

A novel genotyping approach to improve transfusion support for patients with HLA and/or HPA alloantibodies

Susan Rosemary Davey

A thesis submitted in partial fulfilment of the requirements of
the University of the West of England, Bristol for the degree of
Professional Doctorate in Biomedical Science

Faculty of Health and Applied Sciences,
University of the West of England, Bristol

July 2017

Word count 34,168

Acknowledgements.....9

List of Abbreviations10

List of Tables11

List of Figures13

1. Introduction15

 1.1 Human Leucocyte Antigens (HLA)15

 1.1.1 HLA class I gene structure16

 1.1.2 HLA class I nomenclature17

 1.2 Human platelet antigens (HPA)18

 1.3 Clinical relevance of HLA and HPA in transfusion.....22

 1.3.1 Immune platelet refractoriness (IPR)22

 1.3.2 Foetal or neonatal alloimmune thrombocytopenia.....22

 1.3.3 Transfusion support for patients with HLA and or HPA alloantibodies24

 1.3.3.1 Cross-match negative25

 1.3.3.2 Antigen avoidance.....25

 1.3.3.3 HLA matching26

 1.3.3.4 HLA epitope matching26

 1.4 HLA and HPA genotyping – current state26

 1.4.1 Patient genotyping27

 1.4.1.1 Patient HLA typing.....28

 1.4.1.2 Patient HPA typing.....28

 1.4.2 Donor genotyping.....28

 1.4.2.1 Donor HLA typing28

 1.4.2.2 Donor HPA typing29

 1.5 Requirement for the improvement of HLA and HPA genotyping29

 1.5.1 Increasing the resolution of HLA genotyping.....29

 1.5.2 Expansion of HPA genotyping.....30

 1.5.3 HPA whole gene sequencing31

 1.6 Next Generation Sequencing.....31

 1.6.1 HLA genotyping by NGS.....35

 1.6.1.1 HLA imputation from whole genome or exome sequence data35

 1.6.1.2 Targeted HLA sequencing36

 1.6.1.3 Sequencing platforms.....37

 1.6.2 HPA genotyping by NGS38

 1.7 Aims and objectives.....38

1.8 Hypothesis.....	38
2. Methods and Materials	39
2.1 Samples	39
2.1.1 Donor samples for HLA genotyping	39
2.1.2 Control DNA for HPA genotyping	39
2.1.3 Patient DNA for HPA genotyping	39
2.2 DNA extraction of donor samples	39
2.3 DNA purification.....	39
2.4 DNA quantification	40
2.4.1 Quant-iT™ Assay kit, Broad Range (Invitrogen)	40
2.4.2 Qubit® dsDNA BR assay (Life Technologies)	40
2.5 DNA normalisation	40
2.5.1 Manual normalisation	40
2.5.2 Automated normalisation	40
2.6 HLA genotyping by Next Generation Sequencing.....	41
2.6.1 PCR amplification of HLA class I genes	41
2.6.1.1 TruSight™ HLA Sequencing panel	41
2.6.1.2 In-house PCR protocol	41
2.6.2 PCR amplicon assessment	42
2.6.2.1 Agarose gel electrophoresis	42
2.6.2.2 Quantification of PCR amplicons	42
2.6.3 Amplicon pooling.....	42
2.6.4 DNA library preparation	43
2.6.3.1 TruSight™ HLA sequencing panel library preparation	44
2.6.3.2 NGSgo® Library preparation.....	45
2.6.5 Assessment of the pooled amplicon libraries.....	45
2.6.5.1 2100 Bioanalyser	45
2.6.5.2 Qubit®	45
2.6.6 Sequencing on the MiSeq.....	46
2.6.7 Data analysis.....	46
2.7 HPA genotyping by Targeted Next Generation Sequencing	46
2.7.1 Assay Design.....	46
2.7.2 Targeted enrichment.....	47
2.7.3 Pooling of libraries.....	48
2.7.4 Sequencing on the MiSeq.....	48
2.7.5 Data analysis.....	49

3. HLA typing of platelet donors by next generation sequencing.....	50
3.1 Introduction	50
3.2 Materials & Methods	51
3.2.1 Sample preparation	51
3.2.2 PCR amplification	51
3.2.3 PCR assessment and amplicon pooling.....	52
3.2.4 Library preparation.....	52
3.2.5 Sequencing on the MiSeq.....	53
3.2.6 Data analysis.....	53
3.2.7 HLA concordance with historical HLA types	54
3.3 Results.....	55
3.3.1 PCR amplification	55
3.3.2 Library quantification	56
3.3.3 Sequencing quality	57
3.3.4 HLA typing results.....	59
3.3.4.1 Low sequencing depth.....	61
3.3.4.2 HLA ambiguity	62
3.3.4.3 HLA allele imbalance	63
3.3.4.4 Novel exon mutations	64
3.3.4.5 Intron mismatches.....	64
3.3.5 HLA Concordance	66
3.4 Discussion.....	67
3.4.1 Amplification failure	67
3.4.2 Ambiguity	67
3.4.3 Allele imbalance	68
3.4.4 Low coverage	68
3.4.5 Intron mismatches.....	71
3.4.6 Exon mismatches.....	71
3.5 Conclusions	71
4. HLA epitope matching: allele prediction or NGS typing?.....	73
4.1 Introduction	73
4.2 Materials and methods	74
4.2.1 HLA allele prediction	74
4.2.1.1 Selecting the first allele in the HLA string	74
4.2.1.2 Selecting the most frequent allele.....	74
4.2.2 Comparison of predicted versus defined HLA alleles	74

4.2.3 Determining the impact of HLA allele prediction	74
4.3 Results	75
4.3.1 HLA allele predictions	75
4.3.2 Ethnicity and allele predictions	79
4.3.3 Haplotype frequencies and allele prediction	79
4.3.4 Impact of HLA allele prediction errors	81
4.4 Discussion	82
4.4.1 Accuracy of allele prediction	82
4.4.2 Impact of prediction errors	85
5. HLA frequencies and population analysis	87
5.1 Introduction	87
5.2 Materials & Methods	87
5.2.1 Confirmatory sequencing of samples with putative novel mutations	88
5.2.1.1 Sample selection	88
5.2.1.2 DNA preparation, PCR amplification and amplicon pooling	88
5.2.1.3 Library preparation and sequencing	88
5.2.2 Re-sequencing of samples with HLA types previously rejected	89
5.2.2.1 Sample selection	89
5.2.2.2 DNA preparation, PCR amplification and amplicon pooling	89
5.2.2.3 Library preparation and sequencing	89
5.2.3 Sequence data analysis	89
5.2.4 Population data analysis	89
5.2.4.1 HLA allele frequencies	89
5.2.4.2 HLA haplotype frequencies	90
5.3 Results	90
5.3.1 Repeat amplifications	90
5.3.2 Sequencing libraries	91
5.3.3 HLA sequencing results	91
5.3.3.1 Confirmatory sequences of samples with putative novel mutations	91
5.3.3.2 Impact of Exon mutations	93
5.3.3.3 Re-sequencing samples with HLA types previously rejected	93
5.3.4 Population data analysis	95
5.4 Discussion	98
5.4.1 Outcome of re-sequencing	98
5.4.2 Confirmation of novel mutations	99
5.4.3 Population data analysis	100

6. HPA genotyping by NGS	101
6.1 Introduction	101
6.2 Materials and methods	102
6.2.1 Samples	102
6.2.1.1 Control DNA	102
6.2.1.2 Patient DNA	103
6.2.2 Historical HPA genotyping	103
6.2.3 Assay Design	103
6.2.4 Targeted enrichment and sequencing	103
6.2.4.1 DNA sample preparation	103
6.2.4.2 DNA library preparation	103
6.2.5 Data analysis	104
6.2.5.1 HPA genotyping	104
6.2.5.2 Novel mutation detection	105
6.3 Results	105
6.3.1 Assay design	105
6.3.3 Restriction enzyme digestion	107
6.3.4 Validation of DNA library preparation	107
6.3.4 HPA genotyping	108
6.3.5 Sequence quality	110
6.3.5.1 Sequencing coverage	110
6.3.5.2 Coverage per HPA system	111
6.3.6 Impact of GC content on read depth	112
6.3.7 Allele balance	114
6.3.8 Novel allele detection	115
6.4 Discussion	116
6.4.1 Assay design	116
6.4.2 Sequence quality	117
6.4.2.1 Depth of coverage	117
6.4.2.2 Depth of coverage per HPA system	117
6.4.2.3 Mean Depth of coverage per sample	118
6.4.2.4 Allele balance	119
6.4.3 HPA genotyping	119
6.4.4 Novel mutation detection	119
7. Discussion	121
7.1 HLA typing and NGS	121

7.1.1 HLA typing platelet donors by NGS	121
7.1.2 HLA typing platelet recipients by NGS	125
7.2 Impact of NGS on HEM	125
7.3 The wider impact of HLA typing by NGS	127
7.3.1 Whole gene HLA sequencing	127
7.3.2 HLA allele and haplotype frequencies	127
7.3.3 Novel HLA allele detection	128
7.4 HPA genotyping by NGS	128
7.5 Identification of platelet donors with rare genotypes	129
7.6 Novel HPA allele detection	129
7.7 Conclusions	130
8. References	131
9. Publications 2013-2017	144
9.1 Papers	144
9.2 Abstracts/Orals/Posters	144
10. Appendix	146
10.1 Response from FREC	146
10.2 Donor health check	148
10.3 Welcome booklet	149
10.4 Page 9 of User guide for H&I diagnostic services	150
10.5 INF256 – Histocompatibility testing for platelet transfusion patients	151
10.6 INF283 Platelet groups and antibodies in pregnancy	152

Abstract

Patients who require platelet transfusion support but have become sensitised to Human Leucocyte Antigens (HLA) or Human Platelet Antigens (HPA) require suitably matched or selected products to ensure an adequate increase in their platelet count following transfusion. In England this affects approximately 1500 individuals per annum, with over 25,000 matched or selected platelet units provided for these patients.

Provision of compatible products is often challenging, and requires significant resources from the blood service, with approximately 5000 apheresis platelet donors genotyped for HLA and HPA each year. Current typing technology results in restricted HLA and HPA genotypes for both patients and donors, thereby limiting the ability to provide fully compatible products. This study set out to develop and implement next generation sequencing (NGS) technology to enhance the HLA and HPA definition of both platelet donors and recipients.

An NGS based method was designed and developed for high throughput, allele level HLA class I genotyping and used to evaluate the impact of NGS technology on the selection of platelet donors using HLA epitope matching (HEM). In addition, an alternative NGS approach was designed to simultaneously sequence the six genes that code for glycoproteins expressing HPA in order to define all known HPA systems in both donor and patient samples.

Allele level HLA-A, -B and -C genotypes were generated for 519 platelet donors by NGS. A critical evaluation of algorithms used to predict alleles from low to medium resolution HLA types demonstrated that NGS was more accurate when determining HLA epitopes for the selection of platelets by HEM. The HLA genotyping data obtained was used to establish previously undefined HLA allele and haplotype frequencies in the English platelet donor population. This thesis also includes the first reported NGS based method for the simultaneous genotyping of HPA-1 to HPA-29, with the additional capability of novel HPA detection.

NGS has been shown to significantly improve the definition of both HLA and HPA genetic systems and will provide a number of future benefits for laboratories and the patients they support, including provision of well matched transfusion products, the detection of rare or novel polymorphisms and increased knowledge of HLA and HPA frequencies.

Acknowledgements

I would like to thank my supervisors, Professor Michael Lodomery, Dr Colin Brown and Dr Cristina Navarrete for their support and guidance provided over the past five years.

I would also like to thank the laboratory staff and colleagues at NHSBT for their help and encouragement during my studies. In particular, I would like to mention John Ord who provided assistance with data management, as well as Zareen Deplano, Monica Kyriacou and Jasmine Rizvi who helped me with the automation and rescued my final library preparation during a power failure!

Finally, thanks to my friends and family who have supported me through the stresses and strains of doctoral studies. A special thank you to my husband Nick for proof reading, James for helping me with some figures and Will for keeping me company when we were doing our 'homework' together. I look forward to having my weekends back and spending more time with them.

List of Abbreviations

bp	Base pairs
DNA	Deoxyribose nucleic acid
gDNA	Genomic DNA
DoC	Depth of coverage
DSA	Donor specific antibody
ECD	Enrichment control DNA
EQA	External quality assessment
FNAIT	Foetal and neonatal alloimmune thrombocytopenia
GP	Glycoprotein
HEM	HLA epitope matching
HLA	Human leucocyte antigen
HPA	Human platelet antigen
HWE	Hardy-Weinberg equilibrium
IMGT	IMMunoGeneTics
IPD	Immunopolymorphism database
IPR	Immune platelet refractoriness
MAIPA	Monoclonal antibody immobilisation of platelet antigens
MHC	Major histocompatibility complex
MSR	MiSeq reporter
NFW	Nuclease free water
NGS	Next generation sequencing
NHSBT	National Health Service Blood and Transplant
PCR	Polymerase chain reaction
PCR-SBT	Polymerase chain reaction sequencing based typing
PCR-SSOP	Polymerase chain reaction with sequence specific oligonucleotide probes
PCR-SSP	Polymerase chain reaction with sequence specific primers
SNP	Single nucleotide polymorphism
SPB	Sample purification beads
TIFF	Tag Image File Format
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organisation
H&I	Histocompatibility & Immunogenetics

List of Tables

Table 1.1	Levels of resolution for HLA typing
Table 1.2	Genes that encode glycoproteins (GP) expressing human platelet antigens
Table 1.3	The human platelet antigen (HPA) system
Table 1.4	Summary of options for selecting compatible platelets
Table 1.5	Molecular techniques used for HLA and HPA genotyping
Table 1.6	Advantages and disadvantages of NGS approaches reported for HLA typing
Table 2.1	Comparison of commercial kits employed for DNA library preparation
Table 2.2	Target sequences employed for the bespoke HaloPlexHS design
Table 3.1	The PCR protocol and library preparation methods used for each plate of DNA
Table 3.2	Details of the pooled amplicon libraries combined for each MiSeq run
Table 3.3	Quality parameters and respective acceptance criteria for data analysis using NGSengine analysis software
Table 3.4	Mean concentration of PCR amplicons obtained for each HLA locus
Table 3.5	Mean concentrations and fragment sizes determined for pooled DNA libraries prepared from each plate
Table 3.6	MiSeq parameters observed for each sequencing run
Table 3.7	Summary of quality parameters observed for HLA sequence data
Table 3.8	The number of successful HLA types obtained per plate
Table 3.9	Putative novel mutations observed in plates #464 and #467
Table 3.10	HLA alleles resolved using NGSengine v1.10
Table 3.11	Details of NGS results found to be discrepant with the historical typing data
Table 4.1	Example of HLA alleles predicted from a medium resolution HLA type
Table 4.2	Impact of donor ethnicity on HLA allele prediction using population frequencies
Table 4.3	Prediction of HLA-A*29 alleles from presumed haplotypes
Table 4.4	Prediction of HLA-B*44 alleles from presumed haplotypes
Table 4.5	Mismatched epitopes for each allele prediction discrepancy observed
Table 5.1	HLA alleles resolved following re-sequencing
Table 5.2	Number of HLA sequences observed with novel mutations

Table 5.3	Novel HLA alleles with fully phased sequences
Table 5.4	Effect of nucleotide substitutions for the three novel exon mutations identified
Table 5.5	Reasons for exclusion of donors from the population data analysis
Table 5.6	HLA-A, -B and -C allele frequencies in English platelet donors
Table 5.7	P-values observed for deviation from Hardy-Weinberg equilibrium
Table 6.1	HPA systems previously defined in the sample cohort
Table 6.2	Detail of samples used for each HaloPlexHS experiment
Table 6.3	Regions not amplified by the original HaloPlexHS design
Table 6.4	Summary of final HaloPlexHS design ID 28048-1446018088
Table 6.5	Mean read depth and percentage GC content in the encoding exon for each HPA system
Table 6.6	Nucleotide differences identified by SureCall using pair analysis between samples from mother and baby of a suspected FNAIT case

List of Figures

- Figure 1.1 Schematic showing the HLA regions of the human MHC
- Figure 1.2 The number of HLA alleles recorded on the IMGT/HLA database
- Figure 1.3 Cartoon showing a typical HLA class I gene and the structures each exon encodes
- Figure 1.4 Nomenclature of HLA alleles
- Figure 1.5 Cartoon representing the glycoprotein GPIIIa/GPIIb and respective HPA systems
- Figure 1.6 Summary of the laboratory investigation for platelet refractoriness
- Figure 1.7 DNA library preparation for the Illumina MiSeq
- Figure 1.8 Cluster formation on a glass slide following clonal amplification
- Figure 1.9 Alignment and analysis of sequencing reads
- Figure 1.10 A typical primer map for the amplification of HLA class I genes
- Figure 2.1 Location map of in-house PCR primers for HLA class I amplification
- Figure 2.2 Summary of the PCR amplification and pooling workflow
- Figure 2.3 Summary of workflow for DNA library preparation
- Figure 2.4 Overview of the HaloPlex target-enrichment sequencing sample preparation workflow
- Figure 3.1 Summary of donor sample preparation prior to PCR amplification
- Figure 3.2 Example of HLA PCR amplicons run on a 1% agarose gel
- Figure 3.3 Atypical fragment size distribution for pooled amplicon libraries prepared from plates #462 and #464
- Figure 3.4 Percentage mappability of sequences generated to the respective HLA reference sequences
- Figure 3.5 The percentage of sequences failing to meet quality acceptance criteria
- Figure 3.6 The number of samples with rejected sequences
- Figure 3.7 A summary of reasons for sequence rejection
- Figure 3.8 Percentage of sequences rejected due to poor per base depth of coverage
- Figure 3.9 Ambiguity with HLA-B*44 alleles observed following TruSightHLA amplification
- Figure 3.10 A screenshot of an imbalanced sequence
- Figure 3.11 Reasons for intron mismatches to the reference sequence

- Figure 3.12 Insert size, sequence mappability and read depth
- Figure 4.1 Discrepancies in HLA allele prediction
- Figure 4.2 Discrepant alleles observed and number of samples affected
- Figure 4.3 Number and type of epitope mismatch between NGS type versus predicted HLA allele
- Figure 5.1 Self-declared ethnicity of the 540 apheresis platelet donors HLA typed in this study
- Figure 5.2 Gel image of amplicons taken from the first and last row of each repeat plate
- Figure 5.3 Example of sequence containing a novel mutation with lack of phase
- Figure 5.4 The percentage of sequences rejected for each HLA locus in plate #RPT2
- Figure 5.5 The number of phased regions observed for sequences in plate #RPT2
- Figure 5.6 A example of uneven coverage and resulting poor phasing across the gene
- Figure 5.7 A comparison of the three most frequent HLA allele groups
- Figure 6.1 HPA genotypes of control DNA
- Figure 6.2 Custom HaloPlexHS design quality control
- Figure 6.3 Validation of restriction enzyme digestion using the Bioanalyser 2100
- Figure 6.4 Validation Bioanalyser profile of the pooled DNA library from experiment 3
- Figure 6.5 HPA genotyping results determined using the bespoke HaloPlexHS design
- Figure 6.6 Polymorphisms of the HPA systems sequenced using HaloplexHS
- Figure 6.7 Range of mean per base depth of coverage observed for each sample
- Figure 6.8 Depth of coverage observed for each HPA system
- Figure 6.9 Correlation between mean per base depth of coverage and % GC content
- Figure 6.10 Mean allele balance observed for each heterozygous HPA system
- Figure 6.11 Cartoon showing the location of novel SNP in exon 10 of the *ITGB3* gene detected in patient BB
- Figure 7.1 The current NGS process flow for HSCT donors alongside the NGS protocol designed for platelet donors in this study

1. Introduction

For the majority of patients who require a transfusion of blood or blood components the main consideration for compatibility is likely to be the ABO and Rh blood group systems (Williamson and Devine, 2013). However, following exposure during transfusion, transplantation or pregnancy some individuals may become sensitised to other blood borne antigens such as those expressed on leucocytes and platelets i.e. Human Leucocyte Antigens (HLA) and Human Platelet Antigens (HPA), with the majority of individuals producing antibodies to HLA rather than HPA (Vassallo and Norris, 2016). These patients then require HLA and/or HPA matched or selected products to avoid adverse transfusion reactions resulting from the presence of these donor specific antibodies (Brown and Navarrete, 2011). The provision of these products is often difficult and can require significant resources from the blood service (Kopko *et al.*, 2015).

1.1 Human Leucocyte Antigens (HLA)

Human leucocyte antigens (HLA) are transmembrane glycoproteins intimately involved in the adaptive immune response, presenting processed peptide from pathogens or altered self to the immune system (Dyer *et al.*, 2013). There are two main types of HLA molecule, HLA class I and HLA class II, differentiated by their molecular structure and function and both characterised by their extensive polymorphism (Dyer *et al.*, 2013). The genes coding for the HLA are located within the Major Histocompatibility Complex (MHC) on the short arm of chromosome 6 (Figure 1.1).

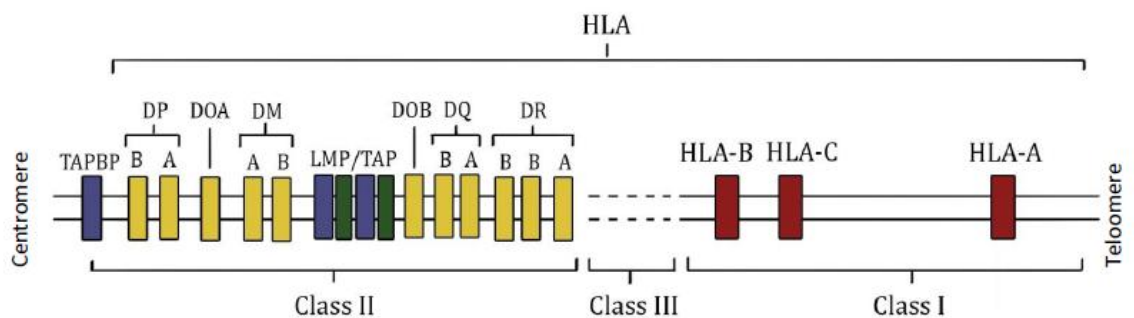


Figure 1.1 Schematic showing the HLA regions of the human MHC. Organisation of the classical HLA genes contained with the class I and II regions of the human MHC with three main class I genes called HLA-A, -B and -C shown in red. HLA class II genes HLA-DR, -DQ and -DP are shown in yellow and are clustered with genes encoding molecules involved with antigen processing and presentation. The so-called class III region encodes other proteins associated with the innate immune system. Copyright © 2008 From Janeway's immunobiology by Murphy *et al.* Adapted with permission of Garland Science/Taylor & Francis Group LLC.

The MHC was first discovered in mice as a tumour resistant locus over 60 years ago, with its equivalent in humans identified and named HLA following the description of alloantibodies

against antigens expressed by human leucocytes (Shiina *et al.*, 2017). The MHC has now been sequenced in 70 mammalian and non-mammalian species, with more than 7000 non-human alleles reported to the ImmunoPolymorphism Database (IPD)-MHC database (Maccari *et al.*, 2017). In comparison, over 16,000 HLA alleles have been recorded to date (Robinson *et al.*, 2015) with the majority reported in just the past eight years (Figure 1.2). Expression of HLA class II molecules is mainly restricted to, but not exclusively, specialized antigen presenting cells whereas the classical HLA class I molecules, namely HLA-A, -B & -C, are expressed by the majority of tissues and nucleated blood cells, (Brown and Navarrete, 2011). Non-nucleated platelets also express HLA class I molecules but not HLA class II, hence the focus on HLA class I for this study.

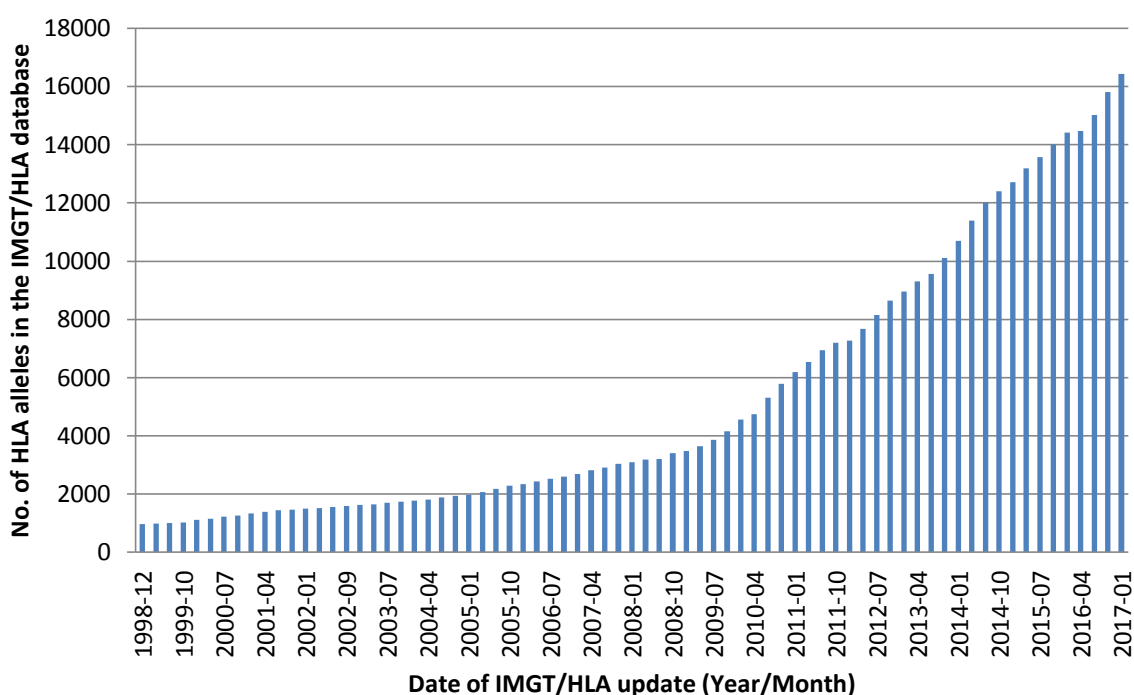


Figure 1.2 The number of HLA alleles recorded on the IMGT/HLA database. The total number of HLA alleles recorded in the IMMunoGeneTics(IMGT) HLA database following each update, from release v1 in December 1998 to v3.27 in January 2017 (Robinson *et al.*, 2015).

1.1.1 HLA class I gene structure

HLA class I genes code for the heavy chain of the HLA class I molecule, which is a heterodimer consisting of a heavy chain and light chain (beta-2 microglobulin). HLA-class I genes consist of seven or eight exons. Exon 1 encodes the leader sequence with exons 2 and 3 encoding the alpha-1 and alpha-2 domains, which form the peptide binding groove. Exon 5 encodes the transmembrane region with the cytoplasmic tail encoded by exons 6 and 7 (Figure 1.3).

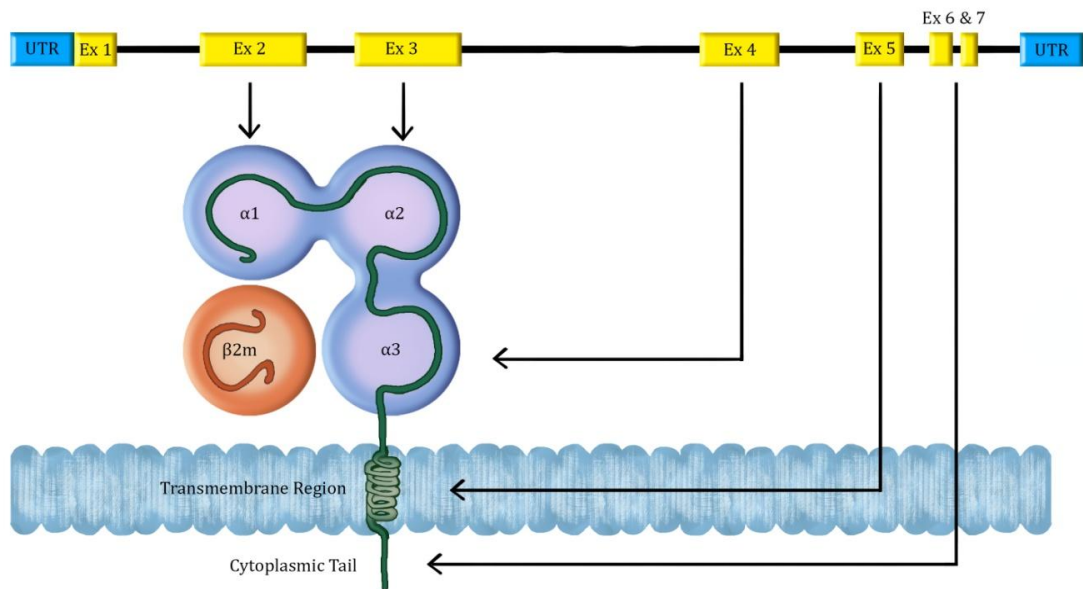


Figure 1.3 Cartoon showing a typical HLA class I gene and the structures each exon encodes. Exons 2 and 3 encode the alpha 1 and 2 domains, which form the peptide binding groove. The alpha 3 domain is encoded by exon 4, with the transmembrane region and cytoplasmic tail encoded by exons 5 and 6/7, respectively.

1.1.2 HLA class I nomenclature

HLA nomenclature follows a prescribed format, with the naming of HLA alleles the responsibility of the WHO Nomenclature Committee for Factors of the HLA System (Marsh *et al.*, 2010). All alleles begin with the HLA prefix, followed by the gene name. Thereafter, each allele is assigned at least a four digit number, with the first field indicating the allele group and second field the specific HLA protein (Figure 1.4).

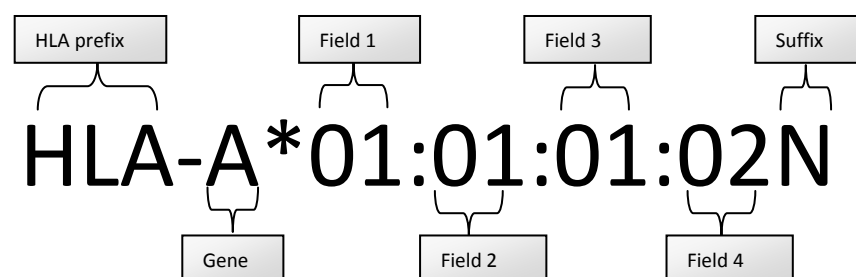


Figure 1.4 Nomenclature of HLA alleles. All alleles begin with the HLA prefix followed by the gene name. At least two fields follow the separator, the first denoting the HLA allele group and the second the specific HLA protein. Some alleles also have a third field which represents synonymous nucleotide substitutions in the coding regions and the fourth field is reserved for differences in non-coding regions. Any suffix present relates to changes in expression. Adapted with permission of AMERICAN SOCIETY OF HEMATOLOGY from Definitions of histocompatibility typing terms, Nunes *et al.*, 118, 23, 2011; permission conveyed through Copyright Clearance Centre, Inc.

The third field relates to synonymous DNA substitutions within the coding region and the fourth field is reserved for non-coding differences in the 5' or 3' un-translated region (UTR) or introns. Some HLA alleles also have a suffix which denotes changes in expression, the most common being 'N' representing a null allele (Marsh *et al.*, 2010). HLA can be determined at different levels of resolution depending on the method employed (Table 1.1). Low resolution DNA based typing is defined as first field or 2-digit resolution and generally equates to serological antigens. Medium or intermediate resolution is the reporting of a subset of HLA alleles in the form of an HLA string. HLA alleles that encode the same amino acid sequence are said to be at high resolution and can be reported as 'P-groups' or 'G-groups', where alleles included have identical amino acid or nucleic acid sequences, respectively, in exons 2 and 3. HLA defined at the allele level has a unique DNA sequence (Nunes *et al.*, 2011).

Resolution	Description	Example
Low	First field only	B*39
Medium	'String' of possible HLA alleles	B*39:01:07/39:06:01/39:06:02/39:34/39:42/39:62/39:64/39:73
High	'P group' – alleles with identical amino acid sequence in exons 2 & 3	B*39:02P (B*39:02:01/39:02:02:01/39:02:02:02/39:02:02:03)
High	'G group' – alleles with identical nucleotide sequence in exons 2 & 3	B*39:06:02G (B*39:06:02:01/39:06:02:02/39:06:02:03)
Allele	Unique nucleotide sequence	B*39:01:01:01

Table 1.1 Levels of resolution for HLA typing. HLA types can be reported at low, medium, high or allele level, depending on the method used to define and the resolution required. The alleles included in the example 'P' and 'G' groups are shown in parenthesis (Marsh *et al.*, 2010).

1.2 Human platelet antigens (HPA)

Platelets are anuclear cell fragments derived from large precursor cells called megakaryocytes in the bone marrow that express numerous cell surface receptors and adhesion molecules which assist interaction with white blood cells (leucocytes) and damaged endothelium (George, 2000). Some of the most important molecules expressed on platelets are integrins, formed by heterodimeric transmembrane polymorphic glycoproteins involved in cell signaling (Jenne *et al.*, 2013). These glycoproteins include GPIa/IIa and GPIIb/IIIa which bind collagen and fibrinogen respectively as well as the GPIb-V-IX complex that mediates binding to von Willebrand factor, a critical factor in the maintenance of haemostasis by promoting platelet aggregation (Jenne, Urrutia and Kubes, 2013). Defects in GPIIIa and its associated glycoprotein IIb can lead to the rare autosomal recessive bleeding disorder, Glanzmann thrombasthenia

(Buitrago *et al.*, 2015). The glycoproteins expressing HPA are encoded by six genes, detailed in Table 1.2.

Glycoprotein	Gene	Chromosome	No. of exons	Size of gene
GPIIIa	ITGB3	17	15	97,541bp
GPIIb	ITGA2B	17	30	24,324bp
GP1a	ITGA2	5	30	112,454bp
GP1bα	GP1BA	17	2	9734bp
GP1bβ	GP1BB	22	2	8232bp
CD109	CD109	6	33	132,533bp

Table 1.2 Genes that encode glycoproteins (GP) expressing human platelet antigens. Includes details of the respective chromosomes, number of exons present and total number of bases per gene.

According to the ImmunoPolymorphism Database (IPD) there are currently 29 HPA systems described (Robinson *et al.*, 2013). The majority of HPA are localised on the IIIa subunit of the highly abundant glycoprotein, GPIIb/IIIa (Curtis and McFarland, 2014), as represented in Figure 1.5. With the exception of HPA-14bw, which is defined by a 3 base pair (bp) deletion, each HPA system is characterised by a single nucleotide polymorphism (SNP) resulting in an amino acid substitution in the corresponding protein (Lucas, 2013), summarised in Table 1.3. HPA nomenclature is sequential, based on the order of discovery and the development of alloantibodies produced by exposed individuals lacking the respective antigen. The major allele of each system is designated 'a' and the minor, less frequent allele called 'b'; for example HPA-1a and HPA-1b with a frequency of 98% and 28% respectively in Caucasian populations (Curtis and McFarland, 2014). The 'w' assignment after the antigen name (e.g. HPA-14bw) indicates no reported alloantibody against the antithetical antigen (Metcalf *et al.*, 2003).

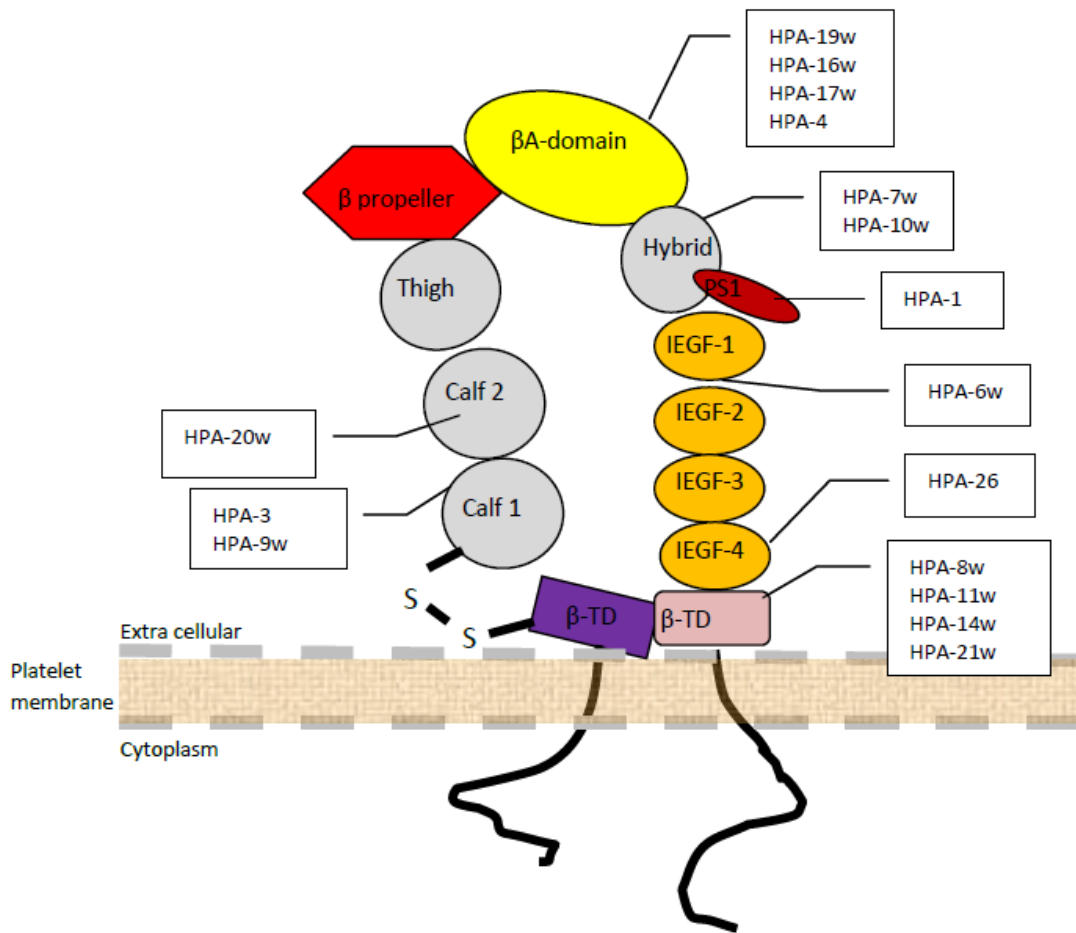


Figure 1.5. Cartoon representing the glycoprotein GPIIIa/GPIIb and respective HPA systems. The approximate location of 16 of the 29 human platelet antigen (HPA) systems expressed on glycoprotein GPIIIa/GPIIb. PSI=plexin/semaphoring/integrin; IEGF =Integrin epidermal growth factor domain; β -TD= β terminal domain. Adapted with permissions from *Transfusion and Transplantation Science* 2003, Page 192, Oxford University Press, and *Practical Transfusion Medicine Third Edition*, 2009, John Wiley and Sons, Inc.

System	Gene	Antigen	Amino acid	Nucleotide	rs number (dbSNP)	Exon	Chr position (GRCh37)
HPA-1	<i>ITGB3</i>	HPA-1a	Leu33	176T	rs5918	3	45360730
HPA-1	<i>ITGB3</i>	HPA-1b	Leu33>Pro	176T>C	rs5918	3	45360730
HPA-1	<i>ITGB3</i>	HPA-1c	Leu33>Val	175C>G	None	3	45360729
HPA-2	<i>GP1BA</i>	HPA-2a	Thr145	482C	rs6065	2	4836381
HPA-2	<i>GP1BA</i>	HPA-2b	Thr145>Met	482C>T	rs6065	2	4836381
HPA-3	<i>ITGA2B</i>	HPA-3a	Ile843	2621T	rs5911	26	42453065
HPA-3	<i>ITGA2B</i>	HPA-3b	Ser843	2621T>G	rs5911	26	42453065
HPA-4	<i>ITGB3</i>	HPA-4a	Arg143	506G	rs5917	4	45361953
HPA-4	<i>ITGB3</i>	HPA-4b	Arg143>Gln	506G>A	rs5917	4	45361953
HPA-5	<i>ITGA2</i>	HPA-5a	Glu505	1600G	rs1801106	13	52358757
HPA-5	<i>ITGA2</i>	HPA-5b	Glu505>Lys	1600G>A	rs1801106	13	52358757
HPA-6w	<i>ITGB3</i>	HPA-6bw	Arg489>Gln	1544G>A	rs13306487	10	45369788
HPA-7w	<i>ITGB3</i>	HPA-7bw	Pro407>Ala	1297C>G	rs121918448	10	45369541
HPA-8w	<i>ITGB3</i>	HPA-8bw	Arg636>Cys	1984C>T	rs151219882	12	45377914
HPA-9w	<i>ITGA2B</i>	HPA-9bw	Val837>Met	2602G>A	rs74988902	26	42453084
HPA-10w	<i>ITGB3</i>	HPA-10bw	Arg62>Gln	263G>A	rs200358667	3	45360817
HPA-11w	<i>ITGB3</i>	HPA-11bw	Arg633>His	1976G>A	rs377302275	12	45377906
HPA-12w	<i>GPIBB</i>	HPA-12bw	Gly15>Glu	119G>A	rs375285857	2	19711485
HPA-13w	<i>ITGA2</i>	HPA-13bw	Thr799>Met	2483C>T	rs79932422	20	52369001
HPA-14w	<i>ITGB3</i>	HPA-14bw	del_Lys	del1909-1911	n/a	11	45376893-5
HPA-15	<i>CD109</i>	HPA-15a	Ser682	2108C	rs10455097	19	74493432
HPA-15	<i>CD109</i>	HPA-15b	Tyr682	2108A	rs10455097	19	74493432
HPA-16w	<i>ITGB3</i>	HPA-16bw	Thr140>Ile	497C>T	rs74708909	4	45361944
HPA-17w	<i>ITGB3</i>	HPA-17bw	Thr195>Met	662C	rs770992614	5	45363673
HPA-18w	<i>ITGA2</i>	HPA-18bw	Gln716>His	2235G>T	rs267606593	17	52366090
HPA-19w	<i>ITGB3</i>	HPA-19bw	Lys137>Gln	487A>C	rs80115510	4	45361934
HPA-20w	<i>ITGA2B</i>	HPA-20bw	Thr619>Met	1949C>T	rs78299130	20	42455875
HPA-21w	<i>ITGB3</i>	HPA-21bw	Glu628>Lys	1960G>A	rs70940817	12	45377890
HPA-22w	<i>ITGA2B</i>	HPA-22bw	Lys164>Thr	584A>C	rs142811900	5	42462694
HPA-23w	<i>ITGB3</i>	HPA-23bw	Arg622>Trp	1942C>T	rs139166528	12	45377872
HPA-24w	<i>ITGA2B</i>	HPA-24bw	Ser472>Asn	1508G>A	rs281864910	15	42457790
HPA-25w	<i>ITGA2</i>	HPA-25bw	Thr1087>Met	3347C>T	rs771035051	28	52382870
HPA-26w	<i>ITGB3</i>	HPA-26bw	Lys580>Asn	1818G	None	11	45376801
HPA-27w	<i>ITGA2B</i>	HPA-27bw	Leu841>Met	2614C>A	rs149468422	26	42453072
HPA-28w	<i>ITGA2B</i>	HPA-28bw	Val740>Leu	2311G>T	rs368953599	23	42453713
HPA-29w	<i>ITGB3</i>	HPA-29bw	Thr33>Met	98C>T	None for CRCh37	2	45351803

Table 1.3 The human platelet antigen (HPA) system: Details of the genes encoding each HPA antigen alongside the respective amino acid and nucleotide variants. Also listed is the location of each HPA polymorphic position, including the respective exon, chromosomal position (Chr position) and rs number (where available from dbSNP), based on data from the human genome assembly GRCh37.

1.3 Clinical relevance of HLA and HPA in transfusion

Individuals can become alloimmunised to HLA and or HPA following exposure during transfusion, pregnancy or transplantation (Brown and Navarrete, 2011), leading to the production of HLA and or HPA antibodies. HLA antibodies in the transfused blood product can result in the rare but potentially fatal transfusion associated acute lung injury (Brown and Navarrete, 2011). Post transfusion purpura is a rare transfusion-related complication due to alloimmunisation of the patient against platelet antigens leading to acute thrombocytopenia. In 2016, there were no reported cases of either transfusion associated acute lung injury or post transfusion purpura in the UK (Bolton-Maggs, 2017). More commonly, HLA or HPA alloantibodies present in the patient lead to complications such as immunological platelet refractoriness and foetal or neonatal alloimmune thrombocytopenia (Lucas, 2013). Binding of HLA or HPA alloantibodies to platelets can lead to a decrease of platelet function and survival (Pavenski *et al.*, 2012).

1.3.1 Immune platelet refractoriness (IPR)

Platelet refractoriness is a failure to achieve a satisfactory response following random platelet transfusions (Stanworth *et al.*, 2015) and is usually defined as an incremental rise in platelet count of $10 \times 10^9/l$ or less, 1 hour or up to 24 hours post transfusion (Brown and Navarrete, 2011). This can lead to a number of adverse outcomes, including increased risk of bleeding, decreased patient survival as well as longer hospital stays and associated in-patient costs (Stanworth *et al.*, 2015). Platelet refractoriness can result from non-immune causes such as an enlarged spleen, bleeding or infection leading to decreased platelet survival, which need to be excluded prior to further investigation (Slichter *et al.*, 2005).

Immune platelet refractoriness (IPR) usually results from the presence of donor-specific HLA antibodies and can be successfully treated by the provision of HLA selected platelets (Brown and Navarrete, 2011). A small percentage of patients fail to respond to HLA selected products and will be subsequently examined for the presence of antibodies to HPA, which usually occur in conjunction with HLA antibodies although they can occur independently (Stanworth *et al.*, 2015). If necessary, patients can be provided with HLA and/or HPA compatible platelet support (Lucas, 2013), as summarised in Figure 1.6. In the financial year 2016/17, over 25,000 HLA and or HPA selected platelet units were issued for patients in England.

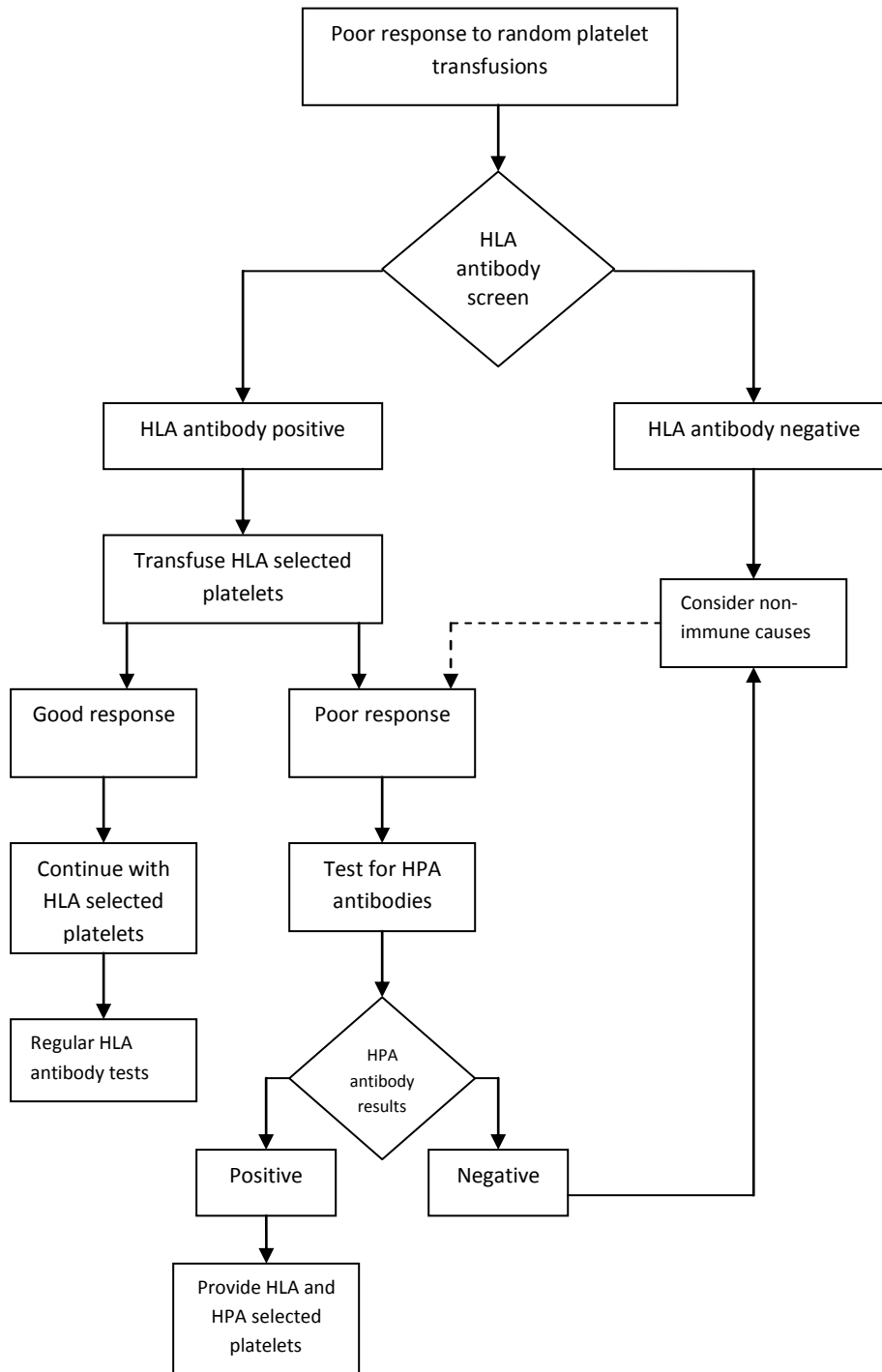


Figure 1.6 Summary of the laboratory investigation for platelet refractoriness. Patients failing to respond to random platelet transfusions are referred for HLA antibody screening. Where HLA antibodies are detected, HLA selected platelets are provided. If HLA antibodies are absent, or the patient is failing to respond to HLA selected platelets, the patient will be screened for antibodies to HPA and provided with HLA and HPA selected platelets as appropriate. Adapted with permission from Brown and Navarrete, *Clinical relevance of the HLA system in blood transfusion*, John Wiley and Sons. © 2011 International Society of Blood Transfusion.

1.3.2 Foetal or neonatal alloimmune thrombocytopenia

Foetal or neonatal alloimmune thrombocytopenia (FNAIT) is a severe bleeding disorder of the foetus and neonate resulting from the destruction of platelets by maternal alloantibodies directed against paternally inherited antigens expressed on foetal platelets (Bertrand and Kaplan, 2014). Alloantibodies involved in FNAIT are predominantly against HPA (Lucas, 2013), although other antigens such as GPIV have been implicated (Xia *et al.*, 2014). FNAIT occurs in approximately 1 in 1000 live births (Lucas, 2013) and can result in intracranial haemorrhage leading to death or disability, with severity linked to the number of pregnancies (Delbos *et al.*, 2016) and history of earlier siblings (Wienzek-Lischka *et al.*, 2015; Wienzek-Lischka *et al.*, 2017). Diagnosis of FNAIT is vital to manage both the index case and any subsequent pregnancies and usually follows presentation in a neonate less than 48 hours old with unexplained bruising and petechiae (red or purple spots on the skin) but who otherwise appears healthy (Bertrand and Kaplan, 2014).

Maternal antibodies to HPA-1a are the most common cause of FNAIT, with HPA-1a immunisation occurring in 1 in 365 pregnancies (Ghevaert *et al.*, 2007). HPA antibody screening was performed on over 500 maternal samples for suspected FNAIT in the financial year 2016/17 in England. Overall, greater than 95% of serologically confirmed FNAIT cases result from maternal-foetal incompatibility for HPA-1, -2, -3, -5 and -15 with the remainder caused by low frequency HPA (Pavenski *et al.*, 2013). Transfusion of the affected foetus or neonate with platelets that do not express the antigen implicated (see section 1.3.3.) results in prolonged platelet survival compared to transfusion with random platelets (Lucas, 2013).

1.3.3 Transfusion support for patients with HLA and or HPA alloantibodies

Patients refractory to platelets and those affected by FNAIT require HLA and/or HPA matched platelets in order to obtain a satisfactory incremental rise in their platelet count (Pavenski *et al.*, 2012). Numerous strategies are available for the selection of platelets to support these patients including the use of cross-match negative donors, of antigen avoidance and of HLA and or HPA matched products (Kopko *et al.*, 2015), summarised in Table 1.4.

Method of platelet selection	Advantages	Disadvantages
Cross-match negative	<ul style="list-style-type: none"> • Rapid availability • Does not require a pool of HLA & HPA typed donors 	<ul style="list-style-type: none"> • Risk of further sensitisation • Not suitable for long term platelet support • Challenging for highly sensitised patients
Antigen avoidance	<ul style="list-style-type: none"> • Increases number of potential donors 	<ul style="list-style-type: none"> • Risk of further sensitisation • Challenging for highly sensitised patients
HLA/HPA matching	<ul style="list-style-type: none"> • Suitable for long term platelet support • Reduced risk of sensitisation 	<ul style="list-style-type: none"> • Time consuming • Patient must be HLA & HPA typed • Requires large panel of HLA & HPA typed donors
HLA epitope matching	<ul style="list-style-type: none"> • Increases number of potential donors 	<ul style="list-style-type: none"> • Requires high resolution HLA typing of patient and donor

Table 1.4 Summary of options for selecting compatible platelets. Advantages and disadvantages of the four approaches used for the selection of platelets for patients with immune platelet refractoriness resulting from antibodies to HLA

1.3.3.1 Cross-match negative

Cross-matching is performed by incubating donor platelets with recipient serum followed by detection of any bound antibody using labelled antiglobulins (Vassallo *et al.*, 2014). Cross-matched platelets have the advantage of being readily available (Rioux-Masse *et al.*, 2014) and are cheaper and easier to obtain than having a pool of HLA typed donors (Lee and Ayob, 2015). However, there are concerns that not matching for HLA risks further alloimmunisation as platelets are likely to be incompatible with the recipient's own HLA type (Vassallo *et al.*, 2014) and is therefore not suitable for patients requiring long term platelet support (Brown and Navarrete, 2011). Sixty percent of US blood centres surveyed provide a cross-matched platelet product, although many reported problems with availability of cross-matching assays (Kopko *et al.*, 2015).

1.3.3.2 Antigen avoidance

Selecting platelets on the basis of avoiding specificities for which the recipient has antibodies has been shown to be as effective as HLA matching (Petz *et al.*, 2000). This approach has the advantage of increasing the number of donors available for patients compared to HLA matching, although it may be problematic when supporting highly sensitised patients (Kopko

et al., 2015). As with crossmatching, there are risks associated with broadening a recipient's antibody profile, as platelet transfusions may not be well matched (Petz *et al.*, 2000).

1.3.3.3 HLA matching

Generally HLA selection is performed for HLA-A and –B only as the role of HLA-C in IPR remains uncertain (Brown and Navarrete, 2011). HLA antigen matched platelets are provided according to two match grades: an A grade when donor and recipient are HLA compatible or a B grade where one or more mismatched HLA antigens are present, denoted by B1, B2, B3 and B4 for one, two, three or four mismatches, respectively (Brown and Navarrete, 2011). Searching for the best donor using this strategy can be a protracted process, particularly for blood centres without large donor panels (Kopko *et al.*, 2015) and also increases the cost of managing patients (Lee and Ayob, 2015). National Health Service Blood and Transplant (NHSBT) has a panel of over 11,000 donors who regularly give their own platelets via an automated cell separating process known as apheresis, to facilitate HLA matching for patients with IPR (Brown and Navarrete, 2011).

1.3.3.4 HLA epitope matching

HLA epitope-based matching (HEM) is a more recent approach based on compatibility of antigenic determinants (Duquesnoy, 2011). HLA epitope matching employs a predictive computer algorithm, such as 'HLAMatchmaker', for donor and recipient compatibility, quantifying mismatched antigenic determinants, so called 'eplets' (Duquesnoy, 2008) or antibody-verified epitopes recorded in the International HLA Epitope Registry website (Duquesnoy *et al.*, 2014). Eplets are described as polymorphic linear or discontinuous amino acid residues at critical antigen/antibody binding sites of the HLA molecule. The HEM algorithm is based on the assumption that patients will not make antibody against eplets or epitopes shared between the donor and their own HLA type or to inaccessible parts of the HLA molecule (Duquesnoy, 2008). Previous studies have shown that the number of eplet mismatches is inversely proportional to the increase in platelet count following platelet transfusion (Brooks, MacPherson and Fung, 2008). The sharing of epitopes between different alleles enables permissive mismatching and should only require donor panels of 'reasonable size' compared to using fully HLA matched platelets (Rubinstein, 2010).

1.4 HLA and HPA genotyping – current state

Patients refractory to platelets and those affected by FNAIT require HLA and/or HPA matched platelets in order to obtain a satisfactory incremental rise in their platelet count (Pavenski *et al.*, 2012). Hence techniques that can accurately and cost effectively define the HLA and HPA type of both patients and donors are required. As platelets only express HLA class I antigens,

genotyping of platelet donors and recipients is limited to HLA-A and HLA-B in most laboratories, due to uncertainty about the clinical relevance of HLA-C in IPR (Brown and Navarrete, 2011). Molecular techniques for HLA typing are well established, having been in routine use for over 20 years (Erlich, 2012). Due to lack of highly specific antisera free from HLA antibodies and the advent of the polymerase chain reaction (PCR), serological HPA phenotyping has also been largely superseded by DNA-based genotyping (Curtis and McFarland, 2009).

NHSBT laboratories receive both patient and donor samples from hospitals and blood collection centres. Different approaches are taken for HLA and HPA genotyping dependent on the number of samples, resolution required and clinical urgency, summarised in Table 1.5.

Method	Genotyping application	Sample type	HLA Resolution	Turnaround time (hours)	Cost per test	Sample throughput
PCR-SSP	HLA HPA	Patient	Low	2-3	££	Low
PCR-SSOP	HLA	Donor & Patient	Low/Medium	6-8	£	Low or high
PCR-SBT	HLA HPA	Patient	High	24-72	£££	Low
Taqman	HPA	Donor	n/a	48	£	High

Table 1.5 Molecular techniques used for HLA and HPA genotyping. The genetic system and sample type tested by each Polymerase chain reaction (PCR) based method, alongside the typical time taken to obtain a result. Cost per test is indicated using a comparative approach, with £ = inexpensive, ££ = medium cost and £££ = most expensive. The throughput of samples for each approach is designated high or low, with high throughput determined as processing in batches of up to 96 samples and low throughput as processing single or small numbers of samples. PCR-SSP = PCR using sequence specific primers; PCR-SSOP = PCR using sequence specific oligonucleotide probes; PCR-SBT = PCR by sequencing based typing

1.4.1 Patient genotyping

Patients are referred to NHSBT for the investigation of IPR and FNAIT which can be urgent in nature, for example when the patient is bleeding (West, 2016). In addition, samples from individuals requiring long-term platelet support are sent for HLA and HPA genotyping to ensure suitably matched products are provided, thereby avoiding the risk of sensitisation (Murphy *et al.*, 1986). In England, approximately 900 patients are HLA typed each year during investigation for platelet refractoriness.

1.4.1.1 Patient HLA typing

Patients with IPR are HLA typed by PCR based methods utilising sequence specific primers (PCR-SSP) or sequence specific oligonucleotide probes (PCR-SSOP) (Brown and Navarrete, 2011). PCR-SSP uses gel electrophoresis to determine the presence or absence of amplicons for each HLA allele group tested. In contrast, PCR-SSOP requires the hybridisation of HLA locus specific amplicons with sequence specific oligonucleotide probes attached to a solid support followed by fluorescence based detection to determine the HLA type. These methods have a quick turnaround time and are suitable for small sample numbers. They produce low to medium resolution HLA types by focusing on polymorphisms within exons 2 and 3 of the HLA class I genes, where the majority of allelic diversity is located (Erlich, 2012).

1.4.1.2 Patient HPA typing

HPA typing, alongside screening for HPA antibodies, is performed only if IPR patients subsequently fail to increment with HLA matched platelets; it is uncommon for HPA antibodies to cause IPR independently (Brown and Navarrete, 2011). PCR-SSP is a commonly used method for HPA genotyping these patients, being both fast and reliable (Nogués, 2011).

Suspected FNAIT cases are investigated by determining maternal and paternal disparity for the more frequent HPA genotypes, also using methods such as PCR-SSP (Arinsburg *et al.*, 2012), although at NHSBT PCR sequencing based typing (PCR-SBT) of HPA-1, -2, -3, -4, -5, -6, -9 and -15 is routinely employed for FNAIT investigations (West, 2016). Further analysis is required should no difference be detected for the common variants. This may involve extensive study, sequencing entire genes which can be a protracted process when searching the large HPA genes (Poles *et al.*, 2013; Stafford *et al.*, 2008). For example, the gene encoding GPIIIa, on which the majority of HPA are expressed, consists of 15 exons and is almost 100,000 bases long (see Table 1.2).

1.4.2 Donor genotyping

Providing suitable platelet transfusions for patients adversely affected by the presence of alloantibodies requires a large panel of HLA- and HPA-defined platelet donors (Pavenski *et al.*, 2012). Approximately 5000 apheresis donors per annum are currently recruited by NHSBT for this purpose, with each donor typed for both HLA and HPA on two separate occasions for quality reasons. This level of commitment to donor enrolment requires high-throughput genotyping techniques (Lucas, 2013).

1.4.2.1 Donor HLA typing

HLA class I typing of platelet donors is currently performed using reverse PCR-SSOP in a 96 well plate format, making it suitable for automation for high throughput typing. In separate PCR

reactions, amplification of exons 2 and 3 of the HLA-A, -B and -C genes is followed by hybridisation with locus specific, fluorescently tagged microspheres that are coated with up to 100 probes, targeting sequence motifs across the respective exons. Where complementary sequences are present, bound PCR amplicons are detected by utilising xMAP[®] technology followed by analysis of probe patterns using interpretation software (Dalva and Beksac, 2007). PCR-SSOP typically produces a low to medium resolution HLA type (Nunes *et al.*, 2011) which is sufficient for the selection of platelets for patients with IPR, based on their HLA type and antibody profile (Brown and Navarrete, 2011).

1.4.2.2 Donor HPA typing

To date, only genotypes of the HPA-1, -2, -3, -4, -5 and -15 systems are determined for apheresis platelet donors, selected originally due to their frequency and clinical relevance (Lucas, 2013). The Taqman™ assay is utilised for this high-throughput HPA typing due to its capacity for automated readout (Lucas, 2013). Briefly, for each HPA tested, PCR amplification using primers flanking the SNP of interest is performed in the presence of two labeled probes, one corresponding with the 'a' allele and the other specific for the 'b' allele. Displacement of an annealed probe by Taq polymerase during primer extension generates a fluorescent signal that can be measured. Using two different dyes for the two allele specific probes enables distinction between the alleles by their emission spectra (Nunes *et al.*, 2011).

1.5 Requirement for the improvement of HLA and HPA genotyping

1.5.1 Increasing the resolution of HLA genotyping

A prospective, randomised non-inferiority trial is currently underway within NHSBT to compare the efficacy of HLA epitope-matched (HEM) platelets for transfusion with standard HLA selected platelets for alloimmunised patients who require regular platelet transfusions but for whom a fully HLA matched donor is not available (ISCTRN23996532). The necessity for utilising HEM may increase as a result in a reduction in the availability of single platelet units. NHSBT recently announced a 'Platelet Supply' project, initiated following recommendations by The Advisory Committee on the Safety of Blood, Tissues and Organs to remove the current requirement of collecting 80% of platelets by apheresis. As part of a drive for efficiency savings in the blood supply chain, the plan was to reduce apheresis to 60% of existing collection levels by the end of 2015/16, with the further potential to reduce to as low as 40% thereafter (Ronaldson and Ashford, 2014). Whilst this strategy will align the percentage of platelet concentrates collected by apheresis in England with other European countries (Berger *et al.*, 2016), it will limit available units for patients with HLA and or HPA antibodies. Using HEM

should increase the number of platelet donors suitable for these patients when fully HLA matched units are unavailable (Pai *et al.*, 2010).

If HEM becomes routine practice, high resolution (i.e. allele level) HLA typing of platelet donors may be necessary because the matching algorithm requires information on amino acid sequence (Duquesnoy, 2008). However, implementation of this new approach will require a change to the current method used, described in 1.4.2.1, due to inadequate resolution of HLA class I typing. Pai *et al.*, 2010 reported using HEM for the selection of platelet donors based on epitopes predicted from low resolution HLA types produced by PCR-SSOP (Pai *et al.*, 2010). However, accurate comparison of donor and patient intra- and inter-locus HLA is likely to require high resolution or allele level typing (Duquesnoy, 2011). Conventionally, high resolution HLA typing is achieved using Sanger sequencing but due to the increasing number of characterised HLA alleles this method often produces ambiguous results due to limited sequence data generated for regions of the HLA genes and lack of phase within and between exons for heterozygous samples (Erich, 2015). In addition, costs and logistical constraints make sequencing based typing (SBT) unsuitable for high throughput HLA genotyping.

1.5.2 Expansion of HPA genotyping

There is increasing evidence that HPA genotypes not routinely defined in patients or apheresis platelet donors are more clinically significant than originally thought, with a growing number of publications reporting maternal alloimmunisation against 'rare' platelet antigens (Peterson *et al.*, 2012; Jallu *et al.*, 2013; Bertrand *et al.*, 2013a; Lucas *et al.*, 2016; Sullivan *et al.*, 2015). In one study, 6.6% of fathers of infants from 244 suspected cases of FNAIT were found to have low frequency HPA following sequencing of genes encoding glycoproteins GPIIIa/IIb (Peterson *et al.*, 2014). This percentage of FNAIT due to rare HPA antibodies was considerably greater than the 0.7% identified in 1054 cases, reported several years earlier (Ghevaert *et al.*, 2009). It has been suggested in the past (Ouweland, 2005) and again more recently (Santoso and Tsuno, 2015) that laboratory investigations should not be limited to the more common HPA systems.

Extending the existing PCR-SSP and Taqman™ techniques to include all 29 HPA systems would prove problematic as some HPA SNPs are adjacent to other polymorphic positions which can result in mistyping (Kengkate *et al.*, 2015). Therefore, expansion of the current HPA repertoire will require an alternative approach, capable of defining all known HPA SNPs alongside scalability for high-throughput.

1.5.3 HPA whole gene sequencing

Occasionally, an investigation for a rare or possibly unique HPA allele is required in cases of suspected FNAIT, where routine HPA typing and antibody screening proves uninformative (Lucas, 2013). As indicated in section 1.4.1, this can necessitate the sequencing of one or more HPA genes, which is complex and time consuming due to their length. Investigation of suspected FNAIT cases should be performed with urgency (Lucas, 2013), so reducing the time taken to search for a candidate mutation would be advantageous. It is possible to develop Sanger sequencing to determine the majority of, if not all, known HPA alleles as a first line test (Hong *et al.*, 2017), but this requires significant development and validation of multiple primer sets due to the location of each HPA system (Lane *et al.*, 2016). In addition, without sequencing each entire gene, Sanger sequencing may also fail to identify and or define novel HPA mutations without further development (Finning *et al.*, 2016).

1.6 Next Generation Sequencing

Next Generation Sequencing (NGS) is a recent innovation in sequencing technology with the ability to produce large amounts of data relatively quickly and cheaply, capable of generating a high proportion of an individual's genetic sequence in a single experiment (Metzker, 2010). When combined with novel DNA bar-coding technology, which consists of sample specific combinations of short oligonucleotides, NGS can be used to target specific regions of the genome from many different individual samples in one pool (Shiina *et al.*, 2012).

The term 'next generation sequencing' refers to a range of sequencing technologies developed during the past decade rather than a single technique (Muzzey, Evans and Lieber, 2015) and typically consists of (i) template preparation (ii) sequencing and imaging followed by (iii) data analysis, with each NGS technology taking a distinct approach (Metzker, 2010). DNA sequencing by NGS requires prepared templates of high molecular weight double stranded DNA that have been fragmented using methods such as sonication, nebulisation or enzymatic fragmentation (Knierim *et al.*, 2011). Unique sample identifiers, so called DNA barcodes, along with platform-specific adapters are then ligated to the fragments generated, which facilitates immobilisation of each fragment to a solid surface or support and downstream sequencing (Figure 1.7).

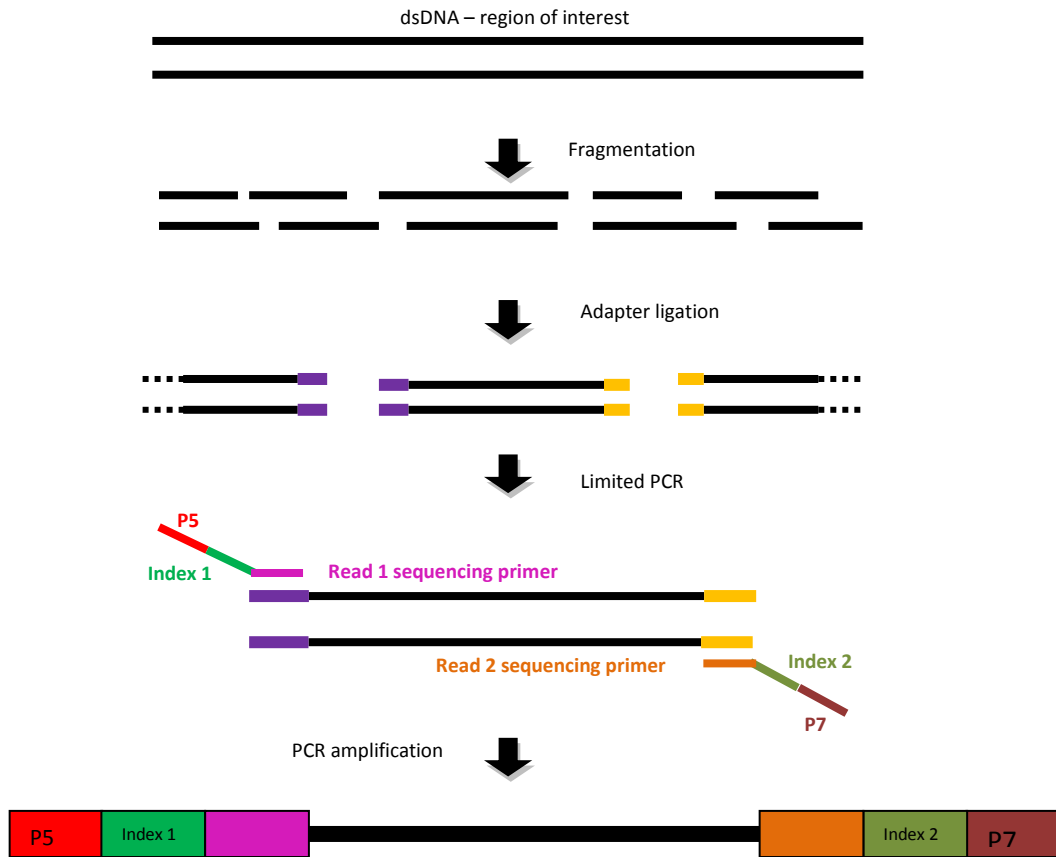


Figure 1.7 DNA library preparation for the Illumina MiSeq. Fragmentation of double stranded DNA is followed by the ligation of partial adapters that serve as primer binding sites. A limited cycle PCR then adds sample specific indices incorporating P5 and P7 ends to each fragment, along with the MiSeq sequencing primer sequences, to allow clonal amplification and paired end sequencing on a MiSeq flow cell.

The majority of NGS approaches then require in-situ clonal amplification of each prepared DNA template in order to provide signal enhancement prior to sequencing (Metzker, 2010). For the Illumina MiSeq, clonal amplification of each bound fragment occurs on a glass slide which allows positional separation of the clusters that form (Figure 1.8) that remain constant during the sequencing reaction (Muzzey, Evans and Lieber, 2015).

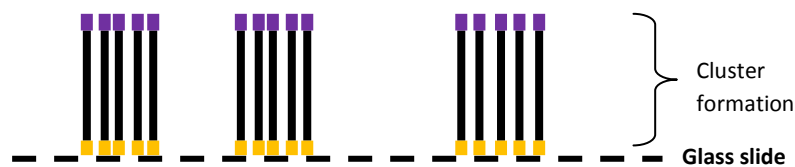


Figure 1.8 Cluster formation on a glass slide following clonal amplification. Each fragment bound to the MiSeq flow cell forms discrete clusters in fixed positions following clonal amplification. This facilitates signal amplification during the subsequent sequencing reaction.

Sequencing of the amplified templates consists of sequencing by synthesis (akin to Sanger sequencing), sequencing by ligation or using chemistries such as ion-semiconductor sequencing, depending on the platform employed (Muzzey, Evans and Lieber, 2015). This results in the production of multiple overlapping short sequences that can be subsequently assembled and analysed (Zhang *et al.*, 2011), an example of which is shown in Figure 1.9, hence the alternative title 'massively parallel sequencing' sometimes cited (Yang *et al.*, 2017). Data analysis generally consists of either variant calling following the alignment of sequencing reads to a reference genome and/or utilising denovo genome assembly algorithms (Sims *et al.*, 2014).

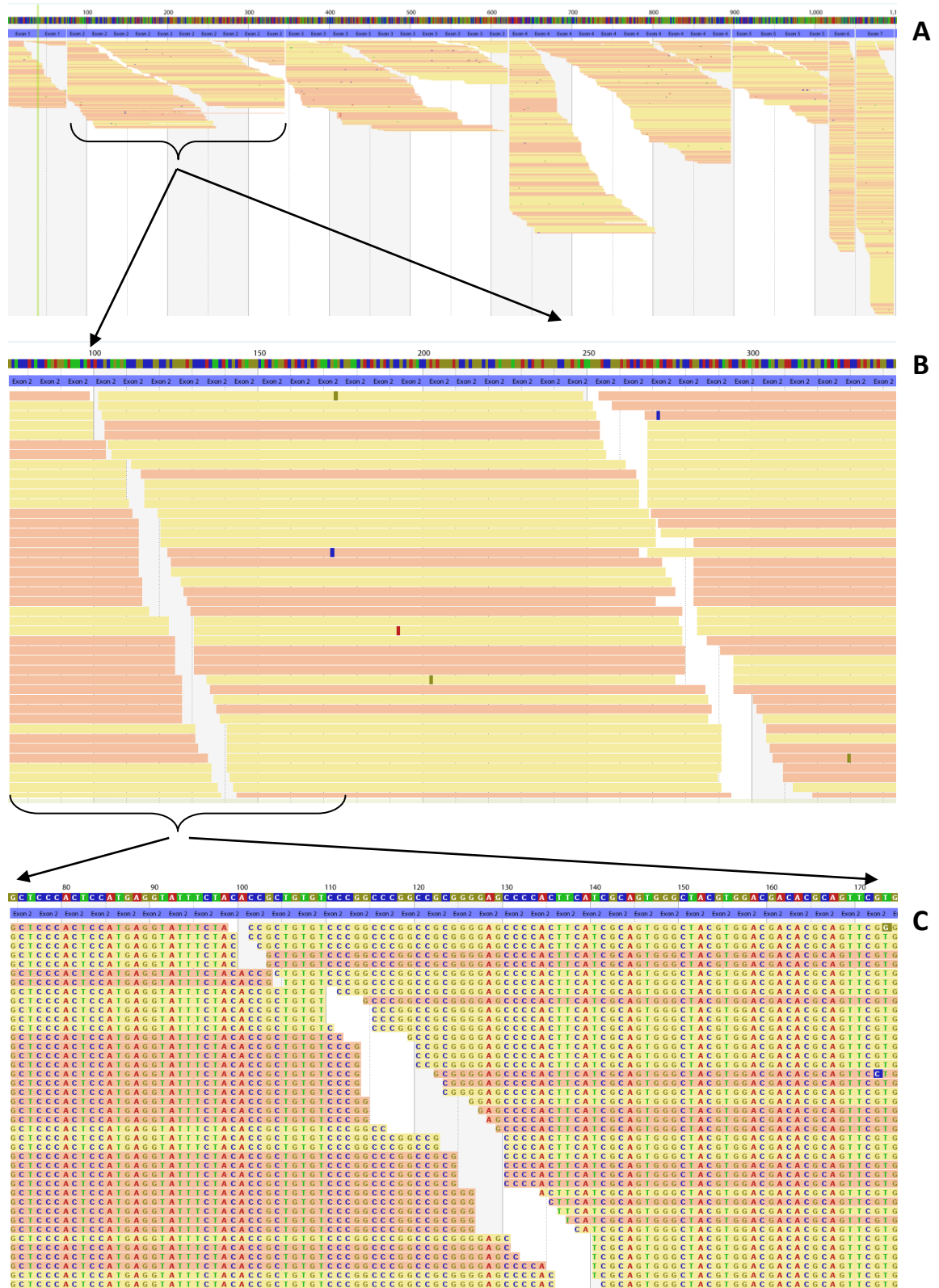


Figure 1.9 Alignment and analysis of sequencing reads. (A) shows the overlapping assembly of sequencing reads spanning an entire HLA gene; (B) shows more detail of the overlapping assembly for exon 2; (C) shows the individual base calls for each overlapping sequence. The pink and yellow colours of each fragment in the images represent fragments from two different sequences in a heterozygous sample.

1.6.1 HLA genotyping by NGS

Over the past five years, a number of approaches to HLA typing by NGS have been reported, ranging from the imputation of HLA types from genomic sequence data to targeted HLA sequencing using bench top NGS platforms, each with their respective advantages and disadvantages summarised in Table 1.6.

NGS approach	Advantages	Disadvantages
WGS/WES	<ul style="list-style-type: none"> • Provides MHC haplotype data • Use of generic WGS/WES protocol 	<ul style="list-style-type: none"> • Requires complex imputation algorithms • Slow analysis • Up to 20% error rate • Not suitable for high throughput
Targeted exon	<ul style="list-style-type: none"> • Fragmentation not required • Fast preparation protocol • Suitable for high throughput • Cost effective 	<ul style="list-style-type: none"> • Results in ambiguous HLA types • Susceptible to PCR artefacts • Risk of allele imbalance or dropout
Whole gene	<ul style="list-style-type: none"> • Unambiguous 3rd or 4th field level typing • Suitable for high throughput 	<ul style="list-style-type: none"> • Protracted library preparation • Risk of allele imbalance or dropout

Table 1.6 Advantages and disadvantages of NGS approaches reported for HLA typing. Reported NGS based approaches for HLA typing have included whole genome sequencing (WGS), whole exome sequencing (WES), targeted HLA exon sequencing or whole HLA gene sequencing.

1.6.1.1 HLA imputation from whole genome or exome sequence data

Large scale whole genome sequencing (WGS) or whole exome sequencing (WES) endeavours are not primarily intended for individual genotyping but rather for population genetic studies (Major *et al.*, 2013). However, is it possible for classical HLA types to be imputed from such data by using statistical prediction algorithms (Bauer *et al.*, 2016). Earlier reports indicated unacceptably high prediction errors of up to 20% due to insufficient read depth (Major *et al.*, 2013) or lack of reference data from ethnically diverse populations (Dilthey *et al.*, 2013). Some mistyping also resulted from the homology of HLA-A, -B and -C genes with pseudogenes, leading to mis-alignment of sequence fragments (Major *et al.*, 2013). More recently, Dilthey *et al.* described use of 'HLA*PRG', a computer algorithm that infers HLA types from WGS data. This program considers both genes and pseudogenes and has a reported accuracy of 99.4%. However, the requirement of high computational power taking 30-250 central processing unit hours per sample (Dilthey *et al.*, 2016) is not acceptable for routine clinical application.

1.6.1.2 Targeted HLA sequencing

For the majority of HLA typing laboratories, facilities for WGS or WES are unavailable. A more practical approach to HLA typing by NGS is the use of targeted sequencing. This requires the isolation of all or part of the HLA gene, typically by PCR amplification, although targeted enrichment capture assays have also been reported (Wittig *et al.*, 2015). Numerous approaches have been described for PCR-based HLA typing by NGS, from targeting one or more exon to sequencing the entire gene (Figure 1.10)

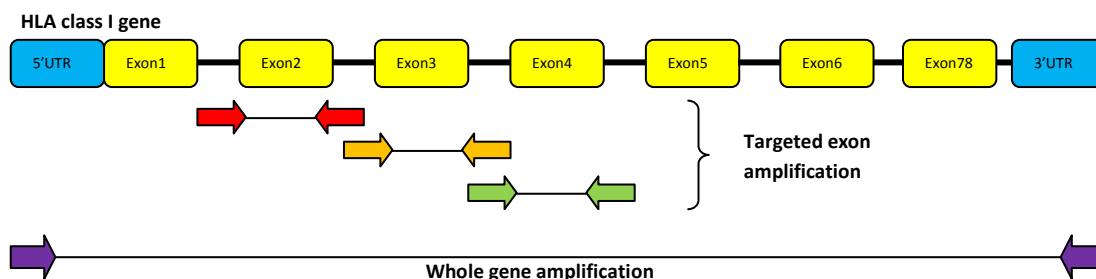


Figure 1.10 A typical primer map for the amplification of HLA class I genes. PCR primers can be designed to amplify exons individually, as represented by the red, orange and green arrows, or can target the entire HLA gene, indicated with the purple arrows.

One of the benefits of using exon specific amplification is that template fragmentation and or ligation steps are not usually required due to the size of the amplicon and inclusion of multiplex identifiers during amplification (De Santis *et al.*, 2013). However, the amplification strategy employed for exon only sequencing will determine the level of HLA typing resolution achieved. Lange *et al.* reported a cost effective, high throughput exon-based NGS method using micro fluidics for sample preparation. However, because only exons 2 and 3 of the HLA genes were sequenced, the results obtained were at a resolution equating to that obtained by Sanger sequencing (Lange *et al.*, 2014). Similarly, sequencing just exons 2, 3 and 4 of the HLA class I genes produces ambiguous results in over 85% of samples (Trachtenberg and Holcomb, 2013). It is possible to increase the HLA typing resolution to the 'G' group level by targeting more exons (Cereb *et al.*, 2015), but primer design will be limited by polymorphisms within introns of the HLA genes (De Santis *et al.*, 2013). Exon specific amplicon approaches to HLA typing by NGS are also susceptible to sequencing artefacts such as primer dimers and PCR mediated recombination (Holcomb *et al.*, 2014; Schofl *et al.*, 2017).

Restricting the sequencing of HLA genes to exon regions will inevitably result in ambiguity due to lack of phase. However, using whole HLA gene amplification allows phasing across the entire region and therefore reduces the likelihood of any HLA ambiguity. Primers are located in

less polymorphic positions, with exons and introns amplified in a single fragment (Wang *et al.*, 2012). Early reports of HLA typing by NGS based on whole gene sequencing suggested this approach may have a limited throughput, with only a few samples included in a sequencing run (Shiina *et al.*, 2012; Hosomichi *et al.*, 2013). This was in part due to the platform employed but also because each amplicon had separate DNA barcode identifiers. Although this made data analysis more straightforward by reducing the risk of misalignment of fragments to homologous regions of other class I genes, it limited the number of samples that could be sequenced simultaneously. More recently others have used identical DNA barcodes to label multiple amplicons from the same sample, enabling the pooling of amplicons and demonstrating it is possible to increase the throughput to 96 samples or more (Ehrenberg *et al.*, 2014; Nelson *et al.*, 2015).

1.6.1.3 Sequencing platforms

HLA typing results obtained by NGS are also influenced by the platform employed. One of the first NGS-based methods for HLA typing was developed on the Roche 454 FLX (Lind *et al.*, 2010). This platform had the advantage of generating long sequencing reads, which is preferable for HLA typing to ensure sequence reads span each exon, and was the first commercial NGS system for HLA typing (Gabriel *et al.*, 2014). However, the Roche 454 system produced far fewer reads per run than its competitors, making it much more expensive, and has been phased out since 2016 (Erlich, 2015). The Ion Torrent (Thermo Fisher Scientific) produces shorter reads to a maximum of 400bp (Erlich, 2015) and also has reported sequencing errors from insertions or deletions (indels), particularly at homopolymer sites. However, the Ion Torrent has the advantage of a short sequencing run time of approximately 2 hours and has potential for scalability, with a range of different chips sizes, making it a more economical option for some laboratories (Gabriel *et al.*, 2014). The Illumina MiSeq has the advantage of using pair end sequencing, which facilitates interpretation of HLA types due to improved haplotype phasing (Carapito, Radosavljevic and Bahram, 2016). It is also the platform of choice for the majority of commercial HLA typing kits currently available.

Instruments such as the PacBio RS11 (Pacific Biosciences) and the MinION (Oxford Nanopore) are more recent innovations. Both have the capability of producing reads of up to 40kb and 50kb, respectively, potentially allowing sequencing of entire MHC haplotypes in phase. However, errors rates are reportedly much higher than the more established platforms (Carapito, Radosavljevic and Bahram, 2016). Approaches to HLA typing using the PacBio system have been published by a number of groups (Cereb *et al.*, 2015; Mayor *et al.*, 2015), but due to the large size of the instrument and high procurement costs this platform is unlikely to be implemented by many routine laboratories. In contrast, the MinION (Oxford Nanopore) is a

small, portable device that uses very different technology to other NGS sequencers and, whilst still in the development phase for HLA genotyping, has the potential to provide sequencing data cheaply and quickly (Carapito, Radosavljevic and Bahram, 2016).

1.6.2 HPA genotyping by NGS

With its ability to scrutinise large portions of the genome, NGS has the capacity to simultaneously sequence the six genes encoding HPA glycoproteins within a few days, rather than the weeks that may be required for investigation by conventional methods. Lane *et al.* described the prediction of red blood cell and platelet antigens from an individual's genome obtained by using paired-end WGS on the Illumina HiSeq platform. However, whilst WGS did provide sequence data for each of the six genes coding for HPA, the authors acknowledged this approach was impractical and unaffordable for most laboratories and advocated using a targeted NGS panel as an alternative method for platelet genotyping (Lane *et al.*, 2016). Targeted NGS methods have been reported for HPA genotyping but approaches published to date have focused on the more common HPA systems included in a genotyping panel alongside blood group antigens (Wienzek-Lischka *et al.*, 2015; Orzinska *et al.*, 2017).

1.7 Aims and objectives

This study set out to exploit the advantages offered by the novel NGS technology with the aim of enhancing the HLA and HPA definition of both platelet donors and recipients. The objectives by which this would be achieved were:

- Comparison of a commercial HLA amplification protocol (Illumina TruSight™ HLA) with an in-house whole gene approach for PCR template generation
- Establishing a high throughput NGS based method for HLA-A, -B and -C genotyping platelet donors and assess its potential impact on the selection of platelet donors using HLA epitope-matching (HEM)
- Designing and developing a multiplex, high throughput NGS based method for HPA genotyping
- Designing and developing an NGS based method to sequence the genes coding for glycoproteins that express HPA.

1.8 Hypothesis

The use of NGS for HLA and HPA genotyping will enhance the provision of suitable platelet transfusions for patients with HLA and/or HPA alloreactive antibodies.

2. Methods and Materials

2.1 Samples

2.1.1 Donor samples for HLA genotyping

Informed consent was obtained from each donor in accordance with organisational protocols (see Appendix). EDTA blood samples were collected from random apheresis platelet donors at NHSBT clinics throughout England. All donors used in this study had been previously typed for HLA-A, -B and -C at intermediate resolution using Luminex-based sequence specific oligonucleotide technology (LABType®, One Lambda).

2.1.2 Control DNA for HPA genotyping

External quality assessment (EQA) DNA samples were selected for their range of genotypes from NHSBT DNA archives. EQA samples were originally distributed by the National Institute for Biological Standards and Controls, with HPA genotypes of each sample determined by over 40 international laboratories. Additional anonymised control DNA samples were provided by the Australian Red Cross Blood Service.

2.1.3 Patient DNA for HPA genotyping

DNA samples previously genotyped for HPA were retrieved from NHSBT DNA archives. They were originally obtained from patients or family members following referral to NHSBT for investigation of FNAIT or thrombasthenia. Informed consent was obtained by the hospital clinical team (see Appendix).

2.2 DNA extraction of donor samples

DNA was extracted from 7ml EDTA blood samples and using a DNA mini kit (Qiagen) on the QIASymphony® SP automated platform. Following extraction from a 200µl aliquot of EDTA blood, each sample was dissolved in 200µl of elution buffer and dispensed into 96-well plates, with 90 samples contained within each plate. Following DNA extraction, sample identifiers were automatically added to an XML rack file by the QIASymphony software v4.0.1. A local Microsoft (MS) Access program was then used to extract the sample IDs and respective plate positions from the rack file to generate an MS Excel 'plate map' used to facilitate downstream sample tracking.

2.3 DNA purification

Purification of DNA was performed on samples prior to HLA genotyping only and was performed either manually or utilising a Biomek FXP laboratory automated workstation (Beckman Coulter) prior to quantification. Briefly, 50µl each of DNA and AMPure® XP beads (Beckman Coulter) were incubated together for 5 minutes followed by separation from

solution using a magnet and the supernatant containing impurities discarded. This was followed by two wash steps in 80% ethanol, freshly made from 100% Ethanol (Sigma-Aldrich). Purified DNA was finally eluted in 30µl of Ambion® nuclease free water (NFW) purchased from Thermo Fisher Scientific. Any wells that did not contain purified DNA were filled with NFW and served as negatives controls.

2.4 DNA quantification

The concentration of double stranded DNA (dsDNA) present in each sample was determined using one of two fluorescence based assays. This approach for DNA quantification was selected over conventional UV absorbance because it is highly selective for dsDNA and therefore provided a more accurate assessment of the target concentration required for NGS. In general, samples in a 96 well plate format were quantified with the Quant-iT™ kit and concentration of DNA contained in tubes was measured using the Qubit®.

2.4.1 Quant-iT™ Assay kit, Broad Range (Invitrogen)

A standard curve was generated in accordance with the Quant-iT™ protocol, using 10µl of each DNA standard provided (range 0-100ng/µl) diluted in 200µl of working solution (1:200 dilution of Quant-iT dye in buffer). Test samples were prepared by adding 4µl of DNA to 206µl of Quant-IT BR working solution into the respective well of a black 96-well assay plate. Both Standards and test samples were then measured on a FiltermaxF3 plate reader (Molecular Devices) at 485nm and analysed using associated Softmax Pro software v6.3.

2.4.2 Qubit® dsDNA BR assay (Life Technologies)

Between 2 and 5µl of each DNA sample (dependant on material available) was made up to 200µl with Qubit® working solution (1:200 dilution of Qubit® dye in buffer) in thin-walled 0.5ml tubes (Qubit® Assay tubes). DNA sample concentrations were determined using a Qubit® 2.0 Fluorometer in accordance with the Qubit® dsDNA BR assay protocol.

2.5 DNA normalisation

DNA was normalised to the desired concentration either manually or using an automated protocol. Manual normalisation was used to prepare DNA samples for the HPA genotyping protocol, whereas DNA for HLA sequencing was normalised using the Biomek workstation.

2.5.1 Manual normalisation

DNA was diluted to the required concentration with 10mM Tris buffer (Buffer EB, Qiagen).

2.5.2 Automated normalisation

Values for DNA concentration obtained from the FiltermaxF3 in step 2.4.1 were imported into a spreadsheet with embedded macros which calculated the volumes of sample and NFW required for dilution of DNA to the desired concentration. The macro produced a Tag Image

File Format (TFF) file which was utilised by the Biomek® FXP liquid handling workstation (Beckman Coulter) to normalise each sample. Where the concentration was below the required input values, DNA was transferred neat.

2.6 HLA genotyping by Next Generation Sequencing

2.6.1 PCR amplification of HLA class I genes

PCR amplification of the HLA-A, -B and -C genes was performed on each sample using either a modified TruSight™ HLA Sequencing panel approach (Illumina) or with an in-house protocol. The primer binding sites for the TruSight HLA sequencing panel were not disclosed by Illumina, but the in-house protocol used primers designed to amplify the entire HLA gene (Figure 2.1).

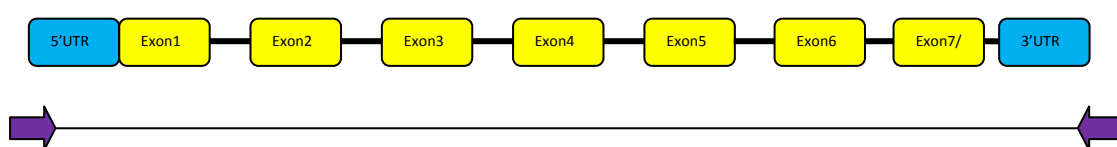


Figure 2.1 Location map of in-house PCR primers for HLA class I amplification. Previously reported locus specific primers (Shiina *et al.*, 2012) were located in conserved regions flanking each target (represented by the purple arrows) to amplify all exons, introns and un-translated regions (UTR) of each HLA gene.

2.6.1.1 TruSight™ HLA Sequencing panel

DNA was amplified using a modified version of the TruSight™ HLA PCR amplification protocol for HLA-A, B and -C, using a reduced reaction volume of 10µl compared to the prescribed 50µl. PCR reactions for each HLA locus were performed in 96-well PCR plates (4titude®) and consisted of 5µl of HLA PCR mix, 0.4µl of MasterAmp Extra-Long DNA polymerase, 1µl HLA locus specific primer, 2.6µl of NFW and 1µl of DNA at 10ng/µl. Amplification was performed in a 9700 thermal cycler (Perkin Elmer) with a ramp rate set to 9600 using the following program: 94°C for 3 minutes; 35 cycles of 94°C for 30 seconds, 60°C for 2 minutes and 68°C for 15 minutes; 68°C for 10 minutes; hold at 10°C.

2.6.1.2 In-house PCR protocol

Whole gene amplification of HLA-A, -B and -C in 96-well PCR plates (4titude®) was performed using previously described primers (Shiina *et al.*, 2012) but with a modified protocol. Each 10µl PCR reaction for HLA-A and -B consisted of 50ng purified DNA, 0.5U PrimeSTAR GXL DNA polymerase (TaKaRa Bio Inc), 1X PrimerSTAR GXL buffer, 200µM of each dNTP and 4pmol of each respective HLA locus specific primer. PCR reactions for HLA-C were as for HLA-A and -B but with the addition of 5% DMSO (Sigma-Aldrich). PCR amplification was performed using a

C1000 thermal cycler (Bio-Rad). HLA-A and -B conditions consisted of a 2 minute denaturation at 94°C followed by 30 cycles of 98°C for 10 seconds, 62°C for 20 seconds and 68°C for 5 minutes. HLA-C amplification conditions were 94°C at 2 minutes followed by 30 cycles of 98°C for 10 seconds, 56°C for 20 seconds and 68°C for 5 minutes. All amplicons were held at 4°C until removed from the thermal cycler.

2.6.2 PCR amplicon assessment

Prior to pooling, PCR products from the first and last row of each plate were either assessed using agarose gel electrophoresis or quantified using a fluorescence-based assay to determine if the amplification was successful. In general, gel electrophoresis was used for amplicon assessment prior to library preparation with TruSightHLA kits whereas quantification was performed if the DNA libraries were prepared using the GenDx protocol. Where either method indicated failure of amplification, the PCR for the respective locus was repeated, as described in 2.6.1.

2.6.2.1 Agarose gel electrophoresis

For each sample evaluated, 2µl of PCR product was run on a 1% agarose gel (w/v) in 0.5X TBE buffer (Thermo Fisher Scientific) stained with SafeView nucleic acid stain (NBS Biologicals Ltd). Gels were visualised under UV light using the Alpha Imager system and Alpha View software v1.3.0.7 to determine the presence or absence of an amplicon.

2.6.2.2 Quantification of PCR amplicons

PCR amplicons were quantified with the Quant-iT™ BR Assay (Invitrogen) according to the protocol described in 2.4.1.

2.6.3 Amplicon pooling

Seven microlitres of amplicon for HLA-A, B and -C from respective PCR plates for each sample were added to a clean PCR plate either manually or using a Biomek® NXP multichannel liquid handling workstation (Beckman Coulter) so that the resulting pooled plate containing 21µl of pooled amplicon in each well of a 96 well plate (Figure 2.2)

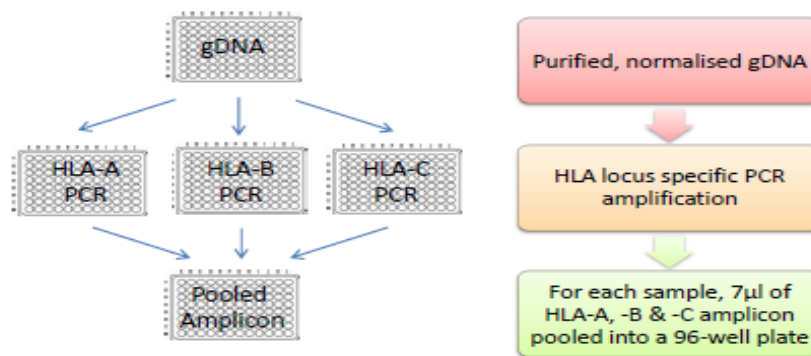


Figure 2.2 Summary of the PCR amplification and pooling workflow. Once purified and normalised, each genomic DNA (gDNA) sample was amplified for HLA-A, -B and -C. Following successful amplification, 7µl of each amplicon was pooled into the respective wells of a single 96 well plate.

2.6.4 DNA library preparation

Two commercial kits were used to prepare pooled amplicon libraries for sequencing on the MiSeq. Both approaches employed enzymatic fragmentation followed by a limited PCR amplification before combining each prepared sample to form a pooled amplicon library as outlined in Figure 2.3.

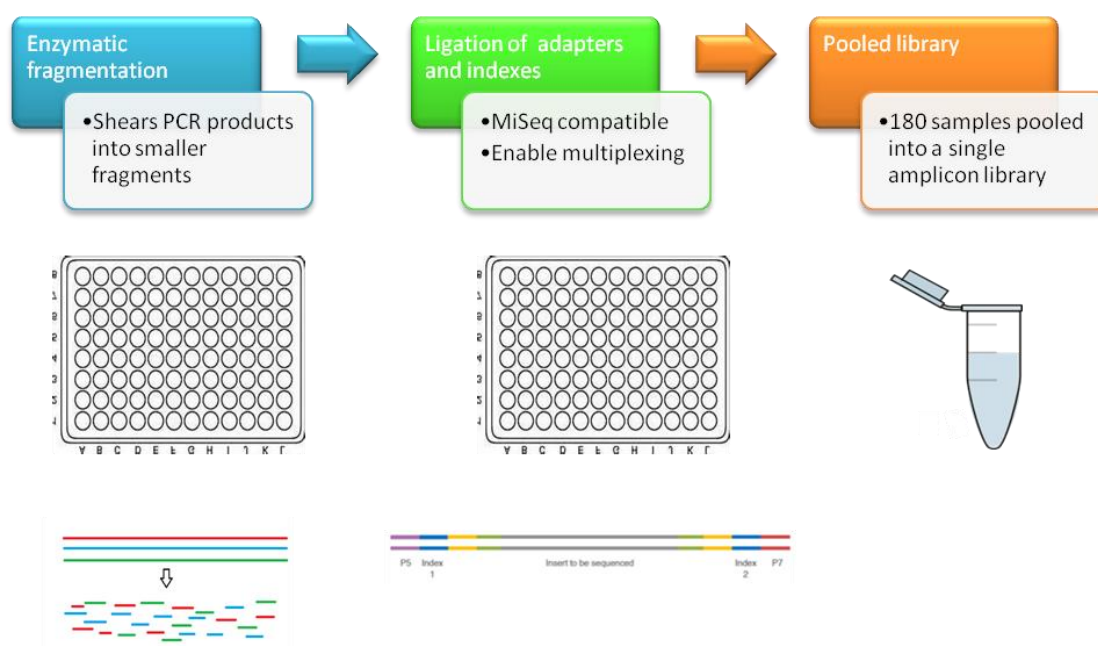


Figure 2.3 Summary of workflow for DNA library preparation. Enzymatic fragmentation of prepared amplicons was followed by adapter ligation. A limited PCR then attached unique sample multiplex identifiers and Miseq compatible sequences. Finally each prepared library was pooled into a single tube prior to downstream sequencing. A maximum of 180 samples were pooled for a single MiSeq run, consisting of libraries from 2 plates with 90 samples per plate.

Pooled DNA amplicons were prepared for sequencing by following either a modified version of the TruSight™ HLA preparation protocol (Illumina) or using NGSgo® library preparation kits (GenDx, Utrecht). The TruSight HLA protocol was initially employed as part of beta testing for Illumina, and provided free of charge. During the early stages of this study it was also the only library preparation protocol that had more than 96 indexes available. NGSgo® was selected for library preparation following the release of a 384 index set by GenDx. The key differences between each protocol are shown in Table 2.1 but each approach resulted in DNA fragments prepared from whole HLA gene amplicons labelled with sample-specific combinations of indices linked to MiSeq specific adapters in preparation for downstream sequencing.

Process step	TruSight HLA (Illumina)	NGSgo® (GenDx)
Quantification of amplicons	N/A	Yes
Purification of amplicons	Bead based	N/A
Normalisation of amplicons	Bead based	Biomek workstation
Tagmentation	Yes	Yes
Index PCR	Yes	Yes
Pooling	Yes	Yes

Table 2.1 Comparison of commercial kits employed for DNA library preparation. Similarities and differences are shown for the key steps of each protocol prescribed by Illumina and GenDx for their respective DNA library preparation kits, TruSightHLA™ and NGSgo®, respectively. The process starts with pooled whole HLA gene amplicons for each sample and results in a pool of up to 90 samples prepared for sequencing HLA-A, -B and -C on the MiSeq NGS platform.

2.6.3.1 TruSight™ HLA sequencing panel library preparation

Pooled amplicons were treated to remove impurities and small DNA fragments by following the TruSight™ HLA PCR clean up protocol but adding only 21µl of sample purification beads (SPB) rather than the 45µl indicated to maintain the 1:1 bead to amplicon ratio. Thereafter, DNA libraries were prepared according to the TruSight™ HLA sequencing panel library preparation guide, version August 2014.

Briefly, normalisation was achieved using a bead based method with subsequent tagging and fragmentation, so called ‘tagmentation’, performed whilst the amplicons were still attached to the normalisation beads. Eluted fragments were purified using 0.5X SPB to remove excess reagents and smaller fragments. This was followed by PCR amplification of the tagmented DNA, utilising Nextera XT Index sets A, B or C (Illumina). Following a final clean up with 0.5X

SPB, 5µl from each well were pooled into a single 1.5ml tube for each plate processed to form a pooled amplicon library.

2.6.3.2 NGSgo® Library preparation

Pooled amplicons were quantified using the Quant-iT™ BR Assay, as described in 2.6.2.1. Values obtained from the FiltermaxF3 were then imported into a spreadsheet with embedded macros which calculated the volumes of sample and NFW required for normalisation. The macro produced TFF which was utilised by a Biomek NXP Span-8 workstation (Beckman Coulter) to normalise each sample to 10ng/µl. Where the concentration was below 10ng/µl amplicon was transferred neat.

Following purification and normalisation, DNA libraries were prepared with NGSgo® reagents in accordance with manufacturer's instructions. Briefly, fragmentation, end repair and dA tailing of each sample was performed using NGSgo®FragX reagents. Adapter ligation with NGSgo®LibrX kits was followed by purification and size selection using 0.45X AMPure beads (Beckman Coulter). Indexing PCR utilised unique combinations of barcode indices from the NGSgo®IndX 384 set. Following a final clean up with 0.6X AMPure beads, 5µl from each well were pooled into a single 1.5ml tube for each plate processed to form a pooled amplicon library (PAL).

2.6.5 Assessment of the pooled amplicon libraries

Each PAL was assessed by one or both methods prior to sequencing to ensure it contained sufficient prepared library of the correct size to obtain a quality sequence. The 2100 Bioanalyser (Agilent) provided data on fragment size and molarity and was mainly employed in earlier experiments and for troubleshooting purposes due to the cost of the Bioanalyser chips. The Qubit® measured the total concentration of the library but did not provide data on quality.

2.6.5.1 2100 Bioanalyser

A 1µl aliquot of each PAL was assessed using the **Agilent High Sensitivity DNA Kit** on the 2100 Bioanalyser system (Agilent Technologies, Inc) in accordance with the manufacturer's instructions. The mean fragment size and profile was determined using the associated 2100 Expert software v1.03 (Agilent Technologies, Inc).

2.6.5.2 Qubit®

Determination of concentration was performed in triplicate using the Qubit® dsDNA HS assay (Life Technologies). For each PAL, a 5µl aliquot was made up to 200µl with Qubit® working solution in thin-walled 0.5ml tubes (Qubit® Assay tubes) and measured using a Qubit® 2.0 Fluorometer in accordance with the Qubit® dsDNA HS protocol.

2.6.6 Sequencing on the MiSeq

Each PAL was diluted in NFW as required to produce a final target concentration of 17pM. Ten microlitres of diluted PAL was added to an equal volume of freshly prepared 0.1N NaOH in a 1.5ml tube, mixed and incubated for 5 minutes at room temperature. This was followed by the addition of 980µl of Hybridisation buffer (Illumina) to form the diluted amplicon library (DAL), inverted to mix. Following preparation and loading of the flow cell, 600µl of DAL was added to a thawed v2 MiSeq reagent cartridge and loaded onto the MiSeq platform (Illumina), in accordance with the MiSeq system user guide (Illumina). Sample identifiers and their respective multiplex indices were recorded along with run parameters on the linked sample sheet. MiSeq run parameters were set to 2 x 251 cycles of paired end sequencing using a FASTQ only analysis application with adapter trimming.

2.6.7 Data analysis

Primary data analysis was performed using MiSeq Reporter v2.5.1 (Illumina) to generate a pair of FASTQ files for each sample. FASTQ files generated from the MiSeq run were then analysed using HLA specific analysis software, NGSengine (GenDX). The version of software and IMGT/HLA database used was dependent on the date analysis was performed. Analysis preferences were set to process a maximum of 100,000 reads per FASTQ file, using the standard phasing algorithm with a minimum read depth threshold defaulted to 20 bases and balance threshold set to 20%.

2.7 HPA genotyping by Targeted Next Generation Sequencing

2.7.1 Assay Design

A custom HaloPlex high sensitivity (HS) panel (Agilent Technologies, UK) was created using Agilent's online SureDesign HaloPlex advanced design wizard. Designs were based on hg19 GRCh37 reference sequences (Table 2.2), with target sources obtained from the RefSeq database (O'Leary *et al.*, 2016). Target parameters for each gene were set to include all coding exons with an extension of 50 bases from the respective 5' and 3' ends of each region.

Gene	RefSeq Target ID (based on GRCh37 reference sequence)
ITGB3	NM_000212
GP1BA	NM_000173
ITGA2B	NM_000419
ITGA2	NM_002203
GP1BB	NM_000407
CD109	NM_133493

Table 2.2 Target sequences employed for the bespoke HaloPlexHS design. RefSeq target identifiers (ID) shown were entered into the Agilent SureDesign Haloplex advanced design wizard. Target IDs were based on the GRCh37 reference sequences to ensure compatibility with the SureDesign software.

2.7.2 Targeted enrichment

Using the bespoke HaloPlexHS design as described in 2.7.1, DNA libraries were prepared by following the HaloPlex HS target enrichment system protocol vC0 December 2015, outlined in figure 2.4. Briefly, genomic DNA (gDNA) was quantified as described in section 2.4 and diluted to 1.8ng/μl in 10mM Tris buffer pH 8.5. Each diluted gDNA sample was digested with a panel of 16 restriction enzymes in the form of eight double digests for 30 minutes at 37°C along with an enrichment control DNA sample (ECD) provided with the kit. The restriction digest was validated by analysing 1:1 dilutions of inactivated ECD reactions on the 2100 Bioanalyser system (Agilent Technologies, Inc) using the Agilent High Sensitivity DNA Kit in accordance with the manufacturer's instructions.

Once digestion was confirmed, gDNA restriction fragments from each sample were pooled and hybridised to the custom HaloPlex HS probe library, with simultaneous incorporation of unique indexing primers occurring at 58°C for 2 hours. Following removal from the hybridisation buffer using AMPure XP beads (Beckman Coulter, UK), ligation reagents were added to the circularised fragments and incubated in a PTC-225 thermal cycler (MJ Research) at 55°C for 10 minutes to close nicks in the probe-target DNA hybrids.

Streptavidin labelled Dynabeads (Thermo Fisher Scientific, UK) were used to capture the biotinylated targets which were then amplified in a 9700 thermal cycler (Perkin Elmer) using the following PCR program: 98°C for 2 minutes; 25 cycles of 98°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute; 72°C for 10 minutes and held at 8°C. PCR products were subsequently purified with AMPure XP beads and evaluated with a Bioanalyser 2100 (Agilent Technologies, UK) or quantified using a Qubit (Thermo Fisher Scientific, UK), as described in 2.6.5.

2.7.3 Pooling of libraries

Prepared DNA Libraries were combined for downstream sequencing by either equimolar pooling or pooling in equal volumes. Each pooled library was validated with a Bioanalyser 2100 to assess fragment size and concentration (see 2.6.5) and normalised to approximately 4nM.

2.7.4 Sequencing on the MiSeq

Based on the Illumina MiSeq System Denature and Dilute Libraries Guide, 10pM of the pooled library was then loaded onto a MiSeq (Illumina) for 2 x 151 paired-end sequencing using either a MiSeq compatible Nano or Standard v2 cartridge and flow cell. The MiSeq Reporter (MSR) software (Illumina, Cambridge) settings were adjusted to generate FASTQ files for index reads.

2.7.5 Data analysis

Primary data analysis was performed using MSR to generate a pair of FASTQ files for each sample. FASTQ files generated were then analysed using SureCall NGS data analysis software v3.0.3.1 (Agilent Technologies, UK).

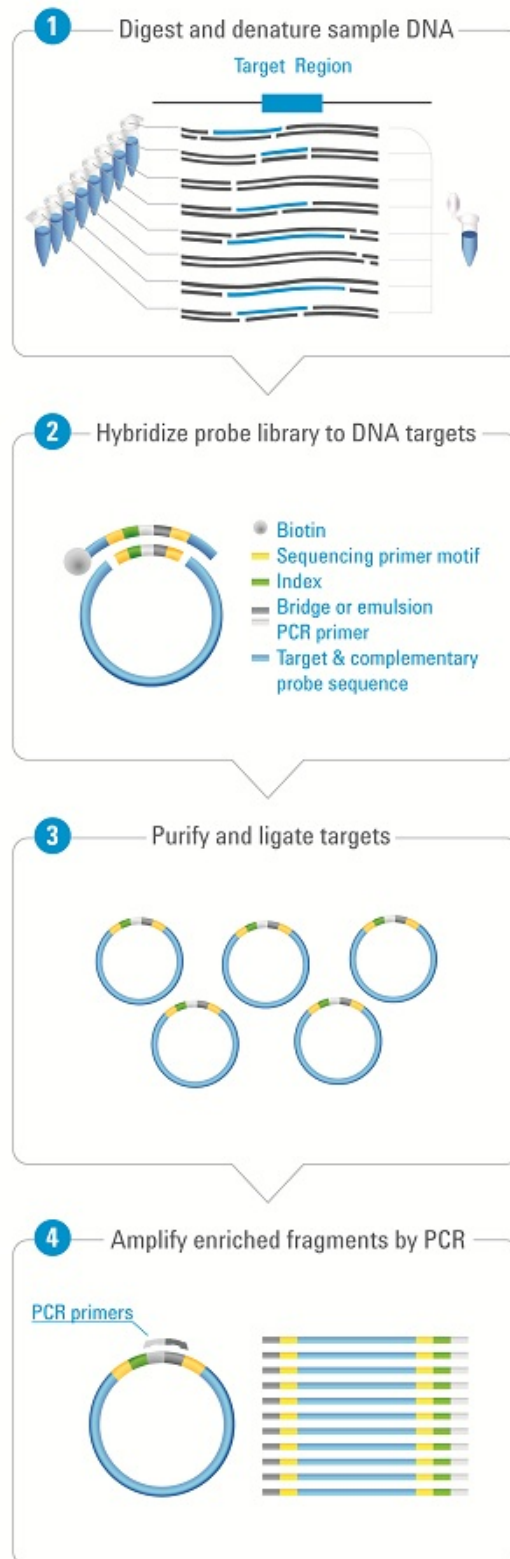


Figure 2.4 Overview of the HaloPlexHS target-enrichment sequencing sample preparation workflow. Sample DNA was first digested and denatured followed by hybridisation with the bespoke HaloPlexHS probe library when unique indexing primers were also incorporated. This was followed by purification and ligation of targets with each fragment then amplified using circularised PCR to prepare libraries for downstream sequencing. ©Agilent Technologies, Inc. 2nd February 2017. Reproduced with Permission, Courtesy of Agilent Technologies, Inc.

3. HLA typing of platelet donors by next generation sequencing

3.1 Introduction

Providing suitable platelet transfusions for patients who are refractory to random platelet transfusions due to the presence of HLA alloantibodies requires a large panel of HLA defined platelet donors (Pavenski, Freedman and Semple, 2012). Approximately 2500 apheresis donors per annum are currently recruited by NHSBT for this purpose, with each donor typed for HLA on two separate occasions for quality reasons. This level of commitment to maintain an active and accurately typed donor panel is high and requires high-throughput genotyping techniques (Lucas, 2013).

Numerous strategies are available for the selection of platelets to support refractory patients including the use of (a) cross-match negative (b) antigen avoidance and (c) HLA matched products (Kopko *et al.*, 2015). HLA epitope-based matching (HEM) is a more recent approach based on compatibility of antigenic determinants, but accurate comparison of donor and patient intra and inter locus HLA is likely to require high resolution or allele level typing (Duquesnoy, 2011). High resolution HLA typing is typically performed by Sanger sequencing but this method is unsuitable for high throughput due to costs and logistical constraints.

Next Generation Sequencing (NGS) is a novel approach to sequencing, with a massive increase in capacity over conventional technology (Metzker, 2010). Consequently, NGS been exploited for HLA typing in recent years, with laboratories adopting a variety of strategies ranging from imputation of HLA types from WGS data (Bauer *et al.*, 2016) to targeted sequencing of the HLA gene (Ehrenberg *et al.*, 2014), reviewed in section 1.6.1. Some of the reported approaches define HLA to the allele level, as required for HEM (Duquesnoy, 2008).

The aim of this initial work was to investigate the feasibility of applying NGS technology for HLA typing platelet donors. Included was an investigation of different approaches for PCR template generation and for preparation of DNA libraries prior to sequencing on the MiSeq NGS platform. Validity of results was provided by the development of stringent acceptance criteria and concordance with historical HLA data.

3.2 Materials & Methods

3.2.1 Sample preparation

Blood samples were collected from 540 random English apheresis platelet donors who had been HLA typed on a previous occasion. Samples were sent to the H&I laboratory at NHSBT, Colindale and DNA extracted from 200 μ l blood in batches of 90 samples using the QIASymphonySP. Following extraction DNA was purified using AMPure beads, quantified using the Quant-iT™ broad range assay kit and subsequently normalized to 25ng/ μ l using the Biomek FXP laboratory workstation (Figure 3.1), detailed in sections 2.1 to 2.2.

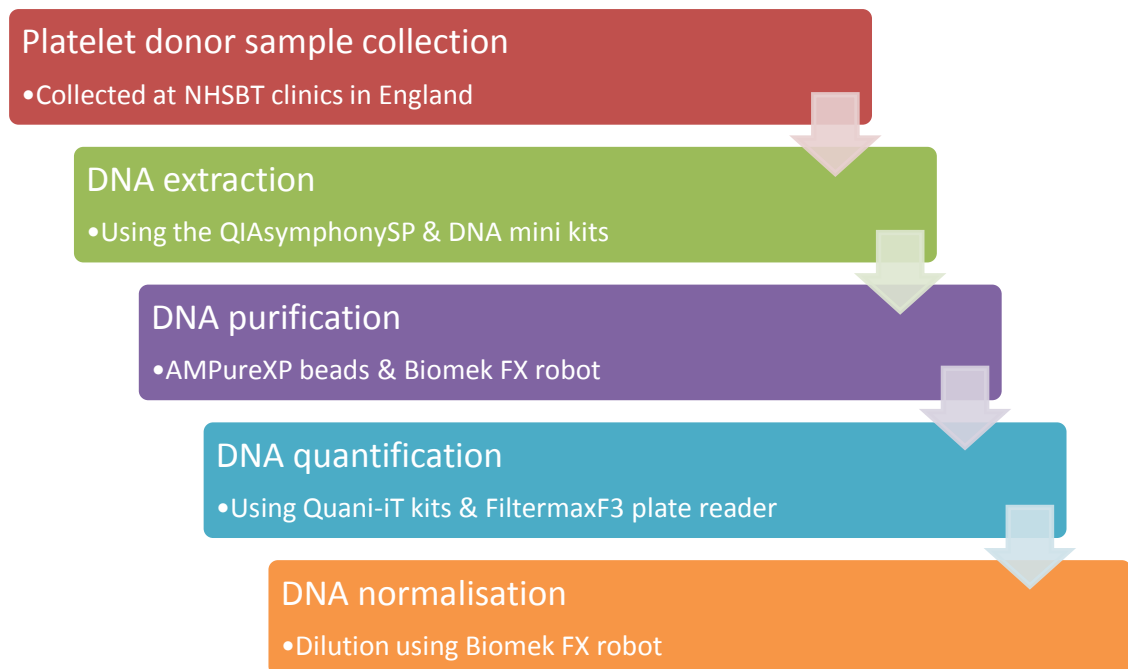


Figure 3.1. Summary of donor sample preparation prior to PCR amplification. Following platelet donor blood sample collection at NHSBT donor clinics, DNA is extracted, purified, measured and normalised to ensure the quantity and quality is optimal for whole HLA gene PCR amplification.

3.2.2 PCR amplification

The first 90 samples prepared (plate #457) were amplified using the TruSight HLA primers, which was performed as part of an early access trial using kits supplied by Illumina. However, due to the limited amount of reagents provided, instead of using the prescribed 50 μ l volumes the PCR reaction was reduced to 10 μ l and set up as described in section 2.6.1.1. The remaining 450 samples contained within 5 plates (#459, #462, #464, #467 & #469) were amplified using previously published primers (Shiina *et al.*, 2012) but with an in-house method (section 2.6.1.2). Modifications from the original conditions were made during optimization and

included a change to the reaction volume (10µl instead of 20µl) and adaptation of thermal cycling parameters. The addition of 5% DMSO (v/v) in the reaction mix was also required for HLA-C reactions.

3.2.3 PCR assessment and amplicon pooling

A 2µl aliquot was taken from the first and last row of each amplification plate to determine if the amplification had been successful. The method used for this assessment was determined by the downstream library preparation protocol, with agarose gel electrophoresis employed prior to TruSight library preparation and quantification used before the GenDx library preparation protocol (section 2.6.2). Any plates that indicated poor or failed amplification were repeated before proceeding. Following successful amplification, 7µl of PCR product generated for HLA-A, -B and -C for each sample were pooled to form a single plate of pooled amplicon (section 2.6.3).

3.2.4 Library preparation

The TruSight HLA library preparation method (section 2.6.3.1) was used to prepare the first two plates as reagents were provided free of charge by Illumina as part of an early access trial. At that time, Illumina were also the only provider of sufficient indices to allow multiplexing of >96 samples in a single MiSeq run. Subsequent libraries were prepared using the NGSgo[®] protocol (section 2.6.3.2) following the release of a 384 set of compatible indices by GenDx. In addition this kit had recently been validated for use by the H&I laboratory for sequencing adult volunteer haematopoietic stem cell donors for HLA-A, -B, C-, DRB1 and -DQB1. Table 3.1 summaries the protocol used for each plate.

Plate ID	Amplification protocol	Library preparation
#457	TruSightHLA (Illumina)	TruSightHLA (Illumina)
#459	In-house	TruSightHLA (Illumina)
#462	In-house	NGSgo [®] (GenDx)
#464	In-house	NGSgo [®] (GenDx)
#467	In-house	NGSgo [®] (GenDx)
#469	In-house	NGSgo [®] (GenDx)

Table 3.1. The PCR protocol and library preparation methods used for each plate of DNA. Six plates were amplified, with the first plate using Illumina’s PCR primers and the remainder by the in-house approach. Library preparation was performed using the TruSightHLA™ kits for plates #457 and #459 with the other plates prepared with NGSgo® reagents.

Both library preparation methods required some modification from their prescribed protocols, as each was designed by the respective manufacturer to process amplicons from a greater number of HLA loci and/or a larger PCR reaction volume than was produced by HLA class I

amplifications in this study. Therefore, rather than the preferred approach of combining amplicons based on concentration to avoid excess sequencing coverage of one locus to the detriment of another, HLA-A, -B and -C were pooled in equal volumes to ensure the minimum volume required by the downstream NGSgo[®] library preparation protocol was met.

For the TruSightHLA protocol, PCR products were pooled prior to purification and normalization, which differed from Illumina’s prescribed process which instructs preparation of amplicons separately. This adjustment was made to maximize the number of samples that could be sequenced with the reagents provided by Illumina. This change also required a modification to the volume of sample purification beads used for the initial purification of pooled amplicon which had to be reduced from 45µl to 21µl to maintain the correct 1:1 bead to amplicon ratio.

In order to maintain sample integrity, the index combinations used for each library preparation were carefully selected to ensure that each set provided a unique combination for each sample sequenced on a single MiSeq flow cell. Details of index combinations used were recorded on the respective sample sheets.

3.2.5 Sequencing on the MiSeq

To maximize the sequencing capacity of each run of the MiSeq, pooled amplicon libraries (PALs) prepared from two plates were combined together for sequencing (Table 3.2). Each PAL was assessed and diluted (as detailed in sections 2.6.5 and 2.6.6) prior to combining with its pair in equal volumes. The final PAL was then quantified using the Qubit[®] dsDNA HS assay prior to dilution, denaturation and loading onto the MiSeq.

MiSeq run	Plate IDs of the combined PALs
1	#457, #459
2	#462, #464
3	#467, #469

Table 3.2 Details of the pooled amplicon libraries (PALs) combined for each MiSeq run. Plate IDs are listed for each PAL and the combination of prepared libraries used for the three MiSeq runs.

3.2.6 Data analysis

FASTQ files generated by Miseq Reporter (MSR) software were transferred from the MiSeq to a high specification PC for analysis using HLA specific analysis software. Plates #457 & #459 were originally analysed with NGSengine v1.8 but were subsequently reanalysed with the later release v1.9 along with the remaining plates, to ensure the data was compared to the latest sequence data made available with IMGT/HLA 3.20. Each result was scrutinised and reviewed

against the acceptance criteria listed in Table 3.3. Mismatched nucleotide positions in either exons or introns were inspected to determine if they were genuine mutations or merely artefacts, assessing base calls in conjunction with adjacent positions. All homozygous results were scrutinised using the statistics data tab of the NGSengine software, which provided a visual representation of allele balance across each gene.

Quality parameter	Acceptance criteria
Average depth of coverage	>100 bp
No. of genotype matches	1
No. of exon mismatches	0
No. of intron mismatches	0
Homozygous	None
Atypical HLA associations	None
Presence of rare HLA alleles	None

Table 3.3 Quality parameters and respective acceptance criteria for data analysis using NGSengine analysis software. Each result was scrutinised against these parameters to review sequence quality and accuracy of HLA allele assignment.

3.2.7 HLA concordance with historical HLA types

Following scrutiny of each result, a concordance check was performed between the HLA types generated by NGS and the historical HLA results. Where the historical HLA type consisted of an HLA string rather than an individual allele, NGS results were considered concordant as long as the determined allele (or alleles if ambiguous) was contained within the historical HLA string. For example, an NGS result of HLA-C*03:04:01:01 was considered concordant with the original type as it is included in the HLA string determined by PCR-SSOP below, as underlined in bold:

C*03:02:01;*03:02:02:01;*03:02:02:02;*03:02:02:03;*03:02:03;*03:02:04;*03:02:05;*03:02:06;*03:02:07;*03:02:08;*03:02:11;*03:02:12;***03:04:01:01**;*03:04:01:02;*03:04:02;*03:04:03;*03:04:04;*03:04:05;*03:04:06;*03:04:07;*03:04:08;*03:04:09;*03:04:10;*03:04:12.

For the few historical results that were at first field only, results were considered concordant if the first two digits were the same as the original HLA type.

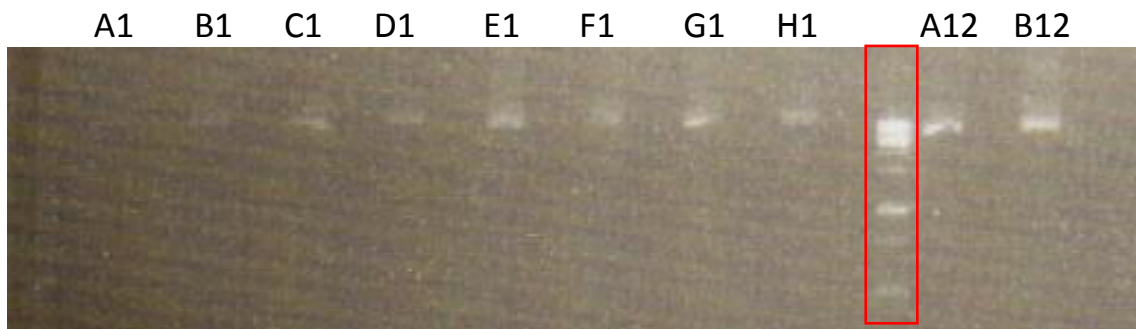
3.3 Results

DNA was extracted and purified resulting in an average concentration of 17.2ng/μl (range from 0 to 47.9ng/μl). PCR amplification was performed on DNA normalised to 25ng/μl (or used neat if below target concentration)

3.3.1 PCR amplification

Agarose gel electrophoresis was used to determine the success of PCR amplification for plates #457 and #459 only. PCR products of the expected length were observed in aliquots taken from the first and last row of each successfully amplified plate (Figure 3.2)

A: TruSightHLA amplicons from #457 HLA-A



B: In-house amplicons from #459 HLA-C

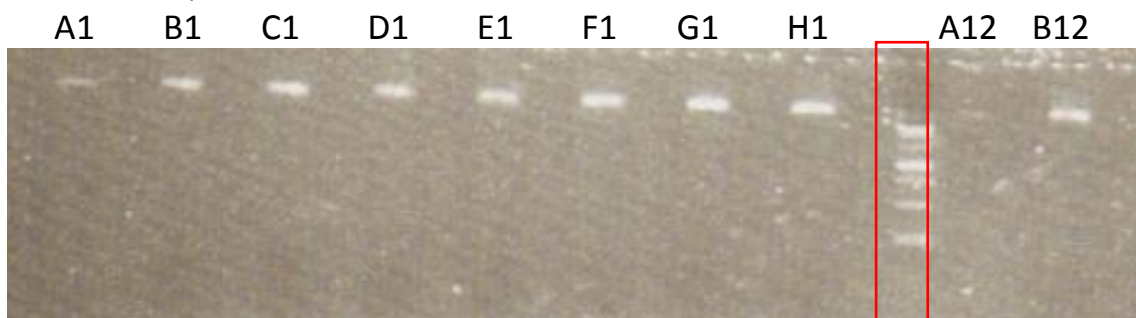


Figure 3.2 Example of HLA PCR amplicons run on a 1% agarose gel. Each lane is labelled with the respective plate position. Molecular weight marker (100bp) is highlighted by the red box. The top image shows PCR products taken from plate #457, amplified with Illumina's TruSightHLA primers for HLA-A. The bottom image shows HLA-C amplicons from plate #459, generated using in-house PCR. Specific PCR products are approximately 4kb in size. No positive control was included as all samples had been previously HLA typed and expected to contain a PCR product. Negative controls are not shown.

Amplicons from the first and last rows of the remaining plates were quantified using the Quant-iT BR assay prior to pooling and normalisation (Table 3.4). The concentration of HLA-C amplifications shown for plate #467 were those obtained from repeat amplification as the original PCR had failed due to poor plate sealing.

Plate ID	HLA-A amplicon mean concentration (ng/μl)	HLA-B amplicon mean concentration (ng/μl)	HLA-C amplicon mean concentration (ng/μl)
#462	20.5	45.9	29.9
#464	40.1	71.2	28.7
#467	64.9	82.6	42.8
#469	68.9	72.4	44.3

Table 3.4 Mean concentration of PCR amplicons obtained for each HLA locus. A 2μl aliquot of each amplicon was taken from the first and last rows of the respective plates following PCR amplification. The average concentration was determined from the values obtained for each plate.

3.3.2 Library quantification

Concentration and fragment size determined for each pooled amplicon library are shown in Table 3.5. The concentration obtained for each library was quite varied, ranging from 0.6 to 11.3ng/μl based on Qubit values with no correlation to library preparation method. The average fragment size of each library did appear to relate to the kit used, with GenDx library preparation producing smaller fragments on average. The fragment size distribution profile also looked atypical for PAL produced from plates #424 and #464 with NGSgo[®] kits, containing smaller fragments when compared to those generated by TruSightHLA, shown in Figure 3.3, although the reasons for this observation were not clear. This data was not available for the remaining plates prepared using NGSgo[®] kits.

Plate ID	Qubit [®] values (ng/μl)	Bioanalyser values (ng/μl)	Fragment size (kb)	Library preparation kit
#457	2.7	3.08	1764	TruSightHLA
#459	0.6	0.51	1434	TruSightHLA
#462	2.5	0.998	567	GenDX
#464	1.9	1.11	576	GenDX
#467	11.3	n/a	n/a	GenDX
#469	8.6	n/a	n/a	GenDX

Table 3.5 Mean concentrations and fragment sizes determined for pooled DNA libraries prepared from each plate. Concentrations were determined using the Qubit and 2100 Bioanalyser assays with the fragment size determined using the Bioanalyser assay only. (n/a = no data obtained)

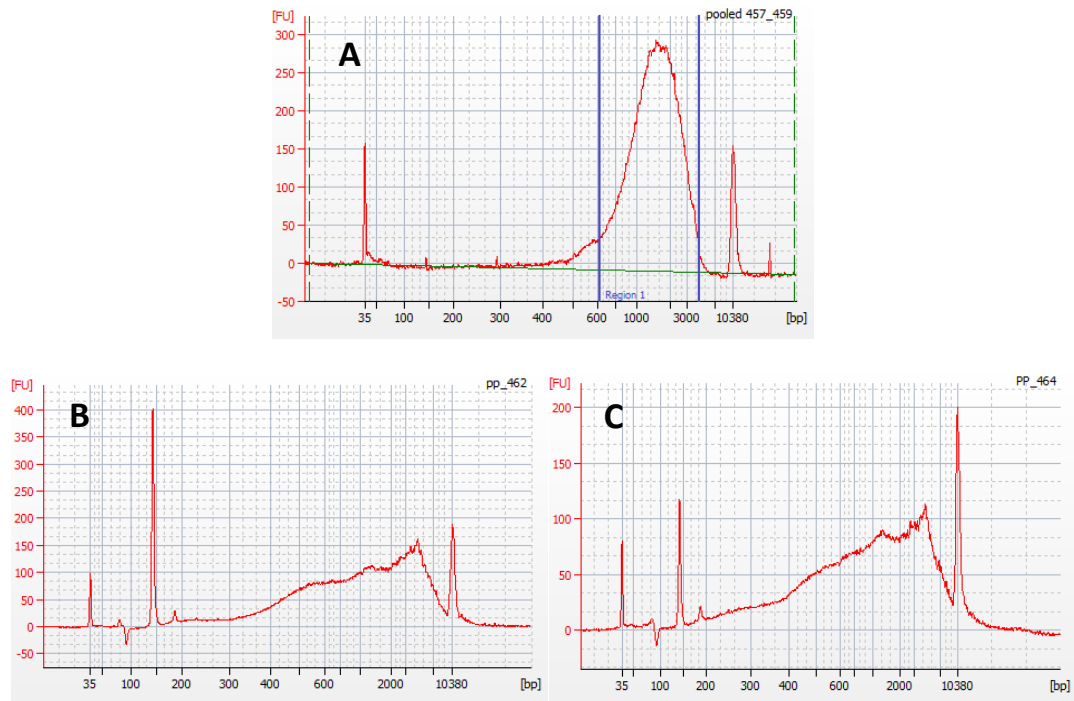


Figure 3.3 Atypical fragment size distribution for pooled amplicon libraries prepared from plates #462 and #464. (A) shows the expected bioanalyser trace, as observed for pooled amplicon library (PAL) prepared from plates #457 and #459. The bioanalyser profiles observed for PALs prepared from plates (B) #462 and (C) #464 had lower fluorescent units (FU) and had a lower and broader size range (bp).

3.3.3 Sequencing quality

The cluster density of the flow cell determines the amount of sequencing data generated, with lower cluster density producing a lower output. However, over clustering can also result in suboptimal data acquisition as fewer reads pass the filter due to reduced signal purity. This leads to higher background, poor base calling and, if excessive, over clustering can ultimately cause run failure (Illumina, 2016a). Cluster density on the flow cell was variable, ranging from 345 to 960 K/mm², with clusters passing filter > 88% on each occasion (Table 3.6).

Run No. & ID	Method of library preparation	Cluster density	% CPF	Mean read depth
1 #457#459	TruSightHLA	820	94.3%	1695
2 #462#464	NGSgo	960	88.1%	222
3 #467#469	NGSgo	345	95.9%	498

Table 3.6 MiSeq parameters observed for each sequencing run. Cluster density, % clusters passing filter (CPF) and mean read depth obtained according to MiSeq run and method used for library preparation

Following the completion of each run, FASTQ files were generated by MSR for all samples and successfully transferred for analysis on the PC using NGSengine. Quality indicators were reviewed for each sample and are summarised in Table 3.7. The number of reads analysed varied according to sample and HLA locus, with an overall average of 19032 reads processed by NGSengine but ranging from the maximum of 200,000 to 0 where no suitable sequence data had been generated. For accepted sequences the minimum number of reads processed per HLA locus was 998. An average of 862x per base coverage depth was observed, although this varied between the three experiments performed, with a mean of 1910, 207x and 369x respectively. Read length ranged from 38bp to a maximum of 250bp although for sequences that were subsequently accepted the range was narrower with a minimum of 105bp for each read.

Quality parameter	All sequence data	Accepted sequences only
Number of reads	19032 (0-200,000)	18799 (998-200,000)
Read length (bp)	190 (38-250)	192 (105-236)
Insert size (bp)	339 (0-633)	340 (59-633)
% Mappability to HLA genes	74.6% (0-99%)	75.2% (9-99%)
Read depth (bp)	862 (0-8752)	807 (36-7788)
No. of phased regions	1.9 (1-39)	1.6 (1-21)

Table 3.7. Summary of quality parameters observed for HLA sequence data. Results from all sequences generated from the three MiSeq runs, regardless of quality, compared with quality parameters observed with valid sequences only. Values are shown as mean average for each parameter with range in parentheses.

The percentage of reads that map uniquely to a reference sequence, so called ‘mappability’, can have a significant impact on read depth (Sims *et al.*, 2014). The range of mappability observed was very wide, even for sequences that met all acceptance criteria. However, there was a clear difference between the % mappability of reads generated using the TruSightHLA library preparation protocol used for plates #457#459 compared to those prepared using the NGSgo[®] kits, with an average of 94.0% and 64.8% respectively (Figure 3.4).

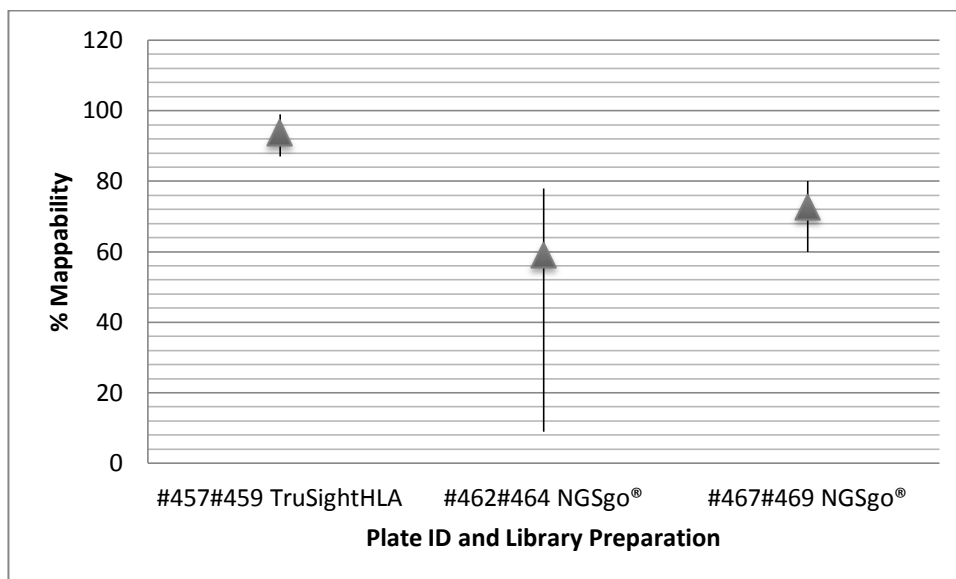


Figure 3.4 Percentage mappability of sequences generated. The range and average mappability of reads generated from each sequencing experiment, with the mean value represented by the triangle.

3.3.4 HLA typing results

In this study 540 samples were sequenced for HLA-A, -B and -C using the NGS protocols described above. Of the 1620 sequences performed, a total of 1466 (90.5%) met the defined quality acceptance criteria (Table 3.8).

Run ID	HLA-A	HLA-B	HLA-C
#457	89 (98.9%)	49 (54.4%)	89 (98.9%)
#459	86 (95.6%)	90 (100%)	89 (98.9%)
#462	76 (84.4%)	83 (92.2%)	85 (94.4%)
#464	83 (92.2%)	85 (94.4%)	82 (91.1%)
#467	88 (97.8%)	88 (97.8%)	86 (95.6%)
#469	75 (83.3%)	72 (80%)	70 (77.8%)
Total	497 (92.0%)	468 (86.7%)	501 (92.8%)

Table 3.8 The number of successful HLA types obtained per plate. Results are presented according to the run identifier (ID) and HLA locus. The percentage of successful results obtained for each plate are shown in parentheses.

In total 441 samples (82.7%) produced an allele level genotype for HLA-A, -B and -C. Of the remaining samples 69 failed to sequence one HLA locus, 5 failed for two HLA loci and 25 failed to produce valid sequence for any HLA locus, summarised in Figure 3.5.

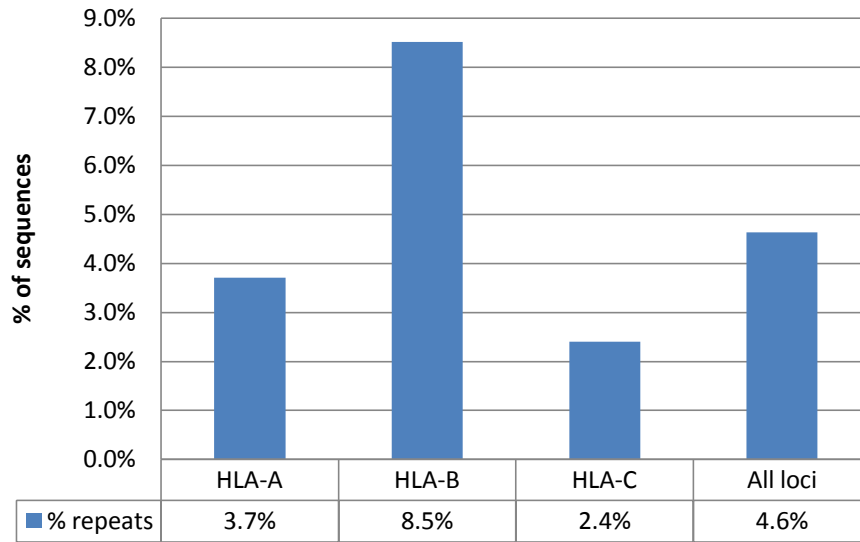


Figure 3.5. The percentage of sequences failing to meet quality acceptance criteria. Results are displayed according to HLA locus. The bar labelled 'all loci' shows the percentage of samples that failed to sequence for any HLA locus.

The number and HLA loci of failed sequences varied between each plate of DNA. Notable differences included a high rejection rate for HLA-B sequences in plate #457 and a higher than average number of samples failing for all HLA loci on plate #469 (Figure 3.6).

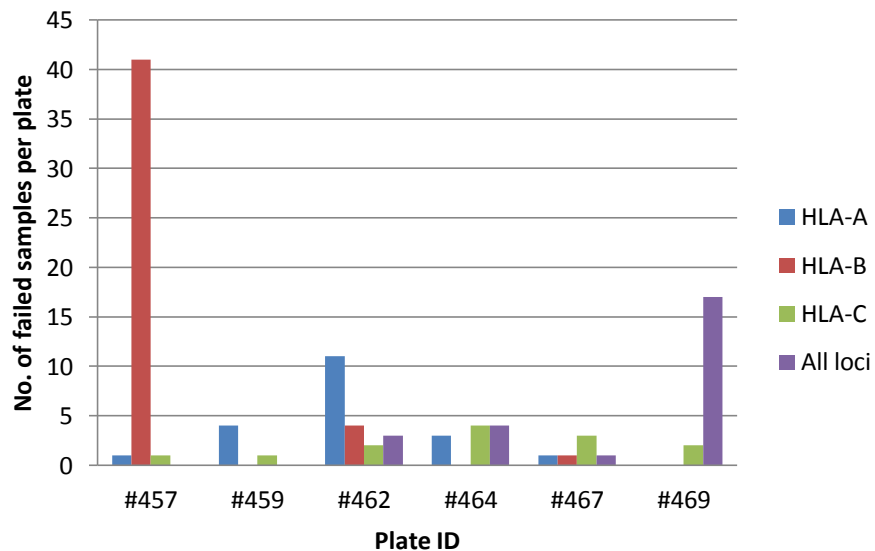


Figure 3.6. The number of samples with rejected sequences. Results are displayed according to plate identifier (ID) and HLA locus.

The reasons for rejected sequence also varied according to HLA locus (Figure 3.7). Low coverage was the primary reason for samples failing for all HLA loci and for HLA-A sequences deemed invalid, whereas HLA-B sequences were predominantly rejected due to the production

of ambiguous results. Allele imbalance was observed in some samples for HLA-B and -C but not for HLA-A. Three samples were rejected due to the presence of putative novel mutations in exon regions of the respective HLA-B or HLA-C genes (see 3.3.4.4).

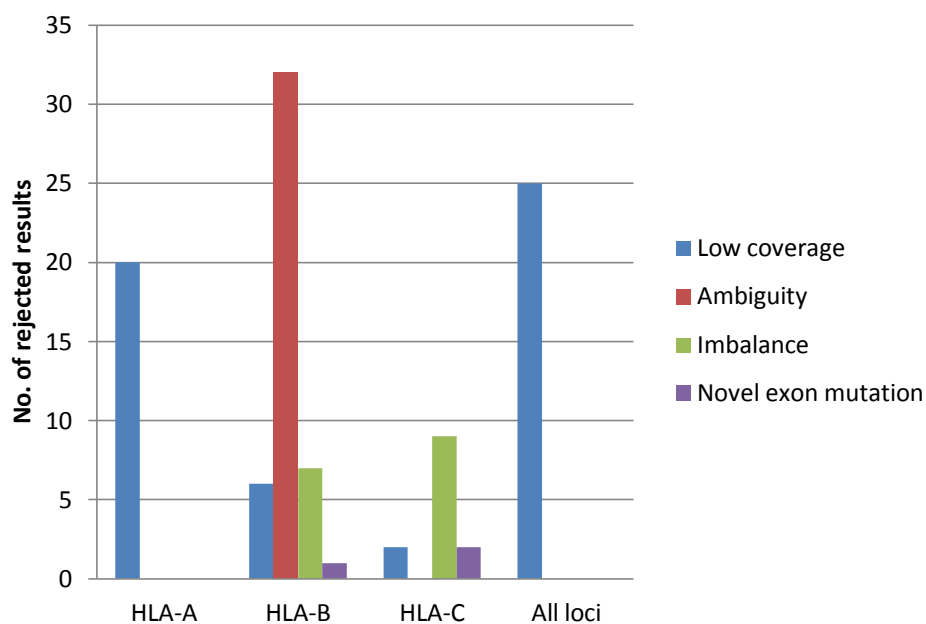


Figure 3.7 A summary of reasons for sequence rejection. Results are displayed according to HLA locus and reason for rejection.

3.3.4.1 Low sequencing depth

The sequence alignment of all samples with an average per base depth of coverage of <100 were carefully scrutinised. Where coverage was even and other quality parameters met, the result was accepted. However, NGSengine did not assign base calls for positions below the 20bp default cut-off so these results, or those where other quality flags were present alongside low coverage, were rejected. Poor per base depth of coverage was the primary reason for samples that failed for all HLA loci. Low coverage was also the cause of all failures at the HLA-A locus, for 6 samples for HLA-B and 2 for HLA-C. Variation in read depth was also observed for each plate, with those prepared using the TruSightHLA kits resulting in fewer rejected sequences due to low coverage compared to the NGSgo kits (Figure 3.8).

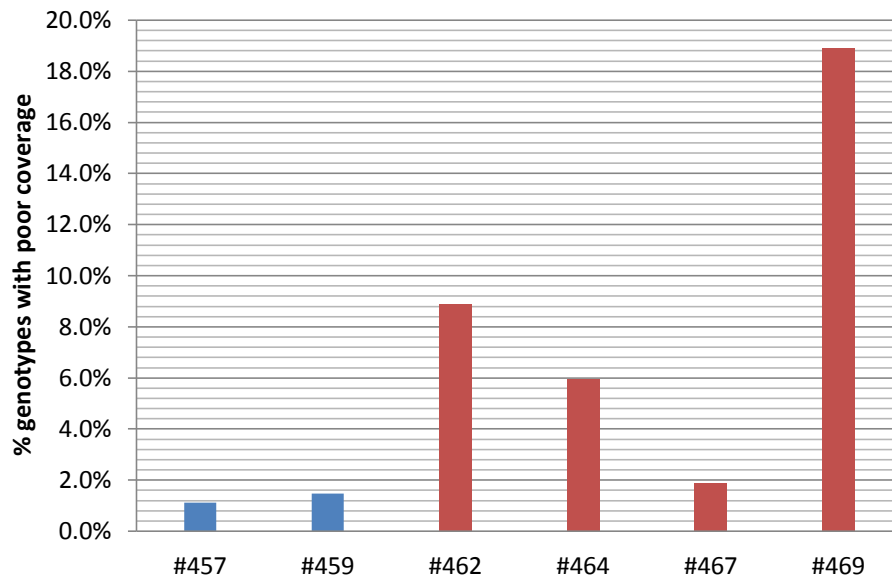


Figure 3.8 Percentage of sequences rejected due to poor per base depth of coverage. Results are displayed according to plate ID. Bars in blue represent libraries prepared with the TruSightHLA kits. The red bars signify plates prepared with NGSgo® reagents.

3.3.4.2 HLA ambiguity

The primary reason for rejected sequences at the HLA-B locus was ambiguity, observed in 32 samples. Ambiguity was indicated when the number of HLA genotypes matching to the NGS-generated nucleotide sequence was >1 and all other sequencing criteria were met. All examples resulted from TrusightHLA HLA-B amplification products. Of these, 20 sequences analysed were unable to resolve HLA-B*44:01:01:01 and B*44:19N due to NGS base calls beginning 10 nucleotides from the start of exon 1. In the example shown, one of the alleles (B*40:01:02) has been determined but the software cannot unambiguously assign the second allele, as it is unable to discriminate between B*44:02:01:01 and the null allele B*44:19N which would be defined by Guanine or a deletion at position 5, respectively (Figure 3.9). The remaining 12 ambiguous types included three instances of an HLA-B*27:05/B*27:13 ambiguity with the remaining HLA-B alleles ambiguous at fourth field.

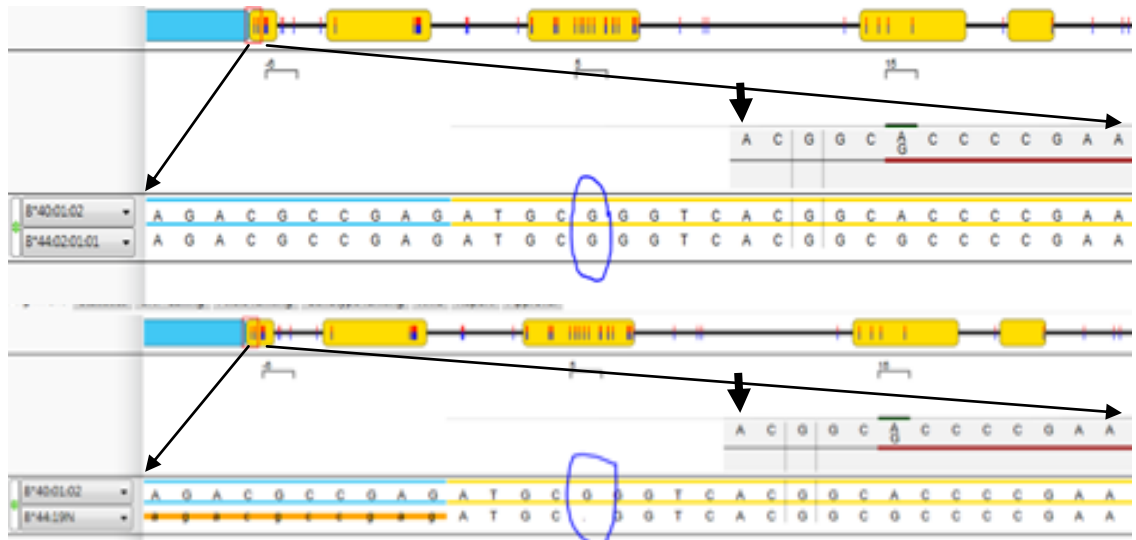


Figure 3.9 Ambiguity with HLA-B*44 alleles observed following TrusightHLA amplification. Screenshots from NGSengine showing the beginning of exon 1 in a sample with ambiguity between B*44:02:01:01 and B*44:19N. The top panel shows the expected base call at position 5 if the second allele is a B*44:02:01:01 and the bottom panel shows a deletion at position 5 required to assign B*44:19N, both circled in blue. The black arrows indicate the beginning of the aligned sequences generated by NGS at position 10 of exon 1.

3.3.4.3 HLA allele imbalance

In cases of imbalanced amplification where heterozygous positions for the minor allele were below the default 20% balance threshold, samples typed as apparent HLA homozygotes. However, the presence of a ‘hidden’ second allele could be visualized (Figure 3.10) and the allele balance threshold subsequently amended, although a lower limit of 10% for imbalance was maintained for all analyses.

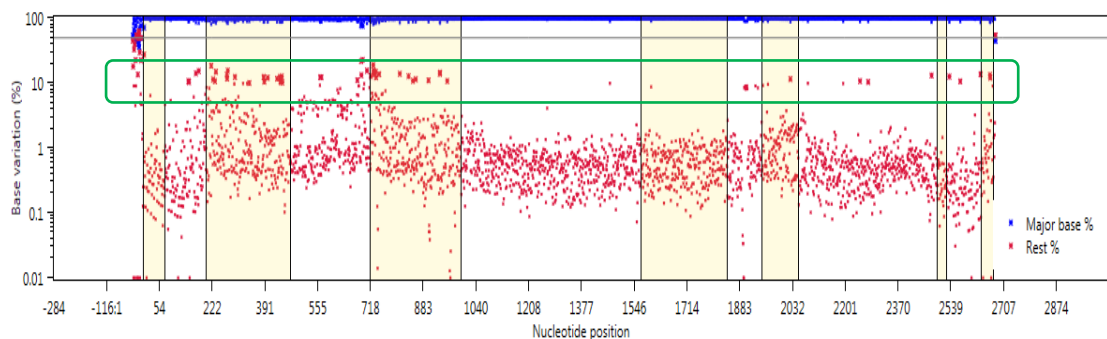


Figure 3.10 A screen shot of an imbalanced sequence. The ‘hidden’ minor allele base positions in red are highlighted by the green box, showing as approximately 10% of the overall sequence. The major allele alongside homozygous positions are shown in blue at the top of the image. The remaining red base positions towards the bottom of the image represent background noise.

A total of 15 results were rejected due to apparent bias in amplification, where one allele in a heterozygous sample failed to reach the minimum threshold of 10%. Of these, seven were HLA-B sequences derived from amplicons generated using TruSightHLA primers from Illumina, with the majority resulting from poor amplification of HLA-B*40:02:01. The remaining eight imbalanced sequences were observed in samples heterozygous for HLA-C, all due to inefficient amplification of HLA-C*07:04:01 using in-house PCR.

3.3.4.4 Novel exon mutations

Putative novel mutations in the exon regions of three samples from plates #464 and #469 were observed, as detailed in Table 3.9.

Plate	Sample ID	HLA allele	Nucleotide position	Exon	Mutation identified
#464	S142	C*03:03:03	2568	6	G>A
#467	S70	B*15:10:01	42	1	G>A
#467	S33	C*03:04:01:01	1810	4	T>A

Table 3.9 Putative novel mutations observed in plates #464 and #467. Identifiers (ID) of each sample containing a putative novel HLA allele along with each respective mutation and its location in the HLA allele affected.

3.3.4.5 Intron mismatches

Mismatches with the reference sequence within introns were observed in 220 sequences analysed using NGSengine v1.9 that were otherwise considered acceptable. Reasons for intronic mismatching were variable and included the presence of putative novel mutations, artefacts or inconclusive base calls and from poor phasing of sequencing fragments. Another common reason for apparent mismatch with the reference was lack of sequence data available from IMGT/HLA, where GenDx had inserted 'presumed' sequences for affected alleles but which was discordant with the sequences obtained in this cohort (Figure 3.11).

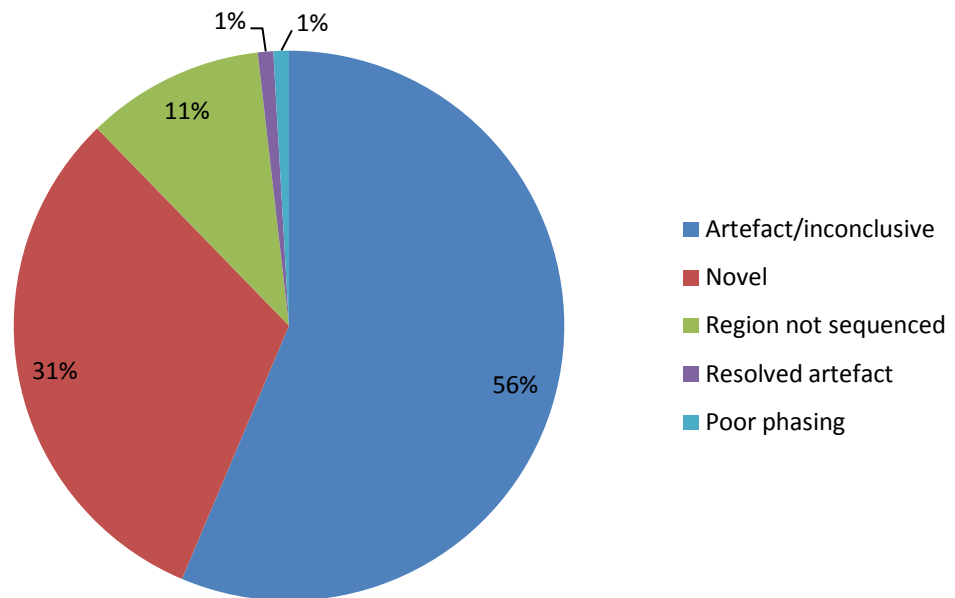


Figure 3.11. Reasons for intron mismatches to the reference sequence. Intron mismatches were detected in 220 sequences analysed using NGSengine v1.9. An explanation for these observed mismatches are displayed as a percentage of the overall number of intron mismatches.

Sequences with intron mismatches were subsequently re-analysed following the release of NGSengine v1.10 soon after the original data was produced, resulting in 28 (12.7%) of these anomalies being resolved. Of these, two samples were typed following removal of intron artefacts. A further two samples containing the allele HLA-B*44:27 were successfully reanalysed using the updated software, both previously having failed to phase correctly. The remaining 24 samples were resolved because previously undefined HLA sequences had been added to the updated IMGT/HLA database (v3.22) which was included in the v1.10 NGSengine release (Table 3.10).

HLA alleles resolved with NGSengine v1.10	No. of examples in cohort
A*03:01:01:05	1
A*03:01:01:06	1
A*24:02:01:04	1
A*24:02:01:05	1
A*68:01:02:03	2
B*14:02:01:02	1
B*15:01:01:04	1
B*15:03:01:02	1
B*44:03:01:02	1
B*47:01:01:03	1
C*04:01:01:06	11
C*17:01:01:05	2

Table 3.10 HLA alleles resolved using NGSengine v1.10. The alleles in this table previously failed to type using NGSengine v1.9 indicating intron mismatches. Due to the inclusion of previously undefined sequences in IMGT/HLA release v3.2.2, these alleles were subsequently resolved.

3.3.5 HLA Concordance

Concordance with historical HLA results for the 1555 genotypes determined by NGS was very strong, at 99.2%, 97.9% and 97.9% for HLA-A, HLA-B and HLA-C respectively. Of the 26 results found to be discordant, 22 NGS derived sequences were of poor quality and were rejected. The remaining 4 discrepancies resulted from errors in the original typing, detailed in Table 3.11.

Sample ID	Batch ID	NGS result	Historical Typing result
S17	PP462	B*27:05:02; B*52:01:01:02	B27; B51
S17	PP462	C*01:02:01; C*12:02:02	Cw1; #
S54	PP462	B*07:02:01; B*40:01:02	B*07:02; B*40:02
S174	PP469	C*04:01:01:01; C*08:92	C*04:01:01:01/*04:01:01:02/*04:01:01:03/*04:01:01:04/*04:01:01:05/*04:01:01:06/*04:01:01:07/*04:01:01:08/*04:01:01:09/*04:01:01:10/*04:01:01:11/*04:01:01:12/*04:01:01:13/*04:01:01:14/*04:01:01:15/*04:01:01:16/*04:01:01:17/*04:01:01:18/*04:01:01:19/*04:01:01:20/*04:01:01:21/*04:01:01:22/*04:01:01:23/*04:01:01:24/*04:01:01:25/*04:01:01:26/*04:01:01:27/*04:01:01:28/*04:01:01:29/*04:01:01:30/*04:01:01:31/*04:01:01:32/*04:01:01:33/*04:01:01:34/*04:01:01:35/*04:01:01:36/*04:01:01:37/*04:01:01:38/*04:01:01:39/*04:01:01:40/*04:01:01:41/*04:01:01:42/*04:01:01:43/*04:01:01:44/*04:01:01:45/*04:01:01:46/*04:01:01:47/*04:01:01:48/*04:01:01:49/*04:01:01:50/*04:01:01:51/*04:01:01:52/*04:01:01:53/*04:01:01:54/*04:01:01:55/*04:01:01:56/*04:01:01:57/*04:01:01:58/*04:01:01:59/*04:01:01:60/*04:01:01:61/*04:01:01:62/*04:01:01:63/*04:01:01:64/*04:01:01:65/*04:01:01:66/*04:01:01:67/*04:01:01:68/*04:01:01:69/*04:01:01:70/*04:01:01:71/*04:01:01:72/*04:01:01:73/*04:01:01:74/*04:01:01:75/*04:01:01:76/*04:01:01:77/*04:01:01:78/*04:01:01:79/*04:01:01:80/*04:01:01:81/*04:01:01:82/*04:01:01:83/*04:01:01:84/*04:01:01:85/*04:01:01:86/*04:01:01:87/*04:01:01:88/*04:01:01:89/*04:01:01:90/*04:01:01:91/*04:01:01:92/*04:01:01:93/*04:01:01:94/*04:01:01:95/*04:01:01:96/*04:01:01:97/*04:01:01:98/*04:01:01:99/*04:01:01:100/*04:01:01:101/*04:01:01:102/*04:01:01:103/*04:01:01:104/*04:01:01:105/*04:01:01:106/*04:01:01:107/*04:01:01:108/*04:01:01:109/*04:01:01:110/*04:01:01:111/*04:01:01:112/*04:01:01:113/*04:01:01:114/*04:01:01:115/*04:01:01:116/*04:01:01:117/*04:01:01:118/*04:01:01:119/*04:01:01:120/*04:01:01:121/*04:01:01:122/*04:01:01:123/*04:01:01:124/*04:01:01:125/*04:01:01:126/*04:01:01:127/*04:01:01:128/*04:01:01:129/*04:01:01:130/*04:01:01:131/*04:01:01:132/*04:01:01:133/*04:01:01:134/*04:01:01:135/*04:01:01:136/*04:01:01:137/*04:01:01:138/*04:01:01:139/*04:01:01:140/*04:01:01:141/*04:01:01:142/*04:01:01:143/*04:01:01:144/*04:01:01:145/*04:01:01:146/*04:01:01:147/*04:01:01:148/*04:01:01:149/*04:01:01:150/*04:01:01:151/*04:01:01:152; C*08:02:04/*08:02:07

Table 3.11 Details of NGS results found to be discrepant with the historical typing data. Discrepant alleles/antigens are highlighted in red. # = presumed homozygous.

3.4 Discussion

A total of 540 samples obtained from English apheresis platelet donors were sequenced for HLA-A, -B and -C using an NGS approach as described above, with over 90% of the 1620 genotypes performed meeting the stringent acceptance criteria. Reasons for rejection of the remaining 154 sequences varied according to amplification protocol and the method used for library preparation.

3.4.1 Amplification failure

Although the majority of samples amplified successfully, it is likely that some samples failed to sequence for one or more HLA locus because of poor amplification. However, only a proportion of PCR products were checked either by gel electrophoresis or quantification so it is not possible to identify those samples affected. Ideally all amplicons should have been checked prior to pooling and library preparation, but it was logistically challenging to repeat adhoc failures when all samples are contained within a 96-well plate. Therefore a decision to continue with library preparation was made unless the sampling of PCR amplicons indicated the entire plate of DNA failed to amplify, such as occurred for plate #467 for HLA-C.

3.4.2 Ambiguity

No ambiguous HLA results were produced following amplification with the in-house PCR protocol. However, use of the TruSightHLA primers for amplifying DNA from plate #457 resulted in ambiguity for 12% of HLA-B genotypes. This can be explained by the primers used by the Illumina kit. Although the exact binding site and primer sequence is not disclosed by the manufacturer, the expected size of amplicon for HLA-B of 2.6kb quoted in the TruSight HLA sequencing panel library preparation guide is significantly smaller than the 4.1kb amplicon for HLA-A and 4.2kb for HLA-C. Scrutinising the results in more detail it was clear the HLA-B primers do not amplify the entire gene and thus some crucial polymorphic positions in exon 1 were not sequenced, as shown in section 3.3.4.2. Conversely, in-house PCR amplification uses primers that bind outside the coding regions for each HLA class I gene, and therefore all positions that affect HLA allele assignment are determined.

As the TruSightHLA primers were provided as part of an early access program, concerns regarding this ambiguity were raised with Illumina, noting that the in-house amplification protocol was superior to the commercial kit. It is worth noting that the later HLA sequencing panel released by Illumina in 2016 (TruSightHLA v2) claimed to amplify exons 1 to 6, producing a slightly longer product of 2.8kb for HLA-B (Illumina, 2016b). Although these primers have not been tested at NHSBT, a recent paper suggests the ambiguity for HLA-B has been improved in the updated protocol, with unambiguous typing at third field increasing from 84% using

TruSightHLA v1 PCR primers to 91% with the later version of the TruSightHLA kit (Profaiser *et al.*, 2017).

3.4.3 Allele imbalance

Both amplification approaches resulted in allele bias for a number of samples, but the affected HLA loci differed according to protocol. TruSightHLA amplification produced imbalanced HLA-B sequences whereas in-house PCR affected HLA-C. Allele bias during amplification of a biallelic system such as HLA is a well documented issue (Lange *et al.*, 2014; Ehrenberg *et al.*, 2014; Nelson *et al.*, 2015; Carapito, Radosavljevic and Bahram, 2016). During the optimisation phase of the in-house protocol, allele drop out of HLA-C*07:04:01 in heterozygous samples was observed. In an attempt to resolve this anomaly, both betaine and DMSO were tried as potential additives to the PCR reaction mix as they are well documented PCR enhancing agents (Chakrabarti and Schutt, 2001; Kang, Lee and Gorenstein, 2005; Strien, Sanft and Mall, 2013). Betaine appeared to have an inhibitory effect for HLA-C as no PCR product was obtained when it was added to the reaction mix. Use of 5% DMSO was then tested as modifications such as the addition of DMSO in HLA class I amplification reactions using the same primer designs has been previously reported to reduce allele bias (Ehrenberg *et al.*, 2014). However although the addition of DMSO did result in some successfully amplified HLA-C*07:04:01 alleles, it did not produce balanced amplification in all samples, with only 4 of the 12 HLA- C*07:04:01 amplifications meeting the minimum 10% threshold. This is the subject of further work to resolve.

3.4.4 Low coverage

Both methods used for library preparation produced even coverage across the HLA class I genes, comparing favourably with alternative methods such as the generic Illumina kit, Nextera XT, which demonstrates extreme bias across the HLA class I genes with coverage loss in GC rich regions (Lan *et al.*, 2015).

Average per base coverage depth for the majority of samples was above the arbitrary threshold of 100bp. This cut off was based on dialogue with colleagues although there is no consensus of opinion on minimum coverage depth for HLA typing by NGS (Shiina *et al.*, 2012; Wang *et al.*, 2012; Gabriel *et al.*, 2014; Gabriel *et al.*, 2014). The latest amendments to accreditation standards from the American Society for Histocompatibility and Immunogenetics do not specify a minimum average coverage but stipulate these values must be obtained from 'laboratory-generated validation data' (ASHI, 2014). A more detailed assessment of results suggested that read depth <100bp alone did not invalidate the result. Consequently sequences with low coverage were accepted if other key quality parameters were met and an unambiguous HLA type was assigned. Interestingly, there was a higher percentage of

sequences rejected due to low coverage from DNA plates prepared with the NGSgo kits compared to the TruSightHLA library preparation protocol. The reasons for this are not clear, and do not seem to directly correlate with cluster formation on the flow cell.

Samples that failed to produce coverage for any HLA locus were most likely due to lack of amplicon template or errors made during library preparation. Lack of template may have resulted from problems with the initial PCR, amplicon pooling or normalisation. The TruSightHLA protocol uses a bead-based normalisation compared to the quantification and subsequent dilution of pooled amplicons required prior to NGSgo[®] library preparation. Differences in approach to normalisation for the two protocols may affect the accuracy of template added to the tagmentation steps. The size distribution of the pooled amplicon libraries shown in Figure 3.3 certainly suggests that the fragmentation process was less optimal when using NGSgo[®] library preparation kits compared to TruSight[™]HLA. Lack of sufficient template for tagmentation may result in over fragmentation resulting in reduced library yield and coverage drop out as small fragments are removed during the purification steps. Likewise, insufficient tagmentation may produce larger libraries that do not cluster efficiently on the flow cell (Illumina, 2015). It is possible that the bead-based normalisation is more accurate than using a manual normalisation process, resulting in a lower failure rate.

An alternative explanation of the higher failure observed with the NGSgo[®] protocol might relate to the very small reagent volumes used compared to the TruSightHLA protocol, for example 1.25µl versus 5µl of each index added, respectively. Difficulties with pipetting during library preparation may have caused inaccurate measurement of reagents. Failure to add sufficient indexing primers may lead to inefficient amplification or failure of the index PCR step, with fragments not attached to the MiSeq I5 and I7 adapter sequences and therefore unable to bind to the flow cell. This would lead to reduced sequence generated for some or all DNA fragments from affected samples.

Low coverage may also result from poor mappability of sequences, where short sequences are rejected as the software algorithm fails to map them to a unique part of the reference sequence because of shared sequence motifs between HLA class I genes (Erlich, 2012). This is supported by data shown in figure 3.12 which demonstrates a strong correlation between size of sequencing insert and % mappability and the consequence of low mappability on read depth.

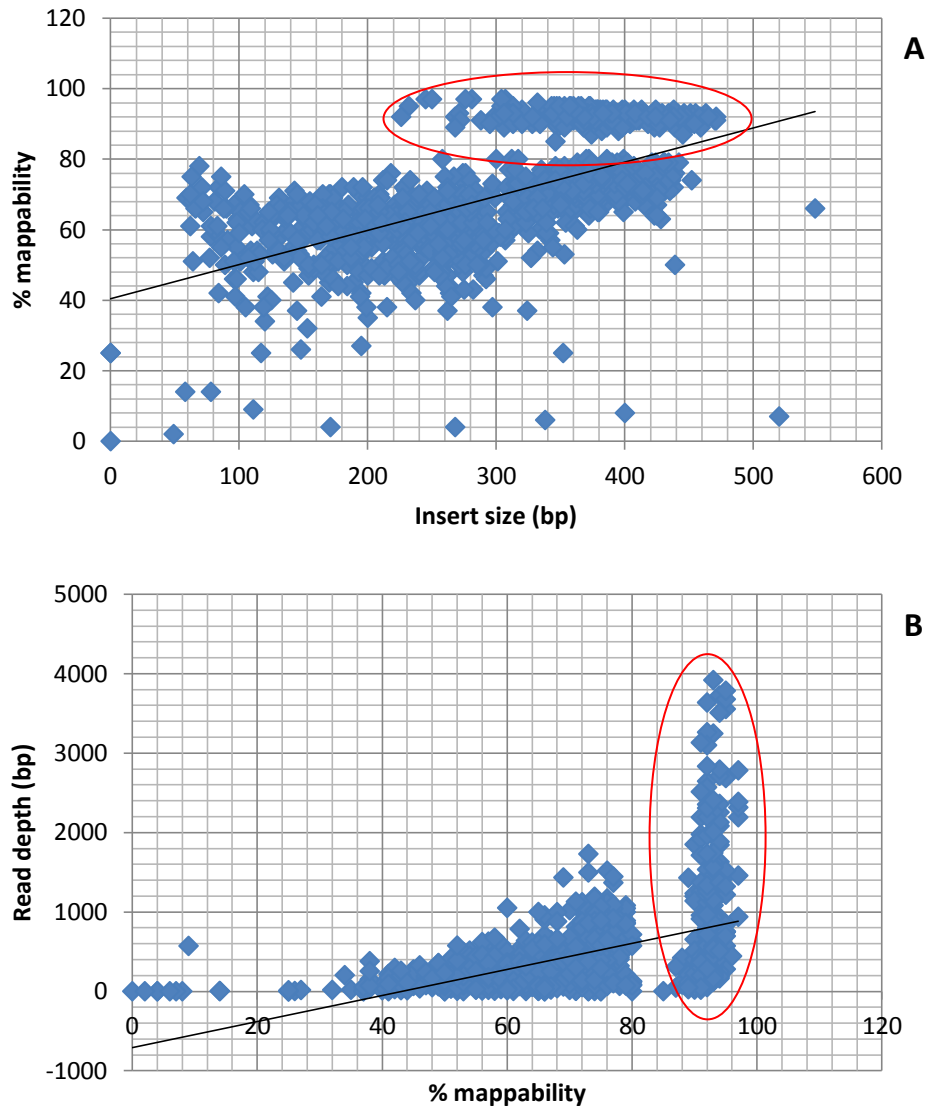


Figure 3.12 Insert size, sequence mappability and read depth. Graph (A) shows insert size versus % mappability observed for all HLA sequences in the cohort, indicating that larger fragments map more accurately to their respective HLA class I genes. Graph (B) demonstrates the effect that % mappability has on read depth, with deeper sequencing observed when a high percentage of fragments are mapped to the respective HLA reference sequence. Data circled in red was generated using the TruSightHLA library preparation kits with the remainder obtained using NGSgo® reagents

Read length will also impact on the ability of the analysis software to phase reads correctly, required to determine whether polymorphic positions are in *cis* or *trans* in order to avoid ambiguity (Tu *et al.*, 2017). Reads of less than 100bp in length are reportedly a problem as they are unable to phase non-polymorphic regions between heterozygous positions (Huang *et al.*, 2015) which will impact on the accuracy of HLA typing (Profaiser *et al.*, 2015). However, it is still possible to produce a single genotype, despite lack of phase (Lind *et al.*, 2013).

3.4.5 Intron mismatches

Intron mismatches with the reference sequence were observed in 15% of valid sequences. Whereas results with mismatches in the exons were rejected, those high quality sequences with intron mismatches were deemed acceptable as this would not impact on amino acid sequence. However, the result would only be considered and reported at third field.

More than half of these 'mismatches' were probably artefacts resulting from poor phasing or inaccurate mapping to the reference sequence. Intron mismatches identified in 22 sequences were due to the analysis software substituting regions of missing reference data with surrogate nucleotide sequence leading to erroneous intron mismatches, a known limitation of NGSengine (Duke *et al.*, 2016). More recent versions of NGSengine allow these mismatches to be ignored by the user.

A significant proportion (32%) appeared to be genuine novel mutations. HLA class I sequences are named by the WHO HLA Nomenclature Committee for Factors of the HLA System following attainment of a unique accession number from either EMBL, GenBank or the DDBJ databanks (Robinson *et al.*, 2015). Once accepted the submitted sequence are stored in the IMGT/HLA database and, until recently, most HLA class I sequences submitted to the IMGT/HLA database were based on data from only exons 2 and 3 due to historical sequencing techniques (Lind *et al.*, 2013). Although full genomic sequences are available for the most common HLA class I types, much of the intronic data is still missing but with more laboratories using NGS for HLA typing it is anticipated that many of the gaps in the data will be filled (Robinson *et al.*, 2015). Indeed, re-analysing novel sequences with a more up to date version of NGSengine produced valid genotypes due to release of identical sequences in IMGT/HLA v3.22. The remaining putative novel mutations are discussed in more detail in chapter 5.

3.4.6 Exon mismatches

Three sequences were found to contain novel mutations in the exons and are discussed in more detail in chapter 5.

3.5 Conclusions

The results presented in this chapter demonstrate that NGS has the capacity to accurately HLA type 180 platelet donor samples simultaneously at the allele level. The use of stringent acceptance criteria supported by a very high level of concordance with historical HLA typing indicates that NGS is a valid approach for high throughput, high resolution HLA typing of platelet donors. The use of an in-house amplification protocol was shown to be superior to the

commercial primers, providing a template for whole gene sequencing of HLA-A, -B and -C. Some variability in outcome was observed between the two methods employed for library preparation, although insufficient data was available to confidently determine that one method was superior to another. Further sequencing experiments using both library preparation protocols to prepare DNA from the same starting amplicons would provide a more realistic comparison.

It is worth mentioning that since the experiments described in this chapter were performed, the routine laboratory at NHSBT transferred from NGSgo[®] kits to library preparation with TruSightHLA reagents for stem cell registry donor typing. This followed a European tender where logistics, data quality and costs were considered. TruSightHLA had the added benefit of a developed automated protocol on the Biomek FX robot that was subsequently implemented in August 2016.

The application and benefits of this method for HLA typing apheresis platelet donors to support HLA epitope matching of platelet transfusions will be the subject of further analysis and discussion in Chapter 4.

4. HLA epitope matching: allele prediction or NGS typing?

4.1 Introduction

The provision of HLA selected platelets is a proven treatment for patients with immunological platelet refractoriness (Kopko *et al.*, 2015) but utilising HLA epitope matching (HEM) may be more relevant than matching the whole HLA, with the use of programs such as HLAMatchmaker providing an improved evaluation of histocompatibility (Duquesnoy, 2008). In addition, HEM may also increase the pool of available donors for patients with HLA alloantibodies, which would be of significant benefit as maintaining a large platelet donor panel for the provision of HLA matched platelets is costly (Pavenski *et al.*, 2013).

Epitope matching algorithms require HLA defined at second field i.e. at the amino acid level to accurately determine the HLA epitopes present in a donor or patient (Duquesnoy, 2008). As the majority of English platelet donors are currently typed using Luminex PCR-SSOP, it would be necessary to convert the medium resolution HLA-types obtained to the second field level, an approach used in previous studies reporting on the efficacy of HEM (Brooks, MacPherson and Fung, 2008; Pai *et al.*, 2010). Results discussed in Chapter 3 demonstrate that it is possible to define HLA to the allele level using NGS but this is likely to incur additional costs over conventional HLA typing technology. Since second field HLA resolution can be achieved using a computer algorithm, an assessment of the accuracy of allele prediction from medium resolution typing is required to determine the cost benefit of moving to NGS for routine platelet donor typing (Brooks, MacPherson and Fung, 2008; Pai *et al.*, 2010; Nambiar *et al.*, 2006).

At NHSBT, the laboratory information management system (LIMS) used by the H&I department is Hematos IIG (Savant, Cumbria UK). This system enables users to search NHSBT stocks for suitable platelet donations for patients requiring HLA and or HPA selected platelets (Mwandoro *et al.*, 2015). Currently the standard search determinates are based on conventional HLA match grades (Brown and Navarrete, 2011). However, Hematos also has the capability of using HEM for platelet selection, with the intention of routine implementation following the successful outcome of an ongoing double blind, non-inferiority trial (ISCTRN23996532). The current Hematos algorithm for platelet donor searching using HEM defaults to the 'most likely HLA allele', deemed to be the first allele in the string, whenever the HLA type of the donor or patient is not typed to at least second field (Kallon, 2015).

In this chapter, the HLA typing results from 540 random apheresis platelet donors were used to determine the accuracy of HLA allele prediction. HLA alleles were predicted from their respective historical low to medium resolution HLA types and compared with the alleles

determined for each donor by NGS to assess the accuracy of allele prediction and the potential impact of any observed disparity.

4.2 Materials and methods

The HLA typing results were obtained from 540 random apheresis platelet donor samples, as described in Chapter 3.

4.2.1 HLA allele prediction

Historical HLA types were used to predict the HLA-A, -B and -C alleles for each donor using the algorithms described in 4.2.1.1 and 4.2.1.2. Both methods of prediction were performed manually and all predicted HLA alleles were defaulted to second field due to lack of available frequency data at third and fourth field.

4.2.1.1 Selecting the first allele in the HLA string

Where historical results consisted of an HLA string, the first allele in the string was selected as the most likely HLA allele. This algorithm was based on the current HLA allele prediction approach embedded in the Hematos epitope matching program.

4.2.1.2 Selecting the most frequent allele

The most frequent allele for each historical HLA type was predicted using the EpHLA-Converter component of the on-line software application 'EpHLA' (Sousa, Luiz Cláudio Demes da Mata *et al.*, 2011). This software enabled conversion of low to medium resolution results into second field HLA alleles. The population 'EUR' was selected for each HLA type converted, which employed Caucasian HLA allele frequency data based on American populations (Maiers, Gragert and Klitz, 2007). This algorithm was based on the approach used by HLAMatchmaker (Brooks, MacPherson and Fung, 2008).

4.2.2 Comparison of predicted versus defined HLA alleles

HLA typing for HLA-A, -B and -C was performed by NGS on DNA extracted from 540 English apheresis platelet donors, as described in Chapter 3, section 3.2. The sequencing results obtained were compared with the HLA alleles predicted using both algorithms described in 4.2.1 to determine if the prediction was correct. Comparisons were performed at second field only.

4.2.3 Determining the impact of HLA allele prediction

Any discrepancies between the HLA alleles predicted from the historical typing and NGS defined alleles were scrutinised to determine the reason for the inconsistency. Differences due to poor sequencing were excluded from any further analysis. The remaining allele

discrepancies were investigated for any potential mismatched HLA epitopes between the predicted and NGS results using the ABC database from the International Registry of Antibody-Defined HLA Epitopes (Duquesnoy *et al.*, 2013b). Each HLA locus was inspected separately with the predicted allele entered as the recipient type and the NGS result entered into the search field. Following application of the filter, mismatched epitopes displayed were recorded, noting in particular whether any mismatches were against exposed epitopes and if they were antibody verified (Duquesnoy, 2014).

4.3 Results

4.3.1 HLA allele predictions

HLA predictions based on historical HLA types were performed on all 540 samples in the cohort. Of these, a comparison of the predicted results with NGS defined alleles was performed on 501, 503 and 509 results for HLA-A, -B and -C respectively, with remaining samples missing data. Discrepant predictions were either due to (a) the historical HLA type containing a rare allele as its first allele in the string, (b) the NGS result being a less frequent allele, (c) poor sequencing data or (d) errors with the historical typing. The overall number of discrepancies between the predicted and NGS defined alleles was greater when the first allele in the string algorithm was used, compared to those based on allele frequencies and regardless of the HLA locus (Figure 4.1).

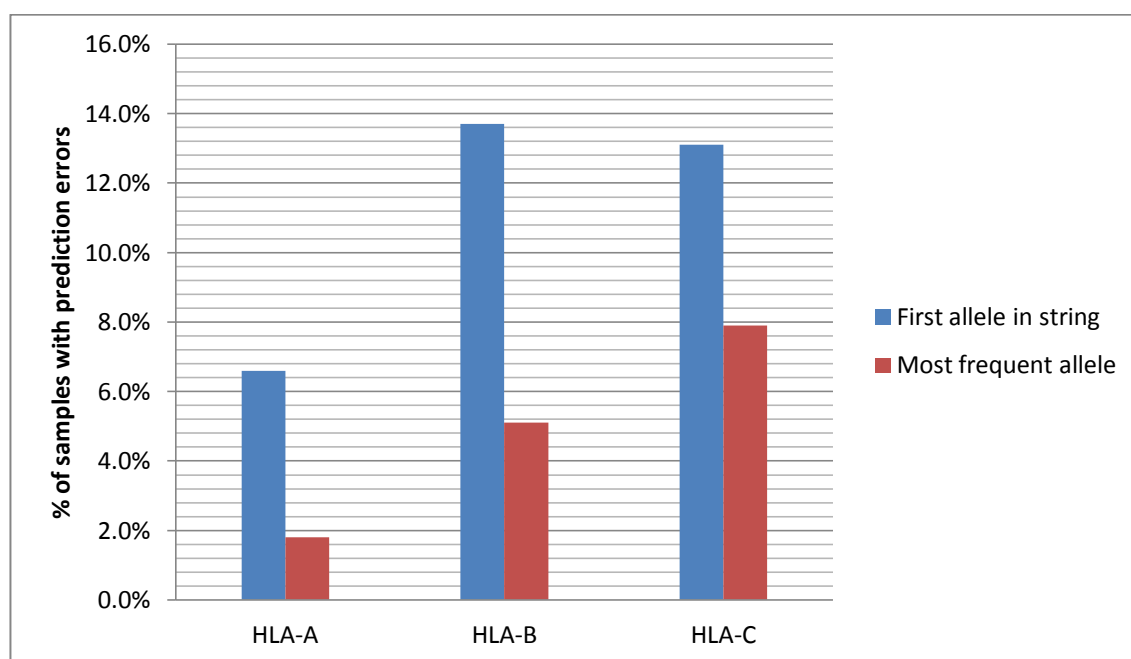


Figure 4.1 Discrepancies in HLA allele prediction. The percentage of samples with discrepancies observed between the predicted allele and the NGS result, using algorithms based on either the first allele or the most frequent allele contained within the historical HLA string for HLA-A, -B and -C.

Table 4.1 shows examples of allele predictions for each HLA locus made using both the first allele in the string and HLA frequency algorithms based on second field results. Those alleles determined as the most frequent in their respective allele group were HLA-A*29:02, -B*44:02 and -C*07:01, having a frequency of 3.279%, 9.011% and 16.658%, respectively. This compares with the second most common alleles in the same allele groups, namely HLA-A*29:01, -*44:03 and -C*07:02, which have frequencies of 0.216%, 4.963% and 15.006%, respectively (Maiers, Gragert and Klitz, 2007).

Historical medium resolution type	First HLA allele in string	Most frequent HLA allele
A* 29:01 :01:01/29:01:01:02N/29:01:02/29:01:04/29:01:05/ 29:02 :01:01/29:02:01:02/29:02:02/29:02:04/29:02:06/29:02:07/29:02:08/29:02:09/29:02:10/29:02:13/29:04/29:06/29:08N/29:10/29:12/29:16/29:17/29:21/29:23/29:25/29:26/29:27/29:29/29:36/29:38/29:41/29:42/29:43/29:44/29:45/29:46/29:50/29:54/29:55/29:58	A*29:01 (0.00216)	A*29:02 (0.03279)
B* 44:02 :11/*44:03:01/*44:03:03/*44:03:04/*44:03:05/*44:03:07/*44:03:08/*44:03:09/*44:03:10/*44:03:11/*44:03:12/*44:03:14/*44:03:18/*44:13/*44:26/*44:35/*44:36/*44:38/*44:39/*44:73/*44:85/*44:89/*44:92/*44:94/*44:98/*44:103/*44:108N/*44:114/*44:115/*44:122/*44:125/*44:141/*44:142/*44:147/*44:155/*44:157/*44:159/*44:161/*44:164/*44:165/*44:167/*44:174/*44:175/*44:178	B*44:02 (0.09011)	B*44:02 (0.04963)
C* 07:01 :25/* 07:02 :01:01/*07:02:01:02/*07:02:01:03/*07:02:01:04/*07:02:01:05/*07:02:02/*07:02:04/*07:02:05/*07:02:06/*07:02:07/*07:02:08/*07:02:09/*07:02:11/*07:02:12/*07:02:14/*07:02:15/*07:02:16/*07:02:17/*07:02:18/*07:02:19/*07:02:21/*07:02:22/*07:02:23/*07:02:24/*07:02:25/*07:02:26/*07:02:27/*07:02:29/*07:02:30/*07:02:31/*07:02:32/*07:02:33/*07:02:36/*07:02:37/*07:02:38/*07:02:39/*07:05/*07:13/*07:27:01/*07:29/*07:31:02/*07:38:01/*07:38:02/*07:39/*07:42/*07:46/*07:47/*07:48/*07:50/*07:51/*07:54/*07:56:01/*07:56:02/*07:61N/*07:62/*07:66/*07:72/*07:74/*07:75/*07:80/*07:84/*07:85/*07:87/*07:88/*07:90/*07:97/*07:99/*07:100/*07:105/*07:117/*07:123/*07:125/*07:126/*07:130/*07:133/*07:135/*07:137:01/*07:137:02/*07:143/*07:144/*07:145/*07:146/*07:147/*07:152N/*07:155/*07:157/*07:159/*07:160/*07:163/*07:167/*07:168/*07:169/*07:172/*07:174/*07:175/*07:178/*07:183/*07:185/*07:187/*07:192/*07:193/*07:195/*07:198N/*07:202/*07:208/*07:209/*07:211/*07:213/*07:217/*07:218/*07:221/*07:225/*07:226/*07:229/*07:234/*07:239/*07:240/*07:243/*07:244/*07:245/*07:251/*07:252/*07:258/*07:259/*07:261/*07:262/*07:270/*07:273/*07:274/*07:275/*07:283/*07:284/*07:286/*07:287/*07:288/*07:289/*07:290/*07:291/*07:292	C*07:01 (0.16658)	C*07:02 (0.15006)

Table 4.1 Example of HLA alleles predicted from a medium resolution HLA type. Alleles were predicted from the same historical HLA result based on either the first allele in the HLA string or the most frequent allele, defaulted to second field. The respective allele frequencies are shown in parentheses (Maiers, Gragert and Klitz, 2007). Predicted allele(s) are underlined in bold in the historical HLA string.

Figure 4.2 shows the number and type of discrepancies observed for each HLA locus. The dominant discrepancy for HLA-A, using the 'first allele in the string' algorithm was observed in 29 samples typed originally as an HLA-A*29 string by Luminex, resulting in an allele prediction of A*29:01, but where the NGS result was determined as A*29:02. However, this discrepancy was not observed when utilising frequency data for allele prediction. Conversely, when using allele frequencies, the opposite occurred in three samples, where NGS typed as HLA-A*29:01 but the most frequent and therefore predicted allele was actually HLA-A*29:02. For HLA-B, the most common discrepancy was the incorrect prediction of B*44:02 instead of B*44:03. This disparity between predicted HLA allele and NGS result was observed in 17 samples using either algorithm, as HLA-B*44:02 was both the first allele and the most frequent in the respective HLA-B string.

HLA-C had the highest allele prediction error of the three loci. The two most common discrepancies observed were between the predicted alleles HLA-C*07:01 and -C*03:02 and their respective NGS defined types of HLA-C*07:02 and -C*03:04. The HLA-C*03:02 vs C*03:04 prediction error was observed in 30 samples when using the first allele in the string. However, this discrepancy was resolved using allele frequency data, as HLA-C*03:04 is more common than HLA-C*03:02 with a frequency of 8.215% compared to 0.146%, respectively. Thirty four samples were incorrectly predicted to be HLA-C*07:01 rather than -C*07:02 using either algorithm due to C*07:01 being both the first allele in the string and the most frequent HLA-C*07 allele.

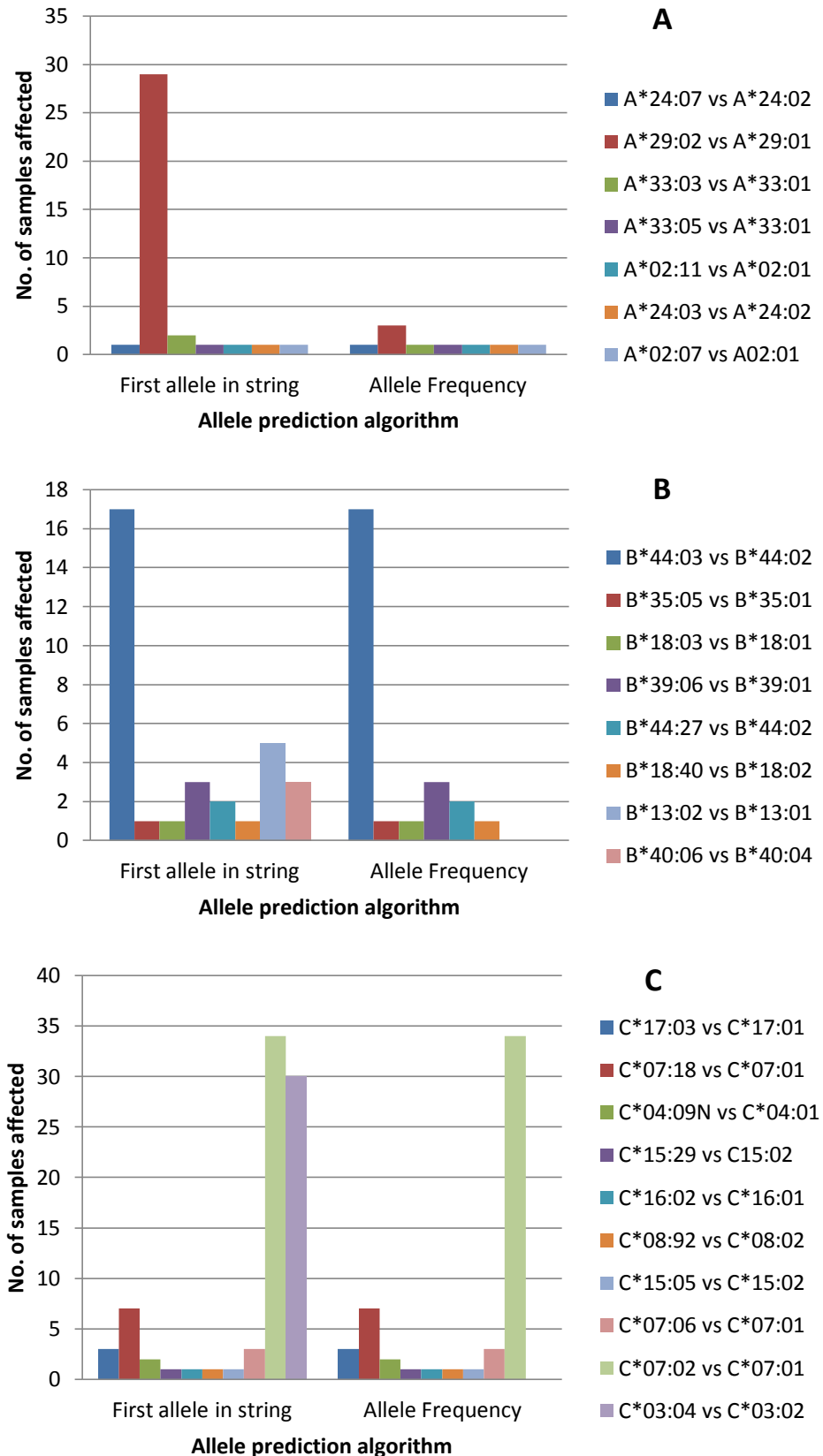


Figure 4.2 Discrepant alleles observed and number of samples affected. The discrepancies observed between the HLA type determined by NGS and the predicted allele. The number of samples affected by the respective prediction errors are shown in graph (A) for HLA-A, (B) for HLA-B and (C) for HLA-C.

4.3.2 Ethnicity and allele predictions

Allele prediction based on HLA frequencies employed data obtained from the American Caucasian population (Maiers, Gragert and Klitz, 2007). However, because allele frequencies are known to differ between ethnic groups (Dyer, 1988), the donor ethnicity was interrogated for all samples indicating prediction errors using the allele frequency algorithm. In cases where the ethnicity was other than white, HLA frequency data was further scrutinised using EpHLA converter (Sousa, Luiz Cláudio Demes da Mata *et al.*, 2011) to determine whether the allele prediction would have changed when considering the ethnic origin of the donor.

Although 13 of the 75 samples with prediction discrepancies using the allele frequency algorithm came from donors with a self declared ethnicity that was not Caucasian, only three predictions would have changed if donor ethnicity was considered (Table 4.2). Allele prediction for the other nine samples remained the same, regardless of the donor's ethnic origin.

Historical HLA result	Donor Ethnicity	Predicted allele: Caucasian frequency data [‡]	Predicted allele: Donor ethnicity frequency data [‡]	NGS result
A*29 allele string	Asian	A*29:02	A*29:01	A*29:01:01
C*07 allele string	Chinese	C*07:01	C*07:02	C*07:02:01:01
C*15 allele string	Black	C*15:02	C*15:05	C*15:05:02

Table 4.2 Impact of donor ethnicity on HLA allele prediction using population frequencies. Three occurrences where the allele prediction using HLA frequency data based on the donor's ethnicity differed to that using Caucasian frequencies [‡]frequency data obtained from (Maiers, Gragert and Klitz, 2007)

4.3.3 Haplotype frequencies and allele prediction

In order to determine if using HLA haplotype frequencies would improve the accuracy of allele prediction, the most frequent errors observed for HLA-A and HLA-B when applying the allele frequencies algorithm for prediction were re-examined. Frequency data employed was based on haplotypes determined from a north west England cohort of 298 individuals (Alfirevic *et al.*, 2012), obtained from www.allelefreqencies.net, selected because it was the largest English population with haplotype frequencies available. Each HLA type affected was reviewed and the most frequent haplotype used to predict the allele from the historical HLA strings. As when employing allele frequencies, all results were considered at second field resolution only. For the three samples typed as HLA-A*29:01 by NGS but predicted to be the more frequent allele HLA-A*29:02, no additional benefit was gained by using haplotype frequency data. Two contained presumed haplotypes that indicated the presence of HLA-A*29:02 and the third sample did not appear to contain a haplotype including HLA-A*29 that was present in the

North West England cohort. Therefore allele prediction based on haplotype frequencies was not possible for this sample (Table 4.3)

Sample ID	Historical HLA-A type	Presumed haplotype and predicted HLA-A*29 allele
6883	A*01,*29; B*44,*57; C*06,*16	A*29:02-B*44:03-C*16:01
711N	A*11,*29; B*07,*55; C*03,*07	A*29:02-B*07:02-C*03:03
0993	A*02,*29; B*07,*07; C*07,*15	N/A

Table 4.3 Prediction of HLA-A*29 alleles from presumed haplotypes. The HLA type of each sample is displayed as first field for presentation purposes only, with the presumed haplotype shown in red; prediction was based on historical medium resolution HLA typing. The predicted haplotype and respective HLA-A*29 allele for each sample were based on the most frequent haplotypes found in North West England (Alfirevic et al., 2012). N/A = no haplotype data available

Using haplotype data proved more accurate than allele frequencies alone for samples typing as HLA-B*44:03 but predicted as the more frequent HLA-B*44:02. Twelve of the seventeen samples affected (70.5%) were correctly predicted as HLA-B*44:03 when employing haplotype data. It was not possible to predict the HLA-B*44 allele for the remaining five samples as there was no appropriate haplotype present in the west English cohort (Table 4.4)

Sample ID	Historical HLA-A type north	Presumed haplotype and predicted HLA-B*44 allele
614X	A*02,*02; B*27,*44; C*01,*04	N/A
496J	A*02,*23; B*27,*44; C*02,*04	A*23:01-B*44:03-C*04:01
6891	A*02,*23; B*27,*44; C*02,*04	A*23:01-B*44:03-C*04:01
152P	A*02,*24; B*27,*44; C*02,*04	N/A
1927	A*01,*29; B*44,*57; C*06,*16	A*29:02-B*44:03-C*16:01
642O	A*01,*23; B*44,*52; C*04,*12	A*23:01-B*44:03-C*04:01
753C	A*02,*23; B*15,*44; C*04,*04	A*23:01-B*44:03-C*04:01
2201	A*02,*32; B*44,*56; C*01,*16	A*02:01-B*44:03-C*16:01
6883	A*01,*29; B*44,*57; C*06,*16	A*29:02-B*44:03-C*16:01
597T	A*26,*30; B*13,*44; C*06,*16	N/A
871X	A*23,*29; B*44,*44; C*04,*16	A*23:01-B*44:03-C*04:01 A*29:02-B*44:03-C*16:01
078E	A*02,*29; B*35,*44; C*04,*16	A*29:02-B*44:03-C*16:01
188N	A*32,*32; B*27,*44; C*01,*04	N/A
021F	A*02,*23; B*40,*44; C*03,*04	A*23:01-B*44:03-C*04:01
629B	A*02,*23; B*15,*44; C*04,*07	A*23:01-B*44:03-C*04:01
793D	A*24,*31; B*14,*44; C*04,*08	N/A
494J	A*01,*23; B*44,*57; C*04,*06	A*23:01-B*44:03-C*04:01

Table 4.4 Prediction of HLA-B*44 alleles from presumed haplotypes. The HLA type of each sample is displayed as first field for presentation purposes only, with the presumed haplotype shown in red; prediction was based on historical medium resolution HLA typing. The predicted haplotype and respective HLA-B*44 allele for each sample were based on the most frequent haplotypes found in north west England (Alfirevic et al., 2012). N/A = no haplotype data available

4.3.4 Impact of HLA allele prediction errors

For each prediction error, the number of potential epitope mismatches between the two alleles was determined, with results summarised in Figure 4.3. Of the 25 different predicted allele discrepancies identified, 10 (40%) were found not to differ at the epitope level. Of the remaining 15 mismatched allele pairs, only two mismatched alleles (both HLA-C) were found to have mismatches with antibody verified epitopes. However, one of these (HLA-C*07:02 vs. C*07:01) was the most frequently observed discrepancy, with 34 donors affected. This mismatched pair also had five other epitope mismatches, four of which were in exposed regions of the HLA molecule.

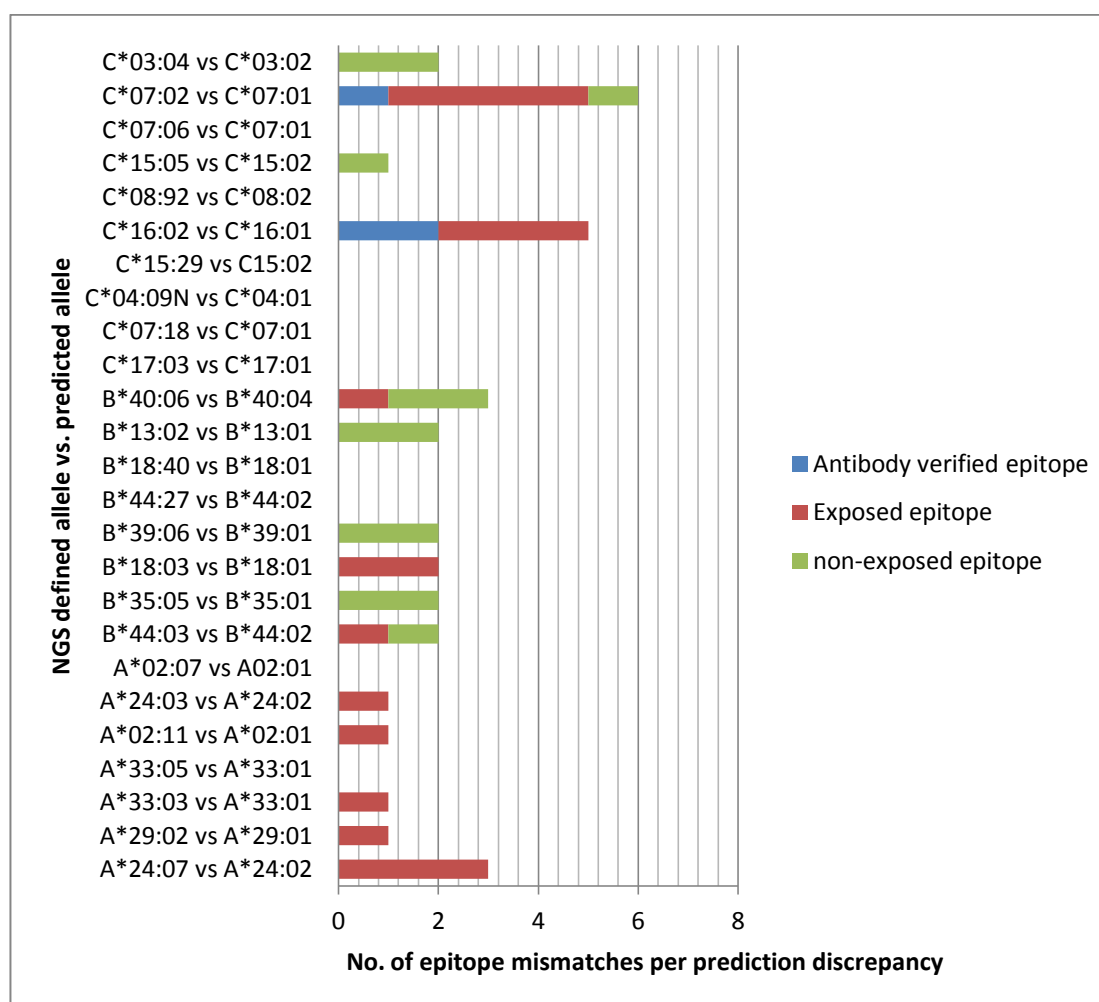


Figure 4.3 Number and type of epitope mismatches between NGS type versus predicted HLA allele. For those mismatched pairs, epitope mismatches have been categorised into antibody verified, exposed and non-exposed epitopes.

Overall, 40% of prediction errors were found to have mismatches to exposed epitopes, with five, three and two mismatches determined at HLA-A, -B and -C, respectively. Eight (32%) predicted allele discrepancies also had mismatches to inaccessible epitopes. The epitopes mismatched between each predicted allele versus NGS-defined type are listed in Table 4.5.

NGS vs. predicted allele	Antibody verified epitopes	Exposed epitopes	Inaccessible epitopes
A*02:11 vs. A*02:01		73ID	
A*24:03 vs. A*24:02		163TEW	
A*24:07 vs. A*24:02		69AQT; 70QT; 71QS	
A*29:02 vs. A*29:01		102DV	
A*33:03 vs. A*33:01		170RY	
B*13:02 vs. B*13:01			94TW; 97T
B*18:03 vs. B*18:01		72QTD; 73TDE	
B*35:05 vs. B*35:01			94TL; 97S
B*39:06 vs. B*39:01			94TW; 97T
B*40:06 vs. B*40:04		102DV	94TW; 97T
B*44:03 vs. B*44:02		156LA	156L
C*03:04 vs. C*03:02			94II; 116Y
C*07:02 vs. C*07:01	65QKR+76VS	62RK; 63EK; 65QKR; 66K	99S
C*15:05 vs. C*15:02			116F
C*16:02 vs. C*16:01	80K; 80K+14R	71ATN; 73TN; 73TVN	

Table 4.5 Mismatched epitopes for each allele prediction discrepancy observed. Epitopes for each mismatch were obtained from the ABC database located within the International Registry of Antibody-Defined HLA Epitopes (Duquesnoy *et al.*, 2013a).

4.4 Discussion

Apheresis platelet donors recruited by NHSBT are currently HLA typed using Luminex PCR-SSOP technology (Brown and Navarrete, 2011) which defines HLA to medium resolution, reporting ambiguous results in the form of allele strings (Marsh *et al.*, 2010). However, to accurately employ HEM for the selection of platelets for patients with immune platelet refractoriness, HLA needs to be defined at high resolution (Duquesnoy, 2008). Previous studies investigating HEM for platelet selection have reported using HLA allele predictions from the Luminex results as an alternative to allele level typing (Brooks, MacPherson and Fung, 2008; Pai *et al.*, 2010; Nambiar *et al.*, 2006). This chapter investigated the accuracy of HLA allele prediction when compared to typing by NGS.

4.4.1 Accuracy of allele prediction

The prediction algorithm programmed in the NHSBT LIMS system Hematos, is currently based on the first allele contained within an HLA string when an allelic type is unavailable. Results from this study demonstrate the inherent inaccuracy of this approach, producing an overall error rate of 11.1%. In 2015, NHSBT provided 19,493 doses of HLA selected platelets. Based on

this data, if HLA epitope matching had been exploited, HLA type predictions would have been incorrect for 2,163 units if the first allele of the Luminex string was used to determine the HLA type and defining epitopes. However, this error rate reduces to 4.9%, equivalent to 961 doses of HLA selected platelets, when using HLA frequencies from Caucasian populations to predict the high resolution type.

The high discrepancy rate for predicted versus NGS allele using the first allele in the string algorithm was due to the first allele being the less common HLA allele, although this varied dependant on donor ethnicity. For example, the HLA-A*29 string shown in table 4.2 contained A*29:01:01:01 as the first allele in the string but in the Caucasian population HLA-A*29:02 is more common than HLA-A*29:01, with a frequency of 3.28% compared to 0.15%, respectively (Maiers, Gragert and Klitz, 2007). This resulted in 29 donors predicted to be A*29:01 when they were actually A*29:02. Conversely, when alleles were predicted based on population frequencies, the majority of predictions were accurate. Of the three samples that did type by NGS as A*29:01, one donor self declared as Asian and would have been predicted correctly if ethnicity was considered alongside allele frequencies, as A*29:01 is more common in the Asian population than A*29:02 (Maiers, Gragert and Klitz, 2007). The other two donors were recorded as Caucasian so the prediction would have remained unchanged. However, it is worth noting that studies exploring ancestral informative markers question the reliability of self-declared ethnicity as indicators of genomic ancestry (Ramos *et al.*, 2016; Cardena *et al.*, 2013).

Even when the most frequent allele appears to be the first in the Luminex string, errors still occurred. For example, the allele B*44:02 from a HLA-B*44 string was incorrectly predicted in 17 donors using either algorithm rather than B*44:03, as determined by NGS. HLA-B*44:02 is the more common B*44 allele, being nearly twice as frequent in Caucasians as HLA-B*44:03 (Maiers, Gragert and Klitz, 2007). However, the Luminex string began with B*44:02:11 which, based on the NGS data obtained from this cohort, is not a common allele in the English population; none of the 540 donors typed as B*44:02:11 but over 100 donors sequenced as either B*44:02:01:01 or B*44:02:01:03 by NGS. A similar error also occurred with a Luminex HLA-C*07 string, where the first allele in the string was C*07:01:25 but NGS typed as C*07:02:01:01 or C*07:02:01:03, with both prediction algorithms calling the discrepant C*07:01 allele, being both the allele first in the string and the most frequent at second field (Maiers, Gragert and Klitz, 2007). If the third field result had been considered in either case, prediction based on allele frequencies would have been correct the majority of the time but unfortunately the EpHLA converter software only provided frequency data at second field (Sousa, Luiz Cláudio Demes da Mata *et al.*, 2011). The Allele Frequencies Net Database (Santos *et al.*, 2016) does contain some population data at third or fourth field but a decision not to

utilise this information was based on the populations included (none were from the UK or Northwest Europe) and or the small cohort when compared to data available from EpHLA converter, which is based on over 6000 European Americans (Maiers, Gragert and Klitz, 2007). A recent publication reporting common and well-documented HLA alleles in Europe was also consulted but this reference defaulted frequency data to second field (Sanchez-Mazas *et al.*, 2017).

The use of haplotype frequencies was investigated to determine if this was a more accurate approach for prediction of HLA alleles compared to using allele frequency data. Due to the complexity of the analysis, which was performed manually, only a small data set was examined for the purposes of this study. Consequently, investigation was restricted to the most frequent prediction errors for HLA-A and HLA-B only. Use of haplotype frequencies did not prove useful for the HLA-A*29 prediction error. This may have been by chance, as only three samples were affected and otherwise using HLA-A allele frequency data only proved reliable in most cases. However, for the HLA-B*44 prediction error example, haplotype analysis proved to be significantly more accurate than using allele frequencies, with over two thirds of prediction errors resolved.

Haplotype frequency data used was obtained from a North West English population, although it is acknowledged the cohort was small, consisting of less than 300 individual HLA types. It might have been more appropriate to use a similar data source as used for HLA allele frequencies (i.e. American Caucasians data). However, although this cohort included HLA types from over 1.2 million bone marrow donors available from www.allelefrequencies.net, there were a very limited number of haplotypes at second field compared to the North West England population (Santos *et al.*, 2016) and was therefore less informative.

Despite the limited data set examined, use of haplotype frequencies does appear to provide accurate allele prediction, certainly for HLA-B*44. This approach for allele prediction is supported by Geneugelijk *et al.* who recently reported development of PIRCHE II software, designed to estimate risk of solid organ transplants based on HLA epitopes. PIRCHE II requires high resolution HLA types to accurately analyse risk and has embedded algorithms that convert low resolution types into HLA alleles using haplotype frequencies (Geneugelijk *et al.*, 2017). However, despite the stated reliability of PIRCHE II estimations, the group acknowledged that this approach was not suitable for all donor-recipient pairs. They suggest that additional NGS-based data is required for determining HLA haplotype frequencies, particularly if sequences are extended beyond exons 2 and 3 for HLA class I alleles, leading to more reliable epitope matching (Geneugelijk *et al.*, 2017). However, use of NGS for HLA typing donors and recipients would completely negate the need for HLA allele prediction.

4.4.2 Impact of prediction errors

Twenty five HLA allele predictions were found to be incorrect, affecting 209 HLA types at HLA-A, -B or -C when using the first in the HLA string algorithm, compared to 75 HLA types when alleles were determined from HLA frequency data.

Following interrogation of the International Registry of Antibody-Defined HLA epitopes (Duquesnoy *et al.*, 2013b), forty percent of the allele prediction errors did not translate to epitope mismatches. A proportion of these were found to have mismatches within HLA 'G groups', such as HLA-C*07:01, -C*07:06 and -C*07:18, where nucleotide sequences are identical in exons 2 and 3 of the respective gene that encodes the peptide binding domain of the HLA molecule (Marsh *et al.*, 2010). Others lacked epitope data for a particular HLA allele, including HLA-B*18:40 and HLA-C*08:92 and in these cases it is possible epitope mismatches were present but have not yet been defined. Although the HLA Epitope Registry is regularly updated, its curators admit that data is incomplete and acknowledge the need for more studies (Duquesnoy *et al.*, 2016). The two HLA-A prediction errors, namely A*02:07 vs. A02:01 and A*33:05 vs. A*33:01 appear to have identical epitopes, despite amino acid differences in the $\alpha 1$ and $\alpha 2$ domain protein sequence (Robinson *et al.*, 2015). The one other example of zero epitope mismatches was due to the presence of HLA-C*04:09N instead of the predicted HLA-C*04:01 allele. As a null allele, HLA-C*04:09N would result in lack of expression of the HLA molecule on the cell surface so any donor specific antibodies present would not give cause for concern if a donor with this type was transfused into a patient with antibodies to epitopes expressed on HLA-C*04:01 (Wang *et al.*, 2002).

Sixty percent of prediction errors would have resulted in epitope mismatches if they had been subsequently used to select platelets for transfusion. The majority of epitope mismatches identified were against exposed regions of the HLA molecule and are potentially clinically relevant being in antibody-accessible regions, although only three were antibody verified epitopes (Duquesnoy, 2014). The most frequently observed example of epitope mismatching using either algorithm resulted from the prediction of HLA-C*07:01 instead of -C*07:02, affecting 34 donors (6.7% of the cohort). HLA-C is not generally currently considered when using HLA match grading to select platelets for refractory patients as the relevance of HLA-C antibodies in platelet refractoriness is still undetermined (Stanworth *et al.*, 2015). However, two of the six epitopes mismatched between HLA-C*07:01 and -C*07:02, namely 62EK and 66K, are also present on common HLA-A antigens such as HLA-A2 ((Duquesnoy, 2014). It is therefore important to consider all HLA class I compatibility when using HEM (Duquesnoy, 2017). Even if HLA-C antigens prove less relevant in platelet transfusion, antibodies to HLA-C

epitopes may still cross react with those expressed by HLA-A and or HLA-B and thus could be clinically relevant. Indeed, Lomago and colleagues reported a case of an HLA-C*07:04 induced alloantibody cross reacting with shared epitope 156DA expressed by HLA-B*44:02 in a renal transplant recipient typing as HLA-B*44:03 who had never been exposed to HLA-B*44:02 leading to an episode of acute cellular rejection (Lomago *et al.*, 2010). Others have similarly postulated antibody mediated rejection can result from shared epitopes with the immunising antigen (Mongkolsuk *et al.*, 2014).

The most frequent HLA-B prediction error affected 17 donors, with two epitope mismatches between B*44:02 and the NGS defined allele B*44:03, although only one of the two mismatches was against exposed epitopes. For HLA-A, the most frequent discrepancy was between A*29:02 and A*29:01, with just one exposed epitope mismatch between the two alleles. The second most common HLA-C prediction error, with C*03:02 predicted instead of C*03:04, occurred in 30 donors using the first allele in the string algorithm. However, the two epitope mismatches between these alleles appear inaccessible to antibody. The clinical relevance of such non-exposed epitopes is questionable, but it has been suggested that because they reside in the peptide binding groove these polymorphic positions may affect the conformation of adjacent residues leading to the creation of antibody-recognisable epitopes (Duquesnoy *et al.*, 2014). However it is not clear whether such positions would be immunogenic (Duquesnoy, 2014).

5. HLA frequencies and population analysis

5.1 Introduction

As discussed in Chapter 4, there is a lack of HLA allele and haplotype frequency data available beyond second field resolution in the English population. Limited frequency data can prevent accurate allele prediction from HLA types defined at a lower resolution (Geneugelijk *et al.*, 2017). In addition, the availability of HLA frequencies at the allele level will provide useful data to the wider community, including the investigation of HLA allele associations with adverse drug reactions and in anthropology studies (Gonzalez-Galarza *et al.*, 2015b). HLA frequency data can also provide estimates for the likelihood of finding an HLA compatible donor in the transplantation and transfusion settings (Pingel *et al.*, 2013).

NGS has the capability of HLA typing to the allele level (Erlich, 2015) and a protocol was developed and used to type 540 English apheresis platelet donors, as described and discussed in Chapter 3. The aim of this part of the study was to determine the HLA allele and haplotype frequencies of the platelet donor cohort based on HLA types obtained by NGS. Results from this analysis would then be submitted to an international database that stores allele frequencies from polymorphic areas in the human genome (Gonzalez-Galarza *et al.*, 2015b).

However, as the initial sequencing only produced complete HLA-A, -B and -C genotypes for 441 (81.7%) donors, a decision was taken to re-sequence as many samples as possible that had previously failed for one or more HLA locus. This was to ensure that the maximum number of samples was included in the frequency data analysis. In addition, because some of the rejected sequences described in Chapter 3 were associated with particular alleles or allele groups, resolution was necessary to avoid skewing frequency analysis by excluding particular allele groups with poor sequence data. All sequences containing putative mutations in the first round of NGS typing were also repeated to confirm the presence of novel sequences.

5.2 Materials & Methods

HLA Typing was performed on DNA obtained from 540 apheresis platelet donors, as described in Chapter 3. Collected at English blood donor centres, the individuals in this cohort were of fixed ethnicity, with the majority (91.1%) of donors self declaring as white (Figure 5.1). In order to produce sufficient data for determining valid HLA allele and haplotype frequencies in this cohort, samples that originally failed to type for one or more HLA locus and those containing putative novel mutations were re-sequenced.

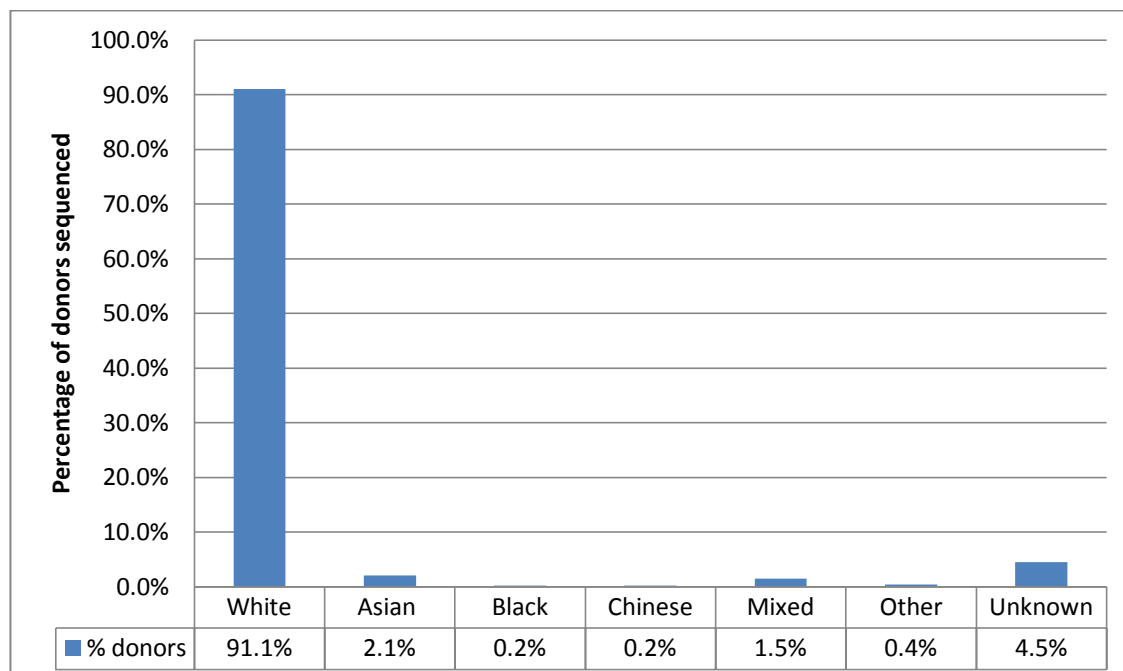


Figure 5.1 Self-declared ethnicity of the 540 apheresis platelet donors HLA typed in this study.

5.2.1 Confirmatory sequencing of samples with putative novel mutations

5.2.1.1 Sample selection

From the original results described in Chapter 3, a total of 45 samples were identified as containing putative novel mutations in one or more HLA sequence that was otherwise valid. Fifty microlitres of each DNA was transferred from their original plates into a clean 96-well plate (#RPT1) using a bespoke cherry-picking program designed for the epMotion® 5075 liquid handling robot (Eppendorf).

5.2.1.2 DNA preparation, PCR amplification and amplicon pooling

DNA was purified manually using AMPure beads as previously described. However, due to the low DNA concentrations observed following the original DNA purification, it was decided not to normalise the DNA but use it neat. PCR amplifications for HLA-A, -B and -C were then set up for all 45 samples using the in-house PCR protocol, as described in 2.6.1.2. Success of amplification was determined by agarose gel electrophoresis (section 2.6.2.1), followed by pooling of 7µl of each amplicon into respective wells of a clean plate.

5.2.1.3 Library preparation and sequencing

Pooled amplicon was prepared for sequencing using the NGSgo® protocol as described in 2.6.3.2. The PAL was assessed using both the Qubit and Bioanalyser and subsequently denatured and diluted before loading onto the MiSeq and sequencing using 2 x 251 chemistry (section 2.6.6).

5.2.2 Re-sequencing of samples with HLA types previously rejected

5.2.2.1 Sample selection

Eighty nine samples that previously failed to sequence for one or more HLA locus had sufficient DNA remaining for re-sequencing. Fifty microlitres of each DNA sample was transferred from their original plates into a clean 96-well plate (#RPT2) using the epMotion[®] 5075 liquid handling robot as described in 5.2.1.1.

5.2.2.2 DNA preparation, PCR amplification and amplicon pooling

DNA was purified on the Biomek FX liquid handling robot using AMPure beads as previously described and neat purified DNA was used to set up PCR amplifications for HLA-A, -B and -C using in-house PCR. Success of amplification was determined using agarose gel electrophoresis and was followed by the pooling of 7µl of amplicon from each HLA locus into respective wells of a clean plate.

5.2.2.3 Library preparation and sequencing

Pooled amplicon was prepared for sequencing using the TruSightHLA protocol, as described in 2.6.3.1. This approach was selected because, at the time this experiment was performed, the Illumina library preparation kits were in routine use at NHSBT. The TruSightHLA protocol was initiated on a Biomek FX robot but due to a power interruption during the amplicon normalisation step, the remainder of the method was performed manually. To ensure that library preparation had not been adversely affected by failure of the robotics, the PAL was assessed using both the Qubit and Bioanalyser. Following denaturation and dilution of the PAL, sequencing was performed on the MiSeq as before but this time using 2 x 151 paired end sequencing on a standard flow cell, which was cheaper and quicker than the 2 x 251 chemistry used previously and had been shown by the routine laboratory to provide acceptable results.

5.2.3 Sequence data analysis

FASTQ files generated by MSR were analysed using NGSengine v2.1, IMGT/HLA 3.24. Samples with remaining intron or exon mismatches were subsequently reanalysed following the later release of NGSenginev2.4 in January 2017, to check against sequences from IMGT/HLA v3.26. Analysis preferences were as described in section 2.6.7 and success of sequencing was determined using the acceptance criteria discussed in Chapter 3.

5.2.4 Population data analysis

5.2.4.1 HLA allele frequencies

Allele frequencies for HLA-A, -B and -C were calculated from data obtained from 519 donors sequenced. Frequencies were determined at third field only and calculated by direct counting.

5.2.4.2 HLA haplotype frequencies

Arlequin software v3.5, an integrated software package for population genetics analysis (Excoffier and Lischer, 2010), was employed to determine both HLA haplotype frequencies and calculate Hardy-Weinberg equilibrium (HWE). Haplotype frequencies were estimated using the expectation-maximum algorithm and the exact test using the Markov chain method (Guo and Thompson, 1992) was used for HWE, as recommended in the Arlequin35 user guide. HLA allele frequencies were also subsequently confirmed using Arlequin.

5.3 Results

5.3.1 Repeat amplifications

Agarose gel electrophoresis indicating successful amplification was achieved for the majority of samples, based on an aliquot of each amplicon taken from the first and last rows of each repeat plate. Although the molecular weight marker was mistakenly omitted from the gel, all visible bands were of the same size and intensity indicating a successful amplification for all but one sample (Figure 5.2).

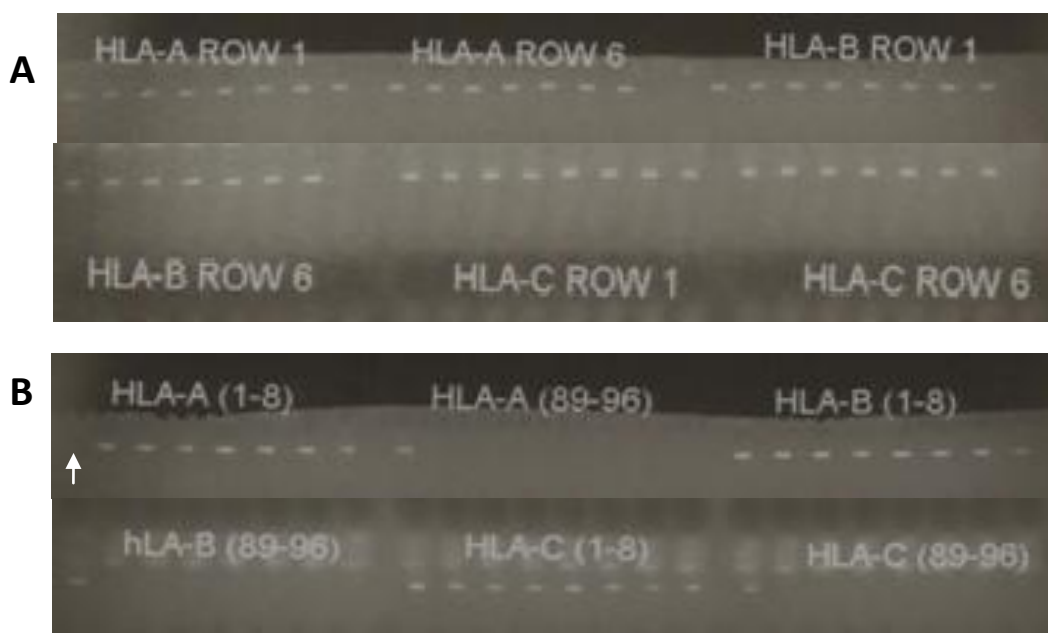


Figure 5.2 Gel images of amplicons taken from the first and last row of each repeat plate. Image (A) shows amplicons from plate #RPT1 taken from row 1 and row 6 of each plate for HLA-A, -B and -C. (B) shows amplicons from plate #RPT2, taken from row 1 (wells 1 to 8) and row 12 (rows 89-96) of each plate for HLA-A, -B and -C. Note that for each row 12, a PCR product was only expected in lane 89. All wells contain a visible band of the same size, except for sample taken from well 1 of the HLA-A second repeat plate (white arrow) where no band was present. Note that no molecular weight markers were run.

5.3.2 Sequencing libraries

Bioanalyser results indicated that each library was an acceptable size, with an average of 1781bp and 1323bp obtained for plates #RPT1 and #RPT2, respectively. The first plate #RPT1 had a slightly higher concentration of 3.7ng/μl when compared to 2.7ng/μl obtained for plate #RPT2, with both values determined using the Qubit. The cluster density achieved on the flow cell was similar for each MiSeq run, with 1251 K/mm² and 1361 K/mm² for #RPT1 and #RPT2, respectively. The percentage of clusters passing the filter was 83% for both sequencing experiments.

5.3.3 HLA sequencing results

5.3.3.1 Confirmatory sequences of samples with putative novel mutations

All 45 samples with previously identified novel sequences were successfully retyped for HLA-A, -B and -C. The average read depth was 1985bp (range 225-4391bp) with a mean insert size of 419bp (range 307-563bp) with an average of 90% of sequences mapping to reference data. Nine of the 45 samples were subsequently resolved for all HLA loci due to the designation of new HLA alleles since the original sequencing was performed, detailed in Table 5.1.

ID	Original closest matched allele	Position of mutation	New allele designation
S30	A*02:01:01:01	2798	A*02:01:01:08
S10	A*31:01:02:01	2486	A*31:01:02:04
S39	A*31:01:02:01	2486	A*31:01:02:04
S17	B*08:01:01	2803	B*08:01:01:02
S36	B*08:01:01	2803	B*08:01:01:02
S41	B*15:01:01:01	2324	B*15:01:01:06
S24	B*56:01:01:02	2988	B*56:01:01:03
S23	B*56:01:01:02	2988	B*56:01:01:03
S12	C*02:02:02:01	3005	C*02:02:02:03

Table 5.1 HLA alleles resolved following re-sequencing. HLA sequences containing nucleotide mismatches at the positions indicated that were subsequently resolved following re-sequencing and analysis with NGSengine v2.4.

Of the remaining 36 samples, mutations were confirmed in 38 HLA sequences with three identified in exons, 20 in the intronic regions and 15 mutations located in the UTRs, summarised in Table 5.2.

Locus	Substitution	Deletion	Insertion
HLA-A	12	1	0
HLA-B	10	0	5
HLA-C	10	0	0

Table 5.2 Number of HLA sequences observed with novel mutations. Samples are shown according to HLA locus and the category of mutation observed.

Twenty seven of these 38 mutations were confirmed in fully phased sequences (Table 5.3) with the remaining sequences failing to separate completely into two alleles for the length of the respective genes due to poor phasing. Lack of sequence phase across an entire gene was generally observed when there were large regions of homozygous sequence between the defining heterozygous positions, as illustrated in Figure 5.3.

ID	Affected allele	Nucleotide position of mutation	Gene Location	Expected nucleotide	Observed nucleotide	Mutation
S6	A*01:01:01:01	1049	INT 3	C	A	Substitution
S5	A*02:01:01:01	2949	UTR	C	T	Substitution
S8S2	A*02:05:01	657/670	INT 2	A/G	G/T	Substitution
S7	A*11:01:01:01	-116	UTR	A	G	Substitution
S14	A*23:01:01	multiple	INT 2	MULTIPLE	-	Deletion
S43	A*23:01:01	-219	UTR	T	C	Substitution
S22	A*24:02:01:01	-17	UTR	C	T	Substitution
S22	A*29:02:01:01	2338	INT 5	C	A	Substitution
S16	A*30:02:01:02	2922	UTR	G	C	Substitution
S31	A*68:02:01:01	2263/2266/2268	INT 5	T/A/C	G/T/T	Substitution
S26	B*07:02:01	1426	INT 3	C	T	Substitution
S21	B*08:01:01:01	1051	INT 3	C	T	Substitution
S11	B*18:01:01:02	3010.2	UTR	-	C	Insertion
S28	B*44:02:01:01	-31	UTR	A/A	C/C	Substitution
S20	B*51:01:01:01	666	INT 2	C	T	Substitution
S45	B*51:01:01:01	666	INT 2	C	T	Substitution
S18	B*51:01:01:01	2978	UTR	T	C	Substitution
S3	B*51:01:01:01	2978	UTR	T	C	Substitution
S33	B*51:01:01:01	2289	INT 5	C	G	Substitution
S34	C*03:03:01:01	2568	EX 6	G	A	Substitution
S29	C*03:04:01:01	1810	EX 4	T	A	Substitution
S46	C*05:01:01:02	1505	INT 3	C	A	Substitution
S9	C*05:01:01:02	3057	UTR	C	T	Substitution
S32	C*06:02:01:01	2426	INT 5	G	T	Substitution
S13	C*07:02:01:01	-44	UTR	T	C	Substitution
S25	C*07:02:01:03	189	INT 1	C	T	Substitution
S37	C*08:02:01:01	1117	INT 3	G	C	Substitution

Table 5.3 Novel HLA alleles with fully phased sequences. Details of the mutations observed for each fully phased novel allele detected, including gene location, change to nucleotide sequence and mutation category.

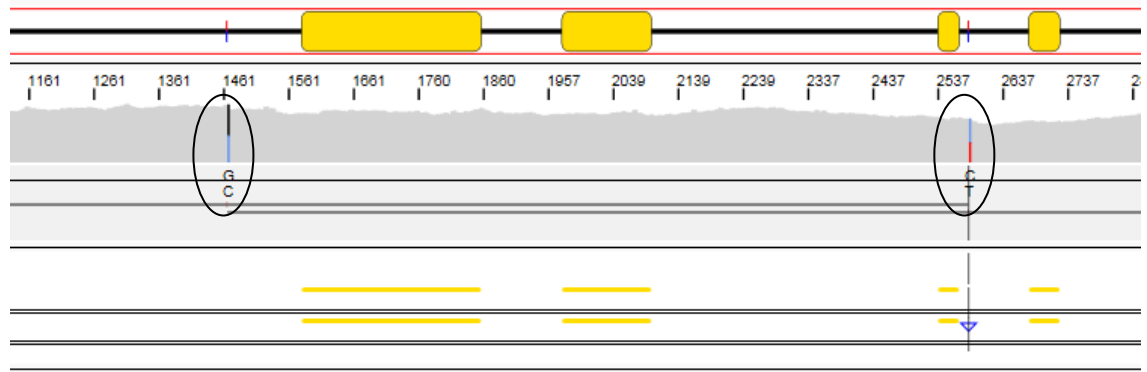


Figure 5.3 Example of sequence containing a novel mutation with lack of phase. The novel mutation is marked with the blue triangle but, due to large homozygous regions between the heterozygous positions (circled in black), this sequence is not fully phased across the gene. Lack of phase is illustrated by the parallel grey lines between the heterozygous positions.

5.3.3.2 Impact of Exon mutations

Of the three sequences identified with novel exon mutations, two would result in changes to amino acid sequence when compared to the reference sequence, with the third being a synonymous substitution (Table 5.4).

Allele	Substitution	Position	Exon	Codon change	Amino acid change
B*15:10:01	G>A	42	1	TCG>TCA	Ser>Ser
C*03:03:01:01	G>A	2568	6	GCG>ACG	Ala>Thr
C*03:04:01:01	T>A	1810	4	TGC>AGC	Cys>Ser

Table 5.4 Effect of nucleotide substitutions for the three novel exon mutations identified.

5.3.3.3 Re-sequencing samples with HLA types previously rejected

A total of 239 of the 267 sequences performed met the acceptance criteria, resulting in 74 (89.5%) samples producing a valid allele level type for HLA-A, -B and -C with the remaining 15 samples having sequences rejected for one or more HLA locus (Figure 5.4). The majority of sequences in plate #RPT2 were rejected due to lack of phase, with an average of 12 phased regions per sequence, resulting in ambiguous allele assignment. Many of the accepted sequences were also poorly phased, with six phased regions per sequence on average (Figure 5.5), despite a mean coverage depth of 1080bp.

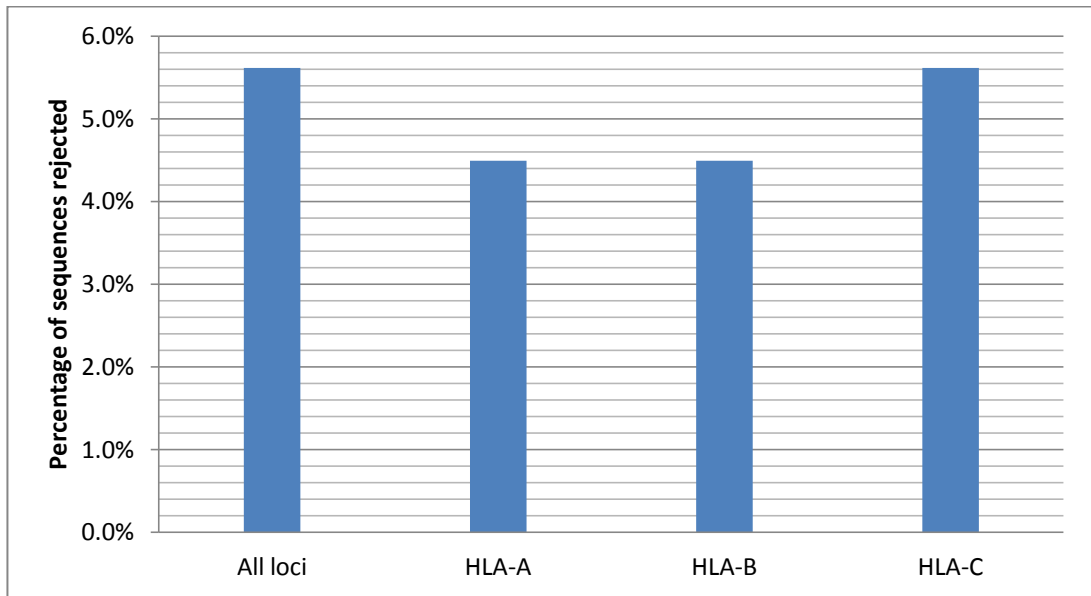


Figure 5.4 The percentage of sequences rejected for each HLA locus in plate #RPT2. Results are presented according to HLA locus.

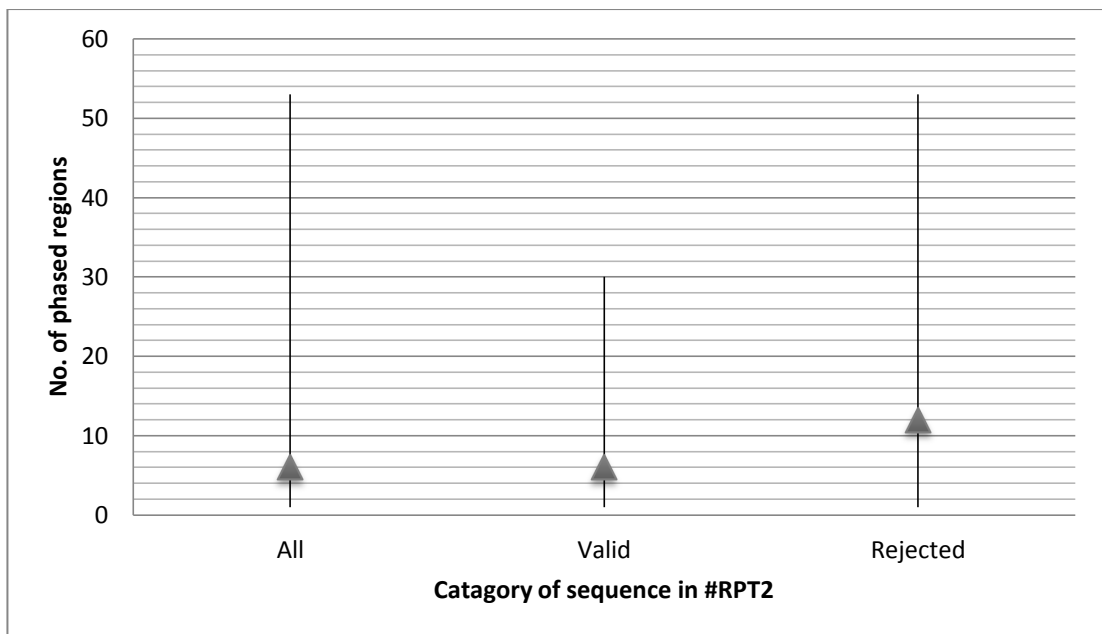


Figure 5.5. The number of phased regions observed for sequences in plate #RPT2. Results are presented according to quality of sequence, with the mean value represented by the triangle

Sequences that were rejected consisted of smaller fragments, although the mean insert size was not markedly different from the valid sequences, with a mean of 345bp and 372bp, respectively. Of note was the uneven coverage observed in sequences from plate #RPT2 (Figure 5.6), suggesting that sequences were either missing or they had failed to map correctly in regions with low sequencing depth. As the average mappability of sequences in both

accepted and rejected sequences were similar (82.8% and 80.6%, respectively), it suggested that sequences in the regions of low coverage had not been generated.

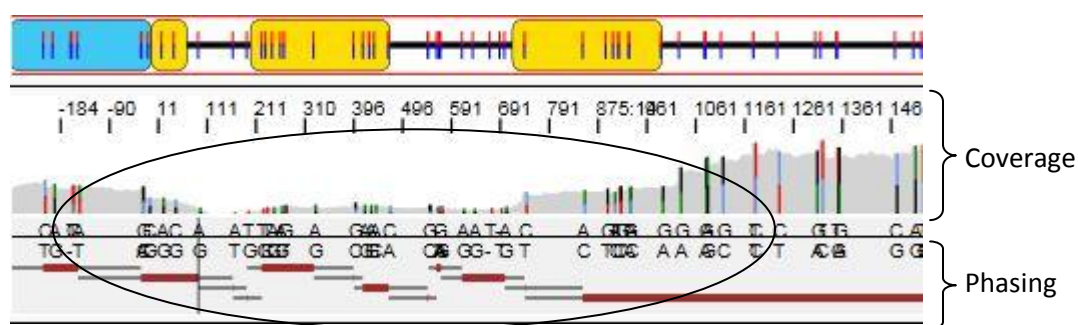


Figure 5.6 An example of uneven coverage and the resulting poor phasing across the gene. Uneven coverage observed in a sample from sequencing run #RPT2. Phased regions are indicated by the solid red line, with grey lines in between representing regions with incomplete phasing, particularly noticeable in regions of low coverage.

5.3.4 Population data analysis

From the starting cohort of 540 donors, a total of 519 samples produced valid sequences for HLA-A, -B and -C, either during the original experiments described in Chapter 3 or in the re-sequence data detailed above. Data from 21 samples were excluded from population analysis due to incomplete HLA sequence information for reasons summarised in Table 5.5.

Reasons for exclusion	No. of samples
Insufficient DNA for repeat	7
All HLA loci failed	3
HLA-A failed	1
HLA-B failed	6
HLA-C failed	4

Table 5.5 Reasons for exclusion of donors from the population data analysis.

Due to the high number of sequences with one or more mismatches to the reference intron data, allele frequencies were determined at third field only, listed in Table 5.6.

HLA-A type	Freq	% individuals	HLA-B type	Frequency	% individuals	HLA-C type	Frequency	% individuals
A*01:01:01	0.2062	37.38%	B*07:02:01	0.1329	25.43%	C*01:02:01	0.0250	5.08%
A*02:01:01	0.3035	50.48%	B*07:05:01	0.0010	0.19%	C*02:02:02	0.0424	8.59%
A*02:03:01	0.0010	0.19%	B*08:01:01	0.1339	25.63%	C*02:10	0.0010	0.20%
A*02:05:01	0.0067	1.35%	B*13:02:01	0.0231	4.62%	C*03:02:02	0.0019	0.39%
A*02:06:01	0.0029	0.58%	B*14:01:01	0.0116	2.31%	C*03:03:01	0.0655	13.28%
A*02:07:01	0.0010	0.19%	B*14:02:01	0.0260	5.01%	C*03:04:01	0.0838	15.82%
A*02:11:01	0.0039	0.77%	B*15:01:01	0.0684	13.49%	C*04:01:01	0.0857	16.80%
A*02:13	0.0010	0.19%	B*15:02:01	0.0010	0.19%	C*04:09N	0.0019	0.39%
A*03:01:01	0.1320	24.86%	B*15:03:01	0.0010	0.19%	C*05:01:01	0.1060	21.09%
A*03:02:01	0.0010	0.19%	B*15:16:01	0.0010	0.19%	C*06:02:01	0.1002	18.55%
A*11:01:01	0.0578	11.37%	B*15:17:01	0.0058	1.16%	C*07:01:01	0.1618	30.86%
A*23:01:01	0.0231	4.62%	B*15:39:01	0.0019	0.39%	C*07:01:02	0.0048	0.98%
A*24:02:01	0.0645	12.33%	B*18:01:01	0.0270	5.39%	C*07:02:01	0.1484	28.13%
A*24:03:01	0.0010	0.19%	B*18:03:01	0.0010	0.19%	C*07:04:01	0.0125	2.54%
A*24:07:01	0.0010	0.19%	B*18:40	0.0010	0.19%	C*07:06	0.0029	0.59%
A*25:01:01	0.0077	1.54%	B*27:02:01	0.0010	0.19%	C*07:18	0.0067	1.37%
A*26:01:01	0.0193	3.85%	B*27:03	0.0010	0.19%	C*08:01:01	0.0010	0.20%
A*29:01:01	0.0010	0.19%	B*27:05:02	0.0376	7.32%	C*08:02:01	0.0356	6.45%
A*29:02:01	0.0308	5.97%	B*27:05:04	0.0019	0.39%	C*08:92	0.0010	0.20%
A*30:01:01	0.0116	2.31%	B*27:10	0.0010	0.19%	C*12:02:02	0.0106	2.15%
A*30:02:01	0.0067	1.35%	B*35:01:01	0.0491	9.44%	C*12:03:01	0.0202	4.10%
A*30:04:01	0.0019	0.39%	B*35:02:01	0.0019	0.39%	C*14:02:01	0.0087	1.76%
A*31:01:02	0.0299	5.97%	B*35:03:01	0.0048	0.96%	C*15:02:01	0.0270	5.27%
A*32:01:01	0.0356	6.74%	B*35:05:01	0.0010	0.19%	C*15:05:02	0.0010	0.20%
A*33:01:01	0.0048	0.96%	B*35:08:01	0.0029	0.58%	C*15:29	0.0010	0.20%
A*33:03:01	0.0029	0.58%	B*37:01:01	0.0212	4.05%	C*16:01:01	0.0356	7.23%
A*33:05	0.0010	0.19%	B*38:01:01	0.0087	1.73%	C*16:02:01	0.0010	0.20%
A*66:01:01	0.0029	0.58%	B*39:01:01	0.0077	1.54%	C*16:04:01	0.0019	0.39%
A*68:01:01	0.0077	1.54%	B*39:06:02	0.0087	1.73%	C*17:01:01	0.0019	0.39%
A*68:01:02	0.0193	3.47%	B*40:01:02	0.0578	10.60%	C*17:03	0.0029	0.59%
A*68:02:01	0.0087	1.73%	B*40:02:01	0.0135	2.70%		1.0000	
A*69:01	0.0010	0.19%	B*40:06:01	0.0029	0.58%			
A*74:01:01	0.0010	0.19%	B*41:01:01	0.0019	0.39%			
	1.0000		B*41:02:01	0.0029	0.58%			
			B*44:02:01	0.1098	21.39%			
			B*44:03:01	0.0549	10.79%			
			B*44:03:02	0.0029	0.58%			
			B*44:05:01	0.0029	0.58%			
			B*44:27:01	0.0019	0.39%			
			B*45:01:01	0.0048	0.96%			
			B*46:01:01	0.0010	0.19%			
			B*47:01:01	0.0010	0.19%			
				0.0145	2.89%			
			B*49:01:01					
				0.0039	0.77%			

B*50:01:01		
B*51:01:01	0.0424	8.29%
B*52:01:01	0.0096	1.93%
B*52:01:02	0.0010	0.19%
B*53:01:01	0.0039	0.77%
B*55:01:01	0.0250	5.01%
B*56:01:01	0.0019	0.39%
B*57:01:01	0.0462	8.86%
B*58:01:01	0.0087	1.73%
	1.0000	

Table 5.6 HLA-A, -B and -C allele frequencies in English platelet donors. The HLA allele frequencies and the % of individuals containing each allele in a cohort of 519 samples collected from English apheresis platelet donors.

A comparison of the three most frequent allele groups from each HLA locus showed similar frequencies observed in 400 healthy volunteers from North West England (Alfirevic *et al.*, 2012). However, this paper did not report unambiguous allelic level HLA types, so some alleles and their respective frequencies were grouped together according to their shorthand 'G' codes e.g. HLA-C*07:01:01, C*07:01:02, C*07:06 and C*07:18 as C*07:01g, which puts together HLA alleles with identical nucleotide sequences across exons 2 and 3 (Marsh *et al.*, 2010), for comparison purposes (Figure 5.7)

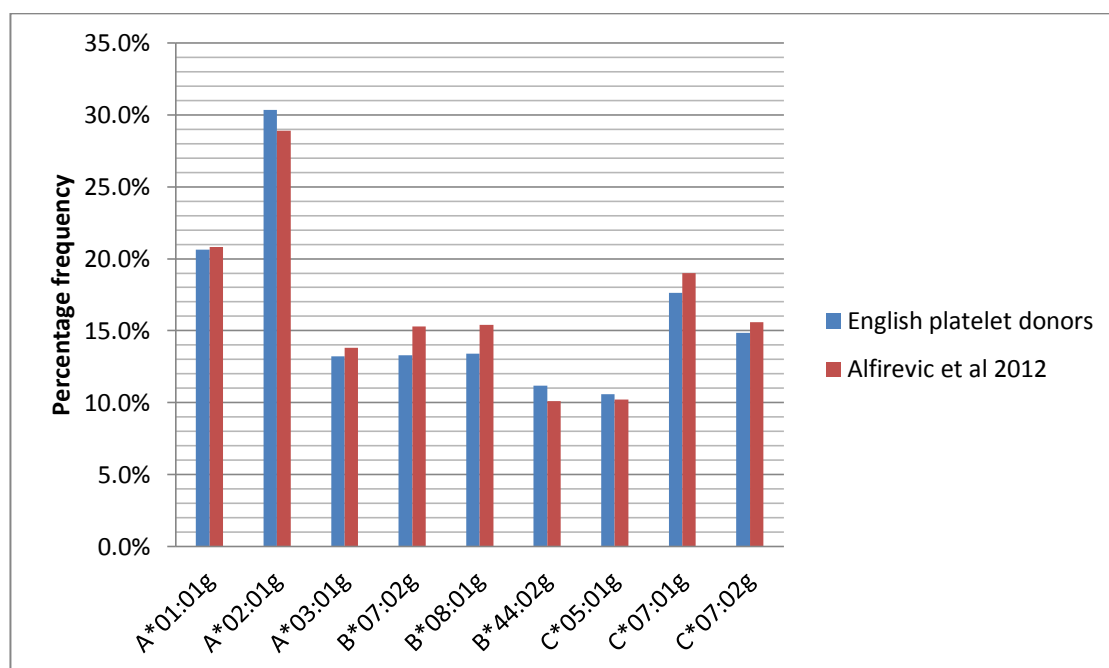


Figure 5.7 A comparison of the three most frequent HLA allele groups. Data is presented per locus by percentage frequency in English platelet donors with HLA frequencies reported by Alfirvic *et al.*, 2012 in 400 healthy volunteers from North West England.

Haplotype frequency estimation determined a total of 1423 haplotypes, with 281 having a frequency of >0.05% in this population. The three most frequent haplotypes observed were HLA-A*01:01:01-B*08:01:01-C*07:01:01 (11.2%), HLA-A*02:01:01-B*44:02:01-C*05:01:01 (6.6%) and HLA-A*03:01:01-B*07:01:01-C*07:02:01 (5.8%). When performed at third field resolution, HLA-A, -B and -C genotypes deviated from the expected Hardy-Weinberg equilibrium. However, when the typing resolution was reduced to the antigenic level, no significant deviation from HWE was observed, with the exception of HLA-B where the deviation is borderline (Table 5.7).

HLA locus	HWE – third field (p-value)	HWE – first field (p-value)	HWE – antigen level (p-value)
HLA-A	0.00071	0.25938	0.56895
HLA-B	0.00369	0.01432	0.04535
HLA-C	0.02545	0.07630	0.07129

Table 5.7 P-values observed for deviation from Hardy-Weinberg equilibrium (HWE). Results for HWE calculation are shown following analysis of HLA types at third field, first field and antigenic level of resolution

5.4 Discussion

Following the re-sequencing of samples that originally failed to produce an allele level type for HLA-A, -B and -C, the HLA allele and haplotype frequencies were determined for the platelet donor cohort HLA typed by NGS in this study.

5.4.1 Outcome of re-sequencing

The majority of samples were successfully re-amplified for HLA-A, -B and -C using the in-house PCR protocol, although because only a sample of amplicons were checked, it was not possible to determine the exact percentage of amplification success. Examining the first and last rows of each amplified plate of DNA by gel electrophoresis indicated that just one out of seventy five (1.3%) amplifications had failed, highlighting the robustness of this approach.

The subsequent library preparation and sequencing was equally successful for the first plate, #RPT1, with all samples producing a sequence of acceptable quality. Unfortunately this was not the case with plate #RPT2, which generated sequences with uneven coverage resulting in poor phasing and low coverage across parts of the gene. This poor quality data was probably due to the failure of automation during the normalisation stage of the library preparation which is likely to have affected the amount of amplicon added to the subsequent tagmentation

step. Adding insufficient template into a tagmentation reaction is known to result in smaller fragments which can lead to coverage drop out (Illumina, 2015). Another possible explanation for the difference is that plate #RPT2 used 2 x 150bp chemistry compared to the 2 x 250 cycle kits employed for all other experiments. However this is not the experience observed by others (Profaiser *et al.*, 2015) or in the routine NHSBT laboratory, where HLA typing results are comparable using either MiSeq chemistry.

5.4.2 Confirmation of novel mutations

Nine of the 45 putative novel mutations confirmed by re-sequencing were found to have designated HLA alleles following re-analysis using later releases of NGSengine and IMGT/HLA. The number of 'novel' sequences subsequently resolved by re-analysing with up to date software, detailed both here and in Chapter 3, is noticeable and can be attributed to the rapid rise of novel full length class I sequences added to the IMGT/HLA database in recent years. This is largely due to the increasing use of NGS for HLA typing (Robinson *et al.*, 2015), enabling characterisation of regions outside exons 2 and 3 of the HLA class I genes that were not previously sequenced by the majority of laboratories using conventional technology (René, Lozano and Eliaou, 2016).

Of the remaining samples, 27 of the 38 novel sequences determined were fully phased and should meet the strict criteria for submission to the IMGT/HLA database once accession numbers have been obtained for each unique sequence, which will be the subject of future work. In order for the ten sequences not 'in-phase' to be assigned a name by the WHO HLA nomenclature committee, samples will need re-sequencing using an alternative approach that is able to produce full length sequence across the HLA class I gene without the need for assembly of shorter reads, as phasing of heterozygous positions is required for IMGT/HLA submission (Lind *et al.*, 2013). This would require the use of an NGS platform such as the PacBio RS II, which has been applied for novel allele confirmation by others (Cereb *et al.*, 2015), but this technology is currently not available at NHSBT.

Of those samples containing novel mutations in exon regions of the HLA class I genes, two of the three identified would result in changes to the amino acid sequence of their respective HLA-C molecules. Whilst the relevance of polymorphism outside exons 2 and 3 is currently unknown (Lamb *et al.*, 2015), of particular interest was the mutation at position 1810 in exon 4 of HLA-C*03:04:01:01 which produces an amino acid substitution of cysteine to serine in the alpha 3 domain of the expressed HLA-C molecule. A recent paper suggested that mutations in the alpha 3 domain of an HLA molecule may affect the binding of CD8 molecules, required by cytotoxic T cells to exert their effect (Zaimoku *et al.*, 2017).

The relevance of the synonymous substitution detected in an HLA-B*15:10:01 allele, as well as those mutations identified outside the coding regions in other samples, is also unclear. Whilst it is possible that these may impact on mRNA expression, very little is known about the effect that HLA polymorphism has on the splicing mechanism because the majority of HLA alleles are identified based on DNA sequence analysis only (Voorter *et al.*, 2016).

5.4.3 Population data analysis

Due to the high number of samples with either novel mutations or artefacts contained within the intronic regions of each gene, it was decided to analyse HLA frequencies at third field to avoid having to exclude a significant proportion of data generated. Nevertheless, it is believed that this is the first reported UK population cohort with HLA typing at a resolution beyond second field. It has been submitted to Allele Frequencies Net database (Gonzalez-Galarza *et al.*, 2015a) under the population name 'England Blood Donors of Mixed Ethnicity', identifier 3392. This data is classed as 'gold standard' with allele frequencies adding up to 1, HLA typing determined to at least second field and a sample size of ≥ 50 (Santos *et al.*, 2016).

HLA allele frequencies were comparable with previous published English data. For example, the genotype frequency of HLA-A*01:01g reported by Alfirevic *et al.* 2012 in 400 volunteer donors was 38.6%, similar to the 37.4% of platelet donors typing as HLA-A*01:01:01 in this study. HLA-A, -B and -C genotypes determined at third field deviated from HWE. This is likely due to the presence of alleles such as A*74:01:01, B*18:40 and C*15:29, that occurred only once in this cohort, resulting in excess heterozygosity which is a known factor causing HWE deviation (Single *et al.*, 2002). Many of the low frequency alleles observed are neither common nor well documented in European populations (Sanchez-Mazas *et al.*, 2017) and may reflect the mixed ethnicity of English platelet donors which is typical of the blood donor population (Lattimore, Wickenden and Brailsford, 2015). However, this is not reflective of the more diverse ethnic distribution of the region, with only 84.9% of individuals in England (and Wales) self declaring as white in 2011 (Office for National Statistics, 2012) compared to 91.1% of platelet donors in this study.

As reported by others, no significant deviation from HWE was observed when HLA genotypes were defaulted to antigen level (Schafer *et al.*, 2016), apart from HLA-B which still deviated significantly. This was probably due to greater variability observed for HLA-B when compared to HLA-A and -C (Single *et al.*, 2002), with 22, 11 and 13 antigens, respectively. Population data reported in this chapter has formed the basis of a short population report (Davey *et al.*, 2017) which provides a structured description of populations along with genetic data and restricted analysis for HLA, KIR, cytokine and MIC genes (Mack and Middleton, 2015).

6. HPA genotyping by NGS

6.1 Introduction

Human platelet antigens (HPA) are located on glycoproteins expressed on the surface of platelets (Curtis and McFarland, 2014). There are currently twenty nine HPA systems described, encoded by the six genes *ITGB3*, *ITGA2B*, *ITGA2*, *GP1BA*, *GP1BB* and *CD109* (Robinson *et al.*, 2015). With the exception of HPA-14bw (which is defined by a three base pair deletion), each HPA is characterised by a single nucleotide polymorphism (SNP) resulting in an amino acid substitution in the corresponding protein (Lucas, 2013).

A variety of molecular techniques have been employed to define the HPA systems (Nogués, 2011), with the majority of HPA typing methods restricted to defining HPA-1, HPA-2, HPA-3, HPA-4, HPA-5 and HPA-15, selected originally for their frequency and clinical relevance (Lucas, 2013). However, due to increasing reports of maternal alloantibodies against 'rare' HPA antigens (Poles *et al.*, 2013; Peterson *et al.*, 2012; Jallu *et al.*, 2013; Bertrand *et al.*, 2013b) it has been suggested that genotyping techniques should be expanded to include all known HPA systems (Santoso and Tsuno, 2015).

Ideally, a single method capable of both detecting all known HPA SNPs and identifying novel mutations implicated in FNAIT should be employed. This would avoid the two tiered approach often required when investigating a rare or possibly unique HPA which can lead to protracted and costly laboratory investigations, requiring full length Sanger sequencing of one or more HPA defining gene (Poles *et al.*, 2013; Wihadmadyatami *et al.*, 2015). However, to date no published HPA genotyping has reported the capacity to define all known HPA systems in a single assay.

Next Generation Sequencing (NGS) has the capacity to produce large amounts of sequence data relatively quickly and cheaply when compared to Sanger sequencing (Metzker, 2010). Lane *et al.* recently described the potential use of whole genome sequencing by NGS for predicting red cell and platelet antigens. They indicated that a targeted NGS approach might be a more affordable option for laboratories with novel DNA bar-coding technology enabling the simultaneous sequencing of specific regions of the genome in multiple individuals (Lane *et al.*, 2016).

In this study, a targeted NGS approach was designed to define all known HPA systems in either patient or donor samples, and assessed for its suitability as a routine approach for patient and donor genotyping, including novel allele detection.

6.2 Materials and methods

6.2.1 Samples

A total of 47 previously HPA genotyped DNA samples were used in this study, as detailed below.

6.2.1.1 Control DNA

Eleven DNA samples were obtained from NHSBT archives of external quality assessment (EQA) material. Samples were selected to ensure they covered as many HPA systems as were available, and included both heterozygous and homozygous examples of HPA genotypes commonly observed in the UK population. An additional fourteen anonymised control DNA samples were provided by the Australian Red Cross Blood Service and tested blind. The Australian samples included those with rarer HPA genotypes not available locally (Figure 6.1).

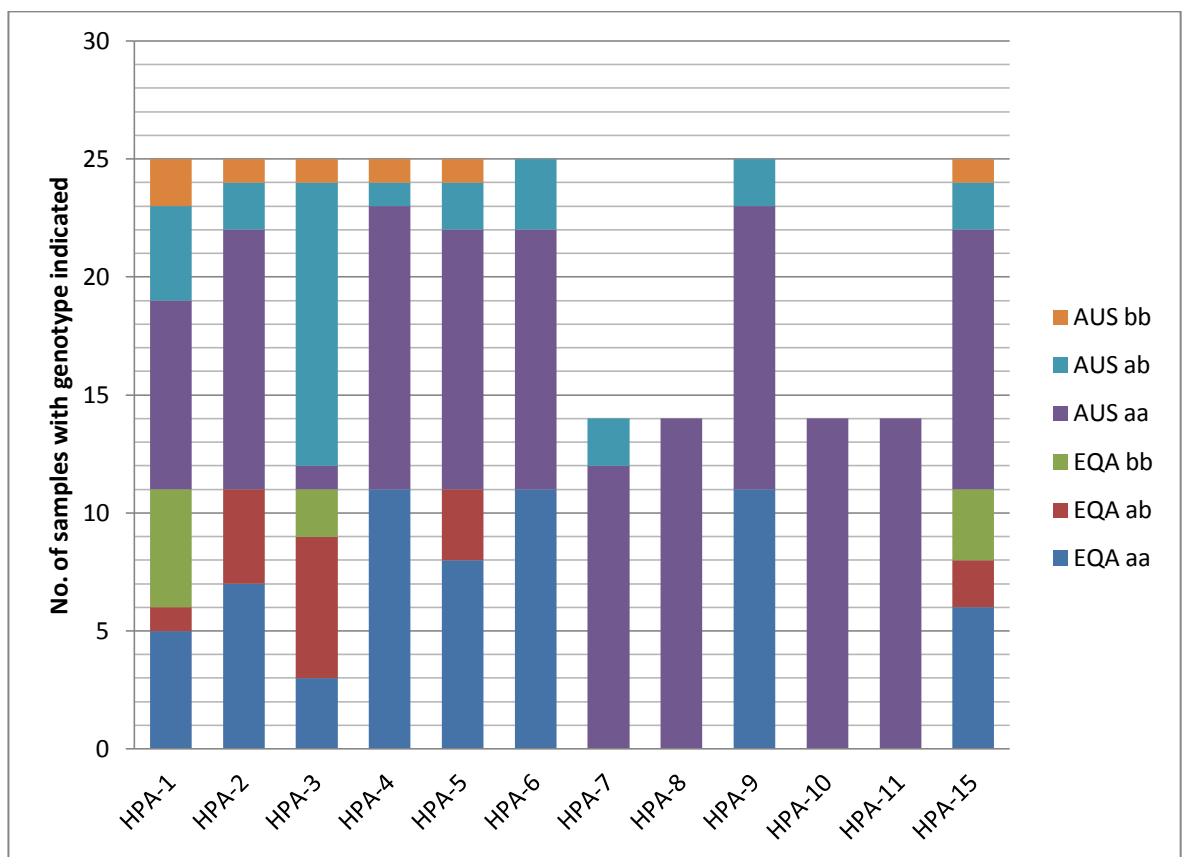


Figure 6.1 HPA genotypes of control DNA. HPA genotypes represented by the external quality assessment (EQA) control DNA and the DNA provided by the Australian Red Cross blood service (AUS). EQA samples had not been previously tested for HPA-7, -8, -10 or -11. All HPA genotypes not represented in this figure were not defined in these control samples before being used in this study.

6.2.1.2 Patient DNA

DNA was provided by the NHSBT Filton H&I laboratory. It had been previously extracted from patient blood samples originally referred for investigation, with informed consent obtained by the clinical team. Blind testing of these 22 samples was performed to assess targeted NGS as a potential method for clinical application. Included in this patient sample cohort were duplicate DNA from three patients and archive material of mother and baby from an unresolved suspected case of FNAIT which had previously undergone investigation by sequencing based typing.

6.2.2 Historical HPA genotyping

All samples in this study had been previously genotyped for HPA-1 to HPA-5 and HPA-15 by TaqMan real-time PCR, PCR-SSP and or PCR-SBT, with some control DNA additionally tested for HPA-6bw and HPA-9bw by the same method. A number of samples were also defined for HPA-7bw, -8bw, -10bw, -11bw, -27bw and or HPA-28w by either PCR-SSP or PCR-SBT (Table 6.1).

HPA system	HPA-1 to HPA-5 and HPA-15	HPA-6w	HPA-7w	HPA-8w	HPA-9w	HPA-10w	HPA-11w	HPA-27w	HPA-28w
Typing method	TaqMan/PCR-SSP/SBT	TaqMan/PCR-SSP/SBT	PCR-SSP	PCR-SSP	TaqMan/PCR-SSP/SBT	PCR-SSP	PCR-SSP	SBT	SBT
No. of samples	47	17	3	3	16	3	3	1	6

Table 6.1 HPA systems previously defined in the sample cohort. The number of DNA samples defined for each HPA system indicated and the methods employed.

6.2.3 Assay Design

A panel of probes to detect the six genes known to encode HPA capturing all exons and flanking regions was designed to determine the feasibility of targeted NGS for HPA genotyping and assess its ability to detect novel HPA, as described in Section 2.7.1.

6.2.4 Targeted enrichment and sequencing

6.2.4.1 DNA sample preparation

Each DNA sample was quantified using either the Quant-iT™ or Qubit® assay, as previously described (Section 2.4). A 50µl aliquot of each DNA sample was then prepared by diluting to 1.8ng/µl with in 10mM Tris buffer pH 8.5 (Buffer EB, Qiagen).

6.2.4.2 DNA library preparation

Indexed DNA libraries were prepared from 50ng of genomic DNA (gDNA) using the bespoke HaloPlex HS target-enrichment assay for Illumina sequencing (Agilent Technologies, UK) as

described in Section 2.7.2. Three separate experiments were set up, with variable sample numbers tested each time (Table 6.2).

Experiment	Sample type	No. of samples
Expt 1	EQA	11
Expt 2	Patient DNA (NHSBT Filton)	21
Expt 3	Control DNA (Australian Red Cross Blood Service) plus one patient (NHSBT Filton)	15

Table 6.2 Detail of samples used for each HaloPlex HS experiment. Sample origin and the number included in each of the three HaloPlex HS experiments performed in this study

The first experiment was performed using EQA samples to ensure the HaloPlex design had the capability of defining the common HPA genotypes. Second and third experiments were set up with DNA provided by external laboratories, which was tested blind and subsequently checked for concordance with the historical HPA genotype determined by the originating centre.

Following preparation, for each experiment individual indexed libraries were combined into pools for downstream sequencing. Equimolar pooling was employed for the first experiment but subsequent pooling was based on equal volumes of each sample, to reduce costs and handling time. Each pooled library was then validated with a Bioanalyser 2100 (section 2.6.5.1) to assess fragment size and concentration and then normalised to approximately 4nM. Pooled libraries were sequenced separately on a MiSeq as described in section 2.7.4, using a Nano Cartridge for the first experiment and the Standard v2 cartridge for subsequent MiSeq runs. The type of cartridge was determined by the number of samples in each MiSeq run to ensure sufficient read depth was obtained.

6.2.5 Data analysis

Primary data analysis was performed using MSR to generate a pair of FASTQ files for each sample. FASTQ files were then analysed with SureCall NGS data analysis software v3.0.3.1 (Agilent Technologies, UK) using the following pipelines:-

6.2.5.1 HPA genotyping

Single sample analysis was performed using the Default HaloPlex Method and chromosomal positions scrutinised using triage view for the respective positions for each HPA system. Nucleotides observed were used to manually assign genotypes for HPA-1 to HPA-29w, in accordance with data available from IPD-HPA. (Robinson *et al.*, 2015) All HPA genotypes generated by NGS were then compared with those HPA systems previously defined in each sample and concordance determined.

6.2.5.2 Novel mutation detection

FASTQ files from both the mother and baby from a suspected FNAIT case were analysed using HaloPlex default for pair analysis (Tumour Normal Method), allowing direct comparison of both sequences. Observed nucleotide differences between mother and baby for the six genes sequenced were examined, in particular those positions where the mother was homozygous and the baby was heterozygous, and therefore may encode a potential target of maternal alloantibody.

6.3 Results

6.3.1 Assay design

The initial HaloPlex HS design based on the HPA gene targets described in section 2.7.1 used the 'maximise specificity' level of stringency to ensure target specificity. However, the HaloPlex design report indicated this would not provide 100% coverage of all six HPA genes, predicting 99.74% and 98.05% coverage for target ID NM_000212 (ITGB3) and NM_002203 (ITGA2B), respectively. Although none of the affected regions encoded known HPA SNPs (Table 6.3), an additional probe group was designed using the 'Maximise coverage' stringency option to ensure sequence was obtained for all coding regions of the six genes, regardless. In total, 1241 amplicons were created, with predicted 100% coverage of 108 target regions covering 27,084kb for design ID 28048-1446018088 (Table 6.4).

Region of interest not amplified	TargetID
CHR5:52351784-52351814	NM_002203
CHR5:52358587-52358587	NM_002203
CHR5:52371095-52371114	NM_002203
CHR5:52385865-52385939	NM_002203
CHR17:45364593-45364602	NM_000212

Table 6.3 Regions not amplified by the original HaloPlex HS design. Target regions of interest that would not be amplified using stringency set to 'Maximize Specificity'

Gene	Genomic Interval	No. of Regions	Size bp	Databases	Coverage	High Coverage*	Low Coverage**	Stringency
GP1BA	chr17:4835850-4837908	1	2059	RefSeq	100	1	0	Maximise specificity
ITGB3	chr17:45331178-45387620	15	3867	RefSeq	99.74	15	0	Maximise specificity
GP1BB	chr22:19711043-19712037	2	821	RefSeq	100	2	0	Maximise specificity
ITGA2B	chr17:42449682-42466891	24	6046	RefSeq	100	24	0	Maximise specificity
ITGA2	chr5:52285249-52386479	29	6529	RefSeq	98.05	27	2	Maximise specificity
CD109	chr6:74405889-74533407	32	7625	RefSeq	100	32	0	Maximise specificity
ITGB3	chr17:45364593-45364602	1	10	CustomRegion	100	1	0	Maximise coverage, optimised for FFPE
ITGA2	chr5:52351784-52385939	4	127	CustomRegion	100	4	0	Maximise coverage, optimised for FFPE

Table 6.4 Summary of final HaloPlex HS Design ID 28048-1446018088. *High Coverage = number of regions where analysable amplicon overlap $\geq 90\%$. **Low Coverage= number of regions where analysable amplicon overlap $< 90\%$

Following manufacturing of the bespoke HaloPlexHS design for HPA genotyping, quality control data provided by Agilent Technologies indicated that amplicons obtained from an enrichment of high-quality DNA were of the expected size range and concentration with no excessive high or low molecular weight product (Figure 6.2).

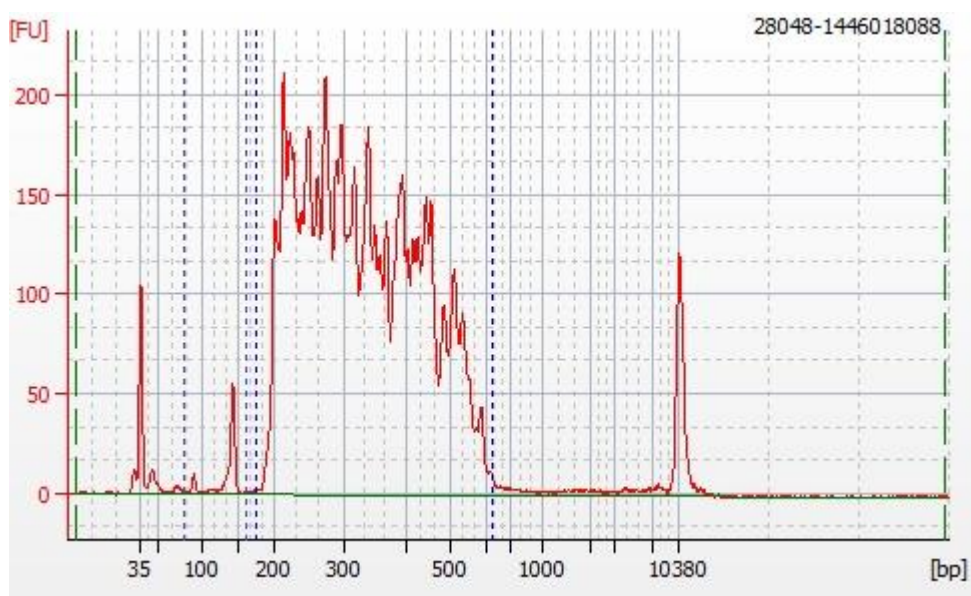


Figure 6.2 Custom HaloPlex HS design quality control. A Bioanalyser electropherogram provided by Agilent Biotechnologies of the PCR product prepared from an enrichment of high-quality DNA amplified with design ID 28048-1446018088

6.3.3 Restriction enzyme digestion

Success of the restriction digest was confirmed by analysing the enrichment control DNA sample (ECD), set up with each experiment. An example of the results obtained from this validation can be seen in Figure 6.3, with each of the eight double digests showing discrete bands at the positions indicated in the HaloPlex HS protocol, version C0 December 2015.

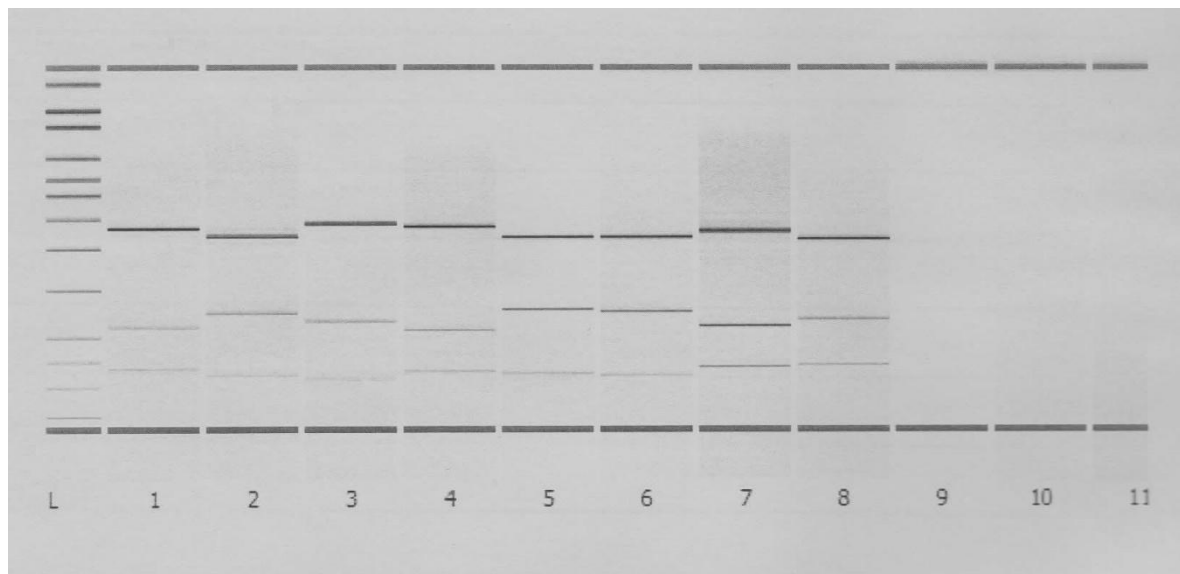


Figure 6.3 Validation of restriction enzyme digestion using the Bioanalyser 2100. Enrichment control DNA (ECD) was digested with a panel of 16 restriction enzymes in the form of two double digests provided with the HaloPlexHS kit. A 1:1 dilution of each ECD digest is shown, run using the Agilent high sensitivity kit on the Bioanalyser 2100, with results comparable to expected profile indicated in the protocol vC0 December 2015. Lane 'L' contains a 50bp ladder; lanes 1-8 contain 1:1 dilutions of the eight EDC restriction digests. Lanes 9-11 contain size markers only.

6.3.4 Validation of DNA library preparation

Pooled DNA libraries were assessed to determine the concentration and molarity using a Bioanalyser 2100 prior to dilution and loading onto the MiSeq platform. The concentration was determined by integration under the peak between 175 and 625bp, with any peak observed below 175bp excluded from the calculation. Peaks at approximately 140bp were observed and, according to the HaloPlex HS protocol, were associated with an adapter-dimer product and could be ignored if less than 10% of the overall peak value. A typical profile obtained is shown in Figure 6.4.

Molarities of the pooled DNA libraries prepared in the first and second experiments were similar, with values of 34.88nM and 32.92nM respectively. However, the third DNA library prepared was approximate half the molarity of the first two, at 18.58nM, although this

difference did not appear to have any consequence given all pooled libraries were eventually diluted to 20pM prior to loading onto the MiSeq.

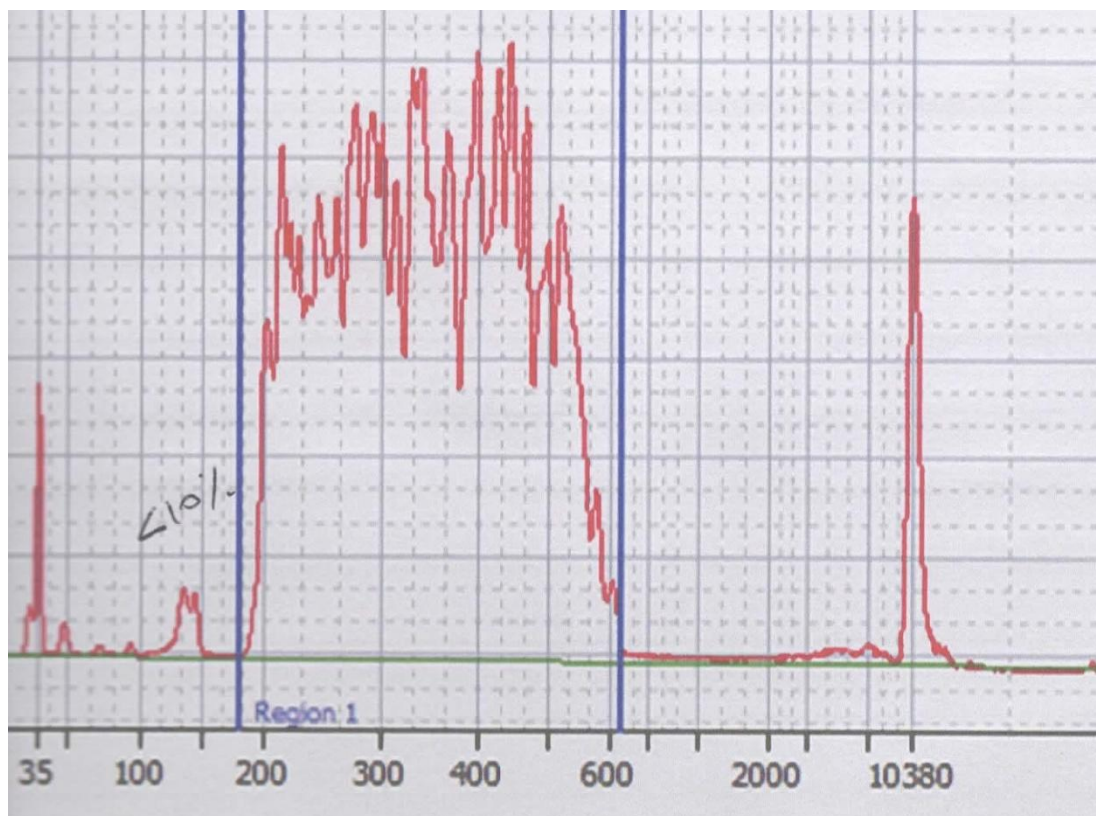


Figure 6.4 Validation Bioanalyser profile of the pooled DNA library from experiment 3. The X axis indicates the fragment size in base pairs. The blue bars (region 1) represent the peak range used to determine concentration and molarity. Values obtained for the peak at ~140bp were less than 10% of the overall concentration and were therefore ignored.

6.3.4 HPA genotyping

Forty six of the forty seven samples sequenced were successfully genotyped for all currently defined HPA systems using the HaloPlex HS assay (Figure 6.5), with one sample excluded due to possible contamination.

All HPA genotypes produced were 100% concordant with historical data. The majority of HPA systems tested were homozygous for the more common 'a' allele observed in the UK population. However, both 'a' and 'b' alleles were represented in one or more sample for HPA-1 to HPA-7bw, HPA-9bw, HPA-15, HPA-27bw and HPA-28bw (Figure 6.6). Unfortunately, it was not possible to test all potential combinations of HPA alleles due to the rarity of many HPA genotypes

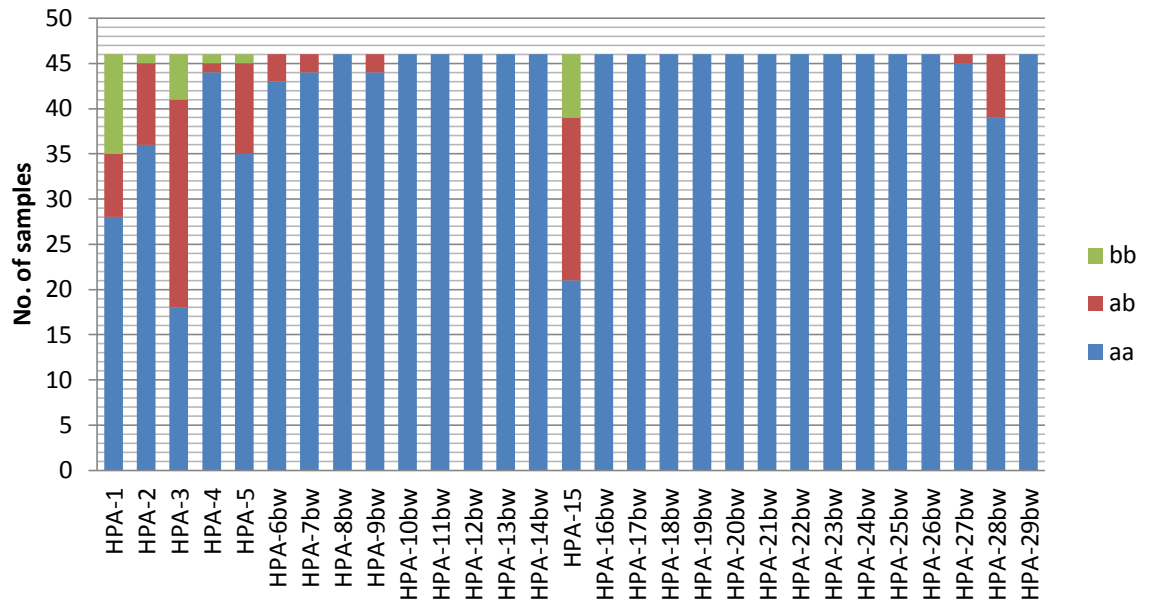


Figure 6.6 Polymorphisms of the HPA systems sequenced using HaloPlexHS. The range of HPA alleles observed in forty six samples sequenced using HaloPlex HS.

6.3.5 Sequence quality

The quality parameters of each sequence obtained by our HaloPlex HS method were scrutinised to verify the accuracy of each HPA genotype assigned.

6.3.5.1 Sequencing coverage

Cluster density on the flow cell ranged from 1115 to 1226K/mm² for the three experiments, with clusters passing filter greater than 87% on each occasion. On average, the per base depth of coverage (DoC) of 1144x was obtained for each sample, although the mean value varied between the three HaloPlex experiments performed, with an average of 584x, 1582x and 978x respectively. Variability in DoC between samples was minimal in the first experiment, but was wider in subsequent experiments (Figure 6.7).

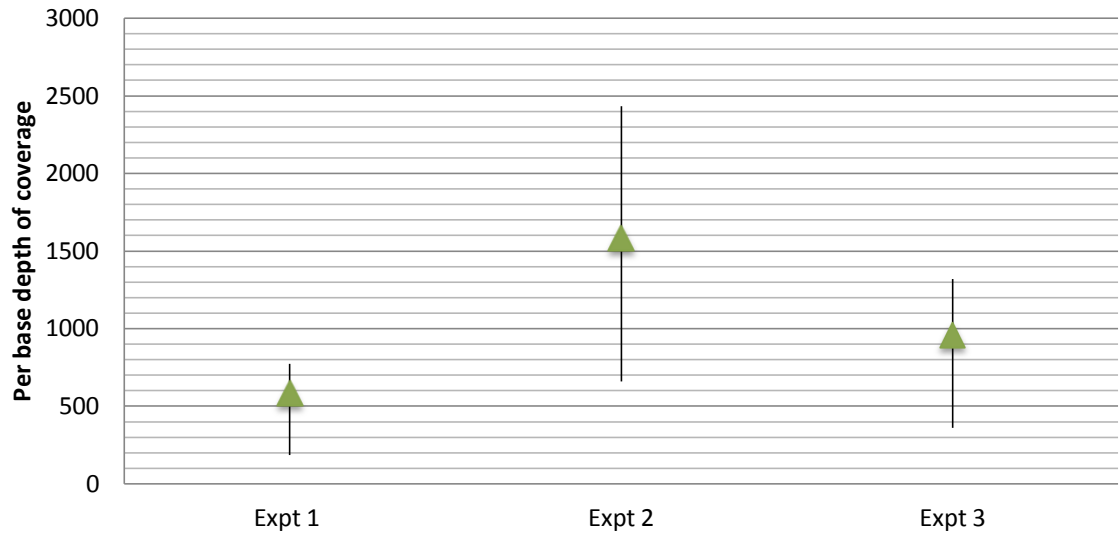


Figure 6.7 Range of mean per base depth of coverage observed for each sample. Average DoC achieved per sample for each HaloPlex HS experiment, with the overall mean represented by the triangle.

6.3.5.2 Coverage per HPA system

The average DoC observed between each HPA system was quite varied, although this variability appeared to be consistent between experiments (Figure 6.8). The highest average DoC was achieved for HPA-23bw, with 2615x (range 469-5849), and the HPA system with lowest average coverage of 335x (range 18-664) was HPA-20bw.

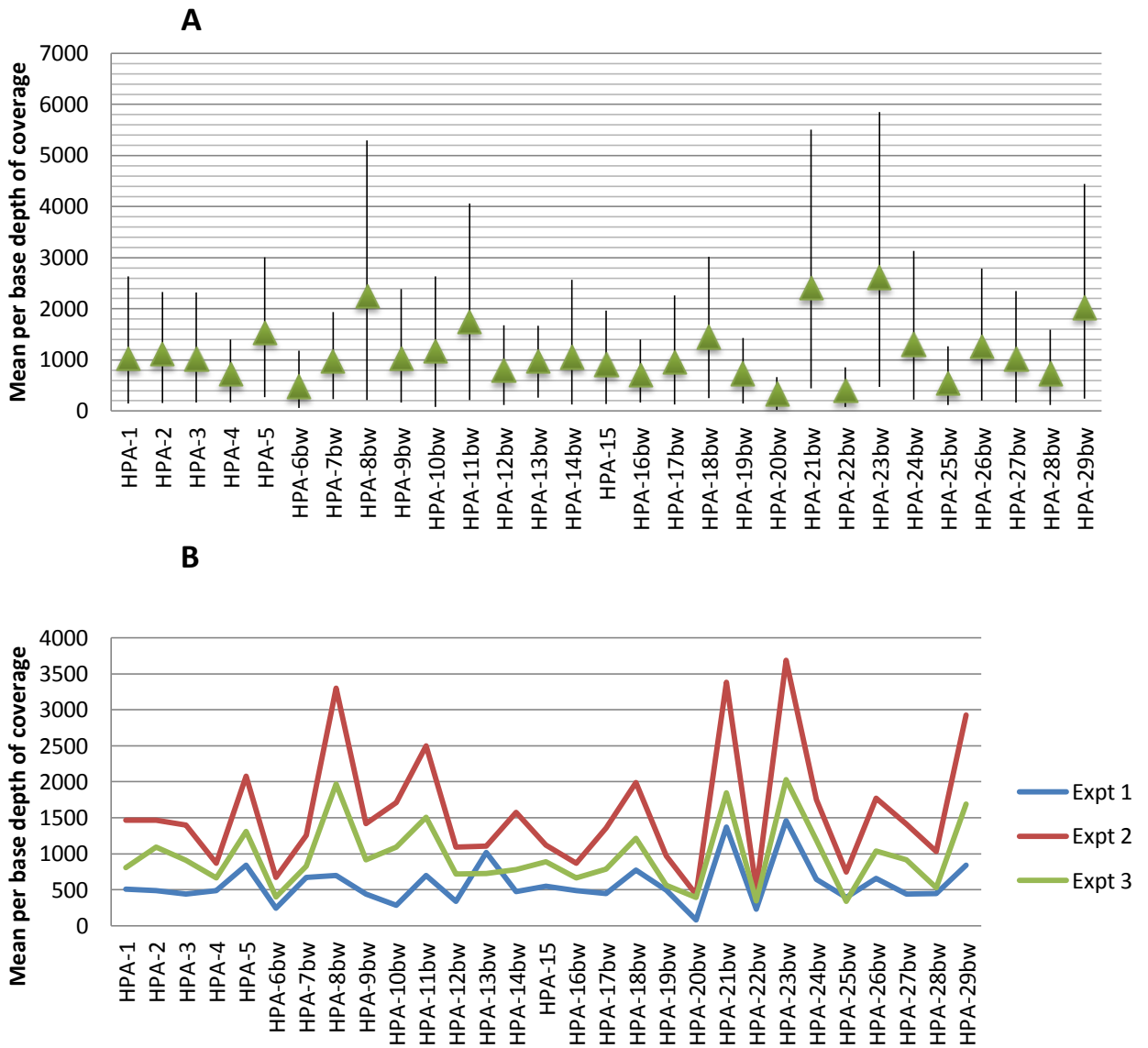


Figure 6.8 Depth of coverage observed for each HPA system. Chart A shows the average and range of per base DoC observed for each HPA system, with the mean value represented by the triangle. Graph B shows the difference in average coverage observed per HPA with each experiment.

6.3.6 Impact of GC content on read depth

Table 6.5 lists the percentage GC content of each exon that encodes each of the respective twenty nine HPA systems, along with the mean DoC observed for each SNP. Fisher’s exact test showed a significant correlation between the mean depth DoC of 1144bp and the GC content of exons encoding each HPA SNP, with a lower read depth associated with a higher GC content (Figure 6.9).

GP1BA	Gene	Exon	Total no. of nucleotides	%GC	Mean read depth (bp)
HPA-1	ITGB3	3	196	61%	926
HPA-2	GP1BA	2	2432	55%	1018
HPA-3	ITGA2B	26	126	65%	915
HPA-4	ITGB3	4	253	52%	675
HPA-5	ITGA2	13	144	42%	1411
HPA-6w	ITGB3	10	430	59%	439
HPA-7w	ITGB3	10	430	59%	918
HPA-8w	ITGB3	12	101	48%	1990
HPA-9w	ITGA2B	26	126	65%	925
HPA-10w	ITGB3	3	196	61%	1028
HPA-11w	ITGB3	12	101	48%	1568
HPA-12w	GP1BB	2	921	73%	718
HPA-13w	ITGA2	20	142	35%	949
HPA-14w	ITGB3	11	223	61%	942
HPA-15	CD109	19	118	43%	853
HPA-16w	ITGB3	4	253	52%	674
HPA-17w	ITGB3	5	163	53%	865
HPA-18w	ITGA2	17	152	40%	1325
HPA-19w	ITGB3	12	101	48%	675
HPA-20w	ITGA2B	20	148	66%	302
HPA-21w	ITGB3	12	101	48%	2202
HPA-22w	ITGA2B	5	50	60%	362
HPA-23w	ITGB3	12	101	48%	2393
HPA-24w	ITGA2B	15	105	55%	1194
HPA-25w	ITGA2	28	90	42%	494
HPA-26w	ITGB3	11	223	61%	1157
HPA-27w	ITGA2B	26	126	65%	922
HPA-28w	ITGA2B	23	81	61%	668
HPA-29w	ITGB3	2	86	63%	1817

Table 6.5 Mean read depth and percentage CG content in the encoding exon for each HPA system.

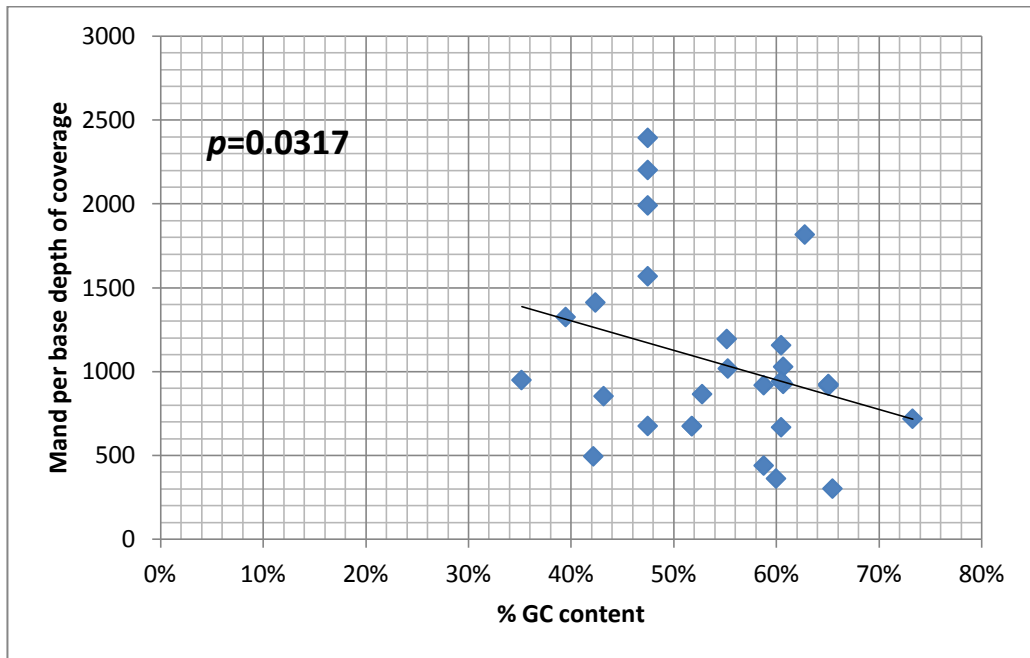


Figure 6.9 Correlation between mean per base depth of coverage and % CG content. The higher the GC content the lower the mean read depth for each HPA SNP. Significant p value determined using Fishers exact test.

6.3.7 Allele balance

The ratio of sequence depth obtained for individual SNPs in all heterozygous samples was even, with a mean of 51% for allele 'a' and 49% for allele 'b', ranging from 41-60% and 40-50% respectively. Those HPA with fewer heterozygous examples in our cohort, such as HPA-6bw, tended to have slightly uneven allele balance (Figure 6.10). However, all heterozygous samples were well within the 20:80 balance ratio used by HLA analysis software for calling heterozygous positions (Section 2.6.7).

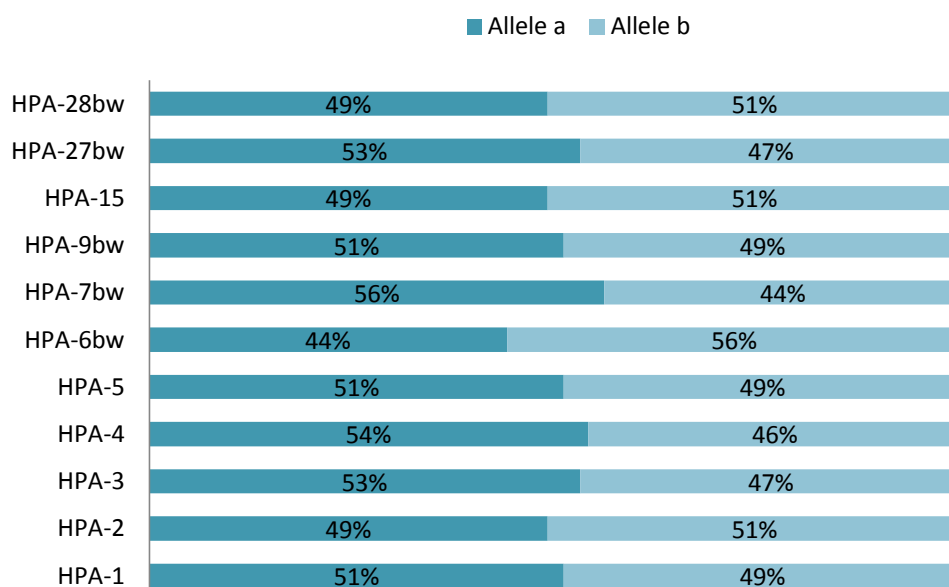


Figure 6.10 Mean allele balance observed for each heterozygous HPA system. The percentage read depth for allele ‘a’ is shown on the left hand side and allele ‘b’ on the right.

6.3.8 Novel allele detection

Full HPA genotyping of both the mother and baby from the unsolved FNAIT case was performed along with other patient samples in the second experiment. HPA-15 was potentially clinically significant, with the mother typing as HPA-15aa and baby determined to be HPA-15ab. However, this HPA system had been previously excluded by NHSBT Filton as causing FNAIT in the original investigation. Consequently, all coding regions of the six genes sequenced were examined using SureCall pair analysis to search for any other potential mutations that might be implicated in FNAIT.

The HaloPlex Tumour Normal method, using the maternal sample as the reference, indicated a total of twenty six differences between mother and baby, with 10 mutations of interest identified where mother was homozygous but the infant was heterozygous (Table 6.6). One of these mutations of interest was the previously discounted HPA-15 disparity, five were excluded as they were non-coding and three mutations were SNPs with no known association to FNAIT (National Center for Biotechnology Information SNP database) so were also ruled out. The remaining difference between mother and baby indicated a novel SNP in the *ITGB3* gene (NM_000212.2:c.1373A>G) present in baby’s sample, resulting in an amino acid change from aspartic acid to glycine. Although there was insufficient DNA to sequence the historical paternal sample, retrospective analysis of original Sanger sequencing data performed by the originating laboratory was able to confirm inheritance of this mutation in baby (A. Poles, personal communication).

Chrom	Pos	ID	Baby	Mother	HGVS(Coding)	HGVS(Protein)
chr5	52337908	rs3212441	T/C	T/C	NM_002203.3:c.186-34C>T	
chr5	52338083	rs1363192	G/T	G/T	NM_002203.3:c.295+32T>G	
chr5	52344610	rs1421933	G	G/A	NM_002203.3:c.630+10A>G	
chr5	52347369	rs1126643	T	T/C	NM_002203.3:c.759C>T	NP_002194.2:p.Phe253=
chr5	52351413	rs1062535	A	A/G	NM_002203.3:c.825G>A	NP_002194.2:p.Thr275=
chr5	52355854	rs2303127	C	C/T	NM_002203.3:c.1312+12T>C	
chr5	52356692	rs3212538	C/T	T	NM_002203.3:c.1313-39T>C	
chr5	52366138	rs2112290	A	A/G	NM_002203.3:c.2235+48G>A	
chr5	52368922	rs984966	A	A/T	NM_002203.3:c.2430-26T>A	
chr5	52370174	rs3212591	A	A/G	NM_002203.3:c.2572-41G>A	
chr5	52379277	rs2303122	T	T/C	NM_002203.3:c.3252C>T	NP_002194.2:p.Phe1084=
chr6	74432923	rs56093139	T/C	C	NM_001159587.2:c.248-51C>T	
chr6	74440104	rs143082026	A/G	A/G	NM_001159587.2:c.314G>A	NP_001153059.1:p.Arg105His
chr6	74466377	rs6453696	T/C	T	NM_001159588.2:c.414C>T	NP_001153060.1:p.Tyr138=
chr6 (HPA-15)	74493432	rs10455097	C/A	C	NM_001159588.2:c.1877A>C	NP_001153060.1:p.Tyr626Ser
chr6	74497009	rs2351528	G/A	G	NM_001159588.2:c.2159A>G	NP_001153060.1:p.Asn720Ser
chr6	74497152	rs5023688	A/G	A	NM_001159588.2:c.2302G>A	NP_001153060.1:p.Val768Ile
chr6	74521947	rs2917862	T/C	C	NM_001159588.2:c.3491C>T	NP_001153060.1:p.Thr1164Met
chr6	74528293	rs3005506	G/A	A	NM_001159587.2:c.4008+35A>G	
chr6	74533192	rs2917887	G	G	NM_001159588.2:c.3942G>T	NP_001153060.1:p.Ala1314=
chr17	4835895	rs2243093	C/T	C	NM_001165417	
chr17	45360730	rs5918	C	C/T	NM_000212.2:c.176T>C	NP_000203.2:p.Leu59Pro
chr17	45368337	rs15908	C	C/A	NM_000212.2:c.1143A>C	NP_000203.2:p.Val381=
chr17	45369617		G/A	A	NM_000212.2:c.1373A>G	NP_000203.2:p.Asp458Gly
chr17	45369777	rs4642	G	G/A	NM_000212.2:c.1533A>G	NP_000203.2:p.Glu511=
chr17	45369789	rs4634	A	A/G	NM_000212.2:c.1545G>A	NP_000203.2:p.Arg515=

Table 6.6 Nucleotide differences identified by SureCall using pair analysis between samples from mother and baby of a suspected FNAIT case. Lines highlighted in yellow are the mutations of potential interest where mother is homozygous and baby heterozygous, with the novel causative mutation highlighted in green.

6.4 Discussion

This is the first reported HPA genotyping method able to simultaneously detect and define all known HPA systems in a targeted NGS-based genotyping assay (Davey, Navarrete and Brown, 2017).

6.4.1 Assay design

A panel of HaloPlex HS probes to detect the six genes that encode all currently defined HPA systems was designed to determine the feasibility of using targeted NGS for HPA genotyping. However, rather than just target the regions containing a restricted set of HPA SNPs, as reported by others (Orzinska *et al.*, 2017), probe groups were designed to capture all exons and 50 base pair flanking regions to determine if this NGS-based approach might also be clinically useful for the investigation of FNAIT, including novel allele detection.

The original design used the 'maximise specificity' stringency option of the SureDesign HaloPlex Advanced Design Wizard; high specificity is important to avoid capturing highly similar but irrelevant regions (Nagy and Mansukhani, 2015). However, this design approach indicated that some target regions in chromosomes five and seven would not be amplified. Although the missed regions did not include those containing HPA SNPs, it was considered important to include all coding regions of each gene sequenced to facilitate detection of both known and novel HPA. Therefore design parameters were modified accordingly to ensure regions had analysable amplicon overlap of $\geq 90\%$, resulting in 100% predicted coverage for the six genes.

Validation of both the restriction digestion and library preparation using data from the Bioanalyser indicated the efficacy of the bespoke HaloPlex HS design with regards sample preparation, although not the specificity of the probe library.

6.4.2 Sequence quality

Sequencing parameters observed, including cluster density and percentage of clusters passing filter, indicated an optimal amount of library had been loaded for each MiSeq run, based on Illumina's recommended values (Illumina, 2016a).

6.4.2.1 Depth of coverage

Sufficient per base depth of coverage (DoC) is critical for NGS to ensure reliable base calling and subsequent alignment to the reference genome by analysis software. It is generally accepted that average depth in excess of $\times 30$ is acceptable for the detection of SNPs (Sims *et al.*, 2014), although there is no consensus for a minimum DoC required. A recent study reported that the accuracy of sequencing actually improves once the read depth goes beyond 10 times. (Pirooznia *et al.*, 2014). For all three HaloPlex experiments, the mean DoC achieved was well above these minimum acceptable levels, although data generated by each MiSeq run was quite disparate.

6.4.2.2 Depth of coverage per HPA system

Substantial variation in the mean DoC for each HPA SNP was observed, and was consistent in each of the three experiments. One factor that might have affected coverage of each SNP is the GC content of the respective region sequenced. It has been reported that sequences with a higher proportion of G and C nucleotides are less likely to denature during PCR (Veal *et al.*, 2012) and this may have affected the efficiency of the enrichment stage of the HaloPlex HS protocol. Indeed, exon 20 of the ITGA2B gene that encodes HPA-20w exhibited the lowest mean coverage and has a GC content of 66%. Conversely, HPA-23bw had the highest observed DoC in each experiment, with just 48% GC content in the encoding exon 12 of the ITGB3 gene.

When all HPA SNPs were considered, Fishers exact test indicated a significant association between mean DoC for each HPA SNP and the GC content of the respective encoding exon ($p=0.0317$).

6.4.2.3 Mean Depth of coverage per sample

As well as variability between each HPA, the overall mean DoC differed markedly between each experiment. The variability of the overall mean DoC observed might be explained by the different number of samples sequenced in each experiment as well as the MiSeq cartridge selected and the amount of PAL loaded for each run. However, differences observed in the second and third experiment are more likely explained by the method used for library pooling. For the first experiment samples were pooled in equimolar amounts. If an outlier sample (which was at a very low concentration) is excluded, the variability of DoC was negligible, ranging from just 539-775bp compared to 185-775bp observed in the 11 samples overall. In contrast, the range of coverage depth increased considerably in subsequent experiments when samples were pooled by equal volume, suggesting that equimolar pooling is beneficial. Pooling in equimolar concentration required the use of a Bioanalyser 2100 to determine the molarity of each library. Whilst the Bioanalyser offered the additional benefit of assessing the quality of each library, this process is both time consuming and expensive when compared to pooling by equal volumes (De Leeneer *et al.*, 2015). The costs further increase when experiments contain more than 11 samples due to the limited number of libraries that can be loaded onto a single bioanalyser chip. An alternative approach might be to estimate the molarity, given all libraries should be a similar composition as a result of AMPure bead size selection. By using an estimation of molarity, quantification of each library could be performed using a Qubit or similar fluorescent based technology which is quicker and cheaper. Subsequent pooling could then be performed by equal concentration rather than molarity or volume. This is likely to produce similar results to those observed using equimolar pooling without the additional cost of individual library assessment using a Bioanalyser.

Others have also reported lack of coverage uniformity across regions using the HaloPlex target enrichment system (Sutton *et al.*, 2015; Wendt *et al.*, 2016; Zhang *et al.*, 2016). The impact of uneven coverage can be limited by ensuring the DoC reaches the required minimum for the regions with the lowest coverage, although this may restrict the number of samples that can be multiplexed. Reducing variability between samples will allow a better estimate of the maximum number of samples that can be pooled and still maintain the deep sequencing required to produce accurate SNP detection.

6.4.2.4 Allele balance

As well as ensuring there is sufficient read depth for each HPA SNP, the accuracy of base calls for heterozygous positions is dependent on allele balance, with errors increasing as the ratio of alleles falls below 20:80 (Pirooznia *et al.*, 2014). The allele balance observed for all heterozygous HPA observed in the samples sequenced was well within this acceptable range, with the least balanced HPA-6bw still maintaining an allele ratio of 41:59. Although heterozygous examples were not available for all HPA systems in the sample cohort, the evenness in allele balance observed with those sequenced suggests rarer HPA would be detected.

6.4.3 HPA genotyping

Forty six samples were successfully HPA genotyped using the custom design HaloPlex HS target enrichment assay, with all results produced being 100% concordant with historical data. Although it was not possible to test all permutations of HPA polymorphism using this HaloPlex HS assay due to the scarcity of samples with rarer HPA genotypes, the results generated provides compelling evidence that a targeted NGS approach is a viable option for simultaneously HPA typing both patient and donor samples for all known HPA systems.

There are several advantages of a targeted NGS approach when compared to conventional Sanger sequence-based HPA typing. As well as being a one-hit assay, benefits include the requirement for only 50ng of genomic DNA (Xu *et al.*, 2009), which is particularly important for clinical investigations where patient material may be very limited due to neutropenia or in the case of a neonate. Another benefit of this approach is that NGS should not be adversely affected by rare mutations that have been known to interfere with the accuracy of some conventional PCR-based HPA genotyping techniques (Curtis and McFarland, 2009) as it is not reliant on sequence specific oligonucleotide primers or probes.

The ability to sequence many samples simultaneously is one of the key benefits of NGS, with up to twenty one samples HPA genotyped simultaneously in this study. However, the HaloPlex HS assay enables up to 96 samples to be sequenced in a single experiment using the current index combinations provided. It is anticipated that optimisation of sample pooling will increase the capacity to multiplex for HPA genotyping, although the actual number of DNA libraries that can be combined without compromising overall sequence quality has yet to be determined.

6.4.4 Novel mutation detection

Included in the sample cohort was DNA from a mother and her baby from an historical unsolved FNAIT case originally investigated by NHSBT Filton. A comparison of the six gene sequences revealed a putative novel mutation in exon 10 of the *ITGB3* gene, resulting in an

amino substitution in the glycoprotein GPIIb/IIIa. This region also encodes HPA-6w and HPA-7w (Figure 6.11), suggesting this novel mutation may have clinical significance and be implicated in FNAIT. This is currently the subject of further investigation by the originating laboratory.

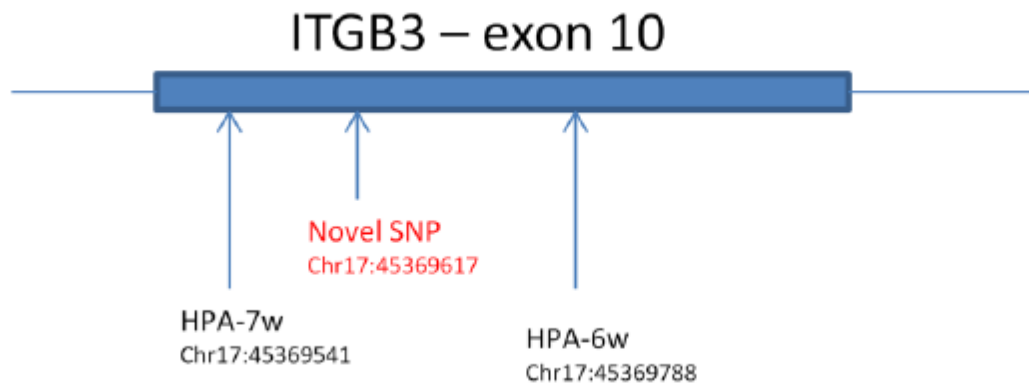


Figure 6.11 Cartoon showing the location of novel SNP in exon 10 of the *ITGB3* gene detected in patient BB. The arrows indicate the proximity of HPA-6w and HPA-7w encoded by the same exon.

7. Discussion

This study investigated the development and implementation of NGS technology to enhance the transfusion support provided to patients with either HLA and or HPA alloantibodies who require appropriately compatible products. Chapters 2 to 6 describe the successful design and development of two distinctive NGS-based methods to define the divergent nature of the HLA and HPA genetic systems and demonstrate the potential clinical application of these novel genotyping approaches.

7.1 HLA typing and NGS

The development of NGS technology has proved to be a paradigm shift in many areas of molecular biology and this is certainly true for the field of histocompatibility and immunogenetics (H&I), where use of NGS for HLA typing has been reported by many groups in recent years (Shiina *et al.*, 2012; Lange *et al.*, 2014; Hosomichi *et al.*, 2013; Ehrenberg *et al.*, 2014). The unique benefit offered by NGS for HLA definition is the ability to produce unambiguous allele level genotypes from a fully phased nucleotide sequence (Erlich, 2015). Application of NGS in H&I laboratories has primarily focused on supporting haematopoietic stem cell transplantation (HSCT) due to the requirement for a high level of matching between donor and recipient (Bravo-Egana and Monos, 2017). The high throughput capability of NGS, alongside economies of scale, makes it particularly suitable for unrelated HSCT donor registry typing and assists scientists and clinicians in identifying suitably matched donors in a timely manner (Lange *et al.*, 2014). It also negates the need for additional extended HLA typing, as may be required for donors defined on registries at low or medium HLA resolution only. Not only does this remove the costs associated with extended typing, but access to allele level donor genotyping data can also significantly reduce the time interval from initiating a donor search to selection of the final donor (Davey *et al.*, 2016). Availability of allele level HLA donor types on registries can also aid clinicians with an options appraisal for a patients' treatment, should a fully matched donor be unavailable (van Rood and Oudshoorn, 2008).

7.1.1 HLA typing platelet donors by NGS

Until now, the clinical application of HLA typing by NGS to support the selection of platelets for transfusion has not been reported. The majority of patients requiring platelet support receive random platelet transfusions but for those patients who fail to respond due to the presence of alloantibodies a specialist product is necessary to avoid HLA and or HPA antigens to which the patient has become sensitised (Brown and Navarrete, 2011). Traditionally, selection of suitable platelets for these patients has consisted of HLA matching based on serologically defined antigens which can be a complex process and requires access to a large panel of HLA typed

platelet donors (Pavenski, Freedman and Semple, 2012). This conventional approach is also limited by considering each HLA locus in isolation, excluding HLA-C from the selection algorithm and assuming homozygosity where only one antigen has been identified at an individual locus.

An alternative option for the selection of platelets for patients with IPR is the use of HLA epitope matching which is currently the subject of a non-inferiority trial (ISCTRN23996532). Benefits of HEM include a theoretically more accurate matching by selecting donors expressing epitopes shared with the recipient thereby avoiding both DSA and risk of further sensitisation. This increases the number of acceptable donors for patients that would otherwise appear mismatched using conventional HLA selection algorithms (Brooks, MacPherson and Fung, 2008). A reduction in the number of platelet donations provided by apheresis is a strategic objective within NHSBT due to the high costs of producing platelet products by apheresis compared to pooled products manufactured when processing red cells. Plans have already been implemented to reduce the proportion of platelets units collected by apheresis from 80% to 60% with further reductions likely in the future (Ronaldson and Ashford, 2014). The implementation of HEM should ensure that patients continue to receive therapeutically beneficial platelet transfusions, despite fewer single donor platelet packs on the shelf.

The matching algorithm for HEM requires both patient and donor HLA types defined to the allele level in order to accurately determine the epitopes expressed (Duquesnoy, 2011) so where HLA types are reported at a lower resolution, conversion into second field results is necessary. Whilst others have indicated this is an acceptable approach (Pai *et al.*, 2010), results presented in chapter 4 show that conversion from low or medium resolution HLA types is error prone. Disparity between predicted and actual alleles present in an individual was found to be particularly high when the first allele in the HLA string is selected as the defined type, with 6.6%, 13.5% and 13.3% errors for HLA-A, -B and -C respectively. This was of concern, as it is the method employed for allele prediction by Hematos (NHSBT's laboratory information management system) when applying HEM for donor selection. This could lead to a platelet unit being selected inappropriately and either fail to improve the patient's platelet count due to the presence of DSA or result in stimulating the production of de novo antibodies.

Further analysis of the data obtained from Chapters 3 & 4 showed that the accuracy of conversion to second field genotypes could be improved by considering HLA allele frequencies to predict alleles from an HLA string, and further still when utilising HLA haplotype frequencies. However, this would require a more complex algorithm embedded in Hematos for allele conversion and would need information on the ethnicity of the donor and patient, which is not always available or reliable. This would then require the use of default Caucasian frequency data, which may not represent the ethnicity for the respective donor or patient, potentially

leading to inaccurate HLA allele predictions. Published HLA allele and in particular haplotype frequencies is also very limited (Geneugelijk *et al.*, 2017), particularly for the UK population. Furthermore, frequencies are often based on data obtained from exons 2 and 3 rather than whole gene analysis (Geneugelijk *et al.*, 2017) and therefore do not always represent the level of resolution required. Frequency data would also need to be regularly updated for any conversion algorithm to ensure the most current allele and haplotype frequencies were utilised.

It is possible to avoid these allele prediction errors altogether by employing NGS for HLA typing. This would provide allele level definition of the HLA class I genes, removing the requirement for allele conversion by Hematos and thus ensuring accurate HEM of platelet donors. In this thesis, chapter 3 described a novel approach to NGS HLA typing that was applied to type 180 platelet donors in a single run on a MiSeq for HLA-A, -B & -C, producing results to third or fourth field resolution. Design of the described pipeline for HLA typing platelet donors occurred alongside development and implementation of an in-house NGS protocol at NHSBT for adult and cord blood donor typing as a service provision for the British Bone Marrow Registry. Some modifications were required to maximise the number of samples typed simultaneously and make optimal use of the sequencing capacity of a MiSeq flow cell. Adaptations included a reduction in the number of steps for DNA preparation, changes to the amplicon pooling process and use of an increased number of indices to facilitate additional multiplexing. In addition, two different approaches for DNA library preparation were trialed, the results of which helped inform the change to reagents kits used by the routine laboratory.

Some of these modifications could also be applied to the existing HSCT donor workflow which would lead to a reduction in the number of steps performed without compromising data quality or quantity (Figure 7.1). This would represent a cost saving due to a reduction in both DNA quantification reagents and in hands-on time. However, HSCT donors are also typed for HLA class II alleles which, based on observations in the Colindale H&I laboratory since 2015, may be more susceptible to variations in DNA quality and concentration in the initial PCR reactions, particularly for HLA-DRB1. This is probably due to the size of the HLA-DRB1 gene which is considerably larger than the HLA class I genes, spanning over 11,000kb compared to HLA-A, -B and -C which are approximately 3500, 4000 and 4300kb in length, respectively (Robinson *et al.*, 2015). Therefore removal of the DNA quantification and normalisation step for HLA class II amplification would need to be carefully validated to ensure it does not adversely affect allele detection or result in amplification bias.

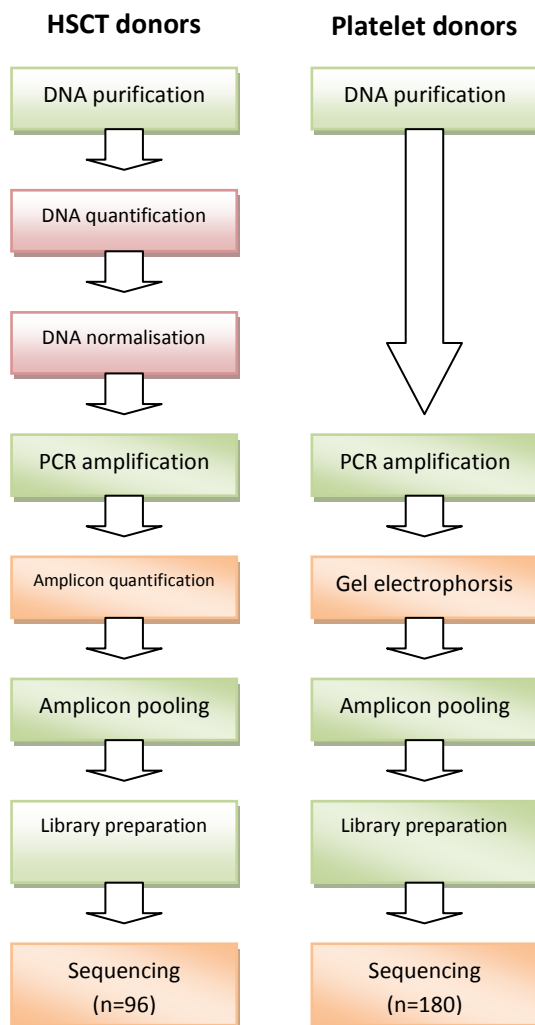


Figure 7.1 The current NGS process flow for HSCT donors alongside the NGS protocol designed for platelet donors in this study. All steps highlighted in green are the same for both protocols. Those parts of the process highlighted in orange differ between HSCT donors and platelet donor type, with gel electrophoresis replacing amplicon quantification in the platelet donor work stream and variation in the total number of samples sequenced simultaneously. Steps shown in red (DNA quantification and normalisation) are not required when processing platelet donor samples.

Due to the success of this study, the laboratory is currently in the process of implementing NGS for routine HLA-A, -B and -C typing of all apheresis platelet donors. While much of the workflow is similar to that used for HSCT donors, some modifications to the laboratory process will be necessary including adaptation of the Biomek robots for automated library preparation. This is to ensure the liquid handling robots are able to manage a reduced volume of pooled amplicon library for each sample and process 180 samples concurrently, compared with the existing 96 sample batch to form the final pooled amplicon library. Despite the variation in NGS protocols required, processing both HSCT and platelet donors samples using the same technology will bring some operational benefits for the laboratory. This will include the ability to dedicate equipment for a single process rather than having to schedule multiple methods within the one laboratory alongside the release of bench space. The training of staff will also be simplified if all samples are processed via the same work stream, although this brings its' own challenges as NGS is a technically demanding process compared to methods such as PCR-SSP or PCR-SSOP.

7.1.2 HLA typing platelet recipients by NGS

NGS is not currently employed for HLA typing patient samples at NHSBT laboratories. The reason is primarily one of logistics as the current work flow has been designed for high throughput. Although a system has been devised to manage requests for ad hoc samples for NGS typing within the routine laboratory, these samples feed into the high throughput donor typing pipeline to maximise the use of sequencing reagents; processing less than 96 samples per batch increases the cost per test significantly. This approach is not compatible with the service level agreement with service users, which is currently 5 working days from receipt of sample to reporting the result. A 5 day turnaround time would be challenging using the existing NGS work flow and does not allow time for any re-work should the initial sequence fail to produce a valid result. However, use of NGS for HLA typing could be appropriate for some patients that require life long HLA matched platelet support.

As discussed at a recent NGS special interest group meeting organised on behalf of the British Society for Histocompatibility and Immunogenetics, other UK laboratories have started to implement NGS for patient HLA typing. However, the majority are using shared sequencing facilities with other departments or employing nearby university sequencing platforms to maximise cost effectiveness. Some laboratories have opted to type all samples by NGS to streamline their workflow as a high resolution HLA type can benefit both HSCT and solid organ transplant programs, although NGS is not suitable for urgent deceased donor typing in its current form (Monos and Maiers, 2015). Of those laboratories using NGS, all but one have chosen to use commercial HLA typing kits which are now readily available, including NGSgo® (GenDx), TruSight HLA (Illumina), Holotype HLA™(Omixon) and NXType™ (One Lambda). These are generally designed for a lower throughput and claim a faster processing time than would be possible sequencing larger numbers of samples. Until HLA typing is performed by NGS for platelet recipients it could be argued that defining platelet donors at a higher resolution is of little benefit when HLA allele prediction will still be required for the majority of patients receiving platelets selected using HEM. A counter argument is that having one individual of the donor recipient pair defined to the allele level improves the overall accuracy of HEM and is still preferable to using prediction algorithms alone (Geneugelijk *et al.*, 2017). NGS could also be performed alongside the faster techniques to enable provision of subsequent products based on allele level HLA typing.

7.2 Impact of NGS on HEM

Results from Chapter 3 have shown that NGS for HLA typing produces accurate allele level typing of platelet donors. Implementation of NGS for routine platelet donor typing will ensure

that the correct epitopes are considered for each donor when using HEM to select donors for a particular patient. The principle of the HEM algorithm for transfusion is to select the platelet donor with the lowest number of epitope mismatches to the patient. Ideally the number of mismatches would be zero but it is not always possible to find a fully matched donor, regardless of the matching algorithm, and there is evidence to suggest that epitope mismatches can be tolerated below a certain threshold (Brooks, MacPherson and Fung, 2008). In reality it is more likely that the type of mismatch is more critical than the actual number per se, as there have been reports of an alloresponse to a single epitope mismatch (Lomago *et al.*, 2010). Whilst NGS will not guarantee to identify donors without any epitope mismatches, it should over time inform scientists about acceptable and unacceptable mismatches. To date, much of the research into HEM and its clinical impact has been around solid organ transplantation. There are clear indications that not all HLA mismatches are clinically significant but understanding which mismatches are permissive is a significant challenge for the discipline; it is hoped the HEM will provide an opportunity to evaluate the immunogenicity between donor and recipient and optimise outcome (Wiebe and Nickerson, 2016). Knowing the exact amino acid sequence of each individual HLA for both donor and recipient will allow scrutiny at the molecular level to hopefully provide more insight into acceptable mismatching (Erich, 2012). A recent paper by Duquesnoy and Marrari explored the use of a program called ElliPro and demonstrated its potential to predict alloantibody responses to HLA-ABC eplets. They reported that ElliPro scores were indicative of an eplets' ability to induce a specific antibody response and that a low ElliPro score assigned to some theoretical eplets suggested they were incapable of inducing an alloantibody response and therefore should be re-classified as 'non-epitopes' (Duquesnoy and Marrari, 2017; Duquesnoy, 2017). Interestingly, some of these so called 'non-epitopes' included 97T and 116F which were identified as mismatched epitopes in some allele prediction discrepancies discussed in Chapter 4 (refer to Table 4.3), suggesting that the mismatches identified are probably not clinically relevant.

Analysis of the impact of HLA epitope matching and mismatching will form part of the ongoing clinical trial into use of HEM for platelet transfusion. It is hoped that results may provide some insight into permissive HLA epitope mismatching in the transfusion setting. In an attempt to find some preliminary evidence of the impact of HEM for platelet selection, the medical history and response to transfusions was examined for a highly sensitised immune refractory patient who was provided HEM selected platelets at the request of their clinician over a period of several months. Initial indications suggested that the patient had an improved platelet increment following HEM selected platelets compared to conventional matching. However, at the time of transfusion with conventionally HLA matched platelets the patient was suffering from sepsis. This co-morbidity is as likely to explain the poor platelet increments as the

approach taken for platelet selection because sepsis is a known cause of thrombocytopenia (Brown and Navarrete, 2011). In addition, the patient had become sensitised to additional HLA antigens as a result of mismatched transfusions but because the patients' HLA antibody screening was not up to date at the time the units were allocated, platelets to which the patient had DSA were provided, resulting in poor increments. This single case serves as a note of caution when interpreting outcome data following platelet transfusions; patient responses are likely to be multi-factorial and results must be considered in their clinical context.

7.3 The wider impact of HLA typing by NGS

Due to the rapidly expanding number of HLA alleles, ensuring that PCR-SSOP and PCR-SSP techniques are able to detect all new alleles will become increasingly challenging and is more likely to require additional testing to resolve the resulting ambiguity (Tait, 2011). The use of NGS for HLA typing resolves these dilemmas, producing a high resolution or allele level result first time.

7.3.1 Whole gene HLA sequencing

The decision to use whole gene amplification in this study rather than taking a restricted exon based approach results in sequencing data for both the coding and non-coding regions of the HLA class I genes. As well as producing an unambiguous HLA genotype, sequencing beyond the conventional exons 2 and 3 includes the transmembrane and cytoplasmic regions of the HLA class I gene, improving the overall characterisation of HLA genes. Sequencing entire HLA genes should eventually provide more insight into the roles these regions play in the immune response, as well as the effect of intronic variation on gene regulation and expression (Monos and Maiers, 2015). Generation of whole gene sequence data may eventually provide other benefits, including application in phylogenetic studies and in assay design, for example in the design of primers located in intronic regions (Lind *et al.*, 2013).

7.3.2 HLA allele and haplotype frequencies

Chapter 5 discussed the HLA-A, -B and -C allele and haplotype frequencies in 519 platelet donors HLA genotyped by NGS. This data was submitted to the Allele Frequency Net Database (Santos *et al.*, 2016) as the first reported cohort of HLA class I frequency data at third field resolution in an English donor population (Davey *et al.*, 2017). One of the anticipated benefits of NGS typing is an increase in the knowledge of allele and haplotype frequencies. Results submitted based on information from this study will hopefully provide impetus to submit more frequency data as it becomes available, particularly from the adult and cord blood populations typed by NHSBT where both HLA class I and class II sequences are produced by NGS. This will in turn assist allele prediction where it is still required, either for patient samples or legacy

donors HLA typed before the implementation of NGS. This will require HLA conversion algorithms that include allele and ideally haplotype frequencies to assist with allele prediction. Indeed, the algorithm employed by Hematos is currently being reviewed by NHSBT in light of data analysed and discussed in Chapter 4 in order to improve accuracy over the existing approach of selecting the first allele in the HLA string.

7.3.3 Novel HLA allele detection

The number of novel mutations discovered in the platelet donors typed by NGS as part of this study was also discussed in Chapter 5, with 38 HLA sequences confirmed as having previously unreported mutations in either exon, intron or UTR regions. Although these putative new HLA alleles have yet to be submitted to the IMGT/HLA database, once complete this additional sequencing data will contribute to the knowledge of HLA polymorphisms.

7.4 HPA genotyping by NGS

This study has resulted in the first published application of a targeted NGS assay to simultaneously genotype patients and donors for all known HPA systems and provides proof of concept that NGS for HPA genotyping can be applied in the clinical setting (Davey, Navarrete and Brown, 2017). One of the initial objectives of this study was to design a method for HPA genotyping platelet donors to replace the existing method using the Taqman assay which is currently limited to defining HPA-1 to -5 and -15. Once designed, it was anticipated that this method would then be applied to the 540 platelet donor cohort to determine HPA-1 to -29 frequencies in the English platelet donor population. Unfortunately due to time and cost constraints this was not possible but will hopefully form part of future work. Understanding the frequencies of the HPA systems not currently defined in platelet donors will inform the strategy for HPA donor genotyping at NHSBT. If the rarer HPA systems cannot be detected in a cohort of over 500 donors it could be argued that it is not necessary to type donors routinely for these less common SNPs. However, until the frequencies are determined it is possible that some of the HPA systems are more frequent than first thought. Indeed, one of the EQA samples used to validate the targeted NGS assay was found to be HPA-28wab, which was unexpected given the HPA-28w system had only been defined relatively recently and had not been detected in 100 blood donors screened for the HPA-28w SNP (Poles *et al.*, 2013).

If the strategic decision is to implement targeted NGS for platelet donor HPA genotyping, a number of developments will be required to facilitate a high throughput workflow, essential to type the 5000 samples received per annum by NHSBT. These will include the automation of the laboratory process in order to minimise hands on time and reduce the risk of sample transfer error. This should be possible using existing Biomek robots installed as the method is based on a 96 well plate format but would require support from the manufacturers of both the kits and

robotics to implement such a protocol. For some of the steps e.g. AMPureXP bead-based clean up, automated protocols already exist in the laboratory. Other processes including the restriction digests are specific to the HaloPlex HS protocol and would need automated programs to be designed. Another critical requirement will be the development of the data analysis pipeline. Although the SureCall software is adequate for limited SNP interpretation, it would be time consuming and potentially error prone if used for high throughput, as navigating to the respective SNPs for each HPA relies on manually entering the chromosomal positions. SureCall does have the capacity to determine user parameters for automated mutation calling and although this option has been discussed with Agilent Technologies as part of this study, time constraints have thus far prevented further investigation. If NGS is to be employed for high throughput HPA typing, additional IT developments will also be required in order to provide an interface between SureCall and the existing LIMS system and enable automated data download, thereby avoiding manual transcription to ensure accuracy of data transfer and reporting of results.

7.5 Identification of platelet donors with rare genotypes

A more limited application of donor HPA genotyping could be used to specifically search for platelet donors with rare HPA genotypes. Screening patients for the presence of HPA antibodies requires techniques such as the monoclonal immobilisation of platelet antigens (MAIPA) assay, which uses a panel of HPA typed platelets (Lucas, 2013). Due to the increasing number of reports of maternal alloimmunisation against rare HPA, it could be argued that focusing only on the more common antigens when investigating suspected FNAIT is inadequate (Santoso and Tsuno, 2015). Genotyping donor samples using targeted NGS could identify apheresis donors with rare HPA genotypes who could be used to supplement the existing HPA typed platelet panel used for antibody screening by MAIPA. In order to reduce costs, one consideration might be to limit NGS genotyping to non-Caucasian donors to increase the likelihood of finding rare HPA genotypes although, as data on HPA frequencies in different populations is very limited (Robinson *et al.*, 2013), this approach might be misguided. A strong argument for implementing NGS for HPA genotyping for all platelet donors would be to increase the knowledge of HPA frequencies in the English population (Robinson *et al.*, 2013), as discussed in section 7.3. Whilst the cost of NGS is more expensive than the existing in-house Taqman technique, because NGS defines all twenty nine HPA systems alongside the potential for novel mutation detection, the cost per HPA is more economical and so could be justified.

7.6 Novel HPA allele detection

By sequencing all exons and flanking regions of the 6 genes included in the HaloPlex design, it was anticipated that this approach would enable detection of novel sequences, as well as

defining all known HPA. To challenge this theory, DNA samples extracted from a mother and her baby from a previously unresolved case of FNAIT were included in this study and sequences compared using the SureCall software paired analysis application for the detection of disease variants. An exciting development of this study was the discovery of a novel SNP in the baby's sample in a region of the ITBG3 gene between two known HPA suggesting a likely causative mutation of FNAIT (Davey, Navarrete and Brown, 2017). Further testing, including recombinant expression, by NHSBT laboratories in Bristol have since confirmed that the mutation identified in this study is a novel HPA, although this still requires ratification by the Platelet Nomenclature Committee. A report is currently in preparation for submission to the journal *Transfusion*, which will reference the NGS investigations completed as part of this study (Poles, A. and Lucas, G., personal communication). This was a significant discovery and it is hoped that targeted NGS developed during this study will be implemented for the detection of rare or novel HPA in NHSBT laboratories as part of routine FNAIT investigation.

7.7 Conclusions

This study demonstrates that NGS can significantly improve the definition of both HLA and HPA genetic systems. If implemented routinely for platelet donor genotyping and for the investigation of FNAIT, these novel approaches will provide a number of tangible benefits for laboratories and the patients they support, as well as the wider community. These include

- Allele level HLA definition to accurately select donor platelets for patients using HEM
- Detection of novel HLA class I polymorphisms, in both coding and non-coding regions
- Increased knowledge of HLA intronic sequences
- Increased knowledge of HLA allele and haplotype frequencies
- Definition of all known HPA systems in a single test
- Increased knowledge of HPA frequencies
- Detection of rare or novel HPA polymorphisms in cases of suspected FNAIT

Some additional development is still required for both HLA and HPA genotyping, particularly in respect of automation, data analysis and reporting of results, but none of these issues are insurmountable. Indeed, it is anticipated that NGS for HLA typing platelet donors will be fully implemented by autumn 2017. The timeline for introducing targeted NGS for routine definition of HPA is less clear but the potential benefits are undeniable, particularly for the detection of rare and novel HPA.

8. References

Alfirevic, A., Gonzalez-Galarza, F., Bell, C., Martinsson, K., Platt, V., Bretland, G., Evely, J., Lichtenfels, M., Cederbrant, K., French, N., Naisbitt, D., Park, B.K., Jones, A.R. and Pirmohamed, M. (2012) In silico analysis of HLA associations with drug-induced liver injury: use of a HLA-genotyped DNA archive from healthy volunteers. *Genome Medicine* [online]. 4 (6), pp.51.

Arinsburg, S.A., Shaz, B.H., Westhoff, C. and Cushing, M.M. (2012) Determination of human platelet antigen typing by molecular methods: Importance in diagnosis and early treatment of neonatal alloimmune thrombocytopenia. *American Journal of Hematology* [online]. 87 (5), pp.525-528.

ASHI (2014) *Standards for Accredited Laboratories*. Available from: <https://c.ymcdn.com/sites/ashi-hla.site-ym.com/resource/resmgr/Docs/Accreditation/Standards /2014ASHIStandardsandGuidance.pdf>.

Bauer, D.C., Zadoorian, A., Wilson, L.O., Melbourne Genomics Health Alliance and Thorne, N.P. (2016) Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Briefings in Bioinformatics* [online].

Berger, K., Schopohl, D., Wittmann, G., Schramm, W., Ostermann, H. and Rieger, C. (2016) Blood Product Supply in Germany: The Impact of Apheresis and Pooled Platelet Concentrates. *Transfusion Medicine and Hemotherapy : Offizielles Organ Der Deutschen Gesellschaft Fur Transfusionsmedizin Und Immunhamatologie* [online]. 43 (6), pp.389-394.

Bertrand, G., Bianchi, F., Quesne, J., Philippe, S., Chenet, C., Martageix, C. and Kaplan, C. (2013a) A severe case of neonatal thrombocytopenia with a fatal outcome, implicating HPA-12bw fetomaternal platelet alloimmunization. *Transfusion* [online]. 53 (2), pp.466-468.

Bertrand, G., Jallu, V., Beranger, T., Bianchi, F., Casale, C., Dufour, V., Chenet, C., Quesne, J., Martageix, C. and Kaplan, C. (2013b) HPA-5 typing discrepancy reveals an Ile503Leu substitution in platelet GPIa (α 2 integrin). *Vox Sanguinis* [online]. 105 (1), pp.73-76.

Bertrand, G. and Kaplan, C. (2014) How do we treat fetal and neonatal alloimmune thrombocytopenia? *Transfusion* [online]. 54 (7), pp.1698-1703.

Bolton-Maggs, P.H.B. (2017) Serious hazards of transfusion - conference report: celebration of 20 years of UK haemovigilance. *Transfusion Medicine (Oxford, England)* [online]. 27 (6), pp.393-400
Bravo-Egana, V. and Monos, D. (2017) The impact of next-generation sequencing in immunogenetics: current status and future directions. *Current Opinion in Organ Transplantation* [online].

Brooks, E.G., MacPherson, B.R. and Fung, M.K. (2008) Validation of HLAMatchmaker algorithm in identifying acceptable HLA mismatches for thrombocytopenic patients refractory to platelet transfusions. *Transfusion* [online]. 48 (10), pp.2159-2166.

Brown, C.J. and Navarrete, C.V. (2011) Clinical relevance of the HLA system in blood transfusion. *Vox Sanguinis* [online]. 101 (2), pp.93-105.

Buitrago, L., Rendon, A., Liang, Y., Simeoni, I., Negri, A., ThromboGenomics Consortium, Filizola, M., Ouwehand, W.H. and Coller, B.S. (2015) alpha11beta3 variants defined by next-generation sequencing: predicting variants likely to cause Glanzmann thrombasthenia.

Proceedings of the National Academy of Sciences of the United States of America [online]. 112 (15), pp.E1898-907.

Carapito, R., Radosavljevic, M. and Bahram, S. (2016) Next-Generation Sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease association studies. *Human Immunology* [online]. 77 (11), pp.1016-1023.

Cardena, M.M., Ribeiro-Dos-Santos, A., Santos, S., Mansur, A.J., Pereira, A.C. and Fridman, C. (2013) Assessment of the relationship between self-declared ethnicity, mitochondrial haplogroups and genomic ancestry in Brazilian individuals. *PLoS One* [online]. 8 (4), pp.e62005.

Cereb, N., Kim, H.R., Ryu, J. and Yang, S.Y. (2015) Advances in DNA sequencing technologies for high resolution HLA typing. *Human Immunology* [online]. 76 (12), pp.923-927.

Chakrabarti, R. and Schutt, C.E. (2001) The enhancement of PCR amplification by low molecular-weight sulfones. *Gene* [online]. 274 (1–2), pp.293-298.

Curtis, B. and McFarland, J. (2009) Detection and identification of platelet antibodies and antigens in the clinical laboratory. *Immunohematology*. 25 (3), pp.125-135.

Curtis, B.R. and McFarland, J.G. (2014) Human platelet antigens - 2013. *Vox Sanguinis* [online]. 106 (2), pp.93-102.

Dalva, K. and Beksac, M. (2007) HLA typing with sequence-specific oligonucleotide primed PCR (PCR-SSO) and use of the Luminex technology. *Methods in Molecular Medicine* [online]. 134 pp.61-69.

Davey, S., Deplano, Z., Ord, J., Quaye, L., Li, Y., Winstone, D., Parkes, G., Navarrete, C. and Brown, C. (2016) Implementation of Whole Gene Next Generation Sequencing for HLA Typing Unrelated Adult and Cord Blood Stem Cell Donors of the British Bone Marrow Registry. WMDA/IDRC Meeting (Poster). .

Davey, S., Ord, J., Navarrete, C. and Brown, C. (2017) HLA-A, -B and -C allele and haplotype frequencies defined by next generation sequencing in a population of 519 English blood donors. *Human Immunology* [online].

Davey, S., Navarrete, C. and Brown, C. (2017) Simultaneous human platelet antigen genotyping and detection of novel single nucleotide polymorphisms by targeted next-generation sequencing. *Transfusion* [online]. pp.n/a-n/a.

De Leeneer, K., Hellemans, J., Steyaert, W., Lefever, S., Vereecke, I., Debals, E., Crombez, B., Baetens, M., Van Heetvelde, M., Coppieters, F., Vandesompele, J., De Jaegher, A., De Baere, E., Coucke, P. and Claes, K. (2015) Flexible, scalable, and efficient targeted resequencing on a benchtop sequencer for variant detection in clinical practice. *Human Mutation* [online]. 36 (3), pp.379-387.

De Santis, D., Dinauer, D., Duke, J., Erlich, H.A., Holcomb, C.L., Lind, C., Mackiewicz, K., Monos, D., Moudgil, A., Norman, P., Parham, P., Sasson, A. and Allcock, R.J. (2013) 16(th) IHIW : review of HLA typing by NGS. *International Journal of Immunogenetics* [online]. 40 (1), pp.72-76.

Delbos, F., Bertrand, G., Croisille, L., Ansart-Pirenne, H., Bierling, P. and Kaplan, C. (2016) Fetal and neonatal alloimmune thrombocytopenia: predictive factors of intracranial hemorrhage. *Transfusion* [online]. 56 (1), pp.59-66; quiz 58.

- Dilthey, A., Leslie, S., Moutsianas, L., Shen, J., Cox, C., Nelson, M.R. and McVean, G. (2013) Multi-population classical HLA type imputation. *PLoS Computational Biology* [online]. 9 (2), pp.e1002877.
- Dilthey, A.T., Gourraud, P.A., Mentzer, A.J., Cereb, N., Iqbal, Z. and McVean, G. (2016) High-Accuracy HLA Type Inference from Whole-Genome Sequencing Data Using Population Reference Graphs. *PLoS Computational Biology* [online]. 12 (10), pp.e1005151.
- Duke, J.L., Lind, C., Mackiewicz, K., Ferriola, D., Papazoglou, A., Gasiewski, A., Heron, S., Huynh, A., McLaughlin, L., Rogers, M., Slavich, L., Walker, R. and Monos, D.S. (2016) Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *Hla* [online]. 87 (3), pp.141-152.
- Duquesnoy, R.J. (2017) Should epitope-based HLA compatibility be used in the kidney allocation system? *Human Immunology* [online]. 78 (1), pp.24-29.
- Duquesnoy, R.J. (2014) Update of the HLA class I eplet database in the website based registry of antibody-defined HLA epitopes. *Tissue Antigens* [online]. 83 (6), pp.382-390.
- Duquesnoy, R.J. (2011) Antibody-reactive epitope determination with HLAMatchmaker and its clinical applications. *Tissue Antigens* [online]. 77 (6), pp.525-534.
- Duquesnoy, R.J. (2008) Structural epitope matching for HLA-alloimmunized thrombocytopenic patients: a new strategy to provide more effective platelet transfusion support? *Transfusion* [online]. 48 (2), pp.221-227.
- Duquesnoy, R.J., Honger, G., Hosli, I., Marrari, M. and Schaub, S. (2016) Detection of newly antibody-defined epitopes on HLA class I alleles reacting with antibodies induced during pregnancy. *International Journal of Immunogenetics* [online]. 43 (4), pp.200-208.
- Duquesnoy, R.J. and Marrari, M. (2017) Usefulness of the ElliPro epitope predictor program in defining the repertoire of HLA-ABC eplets. *Human Immunology* [online].
- Duquesnoy, R.J., Marrari, M., da M Sousa, L.C., de M Barroso, J.R., de S U Aita, K.M., da Silva, A.S. and do Monte, S.J. (2013a) 16th IHIW: a website for antibody-defined HLA epitope Registry. *International Journal of Immunogenetics* [online]. 40 (1), pp.54-59.
- Duquesnoy, R.J., Marrari, M., da M Sousa, L.C., de M Barroso, J.R., de S U Aita, K.M., da Silva, A.S. and do Monte, S.J. (2013b) 16th IHIW: a website for antibody-defined HLA epitope Registry. *International Journal of Immunogenetics* [online]. 40 (1), pp.54-59.
- Duquesnoy, R.J., Marrari, M., Mulder, A., Sousa, L.C., da Silva, A.S. and do Monte, S.J. (2014) First report on the antibody verification of HLA-ABC epitopes recorded in the website-based HLA Epitope Registry. *Tissue Antigens* [online]. 83 (6), pp.391-400.
- Dyer, P. A. (1988) Class I and class II loci of the human major histocompatibility complex. *Baillière's Clinical Rheumatology* [online]. 2(3), pp. 529-546.
- Dyer, P., McGilvray, R., Robertson, V. and Turner, D. (2013) Status report from 'double agent HLA': Health and disease. *Molecular Immunology* [online]. 55 (1), pp.2-7.
- Ehrenberg, P.K., Geretz, A., Baldwin, K.M., Apps, R., Polonis, V.R., Robb, M.L., Kim, J.H., Michael, N.L. and Thomas, R. (2014) High-throughput multiplex HLA genotyping by next-

generation sequencing using multi-locus individual tagging. *BMC Genomics* [online]. 15 pp.864-2164-15-864.

Erlich, H. (2012) HLA DNA typing: past, present, and future. *Tissue Antigens* [online]. 80 (1), pp.1-11.

Erlich, H.A. (2015) HLA typing using next generation sequencing: An overview. *Human Immunology* [online].

Excoffier, L. and Lischer, H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* [online]. 10 (3), pp.564-567.

Finning, K., Bhandari, R., Sellers, F., Revelli, N., Villa, M.A., Muniz-Diaz, E. and Nogues, N. (2016) Evaluation of red blood cell and platelet antigen genotyping platforms (ID CORE XT/ID HPA XT) in routine clinical practice. *Blood Transfusion = Trasfusione Del Sangue* [online]. 14 (2), pp.160-167.

Gabriel, C., Furst, D., Fae, I., Wenda, S., Zollikofer, C., Mytilineos, J. and Fischer, G.F. (2014) HLA typing by next-generation sequencing - getting closer to reality. *Tissue Antigens* [online]. 83 (2), pp.65-75.

Geneugelijck, K., Wissing, J., Koppenaar, D., Niemann, M. and Spierings, E. (2017) Computational Approaches to Facilitate Epitope-Based HLA Matching in Solid Organ Transplantation. *Journal of Immunology Research* [online]. 2017 pp.9130879.

George, J.N. (2000) Platelets. *The Lancet*. 355 (April 29), pp.1531-1539.

Ghevaert, C., Campbell, K., Walton, J., Smith, G.A., Allen, D., Williamson, L.M., Ouwehand, W.H. and Ransinghe, E. (2007) Management and outcome of 200 cases of fetomaternal alloimmune thrombocytopenia. *Transfusion* [online]. 47 (5), pp.901-910.

Ghevaert, C., Rankin, A., Huiskes, E., Porcelijn, L., Javela, K., Kekomaki, R., Bakchoul, T., Santoso, S., Nutland, S., Smyth, D.J., Smith, G.A., McBride, S., Watkins, N.A. and Ouwehand, W.H. (2009) Alloantibodies against low-frequency human platelet antigens do not account for a significant proportion of cases of fetomaternal alloimmune thrombocytopenia: evidence from 1054 cases. *Transfusion* [online]. 49 (10), pp.2084-2089.

Gonzalez-Galarza, F.F., Takeshita, L.Y., Santos, E.J., Kempson, F., Maia, M.H., da Silva, A.L., Teles e Silva, A.L., Ghattaoraya, G.S., Alfirevic, A., Jones, A.R. and Middleton, D. (2015a) Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Research* [online]. 43 (Database issue), pp.D784-8.

Gonzalez-Galarza, F.F., Takeshita, L.Y., Santos, E.J., Kempson, F., Maia, M.H., da Silva, A.L., Teles e Silva, A.L., Ghattaoraya, G.S., Alfirevic, A., Jones, A.R. and Middleton, D. (2015b) Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Research* [online]. 43 (Database issue), pp.D784-8.

Guo, S.W. and Thompson, E.A. (1992) A Monte Carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics* [online]. 51 (5), pp.1111-1126.

Holcomb, C.L., Rastrou, M., Williams, T.C., Goodridge, D., Lazaro, A.M., Tilanus, M. and Erlich, H.A. (2014) Next-generation sequencing can reveal in vitro-generated PCR crossover products:

- some artifactual sequences correspond to HLA alleles in the IMGT/HLA database. *Tissue Antigens* [online]. 83 (1), pp.32-40.
- Hong, X., Chen, S., Ying, Y., Liu, Y., Xu, X., He, J. and Zhu, F. (2017) Simultaneous genotyping of human platelet alloantigen-1 to 28bw systems by multiplex polymerase chain reaction sequence-based typing. *Vox Sanguinis* [online].
- Hosomichi, K., Jinam, T.A., Mitsunaga, S., Nakaoka, H. and Inoue, I. (2013) Phase-defined complete sequencing of the HLA genes by next-generation sequencing. *BMC Genomics* [online]. 14 pp.355-2164-14-355.
- Huang, Y., Yang, J., Ying, D., Zhang, Y., Shotelersuk, V., Hirankarn, N., Sham, P.C., Lau, Y.L. and Yang, W. (2015) HLAreporter: a tool for HLA typing from next generation sequencing data. *Genome Medicine* [online]. 7 (1), pp.25-015-0145-3. eCollection 2015.
- Illumina *Optimising Cluster Density on Illumina Sequencing Systems*. (2016a).
- Illumina (2016b) *TruSight HLA Sequencing Panel v2*. Available from: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/trusight-hla-v2-data-sheet-070-2016-002.pdf> .
- Illumina (2015) *Nextera XT Library Prep: Tips and Troubleshooting*. Available from: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-troubleshooting-guide.pdf .
- IPD- IMGT/HLA (2017) *Conditions for Acceptance of New Allele Sequences*. Available from: <https://www.ebi.ac.uk/ipd/imgt/hla/subs/submit.html> .
- Jallu, V., Bertrand, G., Bianchi, F., Chenet, C., Poulain, P. and Kaplan, C. (2013) The alphaIIb p.Leu841Met (Cab3(a+)) polymorphism results in a new human platelet alloantigen involved in neonatal alloimmune thrombocytopenia. *Transfusion* [online]. 53 (3), pp.554-563.
- Jenne, C.N., Urrutia, R. and Kubes, P. (2013) Platelets: bridging hemostasis, inflammation, and immunity. *International Journal of Laboratory Hematology* [online]. 35 (3), pp.254-261.
- Kallon, D. *Hematos User Requirement Specification H0391: Eplet Matcing*. v1.1 (2015)NHS Blood and Transplant.
- Kang, J., Lee, M.S. and Gorenstein, D.G. (2005) The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: application to in vitro combinatorial selection of aptamers. *Journal of Biochemical and Biophysical Methods* [online]. 64 (2), pp.147-151.
- Kengkate, M., Butthep, P., Kupatawintu, P., Srisuddee, A., Chantratita, W. and Nathalang, O. (2015) Comparison of a simple-probe real-time PCR and multiplex PCR techniques for HPA-1 to HPA-6 and HPA-15 genotyping. *Journal of Clinical Laboratory Analysis* [online]. 29 (2), pp.94-99.
- Knierim, E., Lucke, B., Schwarz, J.M., Schuelke, M. and Seelow, D. (2011) Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One* [online]. 6 (11), pp.e28240.

- Kopko, P.M., Warner, P., Kresie, L. and Pancoska, C. (2015) Methods for the selection of platelet products for alloimmune-refractory patients. *Transfusion* [online]. 55 (2), pp.235-244.
- Lamb, G., Choi, K.-., Selwyn, C., Wheeler, A., Hammond, L., Morgan, J. and Dunn, P.P.J. (2015) Identification of seven novel HLA class I alleles in New Zealand. *International Journal of Immunogenetics* [online]. 42 (5), pp.361-363.
- Lan, J.H., Yin, Y., Reed, E.F., Moua, K., Thomas, K. and Zhang, Q. (2015) Impact of three Illumina library construction methods on GC bias and HLA genotype calling. *Human Immunology* [online]. 76 (2-3), pp.166-175.
- Lane, W.J., Westhoff, C.M., Uy, J.M., Aguad, M., Smeland-Wagman, R., Kaufman, R.M., Rehm, H.L., Green, R.C., Silberstein, L.E. and for the MedSeq Project (2016) Comprehensive red blood cell and platelet antigen prediction from whole genome sequencing: proof of principle. *Transfusion* [online]. 56 (3), pp.743-754.
- Lange, V., Bohme, I., Hofmann, J., Lang, K., Sauter, J., Schone, B., Paul, P., Albrecht, V., Andreas, J.M., Baier, D.M., Nething, J., Ehninger, U., Schwarzelt, C., Pingel, J., Ehninger, G. and Schmidt, A.H. (2014) Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics* [online]. 15 pp.63-2164-15-63.
- Lattimore, S., Wickenden, C. and Brailsford, S.R. (2015) Blood donors in England and North Wales: demography and patterns of donation. *Transfusion* [online]. 55 (1), pp.91-99.
- Lee, C. and Ayob, Y. (2015) Approach to managing platelet refractory patients. *ISBT Science Series* [online]. 10 (S1), pp.89-94.
- Lind, C., Ferriola, D., Mackiewicz, K., Heron, S., Rogers, M., Slavich, L., Walker, R., Hsiao, T., McLaughlin, L., D'Arcy, M., Gai, X., Goodridge, D., Sayer, D. and Monos, D. (2010) Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. *Human Immunology* [online]. 71 (10), pp.1033-1042.
- Lind, C., Ferriola, D., Mackiewicz, K., Papazoglou, A., Sasson, A. and Monos, D. (2013) Filling the gaps - the generation of full genomic sequences for 15 common and well-documented HLA class I alleles using next-generation sequencing technology. *Human Immunology* [online]. 74 (3), pp.325-329.
- Lomago, J., Jelenik, L., Zern, D., Howe, J., Martell, J., Zeevi, A. and Duquesnoy, R.J. (2010) How did a patient who types for HLA-B*4403 develop antibodies that react with HLA-B*4402? *Human Immunology* [online]. 71 (2), pp.176-178.
- Lucas, G. (2013) Human Platelet Antigens (HPA) and Human Neutrophil Antigens (HNA) and their clinical significance. In: Knight, R., ed. (2013) *Transfusion and Transplantation Science*. Great Clarendon Street, Oxford, OX2 6DP, United Kingdom: Oxford University Press, pp.187-216.
- Lucas, G., Poles, A., Wozniak, M.J. and Gilmore, R. (2016) Further observations on the clinical significance and inheritance of the low-frequency platelet antigen HPA-28bw. *Transfusion* [online]. 56 (4), pp.873-877.
- Maccari, G., Robinson, J., Ballingall, K., Guethlein, L.A., Grimholt, U., Kaufman, J., Ho, C.S., de Groot, N.G., Flicek, P., Bontrop, R.E., Hammond, J.A. and Marsh, S.G. (2017) IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Research* [online]. 45 (D1), pp.D860-D864.

- Mack, S.J. and Middleton, D. (2015) Introducing a new manuscript format: Enabling access to immunogenomic population data with short population reports. *Human Immunology* [online]. 76 (6), pp.393-394.
- Maiers, M., Gragert, L. and Klitz, W. (2007) High-resolution HLA alleles and haplotypes in the United States population. *Human Immunology* [online]. 68 (9), pp.779-788.
- Major, E., Rigo, K., Hague, T., Berces, A. and Juhos, S. (2013) HLA typing from 1000 genomes whole genome and whole exome illumina data. *PLoS One* [online]. 8 (11), pp.e78410.
- Marsh, S.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Fernandez-Vina, M., Geraghty, D.E., Holdsworth, R., Hurley, C.K., Lau, M., Lee, K.W., Mach, B., Maiers, M., Mayr, W.R., Muller, C.R., Parham, P., Petersdorf, E.W., Sasazuki, T., Strominger, J.L., Svejgaard, A., Terasaki, P.I., Tiercy, J.M. and Trowsdale, J. (2010) Nomenclature for factors of the HLA system, 2010. *Tissue Antigens* [online]. 75 (4), pp.291-455.
- Mayor, N.P., Robinson, J., McWhinnie, A.J., Ranade, S., Eng, K., Midwinter, W., Bultitude, W.P., Chin, C.S., Bowman, B., Marks, P., Braund, H., Madrigal, J.A., Latham, K. and Marsh, S.G. (2015) HLA Typing for the Next Generation. *PLoS One* [online]. 10 (5), pp.e0127153.
- Metcalfe, P., Watkins, N.A., Ouwehand, W.H., Kaplan, C., Newman, P., Kekomaki, R., De Haas, M., Aster, R., Shibata, Y., Smith, J., Kiefel, V. and Santoso, S. (2003) Nomenclature of human platelet antigens. *Vox Sanguinis* [online]. 85 (3), pp.240-245.
- Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics* [online]. 11 (1), pp.31-46.
- Mongkolsuk, T., Ingsathit, A., Worawichawong, S., Jirasiritham, S., Kitpoka, P. and Thammanichanond, D. (2014) Shared molecular eplet stimulates acute antibody-mediated rejection in a kidney transplant recipient with low-level donor-specific antibodies: a case report. *Transplantation Proceedings* [online]. 46 (2), pp.644-647.
- Monos, D. and Maiers, M.J. (2015) Progressing towards the complete and thorough characterization of the HLA genes by NGS (or single-molecule DNA sequencing): Consequences, opportunities and challenges. *Human Immunology* [online]. 76 (12), pp.883-886.
- Murphy, M.F., Metcalfe, P., Thomas, H., Eve, J., Ord, J., Lister, T.A. and Waters, A.H. (1986) Use of leucocyte-poor blood components and HLA-matched-platelet donors to prevent HLA alloimmunization. *British Journal of Haematology* [online]. 62 (3), pp.529-534.
- Muzzey, D., Evans, E.A. and Lieber, C. (2015) Understanding the Basics of NGS: From Mechanism to Variant Calling. *Current Genetic Medicine Reports* [online]. 3 (4), pp.158-165.
- Mwandoro, N., Kallon, D., North, H. and Goodwin, J. *Hematos User Requirement Specification H0387. New HLA/HPA platelets/Red Cells Donations Search*. (2015)NHS Blood and Transplant.
- Nagy, P.L. and Mansukhani, M. (2015) The role of clinical genomic testing in diagnosis and discovery of pathogenic mutations. *Expert Review of Molecular Diagnostics* [online]. 15 (9), pp.1101-1105.
- Nambiar, A., Duquesnoy, R.J., Adams, S., Zhao, Y., Oblitas, J., Leitman, S., Stroncek, D. and Marincola, F. (2006) HLAMatchmaker-driven analysis of responses to HLA-typed platelet

transfusions in alloimmunized thrombocytopenic patients. *Blood* [online]. 107 (4), pp.1680-1687.

Nelson, W.C., Pyo, C., Vogan, D., Wang, R., Pyon, Y., Hennessey, C., Smith, A., Pereira, S., Ishitani, A. and Geraghty, D.E. (2015) An integrated genotyping approach for HLA and other complex genetic systems. *Human Immunology* [online]. 76 (12), pp.928-938.

Nogués, N. (2011) Molecular testing for platelet and granulocyte antigens. *Isbt Science Series*. 6 pp.7-12.

Nunes, E., Heslop, H., Fernandez-Vina, M., Taves, C., Wagenknecht, D.R., Eisenbrey, A.B., Fischer, G., Poulton, K., Wacker, K., Hurley, C.K., Noreen, H. and Sacchi, N. (2011) Definitions of histocompatibility typing terms. *Blood* [online]. 118 (23), pp.e180-3.

Office for National Statistics, (2012) *Ethnicity and National Identity in England and Wales: 2011*. Office for National Statistics.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D. and Pruitt, K.D. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* [online]. 44 (D1), pp.D733-45.

Orzinska, A., Guz, K., Mikula, M., Kulecka, M., Kluska, A., Balabas, A., Pelc-Klopotowska, M., Ostrowski, J. and Brojer, E. (2017) A preliminary evaluation of next-generation sequencing as a screening tool for targeted genotyping of erythrocyte and platelet antigens in blood donors. *Blood Transfusion = Trasfusione Del Sangue* [online]. pp.1-8.

Ouwehand, W.H. (2005) The dilemma of screening for antibodies against low-frequency human platelet antigens. *Transfusion* [online]. 45 (3), pp.288-289.

Pai, S., Lo, S., Lin Tsai, S., Chang, J., Lin, D., Lin, K. and Lin, L. (2010) Epitope-based matching for HLA-alloimmunized platelet refractoriness in patients with hematologic diseases. *Transfusion* [online]. 50 (11), pp.2318-2327.

Pavenski, K., Freedman, J. and Semple, J.W. (2012) HLA alloimmunization against platelet transfusions: pathophysiology, significance, prevention and management. *Tissue Antigens* [online]. 79 (4), pp.237-245.

Pavenski, K., Rebullá, P., Duquesnoy, R., Saw, C.L., Slichter, S.J., Tanael, S., Shehata, N. and International Collaboration for Guideline Development, Implementation and Evaluation for Transfusion Therapies (ICTMG) Collaborators (2013) Efficacy of HLA-matched platelet transfusions for patients with hypoproliferative thrombocytopenia: a systematic review. *Transfusion* [online]. 53 (10), pp.2230-2242.

Peterson, J., Pechauer, S., Gitter, M., Kanack, A., Curtis, B., Reese, J., Kamath, V., McFarland, J. and Aster, R. (2012) New platelet glycoprotein polymorphisms causing maternal immunization and neonatal alloimmune thrombocytopenia. *Transfusion*. 52 pp.1117-1124.

- Peterson, J.A., Gitter, M., Bougie, D.W., Pechauer, S., Hopp, K.A., Pietz, B., Szabo, A., Curtis, B.R., McFarland, J. and Aster, R.H. (2014) Low-frequency human platelet antigens as triggers for neonatal alloimmune thrombocytopenia. *Transfusion* [online]. 54 (5), pp.1286-1293.
- Petz, L.D., Garratty, G., Calhoun, L., Clark, B.D., Terasaki, P.I., Gresens, C., Gornbein, J.A., Landaw, E.M., Smith, R. and Cecka, J.M. (2000) Selecting donors of platelets for refractory patients on the basis of HLA antibody specificity. *Transfusion* [online]. 40 (12), pp.1446-1456.
- Pingel, J., Solloch, U.V., Hofmann, J.A., Lange, V., Ehninger, G. and Schmidt, A.H. (2013) High-resolution HLA haplotype frequencies of stem cell donors in Germany with foreign parentage: How can they be used to improve unrelated donor searches? *Human Immunology* [online]. 74 (3), pp.330-340.
- Pirooznia, M., Kramer, M., Parla, J., Goes, F.S., Potash, J.B., McCombie, W.R. and Zandi, P.P. (2014) Validation and assessment of variant calling pipelines for next-generation sequencing. *Human Genomics* [online]. 8 pp.14-7364-8-14.
- Poles, A., Wozniak, M.J., Walser, P., Ridgwell, K., Fitzgerald, J., Green, A., Gilmore, R. and Lucas, G. (2013) A V740L mutation in glycoprotein IIb defines a novel epitope (War) associated with fetomaternal alloimmune thrombocytopenia. *Transfusion* [online].
- Profaizer, T., Coonrod, E.M., Delgado, J.C. and Kumánovics, A. (2015) Report on the effects of fragment size, indexing, and read length on HLA sequencing on the Illumina MiSeq. *Human Immunology* [online]. 76 (12), pp.897-902.
- Profaizer, T., Lazar-Molnar, E., Pole, A., Delgado, J.C. and Kumanovics, A. (2017) HLA genotyping using the Illumina HLA TruSight next-generation sequencing kits: A comparison. *International Journal of Immunogenetics* [online].
- Ramos, B.R., D'Elia, M.P., Amador, M.A., Santos, N.P., Santos, S.E., da Cruz Castelli, E., Witkin, S.S., Miot, H.A., Miot, L.D. and da Silva, M.G. (2016) Neither self-reported ethnicity nor declared family origin are reliable indicators of genomic ancestry. *Genetica* [online]. 144 (3), pp.259-265.
- René, C., Lozano, C. and Eliaou, J.-. (2016) Expression of classical HLA class I molecules: regulation and clinical impacts. *HLA* [online]. 87 (5), pp.338-349.
- Rioux-Masse, B., Cohn, C., Lindgren, B., Pulkrabek, S. and McCullough, J. (2014) Utilization of cross-matched or HLA-matched platelets for patients refractory to platelet transfusion. *Transfusion* [online]. 54 (12), pp.3080-3087.
- Robinson, J., Halliwell, J.A., McWilliam, H., Lopez, R. and Marsh, S.G. (2013) *IPD--the Immuno Polymorphism Database*. Available from: <http://www.ebi.ac.uk/ipd/hpa> .
- Robinson, J., Halliwell, J.A., Hayhurst, J.D., Flicek, P., Parham, P. and Marsh, S.G.E. (2015) The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research* [online]. 43 (D1), pp.D423-D431.
- Ronaldson, C. and Ashford, M., (2014) *Platelet Supply Project*. http://www.nhsbt.nhs.uk/download/board_papers/july14/m14_74_Platelet_Supply_Project.pdf: NHS Blood and Transplant.
- Rubinstein, P. (2010) Platelet transfusions to HLA-immunized recipients: how to be or not to be matched. *Transfusion* [online]. 50 (11), pp.2292-2294.

Sanchez-Mazas, A., Nunes, J.M., Middleton, D., Sauter, J., Buhler, S., McCabe, A., Hofmann, J., Baier, D.M., Schmidt, A.H., Nicoloso, G., Andreani, M., Grubic, Z., Tiercy, J.M. and Fleischhauer, K. (2017) Common and well-documented HLA alleles over all of Europe and within European sub-regions: A catalogue from the European Federation for Immunogenetics. *Hla* [online]. 89 (2), pp.104-113.

Santos, E.J.M.d., McCabe, A., Gonzalez-Galarza, F.F., Jones, A.R. and Middleton, D. (2016) Allele Frequencies Net Database: Improvements for storage of individual genotypes and analysis of existing data. *Human Immunology* [online]. 77 (3), pp.238-248.

Santoso, S. and Tsuno, N.H. (2015) Progress and challenges in platelet serology. *ISBT Science Series* [online]. 10 (S1), pp.211-218.

Schafer, C., Sauter, J., Riethmuller, T., Kashi, Z.M., Schmidt, A.H. and Barriga, F.J. (2016) HLA-A, -B, -DRB1 allele and haplotype frequencies of 920 cord blood units from Central Chile. *Human Immunology* [online]. 77 (8), pp.622-623.

Schofl, G., Lang, K., Quenzel, P., Bohme, I., Sauter, J., Hofmann, J.A., Pingel, J., Schmidt, A.H. and Lange, V. (2017) 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned. *BMC Genomics* [online]. 18 (1), pp.161-017-3575-z.

Shiina, T., Blancher, A., Inoko, H. and Kulski, J.K. (2017) Comparative genomics of the human, macaque and mouse major histocompatibility complex. *Immunology* [online]. 150 (2), pp.127-138.

Shiina, T., Suzuki, S., Ozaki, Y., Taira, H., Kikkawa, E., Shigenari, A., Oka, A., Umemura, T., Joshita, S., Takahashi, O., Hayashi, Y., Paumen, M., Katsuyama, Y., Mitsunaga, S., Ota, M., Kulski, J.K. and Inoko, H. (2012) Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. *Tissue Antigens* [online]. 80 (4), pp.305-316.

Sims, D., Sudbery, I., Illott, N.E., Heger, A. and Ponting, C.P. (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews.Genetics* [online]. 15 (2), pp.121-132.

Single, R.M., Meyer, D., Hollenbach, J.A., Nelson, M.P., Noble, J.A., Erlich, H.A. and Thomson, G. (2002) Haplotype frequency estimation in patient populations: the effect of departures from Hardy-Weinberg proportions and collapsing over a locus in the HLA region. *Genetic Epidemiology* [online]. 22 (2), pp.186-195.

Slichter, S.J., Davis, K., Enright, H., Braine, H., Gernsheimer, T., Kao, K.J., Kickler, T., Lee, E., McFarland, J., McCullough, J., Rodey, G., Schiffer, C.A. and Woodson, R. (2005) Factors affecting posttransfusion platelet increments, platelet refractoriness, and platelet transfusion intervals in thrombocytopenic patients. *Blood* [online]. 105 (10), pp.4106-4114.

Sousa, Luiz Cláudio Demes da Mata, Filho, H.L.A.S., Von Glehn, Cristina de Queiroz Carrascosa, da Silva, A.S., Neto, Pedro de Alcântara dos Santos, de Castro, J.A.F. and Monte, S.J.H.d. (2011) EpHLA: An innovative and user-friendly software automating the HLAMatchmaker algorithm for antibody analysis. *Transplant Immunology* [online]. 25 (4), pp.210-216.

Stafford, P., Garner, S., Rankin, A., Kekomaki, R., Watkins, N. and Ouwehand, W. (2008) A single-nucleotide polymorphism in the human ITGB3 gene is associated with the platelet-specific alloantigen Va^a(HPA-17bw) involved in fetal maternal alloimmune thrombocytopenia. *Transfusion*. 48 pp.1432-1438.

- Stanworth, S.J., Navarrete, C., Estcourt, L. and Marsh, J. (2015) Platelet refractoriness ? practical approaches and ongoing dilemmas in patient management. *British Journal of Haematology* [online]. pp.n/a-n/a.
- Strien, J., Sanft, J. and Mall, G. (2013) Enhancement of PCR amplification of moderate GC-containing and highly GC-rich DNA sequences. *Molecular Biotechnology* [online]. 54 (3), pp.1048-1054.
- Sullivan, M.J., Peterson, J., McFarland, J.G., Bougie, D., Aster, R.H. and Curtis, B.R. (2015) A new low-frequency alloantigen (Khab) located on platelet glycoprotein IIIa as a cause of maternal sensitization leading to neonatal alloimmune thrombocytopenia. *Transfusion* [online]. 55 (6pt2), pp.1584-1585.
- Sutton, L.A., Ljungstrom, V., Mansouri, L., Young, E., Cortese, D., Navrkalova, V., Malcikova, J., Muggen, A.F., Trbusek, M., Panagiotidis, P., Davi, F., Belessi, C., Langerak, A.W., Ghia, P., Pospisilova, S., Stamatopoulos, K. and Rosenquist, R. (2015) Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica* [online]. 100 (3), pp.370-376.
- Tait, B.D. (2011) The ever-expanding list of HLA alleles: changing HLA nomenclature and its relevance to clinical transplantation. *Transplantation Reviews (Orlando, Fla.)* [online]. 25 (1), pp.1-8.
- Trachtenberg, E.A. and Holcomb, C.L. (2013) Next-generation HLA sequencing using the 454 GS FLX system. *Methods in Molecular Biology (Clifton, N.J.)* [online]. 1034 pp.197-219.
- Tu, B., Masaberg, C., Hou, L., Behm, D., Brescia, P., Cha, N., Kariyawasam, K., Lee, J.H., Nong, T., Sells, J., Tausch, P., Yang, R., Ng, J. and Hurley, C.K. (2017) Combining one-step Sanger sequencing with phasing probe hybridization for HLA class I typing yields rapid, G-group resolution predicting 99% of unique full length protein sequences. *HLA* [online]. 89 (2), pp.90-97.
- van Rood, J.J. and Oudshoorn, M. (2008) Eleven million donors in Bone Marrow Donors Worldwide! Time for reassessment? *Bone Marrow Transplantation* [online]. 41 (1), pp.1-9.
- Vassallo, R.R., Fung, M., Rebull, P., Duquesnoy, R., Saw, C.L., Slichter, S.J., Tanael, S., Shehata, N. and International Collaboration for Guideline Development, Implementation and Evaluation for Transfusion Therapies (2014) Utility of cross-matched platelet transfusions in patients with hypoproliferative thrombocytopenia: a systematic review. *Transfusion* [online]. 54 (4), pp.1180-1191.
- Vassallo, R.R. and Norris, P.J. (2016) Can we "terminate" alloimmune platelet transfusion refractoriness? *Transfusion* [online]. 56 (1), pp.19-22.
- Veal, C.D., Freeman, P.J., Jacobs, K., Lancaster, O., Jamain, S., Leboyer, M., Albanes, D., Vaghela, R.R., Gut, I., Chanock, S.J. and Brookes, A.J. (2012) A mechanistic basis for amplification differences between samples and between genome regions. *BMC Genomics* [online]. 13 pp.455-2164-13-455.
- Voorter, C.E.M., Gerritsen, K.E.H., Groeneweg, M., Wieten, L. and Tilanus, M.G.J. (2016) The role of gene polymorphism in HLA class I splicing. *International Journal of Immunogenetics* [online]. 43 (2), pp.65-78.

- Wang, C., Krishnakumar, S., Wilhelmy, J., Babrzadeh, F., Stepanyan, L., Su, L.F., Levinson, D., Fernandez-Vina, M.A., Davis, R.W., Davis, M.M. and Mindrinos, M. (2012) High-throughput, high-fidelity HLA genotyping with deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America* [online]. 109 (22), pp.8676-8681.
- Wang, Z.C., Smith, A.G., Yunis, E.J., Selvakumar, A., Ferrone, S., McKinney, S., Lee, J.H., Fernandez-Vina, M. and Hansen, J.A. (2002) Molecular characterization of the HLA-Cw*0409N allele. *Human Immunology* [online]. 63 (4), pp.295-300.
- Wendt, F.R., Zeng, X., Churchill, J.D., King, J.L. and Budowle, B. (2016) Analysis of Short Tandem Repeat and Single Nucleotide Polymorphism Loci From Single-Source Samples Using a Custom HaloPlex Target Enrichment System Panel. *The American Journal of Forensic Medicine and Pathology* [online]. 37 (2), pp.99-107.
- West, A. (2016) *Histocompatibility and Immunogenetics Diagnostic Service User Guide 2015-16*. Available from: <http://hospital.blood.co.uk/media/28854/inf136.pdf> .
- Wiebe, C. and Nickerson, P. (2016) Strategic Use of Epitope Matching to Improve Outcomes. *Transplantation* [online]. 100 (10), pp.2048-2052.
- Wienzek-Lischka, S., Konig, I.R., Papenkort, E.M., Hackstein, H., Santoso, S., Sachs, U.J. and Bein, G. (2017) HLA-DRB3*01:01 is a predictor of immunization against human platelet antigen-1a but not of the severity of fetal and neonatal alloimmune thrombocytopenia. *Transfusion* [online]. 57 (3), pp.533-540.
- Wienzek-Lischka, S., Krautwurst, A., Frohner, V., Hackstein, H., Gattenlohner, S., Brauninger, A., Axt-Fliedner, R., Degenhardt, J., Deisting, C., Santoso, S., Sachs, U.J. and Bein, G. (2015) Noninvasive fetal genotyping of human platelet antigen-1a using targeted massively parallel sequencing. *Transfusion* [online]. 55 (6 Pt 2), pp.1538-1544.
- Wihadmadyatami, H., Heidinger, K., Roder, L., Werth, S., Giptner, A., Hackstein, H., Knorr, M., Bein, G., Sachs, U.J. and Santoso, S. (2015) Alloantibody against new platelet alloantigen (Lap(a)) on glycoprotein IIb is responsible for a case of fetal and neonatal alloimmune thrombocytopenia. *Transfusion* [online]. 55 (12), pp.2920-2929.
- Williamson, L.M. and Devine, D.V. (2013) Challenges in the management of the blood supply. *The Lancet* [online]. 381 (9880), pp.1866-1875.
- Wittig, M., Anmarkrud, J.A., Kassens, J.C., Koch, S., Forster, M., Ellinghaus, E., Hov, J.R., Sauer, S., Schimmler, M., Ziemann, M., Gorg, S., Jacob, F., Karlsen, T.H. and Franke, A. (2015) Development of a high-resolution NGS-based HLA-typing and analysis pipeline. *Nucleic Acids Research* [online]. 43 (11), pp.e70.
- Xia, W., Ye, X., Xu, X., Ding, H., Liu, J., Deng, J., Chen, Y., Shao, Y., Wang, J., Li, H., Fu, Y. and Santoso, S. (2014) Two cases of platelet transfusion refractoriness and one case of possible FNAIT caused by antibodies against CD36 in China. *Transfusion Medicine (Oxford, England)* [online]. 24 (4), pp.254-256.
- Xu, X., Zhu, F., Ying, Y., Tao, S., Hong, X. and Yan, L. (2009) Simultaneous genotyping of human platelet antigen- to 17w by polymerase chain reaction sequence-based typing. *Vox Sanguinis*. 97 pp.330-337.

- Yang, D., Liang, H., Gao, Y., Lin, S., He, Z., Gao, J., Sun, H., Li, Q., Ma, X. and Ou, X. (2017) Noninvasive fetal genotyping of paternally inherited alleles using targeted massively parallel sequencing in parentage testing cases. *Transfusion* [online].
- Zaimoku, Y., Takamatsu, H., Hosomichi, K., Ozawa, T., Nakagawa, N., Imi, T., Maruyama, H., Katagiri, T., Kishi, H., Tajima, A., Muraguchi, A., Kashiwase, K. and Nakao, S. (2017) Identification of an HLA class I allele closely involved in the autoantigen presentation in acquired aplastic anemia. *Blood* [online]. 129 (21), pp.2908-2916.
- Zhang, J., Chiodini, R., Badr, A. and Zhang, G. (2011) The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics = Yi Chuan Xue Bao* [online]. 38 (3), pp.95-109.
- Zhang, X., Ma, D., Zou, W., Ding, Y., Zhu, C., Min, H., Zhang, B., Wang, W., Chen, B., Ye, M., Cai, M., Pan, Y., Cao, L., Wan, Y., Jin, Y., Gao, Q. and Yi, L. (2016) A rapid NGS strategy for comprehensive molecular diagnosis of Birt-Hogg-Dube syndrome in patients with primary spontaneous pneumothorax. *Respiratory Research* [online]. 17 (1), pp.64-016-0377-9.

9. Publications 2013-2017

9.1 Papers

Davey, S., Ord, J., Navarrete, C., Brown, C., (2017) HLA-A, -B and -C allele and haplotype frequencies defined by next generation sequencing in a population of 519 English blood donors, *Human Immunology*, doi:http://dx.doi.org/10.1016/j.humimm.2017.04.001

Davey, S., Navarrete, C. and Brown, C. (2017), Simultaneous human platelet antigen genotyping and detection of novel single nucleotide polymorphisms by targeted next-generation sequencing. *Transfusion*. doi:10.1111/trf.14092

S Davey (2015) Next generation sequencing – Tissue (HLA) typing adult and cord blood stem cell donors for transplantation. *Blood and Transplant Matters*, January 2015

S Davey (2014) Review: Next generation sequencing – considering the ethics. *International Journal of Immunogenetics* **41**, 457-462

9.2 Abstracts/Orals/Posters

Sue Davey, Monica Kyriacou, Jasmine Rizvi, Arthi Anand & Colin Brown (2016) The impact of automation on DNA library preparation for whole gene HLA typing by Next Generation Sequencing (Poster) *BSHI AGM*

S Davey, C Navarrete & C Brown (2016) HLA Epitope matched platelets: is high resolution HLA typing essential? (Poster) *ESPGI, Stockholm*

S Davey, C Navarrete & C Brown (2016) HPA genotyping using targeted next generation sequencing (Poster) *ESPGI, Stockholm*

Sue Davey, Zareen Deplano, Monica Kyriacou, John Ord, Lydia Quaye, Ying Li, David Winstone, Guy Parkes, Cristina Navarrete & Colin Brown (2016) Implementation of whole gene Next Generation Sequencing for HLA typing unrelated adult and cord blood stem cell donors of the British Bone Marrow Registry (Poster) *IDRC/WMDA Singapore*

Lydia Quaye, John Ord, Sue Davey, Monica Kyriacou, Zareen Deplano, Trevor Green, Katy Derbyshire, Colin Brown and Cristina Navarrete (2015) Methods for evaluating and modelling next generation sequencing quality parameters (Poster) *BSHI AGM*

Davey, S., Brown, C. & Navarrete (2015) Next Generation Sequencing to facilitate donor selection using HLA epitope-based matching for patients immunologically refractory to random platelet transfusions (Oral) *ISBT, London*

Sue Davey, Zareen Deplano, John Ord, Monica Kyriacou, Lydia Quaye, Lisa Creary, Colin Brown & Cristina Navarrete (2015) Validation and implementation of next generation sequencing for HLA typing unrelated adult and cord blood stem cell donors of the British Bone Marrow Registry (Poster) *EFI, Geneva – Awarded best poster*

Sue Davey, Zareen Deplano, John Ord, Monica Kyriacou, Lydia Quaye, Cristina Navarrete & Colin Brown (2015) Full gene sequencing by next generation sequencing highlights the limitations of a targeted exon approach for HLA typing (Poster) *EFI, Geneva*

Z. Goburdhun, L.Creary, S.Davey, J.Ord, M.Kyriacou, C.V. Navarrete and C.J.Brown (2014) Automated HLA typing using Next Generation Sequencing (Oral) *BSHI AGM*

Sue Davey, Mark Nightingale, John Ord, Zareen Goburdhun, Monica Kyriacou & Colin Brown (2014) 'Rare' HLA alleles detected by next generation sequencing (Oral) *BSHI AGM*

Sue Davey, Zareen Goburdhun, Lisa Creary, John Ord, Monica Kyriacou, Cristina Navarrete & Colin Brown (2014) Control of change to next generation sequencing: considerations for HLA typing (Poster) *BSHI AGM*

10. Appendix

10.1 Response from FREC

Page 1 of 2

Davey Sue

From: Julie Woodley [Julie.Tonks@uwe.ac.uk]
Sent: 17 March 2014 13:54
To: Davey Sue; Leigh Taylor
Cc: Michael Ladomery; Brown Colin & PA
Subject: RE: Enquiry re: requirement for ethical approval

Dear Sue

Looking through this I would also say that this does not require FREC ethical approval for the same reasons you outline below. You will still need Health and Safety and the usual permissions and sample storage regs but not formal ethics as such. Hope that clarifies
Julie

Dr Julie Woodley
Senior Lecturer -Allied Health
Chair of HAS Faculty Research Ethics Committee
Rm 2K01 Glenside campus UWE
Faculty of Health and Applied Science
Stapleton
Bristol BS16 1DD
+44 (0)117 3288528
UREC Research Ethics Website <http://www1.uwe.ac.uk/research/researchethics>

HAS CPD Website: <http://cpd.hsc.uwe.ac.uk/welcome.aspx>

UWE AHP Web-page: <http://www1.uwe.ac.uk/hls/ahp>

HAVE YOUR SAY  Complete the NSS or SES!
27th January-30th April

From: Davey Sue [mailto:Sue.Davey@nhsbt.nhs.uk]
Sent: 17 March 2014 10:30
To: Leigh Taylor
Cc: Julie Woodley; Michael Ladomery; Brown Colin & PA
Subject: Enquiry re: requirement for ethical approval

Dear Leigh,

I wonder if you could clarify whether or not I need to apply for ethical approval for my research project, part of my Doctorate in Biomedical Science?

My thesis will involve the development of novel genotyping techniques for HLA and HPA (see enclosed RD1 form for more details); all practical work will be performed at NHS Blood and Transplant (NHSBT), my employer & sponsor.

Method development and validation will include the use of DNA obtained from NHSBT archives, previously extracted from blood collected from platelet donors. Should these new methods be successful, I may also be using it to HPA test archive DNA from patients previously tested by NHSBT for a condition (NAIT), where HPA genotyping is required and was performed previously using current techniques.

I am of the understanding that I do not need ethical approval to complete this work for the following reasons:

1. This project will not involve human participants or Human Tissue (as defined by the HTA)
2. The on line Health Research Authority decision tree indicates that NHS REC approval is not required

17/03/2014

3. Platelet donors have provided qualified consent for the use of their blood collected during the donation for HLA and HPA typing, as well as generic consent for use of their samples for service development (see enclosed INF48, FRM421, INF234, INF256).

4. Patients previously investigated for NAIT have provided qualified consent for their samples to be HPA genotyped (see enclosed INF283 & INF136 page 9).

I would therefore be grateful if you could confirm whether or not I do need to seek formal approval from the FREC.

Yours sincerely

Sue Davey MSc DipRCPATH
Laboratory Operations Manager
Dept of H&I
NHSBT
Charcot Road
Colindale
London NW9 5BG

Tel: 0208 957 2997
Fax: 0208 957 2717
e-mail: sue.davey@nhsbt.nhs.uk
UWE e-mail: Susan3.Davey@live.uwe.ac.uk

10.2 Donor health check

Version 5 01/05/12 DONATION RECORD - REGULAR DONOR				:	:
Surname		Group			
Forenames					
Title	DOB	Sex	Donor No		
Address		01 Donated Outcome of Attendance		Donation No	
Tel No		02 Low Hb		Code	
Occupation		03 Other samples only		BBMR	
Tel No. (day)		04 No Donation - No numbers			
		05 No Donation - Numbers issued			
		06 No Donation - Pack labelled			
		PACK HOLD CODE			
		Signature.....			
Total	Attendance	Outcome		Date	
Award				Panel	
Award Given Y <input type="checkbox"/> N <input type="checkbox"/>				Sub Panel	
		P:			
		T:			
DONATION TYPE: Whole Blood <input type="checkbox"/>		Platelets Only <input type="checkbox"/>	Platelets & Plasma <input type="checkbox"/>	Autologous Tears <input type="checkbox"/>	
Plasma for FFP <input type="checkbox"/>		Red Cells Only <input type="checkbox"/>	Granulocytes <input type="checkbox"/>	Other <input type="checkbox"/>	
I have completed my tasks in accordance with SOPs		Signatures		Incident Record	
Registration		Malaria (MA)		TCruzi (TC)	
Health Check <input type="checkbox"/> Hb Pass <input type="checkbox"/> Low <input type="checkbox"/>		R <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		R <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Session Hb result g/L		L <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		L <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Machine/Pack set up		T <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>			
LA	VP	M <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		Discretionary Testing Codes	
Needle Removal		V <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		R = Risk/Exposure	
Dressing/PD advice		F <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		L = Long Stay	
Final Pack/Sample Check		West Nile Virus (WN)		T = Transfused before 1st January 1988 in a malaria area	
Reconciliation		R <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		M = Malaria Symptoms	
		W <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		V = Re-visited malaria area after previous illness/plasma(M)	
				F = Unexplained Fever	
				W = West Nile Virus Illness/symptoms	
				S = Skin Picking	
				J = Jaundice/HBV	
				C = Contact with HBV	
				E = Endoscopy 4-6 months	

Donor Consent - to be signed in the presence of a member of NHSBT staff

- I have today read and understood the Welcome Booklet. I have been given the opportunity to ask questions and they have been answered.
- To the best of my knowledge I am not at risk of infection or of transmitting the infections listed in the Welcome Booklet.
- I agree that my blood donation will be tested for HIV and other conditions listed in the Welcome Booklet. I understand that if my donation gives a positive result for any of these tests I will be informed and asked to participate in a post-test discussion.
- I understand the nature of the donation process and the possible risks involved as explained in the Welcome Booklet.
- I agree to NHS Blood and Transplant holding information about me, my health, my attendances and donations, to contacting my doctor for further information and using my donation for the purposes explained in the Welcome Booklet.
- I give my blood to NHS Blood and Transplant to be used for the benefit of patients. This may be by direct transfusion to a patient or for other purposes as explained in the Welcome Booklet.

Date Donor Signature _____

page 2 of 2 01/05/12 MMZ/2011

10.3 Welcome booklet

Blood Safety – Keep Blood Transfusion Safe
To be read by all donors.

Thank you for wanting to give blood. Blood transfusion is very safe for patients because we take great care in choosing donors. Help us to keep it safe by reading this leaflet before you give blood. It will help you to decide if you should give blood.

If you are not sure, please ask to talk confidentially to a nurse. If you need more time, it is OK to leave. You can talk confidentially to a doctor or nurse by telephoning 0300 123 23 23.

All donors are tested for infection with hepatitis B and C, HIV, Syphilis and HTLV.

These tests do not always show if you are infected. This is why we must take care in choosing donors and why you must not give blood to see if you are infected.

If you cannot give blood, we will explain why. If we find any of these infections in your blood, we will always tell you and offer advice and help. **Please read this leaflet carefully, if you need help, do please ask.**

Thank you.
NHS Blood and Transplant

You must not give blood if:

- You think you need a test for HIV/AIDS or hepatitis.
- You are HIV positive.
- You have hepatitis B.
- You have hepatitis C.
- You have HTLV.
- You are a man who has in the last 12 months had oral or anal sex with another man, even if you used a condom or other protective.
- You have ever received money or drugs for sex.
- You have ever injected, or been injected with, drugs, even a long time ago or only once. This includes body-building drugs. You may be able to give if a doctor prescribed the drugs. Please ask.

You must not give blood for at least 12 months after sex (even if you used a condom or other protective with:

- A partner who is, or you think may be, HIV positive.
 - A hepatitis B carrier.
 - A hepatitis C carrier.
- HTLV positive.
- If you are a woman a man who has ever had oral or anal sex with another man, even if they used a condom or other protective. There are exceptions so please ask.
- A partner who has ever received money or drugs for sex.
- A partner who has ever injected, or been injected with, drugs, even a long time ago or only once. This includes body-building drugs. You may be able to give if a doctor prescribed the drugs. Please ask.
- A partner who has, or you think may have been, sexually active in parts of the world where HIV/AIDS is very common. This includes most countries in Africa. There are exceptions, so please ask.

Warning about HIV or Hepatitis – Who can you tell?
There are several people you can talk to, such as the nurse on duty today or your GP. You can also call the National AIDS Helpline, which is specially set up for advice and support on HIV. Ask them about 0300 123 23 23.

Please read this information if you are a Regular Donor
Please also read the Blood Safety information on the previous page.

BEFORE YOU DONATE:

- You need to fill in the 'Health Check Questionnaire'.
- Have a short health screening interview, which includes your lifestyle and travel history.
- Read and sign the 'Donor Consent'.

This is to be sure that you are well enough to donate blood and it's safe to give to a patient. If you are not able to give blood, we'll tell you why.

You may be asked if it's okay to contact your doctor for further information.

If you don't wish to donate, you can leave at any time.

YOUR BLOOD IS TESTED EACH TIME FOR:

- Blood group.
- HIV, hepatitis B, hepatitis C, syphilis and HTLV. If there is a positive result for any of these, we will contact you and offer further advice. Your donation will have to be destroyed.
- Extra tests are done as required e.g. malaria and sickle cell.

BLOOD USE
The vast majority of donations are used for patients e.g. after blood loss or cancer treatment. However, occasionally blood donations may be used for laboratory work, education, training, research and development by us or other organisations.

RECORDS
Information on donors, donations and test results is kept for at least 30 years. We are committed to protecting donor confidentiality and to meeting our responsibilities under the Data Protection Act 1998. If you want to know more, please read 'Data Protection: A Guide for Donors.'

POSSIBLE CONSEQUENCES OF GIVING BLOOD
Please read 'Possible Consequences of Giving Blood' on page 3.

LET US KNOW
If you become unwell (except for a cold or cold sore) up to two weeks after donating, or if you think there is a reason that your blood should not be used, just call our Donor Helpline on 0300 123 23 23 as soon as you can.

Thank you for reading the donor information.
Please book your next appointment before you leave the session.

blood.co.uk
0300 123 23 23

9622475 Effective Date: 07.11.11 BLS55P 11/12/02



10.4 Page 9 of User guide for H&I diagnostic services

INFORMATION DOCUMENT INF138/2

Effective: 10/05/2013

User Guide for Histocompatibility and Immunogenetics Diagnostics Services

CONSENT

To comply with the Human Tissue Act legislation (Human Tissue Act, 2004), it is the responsibility of the requester to ensure that any patient or donor has been informed of, and has consented to, the tests being requested.

NHSBT may ask the requester to provide a copy of this information. Patients/donors should be informed that any residual material of a sample may be stored as part of required archiving protocols or to enable further investigation for the benefit of the individual. They also must be informed that excess surplus material may be used anonymously for quality control purposes, service development or education, and / or ethics committee approved research projects.

NHSBT H&I laboratories have developed a series of patient information leaflets to assist healthcare professionals to obtain informed consent for diagnostic testing. The leaflets explain what happens to their samples and why the tests are undertaken. In addition, there is a brief explanation of Histocompatibility & Immunogenetics investigations. The leaflets are available to download from the NHSBT hospital website and/ or hard copies can be ordered directly from your local H&I laboratory. The link for patient information leaflets is:

http://hospital.blood.co.uk/diagnostic_services/handi/h&i%20patient%20information%20leaflets/

Table 4: Summary of H&I patient leaflets

Patient information leaflet		Laboratory
Histocompatibility testing for kidney transplant donors	INF253	Local solid organ H&I laboratory
Histocompatibility testing for kidney transplant patients	INF255	
Histocompatibility testing for cardiothoracic transplant patients	INF254	
Histocompatibility testing for platelet transfusion patients	INF256	Local H&I laboratory
Histocompatibility testing for stem cell transplant patients	INF257	
Histocompatibility testing for possible donors or relatives of stem cell transplant patients	INF258	
Immunogenetic markers and diagnosing diseases	INF259	Filton
Heparin-induced thrombocytopenia (HIT) : Your background guide to HIT and the associated laboratory testing	INF260	
Information for mothers about neutrophil blood groups and Neonatal alloimmune Neutropenia (NAIN)	INF261	
Platelet groups & antibodies in pregnancy	INF263	

PACKAGING AND TRANSPORT

It is the responsibility of the sender to ensure that all samples are packaged in accordance with the current European agreement concerning Carriage of Dangerous Goods by Road Regulations, packaging instructions 650, to prevent breakage or spillage in transit. The outside of the box or package containing the samples must be clearly addressed to the appropriate Blood Centre where the testing laboratory is based. This may not be your local blood centre. NHSBT reserves the right to refuse to handle any samples which are inappropriately packaged or labelled; customers sending unsatisfactorily packaged samples will be contacted.

Pre-printed address labels can be provided on request from Customer Services. For advice from the Health and Safety Executive (HSE) on packaging for posting samples see: <http://www.hse.gov.uk/biosafety/blood-borne-viruses/transportation-of-infectious-substances.htm>

(Template Version 07/10/20)

Author(s): Adam West

Page 9 of 40

10.5 INF256 – Histocompatibility testing for platelet transfusion patients

What happens to my sample?

When we no longer need your samples for testing, or we have more samples than we need, the law allows us to use these anonymously for quality control (making sure our tests are working correctly), research (depending on whether this is approved by an ethics committee) or introducing new procedures, or for educating and training doctors, nurses, scientists and other professionals working in healthcare. This helps us maintain accurate testing procedures and improve our knowledge, and so provide the best possible care for all patients.

However, if you do not want us to use your samples for any of the purposes above, you must tell your doctor or the person taking your blood (or both). We will respect your wishes and dispose of any samples we no longer need.



This leaflet is produced by NHS Blood and Transplant, a special health authority within the NHS.

NHS
Blood and Transplant

Histocompatibility testing for platelet transfusion patients

Patient
Information

INF256/1

Effective from 01/05/08

What is histocompatibility testing?

Histocompatibility testing helps decide if we can find a suitable donor to provide specific platelets to patients who do not respond well to some platelet transfusions.

Most platelet transfusions use platelets from the stocks in the blood bank, and in most cases these work well. However, in some cases, the patient's body destroys the platelets very soon after the transfusion, and so the platelets are not replaced as they are meant to be. There are several reasons why this may happen. One of these is because the patient has antibodies in their blood which react with HLAs (human leucocyte antigens) or HPAs (human platelet antigens).

You inherit your HLA type (also called tissue type) and HPA type from your parents. HLAs are found on the surface of most cells in the body. There are many different HLA types and so most people are different from each other. Because of this, it is likely that most of the platelets you normally receive will come from donors whose HLA type is different to your own. This can be a problem if you have HLA antibodies in your blood. HLA antibodies are made by your body's immune system and can destroy platelets you are given from the transfusion. You may have HLA antibodies if you have previously had a blood transfusion, or if you have been pregnant.

We may need to carry out histocompatibility testing if you do not respond well to some platelet transfusions. If you respond poorly to a platelet transfusion, we will give your blood sample to a histocompatibility and immunogenetics (H&I) laboratory to find out whether this is the reason you do not respond well to standard platelet transfusions.

There are two tests which the laboratory will carry out first – HLA typing and HLA antibody testing. Some patients will also need HPA typing and HPA antibody testing.

HLA and HPA typing

We will find your HLA type by testing a sample of your DNA, which we get from your blood cells. The laboratory will store a sample of your DNA in case they need to carry out more tests in the future. In some cases, the laboratory will also use your DNA sample to find your HPA type.

We will give the results of your typing to your consultant haematologist. We may also send copies of your results to the other health-care staff who are treating you (for example, specialist nurses), and to any other hospitals you may transfer to in the future. The H&I laboratory will also keep copies of the results.

HLA and HPA antibody testing

We need a separate blood sample to test for HLA antibodies. HLA antibodies can appear, disappear and reappear over time, so you may need to provide more samples if you need regular platelet transfusions over a long period of time. We will use these samples to check whether any antibodies have formed in your blood.

Some patients may make HPA antibodies as well as, or instead of, HLA antibodies. In these circumstances, the patients may respond poorly to HLA-selected platelet transfusions (see below) as a result of the HPA antibodies which are also found on platelets.

We will give the results to your consultant, who will work with us to decide if you need to receive specially selected platelets.

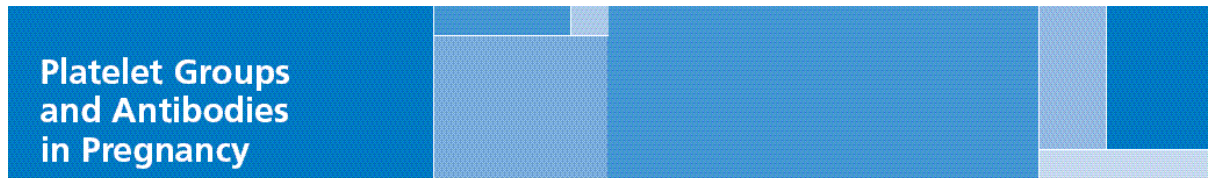
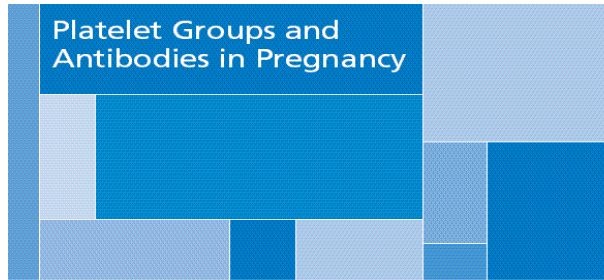
HLA-selected platelets

If the results of your tests show that you have HLA antibodies, you may need to receive HLA-selected platelets. If this is the case, our H&I staff will use the results of your HLA typing and antibody testing to choose platelets from blood donors whose HLA type is suitable for you. This will mean that your HLA antibodies should not react with the specially selected platelets. In most cases, your response to platelet transfusions should improve. In rare cases, this may not happen and you may need more tests to see if there is something else which is affecting these transfusions.

HPA-selected platelets

If the results of tests for HPA antibodies and HPA typing show why you are not responding to HLA-selected platelets, you may need to receive HPA-selected platelets. In some cases, you may need to receive HLA- and HPA-selected platelets that will not be affected by the HLA and HPA antibodies in your blood. As a result, your response to platelet transfusions should improve.

10.6 INF283 Platelet groups and antibodies in pregnancy



Platelet Groups and Antibodies in Pregnancy

This leaflet explains the blood test results that you have been given and what this means to you and your baby. It contains information about the significance of platelet groups and antibodies in a condition known as Neonatal Alloimmune Thrombocytopenia (NAIT)

What are platelets?

Platelets are the smallest type of cells that circulate in the blood. They are important in preventing and stopping bleeding. Sometimes the blood contains fewer platelets than normal, a condition known as Thrombocytopenia.

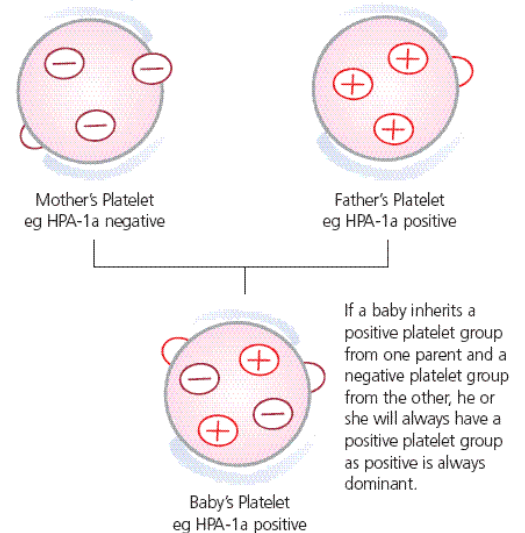
What is a platelet group?

Every platelet has natural proteins on its surface, known as human platelet antigens, which make up your platelet group. Your platelet group is inherited from both your parents. There are many different types of platelet group and scientists are discovering new types of group all the time. The most common groups amongst mothers are HPA-1a and HPA-5b.

How Platelet Groups are Inherited

The diagram opposite shows how platelet groups are inherited. The example given shows a mother who is HPA-1a negative and a father who is HPA-1a positive. However, other platelet groups could be inherited in the same way.

How Platelet Groups are Inherited



What are antibodies?

Antibodies are an important part of the body's immune system, which help us fight disease. They are formed when the body's immune system comes into contact with a 'foreign' substance, for example a virus, vaccine or a different blood group.

How are platelet antibodies formed during pregnancy?

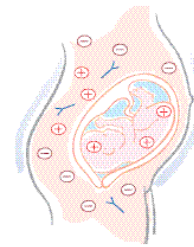
During pregnancy, platelet antibodies are formed if a baby has a different platelet group to that of its mother, and some of the baby's platelets pass into the mother's bloodstream. The mother's body reacts to the presence of the baby's platelet group, and makes antibodies against the baby's platelets.

How could platelet antibodies affect pregnancy?

It is possible that these antibodies could pass through the placenta into the baby's bloodstream, come into contact with the baby's platelets and damage them. This causes the number of platelets in the baby's blood (known as the platelet count) to decrease. This condition is known as Neonatal Alloimmune Thrombocytopenia or NAIT, and occurs in approximately 1 in 1000 pregnancies. The diagram opposite explains this process.

How antibodies form in the mother and how they affect the baby

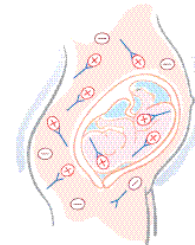
- ⊖ Mother's platelets
- ⊕ Baby's platelets
- ⌞ Mother's antibody



A. A baby might have a different platelet group from its mother, due to the fact that half the platelet group is inherited from its father.



B. Sometimes some of the baby's blood crosses the placenta into the mother's blood. In very rare cases her body reacts against the baby's platelets because they are 'different' or not part of her body. This causes proteins called antibodies to be produced, in order to remove the baby's platelets from the mother's blood.



C. These antibodies cross the placenta into the baby's blood, come into contact with the baby's platelets and damage them. This reduces the number of healthy platelets in the blood, causing Neonatal Alloimmune Thrombocytopenia.