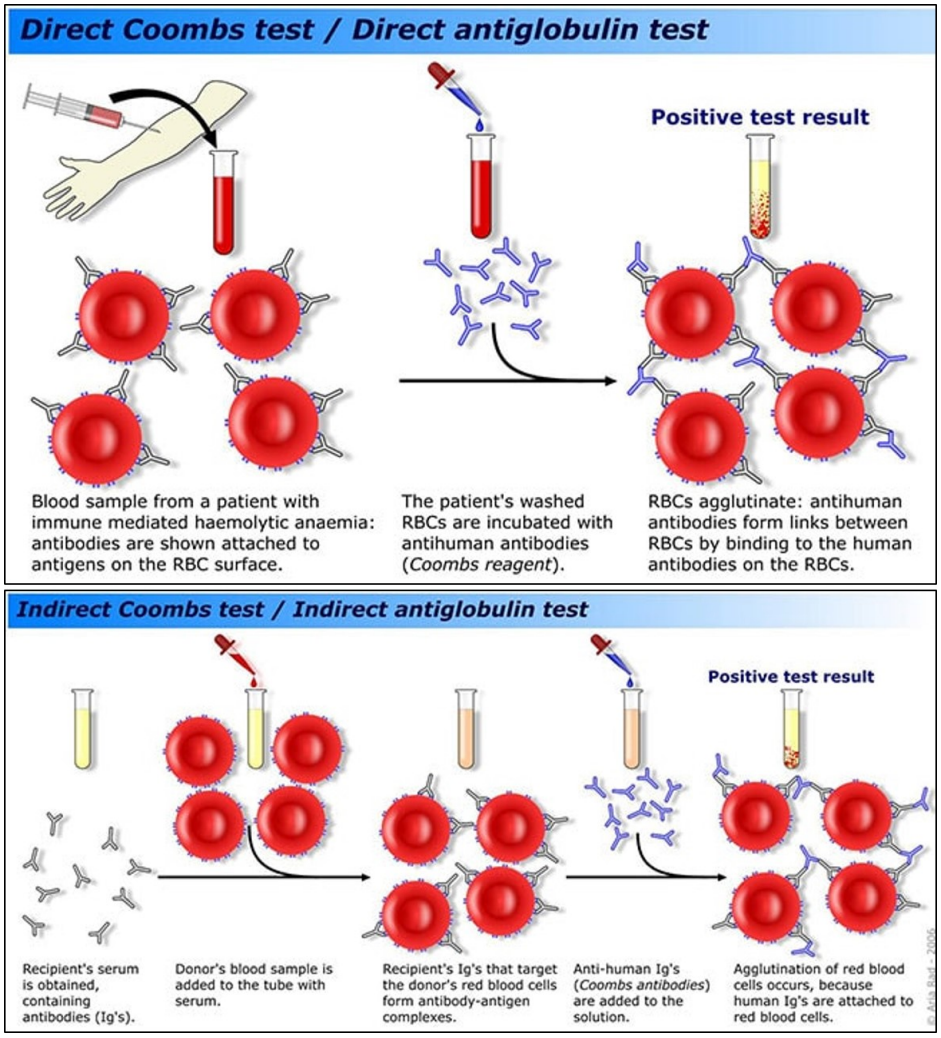


Figure 1.6: Direct and indirect antiglobulin testing. Image courtesy of Wikipedia (2015).

1.10 Blood Group Genotyping (BGG)

1.10.1 BGG Platforms

Although serological tests are fast, cost friendly, and efficient, they are limited by many factors, for example, the availability of antisera (e.g. no reagent is commercially available for Yt^{a}) (Jungbauer, 2011b), affinity and concentration of antibodies, and antigen status like weak or partial antigen expression. Genotyping will provide the freedom to genotype a wider range of blood antigens including low frequency antigens such as Go^{a}, BARC, and Tar (Daniels, 2013a; Boccoz et al., 2018). Complete blood group genotyping (BGG) could be widely used in transfusion practice where serology tests may fail to clarify issues or resolve discrepancies; for instant, in case of positive DAT or IDAT, ABO discrepancies, Rh phenotyping (Jungbauer, 2011a,b).

Extensive efforts to alternatively use molecular testing to genotype blood antigens have been made. These methods range from low, medium to high throughput (Fichou et al., 2014; Avent et al., 2015). Different deoxyribonucleic acid (DNA) microarray-based tests were introduced that enable genotyping of variant SNPs that are blood group gene specific (Hashmi et al., 2005; Avent et al., 2009; Goldman et al., 2015; Chang et al., 2016; Fichou and Férec, 2017; López et al., 2018; Bub and Castilho, 2019). For example, BLOODchip® ID (Progenika Biopharma, a Grifols Company, Derio,

Spain) (Goldman et al., 2015) and HEA BeadChip® (BioArray Solutions, Warren, NJ, USA) (Paccapelo et al., 2015), are glass microarrays that enable genotyping of the most important blood group antigens. IDCore allows genotyping of 10 of the most important blood group antigens which include Rh, Kell, Kidd, Duffy, MNS, Diego, Dombrock, Colton, Cartwright, and Lutheran (Avent et al., 2009; Finning et al., 2016). Although they are very accurate and efficient, these assays have built-in limitations specifically by the number of blood group gene specific SNPs they can detect (Boccoz et al., 2018).

They are designed to target certain nucleotides or DNA regions through polymerase chain reaction (PCR), while novel variants remain unknown (Avent et al., 2015; Fichou et al., 2014), which may require further testing using alternative methods like Sanger sequencing or exome sequencing (Lane et al., 2019). In addition, microarray based tests are limited by their primer and probe specificities. They detect specific SNPs in blood group genes on the assumption that the adjacent sequence including primer binding regions are known. Mutations that might compromise primer/probe sequences may cause false negative results. Therefore, it is important to continuously update these tests (Liu et al., 2014; Finning et al., 2016; Orzińska et al., 2018). Full gene sequencing would be the appropriate method to explore blood group gene variations and resolve issues that might be faced from using other genotyping platforms (Boccoz et al., 2018).

1.10.2 BGG Using NGS

Complete DNA sequencing could be the most relevant technique to thoroughly study blood group variations. Nevertheless, such an approach only allows a few samples to be studied at any one time in a given experiment, and is thus an ineffective implementation for screening. Nowadays, Next-generation sequencing (NGS) is used in HLA testing, which created a strong drive to introduce NGS for BGG (Möller et al., 2016).

NGS is a sequencing technology that has the capacity to sequence full genome (Orzińska et al., 2018), and a powerful technique that would overcome limitations present in current assays used for BGG. NGS of blood group gene specific SNPs would enable the screening of the clinically significant blood antigens for a group of samples in one setup (Orzińska et al., 2018), where serology or current BGG platform failed to provide clear results. For example, Wu et al. (2018) utilised NGS to resolve ABO discrepancies in blood donors and successfully determine *cis/trans* linkage of ABO genes. Full sequencing of blood group genes enables the maintenance and update of SNP targeting assays (Tilley and Grimsley, 2014; Lane et al., 2016; Orzińska et al., 2018) by identifying novel variants and overcomes the issue of primers/probes specificity.

Genotyping could also be utilized to genotype patients with chronic diseases (e.g. sickle cell disease (SCD), thalassaemia) that require multiple

blood transfusions, which place the patients at risk of developing alloantibodies that could be clinically significant (Bakanay et al., 2013) and may lead to further complications discussed in sections 1.8.1 and 1.8.2).

To prevent transfusion reactions in patients with alloantibodies, transfused blood should be lacking the corresponding antigen which patients have developed antibodies against. Therefore, multiple blood units need to be tested for the specific antigen. This process could be very frustrating if a patient has multiple antibodies, antibodies for a high frequency antigen, or a rare blood group.

BGG is a powerful approach for massive blood group genotyping in blood donors (Orzińska et al., 2018). Fully genotyping donors and creating a database would make finding compatible donors and recalling for transfusion easier (Ribeiro et al., 2009; Veldhuisen et al., 2009). Different studies have developed a system to introduce routine BGG for frequent donors to create a database to facilitate finding compatible blood units for patients with allo-antibodies (Perreault et al., 2009; St-Louis et al., 2010). Genotyping will reduce the possibility of patient sensitisation by increasing the compatibility between patients and donors through testing a wider range of blood group antigens (Jungbauer, 2011b).

NGS could also be used to study the molecular interaction of blood antigens with certain pathogens like *Plasmodium vivax* that utilises Duffy gly-

coprotein as a receptor for cell invasion, and study the relationship between blood groups and specific diseases (Möller et al., 2016). There are different techniques in sequencing, starting from first generation sequencing, NGS, and Third generation sequencing (TGS) which will be discussed in depth in the following section 1.11).

1.11 First Generation DNA Sequencing

In the 1970s, different methods were introduced to sequence DNA which included chemical breakdown of the DNA (Maxam and Gilbert, 1977) and the incorporation of the terminating dideoxynucleotide triphosphates (ddNTPs) during DNA synthesis, introduced by Frederick Sanger in 1975 (Sanger et al., 1977). The products were then separated by size on a polyacrylamide gel, smaller products run faster than bigger products on the gel, creating a pattern. To conclude the DNA sequence, the separation pattern on the gel is inspected visually. In the following years, this technique, which was called "Sanger sequencing", was hugely improved and was firstly introduced commercially by Applied Biosystems in 1986 (Anderson and Schrijver, 2010; Pareek et al., 2011).

The sequencing starts first with amplifying the gene of interest either through PCR or by cloning. After successful amplification, the DNA is purified and a specific primer is annealed in the presence of DNA polymerase that will extend the DNA synthesis. During the DNA synthesis, a random addition

of fluorescently labelled ddNTPs terminates the synthesis. Fluorescently labelled DNA of varying sizes is subjected to capillary electrophoresis and the resulting pattern of fluorescent peaks determines the DNA sequence.

The Sanger method can sequence DNA fragments of up to 1200 bp (Zhang et al., 2011). The Sanger method was used in the Human Genome Project (HGP) but in order to sequence longer DNA segments, shotgun sequencing was introduced. Shotgun sequencing is a technology used to sequence the human genome by targeting DNA for random fragmentation and then use overlapping reads to reassemble the genome. Sequencing library preparation included DNA fragmentation into smaller fragments, fragment amplification in vectors, and then sequencing of each fragment separately. Depending on the sequence overlap, the data generated was assembled creating a complete sequence. The human genome was successfully sequenced in 13 years and cost about $ 2.7 billion (Venter et al., 2001; Lander et al., 2001; Margulies et al., 2005; Zhang et al., 2011).

Although Sanger sequencing has > 99.9% raw base accuracy and can sequence for up to 1 kb DNA, its accuracy in sequencing highly polymorphic genomic regions like major histocompatibility complex (MHC) genes is about 10-20%. Very low accuracy in sequencing means it is insufficient to detect infrequent clinically significant alleles (Anderson and Schrijver, 2010).

1.12 Next Generation Sequencing (NGS)

Since the submission of the HGP first draft in 2000 (Lander et al., 2001), there was a need to establish a sequencing technique that is cost efficient, fast, and has high throughput. This would enable scientists to apply DNA sequencing in research and clinical use to allow in-depth genome sequencing to provide thorough understanding of genomic variations in a population, susceptibility to diseases, and pharmacogenomics reactions. These techniques were called NGS or the second generation sequencing (Anderson and Schrijver, 2010; Pareek et al., 2011). In this work, the term NGS is used as it is more commonly used.

In 2000, the Genome Sequencer 20 by 454 Life Sciences was introduced as the first NGS system. The GS 20 combined the single-molecule emulsion PCR and pyrosequencing, which indicates the nucleotide additions by the detection of pyrophosphate release (Pareek et al., 2011). Although the sequencing technology in NGS and Sanger sequencing are different (Figure 1.7), they both require the presence of a specific DNA polymerase that will carry out the reaction to generate a newly synthesized DNA, which is called DNA sequencing-by-synthesis (Chen, 2014). Another mechanism for NGS is DNA sequencing-by-ligation which uses DNA ligase mismatch sensitivity to determine the sequence of a given DNA (Ho et al., 2011; Huang et al., 2012). In this mechanism, single stranded DNA fragments are annealed to an anchor sequence to which the probes and primers bind

and a pool of fluorescent labelled oligonucleotide probes of different length is used with DNA ligase to carry out the sequencing. The DNA ligase seals the nick between the 5′ of the growing strand and the 3′ of the oligo (Ho et al., 2011; Huang et al., 2012). After a nucleotide is incorporated, there is a release of fluorescence, which indicates the nucleotide has been incorporated, the primer and anchor primer disassociate resetting the sequencing and allowing the process to re-initiate (Ho et al., 2011).

Different companies offer different NGS platforms but mostly dominated by Roche's 454, Illumina's Solexa, and Life Technologies' SOLiD (Huang et al., 2012). NGS technology, which is described as massively parallel sequencing, depends on reading randomly through the DNA which is accomplished by breaking the DNA into smaller fragments which are then ligated to adaptors for sequencing-by-synthesis. The read length for NGS is much shorter than the one offered in Sanger sequencing, 50 to 500 continuous bp, are referred to as short reads. These short reads are one of the down sides of NGS as short reads could generate a sequence that cannot be assembled either because it's too short or it matches to different regions in the gene.

Depth of coverage is another critical issue in NGS. Depth of coverage is known as the frequency of short reads that overlap in a defined genomic area; for example, if the depth of coverage for a gene is 30, that means each nucleotide in that gene has been expressed in at least 30 different and

overlapping short reads. Adequate coverage is crucial for precise assembly of the sequence and variant calling (Zhang et al., 2011; Schatz et al., 2010). sequencing quality is another important marker in NGS which is measure by PHRED score. PHRED score is an algorithmic integer value representing the estimated probability of an error in the identification of a nucleic base, for instance, PHRED score of 10 represents a 1/10 (10%) error rate, 20 represents a 1/100 (1%) error rate, and 30 represents a 1/1000 (0.1%) error rate (Ewing et al., 1998; Liao et al., 2017).

In the last few years, NGS has changed the genetic research field by elevating both throughput and data generated, at the same time lowering significantly the cost of sequencing per nucleotide (Fichou et al., 2014) up to 100-fold reduction in cost/bp (Pareek et al., 2011). NGS allows whole genome sequencing (WGS) generating a complete sequence data that will reveal any novel mutations. In BGG, NGS will allow the detection of silencing or weakening expression alleles (Avent et al., 2015). Due to this technological advancement, scientists were successful in decoding the human genome, revealing rare variants and utilizing the genomic advances in cancer researches. Because of the NGS high impact on the research, it was chosen as the method of the year in 2007 (Anderson and Schrijver, 2010). Modern genomic technology has contributed vastly to our understanding of various biological phenomena and altered our understanding of genes linked to specific diseases (Zhang et al., 2011).

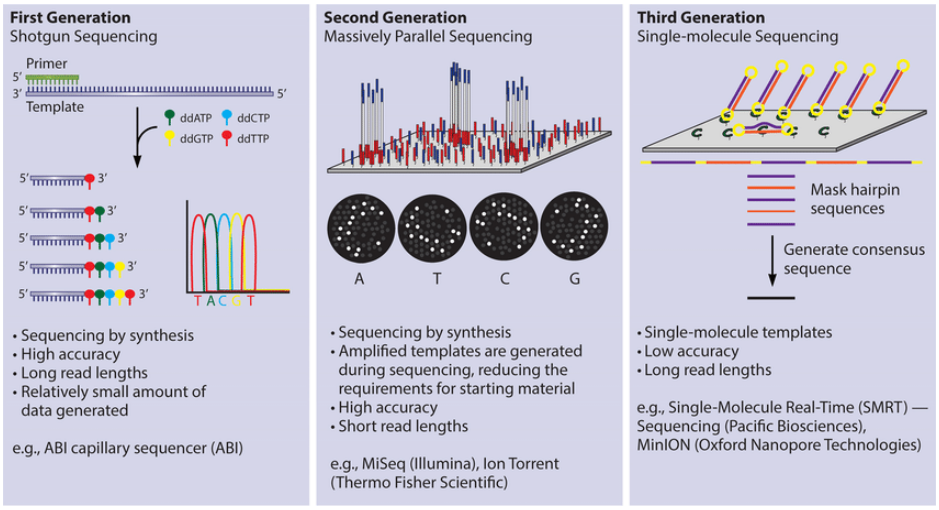


Figure 1.7: Comparison between first, second and third generation sequencing technologies. Image courtesy of Patterson et al. (2020).

1.12.1 Ion Torrent Personal Genome Machine™ (Ion PGM™)

NGS, in this study, was carried out on the Ion Torrent Personal Genome Machine™ (Ion PGM™) (Life Technologies, UK) to genotype for Rh blood group genes. Ion PGM™ is a metal-oxide semiconductor sequencer that uses pH sensitivity as an indicator for nucleotide addition. An ion-sensitive field-effect transistor (ISFET) sensor is used to detect the H^+ ion release by the DNA polymerase during sequencing by synthesis (Lu et al., 2016; Schatz et al., 2010). Wells, 3.5 mm-diameter, with a sensor plate is used to carry out the sequencing (Figure 1.8). In the Ion PGM™ sequencer, all four nucleotides are provided consecutively during DNA synthesis. If a complementary nucleotide was added, the synthesized strand becomes one

base longer causing an increase in negative charge resulting in a single H^+ ion release that alters the surrounding pH by 0.02 pH units (Pourmand et al., 2006; Rothberg et al., 2011). This shift in the pH is detected by a sensor in the bottom of the wells. Through electronic reads and microprocessor software, these changes will be converted to electrical signals which are converted to base calls by using signal processing software (Rothberg et al., 2011). Generated data can then be assembled with a reference sequence or *de novo* and analysed using computer software.

1.13 Third Generation Sequencing (TGS)

Genomics was radically reshaped during the past two decades by the introduction of the first and then the second generation sequencing (Lu et al., 2016). Next generation sequencing is a high-throughput sequencing, depends on PCR amplification, that generates a vast amount of data in a low cost. However, short reads generated during library preparation made *de novo* assembly for large genomes difficult due to repetitive DNA sequence (Lu et al., 2016; Zhang et al., 2011; Schatz et al., 2010), which brought the need for a new sequencing approach that uses longer reads later named TGS and also called single molecule sequencing (SMS) (Lu et al., 2016). Unlike NGS, TGS technology does not rely on PCR and aims for DNA SMS with real time sequencing and data analysis. The PCR free approach in TGS abolished sequencing biases introduced by PCR (Schadt et al., 2010)

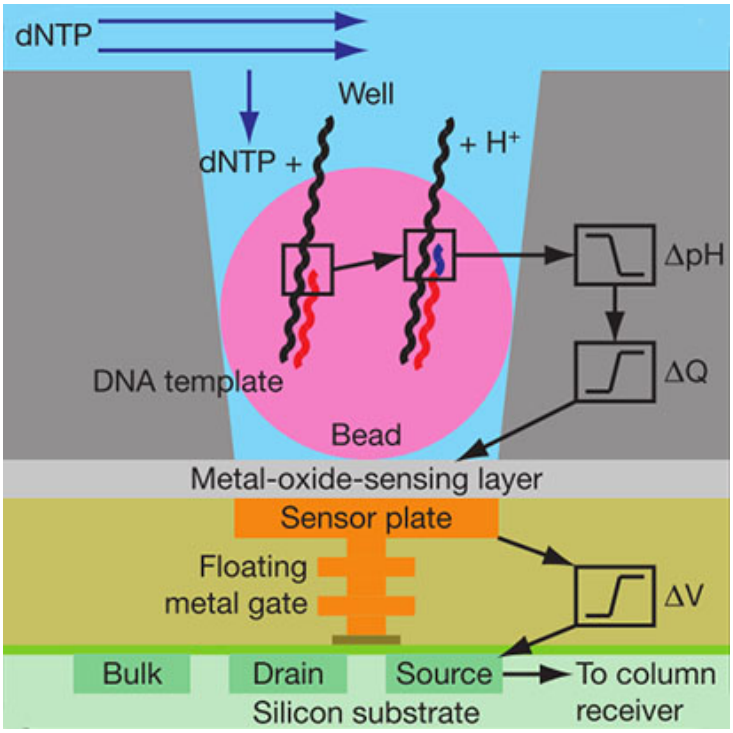


Figure 1.8: Sequencing chemistry of the Ion PGM™. Well, DNA template captured in a bead, an underneath sensor, and electronics are the main elements for sequencing in the Ion PGM™. When a nucleotide is incorporated, a proton H^+ is released altering the pH of the well. Off-chip electronics transform the voltage changes to digital data. Image courtesy Rothberg et al. (2011).

and allowed the detection of epigenetic modifications in the DNA. The advancement of TGS reduced time of library preparation and sequencing from days to hours (Lu et al., 2016).

The first successful single-molecule real time (SMRT) sequencer was established by Pacific Biosciences that produced 0.5-1 gigabytes (GB) per SMRT cell with read lengths of about 10 kb. However, error rate is about 13-15% higher than NGS (Nagarajan and Pop, 2013; Ardui et al., 2018).

1.13.1 Nanopore MinION™

MinION™ introduced by Oxford Nanopore Technologies in 2014 (Meller et al., 2000, 2001) is the smallest portable nanopore based sequencing device on the market that offer different cost-efficient sequencing kits to meet various sequencing needs. MinION™ sequencer technology is based on flowcell containing 512 pores that are derived from *Escherichia coli* Curli (Goyal et al., 2014; Rang et al., 2018). Curli are the main proteinaceous elements of the extracellular matrix produced by different *Enterobacteriaceae* which play essential roles in the pathogenicity of the bacteria (Barnhart and Chapman, 2006). These Curli-derived nanopores are embedded in a synthetic membrane submerged in ionic solution (Meller et al., 2000; Plesivkova et al., 2019). Applying a voltage, DNA molecule is driven through the pores causing changes in the ionic current running through the pores in a distinctive manner, described as "squiggle" (Rang et al., 2018; Plesivkova et al., 2019). These changes are measured by a sensor thousands of times per second (Eisenstein, 2017), which are then translated to nucleotides using software in a process known as basecalling, (Figure 1.9).

MinION™, when first introduced, had a 65-88% accuracy rate (Laver et al., 2015; Lu et al., 2016), however, recent advancement in sequencing chemistry and computational software led error rate to decrease significantly to 5-15 % Rang et al. (2018). Sequencing yield and accuracy, when used for whole microbial genome sequencing, reached 97% for 2D chemistry ex-

periments, and 94% for 1D experiments (Tyler et al., 2018).

MinION™ sequencer has captivated researchers interest, specifically in infectious agent surveillance and clinical diagnosis as they would benefit the most from real-time sequencing technology (Lu et al., 2016). Studies have shown the great potential of the MinION™, for example, during the Ebola virus breakout in Guinea (Quick et al., 2016) and Zika virus epidemic in Brazil in 2015 (Faria et al., 2017). MinION™ was also utilised to sequence the SARS-CoV-2 during COVID-19 pandemic from infected individuals in Ecuador and Colombia (Marquez et al., 2020; Lopez-Alvarez et al., 2020). Different studies have utilised the Nanopore sequencer to detect DNA and RNA modification, such as methylation in bacterial and mammalian genome (Rand et al., 2017; Simpson et al., 2017; Garalde et al., 2018; Xu and Seki, 2020). Although the use of nanopore sequencing has not been widely investigated in blood group genotyping (BGG), it has been utilised and proven effective for clinical genotyping of human HLA in different studies (Lang et al., 2018; Ton et al., 2018; Matern et al., 2020; De Santis et al., 2020). Matern et al. (2017) utilised MinION™ in ABO genotyping by sequencing a 7 kb amplicon, covers the region of exons 6 and 7, which was successful in differentiating between six ABO genotypes (AA, AO, BB, BO, AB, OO).

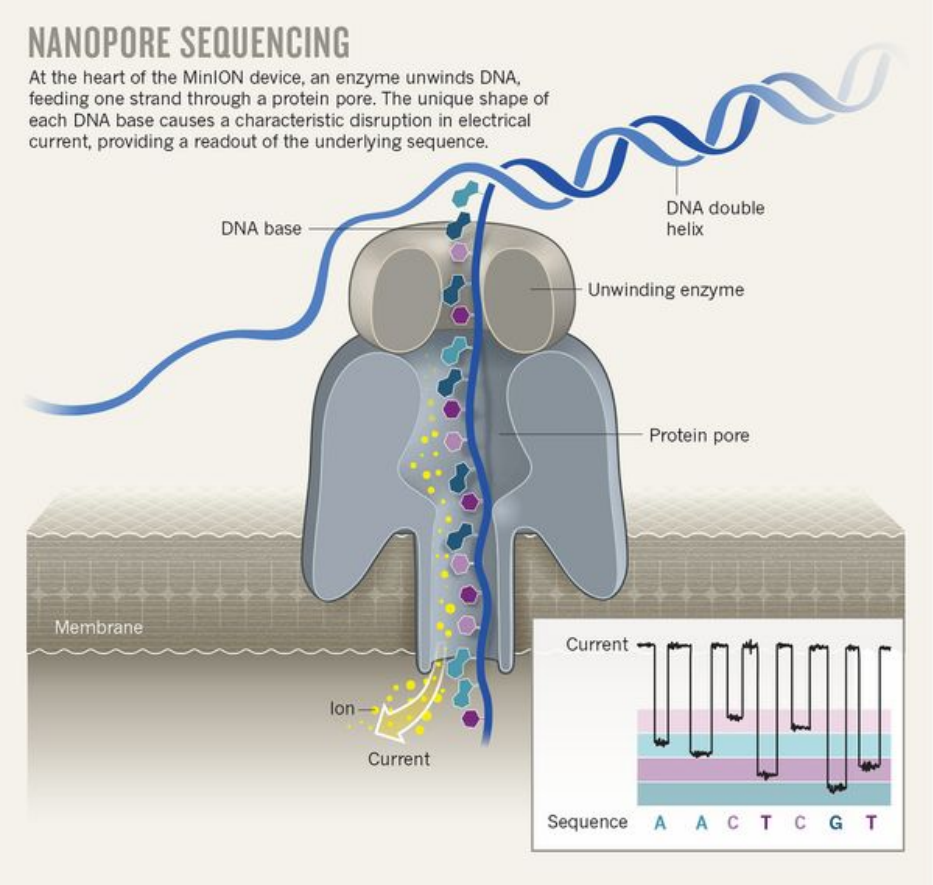


Figure 1.9: MinION™ sequencing chemistry. Protein nanopores are imbibed in a membrane and a voltage is applied to drive a single stranded DNA through the pore. As the DNA passes through the pore, changes in the current are measured by a sensor thousands time per second and recorded as "squiggle" (bottom right image) which is then translated to nucleotides in a process known as basecalling. Image courtesy Eisenstein (2017).

1.14 Study Aims and Objectives

Unlike serological testing, BGG provides the freedom to analyse a wider range of blood antigens including low frequency antigens that are clinically significant and can lead in the case of discrepancy to clinical compli-

cations. BGG could be widely used in transfusion practice, and is specially beneficial to transfusion dependent patients to minimize the risk of sensitisation and risks associated with that. NGS is a powerful technique to be used for BGG as it provides high throughput data and allows exploring the novel alleles of blood group genes. TGS is the new sequencing technology that directly sequences the DNA molecule allowing real time sequencing and data analysis, while reducing preparation time and cost.

This study aimed to use NGS (Ion PGM™) to develop a methodology to fully sequence the Rh genes including *RHD*, *RHCE* and *RHAG* to study variants present in the population and establish Rh haplotype specific reference sequences through studying intronic SNPs suspected to be Rh haplotype specific. We also aimed to use TGS (MinION™) to sequence the *RHD* gene to compare between the two sequencing technologies and assess the use of TGS for BGG. To achieve these aims, objectives were as follow:

1. Design target specific long range-polymerase chain reaction (LR-PCR) primers to amplify the complete *RHD* and *RHCE* genes including promoter, introns and all exons for sequencing. Complete sequencing of the Rh genes will enable the identification of known and novel variants encoded by exonic and/or intronic mutations.

2. All primers will be assessed for specificity to ensure that amplifi-

cation is target specific and not from the other homologous gene. Primers are tested with different Rh alleles for the optimisation of LR-PCR conditions to ensure adequate amplification is achieved from different Rh alleles.

3. Prepare sequencing libraries and sequence the target genes using NGS on Ion PGM™ platform.

4. Map sequence reads to the human genome reference sequence (hg38) and analyse data. Call variants to identify exonic and intronic mutations in the sequence to assess for allele determination.

5. Analyse intronic SNPs and their relation to a specific Rh haplotype and create intronic SNP patterns for these haplotypes to establish Rh allele specific refernce sequences.

6. Use MinION™ to fully sequence the *RHD* gene to test the suitability and efficiency of the MinION™ sequencer in BGG. Data generated from samples sequenced using MinION™ will be compared to the sequence data for the same samples generated by Ion PGM™ which will allow assessment of the MinION™ ability to achieve high quality reads and efficiency in calling variants to aid allele determination.

2. Materials and Methods

2.1 Blood Sample Collection

Donor blood samples (n=146) were supplied in anti-coagulation ethylenediaminetetraacetic acid (EDTA) tubes by the NHSBT, Bristol, United Kingdom. Inclusion criteria for blood samples was either their Rh haplotype (R_1, R_2, R_0, R_z, r, r′, r″) (n=118) or by their D reactivity (weak D) (n=28). Samples were serologically phenotyped for ABO, Rh and other blood groups by the NHSBT and were properly consented, anonymised, and supplied with a full ethical approval. Blood was processed in our laboratory after about 96 hours post collection. Blood tubes were centrifuged at 2500 xg for 10 minutes at room temperature. Plasma on the top layer was carefully disposed and buffy coat was collected into a 1.5 mL tube while the remaining content was discarded.

Genomic DNA (gDNA) samples (n=23) were from the ISBT 1996 work-

shop (ISBT, 1996). These samples were identified as *RHD* variants and were included in this work to determine and compare their sequence to other *RHD* variant alleles. All samples are listed in Table 2.1 including information supplied with them.

gDNA samples (n=35) were supplied by the Finnish Red Cross Blood Service (Helsinki, Finland) by Silja Tammi, Susanna Sainio, and Katri Haimila. Samples were collected from pregnant females who participated in the national screening program for HDFN in Finland between February 2014 and December 2016 (Haimila et al., 2017). Samples tested serologically as D-negative, very weak D or showed inconsistent RhD serology results. Pregnant women with partial D variants and certain weak D variants, as well as those that have inconsistent phenotyping and genotyping results, were considered D-negative, so that they are included in the prophylaxis program, due to the risk of immunisation. The samples were assessed using the SSP-PCR test which showed that all 35 samples were *RHD* gene positive (Tammi, 2019). Figure 2.1 shows the number of samples supplied by NHSBT, ISBT 1996 workshop, and the Finnish Red Cross Blood Service and the number of samples sequenced for each target including *RHD*,

RHCE, and *RHAG*.

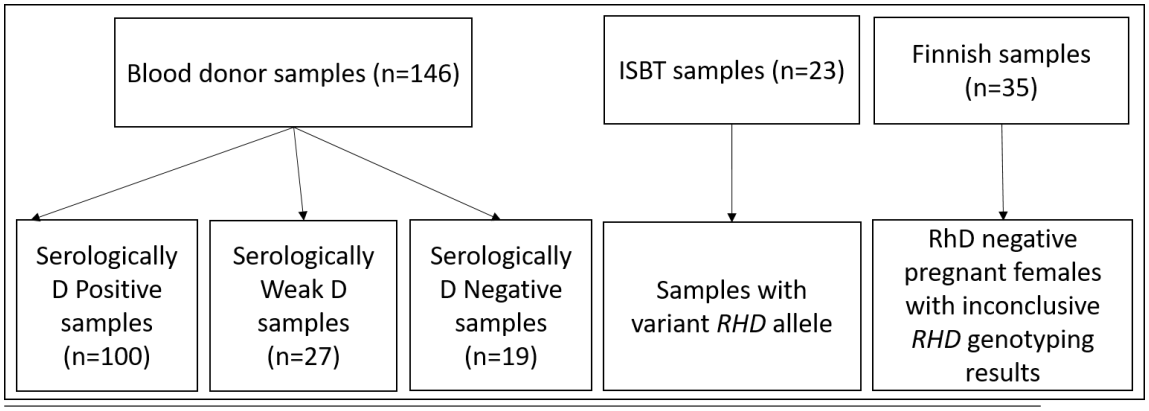


Figure 2.1: Flowchart illustrates the number of samples supplied by NHSBT, ISBT 1996 workshop, and the Finnish Red Cross Blood Service.

Table 2.1: Ethnicity of donors, Rh serology results, and predicted Rh genotype for samples included in the study as supplied.

		Rh Serology Results						
Sample no.	**Ethnicity**	**D**	**C**	**c**	**E**	**e**	**Rh haplotype**	***RHD* allele**
004_01	Caucasian	+	+	-	-	+	R_1R_1	ND
004_02	Caucasian	+	+	-	-	+	R_1R_1	ND
004_03	Other	+	+	-	-	+	R_1R_1	ND
004_04	Caucasian	+	+	-	-	+	R_1R_1	ND
004_05	Caucasian	+	+	-	-	+	R_1R_1	ND
004_06	Caucasian	+	+	-	-	+	R_1R_1	ND
004_07	Caucasian	weak D+	+	-	-	+	R_1R_1	ND
004_08	Caucasian	+	+	+	-	+	R_1r	ND
004_09	Caucasian	+	+	+	-	+	R_1r	ND
004_10	Chinese	+	+	+	-	+	R_1r	ND
004_11	Caucasian	+	+	+	-	+	R_1r	ND
004_12	Caucasian	+	+	+	-	+	R_1r	ND
004_13	Caucasian	+	+	+	-	+	R_1r	ND
004_14	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_15	Caucasian	+	+	+	-	+	R_1r	ND
004_16	Caucasian	weak D+	+	+	-	+	R_1r	ND

continued ...

Sample no.	Ethnicity	D	C	c	E	e	Rh haplotype	Rh Genotype
004_17	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_18	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_19	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_20	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_21	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_22	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_23	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_24	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_25	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_26	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_27	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_28	Caucasian	weak D+	+	+	-	+	R_1r	ND
004_29	Caucasian	+	+	+	+	+	R_1R_2	ND
004_30	Caucasian	+	+	+	+	+	R_1R_2	ND
004_31	Caucasian	+	+	+	+	+	R_1R_2	ND
004_32	Caucasian	+	+	+	+	+	R_1R_2	ND
004_33	Caucasian	+	+	+	+	+	R_1R_2	ND
004_34	Caucasian	+	+	+	+	+	R_1R_2	ND
004_35	Caucasian	weak D+	+	+	+	+	R_1R_2	ND
004_36	Caucasian	+	+	+	+	+	R_1R_2	ND
004_37	Caucasian	+	+	+	+	+	R_1R_2	ND
004_38	Caucasian	+	+	+	+	+	R_1R_2	ND
004_39	Caucasian	+	+	+	+	+	R_1R_2	ND
004_40	Caucasian	+	+	+	+	+	R_1R_2	ND
004_41	Caucasian	+	-	+	+	-	R_2R_2	ND
004_42	Not disclosed	weak D+	-	+	+	-	R_2R_2	ND
004_43	Caucasian	+	-	+	+	-	R_2R_2	ND
004_44	Caucasian	+	-	+	+	-	R_2R_2	ND
004_45	Caucasian	+	-	+	+	-	R_2R_2	ND
004_46	Caucasian	+	-	+	+	-	R_2R_2	ND
004_47	Caucasian	+	-	+	+	-	R_2R_2	ND
004_48	Caucasian	+	-	+	+	+	R_2r	ND
004_49	Caucasian	+	-	+	+	+	R_2r	ND
004_50	Caucasian	+	-	+	+	+	R_2r	ND
004_51	Caucasian	+	-	+	+	+	R_2r	ND
004_52	Caucasian	+	-	+	+	+	R_2r	ND
004_53	Caucasian	+	-	+	+	+	R_2r	ND
004_54	Caucasian	weak D+	-	+	+	+	R_2r	ND
004_55	Caucasian	weak D+	-	+	+	+	R_2r	ND
004_56	Caucasian	weak D+	-	+	+	+	R_2r	ND

continued ...

Sample no.	Ethnicity	D	C	c	E	e	Rh haplotype	Rh Genotype
004_57	Caucasian	weak D+	-	+	+	+	R_2r	ND
004_58	Caucasian	+	-	+	+	+	R_2r	ND
004_59	Caucasian	+	-	+	+	+	R_2r	ND
004_60	Caucasian	weak D+	-	+	+	+	R_2r	ND
004_61	Caucasian	weak D+	-	+	+	+	R_2r	ND
004_62	Caucasian	weak D+	-	+	+	+	R_2r	ND
004_63	Caucasian	+	-	+	-	+	R_0r	ND
004_64	Caucasian	+	-	+	-	+	R_0r	ND
004_65	Caucasian	+	-	+	-	+	R_0r	ND
004_66	Caucasian	+	-	+	-	+	R_0r	ND
004_67	Caucasian	+	-	+	-	+	R_0r	ND
004_68	Caucasian	+	-	+	-	+	R_0r	ND
004_69	Caucasian	weak D+	-	+	-	+	R_0r	ND
004_70	Caucasian	weak D+	-	+	-	+	R_0r	ND
004_71	Caucasian	+	-	+	-	+	R_0r	ND
004_72	Caucasian	+	-	+	-	+	R_0r	ND
004_73	Caucasian	+	-	+	-	+	R_0r	ND
004_74	Caucasian	weak D+	+	+	+	-	R_2R_z	ND
004_75	Caucasian	-	-	+	-	+	rr	ND
004_76	Caucasian	-	-	+	-	+	rr	ND
004_77	Caucasian	-	-	+	-	+	rr	ND
004_78	Caucasian	-	-	+	-	+	rr	ND
004_79	Caucasian	-	-	+	-	+	rr	ND
004_80	Caucasian	-	-	+	-	+	rr	ND
004_81	Caucasian	-	+	+	-	+	r′r	ND
004_82	Caucasian	-	+	+	-	+	r′r	ND
004_83	Caucasian	-	+	+	-	+	r′r	ND
004_84	Caucasian	-	+	+	-	+	r′r	ND
004_85	Caucasian	-	+	+	-	+	r′r	ND
004_86	Black	-	+	+	-	+	r′r	ND
004_87	Black	-	+	+	-	+	r′r	ND
004_88	Caucasian	-	-	+	+	+	r″r	ND
004_89	Caucasian	-	-	+	+	+	r″r	ND
004_90	Caucasian	-	-	+	+	+	r″r	ND
004_91	Caucasian	-	-	+	+	+	r″r	ND
004_92	Caucasian	-	-	+	+	+	r″r	ND
004_93	Caucasian	-	-	+	+	+	r″r	ND
004_94	ND	ND	ND	ND	ND	ND	ND	DIII
004_95	ND	ND	ND	ND	ND	ND	ND	DIII

continued . . .

Sample no.	Ethnicity	D	C	c	E	e	Rh haplotype	Rh Genotype
004_96	ND	+	+	+	-	+	R_1r	DIII
004_97	ND	ND	ND	ND	ND	ND	ND	DVII
004_98	ND	ND	ND	ND	ND	ND	ND	DIIIb
004_99	ND	ND	ND	ND	ND	ND	ND	DIVb
004_100	ND	ND	ND	ND	ND	ND	ND	DIVb
004_101	ND	+	+	+	-	+	R_1r	DVa
004_102	ND	+	+	+	-	+	R_1r	DVI
004_103	ND	+	+	+	-	+	R_1r	DVI
004_104	ND	+	-	+	+	+	R_2r	DVI
004_105	ND	+	-	+	+	+	R_2r	DVI
004_106	ND	ND	ND	ND	ND	ND	ND	DVI
004_107	ND	ND	ND	ND	ND	ND	ND	DVII
004_108	ND	+	+	+	-	+	R_1r	DVII
004_109	ND	+	+	+	-	+	R_1r	DVII
004_110	ND	+	+	+	-	+	R_1r	DVII
004_111	ND	+	+	+	-	+	R_1r	DFR
004_112	ND	ND	ND	ND	ND	ND	ND	DFR
004_113	ND	ND	ND	ND	ND	ND	ND	C^w
004_114	ND	ND	ND	ND	ND	ND	ND	C^w
004_115	ND	ND	ND	ND	ND	ND	ND	C^x
004_116	ND	ND	ND	ND	ND	ND	ND	C^x
004_117	ND	-	+	+	-	+	ND	ND
004_118	ND	-	+	+	-	+	ND	ND
004_119	ND	-	+	+	-	+	ND	ND
004_120	ND	-	+	+	-	+	ND	ND
004_121	ND	-	+	+	-	+	ND	ND
004_122	ND	-	+	+	-	+	ND	ND
004_123	ND	-	+	+	-	+	ND	ND
004_124	ND	weak D+	+	+	-	+	ND	ND
004_125	ND	weak D+	+	+	-	+	ND	ND
004_126	ND	-	+	+	-	+	ND	ND
004_127	ND	-	+	+	-	+	ND	ND
004_128	ND	-	+	+	-	+	ND	ND
004_129	ND	-	+	+	-	+	ND	ND
004_130	ND	-	+	+	-	+	ND	ND
004_131	ND	-	+	+	-	+	ND	ND
004_132	ND	weak D+	+	+	-	+	ND	ND
004_133	ND	-	+	+	-	+	ND	ND
004_134	ND	-	+	+	-	+	ND	ND
004_135	ND	-	+	+	-	+	ND	ND

continued ...

Sample no.	Ethnicity	D	C	c	E	e	Rh haplotype	Rh Genotype
004_136	ND	-	+	+	-	+	ND	ND
004_137	ND	-	+	+	-	+	ND	ND
004_138	ND	-	+	+	-	+	ND	ND
004_139	ND	-	+	-	-	+	ND	ND
004_140	ND	weak D+	+	-	-	+	ND	ND
004_141	ND	weak D+	+	-	-	+	ND	ND
004_142	ND	-	+	+	+	+	ND	ND
004_143	ND	-*	+	+	+	+	ND	ND
004_144	ND	-	-	+	+	+	ND	ND
004_145	ND	-	-	+	+	+	ND	ND
004_146	ND	-	-	+	+	+	ND	ND
004_147	ND	weak D+	-	+	+	+	ND	ND
004_148	ND	+	-	+	-	+	ND	ND
004_149	ND	-	-	+	-	+	ND	ND
004_150	ND	-	-	+	-	+	ND	ND
004_151	ND	-	-	+	-	+	ND	ND
004_152-004_205	Multiple	+	+	+	+	+	R_1R_2	ND

Grey highlight indicates samples were supplied by the NHSBT, Bristol, United Kingdom.
Green highlight indicates samples were from the ISBT workshop 1996.
Blue highlight indicates samples were supplied by the Finnish Red Cross Blood Service, Helsinki, Finland.
ND = No data available.
* Very weak D reaction in serology testing.
** RhD antigen titre after elution +1 as reported by the Finnish research group (Tammi, 2019).

2.2 Genomic DNA Extraction

gDNA was extracted from blood donor samples using the QIAamp DNA Blood Mini kit (Qiagen Ltd, United Kingdom) following manufacturer's

guidelines. All samples were centrifuged on 2500 ×g for 10 minutes to separate blood to three layers; plasma, buffy coat, and packed RBCs. In a 1.5 ml tube, 200 μL of the extracted buffy coat was added to 20 μL of Qiagen Proteinase K. A volume of 200 μL of AL buffer was added, mixed vigorously and incubated at 56 °C for 10 minutes and then centrifuged. 200 μL of absolute ethanol (Fisher Scientific UK, United Kingdom) was added to the mixture and mixed vigorously for 15 seconds. The mixture was transferred to the QIAamp mini spin column, and centrifuged at full speed xg for 1 minute. The filtrate was discarded and the spin column was placed in a new collection tube. Next, 500 μL of AW2 buffer was added to the QIAamp mini spin column tube and was spun at 6000 xg for 3 minutes; the filtrate was discarded afterwards and the QIAamp mini spin column was placed in a clean 2 mL tube. Lastly, gDNA was eluted with 200 μL with nuclease-free water following a 1 minute incubation to increase the yield of gDNA. The tube was centrifuged at 6000 xg for 1 minute. Purified gDNA concentration was determined using Qubit assay (section 2.3). gDNA was finally stored at -20 °C.

2.3 Qubit® dsDNA HS Assay

After gDNA extraction, gDNA concentration was determined using Qubit® dsDNA HS Assay kit (Thermo Fisher Scientific, United Kingdom). The kit included Qubit® dsDNA HS reagent buffer, and two standards, one and two. Qubit™ reagent was diluted with Qubit™ buffer in 1:200 ratio to create the working solution. In a clean Qubit™ assay tube, 190 μL of the working solution was added for each of the standards, and 199 μL for each of the samples. A volume of 10 μL for each standards and 1 μL for each sample were added to reach a total volume of 200 μL. All tubes were mixed vigorously, and incubated for 3 minutes at room temperature. Qubit™ 2.0 Fluorometer (Invitrogen, United Kingdom) instrument was used for reading the standards 1 and 2 (ng/μL), followed by the samples. Using readings acquired from the Qubit™ 2.0 fluorometer, gDNA concentration was calculated. All samples were then diluted to 20 ng/μL using nuclease-free water (Ambion®, Thermo Fisher Scientific, United States) and stored at -20 °C for future use.

2.4 Primer Design

All primers in this study were designed using Primer3 software (Untergasser et al., 2012) and the CLC Main Workbench 9 software (Qiagen Ltd, United Kingdom) to visualize an alignment of the *RHD* and *RHCE* genes to aid in locating primers positions. To ensure primer specificity to the target of interest, primers were assessed using Primer- BLAST on the National Centre for Biotechnology Information (NCBI) website (Altschul et al., 1997).

To investigate 8 *RHD* intronic SNPs detected by Halawani (2015), that were suspected to be R_2 haplotype specific including 2 SNPs in intron 2 (25,611,580 G>A; 25,614,400 C>G), 3 SNPs in intron 3 (25,621,980 C>T; 25,625,471 T>C; 25,627,066 C>G), and 2 SNPs located in intron 8 (25,646,933 T>G; 25,648,349 T>C), *RHD* specific primers were designed to amplify regions of interest for sequencing. SNPs numbers are the position they correlate to in hg19 reference sequence. Six sets of primers were designed, and one pair of primers for SNP (25,646,933) was adapted from Clarke (2016), listed in Table 2.2.

LR-PCR primers for *RHD*, *RHCE*, and *RHAG* are listed in Tables 2.3, 2.4 and 2.5, respectively. Multiple overlapping amplicons were designed to amplify the target gene for sequencing (Figure 2.2). The *RHD* gene was amplified in six PCR amplicons ranging from 9 to 13 kb in size, with approximately 1 kb overlap between each of them. To eliminate amplification from the *RHCE* gene, primers were designed around intronic differences between the *RHD* and the *RHCE* genes positioned at the 3' end to create *RHD* specific primers. Even though exons 8 and 10 for the *RHD* and the *RHCE* are identical, there are intronic differences between the two genes that have been utilised to create *RHD* specific primers.

In a similar manner, the *RHCE* gene was amplified in eight overlapping amplicons (Figure 2.2) listed in Table 2.4, in which RHCE-2 reverse primer was designed by Kelly Sillence and RHCE-3 and RHCE-4 primer sets were adapted from Halawani (2015).

For the *RHAG* gene amplification, three sets of primers were designed to amplify the gene in three overlapping amplicons (Table 2.5).

Another set of LR-PCR *RHD* primers were adopted from Hyland et al.

(2017) (forward primer: 5′-GCTTGGGCTTCCTCACCTCG-3′ ; reverse primer: 5′-TGCCGGCTCCGACGGTATC-3′). Primers were used to amplify one LR-PCR amplicon that covers the region from exon 2 to exon 7 in the *RHD* gene for samples that failed to amplify one or more of the LR-PCR *RHD* amplicons designed in this study. All primers were ordered in a high performance liquid chromatography (HPLC) purified form from Eurofins Genomics (Ebersberg, Germany).

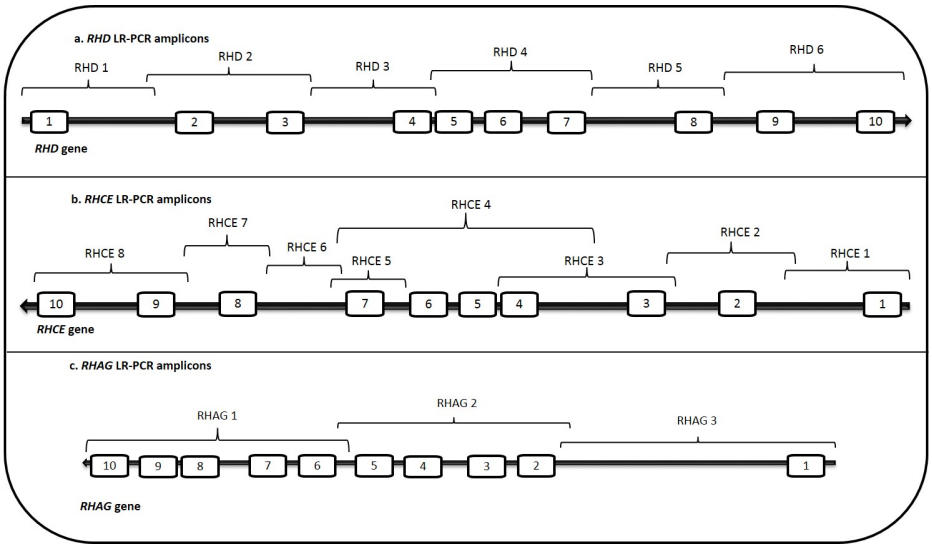


Figure 2.2: The *RHD*, *RHCE*, and *RHAG* genes amplified in LR-PCR amplicons. (a) Six overlapping *RHD* LR-PCR amplicons. (b) The *RHCE* gene amplified in 8 overlapping amplicons. (c) Three overlapping *RHAG* LR-PCR amplicons.

2.5 Polymerase Chain Reaction (PCR)

2.5.1 Conventional PCR

Seven *RHD* intronic SNPs, 2 in intron 2 (25,614,400 C>G, 25,611,580 G>A); 3 in intron 3 (25,621,980 C>T, 25,627,066 C>G, 25,625,471 T>C); and 2 in intron 8 (25,648,349 T>C, 25,646,933 T>G) were suspected to be linked to the R_2 haplotype (Halawani, 2015). All seven SNPs were tested to confirm if they are R_2 haplotype specific. gDNA samples from blood donors of different phenotypes were tested (6 $\mathrm{R}_2\mathrm{R}_2$, 6 $\mathrm{R}_1\mathrm{R}_1$, 1 $\mathrm{R}_2\mathrm{r}$, 1 $\mathrm{R}_1\mathrm{r}$, 6 $\mathrm{R}_1\mathrm{R}_2$, and 8 $\mathrm{R}_0\mathrm{r}$). *RHD* -specific primers were used to amplify the region of interest listed in Table 2.2.

Two different enzymes were used, BioMix™ 2X master mix (Bioline Reagents Limited, United Kingdom) for (25,621,980), (25,625,471), (25,648,349), (25,646,933) (25,648,349), and (25,646,933); and Q5® Hot Start High-Fidelity 2X Master Mix, (New England Biolabs®, United Kingdom) for SNP (25,627,066). Fifty μL PCR reaction contained a 1X master mix, 200ng of gDNA template, 400 nM of the forward primer, and 400 nM of the reverse primer. Veriti thermocycler program was set as following: denaturation at 95 °C

for 10 min, 35 cycles of 95 °C for 30 seconds, annealing temperature (T_a) varied for each primer set (Table 2.2), and 72 °C for 30 seconds. The last extension was at 72 °C for 10 min; finally, samples were held at 4 °C. Agarose Gel Electrophoresis was used to validate the success of PCR amplification, (section 2.6).

Table 2.2: Sequence, annealing temperature (T_a), and product sizes for all primers used for *RHD* intronic SNPs amplification.

Primer name	**Sequence 5′-3′**	**Size (bp)**	T_a, **°C**
25,611,580-Forward	TTTTACTGGACAGCCCTACTCC	558	62
25,611,580-Reverse	CATGGCTATTTATTGTCTAGCAGCA		
25,614,400-Forward	GCTACCATGCCCTGCTAAT	417	63
25,614,400-Reverse	TCCAGTACTTTTCAGAGCC		
25,621,980-Forward	GACCGAACACTTGTTCAATTTTGAAC	442	60
25,621,980-Reverse	CTCTCCCTTTCTTGCTGTGG		
25,625,471-Forward	GGGGCAGCTTCATCTTATCAAGAG	419	62
25,625,471-Reverse	CTCACTGCAACCTCCACCCGTT		
25,627,066 Forward	TGGGATTACAGGCAAAATTAG	834	60
25,627,066-Reverse	AGGTGTGACTTGAAGCCAT		
25,646,933-Forward*	CTTGGGCTGCTGAGAGATTAGAA	507	57
25,646,933-Reverse*	CATATGTATTCCTCCAGCTTCCT		
25,648,349-Forward	TCCAGGAATGACAGGGCT	525	63
25,648,349-Reverse	TGAGGACTGCAGATAGGG		

* Adapted from Clarke (2016).

2.5.2 Long Range-PCR (LR-PCR)

To optimize PCR conditions for LR-PCR, different T_a and primer concentrations were tested to ensure specific amplification from the target gene.

In a 50 μL reaction, 1X master mix of LongAmp Hot Start *Taq* 2X Master Mix (New England Biolabs, United Kingdom) was used with 200 ng of gDNA template. One μM of the forward and reverse primers was used for all amplicons except for *RHD* amplicon 3, where 0.2 μM of the forward and reverse primers was used. Each *RHD* amplification cycle run included *RHD* positive hemizygous control, *RHD* positive homozygous control, *RHD* negative control, and non-template control.

The Veriti Thermal Cycler (Applied Biosystems, United Kingdom) program was set as follows: denaturation at 95 °C for 5 minutes, 30 cycles of 95 °C for 30 seconds, annealing for 30 seconds, and extension at 65 °C 1 minute / kb. T_a varied for each primer set (Tables 2.3, 2.4, 2.5). The last extension was at 65 °C for 10 minutes; finally, samples were held at 4 °C. To validate PCR amplification, PCR products were tested on agarose gel electrophoresis (section 2.6).

Hyland et al. (2017) primers were used to amplify the *RHD* gene from exon 2 to exon 7 in a 22,117 bp LR-PCR amplicon. A final volume of 50 μL reaction contained: 1X master mix of 5X TaKaRa LA Taq Hot Start (Takara, Japan), 100 ng of gDNA template, 1 μM of the forward and re-

verse primers. The reaction was placed on 94 °C for initial denaturation for 30 seconds followed by two-step PCR protocol (30 cycles at 98 °C for 10 seconds followed by 68 °C for 10 minutes) (Hyland et al., 2017).

Table 2.3: Sequence, exons covered, product sizes, and annealing temperature (T_a) for primers used for *RHD* LR-PCR.

Primer name	**Sequence 5′-3′**	**Exons**	**Size, bp**	T_a, °C
RHD-1 Forward	ATCCACTTTCCACCTCCCTGC	1	10,326	62
RHD-1 Reverse	TCTTTGCACTTCTTCTGACAACA			
RHD-2 Forward	CTGGGAGAGTGAAGCTGGGTGTGA	2, 3	13,709	62
RHD-2 Reverse	TTCATACACATCTCTACCCCCCCTC			
RHD-3 Forward	GTTTGAGCCCAGGAGTTAGGGACCGAG	4	10,789	66
RHD-3 Reverse	CCCACTGTGACCACCCAGCATTCTA			
RHD-4 Forward	CATACCTTTGAATTAAGCACTTCAC	5, 6, 7	9,895	66
RHD-4 Reverse	CAGAATGGCCTTTACCAGCCAT			
RHD-5 Forward	GTTCAAGCTGTCAAGGAGACATCACTATACA	8	11,628	65
RHD-5 Reverse	CCAGTTTTAAGAATTTGTCGGCCGGTCG			
RHD-6 Forward	ATACATTCCATCCAGAACTGTTCACC	9, 10	11,284	64
RHD-6 Reverse	AGGCCAAGAGATCCTGGTGAAACTATCC			

Table 2.4: Sequence, exons covered, product sizes, and annealing temperature (T_a) for primers used for *RHCE* LR-PCR.

Primer name	Sequence 5′-3′	Exons (Intron)	Size, bp	T_a, °C
RHCE-1 Forward	TCTTTGCACTTCTTCTGACAGCA	1	10,599	62
RHCE-1 Reverse	GCTAGCGGACATTTAGGTTAGTCTC			
RHCE-2 Forward	AGGTCTTGCTATATTGCCCAGGTTCA	2	10,419	64
RHCE-2 Reverse*	GCTGCTGGAGTATAGTGATGTGATCGTG			
RHCE-3 Forward†	CTACTATCAAGCTCACTGCCCGATTT	3, 4	11,215	62
RHCE-3 Reverse†	ATCCTGGCTCTCCTTCTCA			
RHCE-4 Forward†	CAAGTCCATGTGCAGTGAC	4, 5, 6, 7	17,856	60
RHCE-4 Reverse†	ACAGCCAGCATCTTCTTTCAGTCAG			
RHCE-5 Forward	TTACAGGTGTGCGCCAGAGT	7	3,881	62
RHCE-5 Reverse	GACACGAAATCTTCTGTGACCC			
RHCE-6 Forward	GGCTGAAGTATGAGAATTGCTTGAAC	(7)	6,032	62
RHCE-6 Reverse	GTCACTGCACATGGACTTG			
RHCE-7 Forward	GTAGAGATGGGGTCTCACTATAGGA	8	5,333	64
RHCE-7 Reverse	ATATGCCTAGAAGTGAAATTGTTGAG			
RHCE-8 Forward	GTACTGTTTCCTGTCCCCGA	9, 10	13,966	60
RHCE-8 Reverse	TGTGGTGACATTGGGCATG			

*Designed by Kelly Sillence.

† Adapted from Halawani (2015).

Table 2.5: Sequence, product sizes, and annealing temperature (T_a) for primers used for *RHAG* LR-PCR.

Primer name	Sequence 5′-3′	Exons	Size, bp	T_a, °C
RHAG-1 Forward	TGGTAGGGCTGATTTCCTTGT	6, 7, 8, 9, 10	10,003	62
RHAG-1 Reverse	TGGATGTTTTGGCCCAGCTT			
RHAG-2 Forward	GCTGATCTGAGGGTTACTCCTTT	2, 3, 4, 5	10,519	62
RHAG-2 Reverse	AGGAGGATGGGAACGCTAAG			
RHAG-3 Forward	AATTATTCTGCAGATTTCACCCC	1	15,083	62
RHAG-3 Reverse	GGAGACAAGAATTCCTCCACCTAT			

2.5.3 Droplet Digital PCR (ddPCR)

All samples were tested for zygosity with the aim of knowing the number of *RHD* alleles present for subsequent sequence analysis. The *RHD* zygosity was determined for all samples using droplet digital polymerase chain reaction (ddPCR) to determine if a sample was hemizygous (one copy of the *RHD* gene) or homozygous (two copies of the *RHD* gene) (Sillence et al., 2015, 2017). Samples were tested for *RHD* exon 5 (*RHD5*) and *RHD* exon 7 (*RHD7*) against the reference gene *Argonaute RISC component 1 (AGO1)* on chromosome 1 (Sillence et al., 2015, 2017). Primers and probes used in ddPCR zygosity testing were adapted from Finning et al. (2008); Fan et al. (2009), (Table 2.6). Probes used for the two *RHD* gene targets, *RHD5* and *RHD7*, were labelled with 6-carboxyfluorescein (FAM) on the 5' end and Black Hole Quencher 1 (BHQ-1) on the 3' end, while the reference sequence *AGO1* probe was labelled with Hexachlorofluorescein (HEX) on the 5' end and BHQ-1 on the 3' end.

Each PCR reaction was prepared in 20 μL final volume, which contained: 1X ddPCR™ Supermix for probes (Bio-Rad, United Kingdom), 300 nmol \L of primers, 250 nmol \L probes, and 20 ng of template. To generate

droplets, the 20 μL PCR reaction was transferred to the centre row labelled with "sample" of a DG8™ cartridge (Bio-Rad). To the row labelled with "oil", 70 μL of QX100™ droplet generator oil for probes was placed (Bio-Rad). The cartridge was sealed using DG8™ gaskets (Bio-Rad) and then placed onto the QX100™ droplet generator (Bio-Rad) to generate droplets through oil emersion chemistry.

After the generation of the droplets, 40 μL solution containing droplets was transferred to a 96-well PCR Plate (Sigma-Aldrich, UK). The plate was then sealed using a pierceable foil heat seal (Bio-Rad) on the PX1™ PCR plate sealer (Bio-Rad). The plate was then placed in the C1000 Touch™ Thermal Cycler (Bio-Rad). PCR protocol was as follows; 95 °C for 10 mins, 40 cycles of 95 °C for 30 seconds and 58 °C for 1 minute. Final step was carried out at 98 °C for 10 minutes. Samples were finally held at 12 °C.

Following manufacturer's guidelines, a new experiment was set up using the QuantaSoft™ Software v1.7 (Bio-Rad). After the thermal cycling had finished, the plate was placed into the QX100™ droplet reader (Bio-Rad). The droplet reader in combination with the QuantaSoft™ Software v1.7 analyses the droplets' signals and differentiates between negative and pos-

itive ones creating absolute concentration of desired DNA. The number of *RHD* copies/μL present in a sample was compared to the reference gene *AGO1* copies/μL. A sample presenting *RHD5* or *RHD7*/ *AGO1* ratios ≈ 1 was considered to be homozygous *RHD*-positive, and a sample with ratios ≈ 0.5 were classified as hemizygous *RHD*-positive, and samples showing a ratio ≈ 0 were classified as *RHD*-negative.

Table 2.6: Sequences for primers and probes used in the *RHD* zygosity testing using dPCR.

Gene	Sequence 5′-3′	Size, bp
*AGO1**	Forward: GTTCGGCTTTCACCAGTCT	82
	Reverse: CTCCATAGCTCTCCCCACTC	
	Probe: HEX-CTGCCATGTGGAAGATGATG-BHQ1	
*RHD5***	Forward: CGCCCTCTTCTTGTGGATG	75
	Reverse: GAACACGGCATTCTTCCTTTC	
	Probe: FAM-TCTGGCCAAGTTTCAACTCTGCTCTGCT-BHQ1	
*RHD7***	Forward: CAGCTCCATCATGGGCTACAA	81
	Reverse: AGCACCAGCAGCACAATGTAGA	
	Probe: FAM-AGCTTGCTGGGTCTGCTTGGAGAGATC-BHQ1	

*Adapted from Fan et al. (2009). **Adapted from Finning et al. (2008).

2.6 Agarose Gel Electrophoresis

To validate PCR amplification, PCR products were run on 1% (w/v) agarose gel for *RHD* introns PCR products and on 0.7% (w/v) agarose gel for

LR-PCR products. High resolution standards agarose (AGTC Bioproducts Ltd, Yorkshire, UK) was mixed with 1X Tris-acetate-EDTA (TAE) electrophoresis buffer [40 mM Tris-acetate and 1 mM EDTA pH 8.0] (Sigma-Aldrich Company Ltd, United Kingdom), and with the addition of fluorescent nucleic acid dye (GelRed) 1:10000 dilution (Biotium Inc, United States).

A volume of 5 μL of each sample was mixed with 1 μL of loading buffer [10 mM Tris-hydrochloric acid (pH 7.6), 0.03% (v/v) bromophenol blue, 0.03% (v/v) xylene cyanol FF, 60% (v/v) glycerol, and 60 mM EDTA] and then to a well. Quick-Load® 1 kb Extend DNA Ladder (New England Biolabs®, United Kingdom) was used to assess the amplicon size for LR-PCR products, and 100 bp DNA ladder (NORGEN, Thorold, Canada) was used for *RHD* introns PCR products. The samples were subjected to electrophoresis for 1 hour and 1 hour 30 minutes for *RHD* introns and LR-PCR products, respectively; on 110V constant voltage. The gel was then removed to an EC3 imaging system (Ultra Violet Products Ltd, United Kingdom), and was visualized using the Launch Vision WorksLS.

2.7 QIAquick Gel Extraction

QIAquick Gel Extraction kit (Qiagen Ltd, United Kingdom) was used to purify *RHD* introns amplicons. Fifty μL of PCR reaction was mixed with 10 μL of the loading agent [10 mM Tris-hydrochloric acid (pH 7.6), 0.03% (v/v) bromophenol blue, 0.03% (v/v) xylene cyanol FF, 60% (v/v) glycerol, and 60 mM EDTA] and then subjected to electrophoresis (section 2.6). After the electrophoresis was finished and all bands were clearly separated, the size specific band was excised using a sharp scalpel and placed in a clean 2 ml tube. The weight of the band was then measured on a balance to determine the required of the QIAquick buffer QG. For each volume of the gel (100 mg), 3 volumes of QIAquick buffer QG were added (300 μL). The mixture was then incubated at 50 °C for 10 minutes or until the gel had dissolved completely. After the gel slice had dissolved, one volume of isopropanol (Fisher Scientific UK, United Kingdom) was added to the sample and mixed vigorously. The mixture was then filtered using the provided QIAquick spin column which was centrifuged at 6000 rpm for 1 minute and filtrate was discarded; this step was repeated as needed to filter the complete volume of mixture. The filter spin column (contain-

ing DNA) was placed back to the collection tube for the washing steps.

Because PCR products were for direct sequencing, a wash with 500 μL of QIAquick buffer QG was recommended, the spin column was centrifuged and filtrate was discarded. A volume of 750 μL of the QIAquick buffer PE was added to the column and incubated at room temperature for 5 minutes. The column was then centrifuged and filtrate was discarded, additional centrifugation was recommended to thoroughly remove the ethanol. The QIAquick spin column was placed in a clean 1.5 ml collection tube for elution. The gDNA was eluted with 50 μL of QIAquick buffer EB and then incubated at room temperature for 1 minute to increase the yield of DNA. The samples were then subjected to agarose gel electrophoresis to validate the success of the extraction and purification (section 2.6). One hundred bp DNA ladder (NORGEN, Thorold, Canada) is provided with an estimated concentration for each band which enabled the gDNA concentration for the PCR amplicons to be estimated by comparing the band intensity to the known concentration of DNA in the ladder bands.

2.8 Sanger Sequencing

For *RHD* intronic SNPs investigation, purified amplicons were sent for Sanger sequencing to Eurofins MWG Operon (Ebersberg, Germany) by using Mix2Seq kit. Each sample was distributed into two barcoded tubes, 15 μL in each. A volume of 2 μL of each of the primers was added to a single tube in 10 pmol/μL concentration. Sequencing data was received in (ab1) format, which was then mapped and analysed using the CodonCode Aligner 6 (Aligner, United States). Data was aligned to the GRCh37 chromosome assembly RefSeq (NC_000001.10) which is the genome reference consortium human genome build 37 (known as hg19) (NCBI, 2009). Each trace chromatogram was analysed to determine intronic SNPs zygosity and links to Rh specific haplotype.

2.9 NGS Library Preparation

2.9.1 Agencourt® AMPure® XP

Agencourt® AMPure® XP reagent (Beckman Coulter, United Kingdom) was used to purify LR-PCR products. For each volume of the sample, 1.8 X volume of the purifying agent was added. The mixture was mixed

thoroughly and then incubated at room temperature for 5 minutes. The samples were then placed on the magnetic field (Alpaqua Engineering LLC, United States) for 3 minutes to isolate smaller amplicons, and primer-dimers. After 3 minutes, the supernatant was discarded and the remaining bead pellets were washed two times with 70% ethanol (Fisher Scientific UK, United Kingdom). After the second wash, the bead pellet was left to air dry for 4 minutes. Finally, the samples were eluted with 20 µL of nuclease-free water (Ambion®, Thermo Fisher Scientific, United States). Samples were then quantified using Qubit® dsDNA HS Assay kit, (section 2.3).

2.9.2 Pooling and Fragmentation

An equimolar pool for purified amplicons was created to ensure an equal depth of coverage for the targeted gene. The pooling was prepared with a final concentration of 100 ng in a total volume of 35 µL. One hundred ng was divided by the number of amplicons for each sample, which meant, for example, 16.67 ng of each of the *RHD* amplicons was required to achieve the 100 ng concentration. Final volume was brought to 35 µL by the addition of nuclease-free water (Ambion®, Thermo Fisher Scientific, United States).

After pooling of the different amplicons, samples were fragmented to create a 200 bp read library using the Ion Xpress™ Plus Fragment Library kit (Life Technologies). In an Eppendorf LoBind™ tube, each sample was mixed with 5 µL of Ion Shear™ Plus 10X Reaction Buffer and nuclease-free water (Ambion®, Thermo Fisher Scientific, United States) to bring the final volume to 40 µL, tubes were mixed for 5 seconds, vigorously. Ten µL of the Ion Shear™ Plus Enzyme Mix II was added to each sample to bring the total volume to 50 µL. The mixture was mixed by pipetting up and down 8-10 times. The reaction tube was then incubated on a heat block at 37 °C for 15 minutes. Immediately after incubation, 5 µL of Ion Shear™ Stop Buffer was added to stop the reaction. The fragmented amplicons were purified by Agencourt® AMPure® XP reagent using 1.2X volume of the samples (refer to section 2.9.1). However, in final elution the samples were treated differently. Samples were eluted in 25 µL instead of 50 µL of Low TE (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA). To check the fragment size, 1 µL of the eluted DNA was tested using the Agilent™ 2100 Bioanalyzer™ instrument and Agilent™ High Sensitivity DNA kit, (section 2.9.3).

2.9.3 Agilent® 2100 Bioanalyzer System

The 2100 Bioanalyzer™ instrument with an Agilent™ High Sensitivity DNA kit (Agilent™ Technologies UK Limited, United Kingdom) were used to determine the size distribution and the concentration of the samples. To prepare the gel-dye mix, 15 μL of the dye concentrate was added to the light sensitive DNA gel matrix vial and then was mixed for 10 seconds. The mixture was then placed in a spin filter, and centrifuged at 6000 rpm for 10 minutes. A High Sensitivity DNA chip was placed on the chip priming station, and the plunger was placed at 1 ml. A volume of 9 μL of the gel-dye mix was pipetted into the third well marked with G, the priming station was closed and the plunger was then pressed down until it was held by the clip; a timer was then started for 60 seconds. The plunger was released after 60 seconds and the priming station was opened once the plunger reached 1 mL. A volume of 9 μL of the gel-dye mix was placed into the remaining three wells marked with G. Five μL of the High Sensitivity DNA marker was dispensed into all sample wells including the ladder marked well. One μL of the High Sensitivity DNA ladder was placed in the ladder position. In each sample well, 1 μL of each sample was loaded. The chip was placed in the IKA vortex mixer (IKA®, United

Kingdom) for 60 seconds at 2400 rpm. The chip was then placed in the Agilent 2100 Bioanalyzer, and a run was started following the manufacturer's guidelines.

2.9.4 Adaptor Ligation and Nick Repair

After amplicons fragmentation, fragments were ligated to adaptors and nick-repaired using the Ion Xpress™ Barcode Adapters kit (Life Technologies) for barcoding materials and Ion Plus Fragment Library kit for nick repair. In 0.2 mL PCR tubes, 25 µL of the fragmented amplicons were mixed with 10 µL of 10X Ligase Buffer, 2 µL of Ion P1 adapter, 2 µL of Ion Xpress™ Barcode X , 2 µL of deoxynucleotide triphosphates (dNTPs) Mix, 49 µL of nuclease-free water (Ambion®, Thermo Fisher Scientific, United States), 2 µL of DNA ligase, and 8 µL of Nick Repair Polymerase. Veriti thermocycler was programmed as follows: 25 °C for 15 minutes, then 72 °C for 5 minutes and finally the samples were held at 4 °C for up to 1 hour. The product was then purified using the 1.2X volume of the Agencourt® AMPure® XP reagent (section 2.9.1). The samples were eluted with 20 µL of Low TE (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA). To validate the success of this step, samples were tested on Agilent® 2100 Bioanalyzer System, (section 2.9.3).

2.9.5 Size Selection

After ligation and nick repair, samples were subjected to size selection using Agencourt® AMPure® XP. Right and left sides size selection was performed using different ratio of the SPRIselect (Figure 2.3). The ratio of SPRIselect reagent used dictate the size of the DNA fragments binding. For the left side size selection, 0.8X of the Agencourt® AMPure® XP was mixed with each sample and incubated at room temperature for 1 minute then placed on the magnetic plate for 3-4 minutes to separate. Magnetic beads in this stage bind to small fragments because increasing the ratio of SPRIselect to sample increases the efficiency of binding smaller fragments. The supernatant was removed carefully without disrupting the beads. Two 85% ethanol washes were performed with discarding the supernatant after each wash. After the last wash, the beads were left to dry at room temperature for 5 minutes and then eluted with 50 μL of nuclease-free water (Ambion®, Thermo Fisher Scientific, United States).

For the right side size selection, 0.7X of the SPRIselect was used as decreasing the SPRIselect volume will increase the efficiency of binding larger fragments. After mixing the samples with the SPRIselect, they were in-

cubated at room temperature for 1 minute then placed on the magnetic plate for 3-4 minutes to separate. The supernatant was then collected and placed in a new well, leaving the beads with large fragments behind. At this stage, the supernatant contains the right side size selected sample. To the supernatant, 1.1X of the SPRIselect was mixed with each sample, incubated for 1 minute and then placed on the magnetic rack to separate. The supernatant was discarded and two 85% ethanol washes were performed. After the final wash, the beads were left to air dry at room temperature for 5 minutes. Finally, samples were eluted with 20 μL of nuclease-free water (Ambion®, Thermo Fisher Scientific, United States). The final product was then analysed using Agilent™ 2100 Bioanalyzer™ instrument and Agilent™ High Sensitivity DNA kit, (section 2.9.3). Data obtained from 2100 Bioanalyzer™ included size distribution and molar concentration which were used to calculate the template dilution factor.

2.10 Template Amplification and Enrichment

2.10.1 Pooling of the Libraries and Amplification

To guarantee an equivalent representation of each of the libraries in the sequencing run, sequencing libraries were diluted to a concentration of 26

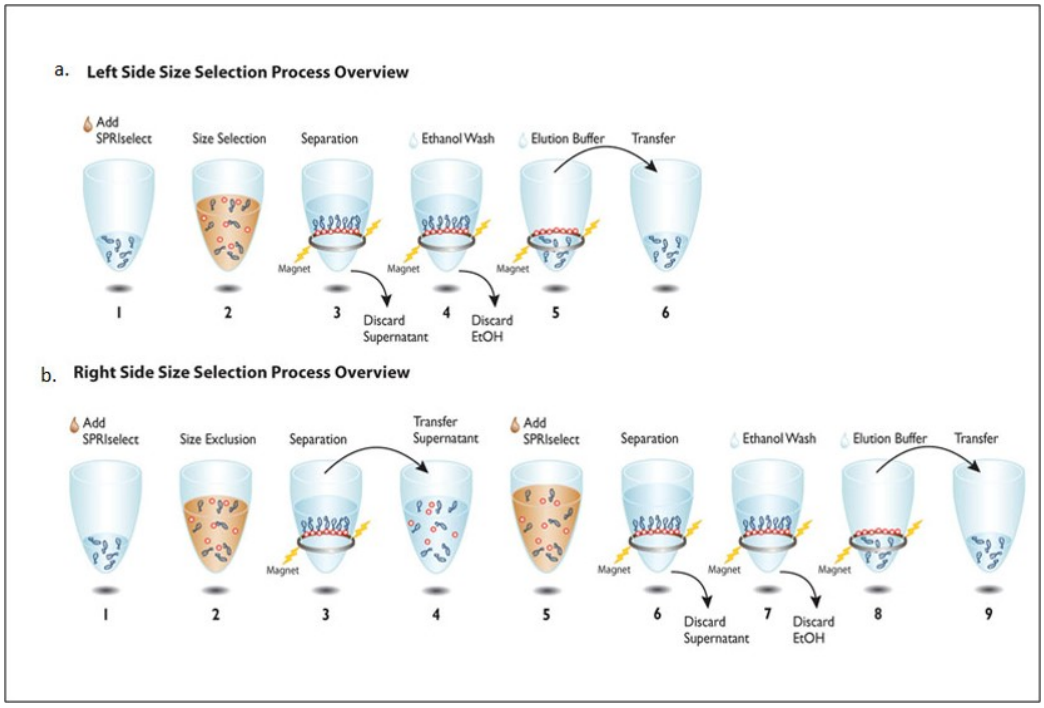


Figure 2.3: **Left and right sides size selection using the Agencourt® AMPure® XP.** (a) Left side size selection was performed using 0.8X Agencourt® AMPure® XP to eliminate smaller than 200 bp fragments; and (b) 0.7X Agencourt® AMPure® XP was used for right side size selection to eliminate fragments larger than 200 bp. Image courtesy of Coulter (2012).

pM. Dilution factor was determined by dividing the library concentration in pM by 26. For instance, if the library concentration was 130 pM, the dilution factor would be 5. Thus, 1 μL of the sequencing library would be diluted in 4 μL of nuclease-free water. If a sample had less than 26 pM concentration, a higher volume of the library was added to reach the required amount. For example, if a sample has 10 pM, then 2.6 μL was added to

reach the amount of DNA required.

For library enrichment, the Ion OneTouch™ was used with the Ion PGM™ Template OT2 200 kit (Thermo Fisher Scientific). Dr. Michele Kiernan conducted this step and subsequent sequencing protocols. The Ion PGM™ OneTouch Plus reaction filter was assembled according to the manufacturer's guidelines. To prepare the amplification solution, 2 μL of the pooled sequencing library was diluted in 23 μL of nuclease-free water. To the diluted library, subsequent solutions were added as follows: (25 μL of nuclease free water, 500 μL of Ion PGM™ Template OT2 200 Reagent Mix, 300 μL of the Ion PGM™ Template OT2 200 PCR Reagent B, 50 μL of the Ion PGM™ Template OT2 200 Enzyme Mix) to bring the final volume to 900 μL. The mixture was then mixed and 100 μL of the Ion PGM™ Template OT2 200 Ion Sphere™ Particles were added and mixed thoroughly. Following the protocol instructions, 1000 μL of the amplification solution was placed into the sample port in the reaction filter, followed by the addition of 1.5 mL of the Ion OneTouch™ Reaction Oil. The reaction filter was then placed in the Ion OneTouch™ 2 instrument and the run was started.

2.10.2 Positive ISPs Recovery and Enrichment

After the completion of the Ion OneTouch™ 2 run, two Ion OneTouch™ recovery tubes were placed in the centrifuge in the Ion OneTouch™ 2 instrument, and were centrifuged for 10 minutes. The Ion OneTouch™ recovery router was discarded at the end of the centrifugation and the tubes were placed in a rack. From each tube, all but 50 μL of the recovery solution was removed carefully without disturbing the ion sphere particles (ISPs) pellet, which was re-suspended in the remaining solution. The ISPs suspension from both tubes was combined in a 1.5 mL tube and 1 mL of the Ion OneTouch™ wash solution was added. In an 8-well strip from the Ion OneTouch™ ES supplies kit, 100 μL of the ISPs-suspension was placed in well 1.

On the Ion OneTouch™ ES, template enrichment was performed to increase template numbers for sequencing. Melt-off reagent was prepared by mixing 280 μL of the Tween™ solution with 40 μL of 1 M sodium hydroxide (NaOH), which was then placed into well 7 in the 8-well strip. 13 μL of the Dynabeads™ MyOne™ Streptavidin C1 Beads was placed in a 1.5 mL tube, which was then placed on DynaMag™-2 magnet for 2

minutes to separate the beads. After two minutes, the supernatant was removed with caution to not disturb the beads pellet. To the beads, 130 μL of MyOne™ beads wash solution was added and then mixed by vortex for 30 seconds. In the 8-well strip, where 100 μL of the ISPs-suspension was placed in well 1, 130 μL of the re-suspended MyOne™ beads in the wash solution was placed in well 2. Wells 3, 4, and 5, were filled with 300 μL of the Ion OneTouch™ wash solution, while wells 6 and 8 were left empty. Following user guidelines, the Ion OneTouch™ ES was prepared for template enrichment. The run was started following user instruction.

2.11 Ion PGM™ Sequencing

After completion of the enrichment stage, the PCR tube containing the enriched ISPs was removed and quality assessment was performed using the Qubit® 2.0 Fluorometer, (section 2.3). Finally, a sequencing run was carried on the Ion PGM™ on Ion 316™ Chip v2 by Dr. Michele Kiernan in the Genomics Facility of the Systems Biology Centre, University of Plymouth. A run is started with creating a planned run in the Torrent Browser connected to the Ion PGM™ system. The Ion PGM™ is cleaned and initiated following manufacturer guidelines.

The enriched ISPs were prepared for sequencing using the Ion PGM™ Sequencing 200 Kit v2. In a PCR tube, 5 µL of Control Ion Sphere™ Particles was added directly to the whole amount of the enriched ISPs and mixed thoroughly by pipetting, followed by centrifugation for 2 minutes at 15,500 × g. Supernatant was carefully removed to not disturb the pellet leaving about 15 µL of liquid in the tube. To the pellet, 12 µL of the Sequencing Primer was added and was mixed by pipetting. The tube was placed in Veriti thermal cycler programmed with 95°C for 2 minutes; 37°C for 2 minutes to anneal sequencing primers.

After thermal cycler, 3 µL of Ion PGM™ Sequencing 200 v2 Polymerase was added, mixed by pipetting, and incubated for 5 minutes at ambient temperature while sequencing Ion 316™ Chip v2 was inspected for any physical damages before loading library. Sequencing chip was prepared and library was loaded following carefully the user guide. The chip was loaded into the Ion PGM™ and a sequencing run was initiated on the Ion PGM™ system. Sequencing programme runs over night (about 8 hours) and sequencing data is extracted in FASTQ files, see section 2.14.1 for data analysis.

Samples genotyped for *RHD*, *RHCE*, and *RHAG* using Ion PGM™ are shown in Figure 2.4. Samples genotyped for the *RHD* gene included 75 blood donor samples that were either hemizygous for the *RHD* gene or phenotyped as weak D by serology. Other homozygous samples were also sequenced for comparison. Thirty-five gDNA samples from D negative pregnant females with inconclusive results were also included in *RHD* genotyping to identify genetic background of *RHD* variants. A further 23 gDNA samples from the ISBT 1996 workshop (ISBT, 1996), previously genotyped as variant *RHD* allele, were included to study the full sequence of different variant alleles and intronic SNPs variations might be associated with these variants.

2.12 MinION™ Library Preparation

The *RHD* gene from gDNA samples (n=12) from Finnish pregnant females with variant *RHD* gene, supplied by the Finnish Red Cross Blood Service (Helsinki, Finland) were sequenced using the MinION™ (Oxford Nanopore Technologies, UK), and one blood donor sample (Figure 2.4). Library preparation was performed following the 1D Native barcoding

gDNA (Oxford Nanopore Technologies, UK) protocol using the Native Barcoding kit 1-12 (Oxford Nanopore Technologies, UK) and the Ligation Sequencing kit (SQK-LSK109) (Oxford Nanopore Technologies, UK) with 1D flow cells R9 version (Oxford Nanopore Technologies, UK).

Six *RHD* LR-PCR amplicons were pooled for each sample to yield a final amount of 1800 ng at 48 μL final volume. For end repair and dA-tailing, 3.5 μL of NEBNext FFPE DNA Repair Buffer (New England Biolabs®, United Kingdom), 2 μL of NEBNext FFPE DNA Repair Mix (New England Biolabs®, United Kingdom), 3.5 μL of Ultra II End-Prep buffer (New England Biolabs®, United Kingdom), and 3 μL of Ultra II End-Prep enzyme mix (New England Biolabs®, United Kingdom) was added and mixed by pipetting. Using a Veriti Thermal Cycler, the reaction was incubated at 20 °C for 5 min and then at 65 °C for 5 minutes, and was then purified using AMPure XP beads (section 2.9.1) and eluted in 25 μL nuclease free water.

Recovered samples were then quantified using Qubit® dsDNA HS Assay Kit (section 2.3). Samples were then diluted to 500 ng in a 22.5 μL final volume in preparation for barcode ligation. To each sample, 2.5 μL of unique barcode and 25 μL of Blunt/TA Ligase Master Mix (New England Bio-

labs®, United Kingdom) was added and mixed by flicking the tube which was then incubated for 10 minutes at ambient temperature. The reaction was then purified using AMPure XP beads (section 2.9.1) and eluted in 26 μL nuclease free water and quantified using Qubit® dsDNA HS assay (Thermo Fisher Scientific, United Kingdom) explained in (section 2.3) before pooling to ensure equal coverage for all samples.

Samples were then pooled in equimolar amount to yield a final concentration of 700 ng in 65 μL final volume. To pooled libraries, 5 μL of the Adapter mix II, 20 μL of NEBNext Quick Ligation Buffer (New England Biolabs®, United Kingdom), and 10 μL of Quick T4 DNA Ligase (New England Biolabs®, United Kingdom) were added and mixed by flicking the tube, which was then incubated for 10 minutes at ambient temperature. Last step purification was carried out using 50 μL of AMPure XP beads. After discarding the supernatant, pellet was washed twice with 250 μL of Long Fragment Buffer. Library was then eluted in 15 μL Elution Buffer. Eluted library was then quantified using Qubit® dsDNA HS Assay Kit, (section 2.3). Prepared library was stored in ice until the sequencing preparation was complete (next section 2.13).

2.13 MinION™ Sequencing

A flow cell was placed in the MinION™ which is then plugged directly into USB3 port on laptop running Windows 10. MinKNOW software (Oxford Nanopore Technologies, United Kingdom) connects to the MinION™ and runs control checks. To prepare priming solution, 30 µL of the flush tether was added to a full tube of flush buffer. To prime flow cells, 800 µL of the priming solution was loaded to the flow cells through the priming port. After 5 minutes, 200 µL of the priming solution was loaded to the flow cells through the priming port while the SpotON cover was open. In preparation for sequencing, 12 µL of DNA library was mixed with 37.5 µL of sequencing buffer, and 25.5 µL of loading beads. The library was loaded drop by drop to the flow cell via the SpotON port. On the MinKNOW software (MinION™ Release 18.12.6), the sequencing run was started and was left running for 12 hours. Data was then collected and analysed (section 2.14.2).

2.14 Data Analysis

2.14.1 NGS Data Analysis

Data generated from the NGS (FASTQ) was analysed using the CLC Main Workbench 10 software (Qiagen Ltd, United Kingdom). Short reads were aligned to the human reference sequence hg38 downloaded from the NCBI database (NC_000001.11) (Tatusova et al., 2016). The *RHCE* gene was masked in the *RHD* gene analysis by converting it into trimmed track i.e (blocked-out) to prevent reads from scattering by mapping to a different region or gene in the chromosome assembly, and vice versa for the *RHCE* data analysis. Variant detection was performed on a minimum coverage of 30 i.e (the numbers of overlapping reads over a single base) and variants detected were analysed on a single base basis considering different parameters including number and percentage of reads and nucleotide count (Nielsen et al., 2011). Variant tracks created were annotated with the exon numbers and compared with known variants for prediction of amino acid changes and splice site alterations. For allele determination, SNPs detected that predicted amino acid changes or splice site changes were compared to known alleles on RhesusBase (Wagner and Flegel, 2014), Erythro-

gene (Möller et al., 2016) and the Blood Group Antigen FactsBook (Reid et al., 2012). The reference SNP number was then found for each SNP detected in the database of SNPs (Sherry et al., 2001).

2.14.2 MinION™ Data Analysis

The MinION™ was run for about 12 hours which produced 49 FAST5 files each containing about 4000 reads. The data analysis workflow is illustrated in Figure 2.5. Guppy basecaller v3.2.4 (Oxford Nanopore Technologies, United Kingdom) was used to call the bases of the raw data (FAST5) files, which divided the read into pass and fail FASTQ files based on quality score determined. Only pass reads were used to carry on the analysis. Files were subjected to sequencing quality analysis using EPI2ME software (Oxford Nanopore Technologies, United Kingdom).

FASTQ files were then concatenated into one FASTQ file for barcoding using Samtools v.1.4.1. Porechop software v.0.2.1 was used for barcoding, which divided reads by barcodes 1-12. Barcodes were then trimmed using SeqKit software v0.7.1. A bash loop script, a programming language that executes command codes subsequently, (Figure 2.6) was created by Dr Vasileios Lenis to analyse the data. Nanopolish software v.0.9.0

was used to index the reference (hg38 chromosome 1 reference sequence (NC_000001.11)), index the FASTQ files, and then map the reads to the reference which generated BAM files. BAM files were then sorted using Samtools v.1.4.1 to generate BAI files. Variants were then called using Nanopolish software v.0.9.0. (Loman et al., 2015). BAM and BAI files were visualised using Integrated Genome Viewer (IGV) v.2.5.3 (Broad Institute and the Regents of the University of California, United States) and CLC Main Workbench 10 software (Qiagen Ltd, United Kingdom). Variant calling was also performed using CLC Main Workbench 10 software at 100x minimum coverage. The data was then compared to the sequencing data obtained from Ion PGM™. Exonic and intronic mutations detected from both platforms were compared and variant track of the same samples were aligned for comparison.

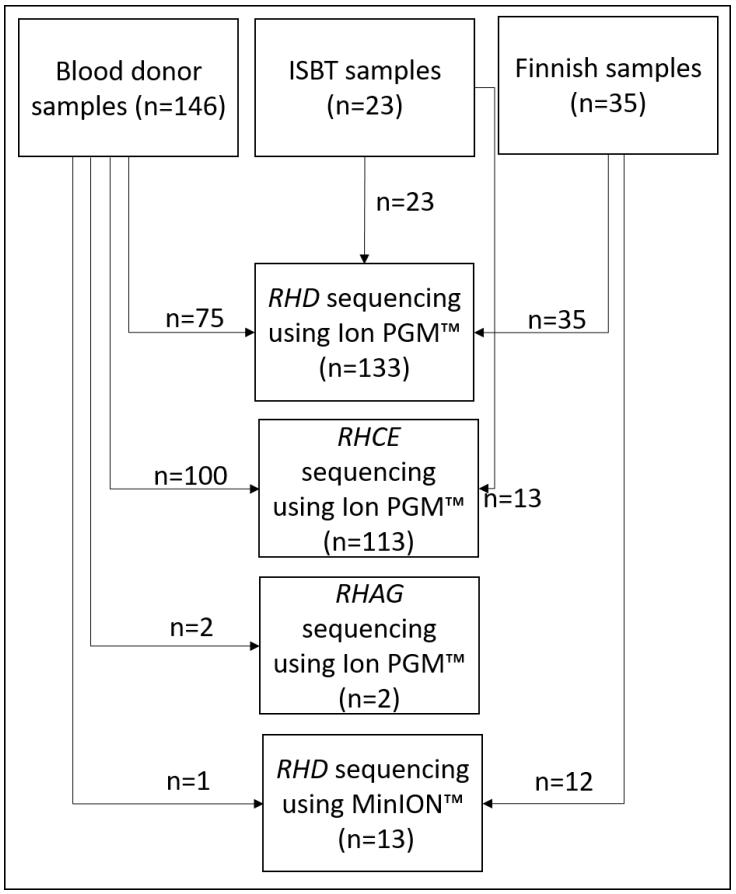


Figure 2.4: Flowchart illustrates the number of samples genotyped on Ion PGM™ for each target gene including *RHD*, *RHCE*, and *RHAG*, and number of samples genotypes for the *RHD* gene using MinION™.

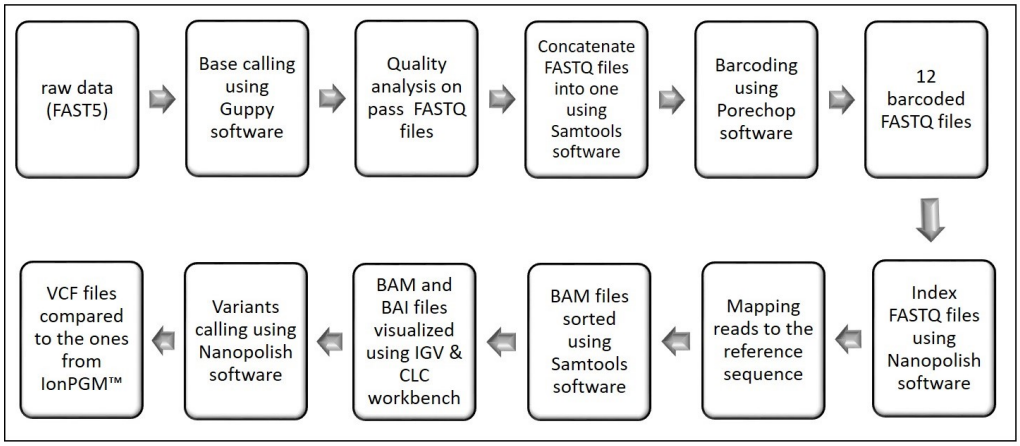


Figure 2.5: MinION™ data analysis workflow. Analysis started with calling bases of the raw data (FAST5) using Guppy basecaller v3.2.4. Pass FASTQ files were subjected to quality assessment using EPI2ME software which were then concatenated into one FASTQ file for barcoding using Samtools software v.1.4.1. Barcoding was then performed using Porechop software v.0.2.1 to divide reads into 12 barcoded files. The reads were then mapped to the *RHD* reference sequence hg38 which were visualised using IGV software v.2.5.3 and CLC Main Workbench 10 software. Calling variants was performed using Nanopolish software v.0.9.0 which were then compared to the ones detected from Ion PGM™.

```
#!/bin/bash

#NP=/home/ubuntu/mybin/nanopolish/
#FAST5=/home/ubuntu/vol/WT_raw_data/fast5/out_1/workspace
#SUM=/home/ubuntu/vol/WT_raw_data/fast5/out_1
#REF=/home/ubuntu/vol/WT_raw_data/fast5/out_1/pass/merged_fastq/barcoded_reads/chr1_ref.fa
echo "bwa index $REF"

[ ! -d $PWD/mapped ] && mkdir -p $PWD/mapped
for i in $(seq -w 01 12)
do
        echo "$NP/nanopolish index -d $FAST5/ -s $SUM/sequencing_summary.txt BC${i}/BC${i}_trim.fastq"
done

for i in $(seq -w 01 12)
do
    echo "bwa mem -x ont2d -t 6 $REF BC${i}/BC${i}_trim.fastq | samtools sort -o mapped/BC${i}.sorted.bam -T reads.tmp"
done

for i in $(seq -w 01 12)
do
    echo "samtools index mapped/BC${i}.sorted.bam"
done

for i in $(seq -w 01 12)
do
    echo "samtools stats mapped/BC${i}.sorted.bam"
done

for i in $(seq -w 01 12)
do
    echo "$TOOLS/nanopolish/nanopolish variants --ploidy 2 -m 0.2 -t $CORES -v --calculate-all-support -w "NC_000016_10:0-
22377" -r $CONS/barcode${BAR1}_trim.fasta -b $MAP/barcode${BAR1}.sorted.bam -g $TMP/${REF}.fa -o
$VCF/barcode${BAR1}_trim.vcf"
done

#echo "Compress and index the vcf"
#bgzip $VCF/barcode${BAR1}_trim.vcf
#tabix -p vcf $VCF/barcode${BAR1}_trim.vcf.gz
```

Figure 2.6: A bash loop script used for MinION™ data analysis. Bash is a programming language that enables command execution subsequently. Passed raw files (fast5) were barcoded and then mapped to the human genome reference sequence (hg38) generating BAM and BAI files. Variants calling was then performed using Nanopolish software v.0.9.0.

3. *RHD* & *RHAG* Next Generation Sequencing and Single Molecule Sequencing in Blood Group Genotyping

3.1 Introduction

The Rh blood group system (ISBT004) is the second most important blood group after ABO (Avent and Reid, 2000; Avent et al., 2006) and one of the most polymorphic blood group systems. The *RHD* and *RHCE* genes located on chromosome 1 (1p33.1_1p36) encode the RhD protein and RhCcEe protein, respectively (Colin et al., 1991; Wagner and Flegel, 2000). The RhD antigen is the most clinically significant antigen in the Rh system due

to its high immunogenicity and being the main cause of HDFN (Le van Kim et al., 1992). The *RHD* and the *RHCE* genes show 93.8% homology in their introns and coding exons (Okuda et al., 1999). The similarities between these two genes give an indication to their evolutionary rise from the same ancestral gene through duplication (Carritt et al., 1997; Okuda et al., 1999, 2000). Recombination, deletion, and point mutations in these two genes generate the eight most common Rh haplotypes which include: $\mathrm{R_1}$ (DCe), $\mathrm{R_2}$ (DcE), $\mathrm{R_0}$ (Dce), $\mathrm{R_z}$ (DCE), r (dce), $\mathrm{r^y}$ (dCE), $\mathrm{r'}$ (dCe), and $\mathrm{r''}$ (dcE) (Noizat-Pirenne et al., 1998).

Serological testing is fast, cost friendly, and efficient; however, it is limited by many factors; for instance, the availability of antisera (Jungbauer, 2011a; Orzińska et al., 2018), reactivity of the antibodies, and the antigen status like weak or partial expression. Current assignment of a partial or a weak D phenotype would require an extensive collection of monoclonal anti-D. Monoclonals to low frequency Rh antigens, such as $\mathrm{Go^a}$, BARC, and Tar, to identify specific partial D phenotypes are still unavailable. Serological testing also leads to predicting the Rh genotype based on the most common haplotype present in the population, which for some

cases is incorrect (Sillence et al., 2017).

Unlike serological testing, genotyping provides the freedom to analyse a wider range of blood group antigens including low frequency antigens, for instance: Go^a, BARC, and Tar which can cause HDFN and allo-immunisation (Avent and Reid, 2000; Daniels, 2013c). Complete BGG could be widely used in transfusion practice where serology tests may fail to clarify issues or resolve discrepancies.

Extensive efforts have been made to use molecular genotyping tests, as an alternative to serological testing, ranging from low to high throughput assays (Fichou et al., 2014). Different DNA microarray-based tests were introduced that enable genotyping of variant blood groups by targeting specific SNPs (Hashmi et al., 2005; Avent et al., 2009; Goldman et al., 2015; Chang et al., 2016; Fichou and Férec, 2017; López et al., 2018; Bub and Castilho, 2019). Although these assays are very accurate, they have some limitations. They are designed to target certain nucleotides or DNA regions through PCR, while novel variants remain unknown (Fichou et al., 2014; Avent et al., 2015). Complete DNA sequencing could be the most relevant technique to thoroughly study blood group variations and over-

come limitations in other assays (Avent et al., 2015; Lane et al., 2016).

Genomics was radically reshaped during the past decades by the introduction of the first generation sequencing in the 1970s (Sanger et al., 1977), and then the introduction of the second (Anderson and Schrijver, 2010; Pareek et al., 2011) and third generation sequencing (Nagarajan and Pop, 2013; Lu et al., 2016; Ardui et al., 2018) less than three decades later. Since it was introduced in 2005, NGS has greatly impacted the genetic research field by elevating both throughput and data generated, and at the same time lowering significantly the cost of sequencing per nucleotide (Fichou and Férec, 2017; Fichou et al., 2014; Pareek et al., 2011; Tilley and Grimsley, 2014; Zhang et al., 2011). NGS is used in HLA testing (Profaizer et al., 2016), which creates a strong impetus to introduce NGS for BGG (Möller et al., 2016). Genotyping could be utilized to genotype blood transfusion dependent patients who are at risk of allo-sensitisation (Bakanay et al., 2013; Fichou et al., 2016; Chou et al., 2017; Dezan et al., 2017). It could also be used to genotype donors and create a database that would make finding and recall of compatible donors for transfusion easier (Denomme and Van Oene, 2005; Ribeiro et al., 2009; Stabentheiner et al., 2011; Fichou and

Férec, 2017).

Different studies have aimed to use NGS in BGG using a variety of approaches. Dezan et al. (2017), Chou et al. (2017) and Schoeman et al. (2018) used whole exome sequencing (WES) to identify Rh variation but the high similarity between the *RHD* and *RHCE* genes could make it challenging to analyse data especially in exons 8 and 10 where there are no differences between the two genes. Hyland et al. (2017) used LR-PCR using primers described by Gassner et al. (1997) to amplify the *RHD* gene from exon 2 to exon 7 but omitting exons 1, 8, 9 and 10.

NGS is a high-throughput sequencing, depends on PCR amplification, that generates a vast amount of data at a low cost. However, short reads generated during library preparation made *de novo* assembly for large genomes difficult due to repetitive DNA sequence (Schatz et al., 2010; Zhang et al., 2011; Lu et al., 2016), which initiated the need for a new sequencing approach that uses longer reads, later called the TGS (Lu et al., 2016).

MinION™ sequencer offered by Nanopore Technologies (Oxford, United Kingdom) is the only sequencing technology that offers direct DNA se-

quencing using long reads up to 2 Mb and real time data analysis while reducing the cost to about $ 1,000 (Branton et al., 2008). MinION™ sequencing chemistry depends on driving the DNA through nanopores, which are embedded in a membrane submerged in ionic solution. By applying a voltage, a DNA molecule passes through the pores which causes changes in the ionic current running through the pores. These changes in the flowing current occur in a distinctive manner which are recorded and then translated to bases (Meller et al., 2000, 2001; Branton et al., 2008).

MinION™ sequencer has been used in different studies including BGG (Matern et al., 2017). Other studies have shown the great potential of the MinION™ sequencer on spot sequencing for example, during the Ebola virus outbreak in Guinea (Quick et al., 2016), Zika virus epidemic in Brazil and the Americas in 2015 (Faria et al., 2017), and during COVID-19 pandemic (Marquez et al., 2020; Lopez-Alvarez et al., 2020).

3.2 Aim and Objectives

This study aimed to fully sequence the *RHD* gene using NGS (Ion PGM™) to study alleles present in the population and identify novel variants. It

also focused on establishing Rh allele specific reference sequences by investigating intronic SNPs and their relation to a specific Rh haplotypes. We also aimed to assess the suitability and efficiency of MinION™ sequencer in BGG and its ability to enable *RHD* allele determination. The *RHAG* gene was also investigated for mutations that might be linked to disturbed RhD expression (weak or null D). To achieve these aims, objectives were as follow:

1. Test all samples for *RHD* zygosity using dPCR to allow the use of hemizygous *RHD* samples to unequivocally establish reference sequences for specific *RHD* alleles. Sequencing *RHD* hemizygous samples will ensure only one allele is amplified and sequenced to establish Rh allele specific reference sequences.

2. Investigate the relation of intronic SNPs to R_2 haplotype which were previously detected by Halawani (2015). The suspected intronic SNPs, that could be haplotype specific, were not confirmed due to low coverage and low number of samples tested in Halawani (2015) work. Specific primers were designed to amplify target region of SNPs for Sanger sequencing to confirm their relation to the R_2 haplotype.

3. Design LR-PCR *RHD* specific primers, to amplify the complete *RHD* gene including promoter, introns and all exons, that would allow specific amplification from the target gene without contamination from the homologous *RHCE* gene.

4. Fully sequence the *RHD* gene using NGS on Ion PGM™ platform and analyse the sequencing data and call variants to determine the *RHD* alleles and intronic SNP patterns for different haplotypes.

5. Utilise *RHD* intronic SNPs to establish Rh haplotype specific reference sequences. Identifying intronic changes in different *RHD* alleles is important as these SNPs could be a useful target to identify *RHD* haplotypes. Intronic changes may also lead to allelic dropout when attempting target sequencing; therefore, identifying these changes is crucial.

6. Design primers to amplify and sequence the *RHAG* gene using Ion PGM™, as mutations in *RHAG* can lead to weak or null D phenotype. The *RHAG* gene sequencing is targeted for samples that show weak D reactivity in serology but where no mutation in *RHD* gene was detected.

7. Use MinION™ to sequence the *RHD* gene to compare with data generated from Ion PGM™ to test the suitability and efficiency of the MinION™ sequencer in BGG.

8. Analyse MinION™ data to achieve allele phasing (which referes to the ability to segregate reads and assign alleles to different allele in heterozygous samples) and accurate variant calling to determine alleles. Variant calling, a process to identify variants in a sequence set when reads are aligned to the reference sequence, is crucial to assess throughput and accuracy before adapting nanopore sequencing to BGG.

3.3 Results

3.3.1 *RHD* Zygosity

Samples (n=206) (Table 3.1) with different Rh genotypes presumed from serology results were first tested using ddPCR to determine *RHD* zygosity. An additional 12 samples (not shown) were also tested for *RHD* zygosity which were included in the investigation of *RHD* intronic variation using Sanger sequencing, see section 3.3.2. The presence or absence of the *RHD* amplification on the ddPCR platform was used to determine whether the samples were hemizygous or homozygous *RHD*-positive or *RHD*-negative, respectively (Figure 3.1). Samples showing *RHD5* or *RHD7/AGO1* ratios ≈1 were determined to be homozygous *RHD*-positive, samples with ratios ≈0.5 were classified as hemizygous *RHD*-positive, and samples showing a ratio ≈0 were classified as *RHD*-negative (Table 3.1).

Samples included 10 $\mathrm{R_1R_1}$ (DCe/DCe), 22 $\mathrm{R_1r}$ (DCe/dce), 10 $\mathrm{R_2R_2}$ (DcE/DcE), 16 $\mathrm{R_2r}$ (DcE/dce), 66 $\mathrm{R_1R_2}$ (DCe/DcE), 13 $\mathrm{R_0r}$ (Dce/dce), 1 $\mathrm{R_2R_z}$ (DcE/DCE), 6 rr (dce/dce), 7 r'r (dCe/dce), 6 r''r (dcE/dce), as determined by serology. RhD negative samples were collected for the purpose of *RHCE* se-

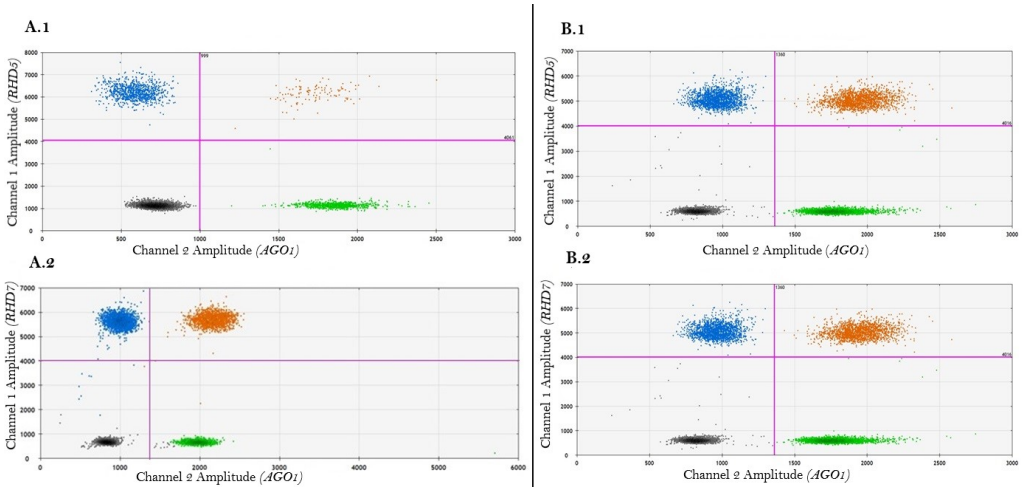


Figure 3.1: Two-dimensional amplitude plot from two representative samples tested for *RHD* zygosity using ddPCR. Blue plot represents *RHD* gene targets (exon5/exon7) positive droplets (FAM), green plot represents reference gene *AGO1* positive droplets (HEX), orange represents droplets positive for the *RHD* and reference gene *AGO1* ,and grey plot represents negative droplets. In A and B, samples were tested for *RHD* exon 5 (A.1/B.1) and *RHD* exon 7 (A.2/B.2) against the reference gene *AGO1*. Sample in A presented a ratio of 1 for both *RHD* targets against the reference gene *AGO1* and was considered *RHD* homozygous, and sample in B presented a ratio of 0.5 for both targets and was considered hemizygous for the *RHD* gene.

quencing. Samples were therefore tested for the presence of the *RHD* gene to confirm deletion. Twenty-three ISBT samples and 35 Finnish samples (Section 2.1) were also tested for *RHD* zygosity.

Zygosity results were compatible with the serologically predicted genotype except for the samples listed in Table 3.2. Sample (004_14), previously classified by serology as being phenotypically $\mathrm{R_1r}$ (DCe/dce), expressed

ratios of 1.06 and 0.99 for the *RHD5* and *RHD7* multiplex reactions, respectively (Table 3.2). This result contradicted previous serological classifications and indicated that the sample expressed two copies of the *RHD* gene. Sample (004_42), previously classified by serology as being phenotypically $\mathrm{R_2R_2}$, expressed ratios of 0.54 and 0.47 for the *RHD5* and *RHD7* multiplex reactions, respectively (Table 3.2). This result contradicted previous serological classifications and indicated that the sample expressed one copy of the *RHD* gene (hemizygous). In a similar manner, samples 004_35, 004_36, 004_37, 004_38, 004_39 and 004_40 were previously classified by serology as being phenotypically $\mathrm{R_1R_2}$. However, given the ratios from the *RHD5* (average 0.51) and *RHD7* (average 0.51) multiplex reactions, these samples only express one copy of the *RHD* gene and are therefore classified as being *RHD* hemizygous. One $\mathrm{R_1R_1}$sample (004_07) showed a discrepancy between hemizygous *RHD5* (ratio 0.54) and homozygous *RHD7* (ratio 1.01) indicating a possible deletion of exon 5 in one of the *RHD* alleles.

Blood donor sample (004_88) classified by serology as r''r showed a hemizygous *RHD* gene on ddPCR with 0.51 and 0.47 ratios for *RHD5* and *RHD7* assays, respectively. Six ISBT samples (004_101- 004_106) showed a

ratio of 0.5 for the *RHD7* assay indicating a hemizygous *RHD* gene and a ratio ≈0 for the *RHD5* assay indicating a deletion in exon 5.

Finnish samples (004_117, 004_118, 004_119, 004_120, 004_121, 004_122, 004_123, 004_126, 004_127, 004_128, 004_129, 004_130, 004_131, 004_133, 004_134, 004_135, 004_136, 004_137, 004_138, 004_139, 004_142, 004_143, 004_144, 004_145, 004_146, 004_149, 004_150, 004_151) were serologically classified as RhD negative presented a ratio ≈0.5 for *RHD5* and *RHD7* assays against the reference gene indicating the presence of a hemizygous *RHD* gene. One Finnish sample (004_125) showed ratios of ≈1 for both *RHD5* and *RHD7* assays against the reference sequence indicating the presence of homozygous *RHD* gene.

Table 3.1: Serologically predicted genotype (Rh serology), ethnicity of donors, dPCR *RHD* zygosity results, and *RHD* allele as determined by NGS (Ion PGM™).

Lab no.	Rh serology[Y]	Ethnicity[Y]	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype [▽]
004_01	R_1R_1	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_02	R_1R_1	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_03	R_1R_1	Other	Homozygous	-	-	-	*RHD*01*	NA
004_04	R_1R_1	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_05	R_1R_1	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_06	R_1R_1	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_07	R_1R_1	Caucasian	Discrepancy	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_08	R_1r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_09	R_1r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_10	R_1r	Chinese	Hemizygous	-	-	-	*RHD*01*	NA
004_11	R_1r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_12	R_1r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_13	R_1r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_14	R_1r	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_15	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_16	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_17	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_18	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_19	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_20	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_21	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_22	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_23	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_24	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_25	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_26	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)

continued ...

Lab no.	Rh serology$^{\Upsilon}$	Ethnicity$^{\Upsilon}$	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype $^{\nabla}$
004_27	R_1r	Caucasian	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_28	R_1r	Caucasian	Hemizygous	8C>G	1	Ser3Cys	*RHD*01W.03*	R_1 (DCe)
004_29	R_1R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_30	R_1R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_31	R_1R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_32	R_1R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_33	R_1R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_34	R_1R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_35	R_1R_2	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	NA
004_36	R_1R_2	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_37	R_1R_2	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_38	R_1R_2	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_39	R_1R_2	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_40	R_1R_2	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_41	R_2R_2	Caucasian	Homozygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_42	R_2R_2	Not disclosed	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_43	R_2R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_44	R_2R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_45	R_2R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_46	R_2R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_47	R_2R_2	Caucasian	Homozygous	-	-	-	*RHD*01*	NA
004_48	R_2r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_49	R_2r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_50	R_2r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_51	R_2r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_52	R_2r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_53	R_2r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_54	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_55	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)

continued ...

Lab no.	Rh serology$^{\Upsilon}$	Ethnicity$^{\Upsilon}$	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype $^{\nabla}$
004_56	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_57	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_58	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_59	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_60	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_61	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_62	R_2r	Caucasian	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_63	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_64	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_65	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_66	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_67	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_68	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_69	R_0r	Caucasian	Hemizygous	602C>G	4	Thr201Arg	*RHD*09.03.01*	R_0 (Dce)
				667T>G	5	Phe223Val		
				819G>A	6	silent		
004_70	R_0r	Caucasian	Hemizygous	602C>G	4	Thr201Arg	*RHD*09.01.02*	R_0 (Dce)
				667T>G	5	Phe223Val		
				744C>T	5	silent		
				1025T>C	7	Ile342Thr		
004_71	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_72	R_0r	Caucasian	Hemizygous	-	-	-	*RHD*01*	NA
004_73	R_0r	Caucasian	Hemizygous	505A>C	4	Met169Leu	*RHD*17.02*	R_0 (Dce)
				509T>G	4	Met170Arg		
				514A>T	4	Ile172Phe		
				544T>A	4	Ser182Thr		
				577G>A	4	Glu193Lys		
				594A>T	4	Lys198Asn		
				602C>G	4	Thr201Arg		
004_74	R_2R_z	Caucasian	Homozygous	-	-	-	*RHD*01*	NA

continued …

Lab no.	Rh serology[Υ]	Ethnicity[Υ]	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype [▽]
004_75	rr	Caucasian	Negative	NA	NA	NA	NA	NA
004_76	rr	Caucasian	Negative	NA	NA	NA	NA	NA
004_77	rr	Caucasian	Negative	NA	NA	NA	NA	NA
004_78	rr	Caucasian	Negative	NA	NA	NA	NA	NA
004_79	rr	Caucasian	Negative	NA	NA	NA	NA	NA
004_80	rr	Caucasian	Negative	NA	NA	NA	NA	NA
004_81	r′r	Caucasian	Negative	NA	NA	NA	NA	NA
004_82	r′r	Caucasian	Negative	NA	NA	NA	NA	NA
004_83	r′r	Caucasian	Negative	NA	NA	NA	NA	NA
004_84	r′r	Caucasian	Negative	NA	NA	NA	NA	NA
004_85	r′r	Caucasian	Negative	NA	NA	NA	NA	NA
004_86	r′r	Black/Caribbean	Negative	NA	NA	NA	NA	NA
004_87	r′r	Black/Caribbean	Negative	NA	NA	NA	NA	NA
004_88	r″r	Caucasian	Hemizygous	399delG	3	Val134fs	Novel variant 1	NA
004_89	r″r	Caucasian	Negative	NA	NA	NA	NA	NA
004_90	r″r	Caucasian	Negative	NA	NA	NA	NA	NA
004_91	r″r	Caucasian	Negative	NA	NA	NA	NA	NA
004_92	r″r	Caucasian	Negative	NA	NA	NA	NA	NA
004_93	r″r	Caucasian	Negative	NA	NA	NA	NA	NA
004_94	No data	No data	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_95	No data	No data	Hemizygous	809T>G	6	Val270Gly	*RHD*01W.01*	R_1 (DCe)
004_96	R_1r	No data	Hemizygous	186G>T	2	Leu62Phe	*RHD*03.01*	R_0 (Dce)
				410C>T	3	Ala137Val		
				455A>C	4	Asn152Thr		
				602C>G	4	Thr201Arg		
				667T>G	5	Phe223Val		
004_97	No data	No data	Hemizygous	602C>G	4	Thr201Arg	*RHD*09.03.01*	R_0 (Dce)
				667T>G	5	Phe223Val		
				819G>A	6	silent		

continued …

Lab no.	Rh serology$^{\Upsilon}$	Ethnicity$^{\Upsilon}$	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype $^{\nabla}$
004_98	No data	No data	Hemizygous	1063G>A	7	Gly355Ser	*RHD*25*	R_1 (DCe)
004_99	No data	No data	Hemizygous	1048G>C	7	Asp350His	*RHD*04.04*	R_1 (DCe)
				1057G>T	7	Gly353Trp		
				1059A>G	7	Gly353Trp		
				1060G>A	7	Ala354Asn		
				1061C>A	7	Ala354Asn		
004_100	No data	No data	Hemizygous	=	=	=	*RHD*04.04*	R_1 (DCe)
004_101	R_1r	No data	Hemizygous	505A>C	4	Met169Leu	*RHD*06.01*$^{\zeta}$	R_2 (DcE)
				509T>G	4	Met170Arg		
				514A>T	4	Ile172Phe		
				544T>A	4	Ser182Thr		
				577G>A	4	Glu193Lys		
				594A>T	4	Lys198Asn		
				602C>G	4	Thr201Arg		
				667T>G	5	Phe223Val		
				697G>C	5	Glu233Gln		
				712G>A	5	Val238Met		
				733G>C	5	Val245Leu		
				744C>T	5	Silent		
				787G>A	5	Gly263Arg		
				800A>T	5	Lys267Met		
004_102	R_1r	No data	Hemizygous	=	=	=	Undetermined$^{\otimes}$	NA
004_103	R_1r	No data	Hemizygous	=	=	=	Undetermined$^{\otimes}$	NA

continued …

Lab no.	Rh serology$^{\Upsilon}$	Ethnicity$^{\Upsilon}$	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype $^{\nabla}$
004_104	R_2r	No data	Hemizygous	505A>C	4	Met169Leu	Undetermined$^{\otimes}$	NA
				509T>G	4	Met170Arg		
				514A>T	4	Ile172Phe		
				544T>A	4	Ser182Thr		
				577G>A	4	Glu193Lys		
				594A>T	4	Lys198Asn		
				602C>G	4	Thr201Arg		
				667T>G	5	Phe223Val		
				676G>C	5	Ala226Pro		
				697G>C	5	Glu233Gln		
				712G>A	5	Val238Met		
				733G>C	5	Val245Leu		
				744C>T	5	Silent		
				787G>A	5	Gly263Arg		
				800A>T	5	Lys267Met		
004_105	R_2r	No data	Hemizygous	=	=	=	Undetermined$^{\otimes}$	NA
004_106	No data	No data	Hemizygous	=	=	=	Undetermined$^{\otimes}$	NA
004_107	No data	No data	Hemizygous	8C>G	1	Ser3Cys	*RHD*01W.03*	R_1 (DCe)
004_108	R_1r	No data	Hemizygous	329T>C	2	Leu110Pro	*RHD*07.01*	R_1 (DCe)
004_109	R_1r	No data	Hemizygous	329T>C	2	Leu110Pro	*RHD*07.01*	R_1 (DCe)
004_110	R_1r	No data	Hemizygous	329T>C	2	Leu110Pro	*RHD*07.01*	R_1 (DCe)
004_111	R_1r	No data	Hemizygous	505A>C	4	Met169Leu	*RHD*17.01*	multiple
				509T>G	4	Met170Arg		
				514A>T	4	Ile172Phe		

continued . . .

Lab no.	Rh serology[Y]	Ethnicity[Y]	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype [▽]
004_112	No data	No data	Hemizygous	329T>C	2	Leu110Pro	Novel variant 2	NA
				505A>C	4	Met169Leu		
				509T>G	4	Met170Arg		
				514A>T	4	Ile172Phe		
				544T>A	4	Ser182Thr		
				577G>A	4	Glu193Lys		
				594A>T	4	Lys198Asn		
				602C>G	4	Thr201Arg		
004_113	No data	No data	Hemizygous	-	-	-	*RHD*01*	NA
004_114	No data	No data	Homozygous	-	-	-	*RHD*01*	NA
004_115	No data	No data	Hemizygous	-	-	-	*RHD*01*	NA
004_116	No data	No data	Homozygous	-	-	-	*RHD*01*	NA
004_117	D-	No data	Hemizygous	1227G>A	9	Splice site	*RHD*01EL.01*	R_1 (DCe)
004_118	D-	No data	Hemizygous	1227G>A	9	Splice site	*RHD*01EL.01*	R_1 (DCe)
004_119	D-	No data	Hemizygous	8C>G	1	Ser3Cys	Novel variant 3	NA
				49delG	1	Ala17fs		
004_120	D-	No data	Hemizygous	952C>T	7	Arg318*	*RHD*01N.61*	R_1 (DCe)
004_121	D-	No data	Hemizygous	712delG	5	Val238Fs	*RHD*01N.33*	R_1 (DCe)
004_122	D-	No data	Hemizygous	885G>T	6	Met295Ile	*RHD*11*	R_0 (Dce)
004_123	D-	No data	Hemizygous	845G>A	6	Gly282Asp	*RHD*15*	R_2 (DcE)
004_124	Weak D+	No data	Hemizygous	1212C>A	9	Asp404Glu	*RHD*01W.72*	Not reported
004_125	Weak D+	No data	Homozygous	48G>C	1	Trp16Cys	Undetermined	NA
				602C>G	4	Thr201Arg		
				667T>G	5	Phe223Val		
004_126	D-	No data	Hemizygous	147delA	1	fs, stop	*RHD*01EL.04*	R_1 (DCe)
				IVS1+6delA	(1)			
004_127	D-	No data	Hemizygous	829G>A	6	Gly277Arg	Novel variant 10	NA
004_128	D-	No data	Hemizygous	829G>A	6	Gly277Arg	Novel variant 10	NA
004_129	D-	No data	Hemizygous	829G>A	6	Gly277Arg	Novel variant 10	NA
004_130	D-	No data	Hemizygous	829G>A	6	Gly277Arg	Novel variant 10	NA

continued ...

Lab no.	Rh serology$^{\curlyvee}$	Ethnicity$^{\curlyvee}$	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype $^{\triangledown}$
004_131	D-	No data	Hemizygous	784delC	5	Gln262fs	Novel variant 5	NA
004_132	Weak D+	No data	Hemizygous	782C>T	5	Pro261Leu	Novel variant 8	NA
004_133	D-	No data	Hemizygous	421delG	3	Val141fs	Novel variant 9	NA
				422T>A	3	Val141Glu		
004_134	D-	No data	Hemizygous	421delG	3	Val141fs	Novel variant 9	NA
				422T>A	3	Val141Glu		
004_135	D-	No data	Hemizygous	421delG	3	Val141fs	Novel variant 9	NA
				422T>A	3	Val141Glu		
004_136	D-	No data	Hemizygous	486+1G>A	(3)	splice site	*RHD*01EL.08*	R_1 (DCe)
004_137	D-	No data	Hemizygous	-	-	-	Undetermined$^{\bowtie}$	NA
004_138	D-	No data	Hemizygous	-	-	-	Undetermined$^{\bowtie}$	NA
004_139	D-	No data	Hemizygous	829G>A	6	Gly277Arg	Novel variant 10	NA
004_140	Weak D+	No data	Hemizygous	8C>G	1	Ser3Cys	*RHD*01W.03*	NA
004_141	Weak D+	No data	Hemizygous	782C>T	5	Pro261Leu	Novel variant 8	NA
004_142	D-	No data	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_143	D-	No data	Hemizygous	1154G>C	9	Gly385Ala	*RHD*01W.02*	R_2 (DcE)
004_144	D-	No data	Hemizygous	1154G>C	9	Gly385Ala	Novel variant 4	NA
				1163T>G	9	Leu388Arg		
004_145	D-	No data	Hemizygous	519C>G	4	Tyr173*	Novel variant 6	NA
004_146	D-	No data	Hemizygous	845G>A	6	Gly282Asp	*RHD*15*	R_2 (DcE)
004_147	Weak D+	No data	Hemizygous	1016G>C	7	Gly339Ala	Novel variant 7	NA
004_148	D+	No data	Hemizygous	602C>G	4	Thr201Arg	*RHD*09.03.01*	R_0 (Dce)
				667T>G	5	Phe223Val		
				819G>A	6	silent		
004_149	D-	No data	Hemizygous	330-331delGT	2	Phe111fs	*RHD*01N.35*	Not reported
004_150	D-	No data	Hemizygous	602C>G	4	Thr201Arg	Undetermined$^{\propto}$	NA
				667T>G	5	Phe223Val		
				819G>A	6	Silent		
				336-486 del	3			

continued ...

Lab no.	Rh serology$^{\curlyvee}$	Ethnicity$^{\curlyvee}$	*RHD* Zygosity	Nucleotide	Exon (Intron)	Amino acid	*RHD* Allele	Linked haplotype $^{\nabla}$
004_151	D-	No data	Hemizygous	395_396 insGG	3	Lys133fs	No designation assigned	NA
004_152	R_1R_2	Multiple	Homozygous	NA	NA	NA	Not sequenced	NA
004_205								

Blue highlight indicates samples show discrepant results between Rh serology predicted genotype and *RHD* zygosity results using dPCR.
*Termination codon.
$^{\curlyvee}$As supplied by the NHSBT, Bristol, United Kingdom, and the Finnish Red Cross Blood Service, Helsinki, Finland.
$^{\nabla}$Rh haplotype linked to the *RHD* allele as found in RhesusBase (Wagner and Flegel, 2014).
** RhD antigen titre after elution +1 as reported by the Finnish research group (Tammi, 2019).
- no SNPs detected.
NA= Not applicable.
= Similar to previous.
$^{\otimes}$ Samples previously genotyped as hybrid *RHD*06.01* allele, but results remain inconclusive due to the lack of coverage in exon 6 (expected conversion).
$^{\zeta}$ *RHD* NGS findings do not agree with existence genotyping data (DVa) provided by ISBT (1996).
$^{\bowtie}$One or multiple failed amplicons in LR-PCR.
$^{\propto}$sample showed homozygous *RHD* gene using ddPCR indicating a compound heterozygote variant *RHD* alleles.

Table 3.2: Samples with discrepant results between the serology predicted Rh genotypes and *RHD* gene zygosity results using dPCR.

Lab no.	**Rh serology***	***RHD5/AGO1* ratio**	***RHD7/AGO1* ratio**	***RHD* zygosity**
004_07	R_1R_1	0.54	1.01	Discrepancy [a]
004_14	R_1r	1.06	0.99	Homozygous [b]
004_35	R_1R_2	0.46	0.51	Hemizygous [c]
004_36	R_1R_2	0.51	0.51	Hemizygous [c]
004_37	R_1R_2	0.53	0.49	Hemizygous [c]
004_38	R_1R_2	0.51	0.53	Hemizygous [c]
004_39	R_1R_2	0.52	0.51	Hemizygous [c]
004_40	R_1R_2	0.53	0.52	Hemizygous [c]
004_42	R_2R_2	0.54	0.47	Hemizygous [d]
004_88	r″r	0.51	0.47	Hemizygous[e]
004_101	R_1r	0.0	0.45	Hemizygous[f]
004_102	R_1r	0.0	0.49	Hemizygous[f]
004_103	R_1r	0.0	0.53	Hemizygous[f]
004_104	R_2r	0.0	0.53	Hemizygous[f]
004_105	R_2r	0.0	0.50	Hemizygous[f]
004_106	No data	0.0	0.58	Hemizygous[f]
004_117	D-	0.50	0.50	Hemizygous[g]
004_118	D-	0.51	0.52	Hemizygous[g]
004_119	D-	0.51	0.50	Hemizygous[g]
004_120	D-	0.53	0.52	Hemizygous[g]
004_121	D-	0.45	0.52	Hemizygous[g]
004_122	D-	0.51	0.49	Hemizygous[g]
004_123	D-	0.51	0.49	Hemizygous[g]
004_125	Weak D+	1.0	1.0	Homozygous[h]
004_126	D-	0.51	0.51	Hemizygous[g]
004_127	D-	0.51	0.52	Hemizygous[g]
004_128	D-	0.52	0.50	Hemizygous[g]
004_129	D-	0.51	0.53	Hemizygous[g]
004_130	D-	0.49	0.50	Hemizygous[g]
004_131	D-	0.47	0.50	Hemizygous[g]
004_133	D-	0.53	0.52	Hemizygous[g]
004_134	D-	0.52	0.49	Hemizygous[g]
004_135	D-	0.54	0.52	Hemizygous[g]
004_136	D-	0.52	0.51	Hemizygous[g]
004_137	D-	0.51	0.51	Hemizygous[g]
004_138	D-	0.51	0.53	Hemizygous[g]
004_139	D-	0.53	0.52	Hemizygous[g]
004_142	D-	0.51	0.50	Hemizygous[g]
004_143	D-	0.52	0.52	Hemizygous[g]
004_144	D-	0.55	0.50	Hemizygous[g]
004_145	D-	0.51	0.53	Hemizygous[g]
004_146	D-	0.53	0.50	Hemizygous[g]

continued ...

Lab no.	Rh serology*	*RHD5/AGO1* ratio	*RHD7/AGO1* ratio	*RHD* zygosity
004_149	D-	0.55	0.53	Hemizygous$^{(g)}$
004_150	D-	0.52	0.54	Hemizygous$^{(g)}$
004_151	D-	0.50	0.50	Hemizygous$^{(g)}$

* As supplied by the NHSBT, Bristol, UK or Finnish Red Cross Blood Service, Helsinki, Finland. (a) Sample shows discrepancy between hemizygous *RHD5* and homozygous *RHD7* meaning that it is likely one of the *RHD* alleles has a deletion in exon 5. Samples showing incompatible dPCR results with serologically predicted genotypes indicating incorrectly assigned genotype by serology include: (b) $\mathrm{R_1r}$ sample shows homozygous *RHD* gene, (c) $\mathrm{R_1R_2}$ samples show hemizygous *RHD* gene, (d) $\mathrm{R_2R_2}$ sample shows hemizygous *RHD* gene, (e) r″r sample genotype predicted by serology as r″r found to be *RHD* hemizygous, (f) Samples presented *RHD5/AGO1* ratio = 0 and *RHD7/AGO1* ≈0.5 indicating a hemizygous *RHD* gene but with a deletion in exon 5, and (g) Samples phenotyped by serology as RhD negative and tested as *RHD* hemizygous. (h) Sample phenotyped by serology as RhD weak and tested as *RHD* homozygous indicating two variant *RHD* alleles.

3.3.2 Sanger Sequencing

Seven intronic *RHD* SNPs were detected through NGS (Halawani, 2015) that were thought to be $\mathrm{R_2}$ haplotype specific. Two SNPs located in intron 2 (25,611,580 G>A; 25,614,400 C>G), three SNPs located in intron 3 (25,621,980 C>T; 25,625,471 T>C; 25,627,066 C>G), and two SNPs located in intron 8 (25,646,933 T>G; 25,648,349 T>C) (Halawani, 2015; Sillence et al., 2017). *RHD* primers were designed to amplify regions of interest. Samples with different Rh phenotypes were sequenced by Sanger sequencing to test if these SNPs were $\mathrm{R_2}$ haplotype specific. The samples tested included 6 $\mathrm{R_1R_1}$, 1 $\mathrm{R_1r}$, 6 $\mathrm{R_2R_2}$, 1 $\mathrm{R_2r}$, 8 $\mathrm{R_0r}$, and 6 $\mathrm{R_1R_2}$.

Target intronic SNPs were not detected in samples genotyped by serology as $\mathrm{R_1R_1}$, and $\mathrm{R_1r}$. In contrast, all seven SNPs were detected in samples genotyped by serology as $\mathrm{R_2R_2}$ and $\mathrm{R_2r}$ samples. Samples genotyped by serology as $\mathrm{R_0r}$ were included to test if they show $\mathrm{R_1}$ or $\mathrm{R_2}$ intronic SNP patterns, and they were all missing the SNPs ($\mathrm{R_1}$ pattern). Six

R_1R_2 samples were tested to confirm the previous results by the presence of two peaks in the trace chromatogram at the SNP position, in which one comes from the R_1 allele and another from the R_2 allele. For SNPs (25,611,580 G>A), (25,614,400 C>G), (25,625,471 T>C), (25,627,066 C>G), and (25,648,349 T>C), all samples but one R_1R_2 sample (004_38) have shown two different peaks at the SNP positions. Sample (004_38), serologically predicted as R_1R_2, was missing all intronic SNPs investigated suggesting that R_2 allele was assigned incorrectly for this sample. Figure 3.2 shows SNP position 25,614,400 C>G in 6 R_1R_2 samples showing two peaks except sample (004_38) that shows one peak (C).

All 6 R_1R_2 samples were missing the SNP (25,621,980 C>T) showing only C at the SNP position. For SNP (25,646,933 T>G), all samples except for (004_38) present the SNP peak (G) in the SNP position while missing the second peak (T); sample (004_38) showed only one peak (T) for that SNP. Sample (004_38) tested as hemizygous for *RHD* gene showing a ratio of 0.51 and 0.53 for *RHD5* and *RHD7* against the reference sequence *AGO1*. This indicates a discrepancy between serologically assigned Rh haplotype and ddPCR zygosity results. Table 3.3 summarizes Sanger sequencing results for *RHD* intronic SNP investigations for R_1, R_2, R_0, and R_1R_2 samples. Table 3.3 also gives the position of these SNPs corresponding to both hg19 and hg38 reference sequences.

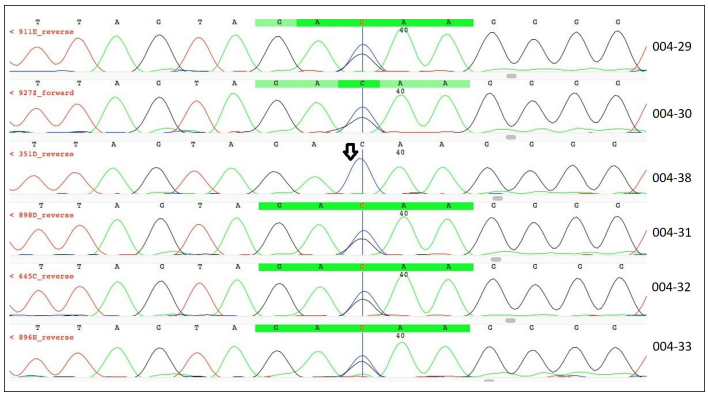


Figure 3.2: Sanger sequence data from reverse primer for R_1R_2 samples for *RHD* intronic SNP 25,614,400 C>G. All samples show two peaks (blue for C and black for G) in the SNP position except for the sample (004_38) that shows only one peak (C), indicated by an arrow.

Table 3.3: Summary of *RHD* intronic SNPs results for R_1, R_2, R_0, and R_1R_2 samples.

Position (hg19)	**Position (hg38)***	SNP	**Intron**	R_1	R_2	R_0	R_1R_2
25,611,580	25,285,089	G>A	2	-	+	-	+/-**
25,614,400	25,287,909	C>G	2	-	+	-	+/-**
25,621,980	25,295,489	C>T	3	-	+	-	-[@]
25,625,471	25,298,980	T>C	3	-	+	-	+/-**
25,627,066	25,300,575	C>G	3	-	+	-	+/-**
25,646,933	25,320,442	T>G	8	-	+	-	+[$]
25,648,349	25,321,858	T>C	8	-	+	-	+/-**

* Position of SNPs for *RHD* intronic SNPs detected using the hg19 reference sequence correlated to position in the hg38 reference sequence.
** All R_1R_2 samples except sample (004_38), that presented a hemizygous *RHD* gene on dPCR, show two peaks at SNP positions.
[$] All R_1R_2 samples show one SNP peak (G) except for sample (004_38), that presented a hemizygous *RHD* gene on dPCR, shows (T) in SNP position (-).
[@] SNP was negative in all 6 R_1R_2 samples indicating that SNP is not R_2 allele specific by Sanger sequencing.

3.3.3 *RHD* NGS Data Quality

To fully sequence the *RHD* gene, LR-PCR amplification was used to amplify the gene in six overlapping amplicons, Figure 3.3. Samples were prepared for sequencing (section 2.9) and sequenced using Ion PGM™, each sequencing run contained between 10 and 39 samples (Table 3.4). Figure 3.4 shows Ion PGM™ sequencing run summary from a single representative run, which indicates (a) high ISPs loading density in 316™ Chip of 80%, (b) recovered reads were over 3.9 million reads, and (c) reads length distribution, which refers to the number of bases sequenced in a DNA fragment, was around 120-130 bp. A higher percentage of polyclonal amplification was noticed in other runs with higher number of samples. ISPs loading density refers to the percentage of chip wells that contain ISP, in which red colour indicates adequate loading density, yellow indicates less-than adequate loading density, and red indicates inadequate loading density or absence of beads.

NGS data was obtained in FASTQ files, which are raw text sequence data with corresponding quality scores, was assessed for sequence quality. CLC workbench 10 software was utilised to generate a quality sequence summary report. Quality analysis on reads included, but was not limited to, reads lengths distribution (Figure 3.5 (A)), GC-content, quality distribution using PHRED score (Figure 3.5 (B)), per-base coverage (Figure 3.5 (C)), and sequence duplication level. High PHRED sequence quality score was

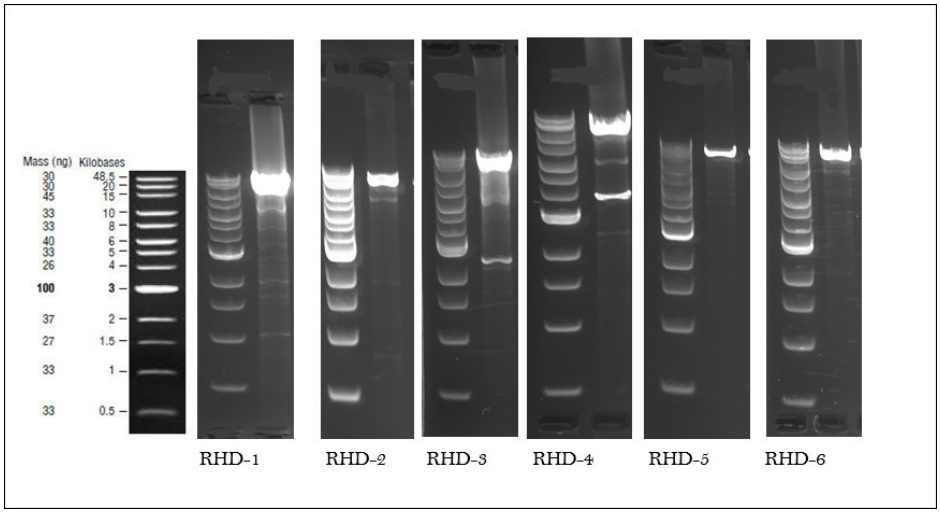


Figure 3.3: Gel electrophoresis of the 6 LR-PCR *RHD* amplicons. Six LR-PCR amplicons RHD1-RHD6 from a single representative *RHD* positive sample next to 1 kb Extend ladder (New England Biolabs®, U.K) was used to indicate amplicon size. All PCR reactions were accompanied with *RHD* positive control, *RHD* negative control (rr), and NTC (not shown).

noticed across all the samples sequenced. Figure 3.5 shows a PHRED score of over 30 for a representative sample that was sequenced indicating >99% accuracy. PHRED score is an algorithmic integer value representing the estimated probability of an error in the identification of a nucleic acid base (section 1.12). There is no recommended PHRED score for NGS; however, higher PHRED score indicates lower probability of error, therefore, higher accuracy.

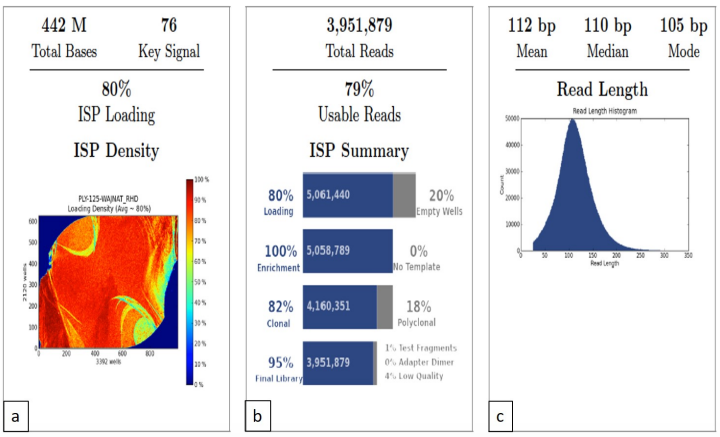


Figure 3.4: Ion PGM™ sequencing run summary from a single representative run. (a) The loading density in 316™ Chip was 80% of the chip capacity, giving a final throughput of 442 Mb. (b) Over 3.9 million reads were recovered after enrichment and polyclonal filters. (c) A histogram shows the distribution of the read length mean around 112bp, where the y-axis and x-axis represent the read count and the read length in bp, respectively.

Reads were mapped to the *RHD* gene in the hg38 reference sequence (Tatusova et al., 2016). The homologous *RHCE* gene was masked in the mapping process to prevent reads from scattering. Detailed mapping report was generated using the CLC workbench 10 which included coverage analysis

(Figure 3.6), mismatches analysis, read length distribution, insertion and deletion level, and quality for match and mismatch distribution.

Variant detection was performed on a minimum coverage (i.e the numbers of overlapping reads over a single base) of 30 and variants detected were analysed on a single base basis considering different parameters including number and percentage of reads and nucleotide count (Nielsen et al., 2011). Full coverage across the gene was achieved and depth of sequencing coverage was adequate at >50 and sufficient for calling variants (Figure 3.7) (Stabentheiner et al., 2011; Sims et al., 2014; Fichou et al., 2014).

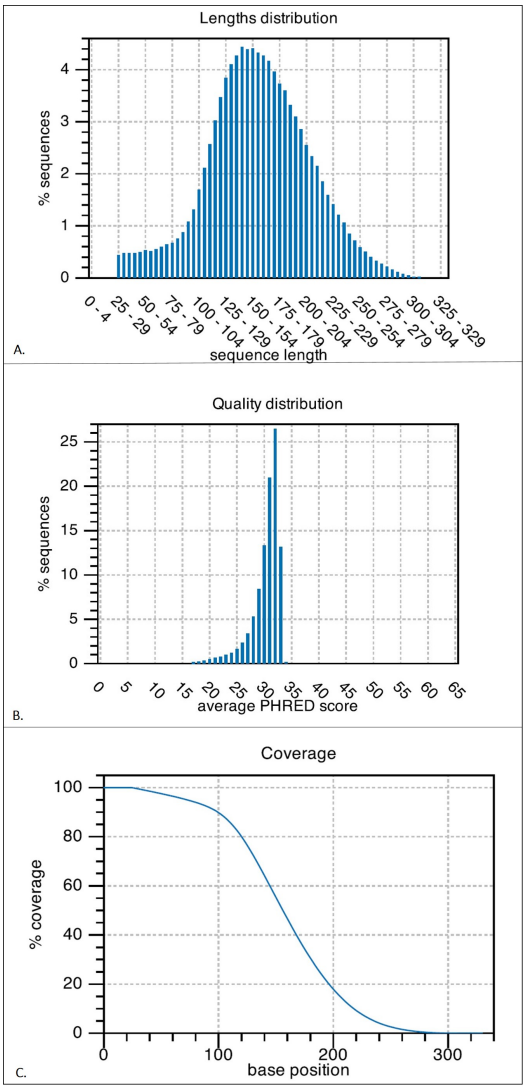


Figure 3.5: *RHD* gene sequence quality summary report generated by the CLC workbench 10 software from a single representative sample. A. Distribution of reads sequence lengths from a single representative sample; x axis: sequence length in base-pairs, y axis: number of reads showing a specific length normalized to the total number of sequences. B. PHRED score for *RHD* gene sequence from a single representative sample. x axis: PHRED-score; y axis: percentage of sequences, normalized to the total number of reads, noted at that quality. C. The number of sequences that cover the individual base positions from a single representative sample, x axis: base position; y axis: number of sequences covering individual base positions normalized to the total number of sequences.

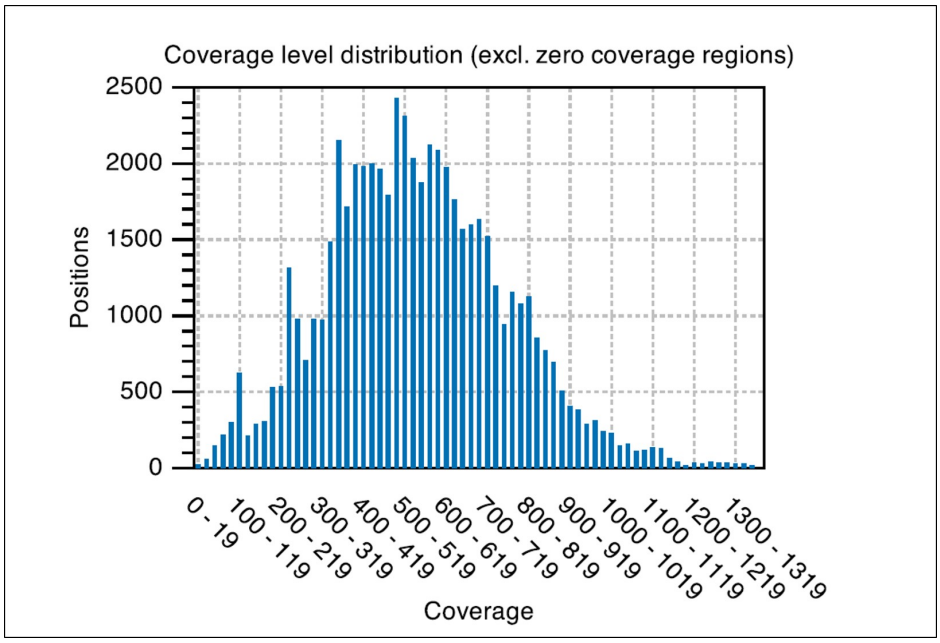


Figure 3.6: Level of coverage across the *RHD* gene from a single representative sample. Level of coverage (x axis) and number of bases observed at that level (y axis). A total of 57,317 positions (*RHD* gene) have coverage between 1 and 1,189. Positions of zero coverage (248,898,842) (chr1 genome) were not included in the graph.

Table 3.4: Ion PGM™ sequencing run summary report details for *RHD* sequencing runs.

Library no.	No. of samples	ISP loading (%)	Polyclonal (%)	Usable reads (%)	Total reads	Average reads /sample
1	10	86	34	60	3,249,115	324,911
2	26	76	30	62	2,967,256	114,125
3	27	84	40	54	2,836,007	105,037
4	33	87	37	57	3,107,386	94,163
5*	36	91	51	39	2,224,525	61,792
6	39	80	18	79	3,951,879	101,330

*High loading density (91%) but lowest usable reads (39%) could be due to variations in either library preparation and concentration of reads used or Ion OneTouch™ performance.

3.3.4 *RHD* Genotyping using NGS

In this study, we have collected (n= 205) samples, from which (n=133) samples were genotyped for the *RHD* gene, 53 $\mathrm{R_1R_2}$ samples were excluded from the sequencing as they tested homozygous for the *RHD* gene, and 18 samples (6 rr, 7 r'r, 5 r''r) tested as *RHD* negative on dPCR. Sequencing *RHD* hemizygous samples will ensure only one allele is sequenced to establish Rh allele specific reference sequences. From (n=133) samples genotyped for the *RHD* gene, 75 were blood donor samples, 35 were from RhD negative/weak Finnish pregnant females, and 23 from the ISBT workshop 1996 (ISBT, 1996), (Figure 3.8).

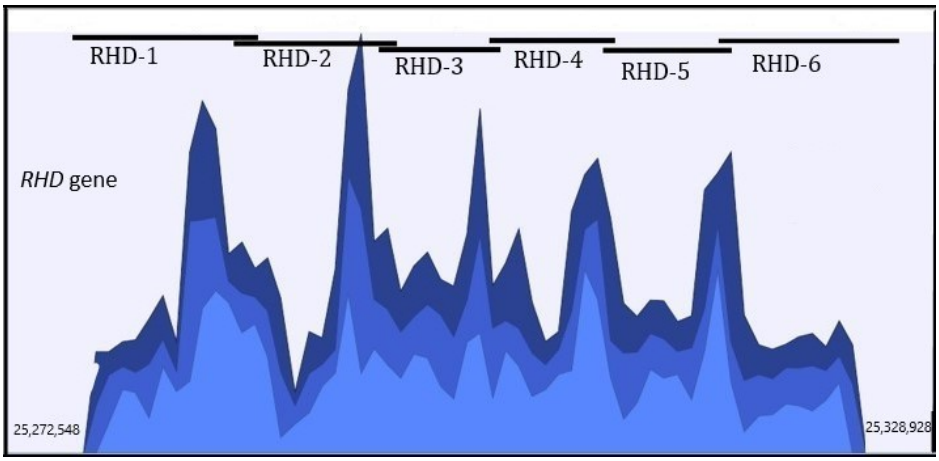


Figure 3.7: Depth of coverage across the *RHD* gene from a single representative sample. Six *RHD* amplicons are mapped to the *RHD* gene (top). Blue peaks indicate continuous and high level of coverage.

In order to establish reference *RHD* allele sequences, we aimed to sequence blood donor hemizygous *RHD* samples to avoid amplification from 2 different *RHD* alleles; nevertheless, *RHD* homozygous samples were also included in the sequence analysis to detect weak D that could be undetectable by serological testing due to the presence of a wild type copy of the *RHD* allele. We also sequenced the *RHD* gene in RhD negative samples or very weak D samples that tested *RHD* positive by ddPCR to determine the genetic reasons behind the negative or weakened expression, this included 1 blood donor sample (004_88) and 35 pregnant female samples supplied by the Finnish Red Cross Blood Service (004_117- 004_151).

Donor samples (n=75) (Table 3.1) with different Rh serologically predicted

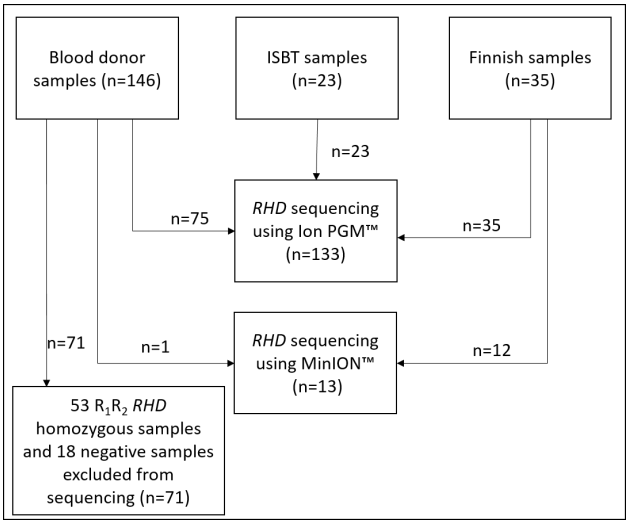


Figure 3.8: Flowchart illustrates number of samples tested on each sequencing platform. Blood donor samples not sequenced included 18 *RHD* negative samples and 53 $\mathrm{R_1R_2}$ samples found to be homozygous for the *RHD* gene using ddPCR.

genotypes were sequenced on the Ion PGM™ including 7 $\mathrm{R_1R_1}$ (DCe/DCe), 21 $\mathrm{R_1r}$ (DCe/dce), 12 $\mathrm{R_1R_2}$ (DCe/DcE), 7 $\mathrm{R_2R_2}$ (DcE/DcE), 15 $\mathrm{R_2r}$ (DcE/dce), 11 $\mathrm{R_0r}$ (Dce/dce) and 1 $\mathrm{R_2R_z}$ (DcE/DCE). One r''r (dcE/dce) sample which showed hemizygous *RHD* gene on ddPCR was also sequenced. We purposely included the 6 $\mathrm{R_1R_2}$ samples (004_35, 004_36, 004_37, 004_38, 004_39, 004_40) that tested as hemizygous for the *RHD* gene and included another set of 6 homozygous $\mathrm{R_1R_2}$ samples (004_29, 004_30, 004_31, 004_32, 004_33, 004_34) for a comparison, which were randomly chosen from the remaining 60 homozygous $\mathrm{R_1R_2}$ samples.

The *RHD* gene was also sequenced from the ISBT workshop 1996 samples (n=23), however Rh serological data was available for only 10 sam-

ples. *RHD* allele was determined and supplied for these samples (n=23). Finnish samples (n=35) with negative and weak D phenotypes by serology were also sequenced for the *RHD* gene to determine the genetic background for the RhD negative expression. Samples included 22 (dCe/dce), 3 (dCe/dCe), 2 (dCe/DcE), 4 (dcE/dce) and 4 (dce/dce).

Data was aligned to the hg38 reference sequence using the CLC workbench 10 software (Qiagen Ltd, United Kingdom). It is noteworthy that the *RHD* reference sequence (NC_000001.11) (Tatusova et al., 2016) is *RHD*DAU0 (RHD*10.00)* that presents a SNP in exon 8 (1136C>T) causing amino acid change Thr379Met; therefore, all samples sequenced presented a SNP in exon 8 (1136T>C) Met379Thr, illustrated in Figure 3.9. Multiple exonic SNPs linked to different known *RHD* alleles were detected across all the samples sequenced (n=133) and are shown in Table 3.5.

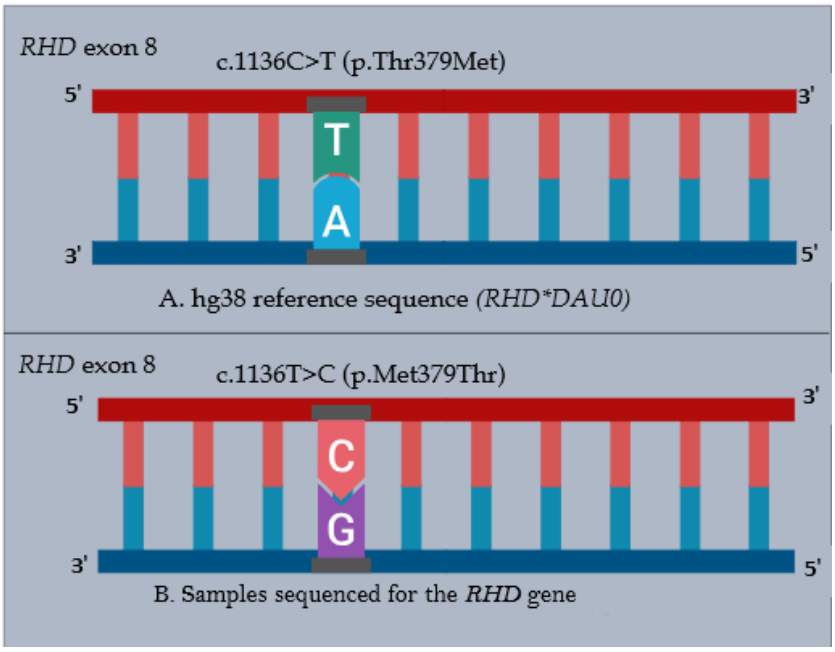


Figure 3.9: The *RHD* hg38 reference sequence is a variant allele. (A) *RHD* hg38 reference sequence is a variant allele that shows c.1136T (p.379Met) in exon 8 that encodes *RHD*DAU0*. (B) All samples sequenced showed c.1136T>C (p.Met379Thr) when aligned to the reference sequence.

Data analysis of the (n= 133) samples (Table 3.1) showed that 48 samples carried the *RHD* wild type allele *RHD*01*. Other samples (n= 85) showed a *RHD* variant allele, in which 59 samples linked to known *RHD* variant alleles, and 17 samples showed novel *RHD* variant alleles. Table 3.5 lists exonic changes and the predicted amino acid alterations they encode for each allele. In the remaining 9 samples (004_102, 004_103, 004_104, 004_105, 004_106, 004_125, 004_137, 004_138, 004_150) the *RHD* genotyping results were inconclusive. Inconclusive genotyping results were due to either the lack of a complete sequence, failure to amplify one or more of

the amplicons, or the presence of two different possible variant alleles.

Known *RHD* variants detected include weak, partial, DEL and null *RHD* alleles. Four different weak D alleles were detected in 24 samples in which: 16 samples (004_07, 004_15, 004_16, 004_17, 004_18, 004_19, 004_20, 004_21, 004_22, 004_23, 004_24, 004_25, 004_26, 004_27, 004_94, 004_95) were genotyped as *RHD*01W.01*, 14 samples (004_35, 004_41, 004_42, 004_54, 004_55, 004_56, 004_57, 004_58, 004_59, 004_60, 004_61, 004_62, 004_142, 004_143) were genotyped as *RHD*01W.02*, 3 samples (004_28, 004_107, 004_140) were genotyped as *RHD*01W.03*, and 1 sample (004_124) genotyped as *RHD*01W.72*.

Eleven partial weak D alleles were found in 17 samples in which 1 sample (004_101), genotyped previously by the ISBT (ISBT, 1996) as DVa, showed SNPs that encode the hybrid allele *RHD*06.01*. Three samples (004_108, 004_109, 004_110) were genotyped as *RHD*07.01*, 3 samples (004_69, 004_97, 004_148) were genotyped as *RHD*09.03.01*, 2 samples (004_123, 004_146) were genotyped as *RHD*15* and 2 samples (004_99, 004_100) were genotyped as *RHD*04.04*. Sample (004_111) genotyped as *RHD*17.01*, sample (004_111) genotyped as *RHD*17.02*, sample (004_98) was genotyped as *RHD*25*, sample (004_96) was genotyped as *RHD*03.01*, sample (004_70) was genotyped as *RHD*09.01.02*, and sample (004_122) was genotyped as *RHD*11*.

Three DEL alleles were detected in 4 samples in which: 2 samples (004_117, 004_118) were genotyped as *RHD*01EL.01*, 1 sample (004_126) was genotyped as *RHD*01EL.04*, and 1 sample (004_136) was genotyped as *RHD*01EL.08*.

Three negative *RHD* alleles were detected in 3 samples, in which: sample (004_149) was genotyped as *RHD*01N.35*, sample (004_120) was genotyped as *RHD*01N.61*, and sample (004_121) was genotyped as *RHD*01N.33*.

In 5 samples (004_102, 004_103, 004_104, 004_105, 004_106), genotyped previously by the ISBT (ISBT, 1996) as *RHD*06.01* (*DVI*), exonic changes linked to *RHCE4-5* insertion were detected. Lack of coverage in exon 6 was noted indicating a deletion of the exon. Adequate coverage was noticed over exons 4 and 5 which are amplified by the same amplicon as exon 6 (amplicon 4). Exon 5 SNP (c.676G>C) was detected in 3 samples (004_104, 004_105, 004_106) but not in samples serologically genotyped as $\mathrm{R_1r}$ (004_102, 004_103). Genotyping results for theses 5 samples remain inconclusive.

Sample (004_125) tested as *RHD* homozygous by dPCR and was serologically phenotyped as weak D. NGS data showed 4 heterozygous exonic changes c.48G>C (p.Trp16Cys) in exon 1, c.602C>G (p.Thr201Arg) in exon 4, c.667T>G (p.Phe223Val) in exon 5, and c.819G>A (silent) in exon 6. Possible alleles encoded by these exonic changes are either *RHD*09.03* and *RHD*01.01* or *RHD*09.04* and *RHD*01*, however, both options include a

wild type *RHD* allele that should produce normal RhD protein which does not explain the weak D reactivity in serology. Definite alleles remain inconclusive.

Table 3.5: Molecular basis of the different *RHD* alleles detected by NGS.

no. of samples	Nucleotide	Exon (Intron)	Amino acid	Allele name	Allele encodes	Haplotype
48	NA	NA	NA	*RHD*01*	Normal D	
1	48G>C	1	Trp16Cys	*RHD*01.01*	Normal D	DCe
16	809T>G	6	Val270Gly	*RHD*01W.01*	Weak D type 1	DCe
14	1154G>C	9	Gly385Ala	*RHD*01W.02*	Weak D type 2	DcE
3	8C>G	1	Ser3Cys	*RHD*01W.03*	Weak D type 3	DCe
1	1212C>A	9	Asp404Glu	*RHD*01W.72*	Weak D type 72	NR
1	505A>C	4	Met169Leu	*RHD*06.01*	DVI type 1	DcE
	509T>G	4	Met170Arg			
	514A>T	4	Ile172Phe			
	544T>A	4	Ser182Thr			
	577G>A	4	Glu193Lys			
	594A>T	4	Lys198Asn			
	602C>G	4	Thr201Arg			
	667T>G	5	Phe223Val			
	697G>C	5	Glu233Gln			
	712G>A	5	Val238Met			
	733G>C	5	Val245Leu			
	744C>T	5	Silent			
	787G>A	5	Gly263Arg			
	800A>T	5	Lys267Met			
1	602C>G	4	Thr201Arg	*RHD*09.01.02*	DAR1.2	Dce
	667T>G	5	Phe223Val			
	744C>T	5	Silent			
	957G>A	7	Silent			
	1025T>C	7	Ile342Thr			
3	602C>G	4	Thr201Arg	*RHD*09.03.01*	DAR3.1	Dce
	667T>G	5	Phe223Val			
	819G>A	6	Silent			
1	48G>C	1	Trp16Cys	*RHD*09.04*	weak D type 4.1	Dce
	602C>G	4	Thr201Arg			
	667T>G	5	Phe223Val			
	819G>A	6	Silent			
3	329T>C	2	Leu110Pro	*RHD*07.01*	DVII	DCe
1	505A>C	4	Met169Leu	*RHD*17.01*	DFR1	multiple
	509T>G	4	Met170Arg			
	514A>T	4	Ile172Phe			

continued ...

no. of samples	Nucleotide	Exon (Intron)	Amino acid	Allele name	Allele encodes	Haplotype
1	505A>C	4	Met169Leu	*RHD*17.02*	DFR2	DCe
	509T>G	4	Met170Arg			
	514A>T	4	Ile172Phe			
	544T>A	4	Ser182Thr			
	577G>A	4	Glu193Lys			
	594A>T	4	Lys198Asn			
	602C>G	4	Thr201Arg			
1	186G>T	2	Leu62Phe	*RHD*03.01*	DIIIa	Dce
	410C>T	3	Ala137Val			
	455A>C	4	Asn152Thr			
	602C>G	4	Thr201Arg			
	667T>G	5	Phe223Val			
1	885G>T	6	Met295Ile	*RHD*11*	Weak partial D 11	DCe
2	845G>A	6	Gly282Asp	*RHD*15*	Weak partial D 15	DcE
1	1063G>A	7	Gly355Ser	*RHD*25*	DNB	DCe
2	1048G>C	7	Asp350His	*RHD*04.04*	DIV type 4	DCe
	1057G>T	7	Gly353Trp			
	1059A>G	7	Gly353Trp			
	1060G>A	7	Ala354Asn			
‘	1061C>A	7	Ala354Asn			
2	1227G>A	9	Lys409Lys splice site	*RHD*01EL.01*	Del	DCe
1	147delA	1	fs, stop	*RHD*01EL.04*	Del	DCe
	IVS1+6delA	(1)				
1	486+1G>A	(3)	splice site	*RHD*01EL.08*	Del	DCe
1	712delG	5	Val238fs	*RHD*01N.33*	D-	DCe
1	330_331delGT	2	Phe111fs	*RHD*01N.35*	D-	NR
1	952C>T	7	Arg318*	*RHD*01N.61*	D-	DCe
1	399delG	3	Val134fs	Novel variant 1	NA	NA
1	329T>C	2	Leu110Pro	Novel variant 2	NA	NA
	505A>C	4	Met169Leu			
	509T>G	4	Met170Arg			
	514A>T	4	Ile172Phe			
	544T>A	4	Ser182Thr			
	577G>A	4	Glu193Lys			
	594A>T	4	Lys198Asn			
	602C>G	4	Thr201Arg			
1	8C>G	1	Ser3Cys	Novel variant 3	NA	NA
	49delG	1	Ala17fs			
1	1154G>C	9	Gly385Ala	Novel variant 4	NA	NA
	1163T>G	9	Leu388Arg			
1	784delC	5	Gln262fs	Novel variant 5	NA	NA
1	519C>G	4	Tyr173*	Novel variant 6	NA	NA

continued . . .

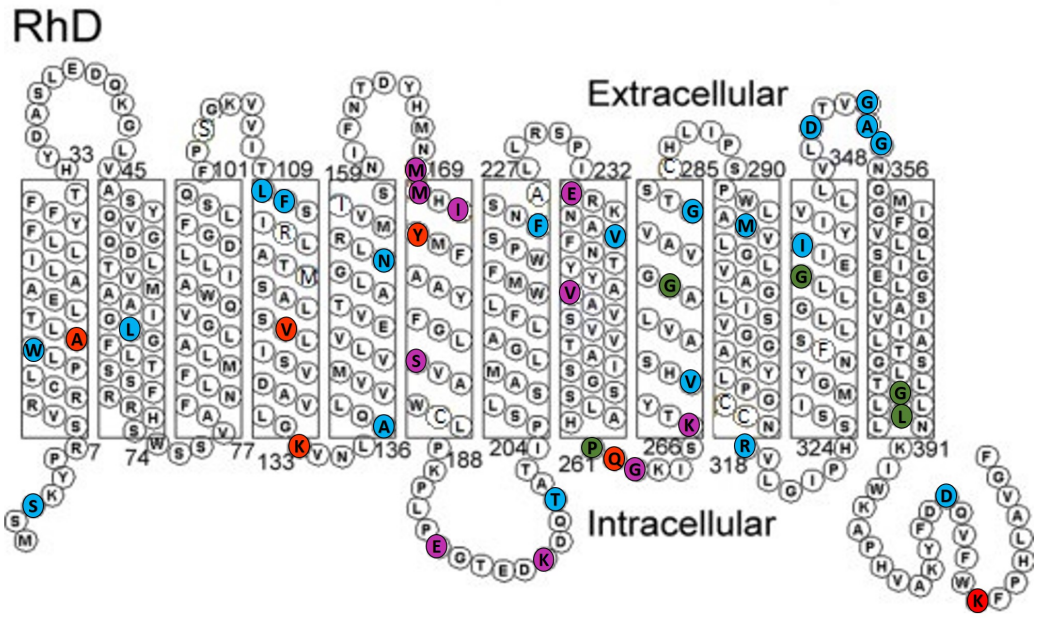
no. of samples	Nucleotide	Exon (Intron)	Amino acid	Allele name	Allele encodes	Haplotype
1	1016G>C	7	Gly339Ala	Novel variant 7	NA	NA
2	782C>T	5	Pro261Leu	Novel variant 8	NA	NA
3	421delG 422T>A	3 3	Val141fs Val141Glu	Novel variant 9	NA	NA
5	829G>A	6	Gly277Arg	Novel variant 10	NA	NA
1	395_396dupGG	3	Lys133fs	*	NA	NA

NA = Not applicable.
NR = Not reported.
*No designation assigned by ISBT.
Blue highlight indicates *RHD* alleles encode weak D.
Red highlight indicates *RHD* alleles encode weak partial D.
Green highlight indicates *RHD* alleles encode DEL phenotype.
Yellow highlight indicates *RHD* alleles encode null phenotype.
Gray highlight indicates novel *RHD* variant alleles.

3.3.4.1 NHSBT Blood Donor Samples

Of the NHSBT blood donor samples (n=75), one $\mathrm{R_1r}$ sample (004_28) showed a SNP in exon 1 (c.8C>G) Ser3Cys that encodes weak D type 3 *(RHD*01W.03)*. Thirteen $\mathrm{R_1r}$ samples (004_15, 004_16, 004_17, 004_18, 004_19, 004_20, 004_21, 004_22, 004_23, 004_24, 004_25, 004_26, 004_27) and 1 $\mathrm{R_1R_1}$ sample (004_07) showed a SNP in exon 6, (c.809T>G) Val270Gly that encodes weak D type 1 *(RHD*01W.01)*. Nine $\mathrm{R_2r}$ samples (004_54, 004_55, 004_56, 004_57, 004_58, 004_59, 004_60, 004_61, 004_62), two $\mathrm{R_2R_2}$ samples (004_41, 004_42), and one $\mathrm{R_1R_2}$ sample (004_35) showed the exon 9 (c.1154G>C) SNP that causes amino acid change p.Gly385Ala which encodes weak D type 2 *(RHD*01W.02)*. Sample (004_69) showed c.186G>T in exon 2, c.410C>T in exon 3, c.455A>C

and c.602C>G in exon 4, and c.667T>G in exon 5 mutations that predict the following amino acid changes p.Leu62Phe, p.Ala137Val, p.Asn152Thr, p.Thr201Arg, and p.Phe223Val, that encode *RHD*09.03.01*. Sample (004_70) showed mutations that encode for the *RHD*09.01.02 (DAR1.2)* allele which include: c.602C>G (p.Thr201Arg) in exon 2, c.667T>G (p.Phe223Val) and c.744C>T (silent) in exon 5, and c.957G>A (silent) and c.1025T>C (Ile342Thr) in exon 7. One sample (004_73) showed SNPs in exon 4 that encode for the variant *RHD*17.02* allele, which include c.505A>C (p.Met169Leu), c.509T>G (p.Met170Arg), c.514A>T (p.Ile172Phe), c.544T>A (p.Ser182Thr), c.577G>A (p.Glu193Lys), c.594A>T (p.Lys198Asn), and c.602C>G (p.Thr201Arg). One donor sample (004_88) phenotyped by serology as r''r was found to be *RHD* hemizygous using ddPCR. NGS data showed 399delG in exon 3 causing Val134fs which is a novel variant, listed here as novel variant 1, that has not been reported in the literature. Figure 3.10 illustrates different predicted amino acid changes on the RhD protein.



Amino acid highlighted were affected as follow: S3C, W16C, L62F, L110P, F111fs, K133fs, V134fs, A137V, V141fs, N152T, M169L, M170R, I172F, Y173, S182T, E193N, K198N, T201R, F223V, Q233Q, V238fs, V238M, V245L, P261L, Q262fs, G263R, K267M, V270G, G277E, G282D, M295I, R318*, G339A, D350H, G353W, A354N, G355S, G385A, L388R, D404E, K409K.

Figure 3.10: Amino acid changes predicted by NGS illustrated on the RhD protein. red: amino acid changes predicted from NGS that lead to stop codon, frame shift or splice site change; green: amino acid changes were linked to novel *RHD* variants; blue: amino acid changes predicted from NGS that link to known *RHD* variant; purple: amino acid changes predicted by NGS from *RHD-RHCE-RHD* hybrid genes. Image courtesy of Conroy et al. (2005).

Of the 28 samples that were serologically phenotyped as weak D, 26 of them were confirmed to be weak D by NGS and the *RHD* allele was determined. One $\mathrm{R_1r}$ sample (004_14) and the $\mathrm{R_2R_z}$ (004_74) sample were serologically predicted to be weak D but no SNPs in the *RHD* gene predicting amino acid changes in the RhD protein were detected by sequencing. For these two samples (004_14 and 004_74), the *RHAG* gene was sequenced to

test if there were any mutations in the *RHAG* gene that could lead to weak D expression, (section 3.3.6).

3.3.4.2 ISBT Samples

ISBT samples (n=23) (Table 3.1) from the 1996 workshop had the *RHD* gene sequenced for the presence of variant *RHD* alleles. Samples were tested for *RHD* zygosity and genotyped using NGS. The *RHD* genotyping results were as follows: 2 samples (004_94 and 004_95) were genotyped as weak D type 1 *RHD*01W.01*, 1 sample (004_96) was genotyped as partial D type *DIIIa* (*RHD*03.01*), 1 sample (004_97) was genotyped as *RHD*09.03.01*, 1 sample (004_98) was genotyped as partial D type *DNB* (*RHD*25*), 2 samples (004_99 and 004_100) were genotyped as partial *DIV* type 4 (*RHD*04.04*). One $\mathrm{R_1r}$ sample (004_101) was genotyped as the hybrid *RHD* gene *DVI* type 1 (*RHD*06.01*) although previously genotyped as *DVa* by ISBT (1996).

Five samples (004_102- 004_106) showed exonic SNPs in exons 4 and 5 associated with *RHCE4-5* insertion in the *RHD* gene, however, results remain inconclusive due to the breakage of coverage in exon 6 which may indicate deletion of that exon, which can be confirmed by exome sequencing PCR. One sample (004_107) was genotyped as weak D type 3 *RHD*01W.03*, 4 samples (004_108, 004_109, 004_110, 004_111) were genotyped as partial D type DVII *RHD*07.01*, and 4 samples (004_113, 004_114, 004_115, 004_116) were genotyped as wild type *RHD* gene *RHD*01*. Sample (004_112) that

showed hemizygous *RHD* gene on ddPCR presented mutations that encode both *RHD*07.01* and *RHD*17.02* which include c.329T>C (p.Leu110Pro) in exon 2 and c.505A>C (p.Met169Leu), c.509T>G (p.Met170Arg), c.514A>T (p.Ile172Phe), c.544T>A (p.Ser182Thr), c.577G>A (p.Glu193Lys), c.594A>T (p.Lys198Asn), and c.602C>G (p.Thr201Arg) in exon 4, which was considered a novel *RHD* variant allele (novel variant 2) consisting of a hybrid of these two alleles (Table 3.5).

3.3.4.3 Finnish Samples

35 samples (004_117- 004_151) (Table 3.1), supplied by the Finnish Red Cross Blood Service, were from serologically considered RhD negative pregnant females in which the genetic variation of the *RHD* allele could not be determined using the SSP-PCR (Tammi, 2019). All samples were tested for *RHD* zygosity using ddPCR, all samples were hemizygous for the *RHD* gene except sample (004_125) which was found to be homozygous (Table 3.2).

Complete sequencing of the *RHD* gene revealed 21 different variant *RHD* alleles in 31 out of 35 samples, (Table 3.1). In total, 13 previously described *RHD* variants were identified in 16 samples. These variants include 3 null alleles, 3 DEL alleles, 3 weak D alleles, 3 partial weak D alleles, and one allele in sample (004_151) has been reported by Karnot et al. (2016) but no designation has been assigned by the ISBT, as found in RhesusBase (Wagner and Flegel, 2014). The 3 null alleles were found in 3 samples

which include: *RHD*01N.61* allele in sample (004_120), *RHD*01N.33* allele in sample (004_121), and *RHD*01N.35* allele in sample (004_149). The 3 DEL alleles were found in 4 samples which include: *RHD*01EL.01* allele in samples (004_117 and 004_118), *RHD*01EL.08* allele in sample (004_136), and *RHD*01EL.04* allele in sample (004_126). The 3 weak D alleles were found in 4 samples which include: *RHD*01W.72* allele in sample (004_124), *RHD*01W.03* allele in sample (004_140), and *RHD*01W.02* allele in samples (004_142 and 004_143). The 3 partial D alleles were found in 4 samples which include: *RHD*11* allele in sample (004_122), *RHD*15* in samples (004_123 and 004_146), and *RHD*09.03.01* allele in sample (004_148). Amino acid changes that encode the different *RHD* alleles are listed in Table 3.5.

Eight different novel alleles identified in 15 samples, listed in Table 3.6, are encoded by 4 novel missense mutations, 3 frame shift mutations caused by deletions, and 1 nonsense mutation (Table 3.5). These mutations include: one SNP c.829G>A in exon 6 that predicts amino acid change p.Gly277Arg which was detected in 5 samples (004_127, 004_128, 004_129, 004_130, 004_139) (novel variant 10). The novel c.829G>A was confirmed using Sanger sequencing by the Finnish research group (Tammi, 2019; Tammi et al., 2020), since the same nucleotide change in the adjacent base, c.830G>A, has been reported to cause the *RHD*01W.12* (*RHD**weak D type 12) allele. Identification of novel SNPs showed consistent results using Sanger se-

quencing, NGS and TGS.

In five samples (004_127, 004_128, 004_129, 004_130, 004_139), serologically phenotyped as RhD negative, one exon 6 mutation c.829G>A that predicts amino acid change p.Gly277Arg was detected (Novel Variant 10). This missense mutation may indicate a *DEL* phenotype, because missense mutations should not cause a null phenotype. Affected amino acid in this novel variant allele Gly277 is the same amino acid mutation that encodes *RHD*01W.12* variant allele, however, the exonic change in this weak D type 12 allele is the adjacent base c.830G>A (p.Gly277Glu). The *RHD*01W.12* allele has been associated with extremely low antigen density (Wagner et al., 2000), which might explain the RhD negative phenotype in these 5 samples. In both exonic changes (c.829G>A and c.830G>A) a non-polar amino acid (Glycine) is changed to, a negatively charged amino acid (Arginine), for c.829G>A or (Glutamic acid) for c.830G>A. Therefore, it is expected that the novel mutation, c.829G>A (p.Gly277Arg), would have a similar effect on the RhD protein as the c.830G>A (p.Gly277Glu) change in *RHD*01W.12*. *RHD* DEL phenotype could be confirmed by D antigen quantifying tests. Figure 3.11 shows 3D modelling of the RhD protein showing (a) 277Arg and (b) 277Glu, where both show similar protein structure.

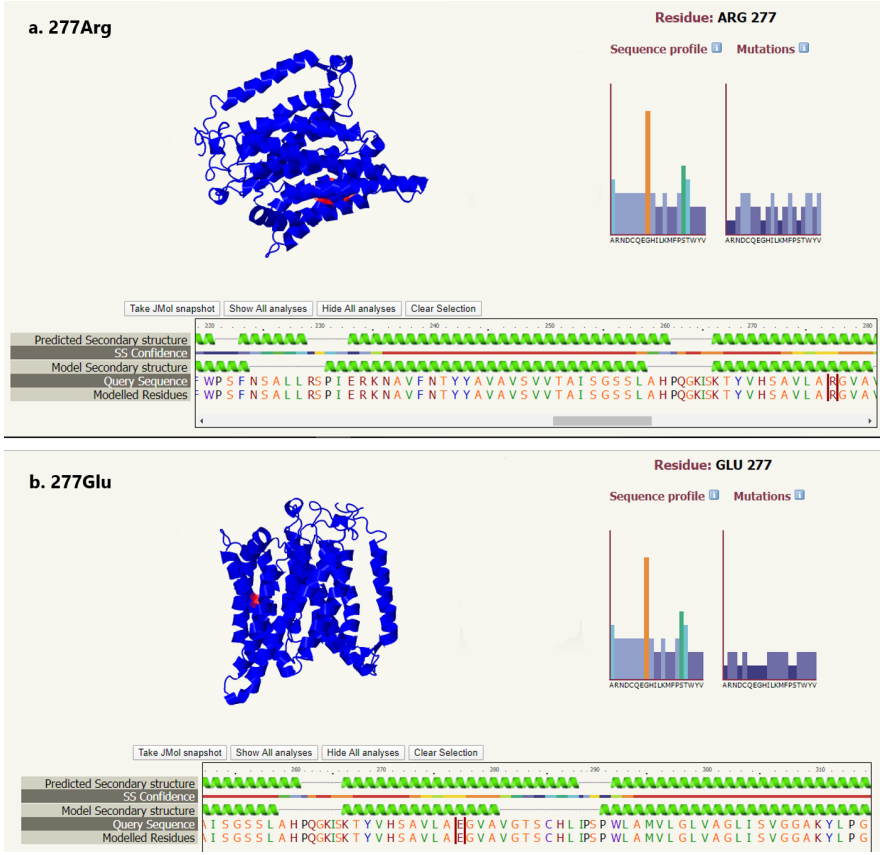


Figure 3.11: 3D model of the RhD protein showing (a) Arginine and (b) Glutamic acid in position 277. Red indicates the position 277 in the protein structure. Protein modelling was performed using the Phyre2 software (Kelley et al., 2015).

Two SNPs were detected in exon 3 c.421delG and c.422T>A that predict a frame shift and amino acid change p.Val141fs and p.Val141Glu which were detected in 3 samples (004_133, 004_134, 004_135) (novel variant 9).

A mutation in exon 5 c.782C>T that predicts amino acid change p.Pro261Leu was detected in 2 samples (004_132, 004_141) (novel variant 8). Two missense mutations were detected in sample (004_144). The missense mutation c.1154G>C was detected in exon 9. This mutation predicts amino acid change p.Gly385Ala which encodes the *RHD*01W.02* allele. However, an additional missense mutation c.1163T>G causing amino acid change p.Leu388Arg was detected in exon 9, for which it is considered a novel *RHD*01W.02* sub-allele (novel variant 4). Mutations c.8C>G and c.49delG in exon 1 which predict amino acid change p.Ser3Cys and a frame shift p.Ala17fs were found in sample (004_119) (novel variant 3). Another frame shift SNP c.784delC in exon 5 predicts p.Gln262fs and was detected in sample (004_131) (novel variant 5), and c.519C>G in exon 4 predicts a stop codon p.Tyr173* and was detected in sample (004_145) (novel variant 6). A mutation in exon 7 c.1016G>C was detected in sample (004_147) and predicts amino acid change p.Gly339Ala (novel variant 7).

Three samples (004_137, 004_138, 004_150) failed to amplify one or more of the *RHD* amplicons which could be due to variations, a hybrid gene or a deletion, in the *RHD* gene that compromised some of the *RHD*-specific LR-PCR primers binding regions.

Samples (004_137 and 004_138) failed to amplify *RHD* amplicon 6, which includes the area between part of intron 8 to exon 10 of the *RHD* gene. The Finnish research group have reported that sample 004_138 tested negative for *RHD* exon 10 in the SSP-PCR test, which indicates a deletion or a hybrid sequence in that region of the *RHD* gene. In sample (004_137), all exons were detected in SSP-PCR as reported by the Finnish research group (Tammi, 2019). For both samples, the *RHD* gene was sequenced from the remaining 5 amplicons but no variations that could affect RhD expression were detected and therefore the *RHD* gene was considered undetermined due to the lack of the full gene sequence.

In sample (004_150), the LR-PCR reactions of *RHD* amplicons 1, 2 and 3 failed, which cover the region between *RHD* exons 1 and 5. The LR-PCR was repeated using primers adapted from Hyland et al. (2017) to amplify the *RHD* region from exon 2 to exon 7. The results showed a complete deletion of exon 3 (336-486del) with a partial deletion of intron 2, which appeared as break of coverage in that region (Figure 3.12), in addition to mutations detected in exons 4, 5, and 6 c.602C>G, c.667T>G, and c.819G>A that predict amino acid changes p.Thr201Arg, p.Phe223Val, and a silent mutation, respectively.

In sample (004_125) that showed two copies of the *RHD* gene according to the ddPCR results (homozygous), 4 exonic mutations c.48G>C, c.602C>G,

c.667T>G, and 819G>A were detected in exons 1, 4, 5, and 6, respectively. These mutations predict the amino acid changes p.Trp16Cys, p.Thr201Arg, p.Phe223Val, and a silent mutation, respectively. In the NGS data, the frequency of the reads containing each mutation were approximately 0.5, which indicates that the mutations are present only in one copy i.e. heterozygous SNPs. Probable alleles encoded by these exonic changes are either *RHD*09.03* and *RHD*01.01* or *RHD*09.04* and *RHD*01*. The latter are the most probable alleles as it matches results acquired through SSP-PCR as reported by the Finnish research group (Tammi, 2019).

Table 3.6: Molecular basis of novel *RHD* variant alleles detected by NGS.

no. of samples	Nucleotide	Exon	Type	Amino acid	Consequence	Genbank accession
1	399delG	3	Del	Val134fs	Frame shift	MN624143
1	329T>C	2	SNV	Leu110Pro	Missense	
	505A>C	4	SNV	Met169Leu	Missense	
	509T>G	4	SNV	Met170Arg	Missense	
	514A>T	4	SNV	Ile172Phe	Missense	
	544T>A	4	SNV	Ser182Thr	Missense	
	577G>A	4	SNV	Glu193Lys	Missense	
	594A>T	4	SNV	Lys198Asn	Missense	
	602C>G	4	SNV	Thr201Arg	Missense	
1	8C>G	1	SNV	Ser3Cys	Missense	MN366001
	49delG	1	Del	Ala17fs	Frame shift	
1	1154G>C	9	SNV	Gly385Ala	Missense	MN365998
	1163T>G	9	SNV	Leu388Arg	Missense	
1	784delC	5	Del	Gln262fs	Frame shift	MN365997
1	519C>G	4	SNV	Tyr173*	Nonsense	MN365999
1	1016G>C	7	SNV	Gly339Ala	Missense	MN366002
2	782C>T	5	SNV	Pro261Leu	Missense	MN366000
3	421delG	3	Del	Val141fs	Frame shift	MN365995
	422T>A	3	SNV	Val141Glu	Missense	
5	829G>A	6	SNV	Gly277Arg	Missense	MN365996

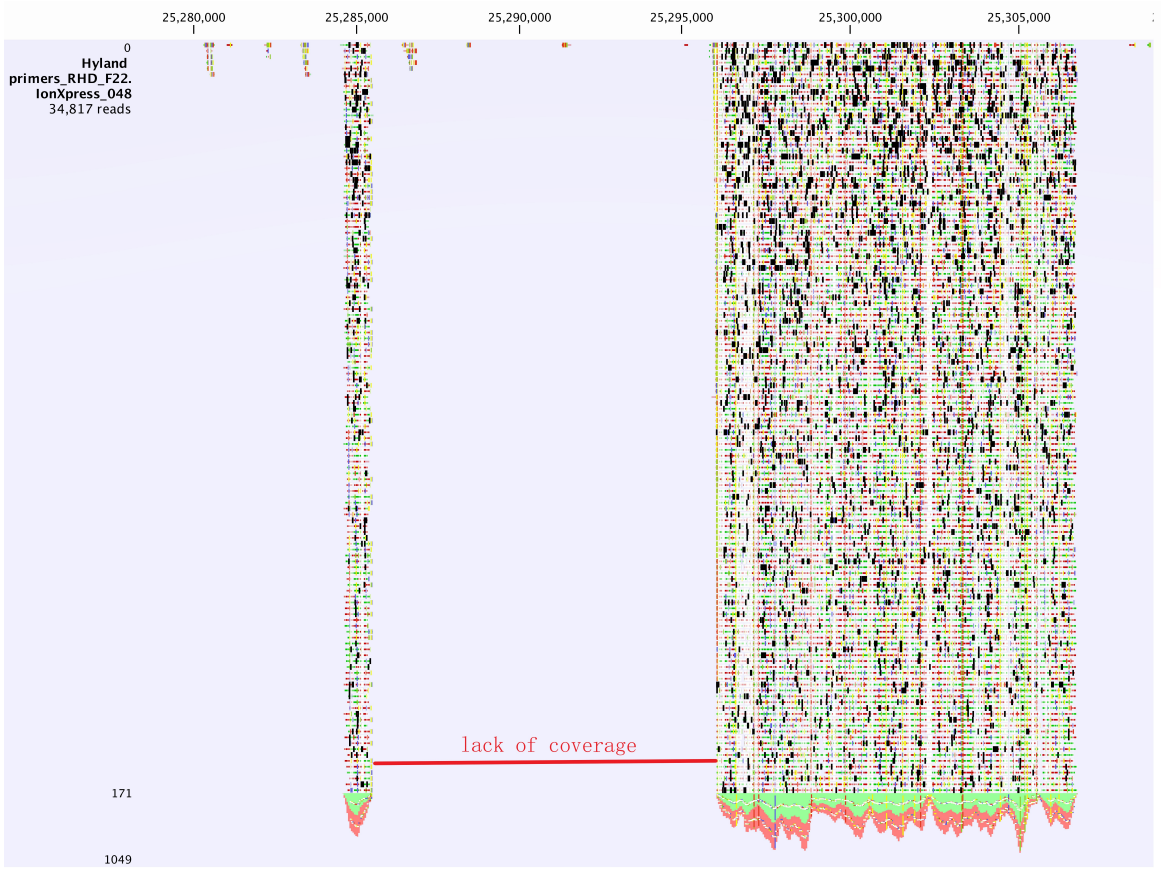


Figure 3.12: Complete deletion of intron 2 and exon 3 in sample 004_150 shown by the lack of coverage in that region. Sample was amplified using Hyland et al. (2017) primers to amplify the *RHD* gene from exon 2 to exon 7. NGS reads aligned against the hg38 reference sequence.

3.3.5 Intronic SNPs

Due to *RHD*DAU0 (RHD*10.00)* being the reference sequence hg38, 21 homozygous SNPs were detected in all samples sequenced, (Table 3.7) which are specific to the reference allele, i.e. *RHD*DAU0 (RHD*10.00)*. Multiple intronic SNPs are suspected to be haplotype specific, for example 23 SNPs, (Table 3.8) were homozygous SNPs in all samples with the $\mathrm{R_2}$ haplotype. They were detected in $\mathrm{R_2R_2}$, $\mathrm{R_2r}$, and in 3 of the 6 $\mathrm{R_1R_2}$ samples (004_35, 004_36, 004_37) which were determined by ddPCR to be hemizygous for the *RHD* gene. These SNPs were also present in 6 $\mathrm{R_1R_2}$ samples (D homozygous) (004_29, 004_30, 004_31, 004_32, 004_33, 004_34), and in the $\mathrm{R_2R_z}$ sample (004_74) as heterozygous SNPs.

Fifteen SNPs, (Table 3.9) were detected as homozygous in all $\mathrm{R_1R_1}$, $\mathrm{R_1r}$ and in 3 out of 6 $\mathrm{R_1R_2}$ samples (004_38, 004_39, 004_40) which were shown by ddPCR to be hemizygous for the *RHD* gene. They were also detected in all 6 $\mathrm{R_0r}$ samples (004_63, 004_64, 004_65, 004_66, 004_67, 004_68). These SNPs were also found as heterozygous SNPs in 6 $\mathrm{R_1R_2}$ samples (D homozygous) (004_29, 004_30, 004_31, 004_32, 004_33, 004_34), and in the $\mathrm{R_2R_z}$ sample (004_74).

Table 3.10 shows the different intronic SNPs detected and their correspondence in $\mathrm{R_2}$ and $\mathrm{R_1, R_0, R_z}$ *RHD* alleles in comparison to the reference sequence. From the 519 intronic SNPs detected most were not conserved

across each haplotype (data not shown). Most of the SNPs in Table 3.10 have been reported and show corresponding reference numbers in the database of SNPs (Sherry et al., 2001). All Rh haplotype specific intronic SNPs were mapped to the *RHD* gene (Figure 3.13).

Intronic SNPs were assessed for the *RHD* variant alleles for the ISBT samples (n=23). Samples genotyped as *RHD*01W.01* (004_94 , 004_95), *RHD*04.04* (004_99 and 004_100), *RHD*01W.03* (004_107), *RHD*07.01* (004_108, 004_109, 004_110), and *RHD*17.01* (004_111) showed $\mathrm{R_1}, \mathrm{R_0}, \mathrm{R_z}$ specific SNPs. The sample genotyped as partial D *RHD*25* (004_98) showed $\mathrm{R_2}$ specific SNPs.

Intronic SNPs pattern for 6 ISBT samples (004_101-004_106), where a hybrid *RHD* allele is expected, and 2 samples genotyped as *RHD*03.01* (004_96, 004_97) was challenging to determine. The *RHD*06.01* and *RHD*03.01* alleles showed some of the $\mathrm{R_1}, \mathrm{R_0}, \mathrm{R_z}$ SNPs and some of the $\mathrm{R_2}$ SNPs. In the 6 ISBT samples, showed exonic SNPs associated with *RHCE4-5* insertion in the *RHD* gene, intronic SNPs located between intron 3 and 6 were not consistent. This variation in intronic SNP patterns is due to *RHCE* insertion in the *RHD* gene. Intronic SNPs located in other introns were consistent and matched with patterns observed previously. Although some ISBT samples were lacking Rh phenotyping results, we attempted to predict the probable correct Rh haplotype using the *RHD* zygosity results, intronic SNP pattern and the *RHCE* sequence data, see chapter 4, section 4.3.4.

Intronic changes were also studied for the Finnish samples (n=35) to determine intronic SNP patterns, and whether $\mathrm{R}_1, \mathrm{R}_0, \mathrm{R}_\mathrm{z}$ SNPs or R_2 SNPs were present. We found that samples showed a clear cut pattern for intronic SNPs, including novel variant alleles. Samples that were found to be hemizygous for the *RHD* gene using ddPCR and were phenotyped as r′r, r′r′, rr showed homozygous $\mathrm{R}_1, \mathrm{R}_0, \mathrm{R}_\mathrm{z}$ intronic SNPs. Sample (004_125) phenotyped as r′r and showed hemizygous *RHD* gene on ddPCR also presented homozygous $\mathrm{R}_1, \mathrm{R}_0, \mathrm{R}_\mathrm{z}$ intronic SNPs. Other hemizygous samples phenotyped as r′r″ and r″r showed homozygous R_2 SNPs.

In 3 samples (004_69, 004_97, 004_148) genotyped as *RHD*09.03.01 (DAR3.1)*, 33 intronic SNPs were detected as homozygous SNPs, (Table 3.11), which could be specific to this allele. Eleven intronic SNPs were detected as homozygous SNPs (Table 3.12) in the 3 samples genotyped as *RHD*09.03.01 (DAR3.1)*, 1 sample genotyped as *RHD*09.01.02* (004_70), and in 1 sample genotyped as *RHD*03.01* (004_96), and as heterozygous SNPs in sample (004_125) that predicted to have two variant *RHD* genes in which one is expected to be *RHD*09.04*.

In Table 3.13 the most probable Rh haplotype was assigned for samples that were incorrectly assigned Rh haplotype or phenotyped as D negative and tested as *RHD* positive using ddPCR. For these samples, intronic SNP pattern, Rh serology results, *RHD* allele as identified by NGS, and linked

haplotype in literature were considered to assign the most probable haplotypes.

Table 3.7: Position of intronic variations and their reference SNP number detected in all samples sequenced. As the reference sequence is *RHD*DAU0 (RHD*10.00)*, these nucleotide changes are predicted to be *RHD*DAU0 (RHD*10.00)* specific.

Position	hg38	All samples	Intron	rs no. [α]
25,277,761	A	G	1	rs28661958
25,286,520	T	C	2	rs183024534
25,286,601	T	A	2	-
25,286,605	A	T	2	-
25,286,674	C	T	2	-
25,286,732	A	G	2	-
25,290,908	T	C	3	rs28521909
25,290,915	G	A	3	rs28562109
25,295,850	A	G	3	rs28451966
25,297,140	G	A	3	rs28786680
25,305,164	G	T	6	rs28703207
25,308,306	T	C	7	rs28374144
25,308,317	T	C	7	rs28719684
25,308,325	G	A	7	rs71493569
25,308,326	C	T	7	rs71493569
25,308,403	C	T	7	rs1801096
25,316,058	A	G	7	rs28453868
25,319,292	T	C	8	rs28397158
25,322,588	A	G	9	rs28435180
25,327,036	G	A	9	rs61777612
25,329,789	A	G	10	rs28654325

[α]From the database of SNPs (Sherry et al., 2001).

(-) Not found in the database of SNPs.

Grey indicates SNPs that were not detected by MinION™ sequencing.

Table 3.8: Intronic SNPs present in all samples with R_2 haplotype.

Position	SNP	Intron	rs no. [α]
25,282,654	A>G	1	rs3866916
25,285,089*	G>A	2	rs675072
25,287,909*	C>G	2	rs28718098
25,295,072	G>A	3	rs372986392
25,295,354	C>T	3	rs2904840
25,295,489*	C>T	3	rs190056379
25,295,708	G>A	3	rs182346769
25,295,731	A>G	3	rs201512625
25,295,739	G>A	3	rs200682399
25,295,753	A>G	3	rs143670081
25,298,980*	T>C	3	rs2904843
25,300,575*	C>G	3	rs2986167
25,305,898	A>G	6	rs12126031
25,307,714	G>A	7	rs2257611
25,308,845	G>C	7	rs2478025
25,311,722	T>A	7	rs796579065
25,316,269	A>G	7	rs2427767
25,320,442*	T>G	8	rs3927482
25,321,858*	T>C	8	rs28669938
25,323,393	C>T	9	rs77160738
25,323,618	G>C	9	rs201304363
25,323,713	G>C	9	rs202154122
25,327,668	A>G	9	-

[α]From the database of SNPs (Sherry et al., 2001).
(-) Not found in the database of SNPs.
* SNPs were investigated previously using Sanger sequencing.
Intronic SNPs were homozygous in all R_2r samples, R_2R_2 samples and in 3 out of the 6 R_1R_2 samples that tested as hemizygous for the *RHD* gene by ddPCR, and as heterozygous in R_1R_2 samples, and the R_2R_z sample.

Table 3.9: Intronic SNPs present in all samples with R_1, R_0, and R_Z haplotypes.

Position	SNP	Intron	rs no. [α]
25,284,544	G>C	1	rs2301153
25,292,953	G>A	3	rs28645510
25,295,317	G>A	3	rs2986157
25,295,797	T>A	3	rs2986163
25,295,800	G>A	3	rs2986164
25,296,764	A>C	3	rs599792
25,297,476	A>G	3	rs1830962
25,298,410	G>C	3	rs1293267
25,301,905	T>G	5	rs28510210
25,304,945	A>T	6	rs28685153
25,307,040	G>C	7	rs3118453
25,311,520	G>A	7	rs2478028
25,311,722	T>G	7	rs796579065
25,320,257	A>C	8	rs28628791
25,329,839	A>T	10	rs28668998

[α]From the database of SNPs (Sherry et al., 2001).
SNPs were present as homozygous in all R_1R_1 samples, R_1r samples, and R_0r samples, and in 3 out of the 6 R_1R_2 samples that tested as hemizygous for the *RHD* gene by ddPCR. SNPs were heterozygous in the 6 homozygous R_1R_2 samples and the R_2R_z sample.

Table 3.10: Position of intronic variations determined by NGS and their corresponding nucleotide in R_2 and R_1, R_0, R_z *RHD* alleles in comparison to the reference sequence (hg38).

Position	rs no. [α]	Intron	hg38	$\mathbf{R_1, R_0, R_z}$	$\mathbf{R_2}$	Position	rs no. [α]	Intron	hg38	$\mathbf{R_1, R_0, R_z}$	$\mathbf{R_2}$
25,277,761	rs28661958	1	A	G	G	25,300,575	rs2986167	3	C	C	G
25,282,654	rs3866916	1	A	A	G	25,301,905	rs28510210	*5	T	G	T
25,284,544	rs2301153	1	G	C	G	25,304,945	rs28685153	*6	A	T	A
25,285,089	rs675072	2	G	G	A	25,305,164	rs28703207	*6	G	T	T
25,286,520	rs183024534	2	T	C	C	25,305,898	rs12126031	*6	A	A	G
25,286,601	-	2	T	A	A	25,307,040	rs3118453	7	G	C	G
25,286,605	-	2	A	T	T	25,307,714	rs2257611	7	G	G	A
25,286,674	-	2	C	T	T	25,308,306	rs28374144	7	T	C	C
25,286,732	-	2	A	G	G	25,308,317	rs28719684	7	T	C	C
25,287,909	rs28718098	2	C	C	G	25,308,325	rs71493569	7	G	A	A
25,290,908	rs28521909	3	T	C	C	25,308,326	rs71493569	7	C	T	T
25,290,915	rs28562109	3	G	A	A	25,308,403	rs1801096	7	C	T	T
25,292,953	rs28645510	3	G	A	G	25,308,845	rs2478025	7	G	G	C
25,295,072	rs372986392	3	G	G	A	25,311,520	rs2478028	7	G	A	G
25,295,317	rs2986157	3	G	A	G	25,311,722 [x]	rs796579065	7	T	G	A
25,295,354	rs2904840	3	C	C	T	25,316,058	rs28453868	7	A	G	G
25,295,489	rs190056379	3	C	C	T	25,316,269	rs2427767	7	A	A	G
25,295,708	rs182346769	3	G	G	A	25,319,292	rs28397158	8	T	C	C
25,295,731	rs201512625	3	A	A	G	25,320,257	rs28628791	8	A	C	A
25,295,739	rs200682399	3	G	G	A	25,320,442	rs3927482	8	T	T	G
25,295,753	rs143670081	3	A	A	G	25,321,858	rs28669938	8	T	T	C
25,295,797	rs2986163	3	T	A	T	25,322,588	rs28435180	9	A	G	G
25,295,800	rs2986164	3	G	A	G	25,323,393	rs77160738	9	C	C	T
25,295,850	rs28451966	3	A	G	G	25,323,618	rs201304363	9	G	G	C
25,296,764	rs599792	3	A	C	A	25,323,713	rs202154122	9	G	G	C
25,297,140	rs28786680	3	G	A	A	25,327,036	rs61777612	9	G	A	A
25,297,476	rs1830962	3	A	G	A	25,327,668	-	9	A	A	G
25,298,410	rs1293267	3	G	C	G	25,329,789	rs28654325	10	A	G	G
25,298,980	rs2904843	3	T	T	C	25,329,839	rs28668998	10	A	T	A

[α] From the database of SNPs (Sherry et al., 2001).
(-) Not found in the database of SNPs.
[x] SNP position shows three different nucleotides, T for reference (hg38), G for R_1, R_0, R_z haplotypes, and A for R_2 haplotype.
*Slight variation was detected in introns 4, 5, and 6 which could be due to their small sizes (427bp, 1635bp, 3137bp), respectively.

Table 3.11: Intronic SNPs and their reference SNP number present in 3 samples genotyped as *RHD*09.03.01 (DAR3.1)*.

Position	SNP	Intron	rs no. [α]
25,294,998	T>C	3	rs565983375
25,295,395	A>G	3	rs113859891
25,295,403	G>A	3	rs112427260
25,302,318	G>del-	5	rs568375574
25,303,820	A>C	6	rs595776
25,306,247	A>G	6	rs148021941
25,307,491	A>G	7	rs149085087
25,308,882	G>C	7	rs369127976
25,308,889	T>C	7	rs2986170
25,308,929	A>T	7	rs377167394
25,308,964	C>T	7	rs374258180
25,309,005	C>A	7	rs182746791
25,309,011	A>G	7	rs371128974
25,309,052	C>T	7	rs368386643
25,309,064	A>G	7	rs375347442
25,309,178	T>C	7	rs150952004
25,309,201	A>G	7	rs1342381686
25,309,243	A>G	7	rs1443666726
25,309,268	C>T	7	rs1424093742
25,309,320	G>A	7	rs1317071825
25,309,349	C>G	7	rs537011312
25,309,511	T>A	7	rs1233220393
25,309,531	A>T	7	rs1282313512
25,309,592	T>G	7	rs1472960480
25,309,699	C>T	7	rs2986171
25,309,702	G>T	7	rs1375423938
25,309,706	A>G	7	rs592372
25,309,775	C>T	7	rs592784
25,309,828	C>G	7	rs116689941
25,309,830	C>T	7	rs750012340
25,317,824	A>T	8	rs528996192
25,318,601	A>-del	8	rs200414156
25,318,602	A>-del	8	rs200414156
25,326,666	A>C	9	-

[α]From the database of SNPs (Sherry et al., 2001)
(-) Not found in the database of SNPs.

Table 3.12: Intronic SNPs and their reference SNP number present in samples genotyped as *RHD*09.03.01*, *RHD*09.01.02*, *RHD*03.01*, and *RHD*09.03*.

Position	SNP	Intron	rs no. [α]
25,297,129	A>G	3	rs565454189
25,297,131	C>T	3	rs192974801
25,298,827	G>A	3	rs111968158
25,302,081	G>A	5	rs145236797
25,304,459	A>G	6	rs28534644
25,308,085	T>C	7	rs376112497
25,310,078	C>T	7	rs187468762
25,313,241	A>T	7	rs111844516
25,316,554	G>A	7	rs536830264
25,320,143	C>A	8	rs151132067
25,320,272	A>G	8	rs115135456

[α]From the database of SNPs (Sherry et al., 2001).

Table 3.13: Rh haplotype corrected for samples where serologically predicted genotypes do not agree with NGS *RHD* genotyping results.

Lab no.	Rh serology[Y]	RhCcEe phenotype[Y]	*RHD* zygosity	*RHD* allele	Linked haplotype [▽]	Intronic SNPs pattern	Most probable haplotype [†]
004_07	R_1R_1	Ce	Discrepancy	*RHD*01W.01*	R_1 (DCe)	R_1, R_0, R_z	R_1R_1 (DCe/DCe)
004_14	R_1r	Ce	Homozygous	*RHD*01*	NA	R_1, R_0, R_z	R_1R_1 (DCe/DCe)
004_35	R_1R_2	CcEe	Hemizygous	*RHD*01W.02*	R_2 (DcE)	R_2	R_2r' (DcE/dCe)
004_36	R_1R_2	CcEe	Hemizygous	*RHD*01*	NA	R_2	R_2r' (DcE/dCe)
004_37	R_1R_2	CcEe	Hemizygous	*RHD*01*	NA	R_2	R_2r' (DcE/dCe)
004_38	R_1R_2	CcEe	Hemizygous	*RHD*01*	NA	R_1, R_0, R_z	R_1r'' (DCe/dcE)
004_39	R_1R_2	CcEe	Hemizygous	*RHD*01*	NA	R_1, R_0, R_z	R_1r'' (DCe/dcE)
004_40	R_1R_2	CcEe	Hemizygous	*RHD*01*	NA	R_1, R_0, R_z	R_1r'' (DCe/dcE)
004_42	R_2R_2	cE	Hemizygous	*RHD*01W.02*	R_2 (DcE)	R_2	R_2r'' (DcE/dcE)
004_88	$r''r$	cEe	Hemizygous	Novel variant 1	NA	R_2	R_2r (DcE/dce)
004_94	No data	No data	Hemizygous	*RHD*01W.01*	R_1 (DCe)	R_1, R_0, R_z	R_1 (DCe/-)
004_95	No data	No data	Hemizygous	*RHD*01W.01*	R_1 (DCe)	R_1, R_0, R_z	R_1 (DCe/-)
004_96	R_1r	Cce	Hemizygous	*RHD*03.01*	R_0 (Dce)	R_1, R_0, R_z	R_1r (DCe/dce) double check
004_97	No data	No data	Hemizygous	*RHD*09.03.01*	R_0 (Dce)	R_1, R_0, R_z	R_0 (Dce/-)
004_98	No data	No data	Hemizygous	*RHD*25*	R_1 (DCe)	R_1, R_0, R_z	R_1 (DCe/-)
004_99	No data	No data	Hemizygous	*RHD*04.04*	R_1 (DCe)	R_1, R_0, R_z	R_1 (DCe/-)
004_100	No data	No data	Hemizygous	*RHD*04.04*	R_1 (DCe)	R_1, R_0, R_z	R_1 (DCe/-)
004_101	R_1r	Cce	Hemizygous	*RHD*06.01*	R_2 (DcE)	R_1	R_1r (DCe/dce)
004_102	R_1r	Cce	Hemizygous	Undetermined	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_103	R_1r	Cce	Hemizygous	Undetermined	NA	R_1, R_0, R_z	R_2 (DcE/dce)
004_104	R_2r	cEe	Hemizygous	Undetermined	NA	R_2	R_2 (DcE/dce)
004_105	R_2r	cEe	Hemizygous	Undetermined	NA	R_2	R_2r (DcE/dce)
004_106	No data	No data	Hemizygous	Undetermined	NA	R_2	R_2 (DcE/-)
004_107	No data	No data	Hemizygous	*RHD*01W.03*	R_1 (DCe)	R_1, R_0, R_z	R_1 (DCe/-)
004_108	R_1r	Cce	Hemizygous	*RHD*07.01*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_109	R_1r	Cce	Hemizygous	*RHD*07.01*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_110	R_1r	Cce	Hemizygous	*RHD*07.01*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_111	R_1r	Cce	Hemizygous	*RHD*17.01*	multiple	R_1, R_0, R_z	R_1r
004_112	No data	No data	Hemizygous	Novel variant 2	NA	R_1, R_0, R_z	-
004_113	No data	No data	Hemizygous	*RHD*01*	Not reported	R_1, R_0, R_z	-
004_114	No data	No data	Homozygous	*RHD*01*	Not reported	R_1, R_0, R_z	-

continued ...

Lab no.	Rh serology[ϒ]	RhCcEe phenotype[ϒ]	*RHD* zygosity	*RHD* allele	Linked haplotype [▽]	Intronic SNPs pattern	Most probable haplotype [†]
004_115	No data	No data	Hemizygous	*RHD*01*	Not reported	R_1, R_0, R_z	-
004_116	No data	No data	Homozygous	*RHD*01*	Not reported	R_1, R_0, R_z	-
004_117	D-	Cce	Hemizygous	*RHD*01EL.01*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_118	D-	Cce	Hemizygous	*RHD*01EL.01*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_119	D-	Cce	Hemizygous	Novel variant 3	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_120	D-	Cce	Hemizygous	*RHD*01N.61*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_121	D-	Cce	Hemizygous	*RHD*01N.33*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_122	D-	Cce	Hemizygous	*RHD*11*	R_0 (Dce)	R_1, R_0, R_z	R_0r' (Dce/dCe)
004_123	D-	Cce	Hemizygous	*RHD*15*	R_2 (DcE)	R_1, R_0, R_z	R_1r (DCe/dce)
004_124	Weak D+	Cce	Hemizygous	*RHD*01W.72*	Not reported	R_1, R_0, R_z	R_1r (DCe/dce)
004_125	Weak D+	Cce	Homozygous	Undetermined	NA	R_1, R_0, R_z	R_1R_0 (DCe/Dce)
004_126	D-	Cce	Hemizygous	*RHD*01EL.04*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_127	D-	Cce	Hemizygous	Novel variant 10	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_128	D-	Cce	Hemizygous	Novel variant 10	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_129	D-	Cce	Hemizygous	Novel variant 10	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_130	D-	Cce	Hemizygous	Novel variant 10	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_131	D-	Cce	Hemizygous	Novel variant 5	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_132	Weak D+	Cce	Hemizygous	Novel variant 8	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_133	D-	Cce	Hemizygous	Novel variant 9	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_134	D-	Cce	Hemizygous	Novel variant 9	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_135	D-	Cce	Hemizygous	Novel variant 9	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_136	D-	Cce	Hemizygous	*RHD*01EL.08*	R_1 (DCe)	R_1, R_0, R_z	R_1r (DCe/dce)
004_137	D-	Cce	Hemizygous	Undetermined	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_138	D-	Cce	Hemizygous	Undetermined	NA	R_1, R_0, R_z	R_1r (DCe/dce)
004_139	D-	Ce	Hemizygous	Novel variant 10	NA	R_1, R_0, R_z	R_1r (DCe/dCe)
004_140	Weak D+	Ce	Hemizygous	*RHD*01W.03*	Not reported	R_1, R_0, R_z	R_1r' (DCe/dCe)
004_141	Weak D+	Ce	Hemizygous	Novel variant 8	NA	R_1, R_0, R_z	R_1r' (DCe/dCe)
004_142	D-	CcEe	Hemizygous	*RHD*01W.02*	R_2 (DcE)	R_2	R_2r' (DcE/dCe)
004_143	D-	CcEe	Hemizygous	*RHD*01W.02*	R_2 (DcE)	R_2	R_2r' (DcE/dCe)
004_144	D-	cEe	Hemizygous	Novel variant 4	NA	R_2	R_2r (DcE/dce)
004_145	D-	cEe	Hemizygous	Novel variant 6	NA	R_2	R_2r (DcE/dce)
004_146	D-	cEe	Hemizygous	*RHD*15*	R_2 (DcE)	R_2	R_2r' (DcE/dCe)
004_147	Weak D+	cEe	Hemizygous	Novel variant 7	NA	R_2	R_2r (DcE/dce)
004_148	D+	ce	Hemizygous	*RHD*09.03.01*	R_0 (Dce)	R_1, R_0, R_z	R_0r (Dce/dce)

continued ...

Lab no.	Rh serology$^{\Upsilon}$	RhCcEe phenotype$^{\Upsilon}$	*RHD* zygosity	*RHD* allele	Linked haplotype $^{\nabla}$	Intronic SNPs pattern	Most probable haplotype †
004_149	D-	ce	Hemizygous	*RHD*01N.35*	Not reported	R_1, R_0, R_z	R_0r (Dce/dce)
004_150	D-	ce	Hemizygous	Undetermined	NA	R_1, R_0, R_z	R_0r (Dce/dce)
004_151	D-	ce	Hemizygous	No designation assigned*	Not reported	R_1, R_0, R_z	R_0r (Dce/dce)

Blue indicates samples were serologically phenotyped as RhD negative.
NA= Not applicable.
$^{\Upsilon}$ As supplied by the NHSBT, Bristol, United Kingdom, and the Finnish Red Cross Blood Service, Helsinki, Finland.
- Rh probable haplotype was not determined due to the lack of Rh serological results.
$^{\nabla}$ Rh haplotypes linked to the variant *RHD* allele (Wagner and Flegel, 2014).
† Probable Rh haplotypes are predicted by the results of Rh serology phenotype, *RHD* zygosity by ddPCR, *RHD* intronic SNPs pattern, and linked haplotype to the variant allele in literature if applicable.
*As found on RhesusBase (Wagner and Flegel, 2014).
**Allele linked to DcE but sample phenotyped as Cce and present intronic R_1, R_0, R_z SNPs pattern, therefore most probable haplotypes are DCe or Dce.

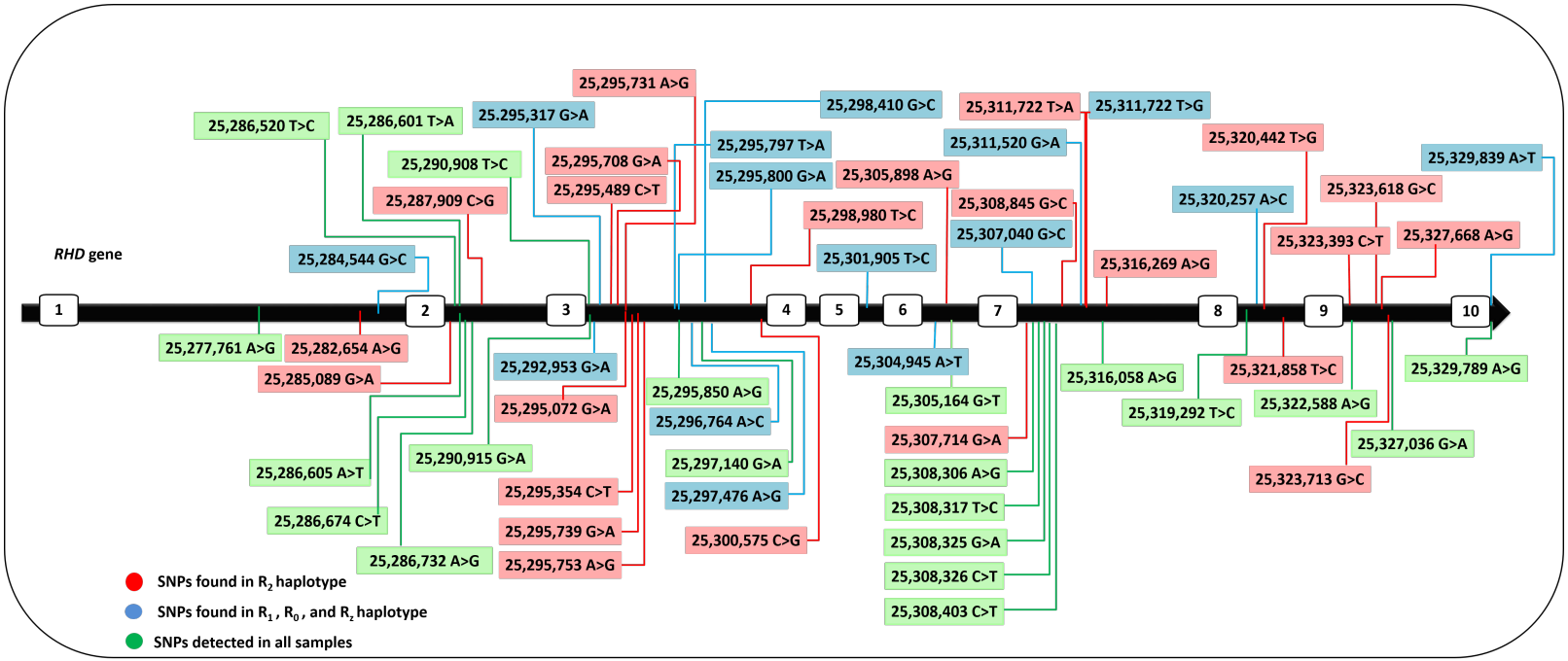


Figure 3.13: *RHD* intronic SNPs mapped to the *RHD* gene.

3.3.6 *RHAG* NGS

The *RHAG* gene was sequenced for two samples (004_14 and 004_74), that were serologically phenotyped as weak D but no *RHD* mutations were detected by NGS that would explain the weak D reactivity. In sample (004_14), two heterozygous exon 6 mutations were detected including c.808G>A (p.Val270Ile) which encodes for the *RHAG*04* allele (Figure 3.14), and c.861G>A (silent). Sample (004_74) showed a wild type *RHAG*01* allele predicting no amino acid changes.

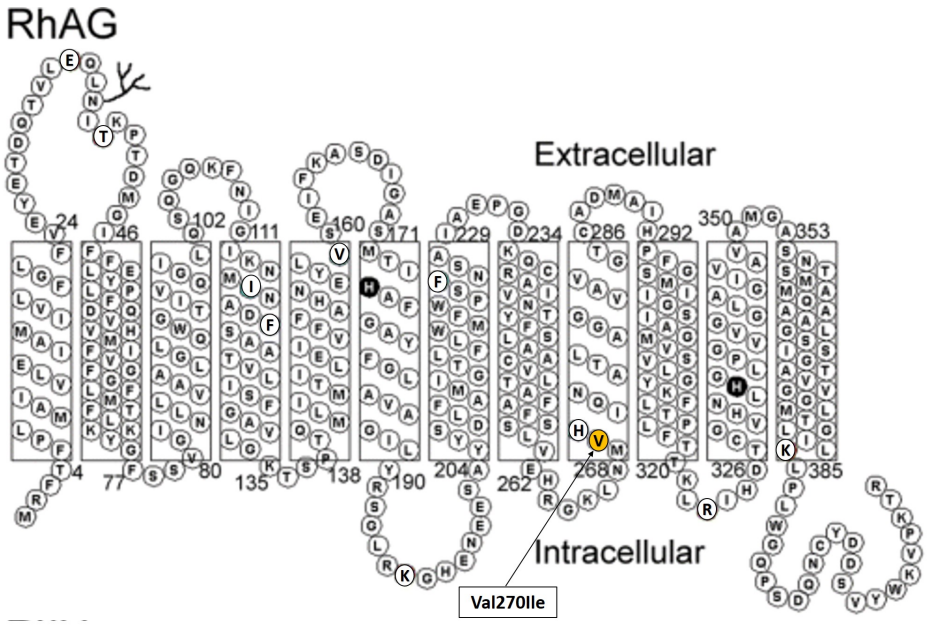


Figure 3.14: RhAG intracellular domains with amino acids illustrated in circles. Amino acid change Val270Ile indicated with an arrow was detected in one R_1r sample, serologically typed as weak D, with no amino acid changes predicted from the *RHD* NGS data. Image courtesy of Conroy et al. (2005).

3.3.7 *RHD* Sequencing Using MinION™

3.3.7.1 MinION™ Sequence Quality

Thirteen samples were genotyped for the *RHD* gene using MinION™, utilising the same LR-PCR amplicons used in the Ion PGM™ sequencing. Sequencing library was prepared and data was analysed as described in sections 2.12 and 2.14.2, respectively. Fast5 files generated from the raw files were subjected to sequence quality analysis using EPI2ME software (Oxford Nanopore Technologies, United Kingdom). Figure 3.15 shows an average quality PHRED-score of 11, and reads length mode of 10,450 bp. A PHRED score of 10 indicates a 1/10 probability of an incorrect base or a 90% accuracy.

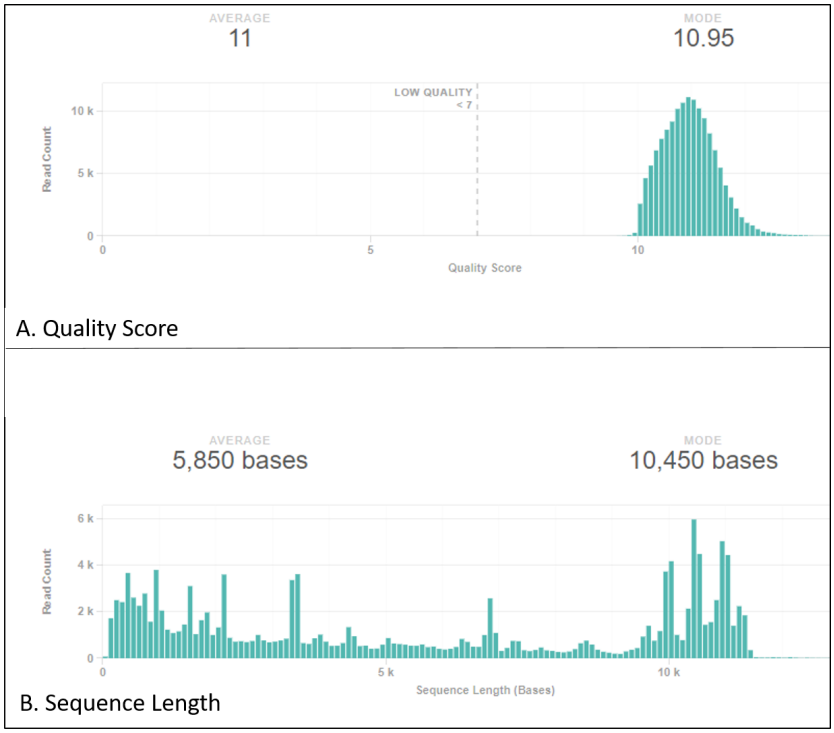


Figure 3.15: Sequence quality analysis performed on MinION™ data using EPI2ME software. A. Quality score (x axis) and number of reads detected at that quality (y axis). Average PHRED score of about 11 was observed. B. Distribution of reads sequence lengths in bp (x axis) and number of reads of a specific length (y axis).

3.3.7.2 *RHD* Genotyping Using MinION™

gDNA samples (n=13) (Table 3.14) in which 12 samples (004_122, 004_123, 004_125, 004_130, 004_131, 004_133, 004_134, 004_143, 004_144, 004_145, 004_146, 004_147) were either serologically phenotyped as RhD negative or weak D (Finnish samples), and 1 R_2R_z sample (004_74) (blood donor sample) phenotyped by serology as weak D. Samples were sequenced using the MinION™ for the *RHD* gene. Data was analysed as described in

section 2.14.2, and mapped to the *RHD* hg38 reference sequence. Mapped data to the hg38 reference sequence were visualised using IGV software, (Figure 3.16). As noted previously, the *RHD* human reference sequence in the hg38 encodes a variant *RHD* allele *RHD*DAU0* encoded by c.1136C>T (p.Thr379Met) in exon 9. Therefore, all 13 samples showed a homozygous SNP in exon 9 c.1136T>C (Met379Thr), (Figure 3.9).

Variants calling was performed on the 13 samples sequenced and a variant track was generated for each sample, which was then compared to the variant track generated from the Ion PGM™ sequencing data for the same samples. Figure 3.17 shows mapped reads from Ion PGM™ and MinION™ aligned against each other. All exonic SNPs detected in the 13 samples agreed with the ones detected from data generated by Ion PGM™ which are listed in Table 3.14. Intronic changes detected were also compared and agreed to the SNPs detected by Ion PGM™ and all Rh haplotype specific SNPs listed in Tables 3.7, 3.8, 3.9 except for 6 SNPs. These 6 intronic SNPs were expected to be specific to the *RHD* reference sequence hg38 *RHD*DAU0* allele, which are highlighted in grey in Table 3.7.

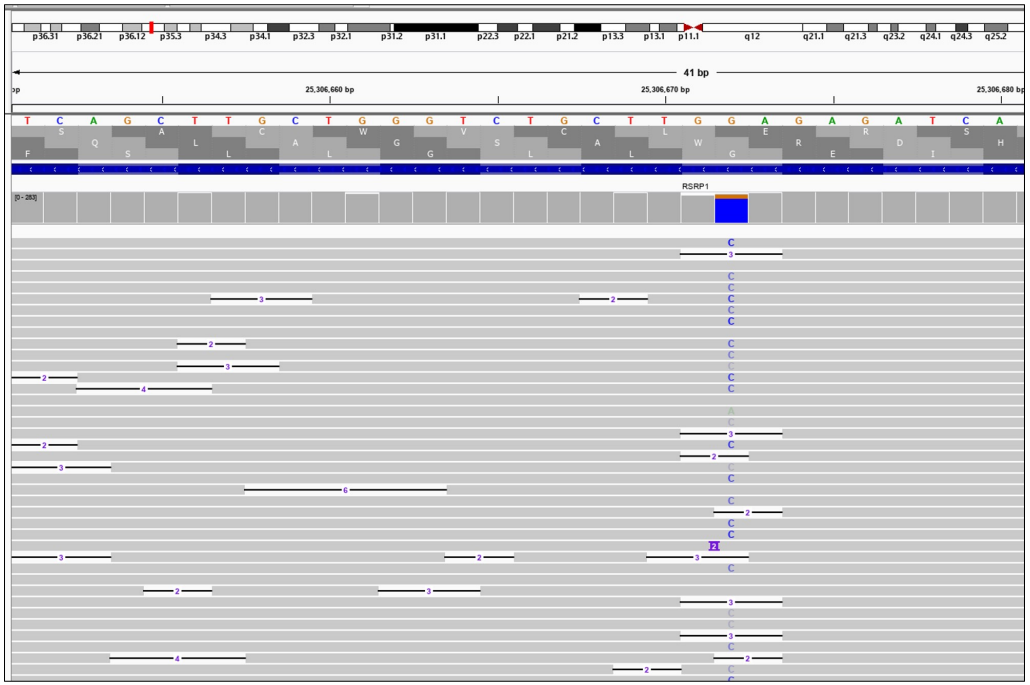


Figure 3.16: MinION™ reads, from a single representative sample, mapped to the hg38 reference sequence visualised using IGV software. Sample shows a SNP (blue) in exon 7 c.1016G>C that predicts amino acid change p.Gly339Ala.

Figure 3.17: Ion PGM™ reads compared to the MinION™ reads using CLC Workbench 10. Ion PGM™ reads (A) and MinION™ reads (B) of the same sample (004_144) mapped to the hg38 reference sequence in an alignment. Two exonic changes in exon 9 c.1154G>C (p.Gly385Ala) (blue highlight) and c.1163T>G (p.Leu388Arg) (yellow highlight) detected from both sets of data.

Table 3.14: Samples sequenced using MinION™

Lab no.	**RhD phenotype**	***RHD* zygosity**	**Exonic SNPs***	***RHD* allele**
004_74	weak D+	Homozygous	-	*RHD*01*
004_122	Negative	Hemizygous	885G>T	*RHD*11*
004_123	Negative	Hemizygous	845G>A	*RHD*15*
004_125	weak D+	Homozygous	48G>C, 602C>G, 667T>G	Undetermined
004_130	Negative	Hemizygous	829G>A	Novel variant 10
004_131	Negative	Hemizygous	784delC	Novel variant 5
004_133	Negative	Hemizygous	421delG, 422T>A	Novel variant 9
004_134	Negative	Hemizygous	421delG, 422T>A	Novel variant 9
004_143	Negative	Hemizygous	1154G>C	*RHD*01W.02*
004_144	Negative	Hemizygous	1154G>C, 1163T>G	Novel variant 4
004_145	Negative	Hemizygous	519C>G	Novel variant 6
004_146	Negative	Hemizygous	845G>A	*RHD*15*
004_147	weak D+	Hemizygous	1016G>C	Novel variant 7

*SNPs detected by MinION™ support the results detected by Ion PGM™.

3.4 Discussion

3.4.1 ddPCR Discrepant Results Characterised by *RHD* NGS

ddPCR was used to test for two targets in the *RHD* gene against the reference gene *AGO1* on chromosome 1. ddPCR has demonstrated high sensitivity when used as a detection method for *RHD* genotyping (Sillence et al., 2015, 2017). All samples included in the sequencing run demonstrated compatible zygosity results with serologically predicted genotype except for samples listed in Table 3.2. Eight samples showed incompatible results with predicted genotype by serological testing which include: one $\mathrm{R_1r}$ sample (004_14) which showed the presence of a homozygous *RHD* gene, six $\mathrm{R_1R_2}$ samples (004_35, 004_36, 004_37, 004_38, 004_39, 004_40) which showed the presence of a hemizygous *RHD* gene, and one $\mathrm{R_2R_2}$ sample (004_42) showed as hemizygous for the *RHD* gene.

One $\mathrm{R_1R_1}$ sample (004_07) showed discrepancy between the *RHD5/AGO1* ratio (0.57) and *RHD7/AGO1* ratio (1.0). This discrepancy between hemizygous *RHD5* and homozygous *RHD7* indicating that one of the *RHD* alleles has a deletion in exon 5. This gene deletion could not be detected through NGS due to the presence of a wild type copy of the other *RHD* allele.

The six $\mathrm{R_1R_2}$ (DCe/DcE) samples which showed only one copy of the *RHD* gene (hemizygous), implied that the samples had incorrect serolog-

ically predicted genotypes by serology findings based on the probability of the gene in the population, in these cases the genotypes are in fact less frequent or occurring with a lower probability.

These lower frequency genotypes are either $\mathrm{R_1r''}$ (DCe/dcE), $\mathrm{R_zr}$ (DCE/dce), $\mathrm{R_2r'}$ (DcE/dCe), or $\mathrm{R_0r^y}$ (Dce/dCE) based on their *RHD* zygosity results. All these Rh genotypes could be inappropriately assigned by serology as $\mathrm{R_1R_2}$ (DCe/DcE) due to gene frequencies in the population. Three of the $\mathrm{R_1R_2}$ (004_35, 004_36, 004_37) samples have the intronic SNPs suspected to be linked to the $\mathrm{R_2}$ haplotype and are missing all the other intronic SNPs that are linked to the $\mathrm{R_1, R_0, R_z}$ haplotypes. Sample 004_35 was genotyped as weak D type 2, and due to the link between the $\mathrm{R_2}$ haplotype and weak D type 2, this sample could only be $\mathrm{R_2r'}$ (DcE/dCe). The other two samples could also be genotyped as $\mathrm{R_2r'}$ (DcE/dCe) as inferred by their intronic SNPs pattern. The correct genotype of the other three hemizygous $\mathrm{R_1R_2}$ samples (004_38, 004_39, 004_40) missing the $\mathrm{R_2}$ specific SNPs could be either $\mathrm{R_1r''}$ (DCe/dcE), $\mathrm{R_zr}$ (DCE/dce), or $\mathrm{R_0r^y}$ (Dce/dCE). Considering the frequency of these alleles (Daniels, 2013c) in the population, in which $\mathrm{R_1r''}$ is 1%, $\mathrm{R_zr}$ is 0.19%, and $\mathrm{R_0r^y}$ is less than 0.01%, it is very likely for these samples to be $\mathrm{R_1r''}$ (DCe/dcE). Based on our zygosity results, the frequency of $\mathrm{R_1r''}$ seems to be higher than anticipated (Daniels, 2013c) in the population. Definitive genotypes for these samples could be confirmed by sequencing the *RHCE* gene, Chapter 4.

The $\mathrm{R_1r}$ (004_14) sample that was homozygous *RHD* by ddPCR showed $\mathrm{R_1/R_0/R_z}$ related SNPs and was missing all $\mathrm{R_2}$ related SNPs which suggests that correct genotype could be $\mathrm{R_1R_0}$ (DCe/Dce). The hemizygous $\mathrm{R_2R_2}$ sample (004_42) was genotyped as weak D type 2 and showed $\mathrm{R_2}$ specific SNPs, therefore, its correct genotype could only be $\mathrm{R_2r''}$ (DcE/dcE).

Samples genotyped as weak D type 2 (004_142 and 004_143) showed intronic SNPs specific to $\mathrm{R_2}$ haplotype and were originally assigned their haplotype by serology as r''r. Weak D type 2 is linked to $\mathrm{R_2}$ (DcE) (Wagner et al., 1999) which means the only correct haplotype for these two samples is $\mathrm{R_2r'}$. Sample (004_124) genotyped as *RHD*01W.72* was originally assigned by serology Rh haplotype as r'r. This sample showed intronic SNPs specific for $\mathrm{R_1, R_0, R_z}$ haplotypes which suggest its correct haplotype to be either $\mathrm{R_1r}$ or $\mathrm{R_0r'}$. Sample (004_140) genotyped as *RHD*01W.03* was serologically assigned Rh haplotype as r'r'. The *RHD*01W.03* allele is linked to the $\mathrm{R_1}$ (DCe) haplotype (Wagner et al., 1999) which indicates that the only probable correct haplotype for this sample is $\mathrm{R_1r'}$. Corrected Rh haplotype for all discrepant samples are listed in Table 3.13.

3.4.2 *RHD* Reference Sequences

We have established a methodology to fully sequence the *RHD* gene including introns and all exons which can be used to study the different *RHD* alleles in the population to establish reference *RHD* allele sequences

and identify novel variants. We sequenced the *RHD* gene in samples that were confirmed to be hemizygous (one copy) *RHD* samples by ddPCR, and compared those sequences with homozygous (two copy) *RHD* genes in samples confirmed to be as such by ddPCR.

Through analysing intronic SNPs, two different haplotype specific patterns were identified in which one is specific for R_1, R_0, R_z haplotypes and the other is R_2 haplotype specific. Based on that, two reference sequences were established and submitted to Genbank and registered with accession numbers MG944308 and MG944309 for the R_1, R_0, R_z haplotypes and the R_2 haplotype, respectively.

3.4.3 *RHD* Genotyping

Data analysis of the (n= 133) samples showed that 48 samples carried the *RHD* wild type allele *RHD*01*. Other samples (n= 85) showed *RHD* variant alleles, in which 59 samples linked to known *RHD* variant alleles, and 17 samples showed novel *RHD* variant alleles. In 9 samples the *RHD* genotyping results were inconclusive.

3.4.3.1 *RHD* Known Variants

Known variants found in 59 samples included weak, partial, *DEL* and null *RHD* alleles. Four different weak D alleles were detected in 24 samples in which: 16 samples were genotyped as *RHD*01W.01*, 14 samples were genotyped as *RHD*01W.02*, 3 samples were genotyped as *RHD*01W.03*, and 1 sample genotyped as *RHD*01W.72*. These results agree with the

common weak D alleles in Caucasian (weak D types 1 and 2) (Daniels, 2013b) that are associated with the most common Rh haplotypes R_1 (DCe) and R_2 (DcE) respectively (Wagner et al., 1999; Flegel, 2007a).

Eleven partial weak D alleles were found in 17 samples in which 1 sample was genotyped as *RHD*06.01* (DVI). DVI is the most common partial D phenotype which results from a hybrid *RHD-RHCE-RHD* gene leading to the loss of multiple D antigen epitopes including epD 1-4, 7-22, 26-29 on the 30-epitope model (Avent et al., 1997). The hybrid allele has originated from a conversion event where *RHD* exons 4 and 5 are replaced by *RHCE* exons 4 and 5. Individuals with a DVI allele could be phenotyped by serology (using polyclonal antisera) as RhD positive which places them at risk of developing RhD antibodies to missing D epitopes when exposed to normal RhD antigen. Pregnant females phenotyped as DVI, with an RhD positive fetus, need to receive prophylactic anti-D to prevent the development of Rh antibodies which are clinically significant and can lead to HTRs and HDFN (Avent et al., 1997), discussed in sections 1.8.1 and 1.8.2. Three samples were genotyped as *RHD*07.01* encoded by c.329C>T in exon 2 predicting amino acid Proline substituting Leucine at position 110 which results in the loss of RhD antigen epitope epD8 (Rouillac et al., 1995).

In three samples, the *RHD* allele was identified as weak partial *RHD* allele in which sample was genotyped as *RHD*11* and two samples were geno-

typed as *RHD*15*. Two samples genotyped as weak D type 15 (*RHD*15*) express a very weak level of antigen D with qualitative changes to its epitopes which mostly fails to react with RhD antisera which explains the D negative serology results (Körmöczi et al., 2005). Partial weak D could be mis-phenotyped by serology as RhD negative due to the fact that some of the RhD epitopes are either altered or completely absent which might affect antigen-antibody binding sites leading to false negative phenotyping results. Incorrectly phenotyped RhD negative donors impose a serious risk of immunisation on RhD negative patients (Avent et al., 1997). Cases of RhD alloantibodies in individuals with weak D type 15 phenotype have been reported (Wagner et al., 2000). Therefore, patients genotyped as *RHD*15* should be transfused with RhD negative blood to prevent possible sensitisation.

One sample was genotyped as partial *RHD*25* allele (DNB) which is the most common partial D allele in whites in Central Europe (Wagner et al., 2002). The DNB allele present a normal positive D in serological phenotyping, however, cases of immunisation (allo-anti-D) in DNB individuals have been reported after immunisation events. DNB is the underlying allele in which the majority of partial D immunisation occur in whites (Wagner et al., 2002).

Two samples were genotyped as *RHD*04.04* (DIV.4) which results from a gene conversion event where part of the *RHD* exon 7 is substituted by the

equivalent region from the *RHCE* gene (Körmöczi et al., 2005; von Zabern et al., 2013). One sample was genotyped as *RHD*17.01* (DFR-1), and one as *RHD*17.02* (DFR-2), which both results from a hybrid *RHD-RHCE-RHD* gene in exon 4 (Rouillac et al., 1995; von Zabern and Flegel, 2007). Substitution in the *RHD* gene results in the loss of some the RhD epitopes leading to partial expression of the D antigen.

Three DEL alleles were detected in 4 samples in which 2 samples were genotyped as *RHD*01EL.01*, 1 sample was genotyped as *RHD*01EL.04*, and 1 sample was genotyped as *RHD*01EL.08*, identified firstly in German blood donors (Wagner et al., 2001). *DEL* phenotype expresses very weak D antigen that could only be detected by adsorption, elution techniques, and molecular genotyping (Kwon et al., 2017), hence, samples are phenotyped by serology as Rh D-negative. Flegel et al. (2009) recommends that blood units genotyped as DEL should be stored with RhD positive blood units and only transfused to RhD positive patients.

It was believed that very weak expression of antigen D most likely will not elicit an immune response, but cases of antigen D allo-sensitisation caused by *DEL* RBCs transfusion to Rh D-negative patients have been reported (Wagner et al., 2005; Yasuda et al., 2005; Shao et al., 2012; Yang et al., 2015). Shao et al. (2012) recommended that Rh D-negative patients should be transfused with only Rh D-negative RBCs and emphasises the importance of correctly genotyping apparently Rh D-negative donors to prevent

anti-D development in Rh D-negative patients.

Three negative *RHD* alleles were detected in 3 samples, in which 1 sample was genotyped as *RHD*01N.35*, 1 sample was genotyped as *RHD*01N.61*, and 1 sample was genotyped as *RHD*01N.33*. The *RHD*01N.35* was first described by Maréchal et al. (2007) in the French population. The other two null alleles *RHD*01N.35* and *RHD*01N.61* were first identified in Germans (Flegel et al., 2009), in which mutations in these alleles lead to premature stop codon with truncated non-functional protein. Individuals carrying null alleles should be considered RhD negative and only transfused with RhD negative blood to prevent immunisation (Maréchal et al., 2007).

In sample (004_41), mutation c.1154G>C in exon 9 was detected which encodes *RHD*01W.02* that predicts amino acid change p.Gly385Ala. The frequency of the reads containing each mutation were approximately 0.5 (i.e. heterozygous SNP), which indicates that one allele is a variant allele *RHD*01W.02* and the other is a wild type *RHD* allele. Wild type *RHD* allele produce a complete RhD protein that will mask the presence of a weak *RHD* allele, which makes it impossible to detect the weak D in this sample by serology as it would be masked by the normal RhD protein.

In samples genotyped as weak D, four samples were serologically phenotyped as either D negative or very weak D. These samples included: sample (004_124) genotyped as *RHD*01W.72*, sample (004_140) genotyped

as *RHD*01W.03*, and two samples (004_142 and 004_143) genotyped as *RHD*01W.02*. Weak D phenotype should express D antigen at a detectable level by serology (Sandler et al., 2017). It has been reported that the presence of the *RHCE*C* expression has a suppressive effect on the RhD antigen (Szymanski, 1987; Araszkiewicz and Szymanski, 1989; Daniels, 2013c).

All four samples were phenotyped by serology as Rh C-positive. In these samples, the D antigen expression was weakened by the exonic SNPs in the *RHD* gene and the RhC suppressive effect on the D antigen. The C suppressive effect on antigen D can cause reduction of the D antigen density and lead to a *DEL* phenotype (Szymanski, 1987; Daniels, 2013c), which makes it challenging to detect the D antigen by routine serology. Probable correct Rh haplotypes were hypothesised for these samples based on their Rh serological phenotype, *RHD* zygosity results, *RHD* allele identified by NGS and intronic SNP patterns detected (Table 3.13).

To compare the C suppressive effect on antigen D expression in different samples, four Finnish samples (004_124, 004_140, 004_142, 004_143) were compared to other blood donor samples of the same *RHD* allele. The C suppressive effect on D antigen expression, known as Ceppellini effect, was first explained in an extensive family phenotyping analysis that showed weak D reactivity may occur without genetic mutations (Ceppellini et al., 1955). Ceppellini et al. (1955) noted that DCe/dce genotype produces normal D expression when c is in *trans* (i.e on the opposite hap-

lotype), and DCe/dCe produces weak D expression when C in in *trans*. Different studies confirmed that the expression of C antigen reduces the D expression, although mechanism is not very clear (McGann et al., 2012). McGann et al. (2012) suggest that the *RHCE* gene may directly suppress the *RHD* gene transcription, or may influence post-translational modifications. It is also possible that the presence of both C and D antigens lead to steric hindrance on the cell membrane leading to reduction of D antigen expression (McGann et al., 2012).

Samples (004_140 and 004_28) were genotyped as *RHD*01W.03* and were both serologically phenotyped as RhC-positive; however, sample (004_140) shows two copies of RHCE*C (*RHCE*Ce*) in *cis* and *trans*; on the contrary, the blood donor sample shows only one copy of the RHCE*C with serologically predicted haplotype $\mathrm{R_1r}$ (DCe/dce) (c in *trans*). The blood donor sample may not be affected by Ceppellini effect as C is in *trans* (i.e on the same haplotype). However, Araszkiewicz and Szymanski (1989) in a D antigen quantitative study found that the lowest number of D antigens (285 per cell) was expressed in people with C in *cis* and *trans* (C homozygous) which may be the reason of very weak D reactivity in sample (004_140). This may indicate that *RHCE*C* presence in *cis* and *trans* (2 copies of C) could severely lower the RhD density on the RBCs making correct serological phenotyping challenging (Araszkiewicz and Szymanski, 1989).

Four samples were phenotyped by serology as D negative but genotyped as weak partial D allele. Variants include: 1 sample genotyped as *RHD*11 (RHD*weak partial 11)*, 2 samples genotyped *RHD*15 (RHD*DNB)*, 1 sample genotyped as *RHD*09.03.01 (RHD*DAR3.01)*, 1 sample genotyped as *RHD*09.01.02*, and 1 sample genotyped as *RHD*03.01*. Weak and partial D occurs due to multiple changes in the RhD protein that have both quantitative and qualitative effects on the protein expression (Wagner et al., 1999, 2000; Rizzo et al., 2012). These changes could make it challenging to correctly phenotype the D antigen as they alter the D antigen epitopes and subsequently the antibody binding ability to bind epitopes. According to Rizzo et al. (2012) and Flegel (2007a), patients with partial D phenotype should be considered RhD negative and transfused with RhD negative blood due to the risk of antibody production to missing or altered epitopes.

One ISBT sample (004_101) was serologically phenotyped as $\mathrm{R_1r}$ and genotyped previously as *DVa* where *RHCE* exon 4 insertion is observed. However, variants detected from NGS data showed exonic SNPs that indicate the *RHD* exon 4 and 5 were replaced with the *RHCE* ones, which encode the *RHD*06.01* allele. This variant allele is linked to the $\mathrm{R_2}$ haplotype Avent et al. (1997) but this sample shows that the hybrid allele also expressed in $\mathrm{R_1}$ haplotype. This allele when expressed in $\mathrm{R_1}$ (DCe) haplotype lacks exon 5 SNP c.676G>C (p.Ala226Pro) that encodes the RhE anti-

gen in the *RHCE* gene. This support the finding that *RHD*06.01* variant allele occurs in R_1 and R_2 haplotypes.

Five ISBT samples (004_102- 004_106), genotyped previously by ISBT (1996)) as *DVI type I*, very weakly amplified *RHD* amplicon 4 that covers the region from intron 4 to intron 7. These samples, when tested for *RHD* gene zygosity using ddPCR, showed ratios of ≈ 0.5 and ≈ 0 for *RHD7/ AGO1* and *RHD5/ AGO1* assays respectively, which indicate the presence of a hemizygous *RHD* gene with a deletion of exon 5. The *DVI type I* allele is a hybrid *RHD* gene, expresses partial RhD protein, where exons 4 and 5 are replaced with the *RHCE* gene correlated exons, *RHD-RHCE(4-5)-RHD*, hence the negative results for *RHD* exon 5 assay. However, NGS data for these samples showed SNPs associated with *RHCE(4-5)* insertion in the *RHD* gene and breakage of coverage in exon 6 indicating a deletion. Deletion could be confirmed using alternative methods such as exome sequencing. It is not clear how these samples were mis-genotyped previously as *DVI type I*, but it is clear that the deletion of exon 6 has occurred on *DVI type I* allele backbone. All 5 samples were phenotyped by serology as RhD positive, therefore, it is expected that the deletion of exon 6 did not completely disturb the RhD protein and a partial expression is expected.

One blood donor (004_88) sample and 28 Finnish samples were serologically phenotyped as D negative (Table 3.13) but tested *RHD* hemizygous using dPCR. In 3 samples out of the 28 Finnish samples, *RHD* genotyping

results were inconclusive, discussed in more detail in section 3.4.4. Out of the remaining 26 samples, 13 RhD negative samples carried novel *RHD* alleles encoded by novel mutations discussed in section 3.4.3.2.

In one sample (004_151) mutation c.395_396dupGG was detected in exon 3 that predicts p.Lys133fs frame shift. This has been reported in the literature by Karnot et al. (2016) but no designation was assigned by the ISBT as found in RhesusBase (Wagner and Flegel, 2014). The insertion most likely causes a frame shift and a truncated non-functional protein that causes a null phenotype. This is consistent with the negative serology results of this sample.

3.4.3.2 *RHD* Novel Alleles

In 18 samples, detected mutations did not link to known *RHD* alleles and these were considered novel *RHD* variants.The 18 samples included 16 serologically D negative Finnish pregnant female samples with undetermined variant *RHD* allele, 1 serologically D negative blood donor sample, and 1 ISBT sample with unknown Rh serology results. All 18 samples tested as hemizygous *RHD* gene on dPCR. Five variants, were considered Rh D-negative alleles in which exonic missense mutations led to a frame shift or a stop codon.

Sample (004_119), presented exon 1 deletion c.49delG predicting frame shift change at p.Ala17fs on *RHD*01W.03* backbone that is encoded by

exon 1 c.8C>G (p.Ser3Cys). Frame shift prevents translation and protein synthesis, which would explain the D-negative phenotype by serology. In three samples two exon 3 mutations were detected c.421delG and c.422T>A that predict amino acid changes p.Val141fs and p.Val141Glu, respectively (Novel variant 9). The novel mutations c.421delG, c.784delC, c.519C>G, and c.399delG are expected to cause null phenotypes, which would be consistent with the observed D-negative results in the RhD phenotyping of the samples.

One hemizygous *RHD* ISBT sample (004_112) showed mutations associated with *RHD*07.01* and *RHD*17.02* variant alleles which include c.329T>C (p.Leu110Pro) in exon 2 and c.505A>C (p.Met169Leu), c.509T>G (p.Met170Arg), c.514A>T (p.Ile172Phe), c.544T>A (p.Ser182Thr), c.577G>A (p.Glu193Lys), c.594A>T (p.Lys198Asn), and c.602C>G (p.Thr201Arg) in exon 4. This sample encodes a novel hybrid *RHD* allele of the two variants *RHD*07.01* and *RHD*17.02* alleles, (Novel variant 2).

In sample (004_144), a novel mutation was identified on a backbone of *RHD*01W.2*. Sample (004_144) showed mutation in exon 9 c.1154G>C predicting amino acid change p.Gly385Ala that encodes *RHD*01W.02*, in addition to another exon 9 mutation c.1163T>G predicting amino acid change p.Leu388Arg (Novel Variant 4), which is located in the transmembrane region of the RhD protein. Whether the phenotype in the sample is DEL, could only be confirmed with antigen quantifying tests.

Novel mutations that present on the backbone of a known variant can alter and reduce the D antigen and can be easily mis-genotyped as in samples (004_119, 004_144). For example, if sample (004_119) was genotyped using single SNP genotyping assays designed to detect weak D mutations, it could be mis-genotyped as a *RHD*01W.03* rather than a D negative sample.

Novel mutations are a concern and need to be included in current genotyping platforms to allow detection of a wider range of blood group variants. It is mostly important to identify apparent RhD negative or weak D donors if they were incorrectly phenotyped as RhD negative, as they pose the risk of immunisation on RhD negative patients. Weak D type 1, 2 and 3 individuals can be safely transfused with D positive blood without being at risk of immunisation (Sandler et al., 2015). However, transfusing RhD negative patient as in sample (004_119) with RhD positive RBCs carries a high risk of sensitisation. Sandler et al. (2017) advise that when serological weak D phenotypes are detected, transfusion facilities ought to determine the *RHD* genotype through further testing, which would mean individuals with a serological weak D phenotype would then be managed as RhD-positive or RhD-negative, according to their *RHD* genotype.

The case of such samples highlights the power of BGG using NGS in detecting novel mutations and not solely dependent on genotyping for

known variants through available genotyping methods such as SSP-PCR. NGS can therefore improve the safety of blood transfusion practices.

3.4.4 Inconclusive Genotyping Results

Four samples failed to amplify one or more of the *RHD* amplicons, which could be due to variations in the *RHD* gene that compromised some of the primer binding regions, discussed in more detail in section 3.4.11.

Two samples (004_137 and 004_138) failed to amplify amplicon 6. The Finnish research group reported that sample (004_138) showed a negative result for exon 10 in the SSP-PCR (Tammi, 2019), indicating that there probably is a deletion or a hybrid gene around exon 10 of the *RHD* gene. Complete or partial deletion of exons 9 and 10 are often associated with DEL phenotype (Fichou et al., 2012; Srivastava et al., 2018; Lopez et al., 2018). Srivastava et al. (2018) identified a 7640bp deletion in the *RHD* gene which includes part of intron 9 and exon 10. This deletion was confirmed (using adsorption and elution techniques) to encode RhD negative allele Srivastava et al. (2018).

The effect of lacking the 8 amino acids encoded by exon 10 is not clear as aromatic and hydrophobic C-terminal amino acids crucial for RhD expression which include valine, phenylalanine, and tryptophan at positions 406, 407, and 408, respectively are present (exon 9) (Srivastava et al., 2018).

Further testing, using a more accurate phenotyping techniques such as elution or flow cytometry, is necessary to determine the effect of exon 10 deletion and the status of the RhD protein.

The *RHD* gene was sequenced for both samples for 5 out 6 amplicons and no variations that could affect RhD expression were detected; therefore, the *RHD* gene was considered undetermined due to the lack of the full gene sequence.

RHD amplicons 1, 2 and 3 failed to amplify in sample (004_150), which cover the region between *RHD* exons 1 and 5. The LR-PCR was repeated using primers adapted from Hyland et al. (2017) to amplify the *RHD* region from exon 2 to exon 7. Data analysis showed 3 exonic mutations detected in exons 4, 5, and a complete deletion of exon 3 (336-486del). This large deletion of exon 3 explains the D negative phenotype in this sample, as the 3 exonic mutations do not encode for a null allele but in fact encode for the *RHD*09.03.01* variant allele. Exon 3 deletion most probably occurred on *RHD*09.03.01* allele. The complete deletion of exon 3 has prevented a successful translation which leads to RhD negative status.

In sample (004_125) that was found to be homozygous for the *RHD* gene on ddPCR, exonic mutations detected included c.48G>C in exon 1, c.602C>G in exon 4, c.667T>G in exons 5, and c.819G>A in exon 6 that predict amino acid changes p.Trp16Cys, p.Thr201Arg, p.Phe223Val, and silent mutation,

respectively. All SNPs detected were heterozygous which indicates presence of two variant *RHD* alleles (compound heterozygote). Possible alleles encoded by these exonic changes are either *RHD*09.03* and *RHD*01.01* or *RHD*09.04* and *RHD*01*. The SSP-PCR CDE kit results reported by Tammi (2019) indicated a presence of weak D 4.0/4.1 allele, which supports NGS results. However, presence of a wild type *RHD* allele that produces a normal RhD protein does not agree with the weak D reactivity in serology. In serology testing, normal D would mask the weak D reactivity and result would be RhD positive instead of weak D. Therefore, presence of an intact copy of either *RHD*01.01* or *RHD*01* allele is unlikely. No other mutations predicting amino acid changes or splice site disruptions were detected. The most possible explanation for this sample is that the wild type *RHD* allele (*RHD*01*) is possibly a variant allele that harbours a deletion or a conversion, masked by the presence of an intact copy of the *RHD* gene (*RHD*09.04*). The possible deletion will lead to a null phenotype. Therefore, only one copy of the *RHD* gene will produce RhD protein which for this sample is the *RHD*09.04* allele which encodes a weak D. Genotyping results for this sample remains inconclusive.

Sample (004_73) failed to amplify *RHD* amplicon 4 that covers the region from intron 4 to intron 7. Different SNPs were detected in exon 4 that indicate a hybrid *RHD-RHCE(4)-RHD* gene where exon 4 of the *RHD* is replaced by *RHCE* exon 4. These changes included c.505A>C; c.509T>G;

c.514A>T; c.544T>A; c.577G>A; c.594A>T; c.602 C>G and predict amino acid changes p.Met169Leu, p.Met170Arg, p.Ile172Phe, p.Ser182Thr, p.Glu193Lys, p.Lys198Asn, p.Thr201Arg which encode the *RHD*17.02* allele. This variant *RHD* allele is linked typically to the $\mathrm{R_1}$ (DCe) haplotype (Faas et al., 1997; Wagner and Flegel, 2014), however, this sample was assigned by serology as $\mathrm{R_0r}$ (Dce/dce). This means that *RHD*17.02* variant allele could also be found in the $\mathrm{R_0}$ (Dce) haplotype. Genotyping results for sample (004_73) remain inconclusive due to the lack of the complete *RHD* sequence.

3.4.5 Rh Haplotype Specific Intronic SNPs

Analysing intronic SNPs, revealed 21 homozygous SNPs present in all samples sequenced.These represent SNP variants of the *RHD*DAU0 (RHD*10.00)* allele, which the hg38 reference sequence encodes. Some intronic SNPs were found to be present in a specific haplotype ($\mathrm{R_2}$), 23 SNPs were homozygous in all $\mathrm{R_2r}$, $\mathrm{R_2R_2}$ and in 3 of the 6 $\mathrm{R_1R_2}$ samples that tested as hemizygous by ddPCR. They were also detected as heterozygous SNPs in all $\mathrm{R_1R_2}$ samples tested as homozygous by ddPCR and in the $\mathrm{R_2R_z}$ sample. Homozygous intronic mutations were detected in all $\mathrm{R_1R_1}$, $\mathrm{R_1r}$, and in the other 3 of the 6 $\mathrm{R_1R_2}$ samples tested as hemizygous in ddPCR.These SNPs were also present in six $\mathrm{R_0r}$ samples and detected as heterozygous SNPs in six $\mathrm{R_1R_2}$ samples tested as homozygous by ddPCR and in the $\mathrm{R_2R_z}$ sample. The similarities of intronic SNPs pattern between differ-

ent haplotypes (R_1, R_0, and R_z) suggest that these haplotypes might have arisen from the same ancestral gene. There were no intronic SNPs specific to each of the R_1, R_0, or R_z alleles, (Figure 3.13).

These intronic SNP patterns were found to be inconsistent in samples where hybrid *RHD* allele is expected. This is due to the *RHCE* insertion, as found in the ISBT samples (004_101-004_106) where exonic SNPs detected indicate hybrid *RHD-RHCE(4-5)-RHD* allele. Intronic SNPs located in introns 3-6 were inconsistent with Rh haplotype assigned by serology, however, other intronic SNPs were consistent with previous patterns identified.

Intronic SNPs related to R_2 haplotype were compared to the previous SNPs that were investigated at the beginning of this research detected by Halawani (2015). Because previous work used the reference sequence hg19, SNPs were correlated to the new accession system used in the reference sequence hg38, (Table 3.3).

3.4.6 *RHD* Allele Specific Intronic SNPs

Three samples (004_69, 004_97, 004_148) genotyped as *RHD*09.03.01*, 1 sample genotyped as *RHD*09.01.02* (004_70), 1 sample (004_96) genotyped as *RHD*03.01*, and 1 sample (004_125) that possibly carries the *RHD*09.03* allele showed 11 intronic SNPs that could be specific to their originating al-

lele. Amino acid sequences of different *RHD* alleles detected in this study were compared and a phylogenetic tree was created using www.phylogeny.fr (Dereeper et al., 2008), to determine if theses variant alleles were originating from the ancestral gene. Figure 3.18 shows that *RHD*09.03.01*, *RHD*09.01.02*, and *RHD*09.03* alleles appear to be originated from the same ancestral gene. Samples (004_69, 004_97, 004_148) genotyped as *RHD*09.03.01 (DAR3.1)*, 33 intronic SNPs were detected as homozygous SNPs in these samples, which were specific to the *DAR3.1* allele.

3.4.7 Rh Haplotype Determination

We attempted to determine the probable correct Rh haplotype for samples where Rh serology results did not agree with the *RHD* genotyping results. To assign the most probable correct haplotype for these samples, we studied the literature to identify if the *RHD* variant allele was linked to a specific haplotype (Wagner et al., 1999; Wagner and Flegel, 2014). We also considered the RhCcEe serology phenotyping results, *RHD* zygosity testing results, and the intronic SNP patterns detected in the sequence.

In samples in which a previously reported variant was identified (Wagner and Flegel, 2014), for example, weak D type 1 *RHD*01W.01* was detected in R_1 (DCe) haplotype samples, weak D type 2 *RHD*01W.02* was found in R_2 (DcE) haplotype samples, and weak D type 3 *RHD*01W.03* was identified in R_1 (DCe) haplotype sample (Wagner et al., 1999), the probable variant haplotypes were consistent with the reported variant haplotypes

in all but two samples (004_73 and 004_123).

The *RHD*15* allele, found in two samples (004_123 and 004_146), is usually associated with antigen E and observed in the R_2 (DcE) haplotype (Wagner and Flegel, 2014) which is the most probable haplotype for sample (004_146), which is supported by the R_2 intronic SNP pattern found in the sequence. However, Ye et al. (2014) observed 2/64 samples identified as *RHD*15* in the ee (Cce) phenotype. In sample (004_123), which showed the $\mathrm{R}_1, \mathrm{R}_0.\mathrm{R}_z$ intronic SNP pattern, the most likely variant haplotype is R_1 (DCe), which supports the observation of *RHD*15* associated also with e.

Sample (004_73) genotyped by NGS as *RHD*17.02*, which is associated with R_1 (DCe) haplotype (Wagner and Flegel, 2014). However, sample was assigned by serology as $\mathrm{R}_0\mathrm{r}$ (Dce/dce). This means that the *RHD*17.02* allele is also associated with R_0 (Dce) haplotype as well as R_1 (DCe).

3.4.8 Evolution of *RHD* Variant Alleles

The evolution of the *RHD* gene, discussed in section 1.4, is thought to be through a duplication event in the *RHCe*ce* allele where the *RHD* gene orientation was altered (Carritt et al., 1997; Okuda and Kajii, 2002). Intronic polymorphism in the *RHCE* intron 2 is a reminder of that genomic event. Different studies have established that the Rh original haplotype was R_0 (Dce) through studying different intronic variations in the *RHD* and *RHCE* genes (Carritt et al., 1997; Okuda and Kajii, 2002).

In this study, *RHD* intronic variation patterns could be linked to the evolution of theses alleles from the same ancestral gene. Figure 3.18 shows a phylogenetic tree of the *RHD* different alleles that were detected in this study. Most of the *RHD* alleles are caused by single nucleotide variations on the backbone of the *RHD* wild type allele, while others occur on the backbone of a variant *RHD* allele.

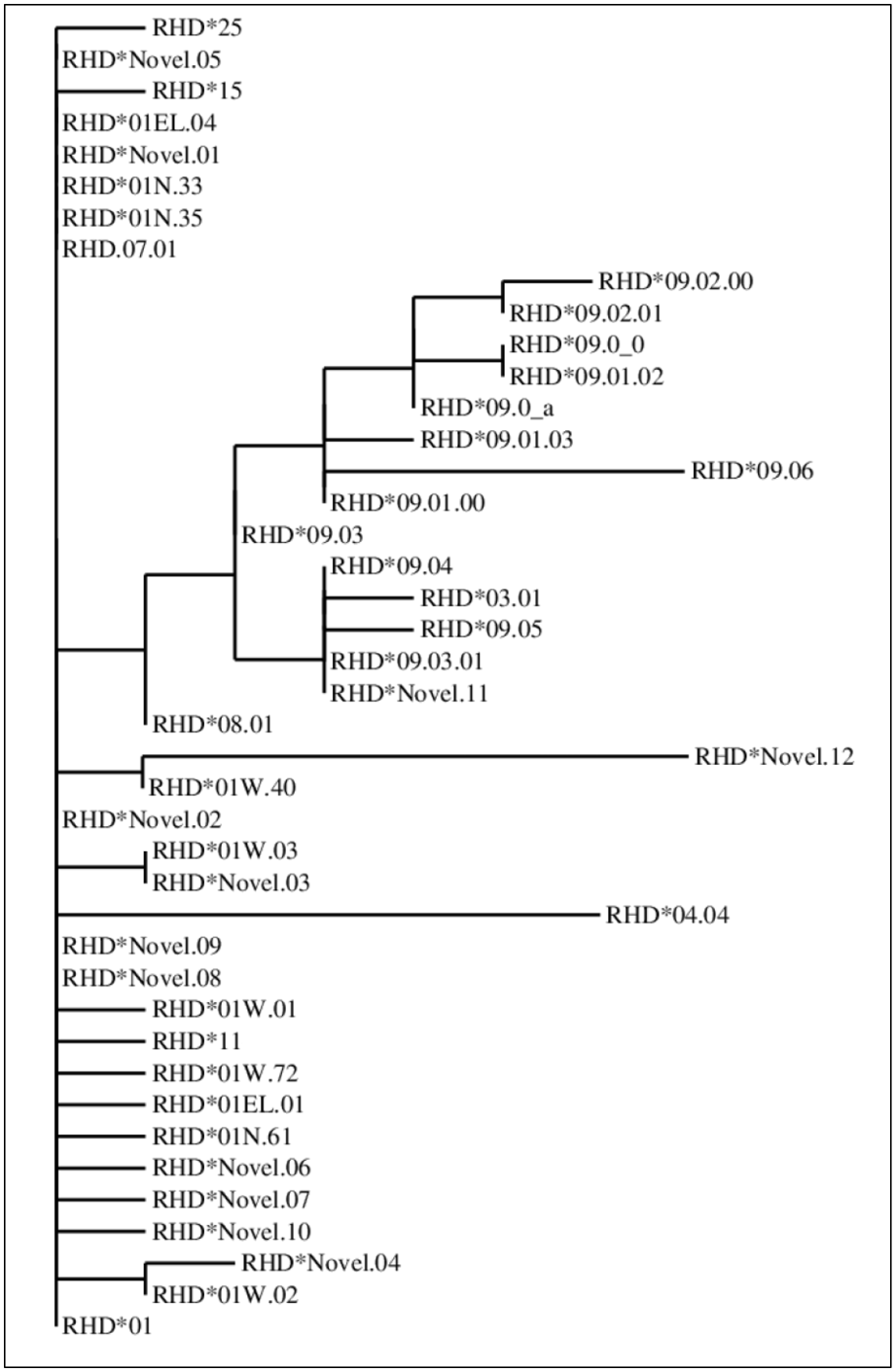


Figure 3.18: Phylogenetic tree of the *RHD* gene different alleles.

3.4.9 *RHAG* NGS

Two samples (004_14 and 004_74) were serologically phenotyped as weak D; however, no amino acid changes were predicted from sequencing of the *RHD* gene. Different mutations in the *RHAG* gene have been reported that disturb the expression of the Rh proteins (Polin et al., 2016; Mu et al., 2019; Wen et al., 2019). Therefore, we sequenced the *RHAG* gene for these samples (004_14 and 004_74) that showed weak D reactivity without finding any alterations in the *RHD* gene. Sample (004_14) showed a heterozygous SNP 808G>A in exon 6 of the *RHAG* gene leading to Val270Ile that encodes the *RHAG*04* allele (Figure 3.14), in addition to a silent SNP in exon 6 861G>A. Although this SNPs is a heterozygous, it could be the main cause of the weak D reactivity, hence no changes were detected from the sequencing of the *RHD* gene in this sample to explain the weak D reactivity. Protein interaction model of the RhD and RhAG proteins was visualised on the CLC workbench 10 to locate the Val270Ils. The Val270Ile is not located on the interaction face of the RhAG protein. Further investigation of this allele and how it may affect the RhD protein is needed to understand the effect of such mutations on the expression on the Rh proteins. Wen et al. (2019) has reported other mutations in the *RHAG* gene that lead to weak D phenotype such as c.572G>A (p.Arg191Gln).

3.4.10 *RHD* SMS

MinION™ sequencer was utilised to sequence the complete *RHD* gene from 13 samples in 6 overlapping amplicons, which were also sequenced on the Ion PGM™. Quality assessment for reads was preformed and an average quality PHRED-score of 11 was detected (Figure 3.15), which indicates accuracy of about 90%. According to Laver et al. (2015) MinION™ error rate per base with a certain quality number does not correspond to error rate per base expected for PHRED value of the same quality results, which means that MinION™ sequencing quality score does not correspond with PHRED-score as NGS technologies. Using R6 MinION™ chemistry, they found that MinION™ has about 40% error on single read sequencing (Laver et al., 2015); however, in current work 1D flow cells R9 version was used for sequencing which shows lower error rate (Tyler et al., 2018).

SMS allows real time sequencing and data analysis while reducing preparation and sequencing time (Schadt et al., 2010; Lu et al., 2016). Error in the MinION™ sequencing occurs in specific sequences with 11% estimated error rate (Magi et al., 2016). However, with 40x depth of coverage MinION™ may cause a false substitution and insertion every 10–50 kb and a false deletion every 1000 bp, which may cause an issue in detection of variations (Magi et al., 2016). However, in this work we did not encounter an issue in that matter as high coverage across the gene was achieved with

up to 500x in some regions. Variation in coverage is expected due to the fact that LR-PCR amplicons are sequenced. In BGG using MinION™, we were able to detect exonic and intronic SNPs and identify alleles which agree with ones found using Ion PGM™

MinION™ library preparation and sequencing time was reduced to a day from 4-5 days in NGS, after PCR amplification and purification which takes 3 days for 20 samples. SMS was successful in BGG and provided sufficient data to call variants and determine alleles. However, data handling, storage and analysis in MinION™ are challenging. The evolving nature of this sequencing technology makes it difficult to establish a user friendly software that would enable fast and accurate data analysis to make it suitable for clinical use.

3.4.11 LR-PCR Limitations

The *RHD* gene was fully sequenced on the Ion PGM™ and MinION™ through LR-PCR amplification. Although LR-PCR is an efficient technique in amplifying the gene for sequencing, the LR-PCR approach is limited. Hybrid *RHD-RHCE* alleles or partial D alleles may not amplify if a primer position is compromised by deletion or mutations. The *RHD*-specific primers in this study were tested with different weak and partial D samples including; *RHD*DVI.01*, *RHD*DNB*, *RHD*DIV.04*, *RHD*DVII.01*, *DFR1*, *DFR2*, and *RHD*DIIIa*. Amplification for all 6 PCR amplicons was achieved in all samples except for samples with the *RHD*DVI.01* allele,

in which amplicon 4 amplified very weakly, which affected the coverage around exon 5.

The *RHD* primers designed in this study also failed to amplify 3 Finnish samples (004_137, 004_138, 004_150). Two samples (004_137, 004_138) failed to amplify the *RHD* amplicon 6, while sample (004_150) failed to amplify *RHD* amplicons 1, 2, and 3. Hyland et al. (2017) LR-PCR primers were used to amplify the *RHD* gene from exon 2 to exon 7 for sample (004_150), which was successful. Data analysis revealed a complete deletion of intron 2 and exon 3, which explains the failure of amplification of the *RHD* amplicons 1, 2, and 3 that cover the *RHD* gene from exon 1 to intron 4.

Amplification issues could be resolved in the future by using a hybrid primer approach; for example, an *RHD* specific forward primer and an *RHCE* specific reverse primer or removing PCR amplification by utilising SMS technology. SMS allow real time sequencing and data analysis while reducing preparation and sequencing time (Schadt et al., 2010; Lu et al., 2016), which was explored in section 3.3.7.2. SMS was successful and provided sufficient data to call variants and determine alleles and reduced preparation time and sequencing to a day.

3.5 Conclusion

We sequenced the complete *RHD* gene in samples (n=133) using NGS to study *RHD* mutations, assess variations present in the population, identify novel variants and establish reference *RHD* allele sequences. Ten novel alleles were identified in 18 samples. Reference sequences for these novel alleles were submitted to Genbank and assigned accession numbers. The use of NGS in BGG to overcome limitations in other genotyping platforms has been promising.

NGS of the complete *RHD* gene was effective in identifying exonic and intronic changes encoding known and novel variants of *RHD* gene. However, in samples where a hybrid *RHD* gene or a large deletion exists, LR-PCR amplification was not successful as one or more of the primers' binding regions have been affected, thus preventing annealing.

NGS also enabled exploring the intronic SNPs and their relation to a specific Rh haplotype. Twenty-one intronic SNPs were identified in all samples sequenced (n=133) indicating their specificity to the *RHD*DAU0 (RHD*10.00)* haplotype which the hg38 reference sequence encodes. Twenty-three intronic SNPs were found to be R_2 specific, and 15 were related to R_1, R_0 and R_z haplotypes. Intronic SNPs may represent a novel diagnostic approach to investigate known and novel variants of the *RHD* and *RHCE* genes. NGS allows the study of intronic and exonic changes in blood group genes

enabling the detection of known and novel mutations, which ultimately improves the safety of blood services.

4. *RHCE* Genotyping Using NGS

4.1 Introduction

The Rh blood group antigens are encoded by two closely related genes, the *RHD* and the *RHCE* genes located on chromosome 1, which encode the RhD protein and RhCcEe protein, respectively (Colin et al., 1991; Wagner and Flegel, 2000). The two closely related genes encode 55 different antigens in the Rh system; five of them are considered the leading antigens, which include RhD, RhC, Rhc, RhE, and Rhe. Although RhD is the most immunogenic and clinically significant antigen in the Rh blood group system, other antigens are immunogenic and antigen incompatibility could lead to HTRs and HDFN, discussed in sections 1.8.1 and 1.8.2, including C, c, E, e antigens (Le van Kim et al., 1992; Ranasinghe et al., 2003; Kumawat et al., 2012; Sheeladevi et al., 2013; Avent, 2018). Single nucleotide variation is the main reason for the *RHCE* gene polymorphism, which gives rise to different antigens.

Six exonic variations in the *RHCE* gene (c.48C>G, c.150C>T, c.178C>A, c.201A>G, c.203A>G, c.307C>T) lead to amino acid alterations (p.Cys16Trp,

p.Leu60Ile, p.Asn68Ser, p.Pro103Ser) that have been linked to the Rhc>RhC change, however, the c.307C>T (p.Pro103Ser) mutation is the key mutation (Mouro et al., 1993; Simsek et al., 1994). Mutation c.676G>C in exon 5 (p.Ala226Pro) is the key mutation for the Rhe>RhE polymorphism. Mutations in the *RHCE* gene lead to different variations of the Rh antigens including weakened and partial expression, for example, C^w, E^w and VS antigen.

Serological tests used for Rh phenotyping cannot differentiate between normal or altered epitopes, especially in partial expression (Moulds et al., 2014). For example, serological tests cannot differentiate between partial RhC and normal expression as long antibody binding sites are not affected. Incorrectly phenotyping patients for Rh antigens could lead to allo-antibody production and therefore make finding compatible blood for transfusion a rigorous mission.

WES has previously been used to sequence the coding regions of the *RHCE* gene to determine the *RHCE* alleles (Chou et al., 2017; Schoeman et al., 2018). However, in the work presented here the whole *RHCE* gene has been sequenced including coding and non-coding regions. To sequence the *RHCE* gene, we used the same approach used for the *RHD* sequencing. Eight overlapping LR-PCR amplicons were used to amplify the *RHCE* gene for sequencing. All samples sequenced for the *RHCE* gene were also sequenced for the *RHD* gene, except for RhD negative samples that were

confirmed to be *RHD* negative by *RHD* zygosity testing using ddPCR.

4.2 Aim and Objectives

We aimed to develop a methodology to fully sequence the *RHCE* gene, from samples that had been sequenced for the *RHD* gene, to establish *RHCE* allele reference sequences by studying intronic SNP patterns in different *RHCE* alleles.To achieve our aim, objectives were as follow:

1. Design LR-PCR *RHCE* specific primers that would allow specific amplification from the target gene.

2. Sequence the *RHCE* gene using Ion PGM™ and analyse data using hg38 as the reference sequence.

3. Call variants and study intronic SNP patterns present in different *RHCE* alleles.

4. Utilise detected intronic SNPs to establish *RHCE* allele specific reference sequences.

5. Compare intronic SNP patterns found in the *RHD* gene with ones present in the *RHCE* gene through creating tree alignments of these sequences.

4.3 Results

4.3.1 *RHCE* NGS Data Quality & Coverage

The *RHCE* gene was sequenced with the same approach used for the *RHD* gene sequencing, in which the *RHCE* gene was amplified in 8 overlapping LR-PCR amplicons, (Figure 2.2). Samples included in the *RHCE* genotyping were also genotyped for the *RHD* gene except for samples that tested negative for the *RHD* gene on ddPCR (n=18). In these samples, *RHD* zygosity testing showed a ratio ≈ 0 for *RHD5* and *RHD7* assays against the reference gene *AGO1*, indicating a deletion of the *RHD* gene.

Three sequencing libraries were prepared and sequenced on the Ion PGM™ as described in section 2.9, in which each sequencing run contained 34, 35, and 47 samples (Table 4.1). ISPs loading density, number of reads after polyclonal filter, and distribution of reads length was reported after each sequencing run from the Ion PGM™. ISPs loading density refers to the percentage of chip wells that contain ISP, in which red colour indicates adequate loading density, yellow indicates less-than adequate loading density, and blue indicates inadequate loading density or absence of beads. Figure 4.1 shows a sequencing run summary in which the sequencing library contained 47 samples. Low ISPs loading density (59%) was noticed in this *RHCE* sequence run. High polyclonal percentage (35%) was also noticed which might be due to high number of reads in the sequence

Table 4.1: Ion PGM™ sequencing run summary report details for *RHCE* sequencing runs

Library no.	No. of samples	ISP loading (%)	Polyclonal (%)	Usable reads (%)	Total usable reads	Average reads /sample
1	34	79	31	63	3,146,839	92,554
2	35	65	47	40	1,616,947	46,198
3	47	59	35	47	3,051,892	63,581

run. Low ISPs density and high polyclonal percentage in this sequence run decreased the number of usable reads to 47% (about 3 million reads), which lead to the loss of reads in some samples (004_99, 004_102, 004_111, 004_113).

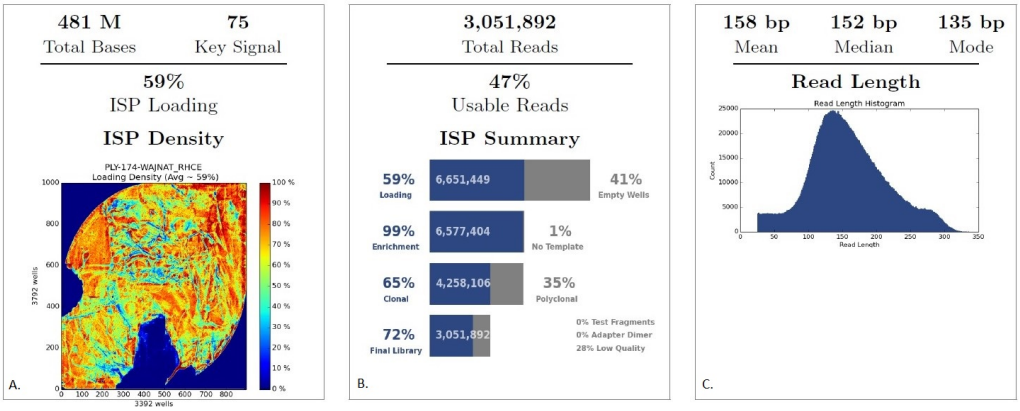


Figure 4.1: Ion PGM™ sequencing run summary for the *RHCE* gene from a single representative run. (a) The loading density in 316™ Chip was 59% of the chip capacity, giving a final throughput of 481 Mb. (b) Over 3 million reads were recovered after enrichment and polyclonal filters. (c) A histogram shows the distribution of the reads length mean around 158 bp, in which the y-axis and x-axis represent the read count and the read length in bp, respectively.

NGS data, obtained in FASTQ files, was assessed for sequence quality.

CLC workbench 10 software was utilised to generate a sequence summary report similar to the ones generated for *RHD* gene sequencing. High PHRED sequence quality score was noticed across all the samples sequenced. PHRED score represents the quality score of each base, which estimates the probability of an error in the sequence. Figure 4.2 shows a PHRED score of over 30 for a representative sample that was sequenced indicating >99% accuracy. Full coverage across the gene was achieved except for a region in intron 7 (25,378,316- 25,381,617) and in some samples breakage of coverage in intron 2 was noticed which could be due to poor PCR amplification of amplicon RHCE-2. Depth of coverage across the *RHCE* coding regions was adequate (Figure 4.3) and sufficient for calling variants, except for a region in intron 7 where breakage of coverage was noticed from 25,378,316 to 25,381,617 (Figure 4.4) made it difficult to call variants from that region.

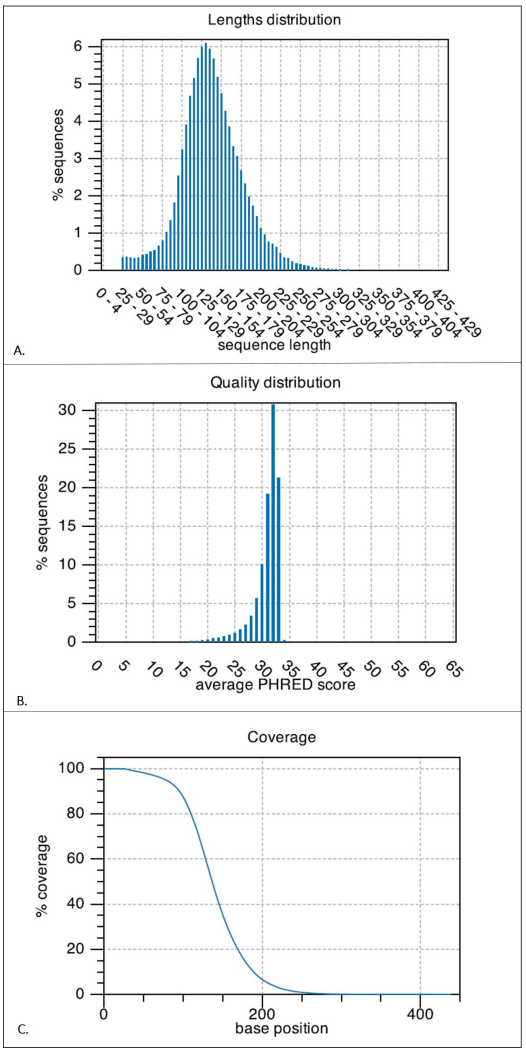


Figure 4.2: *RHCE* gene sequence quality summary report generated by the CLC workbench 10 software from a single representative sample. A. Distribution of reads sequence lengths from a single representative sample; x axis: sequence length in base-pairs, y axis: number of reads showing a specific length normalized to the total number of sequences. B. PHRED score for *RHCE* gene sequence from a single representative sample. x axis: PHRED-score; y axis: percentage of sequences, normalized to the total number of reads, noted at that quality. C. The number of sequences that cover the individual base positions from a single representative sample, x axis: base position; y axis: number of sequences covering individual base positions normalized to the total number of sequences.

FASTQ reads mapped to the *RHCE* gene in the hg38 reference sequence. Genes in the chromosomal assembly including the homologous *RHD* gene were masked in the mapping process to prevent reads from scattering. Detailed mapping report was generated using the CLC workbench 10 similarly to the ones generated in the *RHD* genotyping analysis, which included coverage analysis, mismatches analysis, read length distribution, insertion and deletion level, and quality for match and mismatch distribution.

Variant detection was performed on a minimum coverage of 30 (i.e the numbers of overlapping reads over a single base) and variants detected were analysed on a single base basis considering different parameters including number and percentage of reads and nucleotide count (Nielsen et al., 2011). Breakage of coverage was noticed in intron 7 and intron 2 in some samples, but a full coverage across all 10 exons was achieved and depth of sequencing coverage was adequate and sufficient for calling variants in most samples, (Figures 4.3 & 4.4). In 4 samples (004_99, 004_102, 004_111, 004_113), with low number of reads, adequate coverage over exons was not achieved which made allele determination impossible.

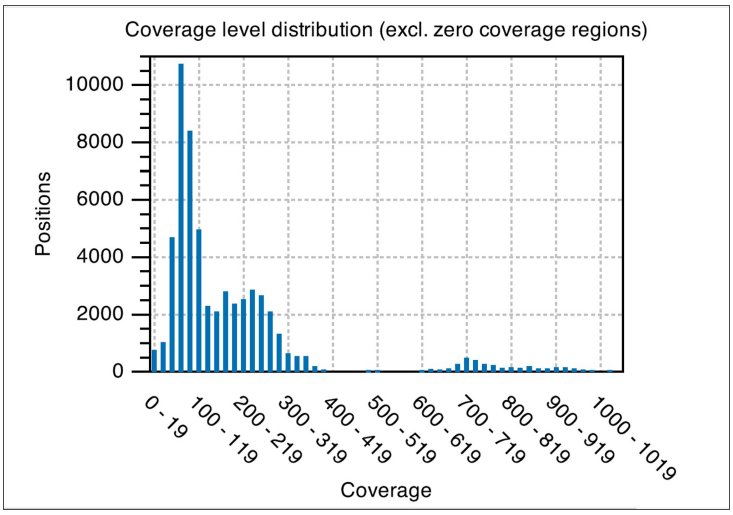


Figure 4.3: Level of coverage across the *RHCE* gene from a single representative sample. Level of coverage (x axis) and number of bases observed at that level (y axis). A total of 54,653 positions (*RHCE* gene) have coverage between 1 and 710. Positions of zero coverage were not included in the graph which include the rest of chromosome 1.

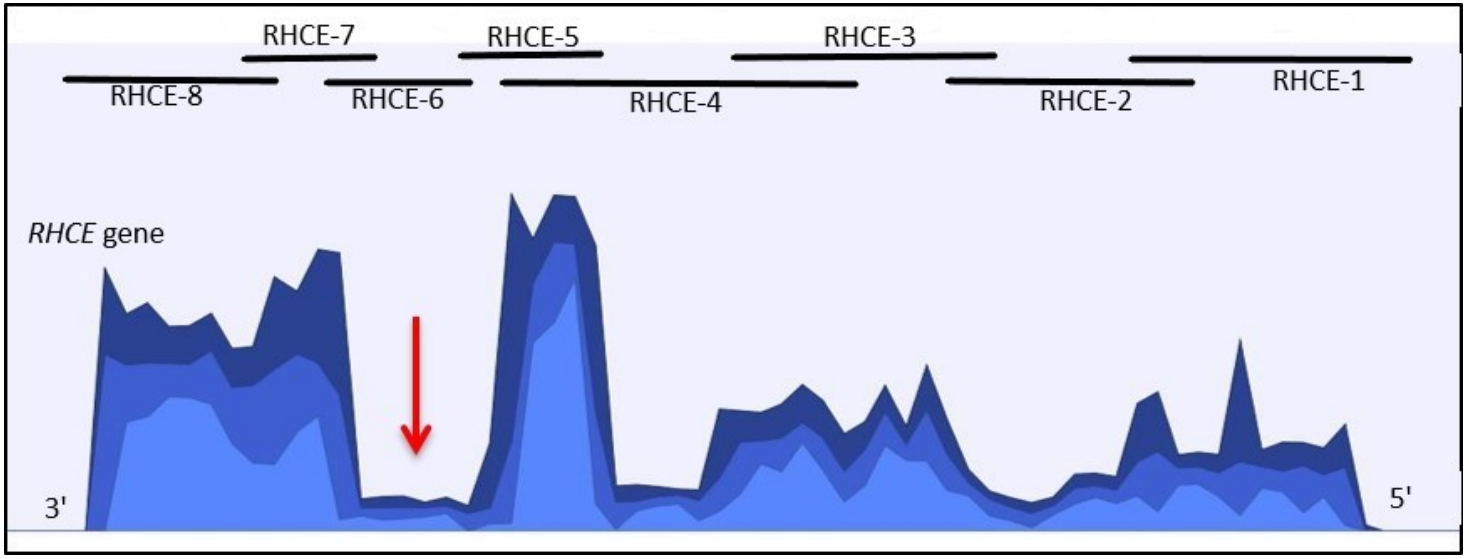


Figure 4.4: Depth of coverage of the *RHCE* gene sequence from a single representative sample. (top) Eight *RHCE* amplicons mapped to the *RHCE* gene (refer to Figure 2.2). ↓ Breakage of coverage noticed in *RHCE* intron 7.

4.3.2 *RHCE* NGS Genotyping

Samples (n=100) were genotyped for the *RHCE* gene using NGS in which (n=87) were blood donor samples and (n=13) were samples from the ISBT workshop ISBT (1996). Blood donor samples included 7 $\mathrm{R_1R_1}$, 20 $\mathrm{R_1r}$, 9 $\mathrm{R_1R_2}$, 6 $\mathrm{R_2R_2}$, 14 $\mathrm{R_2r}$, 11 $\mathrm{R_0r}$, 1 $\mathrm{R_2R_z}$, 6 rr, 7 r'r, and 6 r"r samples. The ISBT workshop samples did not have Rh serological results available. The *RHCE* sequencing data was analysed to determine the *RHCE* alleles for each sample (Table 4.2). Different alleles were compared to determine which SNPs belonged to which allele, especially in heterozygous samples, in an attempt to phase the *RHCE* alleles. The *RHCE* reference sequence in hg38 shows c.307C (p.Pro103) in exon 2, c.676G (p.Ala226) in exon 5, and c.48C (p.Cys16) in exon 1 that encode weak ce allele *RHCE*01.01*. The reference sequence hg38 shows nucleotides associated with Rhc in exon 2 which include: c.150C (silent). c.178C (p.Leu60), c.201A (silent), c.203A (p.Asn68), c.307C (p.Pro103) .

It is important to keep in mind that the hg38 reference sequence is a variant *RHCE* allele *RHCE*01.01* that presents c.48C (p.16Cys) in exon 1. Therefore when sequence data is analysed against the hg38 reference sequence, samples phenotyped by serology as normal e showed a SNP in exon 1 c.48C>G (p.Cys16Trp), (Figure 4.5). On the other hand, other alleles associated with the c.48G>C SNP, for example *RHCE*02* allele, the SNP was not detected as it matches the reference sequence. SNP c.48C>G (p.Cys16Trp) in exon

1 encodes the *RHCE*01* allele. Table 4.3 lists the different *RHCE* alleles detected in samples sequenced by NGS.

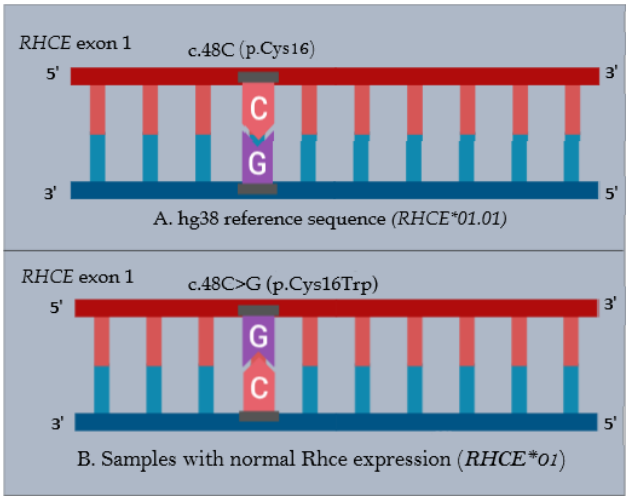


Figure 4.5: A. *RHCE* hg38 reference sequence is a variant allele that shows c.48C (p.Cys16) in exon 1 that encodes *RHCE*01.01*. B. Samples sequenced with normal Rhce expression show c.48C>G (p.Cys16Trp) when aligned to the reference sequence.

4.3.2.1 NHSBT Blood Donor Samples

NGS data was analysed, variants called and the allele they encode are listed in Table 4.2. In seven R_1R_1 (DCe/DCe) samples (004_01- 004_07), 5 homozygous exonic changes were detected in exon 2 including c.150C>T (p.silent), c.178C>A (p.Leu60Ile), c.201A>G (p.silent), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser) which encode for homozygous *RHCE*02 (RHCE*Ce)* allele.

Twenty R_1r (DCe/dce) samples (004_08- 004_023 and 004_25-004_28) and 5 r'r (dCe/dce) samples (004_81- 004_85) showed 6 heterozygous exonic

changes, c.48C>G in exon 1 predicting amino acid change p.Cys16Trp which encode for *RHCE*01 (RHCE*ce)*. Five changes were detected in exon 2 including c.150C>T (p.silent), c.178C>A (p.Leu60Ile), c.201A>G (p.silent), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser) which encode *RHCE*02 (RHCE*Ce)* allele. These samples encode heterozygous *RHCE*02; RHCE*01 (RHCE*Ce; RHCE*ce)* alleles.

Nine samples assigned their genotype by serology as $\mathrm{R_1R_2}$ (DCe/DcE) (004_29, 004_30, 004_31, 004_33, 004_34, 004_35, 004_37, 004_39, 004_40) showed 7 heterozygous exonic changes in exons 1, 2, and 5. Heterozygous SNPs in exons 1 c.48C>G and 5 c.676G>C predicting amino acid changes p.Cys16Trp and p.Ala226Pro, respectively, encode the *RHCE*03 (RHCE*cE)* allele. SNPs in exon 2 c.150C>T (p.silent), c.178C>A (p.Leu60Ile), c.201A>G (p.silent), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser) encode *RHCE*02 (RHCE*Ce)* allele. These samples encode heterozygous *RHCE*02; RHCE*03 (RHCE*Ce; RHCE*cE)* alleles.

Six $\mathrm{R_2R_2}$ (DcE/DcE) samples (004_41, 004_42, 004_43, 004_45, 004_46, 004_47) showed 2 homozygous SNPs in exon 1 c.48C>G and exon 5 c.676G>C that predict amino acid changes p.Cys16Trp and p.Ala226Pro, respectively, which encode homozygous *RHCE*03 (RHCE*cE)*.

Fourteen $\mathrm{R_2r}$ (DcE/dce) samples (004_49- 004_62) and 6 r"r (dcE/dce) samples (004_88- 004_93) showed homozygous SNPs in exon 1 c.48C>G

predicting amino acid change p.Cys16Trp and heterozygous SNP c. 676G>C in exon 5 that predict amino acid change p.Ala226Pro. These samples were genotyped as heterozygous *RHCE*03; RHCE*01 (RHCE*cE; RHCE*ce)*.

From 11 samples assigned Rh genotype by serology as $\mathrm{R_0r}$ (Dce/dce), 9 samples (004_63, 004_64, 004_65, 004_66, 004_67, 004_68, 004_71, 004_72, 004_73) showed heterozygous SNP c.48C>G in exon 1 predicting amino change p.Cys16Trp which indicates that the $\mathrm{R_0}$ allele presents c.48C in exon 1 similar to the reference sequence. These samples were genotyped as heterozygous *RHCE*01; RHCE*01.01 (RHCE*ce; RHCE*ce.01)*. The other two $\mathrm{R_0r}$ samples (004_69 and 004_70) presented multiple exonic changes. Sample (004_69) showed 2 heterozygous SNPs in exon 1 c.48C>G and c.106G>A that predict amino acid changes p.Cys16Trp and p.Ala36Thr, respectively. This sample was genotyped as heterozygous *RHCE*02.09 (RHCE*Ce.09)* encoded by c.106G>A and *RHCE*01 (RHCE*ce)* encoded by c.48C>G change in exon 1. Sample (004_70) showed 6 heterozygous exonic changes in exons 1, 5, and 6. These SNPs include c.48C>G (p.Cys16Trp) in exon 1, c.712A>G (p.Met238Val), c.733C>G (p.Leu245Val), 787A>G (p.Arg 263Gly), and c.800T>C (p.Met267Lys) in exon 5, and c.916A>G (p.Ile306Val) in exon 6. Exon 1 SNPs encode for *RHCE*01 (RHCE*ce)*, while other changes encode for *RHCE*01.04.01 (RHCE*ceAR)*.

One $\mathrm{R_2R_z}$ (DcE/DCE) sample (004_74) showed 6 heterozygous and 1 homozygous exonic changes. Heterozygous SNPs include exon 1 c.48C>G

(Cys16Trp), exon 2 c.150C>T (p.silent), c.178C>A (p.Leu60Ile), c.201A>G (p.silent), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser), and the homozygous SNPs in exon 5 c.676G>C that predict amino acid change p.Ala226Pro which is the key change for RhE expression. This sample genotyped as heterozygous *RHCE*02; RHCE*04 (RHCE*cE; RHCE*CE)*.

Two blood donor samples (004_86 and 004_87) from Black/Caribbean donors were serologically assigned Rh genotype as r'r (dCe/dce). Both samples tested as negative for the *RHD* gene by zygosity testing using ddPCR with a ratio = 0 for both assays *RHD5* and *RHD7* against the reference gene *AGO1*. However, both samples amplified amplicon 1 of the *RHD* gene, indicating the presence of deletion or a hybrid *RHD-CE-RHD* gene. *RHCE* sequencing analysis showed heterozygous SNPs c.48C>G (p.Cys16Trp) in exon 1, c.733C>G (p.Leu245Val) in exon 5, and c.1006G>T (p.Gly336Cys) in exon 7. The SNP in exon 1 encode *RHCE*01 (RHCE*ce)* allele and the SNPs in exon 5 and 7 encode the variant allele *RHCE*ceVS.03 (RHCE*01.02.03)*, a common variant in individuals of Caribbean descent.

Six samples (004_75- 004_80) assigned Rh haplotype by serology as rr (dce/dce), showed 1 homozygous SNP in exon 1 c.48C>G predicting amino acid change p.Cys16Trp which encode for homozygous *RHCE*01 (RHCE*ce)* allele.

Table 4.2: *RHCE* gene NGS results (Ion PGM™).

Lab no.	Rh serology*	Ethnicity*	*RHD* zygosity	*RHD* Allele	*RHCE* SNPs	SNPs zygosity	Exon	Amino acid	*RHCE* Allele
004_01	R_1R_1	Caucasian	Homozygous	*RHD*01*	150C>T	Homozygous	2	silent	*RHCE*Ce; RHCE*Ce*
					178C>A	Homozygous	2	Leu60Ile	
					201A>G	Homozygous	2	silent	
					203A>G	Homozygous	2	Asn68Ser	
					307C>T	Homozygous	2	Pro103Ser	
004_02	R_1R_1	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*Ce*
004_03	R_1R_1	Other	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*Ce*
004_04	R_1R_1	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*Ce*
004_05	R_1R_1	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*Ce*
004_06	R_1R_1	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*Ce*
004_07	R_1R_1	Caucasian	Discrepancy	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*Ce*
004_08	R_1r	Caucasian	Hemizygous	*RHD*01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*Ce; RHCE*ce*
					150C>T	Heterozygous	2	silent	
					178C>A	Heterozygous	2	Leu60Ile	
					201A>G	Heterozygous	2	silent	
					203A>G	Heterozygous	2	Asn68Ser	
					307C>T	Heterozygous	2	Pro103Ser	
004_09	R_1r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_10	R_1r	Chinese	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_11	R_1r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_12	R_1r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_13	R_1r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_14	R_1r	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_15	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_16	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_17	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_18	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_19	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_20	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_21	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_22	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_23	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_25	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*

continued ...

Lab no.	Rh serology*	Ethnicity*	*RHD* zygosity	*RHD* Allele	*RHCE* SNPs	SNPs zygosity	Exon	a.a change	*RHCE* Allele
004_26	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_27	R_1r	Caucasian	Hemizygous	*RHD*01W.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_28	R_1r	Caucasian	Hemizygous	*RHD*01W.03*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_29	R_1R_2	Caucasian	Homozygous	*RHD*01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*Ce; RHCE*cE*
					150C>T	Heterozygous	2	silent	
					178C>A	Heterozygous	2	Leu60Ile	
					201A>G	Heterozygous	2	silent	
					203A>G	Heterozygous	2	Asn68Ser	
					307C>T	Heterozygous	2	Pro103Ser	
					676G>C	Heterozygous	5	Ala226Pro	
004_30	R_1R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_31	R_1R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_33	R_1R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_34	R_1R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_35	R_1R_2	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_37	R_1R_2	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_39	R_1R_2	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_40	R_1R_2	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*Ce; RHCE*cE*
004_41	R_2R_2	Caucasian	Homozygous	*RHD*01W.02*	48C>G	Homozygous	1	Cys16Trp	*RHCE*cE; RHCE*cE*
					676G>C	Homozygous	5	Ala226Pro	
004_42	R_2R_2	Not disclosed	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*cE*
004_43	R_2R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*cE*
004_45	R_2R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*cE*
004_46	R_2R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*cE*
004_47	R_2R_2	Caucasian	Homozygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*cE*
004_49	R_2r	Caucasian	Hemizygous	*RHD*01*	48C>G	Homozygous	1	Cys16Trp	*RHCE*cE; RHCE*ce*
					676G>C	Heterozygous	5	Ala226Pro	
004_50	R_2r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_51	R_2r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_52	R_2r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_53	R_2r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_54	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_55	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_56	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_57	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*

continued ...

Lab no.	Rh serology*	Ethnicity*	*RHD* zygosity	*RHD* Allele	*RHCE* SNPs	SNPs zygosity	Exon	a.a change	*RHCE* Allele
004_58	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_59	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_60	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_61	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_62	R_2r	Caucasian	Hemizygous	*RHD*01W.02*	=	=	=	=	*RHCE*cE; RHCE*ce*
004_63	R_0r	Caucasian	Hemizygous	*RHD*01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*01.01; RHCE*01*
004_64	R_0r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*01.01; RHCE*01*
004_65	R_0r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*01.01; RHCE*01*
004_66	R_0r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*01.01; RHCE*01*
004_67	R_0r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*01.01; RHCE*01*
004_68	R_0r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*01.01; RHCE*01*
004_69	R_0r	Caucasian	Hemizygous	*RHD*09.03.01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*Ce.09; RHCE*ce*
					106G>A	Heterozygous	1	Ala36Thr	
004_70	R_0r	Caucasian	Hemizygous	*RHD*09.03.01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*ceAR; RHCE*ce*
					712A>G	Heterozygous	5	Met238Val	
					733C>G	Heterozygous	5	Leu245Val	
					787A>G	Heterozygous	5	Arg263Gly	
					800T>C	Heterozygous	5	Met267Lys	
					916A>G	Heterozygous	6	Ile306Val	
004_71	R_0r	Caucasian	Hemizygous	*RHD*01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*01.01; RHCE*01*
004_72	R_0r	Caucasian	Hemizygous	*RHD*01*	=	=	=	=	*RHCE*01.01; RHCE*01*
004_73	R_0r	Caucasian	Hemizygous	*RHD*17.02*	=	=	=	=	*RHCE*01.01; RHCE*01*
004_74	R_2R_z	Caucasian	Homozygous	*RHD*01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*cE; RHCE*CE*
					150C>T	Heterozygous	2	silent	
					178C>A	Heterozygous	2	Leu60Ile	
					201A>G	Heterozygous	2	silent	
					203A>G	Heterozygous	2	Asn68Ser	
					307C>T	Heterozygous	2	Pro103Ser	
					676G>C	Homozygous	5	Ala226Pro	
004_75	rr	Caucasian	Negative	NA	48C>G	Homozygous	1	Cys16Trp	*RHCE*ce; RHCE*ce*

continued ...

Lab no.	Rh serology*	Ethnicity*	*RHD* zygosity	*RHD* Allele	*RHCE* SNPs	SNPs zygosity	Exon	a.a change	*RHCE* Allele
004_76	rr	Caucasian	Negative	NA	=	=	=	=	*RHCE*ce; RHCE*ce*
004_77	rr	Caucasian	Negative	NA	=	=	=	=	*RHCE*ce; RHCE*ce*
004_78	rr	Caucasian	Negative	NA	=	=	=	=	*RHCE*ce; RHCE*ce*
004_79	rr	Caucasian	Negative	NA	=	=	=	=	*RHCE*ce; RHCE*ce*
004_80	rr	Caucasian	Negative	NA	=	=	=	=	*RHCE*ce; RHCE*ce*
004_81	r′r	Caucasian	Negative	NA	48C>G	Heterozygous	1	Cys16Trp	*RHCE*Ce; RHCE*ce*
					150C>T	Heterozygous	2	silent	
					178C>A	Heterozygous	2	Leu60Ile	
					201A>G	Heterozygous	2	silent	
					203A>G	Heterozygous	2	Asn68Ser	
					307C>T	Heterozygous	2	Pro103Ser	
004_82	r′r	Caucasian	Negative	NA	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_83	r′r	Caucasian	Negative	NA	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_84	r′r	Caucasian	Negative	NA	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_85	r′r	Caucasian	Negative	NA	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_86	r′r	Black/Caribbean	Negative	NA	48C>G	Heterozygous	1	Cys16Trp	*RHCE*ceVS.03; RHCE*ce*
					733C>G	Heterozygous	5	Leu245Val	
					1006G>T	Heterozygous	7	Gly336Cys	
004_87	r′r	Black/Caribbean	Negative	NA	=	=	=	=	*RHCE*ceVS.03; RHCE*ce*
004_88	r″r	Caucasian	Hemizygous	Novel variant 1	48C>G	Homozygous	1	Cys16Trp	*RHCE*cE; RHCE*ce*
					676G>C	Heterozygous	5	Ala226Pro	
004_89	r″r	Caucasian	Negative	NA	=	=	=	=	*RHCE*cE; RHCE*ce*
004_90	r″r	Caucasian	Negative	NA	=	=	=	=	*RHCE*cE; RHCE*ce*
004_91	r″r	Caucasian	Negative	NA	=	=	=	=	*RHCE*cE; RHCE*ce*
004_92	r″r	Caucasian	Negative	NA	=	=	=	=	*RHCE*cE; RHCE*ce*
004_93	r″r	Caucasian	Negative	NA	=	=	=	=	*RHCE*cE; RHCE*ce*
004_95	No data	No data	Hemizygous	*RHD*01W.01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*Ce; RHCE*ce*
					150C>T	Heterozygous	2	silent	
					178C>A	Heterozygous	2	Leu60Ile	
					201A>G	Heterozygous	2	silent	
					203A>G	Heterozygous	2	Asn68Ser	
					307C>T	Heterozygous	2	Pro103Ser	
004_96	R_1r	No data	Hemizygous	*RHD*03.01*	NA	NA	NA	NA	Undetermined
004_98	No data	No data	Hemizygous	*RHD*25*	48C>G	Homozygous	2	Cys16Trp	*RHCE*cE; RHCE*ce*
					676G>C	Heterozygous	5	Ala226Pro	

continued ...

Lab no.	Rh serology*	Ethnicity*	*RHD* zygosity	*RHD* Allele	*RHCE* SNPs	SNPs zygosity	Exon	a.a change	*RHCE* Allele
004_99	No data	No data	Hemizygous	*RHD*04.04*	NA	NA	NA	NA	Undetermined
004_102	R_1r	No data	Hemizygous	Undetermined	NA	NA	NA	NA	Undetermined
004_104	R_2r	No data	Hemizygous	Undetermined	48C>G	Homozygous	1	Cys16Trp	*RHCE*cE; RHCE*ce*
					676G>C	Heterozygous	5	Ala226Pro	
004_106	No data	No data	Hemizygous	Undetermined	48C>G	Homozygous	2	Cys16Trp	*RHCE*cE; RHCE*ce*
					676G>C	Heterozygous	5	Ala226Pro	
004_109	R_1r	No data	Hemizygous	*RHD*07.01*	48C>G	Heterozygous	1	Cys16Trp	*RHCE*Ce; RHCE*ce*
					150C>T	Heterozygous	2	silent	
					178C>A	Heterozygous	2	Leu60Ile	
					201A>G	Heterozygous	2	silent	
					203A>G	Heterozygous	2	Asn68Ser	
					307C>T	Heterozygous	2	Pro103Ser	
004_110	R_1r	No data	Hemizygous	*RHD*07.01*	=	=	=	=	*RHCE*Ce; RHCE*ce*
004_111	R_1r	No data	Hemizygous	*RHD*17.01*	NA	NA	NA	NA	Undetermined
004_113	No data	No data	Hemizygous	*RHD*01*	NA	NA	NA	NA	Undetermined
004_114	No data	No data	Homozygous	*RHD*01*	122A>G	Heterozygous	1	Gln41Arg	*RHCE*CeCW; RHCE*Ce*
					150C>T	Homozygous	2	silent	
					178C>A	Homozygous	2	Leu60Ile	
					201A>G	Homozygous	2	silent	
					203A>G	Homozygous	2	Asn68Ser	
					307C>T	Homozygous	2	Pro103Ser	
004_116	No data	No data	Homozygous	*RHD*01*	150C>T	Homozygous	2	silent	*RHCE*Ce; RHCE*Ce*
					178C>A	Homozygous	2	Leu60Ile	
					201A>G	Homozygous	2	silent	
					203A>G	Homozygous	2	Asn68Ser	
					307C>T	Homozygous	2	Pro103Ser	

*As supplied by the NHSBT, Bristol, United Kingdom
NA: Not applicable.
= Similar to previous.
Blue indicates samples Rh serology r'r (dCe/dce) does not agree with *RHCE* genotyping results *RHCE*ceVS.03; RHCE*ce* which indicates the presence of a hybrid *RHD-RHCE-RHD* gene that produces *RHCE*Ce* protein. *RHCE*ceVS.03* allele is a common variant in individuals of Caribbean descent.

4.3.2.2 ISBT Samples

From the ISBT workshop 1996, 13 samples were genotyped for the *RHCE* gene. The *RHCE* allele was not determined in 5 samples (004_96, 004_99, 004_102, 004_111, 004_113) which is due to either poor amplification of the target gene by LR-PCR leading to low or lack of coverage or library reads loss as a result of high polyclonal amplification in the library enrichment step.

Three samples (004_95, 004_109, 004_110) showed 6 heterozygous exonic SNPs that encode heterozygous *RHCE*02; RHCE*01 (RHCE*Ce; RHCE*ce)*. These changes include: c.48C>G (p.Cys16Trp) in exon 1 which encode for *RHCE*01 (RHCE*ce)*, c.150C>T (p.silent), c.178C>A (p.Leu60Ile), c.201A>G (p.silent), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser) in exon 2 which encode *RHCE*02 (RHCE*Ce)* allele.

One sample (004_116) was genotyped as homozygous *RHCE*02 (RHCE*Ce)* showing homozygous exonic changes associated with RhC which include c.150C>T (p.silent), c.178C>A (p.Leu60Ile), c.201A>G (p.silent), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser) in exon 2.

Two samples (004_98, 004_106) were genotyped as heterozygous *RHCE*03; RHCE*01 (RHCE*cE; RHCE*ce)*. Homozygous SNPs in exon 1 c.48C>G predicting amino acid change p.Cys16Trp and heterozygous SNP c. 676G>C in exon 5 that predict amino acid change p.Ala226Pro, were detected in

Table 4.3: Molecular basis of the different *RHCE* alleles detected by NGS.

Allele name*	Phenotype*	Nucleotide	Exon (Intron)	Amino acid
*RHCE*01* or *RHCE*ce*	c+e+f+ or RH:4,5,6	48C>G	1	Cys16Trp
*RHCE*01.01* or *RHCE*ce.01*	c+e+	48C	1	Cys16
*RHCE*02* or *RHCE*Ce*	C+e+Ce+ or RH:2,5,7	48C	1	Cys16
		150C>T	2	silent
		178C>A	2	Leu60Ile
		201A>G	2	silent
		203A>G	2	Asn68Ser
		307C>T$^{\Upsilon}$	2	Pro103Ser
		676G	5	Ala226
*RHCE*03* or	c+E+cE+ or	48C>G	1	Cys16Trp
*RHCE*cE*	RH:3,4,27	676G>C**	5	Ala226Pro
*RHCE*04* or	C+E+CE+	48C	1	Cys16
*RHCE*CE*	or RH:2,3,22	150C>T	2	silent
		178C>A	2	Leu60Ile
		201A>G	2	silent
		203A>G	2	Asn68Ser
		307C>T$^{\Upsilon}$	2	Pro103Ser
		676G>C**	5	Ala226Pro
*RHCE*02.08.01*	Partial C C^{w}+ MAR-	48C>G	1	Cys16Trp
or *RHCE*CeCW*	or RH:8,-51	122A>G	1	Gln41Arg
		178C>A	2	Leu60Ile
		203A>G	2	Asn68Ser
		307C>T$^{\Upsilon}$	2	Pro103Ser
*RHCE*01.20.03*	Partial c, partial e, V-VS+hr^{B-}	48C	1	Cys16
or *RHCE*ceVS.03*	or RH:-10,20,-31	733C>G	5	Leu245Val
		1006G>T	7	Gly336Cys
*RHCE*01.04.01*	V+^{w}VS- Hr- hr^{s}	48C	1	Cys16
or *RHCE*ceAR*	or RH:+^{w},-18,-19,-20	712A>G	5	Met238Val
		733C>G	5	Leu245Val
		787A>G	5	Arg263Gly
		800T>A	5	Met267Lys
		916A>G	6	Ile306Val
*RHCE*02.09* or *RHCE*Ce.09*	Partial C+ CX+ MAR- or RH:9-51	106G>A	1	Ala36Thr

$^{\Upsilon}$ C+ key mutation, **E+ key mutation. *Allele name and phenotype as referenced in ISBT (2021) and Reid et al. (2012).

both samples.

One sample (004_106) showed homozygous SNPs c.48C>G (p.Cys16Trp) in exon 1 and c.676G>C (p.Ala226Pro) in exon 5 that encode for homozygous *RHCE*03 (RHCE*cE)*.

One sample (004_114) showed heterozygous SNP c.122A>G in exon 1 that predicts amino acid change p.Gln41Arg and homozygous SNPs c.150C>T, c.178C>A, c.201A>G, c.203A>G, c.307C>T that predict amino acid changes p.silent, p.Leu60Ile, p.silent, p.Asn68Ser, p.Pro103Ser, respectively. Exonic changes encode heterozygous *RHCE*02.08.01; RHCE*02 (RHCE*CeCW; RHCE*Ce)*.

Although ISBT samples were not supplied with Rh serological phenotyping results, through analysing *RHD* and *RHCE* NGS data, we were able to establish Rh haplotype for 8 ISBT samples (004_95, 004_98, 004_106, 004_104, 004_109, 004_110, 004_114, 004_116). Table 4.4 lists the most possible Rh haplotype for these samples based on *RHD* and *RHCE* NGS results.

Table 4.4: Most probable Rh haplotype for IBST samples genotyped for *RHD* and *RHCE* using NGS.

Lab no.	Rh serology	*RHD* zygosity	*RHD* allele	*RHD* intronic SNPs	*RHCE* allele	Rh haplotype
004_95	No data	Hemizygous	*RHD*01W.01*	R_1, R_0, R_z	*RHCE*Ce; RHCE*ce*	R_1r (DCe/dce)
004_98	No data	Hemizygous	*RHD*25*	R_2	*RHCE*cE; RHCE*ce*	R_2r (DcE/dce)
004_104	R_2r	Hemizygous	Undetermined	R_2	*RHCE*cE; RHCE*ce*	R_2r (DcE/dce)
004_106	No data	Hemizygous	Undetermined	R_2	*RHCE*cE; RHCE*ce*	R_2r (DcE/dce)
004_109	R_1r	Hemizygous	*RHD*07.01*	R_1, R_0, R_z	*RHCE*Ce; RHCE*ce*	R_1r (DCe/dce)
004_110	R_1r	Hemizygous	*RHD*07.01*	R_1, R_0, R_z	*RHCE*Ce; RHCE*ce*	R_1r (DCe/dce)
004_114	No data	Homozygous	*RHD*01*	R_1, R_0, R_z	*RHCE*CeCW; RHCE*Ce*	R_1R_1 (DCe/DCe)
004_116	No data	Homozygous	*RHD*01*	R_1, R_0, R_z	*RHCE*Ce; RHCE*Ce*	R_1R_1 (DCe/DCe)

4.3.3 Intronic SNPs

Multiple intronic SNPs were detected in all samples sequenced. Eighty-nine intronic SNPs (Table 4.5), mostly located in intron 2, were present as homozygous SNPs in $\mathrm{R_1R_1}$ (*RHCE*Ce; RHCE*Ce*) samples, and as heterozygous SNPs in $\mathrm{R_1R_2}$ (*RHCE*Ce;RHCE*cE*) samples, $\mathrm{R_1r}$ (*RHCE*Ce; RHCE*ce*) samples, and the $\mathrm{R_2R_z}$ (*RHCE*Ce; RHCE*CE*) sample. From the pattern of these SNPs we concluded that these SNPs are specific to C allele including *RHCE*Ce* and *RHCE*CE* alleles.

Twenty-two intronic SNPs (Table 4.6) located in intron 1 and upstream of the *RHCE* gene were present as homozygous SNPs in $\mathrm{R_2R_2}$ (*RHCE*cE; RHCE*cE*), $\mathrm{R_2r}$ (*RHCE*cE; RHCE*ce*), r''r (*RHCE*cE; RHCE*ce*), and rr (*RHCE*ce; RHCE*ce*) samples, and as heterozygous SNPs in $\mathrm{R_1R_2}$ (*RHCE*Ce; RHCE*cE*) samples, $\mathrm{R_1r}$ (*RHCE*Ce;RHCE*ce*) samples, $\mathrm{R_0r}$ (*RHCE*ce.01; RHCE*ce*) samples, and the $\mathrm{R_2R_z}$ (*RHCE*cE;RHCE*CE*) sample, which indicates that these intronic SNPs are located in $\mathrm{R_2}$, r'' (*RHCE*cE*) and r (*RHCE*ce*) alleles. These two alleles encode c indicating that these SNPs are c specific. All intronic SNPs detected in different Rh haplotypes were compared to the *RHCE* hg38 reference sequence (Table 4.7). No intronic changes specific to RhE were detected, but allele reference sequence was established and submitted to GenBank based on the SNV present in exon 5 that encode Rhe to RhE change.

Table 4.5: Intronic SNPs present in samples with R_1 and R_z haplotypes (RhC specific).

Position	SNP*	Intron	rs no. $^{\alpha}$	Position	SNP	Intron	rs no. $^{\alpha}$
25,409,808	A>T	1	rs28594470	25,407,340	A>T	2	rs185015905
25,409,676	G>A	1	rs71652382	25,407,310	C>G	2	rs150384644
25,409,655	G>C	1	rs71652381	25,407,278	A>G	2	-
25,409,650	A>G	1	rs71652380	25,407,275	G>A	2	-
25,409,635	T>C	1	rs2982326	25,407,267	T>C	2	rs1396520207
25,409,530	T>C	1	rs375284259	25,407,262	T>C	2	rs1166255958
25,409,482	A>G	1	rs71652379	25,407,253	C>A	2	-
25,409,354	T>C	1	rs581970	25,407,249	C>T	2	-
25,409,288	A>G	1	rs597377	25,407,202	C>T	2	rs1210277488
25,409,207	T>C	1	rs581424	25,407,177	C>T	2	-
25,409,194	T>C	1	rs596939	25,407,175	A>G	2	-
25,409,073	G>A	1	rs596447	25,407,148	C>T	2	-
25,408,889	T>C	1	rs184592905	25,407,142	C>T	2	-
25,408,548	C>T	2	rs675936	25,407,129	C>G	2	-
25,408,514	A>del-	2	rs201437378	25,407,128	G>A	2	-
25,408,476	A>G	2	rs593705	25,407,124	C>T	2	rs1461147141
25,408,386	C>T	2	rs147140560	25,407,115	G>A	2	rs374139742
25,408,274	C>T	2	rs189396125	25,407,008	G>A	2	rs141399586
25,408,183	A>C	2	rs1309971753	25,406,976	A>G	2	rs147829334
25,408,164	A>G	2	rs149867242	25,406,914	C>T	2	rs190290858
25,408,149	T>C	2	rs1329812325	25,406,909	G>A	2	rs1214049339
25,408,147	T>C	2	rs1230452984	25,406,909	G>A	2	rs1214049339
25,408,106	A>G	2	rs186132701	25,406,878	A>G	2	rs1252070032
25,408,065	G>A	2	rs1345393229	25,406,877	C>G	2	-
25,408,063	C>G	2	rs1285130086	25,406,876	C>T	2	-
25,408,059	A>G	2	rs1402951394	25,406,875	C>T	2	-
25,408,027	C>T	2	rs181050023	25,406,872	A>G	2	rs28759034
25,407,962	T>A	2	rs1407745287	25,406,867	T>C	2	-
25,407,958	T>C	2	rs1398397158	25,406,868	G>C	2	-
25,407,949	C>A	2	rs1300077586	25,406,847	C>A	2	-
25,407,902	C>G	2	rs200103035	25,406,810	C>A	2	-
25,407,897	G>T	2	rs201812726	25,406,803	G>T	2	-
25,407,837	C>A	2	rs138441784	25,406,787	A>G	2	rs1476265413
25,407,789	C>T	2	rs1296676005	25,406,719	A>G	2	rs185049399
25,407,786	G>C	2	rs1441844028	25,406,718	A>C	2	rs142561320
25,407,761	C>T	2	-	25,406,697	C>T	2	-
25,407,722	T>C	2	rs1418501291	25,406,689	G>C	2	rs1263698499
25,407,694	T>C	2	-	25,406,674	A>G	2	-
25,407,688	G>C	2	-	25,406,665	G>A	2	rs1281889085
25,407,634	A>G	2	rs924675715	25,406,649	A>G	2	rs1313773019
25,407,631	G>C	2	rs966013933	25,406,561	T>del-	2	-
25,407,630	T>A	2	rs912179584	25,406,557	T>C	2	rs192918853
25,407,626	G>C	2	rs1315056073	25,406,534	A>G	2	rs187818188
25,407,447	T>C	2	rs686738	25,406,526	C>T	2	rs148410263
25,407,349	A>G	2	rs138147046				

* Intronic SNPs (hg38) were present as homozygous SNPs in R_1R_1 samples, and heterozygous SNPs in R_1R_2, R_1r, and R_2R_z samples. $^{\alpha}$ From the database of SNPs (Sherry et al., 2001). - Not found in the database of SNPs.

Table 4.6: Intronic SNPs present in samples with R_2, r″, and r haplotypes (Rhc specific).

Position	SNP*	Intron	rs no. [α]
25,421,866	T>C	upstream	rs4649082
25,421,744	G>A	upstream	rs2375313
25,421,164	C>T	upstream	rs2281179
25,421,155	G>A	upstream	rs2072933
25,421,082	C>T	upstream	rs2072932
25,420,072	T>C	1	rs2904813
25,420,908	G>T	1	rs2072931
25,419,823	G>A	1	rs2904815
25,419,154	C>T	1	rs2982360
25,418,713	G>T	1	rs2982358
25,418,570	T>C	1	rs2092326
25,417,977	C>T	1	rs1883427
25,417,674	C>T	1	rs636206
25,416,624	C>T	1	rs2982348
25,416,125	G>A	1	rs76440808
25,416,115	C>T	1	rs77258408
25,416,094	A>T	1	rs112136614
25,415,138	C>A	1	rs28633797
25,414,356	A>C	1	rs28530483
25,413,699	C>T	1	rs2982341
25,413,312	T>C	1	rs600580
25,412,573	G>A	1	rs12048617

*Intronic SNPs were present as homozygous SNPs in R_2R_2, R_2r, r″r, and rr samples, and as heterozygous SNPs in R_1R_2, R_2R_z, R_1r, and R_0r samples. [α] From the database of SNPs (Sherry et al., 2001).

Table 4.7: Position of intronic variations determined by NGS and their corresponding nucleotide in the four main *RHCE* alleles *RHCE*Ce, RHCE*CE* and *RHCE*cE, RHCE*ce* in comparison to the hg38 reference sequence.

Position	**rs no.** [α]	**Intron**	**hg38**	***RHCE*Ce, RHCE*CE***	***RHCE*cE, RHCE*ce***
25,421,866	rs4649082	upstream	T	T	C
25,421,744	rs2375313	upstream	G	G	A
25,421,164	rs2281179	upstream	C	C	T
25,421,155	rs2072933	upstream	G	G	A
25,421,082	rs2072932	upstream	C	C	T
25,420,072	rs2904813	1	C	C	T
25,420,908	rs2072931	1	G	G	T
25,419,823	rs2904815	1	G	G	A
25,419,154	rs2982360	1	C	C	T
25,418,713	rs2982358	1	G	G	T
25,418,570	rs2092326	1	T	T	C
25,417,977	rs1883427	1	C	C	T
25,417,674	rs636206	1	C	C	T
25,416,624	rs2982348	1	C	C	T
25,416,125	rs76440808	1	G	G	A
25,416,115	rs77258408	1	C	C	T
25,416,094	rs112136614	1	A	A	T
25,415,138	rs28633797	1	C	C	A
25,414,356	rs28530483	1	A	A	C
25,413,699	rs2982341	1	C	C	T
25,413,312	rs600580	1	T	T	C
25,412,573	rs12048617	1	G	G	A
25,409,808	rs28594470	1	A	T	A
25,409,676	rs71652382	1	G	A	G
25,409,655	rs71652381	1	G	C	G
25,409,650	rs71652380	1	A	G	A
25,409,635	rs2982326	1	T	C	T
25,409,530	rs375284259	1	T	C	T
25,409,482	rs71652379	1	A	G	A
25,409,354	rs581970	1	T	C	T
25,409,288	rs597377	1	A	G	A
25,409,207	rs581424	1	T	C	T
25,409,194	rs596939	1	T	C	T
25,409,073	rs596447	1	G	A	G
25,408,889	rs184592905	1	T	C	T
25,408,548	rs675936	2	C	T	C
25,408,514	rs201437378	2	A	del-	A
25,408,476	rs593705	2	A	G	A
25,408,386	rs147140560	2	C	T	C
25,408,274	rs189396125	2	C	T	C
25,408,183	rs1309971753	2	A	C	A
25,408,164	rs149867242	2	A	G	A
25,408,149	rs1329812325	2	T	C	T
25,408,147	rs1230452984	2	T	C	T
25,408,106	rs186132701	2	A	G	A
25,408,065	rs1345393229	2	G	A	G
25,408,063	rs1285130086	2	C	G	C
25,408,059	rs1402951394	2	A	G	A
25,408,027	rs181050023	2	C	T	C
25,407,962	rs1407745287	2	T	A	T
25,407,958	rs1398397158	2	T	C	T

continued …

Position	rs no. [α]	Intron	hg38	*RHCE*Ce, RHCE*CE*	*RHCE*cE, RHCE*ce*
25,407,949	rs1300077586	2	C	A	C
25,407,902	rs200103035	2	C	G	C
25,407,897	rs201812726	2	G	T	G
25,407,837	rs138441784	2	C	A	C
25,407,789	rs1296676005	2	C	T	C
25,407,786	rs1441844028	2	G	C	G
25,407,761	-	2	C	T	C
25,407,722	rs1418501291	2	T	C	T
25,407,694	-	2	T	C	T
25,407,688	-	2	G	C	G
25,407,634	rs924675715	2	A	G	A
25,407,631	rs966013933	2	G	C	G
25,407,630	rs912179584	2	T	A	T
25,407,626	rs1315056073	2	G	C	G
25,407,447	rs686738	2	T	C	T
25,407,349	rs138147046	2	A	G	A
25,407,340	rs185015905	2	A	T	A
25,407,310	rs150384644	2	C	G	C
25,407,278	-	2	A	G	A
25,407,275	-	2	G	A	G
25,407,267	rs1396520207	2	T	C	T
25,407,262	rs1166255958	2	T	C	T
25,407,253	-	2	C	A	C
25,407,249	-	2	C	T	C
25,407,202	rs1210277488	2	C	T	C
25,407,177	-	2	C	T	C
25,407,175	-	2	A	G	A
25,407,148	-	2	C	T	C
25,407,142	-	2	C	T	C
25,407,129	-	2	C	G	C
25,407,128	-	2	G	A	G
25,407,124	rs1461147141	2	C	T	C
25,407,115	rs374139742	2	G	A	G
25,407,008	rs141399586	2	G	A	G
25,406,976	rs147829334	2	A	G	A
25,406,914	rs190290858	2	C	T	C
25,406,909	rs1214049339	2	G	A	G
25,406,909	rs1214049339	2	G	A	G
25,406,878	rs1252070032	2	A	G	A
25,406,877	-	2	C	G	C
25,406,876	-	2	C	T	C
25,406,875	-	2	C	T	C
25,406,872	rs28759034	2	A	G	A
25,406,867	-	2	T	C	T
25,406,868	-	2	G	C	G
25,406,847	-	2	C	A	C
25,406,810	-	2	C	A	C
25,406,803	-	2	G	T	G
25,406,787	rs1476265413	2	A	G	A
25,406,719	rs185049399	2	A	G	A
25,406,718	rs142561320	2	A	C	A
25,406,697	-	2	C	T	C
25,406,689	rs1263698499	2	G	C	G
25,406,674	-	2	A	G	A
25,406,665	rs1281889085	2	G	A	G
25,406,649	rs1313773019	2	A	G	A

continued …

Position	rs no. [α]	Intron	hg38	*RHCE*Ce, RHCE*CE*	*RHCE*cE, RHCE*ce*
25,406,561	-	2	T	del-	T
25,406,557	rs192918853	2	T	C	T
25,406,534	rs187818188	2	A	G	A
25,406,526	rs148410263	2	C	T	C

[α] From the database of SNPs (Sherry et al., 2001).

- Not found in the database of SNPs.

4.3.4 Microsatellite Variation in Intron 2

Polymorphism in the *RHCE* intron 2, (AC) and (GCAC) described in Kemp et al. (1999) in samples with Rhc variation, were investigated in samples, with Rhc expression, sequenced in this study. Only samples with homozygous *RHCE*c* genes were analysed for these variations, which included 6 rr (dce/dce) samples (004_75- 004_80), 11 $\mathrm{R_0r}$ (Dce/dce) samples (004_63-004_73), and 6 $\mathrm{R_2R_2}$ (DcE/DcE) samples (004_41- 004_47). Heterozygous *RHCE* genes were not included due to the inability to distinguish *cis/trans* position. The *RHCE* gene in the hg38 reference sequence have 13 AC repeats and 4 GCAC repeats, which were also present in $\mathrm{R_0r}$ samples. Other Rh haplotypes including rr samples showed $(\mathrm{AC})_{11\ -1\ 3}$ and $(\mathrm{GCAC})_6$, $(\mathrm{AC})_9$ and $(\mathrm{GCAC})_6$ were seen in $\mathrm{R_2R_2}$ samples, and $(\mathrm{AC})_{11}$ and $(\mathrm{GCAC})_6$ were seen in r''r samples, results summarised in Table 4.8.

Although there was breakage of coverage in intron 2 for a few samples, which made variant calling for that region challenging, we were able to confirm the presence of the 109 bp insertion in samples with the *RHCE*Ce* allele and the absence of this 109 bp sequence from samples with *RHCE*cE*

Table 4.8: Frequency of microsatellite GCAC and AC repeats in *RHCE* intron 2 in different Rh haplotypes.

Rh haplotype	GCAC repeat	AC repeat
R_0 (Dce)	4	13
r (dce)	6	11-13
R_2 (DcE)	6	9
r″ (dcE)	6	11

allele (Kemp et al., 1999; Zhou et al., 2008). In samples with the *RHCE*cE* allele, a gap of 109 bp was noticed in intron 2 as shown in Figure 4.6, while a continuous sequence was noticed in the same region in samples with the *RHCE*Ce* allele.

Figure 4.6: Insertion detected in intron 2 in samples with *RHCE*Ce* allele (top) and missing from samples with *RHCE*cE* allele (bottom), indicated by a gap in the same region.

4.4 Discussion

4.4.1 *RHCE* Reference Allele Sequences

The *RHCE* NGS data was analysed and exonic and intronic changes detected from samples with different Rh haplotypes were compared. The aim was to create *RHCE* alleles reference sequences. The *RHCE* gene was sequenced from (n=87) blood donor samples and (n=13) ISBT 1996 workshop samples. Blood donor samples included different serologically predicted Rh genotypes including 7 $\mathrm{R_1R_1}$, 20 $\mathrm{R_1r}$, 9 $\mathrm{R_1R_2}$, 6 $\mathrm{R_2R_2}$, 14 $\mathrm{R_2r}$, 11 $\mathrm{R_0r}$, 1 $\mathrm{R_2R_z}$, 6 rr, 7 r'r, and 6 r''r samples. Sequence data was aligned and analysed against the hg38 reference sequence (NC_000001.11) (Tatusova et al., 2016). This reference sequence is *RHCE*ce.01 (RHCE*01.01)* and shows a SNP in exon 1 c.48G>C (p.Trp16Cys). Therefore, samples with r, r'', and $\mathrm{R_2}$ haplotypes showed the SNP c.48C>G (p.Cys16Trp) in exon 1.

Based on the exonic and intronic changes in different *RHCE* alleles, three reference sequences were established and submitted to Genbank and registered with accession numbers MN091965 for *RHCE*Ce* allele, MN091966 for the *RHCE*cE* allele, and MN624142 for the *RHCE*ce* allele.

4.4.1.1 *RHCE* Variant Alleles

The *RHCE* sequencing analysis revealed different exonic mutations that were assessed to determine the *RHCE* allele for each sample (Table 4.2). The *RHCE* alleles were undetermined in four samples (004_97, 004_102,

004_111, 004_112) due to coverage breakage in coding regions. This might be due to high percentage of polyclonal amplification during library enrichment that led to the loss of over 40% of the library reads (Figure 4.1).

The *RHCE* alleles detected in all samples sequenced include *RHCE*ce (RHCE*01)*, *RHCE*ce.01 (RHCE*01.01)*, *RHCE*Ce (RHCE*02)*, *RHCE*cE (RHCE*03)*, *RHCE*CE (RHCE*04)*, *RHCE*ceVS.03 (RHCE*01.20.03)*, *RHCE*CeCW (RHCE*02.08.01)*, *RHCE*Ce.09 (RHCE*02.09)*, and *RHCE*ceAR (RHCE*01.04.01)* (Table 4.3).

Two samples (004_86 and 004_87) from Black/Caribbean donors were assigned their Rh genotype by serology as r'r (dCe/dce) and tested negative for the *RHD* zygosity testing using ddPCR (Table 3.1). The two samples presented three heterozygous exonic changes in the *RHCE* gene that encode heterozygous *RHCE*ce; RHCE*ceVS.03* alleles. The variant *RHCE*ceVS.03* allele is a common allele in RhD negative individuals of Afro/ Caribbean descent (Noizat-Pirenne et al., 2002; Moulds et al., 2014). This haplotype is known as $\mathrm{r'^{s}}$ haplotype (Moulds et al., 2014). The $\mathrm{r'^{s}}$ haplotype have mutated Rh genes including a hybrid *RHD-RHCE-RHD* allele on the *RHD* gene, and $\mathrm{ce^{s}}$ on the *RHCE* gene. The combination of the two mutated genes encode for negative D (Moulds et al., 2014), partial c (Pham et al., 2009), weak partial e known as $\mathrm{e^{s}}$ (Rodrigues et al., 2004), VS (RH20), V (RH10) (Daniels et al., 1998), hrB (RH31) (Beal et al., 1995), and HrB (RH34) antigens. Both $\mathrm{r'^{s}}$ haplotype type 1 and 2 encode for a weak partial C (Lomas et al., 1994). Both samples (004_86 and 004_87)

were phenotyped by serology as RhC positive. Patients, with $\mathrm{r'^{S}}$ haplotype who are incorrectly phenotyped, are at risk of developing Rh alloantibodies to missing or altered epitopes, for instance, anti-C in a seemingly Rh C-positive person (Noizat-Pirenne et al., 2002; Moulds et al., 2014).

Sample (004_113) showed heterozygous SNP c.122A>G (p.Gln41Arg) in exon 1 homozygous SNPs c.48C>G (p.Cys16Trp), c.178C>A (p.Leu60Ile), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser) that encode *RHCE*CeCW*. Rh antigen $\mathrm{C^{w}}$ (Rh8) is one of the many low frequency antigens in the Rh system with about 2% frequency in Whites. This variant has been associated with DCe allele but has been rarely observed in D and/or C negative individuals (Mouro et al., 1995). In sample (004_113), $\mathrm{C^{w}}$ allele was observed in DCe haplotype as heterozygous allele, however, the $\mathrm{C^{w}}$ allele can be present as homozygous alleles *RHCE*CeCW:RHCE*CeCW*. In such cases, MAR (RH51), a high incident Rh antigen, shows negative expression on the RBCs (O'Shea et al., 2001; Orzińska et al., 2016). Individuals with such a haplotype are able to produce anti-MAR-like antibodies after a sensitisation event such as transfusion or pregnancy. This antibody is able to agglutinate RBCs with most Rh phenotypes excluding D- - and $\mathrm{Rh_{null}}$ phenotypes as both are MAR negative. Finding compatible blood products for patients with anti-MAR-like antibodies is very difficult due to the limitation of screening antisera availability and the probability of finding MAR negative blood products (Orzińska et al., 2016). Therefore,

it is especially important to identify $\mathrm{C^w}$ alleles to prevent sensitisation in the first place which could be through sequencing or PCR as performed by Orzińska et al. (2016).

Sample (004_70) showed heterozygous SNPs in exon 5 c.712A>G (p.Met238Val), c.733C>G (p.Leu245Val), c.787A>G (p.Arg263Gly), c.800T>A (p.Met267Lys) and a SNP at exon 6 c.916A>G (p.Ile306Val) which all encode the rare variant allele *RHCE*ceAR*, first described by Hemker et al. (1999) and commonly found in Africans. This sample was phenotyped by serology as $\mathrm{R_0r}$ (Dce/dce) and was genotyped using NGS as *RHCE*ceAR;RHCE*ce*. The variant *RHCE*ceAR* allele is known to express partial e but different studies have described anti-c in individuals with homozygous *RHCE*ceAR* alleles indicating that the rare allele also encodes for a partial c (Peyrard et al., 2009; Hipsky et al., 2010). Serological testing might fail to identify partial c/e in individuals with homozygous variants which might lead subsequently to allo-antibody production in these patients, which stresses the importance of molecular genotyping to prevent immunisation.

4.4.2 *RHCE* Allele Specific Intronic SNPs

Analysing intronic SNPs detected (Table 4.7), showed different patterns of intronic SNPs that are specific to the *RHCE* allele. Samples with Rh haplotype $\mathrm{R_0}$ (Dce) showed similar sequence to the hg38 reference sequence and no intronic SNPs were detected. Eighty-nine intronic SNPs (Table 4.5) were detected in introns 1 and 2 in samples with RhC including $\mathrm{R_1}$

(DCe) and R_Z (DCE). These SNPs are close to the hotspot RhC mutations in exon 2 c.150C (silent), c.178C>A (p.Leu60Ile), c.201A>G (silent), c.203A>G (p.Asn68Ser), c.307C>T (p.Pro103Ser) which may explain the intronic SNP distribution in intron 2. We concluded that these intronic SNPs are specific to the *RHCE* alleles with RhC expression, which include *RHCE*Ce* and *RHCE*CE*.

A further 22 intronic SNPs (Table 4.6) were detected in samples with Rh serologically predicted haplotype R_2 (DcE), r'' (dcE) and r (dce). Samples with serologically predicted haplotype R_2 (DcE) and r'' (dcE) samples showed RhE change in exon 5 c.676G>C (p.Ala226Pro) and Rhc change in exon 1 c.48C>G (p.Cys16Trp). Samples with serologically predicted Rh haplotype r (dce) showed exon 1 Rhc change c.48C>G (p.Cys16Trp). We concluded that these SNPs are specific to *RHCE*ce* and *RHCE*cE*.

No other intronic changes were detected that are RhE specific. This might be due to the development of the RhE allele from a single mutation after the development of the *RHCE*C* allele from *RHCE*ce* and *RHD* genes recombination (Kemp et al., 1999; Zhou et al., 2008). Kemp et al. (1999) also described intronic changes in *RHCE* intron 2 that link to the Rh variants development. Most intronic SNPs detected in this study were in introns 1 and 2 which might be linked to the Rhce change to RhCe which is encoded by c.307C>T in exon 2.

4.4.3 Rh Haplotype Determination Using Intronic Variations

Intronic SNPs in specific Rh haplotypes represent a novel approach in determining Rh haplotypes. These polymorphisms could be targeted through PCR to guide in Rh allele determination. For example, Rh phenotype D+C+c+E-e+ could be assigned by serology as $\mathrm{R_2r}$ (DCe/dce) or $\mathrm{R_2R_0}$ (DCe/Dce). Serology issues, as in this example, are resolved using the frequency of Rh alleles in the population (Daniels, 2013b), which mostly underestimate the rare genotypes or overestimate the most common alleles as we found in some samples that were assigned by serology as $\mathrm{R_1R_2}$ but found to be hemizygous for the *RHD* gene (Sillence et al., 2017), (section 3.3.1).

Different Rh alleles could be present in similar frequencies in a specific population. For example, in Black Africans the homozygous (Dce/Dce) and heterozygous (Dce/dce) haplotypes occur at a similar rate (Kemp et al., 1999). Therefore, assigning these haplotypes using serology alone could lead to incorrectly assigned haplotype. ddPCR has been proven to be an accurate tool in determining *RHD* zygosity (Sillence et al., 2017) which enable the differentiation between such haplotypes (Dce/dce and Dce/Dce).

Targeting intronic repeat regions, by PCR, like (GCAC) and (AC) in intron

2 in samples with c+ antigen (Kemp et al., 1999) and the 109 bp insertion in intron 2 in samples with C+ antigen could resolve such issues and enable better *RHCE* genotyping. Zhou et al. (2008) performed *RHCE* genotyping, through targeting 109 bp intron 2 insertion in the *RHCE* gene and c.48C>G polymorphism in exon 1, for 320 samples by PCR-SSP. Genotyping by targeting more than one polymorphism in the gene, including intronic ones, provides accurate results and eliminates the possibility of false positives or false negatives (Zhou et al., 2008).

In this study we utilised allele specific intronic SNPs in conjunction with serology and *RHD* zygosity results to reassign the Rh genotype for samples that were incorrectly assigned their Rh haplotypes by serology, see section 3.4.7.

For 10 ISBT samples, Rh haplotype was determined (Table 4.4) by utilising *RHD* and *RHCE* genotyping results, *RHD* zygosity testing results, and intronic SNP patterns present in the *RHD* and the *RHCE* genes. Targeting intronic SNPs has proved here to be effective in support of the determination of the Rh haplotype for these samples.

4.4.4 Primer Specificity and Breakage of Coverage

RHCE specific primers were used to amplify the gene for sequencing. Although primers were assessed to ensure their specificity to the target gene using Primer-BLAST on the NCBI website (Altschul et al., 1997), regions

around exons 2 and 8 were very challenging to amplify. Multiple primer pairs were designed for these two regions and were tested and optimised using various annealing temperatures and primer concentrations. Most primers designed amplified samples with the R_2 haplotype but not samples with R_1 haplotype, which shows how different the sequence allele in that region for both haplotypes may be. A similar issue of PCR allele discrimination was noticed in ABO blood group gene genotyping using NGS (Altayar, 2016). To overcome this issue, smaller size amplicons were designed to target these regions. Nevertheless, breakage of coverage in intron 8 and intron 2 in some samples was still noticed.

4.5 Conclusion

In this work, we were able to sequence the *RHCE* gene using NGS from samples (n=100) with different Rh haplotypes. The *RHCE* alleles were identified in 95 samples. Due to low coverage in specific areas of the *RHCE* gene, *RHCE* alleles were not identified in 5 samples. Nevertheless, we were able to identify the *RHCE* alleles in most samples with adequate coverage, over all coding exons and most introns, to call variants.

Through complete NGS of the *RHCE* gene, we were able to study intronic mutations in the *RHCE* gene and their relation to a specific Rh allele. 89 intronic SNPs (Table 4.5) were found in samples assigned their genotype by serology as R_1 and R_z, which lead to the conclusion that

these intronic changes are RhC specific. Twenty-two intronic SNPs (Table 4.6) were found in samples assigned their genotype by serology as R_2, R_0, r , and r", which lead to the conclusion that these SNPs are Rhc specific. These SNPs were not affected by the presence or absence of the *RHD* gene. Fully sequencing the *RHCE* using NGS has led to the establishment of three *RHCE* allele specific reference sequences, in which one is for the RhCe allele, one for the Rhce allele, and one for the RhcE allele.

5. General Discussion and Future Work

5.1 General Discussion

5.1.1 Advantages of NGS in BGG

One of the main aims of pre-transfusion testing is to minimise risk associated with transfusion, whether that is coming from blood transmitted infections or immunological reaction due to blood group antigens incompatibility. The first advancement of genomic testing was put into use to screen blood donors for human immunodeficiency virus (HIV) and hepatitis C virus through nucleic acid testing (NAT) (Hyland et al., 2019). Genomics testing was also introduced for HLA typing, generating a strong impetus to use genomic testing for BGG (Möller et al., 2016).

Routine serological testing for blood group antigen phenotyping and compatibility testing has been used in transfusion facilities for decades (Hyland et al., 2019). Serological testing provides the ability to screen for major blood groups such as ABO and Rh expeditiously. However, identification

of blood group antigens polymorphisms remains an issue that is beyond serological testing ability.

The most important blood groups, ABO and Rh, are the most challenging regarding genomic genotyping and they can inexpensively and efficiently be phenotyped by serological testing (Storry et al., 2019). However, Rh is the most polymorphic blood group system with 55 antigens which could be difficult to identify and easily mis-phenotyped by serology alone, therefore, BGG has an important role to fulfil and ensure correct genotyping when serology fails to deliver accurate results. In this work, NGS was utilised on Ion PGM™ platform to fully sequence the *RHD*, *RHAG*, and *RHCE* genes from samples with different Rh phenotypes.

The results of the use of NGS in BGG to overcome limitations in other genotyping platforms have been promising (Fichou et al., 2016; Chou et al., 2017; Dezan et al., 2017; Tounsi et al., 2018). Exome sequencing utilizing NGS or conventional sequencing (Sanger) could have been used to identify variants; however, exome sequencing cannot identify intronic variations such as intronic insertion associated with *RHD* partial alleles (Chou et al., 2017). NGS allows the use of LR-PCR approach, enabling the amplification of the *RHD* gene (<60 kb) in only 6 overlapping amplicons about 10 kb in size. On the other hand, Sanger sequencing can sequence DNA fragments up to 1 kb (Zhang et al., 2011). Therefore, complete sequencing of the *RHD* gene using NGS was the preferred method, which provided

a complete sequence of the gene in most samples with high coverage that allowed accurate variants calling. Using NGS over Sanger sequencing also enables the sequencing of different blood group genes simultaneously allowing NGS to be used for BGG in reference laboratories to identify blood group variants (Montemayor-Garcia and Westhoff, 2018). Fichoua et al. (2019) utilised LR-PCR amplicons and NGS to establish allele reference sequences for the *ACKR1* gene encoding the Duffy antigens.

We were able to design target specific primers and optimise LR-PCR conditions to amplify genes in multiple overlapping amplicons for NGS. This led to identifying novel *RHD* alleles, establishment of two *RHD* allele reference sequences, one for the $\mathrm{R}_1, \mathrm{R}_0, \mathrm{R}_\mathrm{z}$ haplotypes and one for the R_2 (DcE) haplotype, and establishment of three *RHCE* allele reference sequences for the *RHCE*Ce* allele, the *RHCE*cE* allele, and the *RHCE*ce* allele.

Complete sequencing of the *RHD*, *RHAG*, and *RHCE* genes using NGS generated adequate depth of coverage that was efficient in detecting exonic and intronic changes, enabling the identification of the *RHD*, *RHAG*, and *RHCE* alleles in most samples. For a few samples where complete sequence of the target gene was not achieved due to amplification failure in one or more of the LR-PCR amplicons, the *RHD* or *RHCE* allele was not determined. Amplification through LR-PCR remains one of the limitation of NGS.

Identification of novel variants of blood group genes would not be possible using current BGG platforms such as HEA BeadChip™ (BioArray Solutions Ltd) (Hashmi et al., 2005, 2007; Kappler-Gratias et al., 2011), BLOODChip® (Progenika) (Avent et al., 2007), Genome Lab SNP Stream (Beckman Coulter) (Denomme and Van Oene, 2005), ID CORE XT (López et al., 2018), or HIFI Blood 96™ (Boccoz et al., 2016) as they are designed to target specific SNV that encode specific variant alleles (Liu et al., 2014; Finning et al., 2016; Orzińska et al., 2018), hence, novel variants remain unidentified. The Finnish samples are a good example of that, where novel *RHD* allele variants were not determined using SSP-PCR-PCR testing (Tammi, 2019). Sequencing 35 Finnish samples revealed the presence of 8 novel *RHD* variant alleles that could be specific to the Finnish population. Two other *RHD* novel variants were found in blood donor samples. These two novel alleles are possibly unique to the United Kingdom population.

Identification of novel variants and their genetic background is crucial in expanding our knowledge of blood group gene polymorphisms. This knowledge will aid in developing new BGG assays that rely on high-throughput sequencing. It will also help in updating the current BGG assays, that depend on targeting blood group gene SNPs, to include more blood group gene SNPs. The use of NGS in BGG will supplement serological testing and BGG molecular assays.

The main question asked frequently is when will serology be replaced by molecular genotyping. The simplest answer to that question is this might not happen in any near future, although many will disagree (Avent et al., 2015; Boccoz et al., 2018). Serological phenotyping is efficient, cost effective, and accurate for the most important blood groups like ABO and Rh that are tested routinely in transfusion facilities. However, serology testing alone is just not enough. Some cases are beyond serological testing capabilities such examples include but not limited to ABO discrepancy, cases of multiple antibodies, and inconclusive Rh phenotyping. Serological testing will always be needed to assess blood group antigen status on the RBCs.

Current scenario in blood banks is that samples are serologically phenotyped for most clinically significant blood group antigens (such as ABO and Rh) and when a variant allele is suspected, the sample could be genotyped using the available BGG assay. If results remain inconclusive, then WGS or WES would be the best option for blood group gene allele determination utilising NGS. Serological screening for blood group antigen phenotyping remains an invaluable test in transfusion facilities.

NGS is a high throughput technique that generates a vast amount of data which allows exploration of genetic variations and detection of novel alleles. It is appropriate to consider NGS as a confirmation or identification test when serological testing or other genotyping platforms fail to provide

compatible results (Fürst et al., 2020), as in the case of the Finnish samples where serology and SSP-PCR results were incompatible due to the presence of novel *RHD* variants which were identified after *RHD* NGS.

BGG has been beneficial for patients with chronic illness who are transfusion dependent to lower the risk of immunisation (Bakanay et al., 2013; Dezan et al., 2017). BGG will enable the establishment of blood donor databases of fully genotyped donors that would make finding compatible donors easier, especially in the case of a rare blood group antigen or in the presence of antibodies to a high frequency antigen (Perreault et al., 2009; St-Louis et al., 2010). To justify the genotyping cost for donors, it is important to genotype frequent donors (Jungbauer, 2011b).

In this work we utilised NGS to sequence the Rh genes, *RHD* and *RHCE*, and established unique allele specific reference sequences by studying intronic changes and their relation to a specific Rh haplotype. In the *RHD* gene, it is clear that there is a significant difference between the R_2 (DcE) haplotype and other haplotypes R_1 (DCe), R_0 (Dce), R_z (DCE) in intronic changes that lead to establishing two *RHD* allele specific reference sequences. The reference sequences were registered with Genbank with accession numbers MG944308 for the R_1, R_0, R_z haplotypes and MG944309 for the R_2 haplotype.

In the *RHCE* gene, intronic SNP analysis led to identifying intronic changes

unique to a specific *RHCE* allele. These intronic changes allowed the establishment of 3 *RHCE* allele specific reference sequences, in which one was for the *RHCE*Ce* allele, one for the *RHCE*cE* allele, and one for the *RHCE*ce* allele that were submitted to Genbank and registered with accession numbers MN091965, MN091966, and MN624142, respectively.

Intronic SNP patterns identified in both the *RHD* and *RHCE* genes might be evidence of their evolutionary rise from a single haplotype (R_0) (Carritt et al., 1997; Kemp et al., 1999), (section 1.4). *RHD* and *RHCE* intronic SNPs can be targeted by PCR to distinguish between different Rh haplotypes that cannot be differentiated from each other using serology. For example, the *RHD* allele cannot be distinguished between $\mathrm{R}_1\mathrm{r}''$ (DCe/dcE) and $\mathrm{R}_2\mathrm{r}'$ (DcE/dCe). Targeting R_2 specific intronic SNPs by PCR can differentiate between the R_2 allele and other *RHD* alleles.

5.1.2 SMS in BGG

SMS, also known as TGS, is a sequencing approach that will change genomics and researchers ability to sequence and analyse genomes of any kind. In this work, SMS was utilised through MinION™ sequencing for *RHD* genotyping. MinION™ is a USB portable device that generates 400 megabases of data in a 48-h run and uses reads that can exceed 100,000 bp in length (Karamitros and Magiorkinis, 2015).

Thirteen samples were sequenced and results were compared to the ones

obtained by Ion PGM™. In this work, *RHD* genotyping using MinION™ proved to be successful and alleles determined agreed with the ones identified using NGS. The advantages of using MinION™ over NGS are the faster library preparation and sequencing time and the sequencing of longer reads that would allow of better assembly (Loman et al., 2015). Although LR-PCR amplicons were used to amplify the *RHD* gene for sequencing for both platforms, direct enrichment of any target gene will be possible though biotinylated PCR-generated baits that allow capturing the targeted gene for sequencing (Karamitros and Magiorkinis, 2015).

Currently, there are numerous published papers in scientific journals about the utilisation of MinION™ and data handling and analysis. However, most of these papers focus on micro-organisms genome assembly and analysis (Quick et al., 2016; Faria et al., 2017). The human genome is larger and far complex therefore more work is needed to explore the potential power of this approach in human genome sequencing and analysis to improve sequencing accuracy and develop user-friendly interfaces for data analysis (Magi et al., 2016).

5.1.3 Study Limitations

One of the limitations of high throughput sequencing is the challenge introduced by the vast amount of data generated and issues associated with that such as management of data analysis, data storage, maintenance of

confidentiality, and data ownership (McBean et al., 2014; Tilley and Grimsley, 2014). To overcome some of these challenges, optimisation of bioinformatic tools for data analysis, storage and handling are crucial for future development of BGG using high-throughput sequencing. The amount of data generated from high throughput BGG cannot be logged-in, handled and analysed manually as it will require a long time, a vast amount of effort and is prone to human error; therefore, computerised large-scale genotyping databases are necessary for the future of BGG (Denomme, 2013). Perreault et al. (2009) designed a blood donor database programme that enabled sequencing of 10 555 blood donors for 22 blood group antigens. The programme allowed the generation of barcodes, the storing of genotyping results, finding compatible donors and checking for the availability of components. Clear guidelines need to be established to address legal and ethical challenges in high-throughput sequencing including data protection, privacy, and ownership (Godard et al., 2003; Tilley and Grimsley, 2014; Naveed et al., 2015).

Cost is a major factor in determining the feasibility of applying BGG using NGS. Although cost of sequencing has decreased substantially in recent years, it remains high (McBean et al., 2014) when compared to other BGG assays. Different studies have aimed to estimate the cost of blood transfusion /unit and then breakdown the cost into different areas (Forbes et al., 1991; Shander et al., 2007; Leahy and Mukhtar, 2012; Mazonson et al., 2014;

Stubbs, 2014). Mazonson et al. (2014) calculated the average direct cost of serological testing for antibody-positive patients in four medical centres in the United States, which was $114 per work up and $195 per patient. Serological phenotyping mean cost was about $20 per antigen. Many of the serological testing techniques used to assess blood group antibodies are costly and time consuming, for example, a cell panel costs from $33 to $55 depending on the number of cells in the panel (Mazonson et al., 2014). Therefore, the use of BGG for patients with chronic illnesses will benefit the most from such an advancement to limit, if not prevent, allosensitisation.

Quirino et al. (2019) compared the cost and effectiveness of BGG different microarray platforms including, SSP-PCR, restriction fragment length polymorphism-polymerase chain reaction (RFLP-PCR) and compared results to serological testing (centrifugation gel method) for RHCE, Kidd, Kell, and Duffy blood group systems. They tested 16 samples and showed that serological testing for target blood group antigens cost $32.45 per sample and results were achieved in about 25 minutes, SSP-PCR cost $18.07 per sample and results were available in 3 hours and 20 minutes, and RFLP-PCR cost $38.22 per sample and testing took 19 hours and 30 minutes. Therefore, SSP-PCR is the most economically effective and serological testing is the most time efficient technique.

NGS library preparation and sequencing on Ion PGM™ in this work was

calculated for 20 samples from PCR amplification to the actual sequencing run to be £4160, which is £208 per sample for one gene in addition to the cost of DNA extraction, which is £2.50 per sample using the QIAamp DNA Blood Mini kit (Qiagen Ltd, United Kingdom). When compared, NGS cost per sample, to serological phenotyping cost, $20 per antigen, the huge difference in cost of $188 shows the limitation of NGS affordability.

NGS sequencing is also limited by the considerably long preparation time including PCR, library preparation, enrichment, sequencing, and data analysis. DNA library preparation for NGS is a time-consuming process that requires a lot of manual handling (Tilley and Grimsley, 2014). In this work, manual library preparation from PCR amplification to the point of data analysis, for 20 samples, required about 10 working days. Processing time for NGS is now decreasing due to the introduction of semi automated library preparation machines which are crucial to enable NGS-based BGG (Avent et al., 2015; McBean et al., 2014; Tilley and Grimsley, 2014). For example, SPRIworks Fragment Library Systems (Beckman Coulter, UK) carry out library preparation (size selection and purification) for three NGS platforms including for the Illumina Genome Analyzers, Roche GS FLX DNA Sequencer and Life Technologies SOLiD Sequencers. However, this system processes only 10 samples at a set-up and takes 3-4 hours. Ion Chef™ Instrument (Thermo Fisher Scientific) was introduced to simplify the Ion PGM™ library handling and chip loading and to eliminate incon-

sistency associated with manual handling.

Nevertheless, the whole process of preparation and handling, from DNA extraction to data analysis, is considerably long and cannot compete with other BGG platforms (Tilley and Grimsley, 2014) such as microarrays based tests. However, NGS can be used for selected samples where commercially available BGG assays fail to deliver accurate results. BGG should be updated regularly by adding new blood group gene variants to allow the detection of a wider range of variant alleles.

Another issue faced using NGS for BGG is the homology between some of the blood group genes, such as the *RHD* and *RHCE* genes of the Rh blood group system and the *gyp* genes of the MNS blood group system (Tilley and Grimsley, 2014). Homology levels affect primer specificity as we have had in this study and aligning data could be difficult and complicated. We had difficulties amplifying from hybrid *RHD* genes, see section 3.4.11.

The *RHD* and *RHCE* genes were fully sequenced on the Ion PGM™ through LR-PCR amplification in six and eight overlapping amplicons, respectively. Although LR-PCR is an efficient technique in amplifying target genes for sequencing, the LR-PCR approach is limited. Hybrid *RHD-RHCE* alleles or partial alleles may not amplify if a primer position is compromised by deletion or mutations. For example, some of the *RHD* and *RHCE* primers failed to amplify from samples suspected to have hybrid genes such as

*RHD*DVI.01* samples. This could be resolved using hybrid pairs of primers; for example, an *RHD* specific forward primer and an *RHCE* specific reverse primer or vice versa.

Another issue we faced with the LR-PCR approach was primers discriminating between alleles of the *RHCE* gene, specifically in the region of intron 8. Amplification of these two regions of the *RHCE* gene was very challenging as most of the primers designed only amplified samples with R_2 alleles and failed to amplify other alleles like R_1. Nevertheless of our attempt to design smaller fragments for both regions, low and breakage of coverage still persisted in a region of intron 8 (25,378,316- 25,381,617).

5.2 Future Work

Regarding *RHD* genotyping, including different *RHD* variant alleles would enable the assessment of the encoding sequences and the analysis of their intronic SNPs. Optimising a PCR multiplex protocol for the *RHD* gene, would cut time and labour from library preparation, which would allow testing more samples at ounce. Current *RHD* primers may fail to amplify hybrid *RHD-RHCE-RHD* alleles or partial D alleles may not amplify if a primer position is compromised by deletion or mutations, as observed in some samples. Therefore, designing primers that target hybrid alleles would be useful to sequence these variant alleles by using a hybrid primer approach; for example, an *RHD* specific forward primer and an *RHCE* spe-

cific reverse primer.

For *RHCE* genotyping, future work includes optimising primers for intron 8 to acquire an adequate coverage over that region and eventually optimise a PCR multiplex reaction. Eventually, optimising a PCR multiplex reaction for both Rh genes together would enable fast and efficient amplification and thus faster sequencing of both genes.

MinION™ sequencing of the *RHD* gene and allele determination for all samples sequenced was accomplished successfully. Establishment of a user friendlily software for data assembly and analysis would enable the use of MinION™ more widely in BGG. With future advancement in MinION™, it would enable SMS BGG using a PCR free sequencing approach.

For *RHD* zygosity testing, designing a new target in the *RHD* gene is necessary that is not commonly affected in *RHD-RHCE* hybrid alleles. In this study, *RHD* exons 5 and 7 were targeted for zygosity testing using ddPCR. A few samples presented a deletion in exon 5, mostly affected by a conversion event in the *RHD* gene. Therefore, including more targets for testing is necessary to increase confidence in zygosity results. *RHD* exons 2 and 9 are suitable candidates for testing as they are less affected by deletion and conversion.

In this work, blood group genes including *RHD*, *RHCE*, and *RHAG* were sequenced using NGS (Ion PGM™) to study variation in the population

which led to the establishment of two Rh allele specific reference sequences for the *RHD* gene and three for the *RHCE* gene through intronic variation analysis of both genes. Complete sequencing of the *RHD* gene also resulted in the identification of 10 novel variants of the *RHD* gene that cause weakened and negative expression of the RhD protein. The *RHD* gene was also sequenced using SMS (MinION™) which was successful in determining the *RHD* alleles in all samples sequenced.

References

Altayar, M. A. (2016). *Next Generation Sequencing-Based Genotyping of Human Blood Groups: FY, JK and ABO Genes.* Ph. D. thesis, University of Plymouth.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research 25*(17), 3389–3402.

Anderson, C. C. and N. R. Sinclair (1998). FcR-mediated inhibition of cell activation and other forms of coinhibition. *Critical Reviews in Immunology 18*(6), 525–44.

Anderson, M. and I. Schrijver (2010). Next generation DNA sequencing and the future of genomic medicine. *Genes 1*(1), 38–69.

Araszkiewicz, P. and I. O. Szymanski (1989). Quantitative studies on the D antigen of red cells with the D^u phenotype. *Transfusion 29*(2), 103–105.

Ardui, S., A. Ameur, J. R. Vermeesch, and M. S. Hestand (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Research, 46*(5), 2159–2168.

Avent, N. D. (2018). *Transfusion and Transplantation Science* (2 ed.). Fundamentals of biomedical science. Oxford University Press.

Avent, N. D., W. Liu, J. W. Jones, M. L. Scott, D. Voak, J. W. M Pisacka, and A. Fletcher (1997). Molecular analysis of Rh transcripts and polypeptides from individuals ex-

pressing the DVI variant phenotype: An *RHD* gene deletion event does not generate all DVIccEe phenotypes. *Blood 89*(5), 1779–1786.

Avent, N. D., T. E. Madgett, A. J. Halawani, M. Altayar, M. Kiernan, A. Reynolds, and X. Li (2015). Next generation sequencing: academic overkill or high resolution routine blood group genotyping. *ISBT Science Series 10*(1), 250–256.

Avent, N. D., T. E. Madgett, Z. E. Lee, D. J. Head, D. G. Maddocks, and L. H. Skinner (2006). Molecular biology of Rh proteins and relevance to molecular medicine. *Expert Reviews in Molecular Medicine 8*(13), 1–20.

Avent, N. D., A. Martinez, W. A. Flegel, M. L. Olsson, M. L. Scott, N. Nogués, M. Píšacka, G. L. Daniels, E. Muñiz-Diaz, T. E. Madgett, J. R. Storry, S. Beiboer, P. M. Maaskant-van Wijk, I. von Zabern, E. Jiménez, D. Tejedor, M. López, E. Camacho, G. Cheroutre, A. Hacker, P. Jinoch, I. Svobodova, E. van der Schoot, and M. de Haas (2009). The BloodGen project of the European Union, 2003-2009. *Transfusion Medicine and Hemotherapy 36*(3), 162–167.

Avent, N. D., A. Martinez, W. A. Flegel, M. L. Olsson, M. L. Scott, N. Nogués, M. Písăcka, G. Daniels, E. Van Der Schoot, E. Muñiz-Diaz, T. E. Madgett, J. R. Storry, S. H. Beiboer, P. A. Maaskant-van Wijk, I. Von Zabern, E. Jiménez, D. Tejedor, M. López, E. Camacho, G. Cheroutre, A. Hacker, P. Jinoch, I. Svobodova, and M. de Haas (2007). The BloodGen project: toward mass-scale comprehensive genotyping of blood donors in the European union and beyond. *Transfusion 47*(Suppl 1), 40S–46S.

Avent, N. D. and M. E. Reid (2000). The Rh blood group system: a review. *Blood 95*(2), 375–387.

Avent, N. D., K. Ridgwell, M. J. Tanner, and D. J. Anstee (1990). cDNA cloning of a 30 kDa erythrocyte membrane protein associated with Rh (Rhesus)-blood-group-antigen expression. *Biochemical Journal 271*(3), 821–825.

Bakanay, S. M., A. Ozturk, T. Ileri, E. Ince, S. Yavasoglu, N. Akar, Z. Uysal, and O. Arslan (2013). Blood group genotyping in multi-transfused patients. *Transfusion and Apheresis Science 48*(2), 257–261.

Barnhart, M. M. and M. R. Chapman (2006). Curli biogenesis and function. *Annual Review of Microbiology 60*, 131–147.

Beal, C. L., C. K. Oliver, D. M. Mallory, and P. D. Issitt (1995). The r^s gene is over represented in hrB-negative individuals. *Immunohematology 11*(3), 74–77.

Bhutani, V. K., A. Zipursky, H. Blencowe, R. Khanna, M. Sgro, F. Ebbesen, J. Bell, R. Mori, T. M. Slusher, N. Fahmy, V. K. Paul, L. Du, A. A. Okolo, M.-F. de Almeida, B. O. Olusanya, P. Kumar, S. Cousens, and J. E. Lawn (2013). Neonatal hyperbilirubinemia and Rhesus disease of the newborn: incidence and impairment estimates for 2010 at regional and global levels. *Pediatric Research 74*(Suppl 1), 86–100.

Blanco, S., V. S. Giacomi, L. Slobodianiuk, M. C. Frutos, L. H. Carrizo, G. E. Fanin, J. M. Culasso, and S. V. Gallego (2018). Usefulness of non-invasive fetal *RHD* genotyping towards immunoprophylaxis optimization. *Transfusion Medicine and Hemotherapy 45*(6), 423–428.

Boccoz, S. A., J. Fouret, M. Roche, J. Lachuer, C. Legras-Lachuer, B. P. Corgier, and C. A. Marquette (2018). Massively parallel and multiplex blood group genotyping using next-generation-sequencing. *Clinical Biochemistry 60*, 71–76.

Boccoz, S. A., G. C. L. Goff, C. A. Mandon, B. P. Corgier, L. J. Blum, and C. A. Marquette (2016). Development and validation of a fully automated platform for extended blood group genotyping. *The Journal of Molecular Diagnostics 18*(1), 144–152.

Bowman, J. M. (1988). The prevention of Rh immunization. *Transfusion Medicine Reviews 2*(3), 129–150.

Bowman, J. M. (1998). RhD hemolytic disease of the newborn. *The New England Journal of Medicine 339*(24), 1775–1777.

Branton, D., D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, and J. A. Schloss (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology 26*, 1146–1153.

Bruce, L. J., R. Beckmann, M. L. Ribeiro, L. L. Peters, J. A. Chasis, J. Delaunay, N. Mohandas, D. J. Anstee, and M. J. Tanner (2003). A band 3-based macrocomplex of integral and peripheral proteins in the RBC membrane. *Blood 101*(10), 4180–4188.

Bruce, L. J., H. Guizouarn, N. M. Burton, N. Gabillat, J. Poole, J. F. Flatt, R. L. Brady, F. Borgese, J. Delaunay, and G. W. Stewart (2009). The monovalent cation leak in overhydrated stomatocytic red blood cells results from amino acid substitutions in the Rh-associated glycoprotein. *Blood 113*(6), 1350–1357.

Bub, C. B. and L. Castilho (2019). ID CORE XT as a tool for molecular red blood cell typing. *Expert Review of Molecular Diagnostics 19*(9), 777–783.

Carritt, B., T. J. Kemp, and M. Poulter (1997). Evolution of the human RH (Rhesus) blood group genes: a 50 year old prediction (partially) fulfilled. *Human Molecular Genetics 6*(6), 843–850.

Ceppellini, R., L. Dunn, and M. Turri (1955). An interaction between alleles at the Rh locus in man which weakens the reactivity of the Rh_0 factor (Du). *Proceedings of the National Academy of Sciences of the United States of America 41*(5), 283–288.

Chang, H. E., Y. J. Hong, H. Kim, S. M. Hwang, J. S. Park, S. W. Lee, E. Y. Song, K. U. Park, J. Song, and K. S. Han (2016). Genotyping of 19 red cell antigens, including *RHD*, using liquid bead arrays. *Clinica Chimica Acta 462*, 111–117.

Chen, C.-Y. (2014). DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Frontiers in Microbiology 5*(305), 1–11.

Chou, S. T., J. M. Flanagan, S. Vege, N. L. C. Luban, R. C. Brown, R. E. Ware, and C. M. Westhoff (2017). Whole-exome sequencing for RH genotyping and alloimmunization risk in children with sickle cell anemia. *Blood Advances 1*(18), 1414–1422.

Clarke, K. (2016). The use of intronic *RHD* single nucleotide polymorphisms in determination of *RHD* haplotype. Undergraduate dissertation, University of Plymouth.

Clausen, F. B. (2018). Lessons learned from the implementation of non-invasive fetal *RHD* screening. *Expert Review of Molecular Diagnostics 18*(5), 423–431.

Clausen, F. B., M. Christiansen, R. Steffensen, S. Jørgensen, C. Nielsen, M. A. Jakobsen, R. D. Madsen, K. Jensen, G. R. Krog, K. Rieneck, U. Sprogøe, K. Homburg, N. Grunnet, and M. H. Dziegiel (2012). Report of the first nationally implemented clinical routine screening for fetal *RHD* in D-pregnant women to ascertain the requirement for antenatal RhD prophylaxis. *Transfusion 52*(4), 752–758.

Colin, Y., B. Chérif-Zahar, C. L. V. Kim, V. Raynal, V. V. Huffel, and J. Cartron (1991). Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis. *Blood 78*(10), 2747–2752.

Conroy, M. J., P. A. Bullough, M. Merrick, and N. D. Avent (2005). Modelling the human Rhesus proteins: implications for structure and function. *British Journal of Haematology 131*(4), 543–551.

Coombs, R. R., A. E. Mourant, and R. R. Race (1946). In-vivo isosensitisation of red cells in babies with haemolytic disease. *The Lancet 1*, 264–266.

Coombs, R. R. and F. Roberts (1959). The antiglobulin reaction. *British Medical Bulletin 15*(2), 113–118.

Coombs, R. R. A., A. E. Mourant, and R. R. Race (1945). Detection of weak and "incomplete" Rh agglutinins: a new test. *The Lancet 246*(6358), 15–16.

Coulter, B. (2012). SPRIselect user guide [Online]. Available at www.beckmancoulter.com (accessed on 06.09.2016). Technical report, Beckman Coulter.

Daniels, G. (2005). The molecular genetics of blood group polymorphism. *Transplant Immunology 14*(3), 143–153.

Daniels, G. (2013a). *Low Frequency Antigens*, Chapter 29, pp. 495–499. John Wiley Sons, Ltd.

Daniels, G. (2013b). *Rh and RHAG blood group systems. In: Daniels G, ed. Human Blood Groups* (3 ed.). Wiley-Blackwell.

Daniels, G. (2013c). Variants of RhD – current testing and clinical consequences. *British Journal of Haematology 161*(4), 461–470.

Daniels, G. L., B. Faas, C. A. Green, E. Smart, P. A. Maaskant-van Wijk, N. D. Avent, H. A. Zondervan, A. E. von dem Borne, and C. E. van der Schoot (1998). The VS and V blood group polymorphisms in Africans: a serologic and molecular analysis. *Transfusion 38*(10), 951–958.

Daniels, G. L., A. Fletcher, G. Garratty, S. Henry, J. Jørgensen, W. J. Judd, C. Levene, C. Lomas-Francis, J. J. Moulds, J. M. Moulds, M. Moulds, M. Overbeeke, M. E. Reid, P. Rouger, M. Scott, P. Sistonen, E. Smart, Y. Tani, S. Wendel, T. Zelinski, and I. S. of Blood Transfusion. (2004). Blood group terminology 2004: from the international society of blood transfusion committee on terminology for red cell surface antigens. *Vox Sanguinis 87*(4), 304–316.

de Haas, M., F. F. Thurik, J. M. Koelewijn, and C. E. van Der Schoot (2015). Haemolytic disease of the fetus and newborn. *Vox Sanguinis 109*(2), 99–113.

de Haas, M., F. F. Thurik, C. P. B. van der Ploeg, B. Veldhuisen, H. Hirschberg, A. A. Soussan, H. Woortmeijer, F. Abbink, G. C. M. L. Page-Christiaens, P. G. Scheffer, and C. E. van der Schoot (2016). Sensitivity of fetal *RHD* screening for safe guidance of targeted anti-D immunoglobulin prophylaxis: prospective cohort study of a nationwide programme in the netherlands. *The BMJ 355*(i5789).

De Santis, D., L. Truong, P. Martinez, and L. D'Orsogna (2020). Rapid high-resolution hla genotyping by minion oxford nanopore sequencing for deceased donor organ allocation. *HLA 96*(2), 141–162.

Denomme, G. A. (2013). Prospects for the provision of genotyped blood for transfusion. *British Journal of Haematology 163*(1), 3–9.

Denomme, G. A. and M. Van Oene (2005). High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. *Transfusion 45*(5), 660–666.

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie, and O. Gascuel (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research 1*(2), W465–W469.

Dezan, M. R., I. H. Ribeiro, V. B. Oliveira, J. B. Vieira, F. C. Gomes, L. A. M. Franco, L. Varuzza, R. Ribeiro, K. Z. Chinoca, J. E. Levi, J. E. Krieger, A. C. Pereira, S. F. M. Gualandro, V. G. Rocha, A. Mendrone-Junior, E. C. Sabino, and C. L. Dinardo (2017). *RHD* and *RHCE* genotyping by next-generation sequencing is an effective strategy to identify molecular variants within sickle cell disease patients. *Blood Cells, Molecules, and Diseases 65*, 8–15.

Dinardo, C. L., S. L. Bonifácio, and A. J. Mendrone (2014). Indirect antiglobulin test-crossmatch using low-ionic-strength saline-albumin enhancement medium and re-

duced incubation time: effectiveness in the detection of most clinically significant antibodies and impact on blood utilization. *Immunohematology. 30*(1), 1–5.

Dunstan, R. A. (1986). The expression of ABH antigens during in vitro megakaryocyte maturation: origin of heterogeneity of antigen density. *British Journal of Haematology 62*(3), 587–593.

Eisenstein, M. (2017). An ace in the hole for DNA sequencing. *Nature 550*(7675), 285–288.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green (1998). Base-calling of automated sequencer traces using Phred.i. accuracyassessment. *Genome Research 8*(3), 175–185.

Faas, B. H. W., E. A. M. Beckers, P. Wildoer, P. C. Ligthart, M. A. M. Overbeeke, H. A. Zondervan, A. E. G. K. von dem Borne, and C. van der Schoot (1997). Molecular background of VS and weak C expression in Blacks. *Transfusion 37*(1), 38–44.

Faas, B. H. W., E. A. Beuling, P. C. Ligthart, D. J. V. Rhenen, and C. E. V. D. Schoot (2001). Partial expression of Rhc on the RhD polypeptide. *Transfusion 41*(9), 1136–1142.

Fan, H. C., Y. J. Blumenfeld, Y. Y. El-Sayed, J. Chueh, and S. R. Quake (2009). Microfluidic digital PCR enables rapid prenatal diagnosis of fetal aneuploidy. *American Journal of Obstetrics & Gynecology 200*(5), 543.e1–7.

Faria, N. R., J. Quick, I. M. Claro, J. Thézé, J. G. de Jesus, M. Giovanetti, M. U. G. Kraemer, S. C. Hill, A. Black, A. C. da Costa, L. C. Franco, S. P. Silva, C. H. Wu, J. Raghwani, S. Cauchemez, L. du Plessis, M. P. Verotti, W. K. de Oliveira, E. H. Carmo, G. E. Coelho, A. C. F. S. Santelli, L. C. Vinhal, C. M. Henriques, J. T. Simpson, M. Loose, K. G. Andersen, N. D. Grubaugh, S. Somasekar, C. Y. Chiu, J. E. Muñoz-Medina, C. R. Gonzalez-Bonilla, C. F. Arias, L. L. Lewis-Ximenez, S. A. Baylis, A. O. Chieppe, S. F. Aguiar, C. A. Fernandes, P. S. Lemos, B. L. S. Nascimento, H. A. O. Monteiro, I. C. Siqueira, M. G. de Queiroz, T. R. de Souza, J. F. Bezerra, M. R. Lemos, G. F. Pereira, D. Loudal, L. C.

Moura, R. Dhalia, R. F. França, T. Magalhães, E. T. Marques Jr, T. Jaenisch, G. L. Wallau, M. C. de Lima, V. Nascimento, E. M. de Cerqueira, M. M. de Lima, D. L. Mascarenhas, J. P. M. Neto, A. S. Levin, T. R. Tozetto-Mendoza, S. N. Fonseca, M. C. Mendes-Correa, F. P. Milagres, A. Segurado, E. C. Holmes, A. Rambaut, T. Bedford, M. R. T. Nunes, E. C. Sabino, L. C. J. Alcantara, N. J. Loman, and O. G. Pybus (2017). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature 546*(7658), 406–410.

Fichou, Y., M.-P. Audrézet, P. Guéguen, C. L. Maréchal, and C. Férec (2014). Next-generation sequencing is a credible strategy for blood group genotyping. *British Journal of Haematology 167*(4), 554–562.

Fichou, Y., J.-M. Chen, C. L. Maréchal, D. Jamet, I. Dupont, C. Chuteau, C. Durousseau, M.-J. Loirat, P. Bailly, and C. Férec (2012). Weak D caused by a founder deletion in the *RHD* gene. *Transfusion 52*(11), 2348–2355.

Fichou, Y. and C. Férec (2017). NGS and blood group systems: State of the art and perspectives. *Transfusion Clinique et Biologique 24*(3), 240–244.

Fichou, Y., M. Mariez, C. Le Maréchal, and C. Férec (2016). The experience of extended blood group genotyping by next-generation sequencing (NGS): investigation of patients with sickle-cell disease. *Vox Sanguinis 111*(4), 418–424.

Fichoua, Y., I. Berlivet, G. Richard, C. Tournamille, L. Castilho, and C. Féreca (2019). Defining blood group gene reference alleles by long-read sequencing: Proof of concept in the ACKR1 gene encoding the Duffy antigens. *Transfusion Medicine and Hemotherapy*.

Finning, K., R. Bhandari, F. Sellers, N. Revelli, M. A. Villa, E. Muñiz-Díaz, and N. Nogués (2016). Evaluation of red blood cell and platelet antigen genotyping platforms (ID CORE XT/ID HPA XT) in routine clinical practice. *Blood Transfusion 14*(2), 160–167.

Finning, K., P. Martin, J. Summers, E. Massey, G. Poole, and G. Daniels (2008). Effect of high throughput *RHD* typing of fetal DNA in maternal plasma on use of anti-RhD immunoglobulin in RhD negative pregnant women: prospective feasibility study. *The BMJ 336*(7648), 816–818.

Fisher, R. A. and R. R. Race (1946). Rh gene frequencies in Britain. *Nature 157*, 48–49.

Flegel, W. A. (2007a). Blood group genotyping in Germany. *Transfusion 47*, 47S–53S.

Flegel, W. A. (2007b). The genetic of the Rhesus blood group system. *Dtsch Arztebl 104*(10), 651–657.

Flegel, W. A., I. Von Zabern, and F. F. Wagner (2009). Six years' experience performing *RHD* genotyping to confirm D red blood cell units in Germany for preventing anti-D immunizations. *Transfusion 49*(3).

Flegel, W. A. and F. F. Wagner (2002). Molecular biology of partial D and weak D: implications for blood bank practice. *Clinical Laboratory 48*(1-2), 53–59.

Forbes, J. M., M. D. Anderson, G. F. Anderson, G. C. Bleecker, E. C. Rossi, and G. S. Moss (1991). Blood transfusion costs: a multicenter study. *Transfusion 31*(4), 318–323.

Fürst, D., C. Tsamadou, C. Neuchel, H. Schrezenmeier, J. Mytilineos, and C. Weinstock (2020). Next-generation sequencing technologies in blood group typing. *Transfusion Medicine and Hemotherapy 47*(1), 4–13.

Garalde, D. R., E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, A. Warland, M. Jordan, J. Ciccone, S. Serra, J. Keenan, S. Martin, L. McNeill, E. J. Wallace, L. Jayasinghe, C. Wright, J. Blasco, S. Young, D. Brocklebank, S. Juul, J. Clarke, A. J. Heron, and D. J. Turner (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods 15*, 201–206.

Gardner, B., D. J. Anstee, W. J. Mawby, M. J. Tanner, and A. E. von dem Borne (1991). The abundance and organization of polypeptides associated with antigens of the Rh blood group system. *Transfusion Medicine 1*(2), 77–85.

Garratty, G., W. Dzik, P. D. Issitt, D. M. Lublin, M. E. Reid, and T. Zelinski (2000). Terminology for blood group antigens and genes—historical origins and guidelines in the new millennium. *Transfusion 40*(4), 477–489.

Gassner, C., A. Schmarda, S. Kilga-Nogler, B. Jenny-Feldkircher, E. Rainer, T. H. Muller, F. F. Wagner, W. A. Flegel, and D. Schonitzer (1997). Rhd/ce typing by polymerase chain reaction using sequence-specific primers. *Transfusion 37*(10), 1020–1026.

Godard, B., J. Schmidtke, J. J. Cassiman, and S. Aymé (2003). Data storage and dna banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. a professional perspective. *European Journal of Human Genetics 11*(suppl 2), S88–122.

Goldman, M., N. Núria, and L. Castilho (2015). An overview of the progenika ID CORE XT: an automated genotyping platform based on a fluidic microarray system. *Immunohematology 31*(2), 62–68.

Goyal, P., P. V. Krasteva, N. V. Gerven, F. Gubellini, I. V. den Broeck, A. Troupiotis-Tsaïlaki, W. Jonckheere, G. Péhau-Arnaudet, J. S. Pinkner, M. R. Chapman, S. J. Hultgren, S. Howorka, R. Fronzes, and H. Remaut (2014). Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature 516*, 250–253.

Haimila, K., K. Sulin, M. Kuosmanen, I. Sareneva, A. Korhonen, S. Natunen, J. Tuimala, and S. Sainio (2017). Targeted antenatal anti-D prophylaxis program for RhD-negative pregnant women – outcome of the first two years of a national program in Finland. *Acta Obstetricia et Gynecologica Scandinavica 96*(10), 1228–1233.

Halawani, A. J. (2015). *The future of next generation Sequencing for blood group genotyping and its implications in transfusion medicine.* Ph. D. thesis, University of Plymouth.

Harmening, D. M. (2005). *Modern Blood Banking and Transfusion Practices; The Rh blood group system* (5 ed.). Davis Company.

Hashmi, G., T. Shariff, M. Seul, P. Vissavajjhala, K. Hue-Roye, D. Charles-Pierre, C. Lomas-Francis, A. Chaudhuri, and M. E. Reid (2005). A flexible array format for large-scale, rapid blood group DNA typing. *Transfusion 45*(5), 680–688.

Hashmi, G., T. Shariff, Y. Zhang, J. Cristobal, C. Chau, M. Seul, P. Vissavajjhala, C. Baldwin, K. Hue-Roye, D. Charles-Pierre, C. Lomas-Francis, and M. E. Reid (2007). Determination of 24 minor red blood cell antigens for more than 2000 blood donors by high-throughput DNA analysis. *Transfusion 47*(4), 736–747.

Hassan, M. Z., S. Iberahim, W. S. W. Abdul-Rahman, Z. Zulkafli, R. Bahar, M. Ramli, N. H. Mohd Noor, and R. Mustaffa (2019). Severe anti-D haemolytic disease of fetal and newborn in Rhesus D negative primigravida. *The Malaysian Journal of Pathology 41*(1), 55–58.

Hemker, M. B., P. C. Ligthart, L. Berger, D. J. van Rhenen, C. E. van der Schoot, and P. A. Wijk (1999). DAR, a new RhD variant involving exons 4,5, and 7, often in linkage with ceAR, a new Rhce variant frequently found in African Blacks. *Blood 94*(12), 4337–4342.

Hipsky, C. H., C. Lomas-Francis, A. Fuchisawa, and M. E. Reid (2010). RHCE*ceAR encodes a partial c (RH4) antigen. *Immunohematology 26*(2), 57–59.

Ho, A., M. Murphy, S. Wilson, S. R. Atlas, and J. S. Edwards (2011). Sequencing by ligation variation with endonuclease V digestion and deoxyinosine-containing query oligonucleotides. *BMC Genomics 12*(598), 1–8.

Huang, C.-H. (1998). The human Rh50 glycoprotein gene. structural organization and associated splicing defect resulting in Rh(null) disease. *The Journal of Biological Chemistry 273*(4), 2207–2213.

Huang, Y.-F., S.-C. Chen, Y.-S. Chiang, T.-H. Chen, and K.-P. Chiu (2012). Palindromic sequence impedes sequencing-by-ligation mechanism. *BMS System Biology 6*(S10).

Hyland, C. A., G. M. Millard, H. O'Brien, E. M. Schoeman, G. H. Lopez, E. C. McGowan, A. Tremellen, R. Puddephatt, K. Gaerty, R. L. Flower, J. A. Hyett, and G. J. Gardener (2017). Non-invasive fetal *RHD* genotyping for RhD negative women stratified into *RHD* gene deletion or variant groups: comparative accuracy using two blood collection tube types. *Pathology 49*(7), 757–764.

Hyland, C. A., E. V. Roulis, and E. M. Schoeman (2019). Developments beyond blood group serology in the genomics era. *British Journal of Haematology 184*(6), 897–911.

ISBT (1996). Third international workshop and symposium on monoclonal antibodies and related antigens. annex 3: Rh variant donations. *Transfusion clinique et biologique 3*(6), 535–536.

ISBT (2021). Red cell immunogenetics and blood group terminology [Online]. Available at www.isbtweb.org (accessed on 10.02.2021).

Iwamoto, S. (2005). Molecular aspects of Rh antigens. *Legal Medicine 7*(4), 270–273.

Jungbauer, C. (2011a). Blood group molecular genotyping. *ISBT Science Series 6*(2), 399–403.

Jungbauer, C. (2011b). Routine use of DNA testing for red cell antigens in blood centres. *Transfusion Apheresis Science 45*(1), 61–68.

Kappler-Gratias, S., T. Peyrard, M. Beolet, D. Amiranoff, C. Menanteau, I. Dubeaux, P. Rouger, J.-P. Cartron, P.-Y. L. Pennec, and B.-N. Pham (2011). Blood group genotyping by high-throughput DNA analysis applied to 356 reagent red blood cell samples. *Transfusion 51*(1), 36–42.

Karamitros, T. and G. Magiorkinis (2015, 08). A novel method for the multiplexed target enrichment of MinION next generation sequencing libraries using PCR-generated baits. *Nucleic Acids Research 43*(22), e152–e152.

Karnot, A., S. J. Woestmann, and B. K. Flesch (2016). Weak RhD phenotype caused by compound heterozygosity for DAU-2 and a new RHD*c393-394 dupGG-mutation. *Transfusion Medicine Hemotherapy 43*(Suppl 1), 42–43.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols 10*, 845–858.

Kemp, T. J., M. Poulter, and B. Carritt (1999). Microsatellite variation within the human *RHCE* gene. *Vox Sanguinis 77*(3), 159–163.

Kim, D. S., S. H. Lee, C. H. Ahn, J. Y. Lee, and T. H. Kwon (2006). Disposable integrated microfluidic biochip for blood typing by plastic microinjection moulding. *Lab Chip 6*(6), 794–802.

Klein, H. and D. Anstee (2005). *Blood transfusion in clinical medicine* (11 ed.). Oxford (UK), Blackwell Publishing.

Körmöczi, G. F., E. Förstemann, C. Gabriel, W. R. Mayr, D. Schönitzer, and C. Gassner (2005). Novel weak D types 31 and 32: adsorption-elution-supported D antigen analysis and comparison to prevalent weak D types. *Transfusion 45*(10), 1574–1580.

Kumawat, V., A. Jain, R. R. Sharma, and N. Marwaha (2012). Hemolytic disease of fetus and newborn due to anti-e alloantibody in a newborn of Rh (D)-positive mother. *Asian Journal of Transfusion Science 6*(2), 187.

Kumpel, B. M. (2008). Lessons learnt from many years of experience using anti-D in humans for prevention of RhD immunization and haemolytic disease of the fetus and newborn. *Clinical and Expermintal Immunology 154*(1), 1–5.

Kumpel, B. M. and C. J. Elson (2001). Mechanism of anti-D-mediated immune suppression – a paradox awaiting resolution? *Trends in Immunology 22*(1), 26–31.

Kumpel, B. M., M. J. Goodrick, D. H. Pamphilon, I. D. Fraser, G. D. Poole, C. Morse, G. R. Standen, G. E. Chapman, D. P. Thomas, and D. J. Anstee (1995). Human Rh D monoclonal antibodies (BRAD-3 and BRAD-5) cause accelerated clearance of Rh D+ red blood cells and suppression of Rh D immunization in Rh D- volunteers. *Blood 86*(5), 1701–1709.

Kwon, D. H., S. G. Sandler, and W. A. Flegel (2017). DEL phenotype. *Immunohematology 33*(3), 125–132.

Lander, E., L. Linton, B. Birren, C. Nusbaum, M. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, C. Ponting, and International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature 409*(6822), 860–921.

Landsteiner, K. (1961). On agglutination of normal human blood. *Transfusion 1*(1), 5–8.

Lane, W. J., S. Vege, H. H. Mah, C. Lomas-Francis, M. Aguad, R. Smeland-Wagman, C. Koch, J. M. Killian, C. L. Gardner, M. De Castro, M. S. Lebo, R. M. Kaufman, R. C. Green, C. M. Westhoff, and for the MilSeq Project (2019). Automated typing of red blood cell and platelet antigens from whole exome sequences. *Transfusion 59*(10), 3253–3263.

Lane, W. J., C. M. Westhoff, J. M. Uy, M. Aguad, R. Smeland-Wagman, R. M. Kaufman, H. L. Rehm, R. C. Green, and L. E. M. P. Silberstein (2016). Comprehensive red blood cell and platelet antigen prediction from whole genome sequencing: proof of principle. *Transfusion 56*(3), 743–754.

Lang, K., V. Surendranath, P. Quenzel, G. Schöfl, A. H. Schmidt, and V. Lange (2018). Full-length HLA class I genotyping with the MinION nanopore sequencer. *Methods in Molecular Biology 1802*(155-162).

Laver, T., J. Harrison, P. A. O'Neill, K. Moore, A. Farbos, K. Paszkiewicz, and D. J. Studholme (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification 3*, 1–8.

Le van Kim, C., I. Mouro, B. Chérif-Zahar, V. Raynal, C. Cherrier, J. Cartron, and Y. Colin (1992). Molecular cloning and primary structure of the human blood group RhD polypeptide. *Proceedings of the National Academy of Sciences of the United States of America 89*(22), 10925–10929.

Leahy, M. F. and S. A. Mukhtar (2012). From blood transfusion to patient blood management: a new paradigm for patient care and cost assessment of blood transfusion practice. *Internal Medicine Journal 42*(3), 332–338.

Levine, P., L. Burnham, E. M. Katzin, and P. Vogel (1941). The role of iso-immunization in the pathogenesis of erythroblastosis fetalis. *American Journal of Obstetrics and Gynecology 42*(6), 925–937.

Levine, P. and R. E. Stetson (1939). An unusual case of intra-group agglutination. *Journal of the American Medical Association 113*(2), 126–127.

Liao, P., G. A. Satten, and Y. J. Hu (2017). PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genetic Epidemiology 41*(5), 375–387.

Liu, Z., Y. Chen, R. Mo, C. Hui, J. F. Cheng, N. Mohandas, and C. H. Huang (2000). Characterization of human RhCG and mouse Rhcg as novel nonerythroid Rh glycoprotein homologues predominantly expressed in kidney and testis. *Journal of Biological Chemistry 275*(33), 25641–25651.

Liu, Z., M. Liu, T. Mercado, O. Illoh, and D. R (2014). Extended blood group molecular typing and next-generation sequencing. *Transfusion Medicine Reviews 28*(4), 177–186.

Liu, Z., J. Peng, M. R, H. C, and C. H. Huang (2001). Rh type B glycoprotein is a new member of the Rh superfamily and a putative ammonia transporter in mammals. *Journal of Biological Chemistry 276*(2), 1424.

Loman, N. J., J. Quick, and J. T. Simpson (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods 12*(8), 733–735.

Lomas, C., J. Storry, P. Spruell, and M. Moulds (1994). An apparent alloanti-C produced by r^s people: five examples [abstract]. *Transfusion 34*(10), S25–S25.

Lopez, G. H., R. M. Turner, E. C. McGowan, E. M. Schoeman, S. A. Scott, H. O'Brien, G. M. Millard, E. V. Roulis, A. J. Allen, Y.-W. Liew, R. L. Flower, and C. A. Hyland (2018). A DEL phenotype attributed to *RHD* exon 9 sequence deletion: slipped-strand mispairing and blood group polymorphisms. *Transfusion 58*(3), 685–691.

Lopez-Alvarez, D., B. Parra, and W. J. Cuellar (2020). Genome sequence of SARS-CoV-2

isolate Cali-01, from Colombia, obtained using Oxford Nanopore MinION sequencing. *Microbiology Resource Announcements 9*(26), e00573–20.

López, M., I. Apraiz, M. Rubia, M. Piedrabuena, M. Azkarate, B. Veldhuisen, M. . Vesga, E. van Der Schoot, F. Puente, and D. Tejedor (2018). Performance evaluation study of ID CORE XT, a high throughput blood group genotyping platform. *Blood Transfusion 16*(2), 193–199.

Lu, H., F. Giordano, and Z. Ning (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics 14*(5), 265–279.

Lubart, E., R. Segal, T. N, S. E, and A. Leibovitz (2014). Blood transfusion reactions in elderly patients hospitalized in a multilevel geriatric hospital. *Journal of Aging Research 2014*, 1–3.

Löw, B. and L. Messeter (1974). Antiglobulin test in low-ionic strength salt solution for rapid antibody screening and cross-matching. *Vox Sanguinis 26*, 53–61.

Magi, A., B. Giusti, and L. Tattini (2016). Characterization of MinION nanopore data forresequencing analyses. *Briefings in Bioinformatics 18*(6), bbw077.

Malomgré, W. and B. Neumeister (2009). Recent and future trends in blood group typing. *Anal Bioanal Chem 393*(5), 1443–1451.

Maréchal, C. L., C. Guerry, C. Benech, L. Burlot, B. Cavelier, V. Porra, M. Delamaire, C. Férec, and J.-M. Chen (2007). Identification of 12 novel *RHD* alleles in western France by denaturing high-performance liquid chromatography analysis. *Transfusion 47*(5), 858–863.

Margulies, M., M. Egholm, W. E. Altman, J. Said Attiya, S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, Xavier, V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I.

Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, K. L. L. Jing Li, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, , and J. M. Rothberg (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature 437*(7057), 376–380.

Marquez, S., B. Prado-Vivar, J. J. Guadalupe, B. G. Granja, M. Jibaja, M. Tobar, F. Mora, J. Gaviria, M. Garcia, E. Ligna, F. Espinosa, J. Reyes, V. Barragan, P. Rojas-Silva, G. Trueba, M. Grunauer, and P. Cardenas (2020). Genome sequencing of the first SARS-CoV-2 reported from patients with COVID-19 in Ecuador. preprint. *medRxiv : the preprint server for health sciences 2020.06.11.20128330*.

Matern, B., M. Groeneweg, T. Slangen, M. G. Tilanus, and C. Voorter (2017). ABO blood group typing with the Oxford nanopore MinION. *Human Immunology 78*, 122.

Matern, B. M., T. I. Olieslagers, M. Groeneweg, B. Duygu, L. Wieten, M. G. Tilanus, and C. E. Voorter (2020). Long-read nanopore sequencing validated for human leukocyte antigen class I typing in routine diagnostics. *The Journal of Molecular Diagnostics 22*(7), 912 – 919.

Maxam, A. M. and W. Gilbert (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America 74*(2), 560–564.

Mazonson, P., M. Efrusy, C. Santas, A. Ziman, J. Burner, S. Roseff, A. Vijayaraghavan, and R. Kaufman (2014). The HI-STAR study: resource utilization and costs associated with serologic testing for antibody-positive patients at four United States medical centers. *Transfusion 54*(2), 271–277.

McBean, R. S., C. A. Hyland, and R. L. Flower (2014). Approaches to determination of a full profile of blood group genotypes: single nucleotide variant mapping and massively parallel sequencing. *Computational and Structural Biotechnology Journal 11*(19), 147–151.

McGann, P. T., J. M. Despotovic, T. A. Howard, and R. E. Ware (2012). A novel laboratory technique demonstrating the influences of *rhd* zygosity and the RhCcEe phenotype on erythrocyte D antigen expression. *American Journal of Hematology 87*(3), 266–271.

Meller, A., L. Nivon, E. Brandin, J. Golovchenko, and D. Branton (2000). Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences 97*(3), 1079–1084.

Meller, A., L. Nivon, and D. Branton (2001). Voltage-driven DNA translocations through a nanopore. *Physical Review Letters 86*(15), 3435–3438.

Minon, J. M., C. H. Gerard, F. Chantraine, and M. Nisolle (2015). Anti-D prophylaxis reviewed in the area of foetal *RHD* genotyping. *Journal of Blood Disorders & Transfusion 6*(5), 1–6.

Möller, M., M. Joud, J. R. Storry, and M. L. Olsson (2016). Erythrogene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 genomes project. *Blood Advances 1*(3), 240–249.

Montemayor-Garcia, C. and C. M. Westhoff (2018). The "next generation" reference laboratory? *Transfusion 58*(2), 277–279.

Moulds, J. M., G. T. Noumsi, and K. L. Billingsley (2014). A comparison of methods for the detection of the r^s haplotype. *Transfusion 55*(6pt2), 1418–1422.

Mouro, I., Y. Colin, B. Chérif-Zahar, J. P. Cartron, and C. Le Van Kim (1993). Molecular genetic basis of the human Rhesus blood group system. *Nature Genetics 5*(1), 62–65.

Mouro, I., Y. Colin, P. Sistonen, P. Le Pennec, J. Cartron, and C. Le Van Kim (1995). Molecular basis of the RhCW (Rh8) and RhCX (Rh9) blood group specificities. *Blood 86*(3), 1196.

Mu, S., Y. Cui, W. Wang, L. Wang, H. Xu, O. Zhu, and D. Zhu (2019). A *rhag* point mutation selectively disrupts Rh antigen expression. *Transfusion Medicine 29*(2), 121–127.

Mujahid, A. and F. L. Dickert (2015). Blood group typing: from classical strategies to the application of synthetic antibodies generated by molecular imprinting. *Sensors (Basel) 16*(1), E51.

Nagarajan, N. and M. Pop (2013). Sequence assembly demystified. *Nature Reviews Genetics 14*(3), 157–167.

Naveed, M., E. Ayday, E. W. Clayton, J. Fellay, C. A. Gunter, J.-P. Hubaux, B. A. Malin, and X. Wang (2015). Privacy in the genomic era. *ACM Computing Surveys 48*(1), 6.

NCBI (2009). Genome reference consortium human build 37 (grch37) [internet]. bethesda (md): National library of medicine (us), national center for biotechnology information. *Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF*$_0$00001405.13/ Accessed2016.

NICE (2007). Routine antenatal anti-D prophylaxis for women who are rhesus D negative [Online]. Available at www.nice.org.uk (accessed on 30.06.2020).

Nicolas, V., I. Mouro-Chanteloup, C. Lopez, P. Gane, A. Gimm, N. Mohandas, J. Cartron, C. Le Van Kim, and Y. Colin (2006). Functional interaction between Rh proteins and the spectrin-based skeleton in erythroid and epithelial cells. *Transfusion Clinique et Biologique 13*(1-2), 23–28.

Nielsen, R., J. S. Paul, A. Albrechtsen, and Y. S. Song (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics. 12*(6), 443–451.

Noizat-Pirenne, F., K. Lee, P.-Y. L. Pennec, P. Simon, P. Kazup, D. Bachir, A.-M. Rouzaud, M. Roussel, G. Juszczak, C. Ménanteau, P. Rouger, R. Kotb, J.-P. Cartron, and H. Ansart-Pirenne (2002). Rare *RHCE* phenotypes in black individuals of Afro-Caribbean origin: identification and transfusion safety. *Blood 100*(12), 4223–4231.

Noizat-Pirenne, F., I. Mouro, P. Gane, Y. Okubo, Y. Hori, P. Rouger, P. Y. L. Pennec, and J.-P. Cartron (1998). Heterogeneity of blood group RhE variants revealed by serological analysis and molecular alteration of the *RHCE* gene and transcript. *British Journal of Haematology 103*(2), 429–436.

Oepkes, D., G. C. L. Page-Christiaens, C. J. Bax, M. N. Bekker, C. M. Bilardo, E. M. J. Boon, G. H. Schuring-Blom, A. B. C. Coumans, B. H. Faas, R.-J. H. Galjaard, A. T. Go, L. Henneman, M. V. E. Macville, E. Pajkrt, R. F. Suijkerbuijk, K. H. van Amsterdam, D. V. Opstal, E. J. J. Verweij, M. M. Weiss, E. A. Sistermans, and the Dutch NIPT Consortium. (2006). Trial by Dutch laboratories for evaluation of non-invasive prenatal testing. part I—clinical impact. *Prenatal Diagnosis 36*(12), 1083–1090.

Okuda, H. and E. Kajii (2002). The evolution and formation of RH genes. *Legal Medicine 4*(3), 139–155.

Okuda, H., H. Suganuma, T. Kamesaki, M. Kumada, N. Tsudo, T. Omi, S. Iwamoto, and E. Kajii (2000). The analysis of nucleotide substitutions, gaps, and recombination events between *RHD* and *RHCE* genes through complete sequencing. *Biochemical and Biophysical Research Communications 274*(3), 670–683.

Okuda, H., H. Suganuma, N. Tsudo, T. Omi, S. Iwamoto, and E. Kajii (1999). Sequence analysis of the spacer region between the *RHD* and *RHCE* genes. *Biochemical and Biophysical Research Communications 263*(2), 378–383.

Orzińska, A., K. Guz, B. Michalewska, and E. Brojer (2016). Molecular screening of the

Cw antigen for typing donors compatible with patients with anti-MAR-like antibodies. *Blood Transfusion 14*(6), 573–576.

Orzińska, A., K. Guz, M. Mikula, M. Kulecka, A. Kluska, A. Balabas, M. Pelc-Kłopotowska, J. Ostrowski, and E. Brojer (2018). A preliminary evaluation of next-generation sequencing as a screening tool for targeted genotyping of erythrocyte and platelet antigens in blood donors. *Blood Transfussion 16*(1), 285–292.

O'Shea, K. P., R. Oyen, L. Sausais, V. A. Stephens, G. F. Stillwell, L. A. Bisgard, J. J Martin, and M. E. Reid (2001). A mar-like antibody in a DCWe/DCWe person. *Transfusion 41*(1), 53–55.

Paccapelo, C., F. Truglio, M. A. Villa, N. Revelli, and M. Marconi (2015). HEA BeadChip™ technology in immunohematology. *Immunohematology 31*(2), 81–90.

Pamphilon, D. H. and M. L. Scott (2007). Robin coombs: his life and contribution to haematology and transfusion medicine. *British Journal of Haematology 137*(5), 401–408.

Pareek, C. S., R. Smoczynski, and A. Tretyn (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics 52*(4), 413–435.

Patterson, J., T. Wongsurawat, and A. Rodriquez (2020, 02). A glioblastoma genomics primer for clinicians. *8*, 1–17.

Perreault, J., J. Lavoie, P. Painchaud, M. Côté, J. Constanzo-Yanez, R. Côté, G. Delage, F. Gendron, S. Dubuc, B. Caron, R. Lemieux, and M. St-Louis (2009). Set-up and routine use of a database of 10,555 genotyped blood donors to facilitate the screening of compatible blood components for alloimmunized patients. *Vox Sanguinis 97*(1), 61–68.

Peyrard, T., B. Pham, S. Poupel, S. Martin-Blanc, C. Auxerre, S. Kappler-Gratias, P. Bonin, P. Rouger, and P. L. Pennec (2009). Alloanti-c/ce in a c+cear/ce patient suggests that the rare rhce cear allele (cear) encodes a partial c antigen. *Transfusion 49*(11), 2406–2411.

Pham, B.-N., T. Peyrard, G. Juszczak, C. Auxerre, S. Godin, P. Bonin, P. Rouger, and P.-Y. L. Pennec (2009). Alloanti-c (RH4) revealing that the (C)ce^{s} haplotype encodes a partial c antigen. *Transfusion 49*(7), 1329–1334.

Planelles, G. (2007). Ammonium homeostasis and human Rhesus glycoproteins. *Nephorn Physiology 105*, 11–17.

Plesivkova, D., R. Richards, and S. Harbison (2019). A review of the potential of the minion™ single-molecule sequencing system for forensic applications. *WIREs Forensic Science 1*(1), e1323.

Polin, H., M. Pelc-Klopotowska, M. Danzer, S. Suessner, C. Gabriel, J. Wilflingseder, A. Żmudzin, A. Orzińska, K. Guz, B. Michalewska, and E. Brojer (2016). Compound heterozygosity of two novel *(RHAG)* alleles leads to a considerable disruption of the Rh complex. *Transfusion 56*(4), 950–955.

Poole, J. and G. Daniels (2007). Blood group antibodies and their significance in transfusion medicine. *Transfusion Medicine Reviews 21*(1), 58–71.

Pourmand, N., M. Karhanek, H. H. J. Persson, C. D. Webb, T. H. Lee, A. Zahradníková, and R. W. Davis (2006). Direct electrical detection of DNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America 103*(17), 6466–6470.

Profaizer, T., E. Lázár-Molnár, D. W. Close, J. C. Delgado, and A. Kumánovics (2016). HLA genotyping in the clinical laboratory: comparison of next-generation sequencing methods. *HLA 88*(1-2), 14–24.

Quick, J., N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouédraogo, B. Afrough, A. Bah, J. H. J. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, L. L. Camino-Sánchez, Álvarov Carter, J. Doerrbecker, T. Enkirch, I. G. Dorival, N. Hetzelt, J. Hinzmann,

T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallasch, L. V. Patrono, J. G. Portmann, Jasminev Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, T. Racine, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. J. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O. Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Wölfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keïta, A. Rambaut, P. Formenty, S. Günther, and M. W. Carroll (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature 530*(7589), 228–232.

Quirino, M. G., C. M. Colli, L. C. Macedo, A. M. Sell, and J. E. L. Visentainer (2019). Methods for blood group antigens detection: cost-effectiveness analysis of phenotyping and genotyping. *Hematology, Transfusion and Cell Therapy 41*(1), 44–49.

Ranasinghe, E., E. Goodyear, and G. Burgess (2003). Anti-Ce complicating two consecutive pregnancies with increasing severity of haemolytic disease of the newborn. *Transfusion Medicine 13*(1), 53–56.

Rand, A. C., M. Jain, J. M. Eizenga, A. Musselman-Brown, H. E. Olsen, M. Akeson, and B. Paten (2017). Mapping DNA methylation with high throughput nanopore sequencing. *Nature Methods 14*(4), 411–413.

Rang, F., W. Kloosterman, and J. de Ridder (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology 19*(90).

Reid, M. E., C. Lomas-Francis, and M. L. Olsson. (2012). *The blood group antigen FactsBook* (3 ed.). Boston: Academic Press.

Reid, M. E. and N. Mohandas (2004). Red blood cell blood group antigens: structure and function. *Seminars in Hematology 41*(2), 93–117.

Ribeiro, K. R., M. H. Guarnieri, D. C. da Costa, F. F. Costa, J. J. Pellegrino, and L. Castilho (2009). DNA array analysis for red blood cell antigens facilitates the transfusion support with antigen-matched blood in patients with sickle cell disease. *Vox Sanguinis 97*(2), 147–152.

Ridgwell, K., N. K. Spurr, B. Laguda, C. MacGeoch, N. D. Avent, and M. J. Tanner (1992). Isolation of cDNA clones for a 50 kDa glycoprotein of the human erythrocyte membrane associated with Rh (rhesus) blood-group antigen expression. *The Biochemical journal 287*(1), 223–228.

Rizzo, C., L. Castiglia, E. Arena, S. Gangi, G. Mazzola, C. Caruso, and S. Vasto (2012). Weak D and partial D: our experience in daily activity. *Blood Transfusion 10*(2), 235–236.

Rodrigues, A., M. Rios, F. F. Costa, S. T. O. Saad, J. Pellegrino Jr, and L. Castilho (2004). Weakened expression of 'e' owing to concomitant occurrence of Cys16 and Val245 (VS antigen). *Vox Sanguinis 86*(2), 136–140.

Rosenfield, R. E., F. H. Allen, S. N. Swisher, and S. Kochwa (1962). A review of Rh serology and presentation of a new terminology. *Transfusion 2*(5), 287–311.

Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman,

R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, and J. Bustillo (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature 475*(7356), 348–352.

Rouillac, C., C. Le Van Kim, M. Beolet, J. P. Cartron, and Y. Colin (1995). Leu110Pro substitution in the RhD polypeptide is responsible for the DVII category blood group phenotype. *American Journal of Hematology 49*(1), 87–88.

Sandler, S. G., L. Chen, and W. A. Flegel (2017). Serological weak D phenotypes: A review and guidance for interpreting the RhD blood type using the *RHD* genotype. *British Journal of Haematology 179*(1), 10–19.

Sandler, S. G., W. A. Flegel, C. M. Westhoff, G. A. Denomme, M. Delaney, M. A. Keller, S. T. Johnson, L. Katz, J. T. Queenan, R. R. Vassallo, C. D. Simon, and C. of American Pathologists Transfusion Medicine Resource Committee Work Group. (2015). It's time to phase in *RHD* genotyping for patients with a serologic weak D phenotype. *Transfusion 55*(3), 680–689.

Sanger, F., S. Nicklen, and A. R. Coulson (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America 74*(12), 5463–5467.

Schadt, E. E., S. Turner, and A. Kasarskis (2010). A window into third-generation sequencing. *Human Molecular Genetics 19*(R2), R227–40.

Schatz, M. C., A. L. Delcher, and S. L. Salzberg (2010). Assembly of large genomes using second-generation sequencing. *Genome Research 20*(9), 1165–1173.

Schoeman, E. M., E. V. Roulis, Y.-W. Liew, J. R. Martin, T. Powley, B. Wilson, G. M. Millard, E. C. McGowan, G. H. Lopez, H. O'Brien, J. A. Condon, R. L. Flower, and C. A. Hyland (2018). Targeted exome sequencing defines novel and rare variants in complex blood

group serology cases for a red blood cell reference laboratory setting. *Transfusion 58*(2), 284–293.

Shander, A., A. Hofmann, H. Gombotz, O. M.Theusinger, and D. R. Spahn (2007). Estimating the cost of blood: past, present, and future directions. *Best Practice & Research: Clinical Anaesthesiology 21*(2), 271–289.

Shao, C.-P., J.-H. Maas, Y.-Q. Su, M. Köhler, and T. J. Legler (2002). Molecular background of Rh D-positive, D-negative, D(el) and weak D phenotypes in Chinese. *Vox Sanguinis 83*(2), 156–161.

Shao, C. P., B. Y. Wang, S. H. Ye, W. L. Zhang, H. Xu, N. B. Zhuang, X. Y. Wu, and H. G. Xu (2012). DEL RBC transfusion should be avoided in particular blood recipient in East Asia due to allosensitization and ineffectiveness. *Journal of Zhejiang University SCIENCE B 13*(11), 913–918.

Sheeladevi, C. S., S. Suchitha, G. V. Manjunath, and S. Murthy (2013). Hemolytic disease of the newborn due to anti-c isoimmunization: A case report. *Indian Journal of Hematology and Blood Transfusion 29*(3), 155–157.

Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research 29*(1), 308–311.

Shirey, R., J. Boyd, and P. Ness (1994). Polyethylene glycol versus low-ionic-strength solution in pretransfusion testing: a blinded comparison study. *Transfusion 34*(5), 368–370.

Sillence, K. A., A. J. Halawani, W. A. Tounsi, K. A. Clarke, M. Kiernan, T. E. Madgett, and N. D. Avent (2017). Rapid *RHD* zygosity determination using digital PCR. *Clinical Chemistry 63*(7), 1388–1397.

Sillence, K. A., L. A. Roberts, H. J. Hollands, H. P. Thompson, M. Kiernan, T. E. Madgett, C. R. Welch, and N. D. Avent (2015). Fetal sex and *RHD* genotyping with digital PCR demonstrates greater sensitivity than real-time PCR. *Clinical Chemistry 61*(11), 1399–1407.

Simister, N. E. (2003). Placental transport of immunoglobulin G. *Vaccine 21*(24), 3365–3369.

Simpson, J. T., R. E. Workman, P. C. Zuzarte, M. David, L. J. Dursi, and W. Timp (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods 14*, 407–410.

Sims, D., I. Sudbery, N. E. Ilott, A. Heger, and C. P. Ponting (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics 15*(2), 121–132.

Simsek, S., C. A. de Jong, H. T. Cuijpers, P. M. Bleeker, T. M. Westers, M. A. Overbeeke, R. Goldschmeding, C. E. van der Schoot, and A. E. von dem Borne (1994). Sequence analysis of cDNA derived from reticulocyte mrnas coding for Rh polypeptides and demonstration of E/e and C/c polymorphisms. *Vox Sanguinis 67*(2), 203–209.

Singleton, B. K., C. A. Green, N. D. Avent, P. G. Martin, E. Smart, A. Daka, E. G. Narter-Olaga, L. M. Hawthorne, and G. Daniels (2000). 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 95*(1), 12–18.

Srivastava, K., D. A. Stiles, F. F. Wagner, and W. A. Flegel (2018). Two large deletions extending beyond either end of the *RHD* gene and their red cell phenotypes. *Journal of human genetics 63*(1), 27–35.

St-Louis, M., J. Perreault, J. Lavoie, J. Émond, J. St-Laurent, A. Long, and M. Richard

(2010). Genotyping of 21,000 blood donors in Quebec and *RHD* analysis. *Transfusion Clinique et Biologique 17*(4), 242–248.

Stabentheiner, S., M. Danzer, N. Niklas, S. Atzmüller, J. Pröll, C. Hackl, H. Polin, K. Hofer, and C. Gabriel (2011). Overcoming methodical limits of standard *RHD* genotyping by next-generation sequencing. *Vox Sanguinis 100*(4), 381.

Storry, J. R., M. Jöud, and M. L. Olsson (2019). Automatic for the people: a rapidly evolving movement for the future of genotyping. *Transfusion 59*(11), 3545–3547.

Strobel, E., F. Noizat-pirenne, S. Hofmann, J. P. Cartron, and M. Bauer (2004). The molecular basis of the Rhesus antigen E_w. *Transfusion 44*(3), 407–409.

Stroup, M. and M. MacIlroy (1965). Evaluation of the albumin antiglobulin technic in antibody detection. *Transfusion 5*(2), 184–191.

Stubbs, J. R. (2014). Wrapping out arms around the cost of transfusion. *Transfusion 54*(2), 259–262.

Szymanski, I. O. (1987). Quantitative studies on the Rh-antigen D: Effect of the C gene. *Transfusion 27*(3), 257–261.

Tammi, S. (2019). Next-generation sequencing of maternal *RHD* variants. Master's thesis, University of Helsinki.

Tammi, S. M., W. A. Tounsi, S. Sainio, M. Kiernan, N. D. Avent, T. E. Madgett, and K. Haimila (2020). Next-generation sequencing of 35 *RHD* variants in 16,253 serologically D negative pregnant women in Finnish population. accepted. *Blood Advances 4*(20), 4994–5001.

Tatusova, T., M. DiCuccio, A. Badretdin, V. Chetvernin, E. Nawrocki, L. Zaslavsky, A. Lomsadze, K. Pruitt, M. Borodovsky, and J. Ostell (2016). Ncbi prokaryotic genome annotation pipeline. *Nucleic Acids Research 44*(4), 6614–6624.

Tian, L., N. Song, Z. Q. Yao, M. Huang, and L. Hou (2011). A family study of the Chinese $\mathrm{rh}_{\mathrm{null}}$ individual of the regulator type: a novel single mis-sense mutation identified in *RHAG* gene. *Transfusion 51*(12), 2686–2689.

Tilley, L. and S. Grimsley (2014). Is next generation sequencing the future of blood group testing? *Transfusion and Apheresis Science 50*(2), 183–188.

Ton, K. N. T., S. L. Cree, S. J. Gronert-Sum, T. R. Merriman, L. K. Stamp, and M. A. Kennedy (2018). Multiplexed nanopore sequencing of HLA-B locus in Māori and Pacific Island samples. *Frontiers in Genetics 9*, 152.

Tounsi, W. A., T. E. Madgett, and N. D. Avent (2018). Complete *RHD* next generation sequencing: establishment of reference *RHD* alleles. *Blood Advances 2*(20), 2713–2723.

Tyler, A. D., L. Mataseje, C. J. Urfano, L. Schmidt, K. S. Antonation, M. R. Mulvey, and C. R. Corbett (2018). Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Scientific Reports 8*(10931).

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. Faircloth, M. Remm, and S. Rozen (2012). Primer3 new capabilities and interfaces. *Nucleic Acids Research 40*(15), e115.

Urbaniak, S. and M. Greiss (2000). RhD haemolytic disease of the fetus and the newborn. *Blood Reviews 14*(1), 44–61.

Veldhuisen, B., C. E. Van Der Schoot, and M. de Haas (2009). Blood group genotyping: from patient to high-throughput donor screening. *Vox Sanguinis 97*(3), 198–206.

Venter, J., M. Adams, E. Myers, P. Li, R. Mural, G. Sutton, H. Smith, M. Yandell, C. Evans, R. Holt, J. Gocayne, P. Amanatides, R. Ballew, D. Huson, J. Wortman, Q. Zhang, C. Kodira, X. Zheng, L. Chen, M. Skupski, G. Subramanian, P. Thomas, J. Zhang, G. Gabor Miklos, C. Nelson, S. Broder, A. Clark, J. Nadeau, V. McKusick, N. Zinder, A. Levine, R. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher,

I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. Heiman, M. Higgins, R. Ji, Z. Ke, K. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. Merkulov, N. Milshina, H. Moore, A. Naik, V. Narayan, B. Neelam, D. Nusskern, D. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. Abril, R. Guigó, M. Campbell, K. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan,

J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu (2001). The sequence of the human genome. *Science 291*(5507), 1304–1351.

Visser, G. H. A., G. C. D. Renzo, S. L. Spitalnik, F. C. S. Motherhood, and N. Health (2019). The continuing burden of Rh disease 50 years after the introduction of anti-Rh(D) immunoglobin prophylaxis: call to action. *American Journal of Obstetrics and Gynecology 221*(3), 227.el–227.e4.

von Zabern, I. and W. A. Flegel (2007). IVS5-38del4 deletion in the *RHD* gene does not cause a DEL phenotype: relevance for *RHD* alleles including DFR-3. *Transfusion 47*(8), 1552–1555.

von Zabern, I., F. F. Wagner, J. M. Moulds, J. J. Moulds, and W. A. Flegel (2013). D category IV: a group of clinically relevant and phylogenetically diverse partial D. *Transfusion 53*(Suppl 2), 2960–2973.

Wagner, F. F., N. I. Eicher, J. R. Jørgensen, C. B. Lonicer, and W. A. Flegel (2002). DNB: a partial D with anti-D frequent in Central Europe. *Blood 100*(6), 2253–2256.

Wagner, F. F. and W. A. Flegel (2000). *RHD* gene deletion occurred in the Rhesus box. *Blood 95*(12), 3662–3668.

Wagner, F. F. and W. A. Flegel (2014). The Rhesus site. *Transfusion Medicine and Hemotherapy 41*(5), 357–363.

Wagner, F. F., A. Frohmajer, and W. A. Flegel (2001). *RHD* positive haplotypes in D negative Europeans. *BMC Genetics 2*(10).

Wagner, F. F., A. Frohmajer, B. Ladewig, N. I. Eicher, C. B. Lonicer, T. H. Müller, M. H. Siegel, and W. A. Flegel (2000). Weak D alleles express distinct phenotypes. *Blood 95*(8), 2699–708.

Wagner, F. F., C. Gassner, T. H. Müller, D. Schönitzer, F. Schunter, and W. A. Flegel (1999). Molecular basis of weak D phenotypes. *Blood 93*(1), 385–393.

Wagner, T., G. F. Körmöczi, C. Buchta, M. Vadon, G. Lanzer, W. R. Mayr, and T. J. Legler (2005). Anti-D immunization by DEL red blood cells. *Transfusion 45*(4), 520–526.

Walker, L. (2018). Low-ionic-strength saline solution-antiglobulin test (LISS-AGT). *Immunohematology 34*(2), 57–60.

Watkins, W. M., P. Greenwell, A. D. Yates, and P. H.Johnson (1988). Regulation of expression of carbohydrate blood group antigens. *Biochimie 70*(11), 1597–1611.

Wen, J., O. J. Verhagen, S. Jia, Q. Liang, Z. Wang, L. Wei, H. Luo, G. Luo, G. Vidarsson, E. van den Akker, Y. Ji, and C. E. van der Schoot (2019). A variant RhAG protein encoded by the *RHAG*572A* allele causes serological weak D expression while maintaining normal RhCE phenotypes. *Tranfusion 59*(1), 405–411.

Westhoff, C. M. (2007). The structure and function of the Rh antigen complex. *Semin Hematol. 44*(1), 42–50.

Westhoff, C. M., L. E. Silberstein, D. E. Wylie, M. Skavdahl, and M. E. Reid (2001). 16Cys encoded by the RHce gene is associated with altered expression of the e antigen and is frequent in the R0r haplotype. *British Journal of Haematology 113*(3), 666–671.

Wiener, A. S. (1969). Karl landsteiner, M.D. History of Rh-Hr blood group system. *New York State Journal of Medicine 69*(2), 2915–2935.

Wikipedia (2015). Coombs test [Online]. Available at www.wikipedia.org (accessed on 17.07.2020).

Wilcox, C. R., B. Holder, and C. E. Jones (2017). Factors affecting the FcRn-mediated transplacental transfer of antibodies and implications for vaccination in pregnancy. *Frontiers in Immunology 8*, 1294.

Wu, P. C., S.-C. Pai, and P.-L. Chen (2018). Blood group genotyping goes next generation: featuring ABO, RH and MNS. *ISBT Science Series 13*(3), 290–297.

Xu, L. and M. Seki (2020). Recent advances in the detection of base modifications using the Nanopore sequencer. *Journal of Human Genetics 65*, 25–33.

Yamamoto, F. (2000). Molecular genetics of ABO. *Vox Sanguinis 78*(Suppl 2), 91–103.

Yang, H. S., M. Y. Lee, T. S. Park, S. Y. Cho, H. J. Lee, G. Lim, D. D. Lee, S. H. Oh, D. Cho, and K. U. Park (2015). Primary anti-D alloimmunization induced by "Asian type" *RHD* (c.1227G>A) DEL red cell transfusion. *Annals of Laboratory Medicine 35*(5), 554–556.

Yasuda, H., H. Ohto, S. Sakuma, and Y. Ishikawa (2005). Secondary anti-D immunization by Del red blood cells. *Transfusion 45*(10), 1581–1584.

Ye, S., D. Wu, M. Wang, X. Wu, H. Xu, H. Xu, and C. Shao (2014). A comprehensive investigation of *RHD* polymorphisms in the Chinese Han population in Xi'an. *Blood Transfusion 12*(3), 396–404.

Zhang, J., R. Chiodini, A. Badr, and G. Zhang (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics 38*(3), 95–109.

Zhou, H. Y., Y. Z. Zhang, Q. B. Meng, X. H. Bai, C. R. Wang, Q. Cao, and J. C. Lan (2008). Comparison between genotyping and serological phenotyping in RhCE blood group. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi 25*(1), 66–69.

Appendices

Training and Awards

- Attended the introduction to Nanopore sequencing and data analysis training workshop, Oxford Nanopore Technologies, Oxford, UK, 29-30 November 2018

- Attended the GRADSchool: residential development program for doctoral researchers. Beacons, Wales. 4^{th}– 7^{th} July 2017.

- Achieved the status of Associate Fellow of the Higher Education Academy, 2017.

- Awarded the Professional Development in Academic Practice: (PDAP), February 2017.

- Awarded by the Saudi Cultural Bureau in London the academic excellence award, October 2017.

- Attended the "human tissue act: working with human tissues training session", University of Plymouth, February 2016.

- Attended the pipetting skills training offered by Alpha laboratories, University of Plymouth, May 2016.

- Attended various Doctoral College training sessions which include but not limited to: introduction to Latex, submitting to Pearl, leadership and management, presenting to an audience, being human,

research integrity, the what, the where and the which of flipped classroom and faster, smarter student feedback.

Presentations

- ISBT 36^{th} international congress, virtual 2020. Poster presentation titled: **sequencing of the *RHD* gene using MinION™**.

- University of Plymouth Annual postgraduate research event. Devon, United Kingdom. March 2019. Poster presentation titled: **complete sequencing of the RH blood group genes using next generation sequencing**.

- Research and development Conference. Devon, United Kingdom. June 2018. Poster presentation titled: **complete next generation sequencing for Rh genes: establishment of reference *RHD* and *RHCE* alleles**.

- ISBT 35^{th} international congress. Toronto, Canada. June 2018. Poster presentation titled: **complete next generation sequencing for Rh genes: establishment of reference *RHD* and *RHCE* alleles**.

- University of Plymouth annual postgraduate research event. Devon, United Kingdom. March 2018. Oral presentation titled: **complete sequencing of the RH blood group genes using next generation sequencing**.

- ISBT 27^{th} regional congress. Copenhagen, Denmark. June 2017. Poster presentation titled: **complete sequencing of the *RHD* blood**

group gene using next generation sequencing: determination of *RHD* reference sequences.

- University of Plymouth Annual postgraduate research event. Cornwall, United Kingdom. March 2017. Poster presentation titled: **complete *RHD* next generation sequencing: establishment of reference *RHD* allele sequences**.

- University of Plymouth annual postgraduate research event. Cornwall, United Kingdom. March 2016. Poster presentation titled: **investigation of intronic mutations in the *RHD* blood group gene**.

Abstracts

Silja M Tammi*, Wajnat A. Tounsi*, Susanna Sainio, Michele Kiernan, Neil D. Avent, Tracey E. Madgett, Katri Haimila. Next-generation sequencing of 35 RHD variants in 16,253 serologically D negative pregnant women in Finnish population. ***Blood Advances*****, 2020; 4(20):4994-5001.**

Fetal *RHD* screening for targeted routine antenatal anti-D prophylaxis has been implemented in many countries, including Finland, since the 2010s. Comprehensive knowledge of the *RHD* polymorphism in the population is essential for the performance and safety of the anti-D prophylaxis program. During the first 3 years of the national screening program in Finland, over 16,000 samples from RhD2 women were screened for fetal *RHD*; among them, 79 samples (0.5%) containing a maternal variant allele were detected. Of the detected maternal variants, 35 cases remained inconclusive using the traditional genotyping methods and required further analysis by next-generation sequencing (NGS) of the whole *RHD* gene to uncover the variant allele. In addition to the 13 *RHD* variants that have been previously reported in different populations, 8 novel variants were also detected, indicating that there is more variation of *RHD* in the RhD-Finnish population than has been previously known. Three of the novel alleles were identified in multiple samples; thus, they are likely specific to the original Finnish population. National screening has thus provided

new information about the diversity of *RHD* variants in the Finnish population. The results show that NGS is a powerful method for genotyping the highly polymorphic *RHD* gene compared with traditional methods that rely on the detection of specific nucleotides by polymerase chain reaction amplification.

Wajnat A. Tounsi, Tracey E. Madgett and Neil D. Avent. Complete *RHD* next generation sequencing: establishment of reference *RHD* alleles. *Blood Advances*, 2018; 2:2713-2723.

The Rh blood group system (ISBT004) is the second most important blood group after ABO and is the most polymorphic one, with 55 antigens encoded by 2 genes, *RHD* and *RHCE*. This research uses next-generation sequencing (NGS) to sequence the complete *RHD* gene by amplifying the whole gene using overlapping long-range polymerase chain reaction (LR-PCR) amplicons. The aim was to study different *RHD* alleles present in the population to establish reference *RHD* allele sequences by using the analysis of intronic single-nucleotide polymorphisms (SNPs) and their correlation to a specific Rh haplotype. Genomic DNA samples (n=69) from blood donors of different serologically predicted genotypes including $\mathrm{R_1R_1}$ (DCe/DCe), $\mathrm{R_2R_2}$ (DcE/DcE), $\mathrm{R_1R_2}$ (DCe/DcE), $\mathrm{R_1R_z}$ (DcE/DCE), $\mathrm{R_1r}$ (DCe/dce), $\mathrm{R_2r}$ (DcE/dce), and $\mathrm{R_0r}$ (Dce/dce) were sequenced and data were then mapped to the human genome reference sequence hg38. We focused on the analysis of hemizygous samples, as these by definition will only have a single copy of *RHD*. For the 69 samples sequenced, different exonic SNPs were detected that correlate with known variants. Multiple intronic SNPs were found in all samples: 21 intronic SNPs were present in all samples indicating their specificity to the *RHD*DAU0 (RHD*10.00)* haplotype which the hg38 reference sequence encodes. Twenty-three intronic SNPs were found to be $\mathrm{R_2}$ haplotype specific, and 15 were linked to $\mathrm{R_1}$, $\mathrm{R_0}$, and $\mathrm{R_z}$

haplotypes. In conclusion, intronic SNPs may represent a novel diagnostic approach to investigate known and novel variants of the *RHD* and *RHCE* genes, while being a useful approach to establish reference *RHD* allele sequences.

Kelly A. Sillence, Amr J. Halawani, Wajnat A. Tounsi, Kirsty A. Clarke, Michele Kiernan, Tracey E. Madgett, Neil D. Avent. Rapid *RHD* zygosity determination using digital PCR. *Clinical Chemistry* 63(7), 1388-1397.

BACKGROUND: Paternal zygosity testing is used for determining homo- or hemizygosity of *RHD* in pregnancies that are at a risk of haemolytic disease of the fetus and newborn. At present, this is achieved by using real-time PCR or the Rhesus box PCR, which can be difficult to interpret and unreliable, particularly for black African populations.

METHODS: DNA samples extracted from 53 blood donors were analysed using 2 multiplex reactions for *RHD* specific targets against a reference *(AGO1)* to determine gene dosage by digital PCR. Results were compared with serological data, and the correct genotype for 2 discordant results was determined by long-range PCR (LR-PCR), next-generation sequencing, and conventional Sanger sequencing.

RESULTS: The results showed clear and reliable determination of *RHD* zygosity using digital PCR and revealed that 4 samples did not match the serologically predicted genotype. Sanger sequencing and long-range PCR followed by next-generation sequencing revealed that the correct genotypes for samples 729M and 351D, which were serologically typed as $\mathrm{R_1R_2}$ (DCe/DcE), were $\mathrm{R_2r}$ (DcE/dCe) for 729M and $\mathrm{R_1r''}$ (DCe/dcE), $\mathrm{R_0r^y}$ (Dce/dCE), or $\mathrm{R_zr}$ (DCE/dce) for 351D, in concordance with the digital PCR data.

CONCLUSIONS: Digital PCR provides a highly accurate method to rapidly define blood group zygosity and has clinical application in the analysis of Rh phenotyped or genotyped samples. The vast majority of current blood group genotyping platforms are not designed to define zygosity, and thus, this technique may be used to define paternal RH zygosity in pregnancies that are at a risk of haemolytic disease of the fetus and newborn and can distinguish between homo- and hemizygous *RHD*-positive individuals.

Wajnat A. Tounsi, Garry S. Farnham, Vasileios P. Lenis, Silja M. Tammi, Susanna Sainio, Katri Haimila, Neil D. Avent, Tracey E. Madgett. Sequencing of the *RHD* gene using MinION™. ISBT 36^{th} international congress, virtual, 2020.

The Rh blood group system (ISBT004), second most important blood group after ABO, is encoded by two genes, *RHD* and *RHCE*. Recombination, deletion, and point mutations in these two genes generate Rh allelic diversity. Blood group genotyping (BGG) could decrease Rh mistyping and minimise adverse reactions following blood transfusion. In this work, we explore the possibility of applying single molecule sequencing (SMS) in BGG. MinION™, introduced by Oxford Nanopore Technologies in 2016, depends on driving DNA through nanopores embedded in a membrane using a voltage, which changes as different nucleotides pass through. Studies have shown the great potential of the MinION™ real time sequencing, for example, during outbreaks of Ebola and Zika viruses. MinION™ sequencer has also been utilised in ABO genotyping and was successful in differentiating between six ABO genotypes (AA, AO, BB, BO, AB, OO).

We aimed to genotype the *RHD* gene using MinION™ and compare results for the same samples to the NGS data obtained from the Ion Personal Genome Machine™ (Ion PGM™). Our aim was to test the suitability and efficiency of the MinION™ sequencer in BGG.

The *RHD* human reference sequence in the hg38 assembly encodes a vari-

ant *RHD* allele *RHD*DAU0* encoded by c.1136C>T (p.Thr379Met) in exon 9. Therefore, all 13 samples showed a homozygous SNP in exon 9 c.1136T>C (Met379Thr), i.e. none were *RHD*DAU0*. Exonic changes were detected in 12 of the 13 samples encoding variant *RHD* alleles, which included 4 known variants in 5 samples and 6 novel variants in 7 samples. Known variants included *RHD*11*, *RHD*15* (in two samples), *RHD*01W.02*, and *RHD*09.04*. Novel variants included c.829G>A (p.Gly277Arg) in exon 6, c.784delC (p.Gln262fs) in exon 5, two exon 3 mutations c.421delG (p.Val141fs) and c.422T>A (p.Val141), two mutations in exon 9 c.1154G>C (p.Gly385Ala) and c.1163T>G (p.Leu388Arg), c.519C>G (p.Tyr173*) in exon 4, and c.1016G>C (p.Gly339Ala) in exon 7. Reference sequences for the novel *RHD* alleles were established and submitted to GenBank. All exonic and intronic changes detected were compatible with the ones detected by Ion PGM™.

The *RHD* human reference sequence in the hg38 assembly encodes a variant *RHD* allele *RHD*DAU0* encoded by c.1136C>T (p.Thr379Met) in exon 9. Therefore, all 13 samples showed a homozygous SNP in exon 9 c.1136T>C (Met379Thr). One sample showed a wild type *RHD* allele. Exonic changes were detected in 12 samples encoding variant *RHD* alleles, which included 4 known variants in 5 samples and 6 novel variants in 7 samples. Known variants included *RHD*11*, *RHD*15* (in two samples), *RHD*01W.02*, and *RHD*09.04*. Novel variants included c.829G>A (p.Gly277Arg) in exon 6, c.784delC (p.Gln262fs) in exon 5, two exon 3 mutations c.421delG (p.Val141fs)

and c.422T>A (p.Val141), two mutations in exon 9 c.1154G>C (p.Gly385Ala) and c.1163T>G (p.Leu388Arg), c.519C>G (p.Tyr173*) in exon 4, and c.1016G>C (p.Gly339Ala) in exon 7. Reference sequences for the novel *RHD* alleles were established and submitted to GenBank. All exonic and intronic changes detected were compatible with the ones detected by Ion PGM™.

In this research, the *RHD* gene was sequenced in 13 samples using MinION™ single molecule sequencing and the *RHD* allele was determined in all cases. MinION™ is a powerful tool and was successful here in BGG. With the advantage of allele phasing in single molecule sequencing, it is hoped that the more complicated genetic structures, such as exist for the hybrid *RHD*/*RHCE* genes, can be more easily assessed in the future. However, challenges remain with developing a straightforward user-friendly data analysis pipeline for translating this work into the clinic.

Wajnat A. Tounsi, Tracey E. Madgett, Neil D. Avent. Complete next generation sequencing for Rh genes: establishment of reference *RHD* and *RHCE* allele. University of Plymouth Annual Postgraduate Research Event, Plymouth, United Kingdom, 2019.

The Rh blood group system (ISBT004) is the second most important blood group after ABO. Two closely related genes, *RHD* and *RHCE*, encode 55 different antigens. Blood group genotyping could decrease Rh mistyping and eventually minimise the adverse reactions following blood transfusion, especially for blood transfusion dependent patients.

We aimed to use LR-PCR to amplify the Rh genes to get a full sequence including promoter, introns and all exons. We focused on establishing reference alleles for the *RHD* and *RHCE* genes by studying intronic SNPs in both genes and their relationship to specific Rh haplotypes.

Genomic DNA samples (n=92) from blood donors and the ISBT 1996 workshop of different phenotypes were sequenced using the Ion Personal Genome Machine™ (Ion PGM™). The *RHD* gene was sequenced from all samples and the *RHCE* gene was sequenced from blood donor samples. Data was then mapped to the hg38 reference sequence and analysed using the CLC Workbench 9.5.

Multiple exonic SNPs were detected that encode 12 *RHD* variant alleles. Due to **DAU0* (*RHD*10.00*) being the hg38 reference sequence for the

RHD gene, 21 homozygous SNPs were detected in all samples which are thought to be specific to the reference allele. Compared to the reference sequence, multiple intronic SNPs were detected that are suspected to be a haplotype specific. Twenty-three SNPs were homozygous SNPs in all samples with the R_2 haplotype and 15 SNPs were homozygous in all samples with R_1 haplotype. Intronic SNP analysis of the *RHCE* gene revealed over 100 intronic SNPs and confirmed the presence of the 109 bp insertion in intron 2 in samples with R_1 haplotype. More than 50 intronic SNPs were detected in all R_2 and r samples.

In this research, 92 samples were sequenced for the *RHD* gene and the *RHCE* gene to study different alleles present in the population to establish reference allele sequences by utilising the analysis of intronic SNPs and their correlation to a specific Rh haplotype. Intronic SNPs are suspected to be related to a specific haplotype, which may represent novel diagnostic approaches to investigate known and novel variants of both genes.

Wajnat A. Tounsi, Tracey E. Madgett, Neil D. Avent. Complete next generation sequencing for Rh genes: establishment of reference *RHD* and *RHCE* alleles. Research and Development conference, Plymouth, UK, 2018.

The Rh blood group system (ISBT004) is the second most important blood group after ABO. Two closely related genes, *RHD* and *RHCE*, encode 55 different antigens. Recombination, deletion, and point mutations in these two genes generate Rh allelic diversity, with the eight most common Rh haplotypes being: $\mathrm{R_1}$ (DCe), $\mathrm{R_2}$ (DcE), $\mathrm{R_0}$ (Dce), $\mathrm{R_z}$ (DCE), r (dce), $\mathrm{r^y}$ (dCE), r′ (dCe), and r″ (dcE). The RhD antigen is the most clinically significant antigen due to its high immunogenicity and being the main cause of haemolytic disease of the fetus and newborn.

Serological testing is fast, cost friendly, and efficient; however, limited by many factors like the availability of antisera and reactivity of the antibodies. Blood group genotyping (BGG) could be widely used in transfusion practice where serology tests may fail to clarify issues or resolve discrepancies.

Different DNA microarray-based tests were introduced that enable genotyping of variant blood groups by targeting specific single nucleotide polymorphisms (SNPs). Although these assays are very accurate, they are designed to target certain nucleotides while novel variants remain unknown. Complete DNA sequencing could be the most relevant technique to thor-

oughly study blood group variations.

In this research we aimed to use long-range PCR (LR-PCR) to amplify the complete Rh genes and then next generation sequencing (NGS) to get a full sequence including promoter, introns and all exons, that would enable us to study *RHD* and *RHCE* gene variations and detect novel alleles.

Genomic DNA samples (n=92) from blood donors of different phenotypes were sequenced using the Ion Personal Genome Machine™ (Ion PGM™) for *RHD* and *RHCE*. Data analysis revealed intronic SNPs that link to a specific Rh haplotype and have led to the establishment of reference alleles. NGS may represent a novel diagnostic approach to investigate known and novel variants of both genes.

The Rh blood group system (ISBT004) is the second most important blood group after ABO that is encoded by the *RHD* and the *RHCE* genes. We aimed to establish a sequencing method using next generation sequencing to get a full sequence of the Rh genes including promoter, introns and all exons, that would enable us to study *RHD* and *RHCE* gene variations and detect novel alleles.

Wajnat A. Tounsi, Tracey E. Madgett, Neil D. Avent. Complete next generation sequencing for Rh genes: establishment of reference *RHD* and *RHCE* allele. ISBT 35^{th} International congress, Toronto, Canada, 2018.

The Rh blood group system (ISBT004) is the second most important blood group after ABO. Two closely related genes, *RHD* and *RHCE*, encode 55 different antigens. Recombination, deletion, and point mutations in these two genes generate Rh allelic diversity and generate the eight most common Rh haplotypes which include: R_1 (DCe), R_2 (DcE), R_0 (Dce), R_z (DCE), r (dce), r^y (dCE), r′ (dCe), and r″ (dcE). Blood group genotyping could decrease Rh mistyping and eventually minimise the adverse reactions following blood transfusion, especially for blood transfusion dependent patients. Previous studies have used exome sequencing to identify Rh variation but the high homology between the *RHD* and *RHCE* genes could make it challenging to analyse data, especially exons 8 and 10, where there are no amino acid differences between the two genes.

We aimed to use LR-PCR to amplify the Rh genes to get a full sequence including promoter, introns and all exons. We focused on establishing reference alleles for the *RHD* and *RHCE* genes by studying intronic SNPs in both genes and their relationship to specific Rh haplotypes.

Genomic DNA samples (n=92) from blood donors and the ISBT 1996 workshop of different phenotypes were sequenced using the Ion Personal Genome Machine™ (Ion PGM™). The *RHD* gene was sequenced from all samples

and the *RHCE* gene was sequenced from blood donor samples. Data was then mapped to the hg38 reference sequence and analysed using the CLC Workbench 9.5.

Multiple exonic SNPs were detected that encode 12 *RHD* alleles which include *RHD*01W.1*, *RHD*01W.02*, *RHD*01W.3*, *RHD*03.01*, *RHD*25*, *RHD*04.04*, *RHD*06.01*, *RHD*03.01*, *RHD*07.01*, *RHD*17.01*, and *RHD*17.02*. One novel allele was identified, in which one *RHD* hemizygous sample showed SNP L110P that encodes *RHD*07.01* and *RHD-RHCE(4)-RHD* that encodes *RHD*17.02*, which means that this *RHD* allele is a hybrid between two variant *RHD* alleles. Due to RHD*DAU0 (*RHD*10.00*) being the hg38 reference sequence for the *RHD* gene, 21 homozygous SNPs were detected in all samples which are thought to be specific to the reference allele. Compared to the reference sequence, multiple intronic SNPs were detected that are suspected to be a haplotype specific. Twenty-three SNPs were homozygous SNPs in all samples with the R_2 haplotype and 15 SNPs were homozygous in all samples with R_1 haplotype.

Intronic SNP analysis of the *RHCE* gene revealed over 100 intronic SNPs in intron 2 and confirmed the presence of the 109 bp insertion in intron 2 in samples with the R_1 haplotype. More than 50 intronic SNPs were detected in all R_2 and r samples.

In this research, 92 samples were sequenced for the *RHD* gene and the

RHCE gene to study different alleles present in the population to establish reference allele sequences by utilising the analysis of intronic SNPs and their correlation to a specific Rh haplotype. Intronic SNPs are suspected to be related to a specific haplotype, which may represent novel diagnostic approaches to investigate known and novel variants of both genes.

Wajnat A. Tounsi, Tracey E. Madgett, Neil D. Avent. Complete next generation sequencing for Rh genes: establishment of reference *RHD* and *RHCE* alleles. University of Plymouth Annual Postgraduate Research Event, Plymouth, United Kingdom, 2018.

The Rh blood group system (ISBT004) is the second most important blood group after ABO. Two closely related genes, *RHD* and *RHCE*, encode 55 different antigens. Recombination, deletion, and point mutations in these two genes generate Rh allelic diversity. Complete DNA sequencing could be the most relevant technique to thoroughly study blood group variations. Complete blood group genotyping could decrease Rh mistyping and eventually minimise the adverse reactions following blood transfusion.

We aimed to use LR-PCR to amplify the Rh genes to get a full sequence including promoter, introns and all exons. We focused on establishing reference alleles for the *RHD* and *RHCE* genes by studying intronic SNPs in both genes and their relationship to specific Rh haplotypes.

Genomic DNA samples (n=69) from blood donors and (n=23) from ISBT 1996 workshop of different phenotypes were sequenced for the *RHD* gene and 31 samples of blood donors were sequenced for the *RHCE* gene using the Ion Personal Genome Machine™ (Ion PGM™).

Multiple exonic SNPs were detected that encode 12 *RHD* alleles which in-

clude *RHD*01W.1*, *RHD*01W.02*, *RHD*01W.3*, *RHD*03.01*, *RHD*25*, *RHD*04.04*, *RHD*06.01*, *RHD*03.01*, *RHD*07.01*, *RHD*17.01*, and *RHD*17.02*. One novel allele was identified, in which one *RHD* hemizygous sample showed a hybrid between two variant *RHD* alleles which include *RHD*07.01* and *RHD*17.02*. Multiple intronic SNPs were identified in the *RHD* and *RHCE* genes, which are suspected to be haplotype specific.

In this research, 92 samples were sequenced for the *RHD* gene and 31 for the *RHCE* gene to study different alleles present in the population to establish reference allele sequences by utilising the analysis of intronic SNPs and their correlation to a specific Rh haplotype. Intronic SNPs are suspected to be related to a specific haplotype, which may represent novel diagnostic approaches to investigate known and novel variants of both genes.

Wajnat A. Tounsi, Tracey E. Madgett, Neil D. Avent. Complete sequencing of the *RHD* blood group gene using next generation sequencing. ISBT 27^{th} regional congress, Copenhagen, Denmark, 2017.

The Rh blood group system (004) is the second most important blood group after ABO. Two closely related genes, *RHD* and *RHCE*, encode over 50 different antigens in the Rh system. Recombination, deletion, and point mutations in these two genes generate Rh allelic diversity which makes the Rh blood group the most polymorphic blood group. In the past decade, different DNA microarray-based tests were introduced that enable genotyping of variant blood group specific single nucleotide polymorphisms (SNPs). However, these assays have limitations because they target certain nucleotides or DNA regions through PCR, while novel ones remain unknown. Complete sequencing could be the most relevant technique to thoroughly study blood group variations. Complete blood group genotyping could decrease Rh mistyping and eventually minimise the adverse reactions following blood transfusion especially for blood transfusion dependent patients.

This research aims to use Next Generation Sequencing (NGS) to sequence the *RHD* gene to detect *RHD* variants present in the population, which will help expand the knowledge about the underlying molecular mechanism of these variants. This research also focuses on investigating intronic SNPs that could be linked to a specific haplotype that might be used in the

future to predict Rh haplotype.

Genomic DNA samples from blood donors of different phenotypes including 6 $\mathrm{R_1R_1}$, 6 $\mathrm{R_2R_2}$, 7 $\mathrm{R_1R_2}$, 6 $\mathrm{R_1r}$, 6 $\mathrm{R_2r}$, and 6 $\mathrm{R_0r}$ were sequenced using the Ion Personal Genome Machine™ (Ion PGM™). All samples were tested for *RHD* zygosity using digital PCR. The *RHD* gene was amplified in 6 overlapping amplicons using *RHD*-specific primers. The 200-base pair read sequencing libraries were prepared and then sequenced on the Ion PGM™ using a 316 chip. Data was then mapped to the hg38 human genome reference sequence and analysed using the CLC Workbench 9.5.

In one $\mathrm{R_2R_2}$ sample, one exon 9 SNP 25321889 G>C was detected resulting in the amino acid change Gly385Ala which is linked to weak D type 2. Multiple intronic SNPs were detected in all samples in which 15 homozygous SNPs were present in all 37 samples, which could indicate an error in the human reference sequence. Another 19 SNPs were present in all $\mathrm{R_2R_2}$ and $\mathrm{R_2r}$ samples as homozygous SNPs and in $\mathrm{R_1R_2}$ samples as heterozygous SNPs. Fourteen intronic SNPs were present in all $\mathrm{R_1R_1}$, $\mathrm{R_1r}$ and $\mathrm{R_0r}$ as homozygous SNPs and heterozygous SNPs in all $\mathrm{R_1R_2}$ samples. Sixteen heterozygous intronic SNPs were only present in the $\mathrm{R_2R_2}$ weak D type 2 sample. Intronic SNPs are suspected to be linked to a specific haplotype, which could be used in the future to establish an assay to genotype Rh antigens without the need to fully sequence the Rh genes.

In this research, 37 samples were sequenced on the Ion PGM™ to study *RHD* mutations and assist the *RHD* variations present in the population. Intronic SNPs were also analysed to determine their relation to specific haplotypes. Multiple SNPs are suspected to be related to a specific haplotype; however, more samples and different Rh phenotypes need to be sequenced to confirm this.

Wajnat A. Tounsi, Tracey E. Madgett, Neil D. Avent. Complete sequencing of the *RHD* blood group gene using next generation sequencing. University of Plymouth Annual Postgraduate Research Event, Cornwall, United Kingdom, 2017.

The Rh blood group system (004) is the second most important blood group after ABO. Two closely related genes, *RHD* and *RHCE*, encode over 50 different antigens in the Rh system. Recombination, deletion, and point mutations in these two genes generate Rh allelic diversity which makes the Rh blood group the most polymorphic blood group. This research aims to use Next Generation Sequencing (NGS) to sequence the *RHD* gene to detect *RHD* variants present in the population, which will help expand the knowledge about the underlying molecular mechanism of these variants.

Genomic DNA samples from blood donors of different phenotypes including 6 $\mathrm{R_1R_1}$, 6 $\mathrm{R_2R_2}$, 7 $\mathrm{R_1R_2}$, 6 $\mathrm{R_1r}$, 6 $\mathrm{R_2r}$, and 6 $\mathrm{R_0r}$ were sequenced using the Ion Personal Genome Machine™ (Ion PGM™). Data was then mapped to the hg38 human genome reference sequence and analysed using the CLC Workbench 9.5.

For 37 samples sequenced, one exon 9 SNP 25321889 G>C was detected in one R2R2 sample resulting in the amino acid change Gly385Ala which is linked to weak D type 2. Multiple intronic SNPs were detected in all samples, with some of them suspected to be linked to a specific haplotype. These intronic SNPs could be used in the future to establish an assay

to genotype Rh antigens without the need to fully sequence the Rh genes. To confirm the relation of these intronic SNPs to a specific Rh haplotype, more samples and variant Rh phenotypes need to be sequenced.

Wajnat A. Tounsi, Amr J. Halawani, Tracey E. Madgett, Neil D. Avent. Investigation of intronic mutations in the *RHD* blood group gene. University of Plymouth Annual Postgraduate Research Event, Cornwall, United Kingdom, 2016.

The Rh blood group system is the most important protein blood group system and the most polymorphic one. Rh system includes five leading antigens: RhD, RhC, Rhc, RhE, and Rhe. Only two highly homologous genes, *RHD* (encodes the RhD protein) and *RHCE* (encodes the RhC/c and RhE/e proteins), give rise to 49 different variants.

Previous next generation sequencing (NGS) work on the *RHD* gene in our laboratory has identified seven different *RHD* intronic single nucleotide polymorphisms (SNPs) (Halawani, 2015), suspected to be linked to the R2 (DcE) haplotype. Four of the seven SNPs have been tested to confirm if they are linked to the R2 haplotype. SNPs tested are: 25,614,400 C>G, 25,621,980 C>T, 25,627,066 C>G and 25,648,349 T>C. Genomic DNA samples from blood donors of different phenotypes were tested (6 $\mathrm{R_2R_2}$, 4 $\mathrm{R_1R_1}$, 1 $\mathrm{R_2r}$, and 1 $\mathrm{R_1r}$).

RHD-specific primers amplified the regions around the SNPs and the PCR amplicons were subjected to Sanger sequencing. Results were aligned with the hg38 human genome reference sequence. For all four SNPs analysed, the $\mathrm{R_2R_2}$ and $\mathrm{R_2r}$ samples showed the mutations but not the $\mathrm{R_1R_1}$ and $\mathrm{R_1r}$ samples. These SNPs could be used in the future to predict Rh

antigen status without the need for full *RHD* and *RHCE* genes sequencing.

Future aims are to use NGS to study Rh genetic basis. Improvements will be made to the long-range PCR and the NGS protocols to obtain a better coverage depth of sequences to identify the frequency of mutations and any novel SNPs.