Factors determining the composition of a public cord blood stem cell bank including HLA diversity

by

Juanita Mellet

A dissertation submitted in fulfilment of the requirements for the degree of Magister Scientiae (M.Sc) Immunology

in

The Department of Immunology

Faculty of Health Sciences

University of Pretoria

2013

Supervisor: Prof. MS Pepper

ABSTRACT

The human leukocyte antigen (HLA) is the most polymorphic region in the human genome and accounts for more than 10% of human diversity. This region plays an important role in matching donors and recipients for transplantation. The South African Bone Marrow Registry (SABMR) does not reflect the demographics of the South African population. The large number of polymorphisms resulting from HLA diversity in the Black South African population and their limited representation in the SABMR reduce the chances of finding adequate matches between donors and recipients in this group. Umbilical cord blood is an alternative to bone marrow for the treatment of fatal diseases. Less strict HLA matching is required due to the naive nature of the T cells in cord blood. A public umbilical cord blood bank is a necessity in trying to cater for the diverse population in South Africa. However, the ethnic diversity of the South African population poses a great challenge in constituting a public umbilical cord blood bank that is representative of the entire population. The Roche designed next generation sequencing (NGS) high resolution (HR) HLA typing kit enables sequencing of additional HLA exons and could improve the degree of matching between individuals to ultimately decrease adverse reactions. An extensive study of the literature was performed to establish the demographics, linguistics, and HLA diversity of the South African population to determine how a public cord blood bank should be constituted. In addition, HLA genotyping was performed by 454 NGS on 20 samples that had previously been HLA typed by conventional methods. The 454 NGS technique made use of a Roche designed medium and high resolution HLA typing kit to genotype the samples. It was possible to assign accurate genotypes to 95.5% of the loci of interest for the total number of 20 samples using the MR kit, compared with 98.5% using the HR kit. In conclusion, the present study indicates the extreme HLA diversity in the South African population, and therefore, recommends constituting the first public umbilical cord blood bank in Gauteng on the basis of race or major ethnic groupings. A minimum number of 10 000 cord blood units is needed to initiate the bank. Furthermore, the 454 NGS platform together with the HR HLA typing kit display potential as an alternative method to be used in a public cord blood bank, as well as routine clinical and diagnostic laboratories, to ultimately improve HLA matching between donors and recipients.

ACKNOWLEDGEMENTS

To God, my Saviour.

My Supervisor, Professor Michael Pepper, for the opportunity that you have given me and for your continued guidance and assistance throughout this study period. I truly appreciate the commitment and time that went into this project.

My immediate family and friends, your encouragement, support, and love throughout the past few years have made me reach my goal. I could not have done this without you!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SYMBOLS

α	Alpha
β	Beta
γ	Gamma
µl	Microlitre
A	Adenine
Ad12	Adenovirus 12
AE	Elution buffer
AIDS	Acquired immune deficiency syndrome
APC	Antigen presenting cell
ATP	Adenosine triphosphate
B_2M	B_2-microglobulin
bp	Base pair
C	Cytosine
CIITA	Class II transcription activator
CD	Cluster of differentiation
CLIP	Class II-associated invariant-chain peptide
CMV	Cytomegalovirus
ddH_2O	Double distilled water
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
E	Exon
emPCR	Emulsion PCR
ER	Endoplasmic reticulum

EtOH	Ethanol
g	Gram
G	Guanine
G-CSF	Granulocyte colony-stimulating factor
Gly	Glycine
GVHD	Graft-versus-host disease
HBV	Hepatitis B virus
HGP	Human genome project
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HR	High resolution
HSC	Hematopoietic stem cell
IFN-γ	Interferon gamma
Ig	Immunoglobulin
IMGT	International immunogenetics information system
Kb	Kilobase
kg	Kilogram
LD	Linkage disequilibrium
M	Molar
Mb	Megabase
$MgCl_2$	Magnesium chloride
MHC	Major histocompatibility complex
MID	Multiplex identifier
min	Minute
ml	Millilitre
mM	Millimolar

MR	Medium resolution
MRC	Medical Research Council
MS	Multiple sclerosis
mtDNA	Mitochondrial DNA
ng	Nanogram
NGS	Next generation sequencing
NHLS	National Health Laboratory Service
NK	Natural killer
nm	Wavelength
PBSC	Peripheral blood stem cell
PCR	Polymerase chain reaction
PPi	Pyrophosphate
PTP	Picotiterplate
R^2	Pearson coefficient of determination
RA	Rheumatoid arthritis
RER	Rough endoplasmic reticulum
RFLP	Restriction fragment length polymorphism
RFU	Relative fluorescence unit
RFX	Regulatory factor X
SABMR	South African Bone Marrow Registry
SBT	Sequence-based typing
sec	Second
SNP	Single-nucleotide polymorphism
SPRI	Solid phase reversible immobilisation
SSOP	Sequence-specific oligonucleotide probes
SSP	Sequence-specific primers

STR	Short tandem repeats
T	Thiamine
TAP	Transporter associated with antigen processing
Taq	*Thermus Aquaticus* polymerase
TB	Tuberculosis
TBE	Tris-Borate-EDTA
T cell	T lymphocyte
TCR	T-cell receptor
TE	Tris-EDTA
TNF	Tumour necrosis factor
U	Unit
UCB	Umbilical cord blood
UCT	University of Cape Town
UP	University of Pretoria
USA	United States of America
UV	Ultraviolet
V	Volt
Val	Valine

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The South African population is known as 'the rainbow nation', which is appropriate for a country with a cultural variety emphasised by 11 official languages. South Africa is one of the most diverse and complex countries on the planet with regard to genetic heterogeneity. The high degree of genetic heterogeneity in Africans and South Africans is a result of both inter- and intra-population variation. Inter-population variation occurs between populations, whereas intra-population variation is between members of a population. Variations within populations are mainly due to bottlenecks, which can be explained on the basis of the 'Out of Africa' theory that suggests that humans migrated in isolated groups from Africa to the rest of the world. Genetic diversity decreases with increased distance from the original population. Therefore, when a population migrates, they will be less diverse than the original population.

The South African Bone Marrow Registry (SABMR) does not reflect the demographics of the South African population (SABMR, 2008), since the majority of donors are Caucasian (72%), 9% Coloured, 5% Asian, and 4% Black South Africans. The increased diversity observed in the Black South African population and the under-represented number of these individuals belonging to the SABMR poses a significant challenge in obtaining human leukocyte antigen (HLA) matching donors for these individuals. A public cord blood bank in South Africa could increase the chances of obtaining HLA matching donors for African individuals, locally and globally. Bone marrow transplantation requires matching at 10 HLA alleles (five from each parent) with a 9/10 match between donors and recipients. Umbilical cord blood on the other hand involves matching of six HLA alleles (three from each parent) and a 4/6 match between donors and recipients is considered to be acceptable for transplantation. The naive nature of the T cells in cord blood results in higher tolerance to HLA mismatches. For this reason, there is a greater probability of obtaining an HLA matching donor for cord blood than with bone marrow.

The characteristics of umbilical cord blood render it a suitable alternative to bone marrow- and peripheral blood-derived stem cells for cell-based therapies (*e.g.* bone marrow transplantation), which has led to the establishment of umbilical cord blood stem cell banks

across the globe. Public banks involve altruistic donation in which fetal blood is harvested from the placenta via the umbilical vein and stored in a general facility for use in unrelated (allogeneic) cord blood transplantations. Private cord blood banks were established for autologous use by the donor, either in childhood, or later in life, in the case of disease development. The likelihood that a child will be able to make use of autologous cord blood is very small. The probability of possessing an HLA identical sibling is 25%, compared with 40% for a 4/6 and 75% for a 3/6 match (Beatty *et al*., 2000). Directed cord blood banking is done for the purpose of donating to siblings that are known to have a disease that can be treated by bone marrow transplantation.

Umbilical cord blood is valuable for the reconstitution of hematopoiesis in children with malignant as well as non-malignant disorders, such as in the treatment of hematological malignancies, non malignant blood disorders, and metabolic disorders. Umbilical cord blood is a rich source of hematopoietic stem cells (Rocha *et al*., 2004; Kurtzberg *et al.*, 1996) which undergoes engraftment following transplantation (Broxmeyer *et al.,* 1989). Clinical observations have indicated that acute graft-versus-host disease (GVHD) occurs less frequently in patients who receive cord blood for transplantation than in individuals who receive bone marrow (Cairo and Wagner, 1997). Stem cells from cord blood are known to be "immune naive" and differ from stem cells from the bone marrow due to minimal previous exposure to antigens (Chalmers *et al.,* 1998) and also contain fewer helper T cells (Loetscher *et al.,* 1998). Although cord blood-derived stem cells have certain advantages, the number of cells obtained from a cord blood sample is limited, which also limits the age/size of the patient that can be transplanted. A minimum of 2.5-3.0 x 10^7 nucleated cells are required per kilogram body weight (Migliaccio *et al.,* 2000; Zhang *et al*., 2012). Suboptimal cell numbers reduce the efficacy of engraftment following transplantation.

There are several advantages to using cord blood instead of bone marrow; (1) cord blood is readily available following birth and does not require an invasive procedure for harvesting; (2) it is possible to cater for a greater diversity due to the ability of cord blood stem cells to tolerate a greater degree of HLA disparity; (3) cells have a high proliferative capacity; and (4) there is a decreased rate of acute GVHD following transplantation.

There are also a number of disadvantages in making use of cord blood for transplantations: (1) there are no additional donor cells available, whereas in the case of bone marrow, it is possible to obtain more blood from the donor; (2) the low number of cells available due to the small volumes of cord blood readily available; (3) increased risk of infection, since cord blood cells are immune naive; and (4) in the case where more than one unit is required for transplantation, all units need to match the recipient as well as each other.

South Africa is a multicultural country where a multiracial population of approximately 52 million currently resides. These include: 39 million Black South Africans (80%); 5 million White South Africans (9%); 4 million Mixed ancestry South Africans (8%), and 1 million Asian South Africans (3%) (Statistics South Africa, 2011). In support of the constitution of a public cord blood bank in South Africa, the bank would have to be representative of the entire South African population. It has been estimated that a minimum number of 10 000 cord blood units would be needed to initiate the bank, of which 8 000 (80%) would have to be representative of Black, 900 (9%) of White, 900 (9%) of Mixed ancestry, and 200 (2%) of Asian South Africans.

The majority of the White South African population originally descended from Europe and speaks Afrikaans and English. The Mixed ancestry population mainly speaks Afrikaans, while the Indian population is English-speaking. The Black South African population has four broad groupings, namely: Nguni (which include Zulu, Xhosa, Ndebele, and Swazi), Sotho-Tswana (which include Southern-, Northern-, and Western Sothos), Tsonga, and Venda. The diversity observed renders the Black population the most diverse population in South Africa. The language that one speaks is often an indication of the ethnicity of an individual. South Africa has 11 official languages; however, there are several other languages being spoken in addition to the official languages. Language is often a reflection of the geographic HLA distribution, which reflects cultural groupings and intermarrying. It is evident that intra-population variation is often more diverse than inter-population variation. It is, therefore, important to determine the diversity within the subgroups of a given population in addition to the diversity between the different populations.

Several studies have considered certain population groups to determine the HLA diversity in the Black South African population (du Toit *et al*., 1988; du Toit *et al*., 1988, p. 42; du Toit *et al.,* 1990a; 1990b). These studies primarily focused on either the Xhosa or the Zulu populations, since they make up a large proportion of the Black South African population in this country. These studies have established that various HLA alleles appear to be restricted to the Xhosa population, while others are exclusive to the Zulu population. Some alleles were observed to occur at a higher frequency in Black South African individuals, while others are rare or completely absent. All of the above-mentioned studies made use of serological methods for HLA determination. A more recent study by Paximadis and co-workers also determined the HLA diversity in South African individuals; however, this study made use of a sequence-based method in order to determine the HLA genotypes of individuals. This study included individuals from all the different linguistic groups residing in South Africa (Paximadis *et al.,* 2011).

1.1. Problem Statement

The ethnic diversity of the South African population poses a great challenge in constituting a public umbilical cord blood bank that is representative of the entire population. The large number of polymorphisms resulting from the HLA diversity within the South African population will reduce the probability of finding adequate matches between donors and recipients.

1.2. Aim

The aim of this study was to determine how a public cord blood bank should be constituted in order to cater for the diverse South African population, which includes race, ethnicity, and HLA diversity (conventional techniques *vs*. next generation sequencing).

1.3. Study Objectives

The present study involved an analysis at four levels:

1. Population demographics

Population statistics were obtained from Statistics South Africa (www.statssa.gov.za) regarding the different population groups in the country. The last census was conducted in 2011 and conveys the most recent population statistics. HLA statistics for the South African population were gathered from research that was conducted during the 1980's and a paper published in 2011, which are the most recent papers published concerning the diversity of HLA in South African individuals. There are also some HLA statistics available on the allele frequency website.

2. Linguistic distribution

Statistics on the linguistic distribution were also gathered from Statistics South Africa, concerning the different languages and the frequency at which they are spoken in the nine provinces. This was helpful in determining where the population groups and subgroups are mainly distributed.

3. Conventional (low to high resolution) techniques for HLA typing

Twenty samples were obtained from the National Health Laboratory Service (NHLS), Department of Immunology, Diagnostic Section, at the University of Cape Town (UCT), courtesy of Professor Clive Gray. Low to high resolution typing had previously been performed on the 20 samples prior to the commencement of this study and the different HLA types were blinded for the purpose of this study. This study has made use of a next generation sequencing (NGS) technology to determine the HLA genotypes of the 20 samples at a high resolution.

4. Next generation sequencing at Medium and High Resolution

This part of the analysis involved the use of the GS GHLA typing kit. DNA amplification was performed in a multiplex manner by making use of specifically designed primers for the various HLA regions (HLA-A, B, C, DQB1, and DRB1) that were investigated during this study.

The kit targets the most hypervariable regions of the HLA genes that include exons 2 and 3 for class I genes, and exon 2 for class II genes. This is defined as medium resolution (MR) genotyping by the manufacturer of the kit. High resolution (HR) genotyping includes, in addition, exon 4 for the class I genes and supplementary regions of the class II genes. Medium and high resolution genotyping kits made use of different PCR plates. This study made use of two different primer sets for the different genotyping resolutions. The MR and HR resolution, as described by the manufacturers of the plate, has caused some confusion in that MR and HR are not referring to the different exons being sequenced. It is generally accepted that low to medium resolution techniques (SSOP and SSP) identifies two (A*02) to four digits but often with multiple possibilities at the four-digit level (A*02:02/02:05/02:09/02:40) due to the ambiguous nature of the HLA genes. High resolution techniques identify four digits and higher (A*02:01:01:01), since it reveals the entire nucleotide sequence. Therefore, it is the highest level of resolution. The 'medium' and 'high' resolution terms, as mentioned by the manufacturers of this product, are therefore incorrect.

HLA analysis was performed at the DNA level using the 454 NGS Roche platform. The amplified DNA regions were sequenced by Inqaba Biotec™ with a GS Junior sequencer, in order to determine the HLA genotypes of the various samples. High resolution HLA genotyping (as described by the manufacturers of this product) was performed for the 20 samples.

The objectives for the NGS aspect of the study include:

i. Validation of the HLA kit on selected samples of the South African population.
ii. Determination of the degree of complexity revealed by NGS.
iii. Assessment of the value of NGS in determining the composition of the cord blood bank.

REVIEW OF THE LITERATURE

2.1. The Immune System

The immune system is an ever evolving collection of organs, cells, and tissues that function together to protect the body against invading pathogens, yet recognise and tolerate the body's own cells. It is the body's natural form of defence and is found exclusively in vertebrates, which have developed and evolved into an adaptable host defence system. It has the ability to protect the body through a variety of cells and molecules, capable of recognising and eliminating foreign pathogens in a dynamic network. Viruses, bacteria, and parasites are all classified as pathogenic organisms. These organisms display different molecular patterns that distinguish one pathogen from another. These changes are recognised by the immune cells as 'foreign', after which the invader will be eliminated and neutralised by the immune system. However, pathogens have evolved to such an extent that they have the ability to evade the host's immune system. The immune system of vertebrates is highly complex and consists of two different types of responses that both play a vital role in protection against pathogens. These include the innate and adaptive immune responses which both generate an immune response to pathogens and microbes.

The innate immune system is regarded as the body's first line of defence, also referred to as an individual's natural immunity against an extensive range of microbes. The innate immune system consists of molecular and chemical mechanisms and acts in a non-specific manner to eliminate pathogens. This part of the immune system involves external as well as internal mechanisms of defence. External defence includes the skin, mucus membranes, and secretions, while internal refers to phagocytic cells, antimicrobial proteins, inflammatory responses, and also natural killer (NK) cells. Innate immunity includes the involvement of a variety of white blood cells, while excluding B and T lymphocytes. The cells of the innate immune system are non-specific and, are therefore, able to initiate an immune response to various pathogens that may enter the body. This type of immune response can determine and distinguish 'self' from 'non-self' but is not capable of distinguishing small differences in foreign molecules, which is rather a feature of the adaptive immune response.

The adaptive immune response, in contrast to the innate immune response, is the second line of defence and is characterised as a slow response initially, but on subsequent exposures to a pathogen it acts more specifically and rapidly. The prime characteristics of the adaptive immune system are to develop a response and adapt to, recognise, eliminate, and memorise the pathogen. B and T lymphocytes are restricted to the adaptive immunity together with immunoglobulins, which are products of assembled gene segments. This allows for the increased variability observed in the adaptive immune recognition. The adaptive immune response is much more sensitive and antigen-specific in order to allow T lymphocytes to become activated. Once the T cells are activated, it takes over from the innate immune response for the destruction of pathogens. The ability of the adaptive immune response to recognise small changes is maintained by 'memory cells' of the immune system. Although the adaptive immune response is a well developed system, capable of protecting our bodies against pathogens, it is also capable of initiating unfortunate outcomes, such as autoimmune diseases, allergies, and rejection of transplants. Figure 1 portrays the differences between the innate and adaptive immune responses.

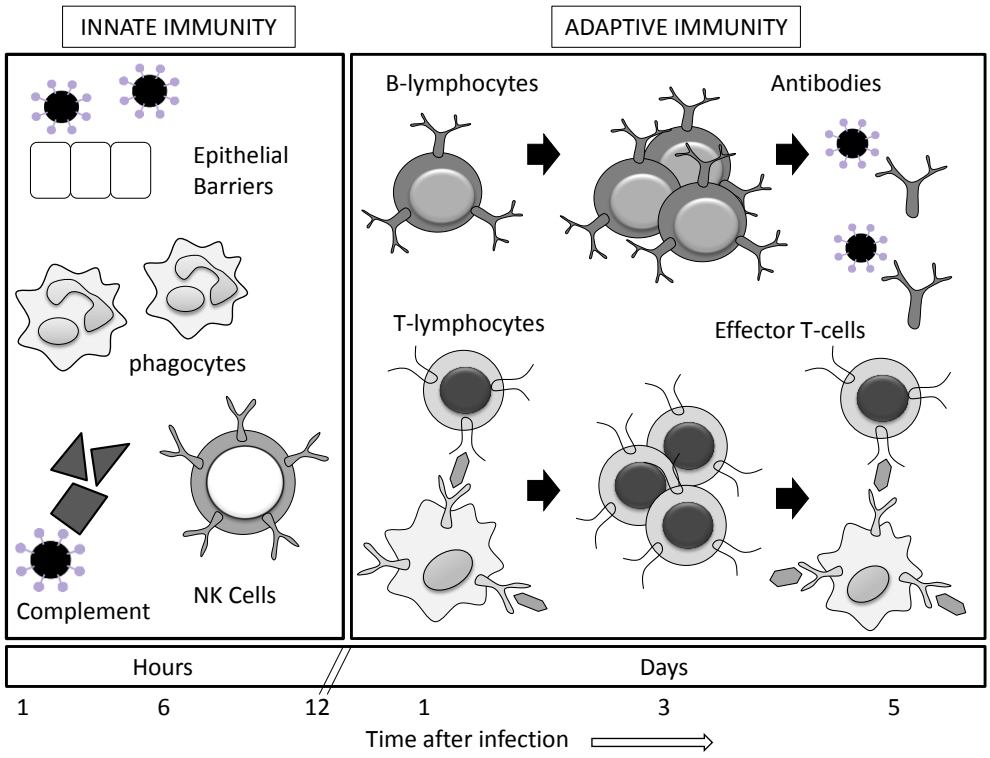


Figure 1: Innate versus adaptive immune response. This figure is a schematic representation of the innate and adaptive immune responses and the cell types involved (Image created by Juanita Mellet, adapted from Figure 27-5 in Townsend et al., 2007).

There are two components of the adaptive immune response that are crucial in differentiation and proliferation of B and T lymphocytes, known as the humoral and cell-mediated immune responses (Figure 2). The humoral part of the immune system involves the interaction of B-cells with antigens and their proliferation and differentiation into antibody-secreting plasma cells and memory B-cells. The immunoglobulins (IgG, IgM, IgE, IgA) on the surfaces of these cells bind to antigens in order to eliminate them from the system. T lymphocytes are activated and generated in response to antigens to form part of the cell-mediated immune response. T lymphocytes proliferate to form T helper and cytotoxic T lymphocytes for the destruction of infected cells. The HLA molecules play a major role in the cell-mediated immune response, where they are displayed on the surfaces of cells for presentation to T lymphocytes.

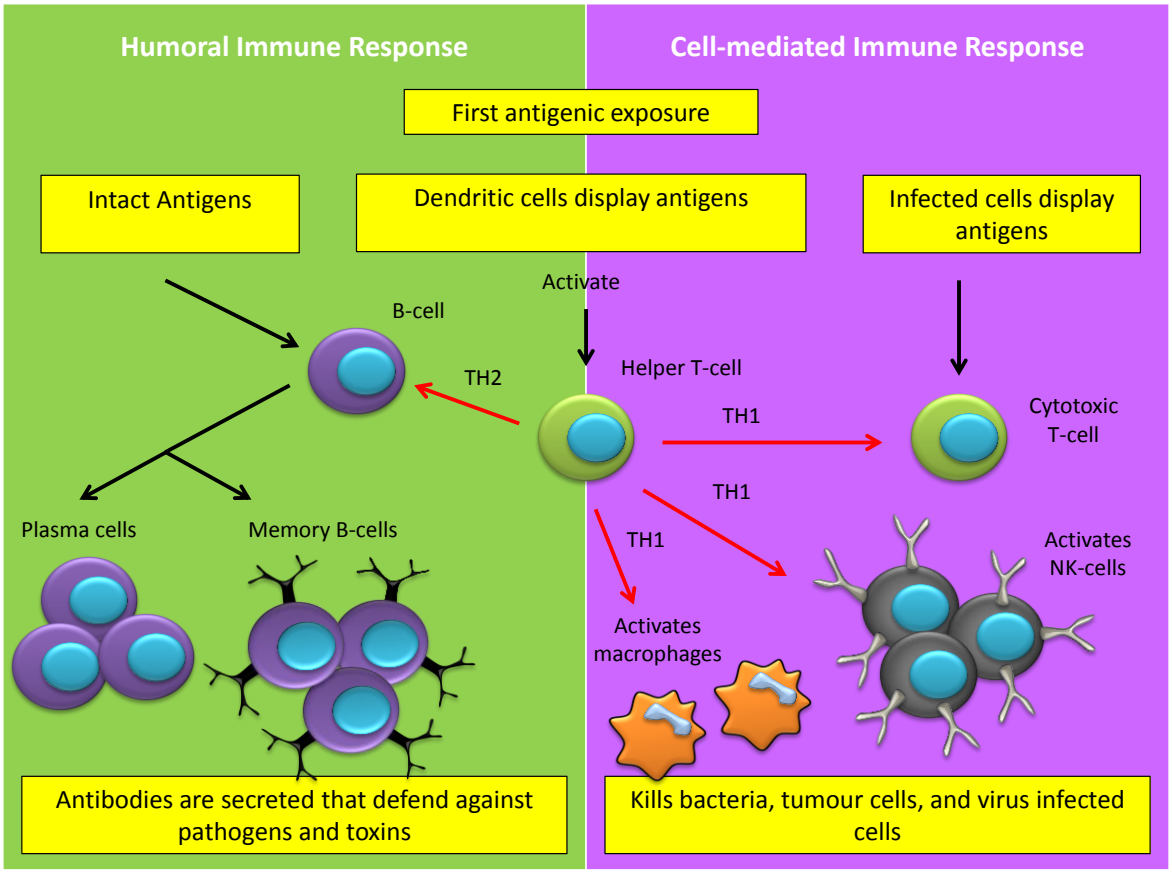


Figure 2: Humoral versus cell-mediated immune response. A schematic representation of the differences between humoral and cell-mediated immune responses (Image created by Juanita Mellet, adapted from Figure 43.14 in Campbell and Reece, 2005).

2.2. The Major Histocompatibility Complex

The major histocompatibility complex (MHC) is a multigene family present within the genomes of all vertebrates. Genes within this region encode glycoproteins that bind peptides of intra- and extracellular origin, that are presented to T lymphocytes. These regions vary significantly between different species due to the differences in gene number, composition, and organisation (Kelley *et al*., 2005; Belov *et al*., 2006). The complete MHC sequence is available for various species, which include mouse and human. The MHC region evolved about 500 million years ago, where the most primitive form is believed to still be present in nurse sharks (Kasahara *et al*., 1992). There are various other factors, such as selection pressures and pathogens that have acted on these regions and have introduced the variability and diversity in the form of gene gain and loss, observed currently in different animals.

The MHC in humans is referred to as the HLA complex, located on the short arm of chromosome 6 (Figure 3). This region encodes cell surface proteins that recognise and bind peptides. These peptides are displayed on the cell surface, where they are presented to T lymphocytes. In the case where peptides are recognised as 'foreign', an immune response is initiated. The HLA region consists of over 128 functional genes, that play a critical role in the innate as well as adaptive immune responses, since 40% of the HLA genes have an immune function (MHC Sequencing Consortium, 1999). These molecules affect the response to antigens of organisms and, are therefore, also linked to disease susceptibility and the development of autoimmune diseases.

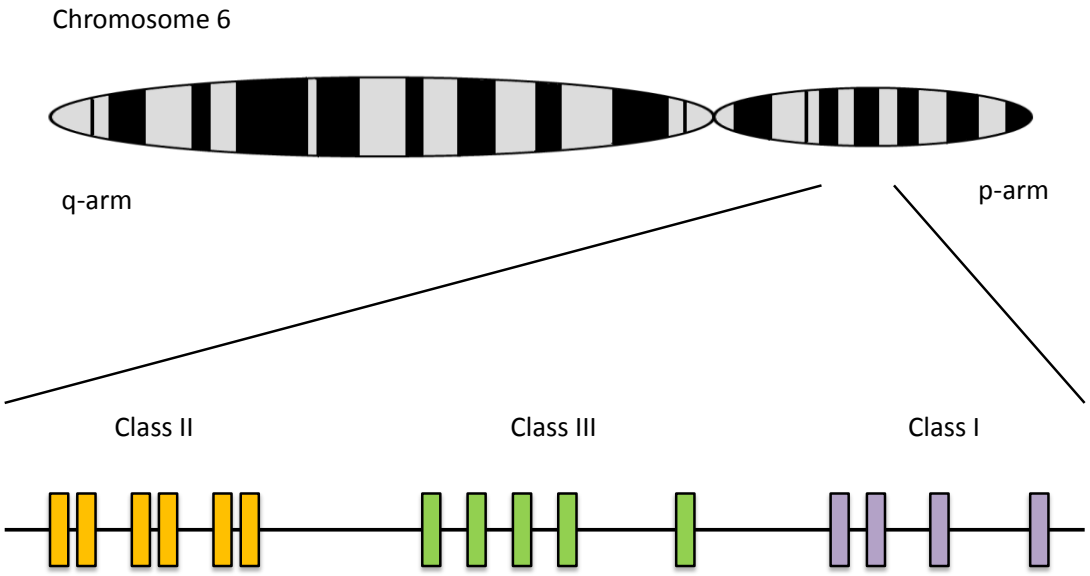


Figure 3: The HLA region on chromosome 6. The HLA region is located on the short arm (p) of chromosome 6 and spans over 3.6 Mb. This region comprises three classes (I, II, and III) that play an important role in immune responses (Image created by Juanita Mellet, adapted from Figure 1 in Mehra and Kaur, 2003).

2.3. Discovery of the MHC/HLA region

The MHC was initially described in 1936 by Peter Gorer after the observation of agglutination of mice erythrocytes by rabbit sera (Gorer, 1937). Research on the MHC was further advanced by George Snell, who discovered that it influences the ability of an organism to either accept of reject transplanted tissue from another member of the same species. This is due to the incompatibility of certain antigens that causes graft rejection. It was only after this particular discovery that Gorer named it the H-2 complex in mice

(Snell, 1986). Jean Dausset established the HLA complex in 1952 when he detected an alloantigen present on human leukocytes, which he called MAC (the initials of the three individuals who played a significant role in the detection of this molecule), today known as HLA-A2. This was confirmed by three independent studies in 1958 by Jean Dausset, Jon van Rood, and Rose Payne (Dausset, 1958; van Rood *et al*., 1958; Payne and Rolfs, 1958), who defined the presence of various other antigens on human leukocytes. These three reports laid the foundation of the information that is currently available for, and known as, the HLA complex. Following his investigation in 1958, Dausset stated that these antigens will still become of great importance in tissue transplantation. This has lead to the ongoing studies by George Snell and the discovery of the histocompatibility locus. However, it was only later in the 1970's, after a publication by Zinkernagel and Doherty on T cell restriction, that the complete function of this complex became apparent. This study revealed that T lymphocytes only recognise viral peptides when displayed on MHC molecules and, will therefore, only be activated in the presence of a combined signal of a MHC molecule together with a pathogenic peptide (Zinkernagel and Doherty, 1974). In 1980, the Nobel Prize in Physiology or Medicine was awarded for the *discoveries concerning genetically determined structures on the cell surface that regulate immunological reactions*. The laureates, Baruj Benacerraf, Jean Dausset, and George Snell jointly held this prestigious award for their work in discovering the HLA molecules (The Nobel Prize in Physiology or Medicine, 1980). Figure 4 depicts the HLA research highlights, from 1936 to 2008.

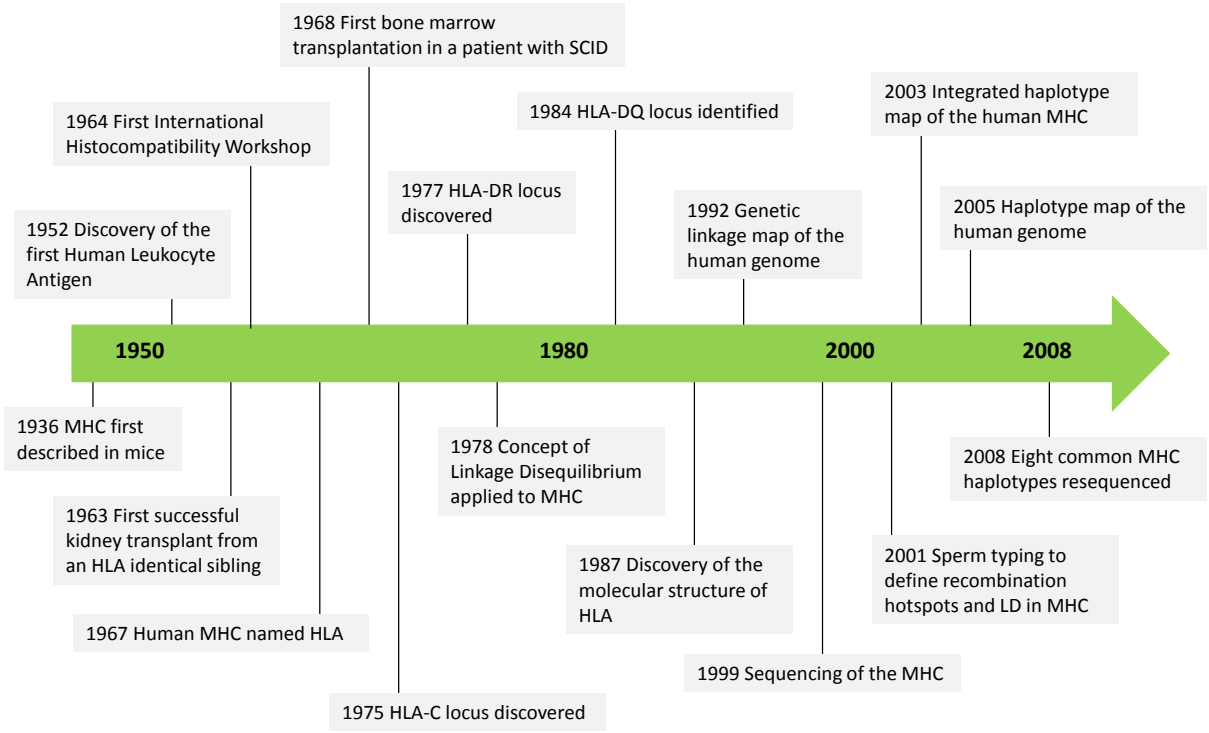


Figure 4: Timeline of research in the MHC. The MHC research highlights, from 1936 to 2008 (Image created by Juanita Mellet).

Over the years, the discovery of the different HLA genes and their importance in immune responses has led to the classification of class I and class II HLA genes. The HLA class I molecules are found on all somatic cells in the body and is involved in CD8-mediated immunity. Class II HLA molecules have a more limited distribution and is involved in CD4-mediated immunity. The main function of the HLA molecules is to present processed peptides (epitopes) to either $CD8^+$ T cells (class I) or $CD4^+$ T cells (class II).

2.4. HLA Structure and Function

The T lymphocytes of the immune system are constantly binding to the HLA molecules on the surfaces of cells in order to determine whether a cell's normal function has been disrupted by a viral infection. Under normal circumstances, HLA molecules display peptides of proteins synthesized and degraded for presentation on the cell surface. In healthy individuals, the immune system will recognise those peptides as 'self'. In the case of infection or disruption of function, an HLA molecule displays foreign peptides in its binding

groove, which activates the T lymphocytes. The immune system reacts by lysing the cell in an attempt to halt further infection by a pathogen or replication of cancer cells.

2.4.1. Class I HLA molecules

The class I HLA region spans over 2 000 Kb and consists of approximately 20 genes. There are three classical HLA genes within the class I region, HLA-A, -B, and -C. The non-classical genes include HLA-E, -F, -G, -J, -X, and various others, some known as pseudogenes, since they do not encode protein products. The non-classical genes within this region are not as polymorphic as the classical genes and also not necessarily peptide-presenting antigens. The HLA-B locus is the most polymorphic of the class I genes (Mungall *et al.,* 2003). The HLA-B locus has 2 862 alleles currently documented in the HLA database, while HLA-A has 2 188, and HLA-C has 1 746 different alleles (Robinson *et al.,* 2013). The classical HLA genes consist of eight exons, while the polymorphisms reside in gene regions that encode the peptide-binding groove. These regions include exons 2 and 3 (Figure 5), and consists of 270 base pairs (bp) and 276 bp, respectively. These two exons encode the α1 and α2 domains, respectively, while exon 4 encodes the α3 domain and consists of 276 bp. Exons 2 and 3 represent the most variable region of the gene (peptide-binding groove), while the rest of the gene is more conserved. Exon 4 also contains some variability but not as much as exons 2 and 3, therefore, it is also regarded as being conserved. All eight exons of the class I HLA genes encode particular components of the cell surface molecule (Table 1) that, in unison, create a fully functional HLA structure.

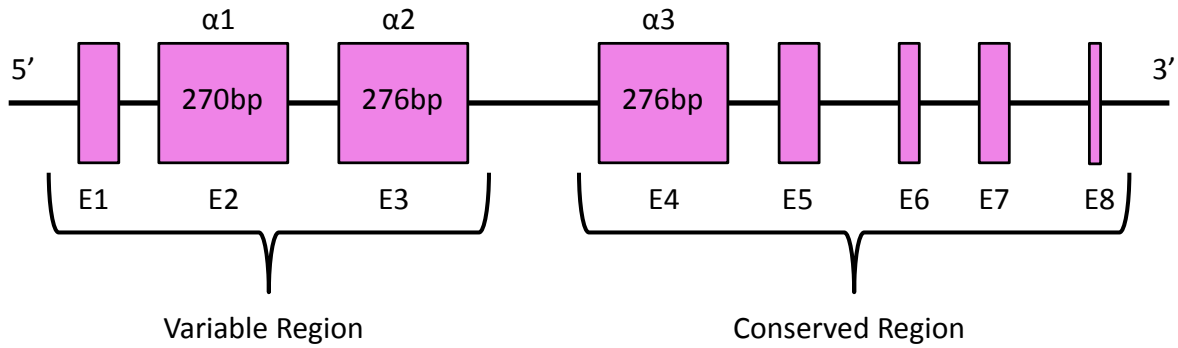


Figure 5: HLA Class I gene structure (Image created by Juanita Mellet, adapted from Figure 2 in Blasczyk, 2003).

Table 1: Class I exons and their coding peptides. The class I genes consist of eight exons that each code for a specific part of the HLA molecule.

Exon	Encodes
1	Leader peptide
2 and 3	Functionally important α1 and α2 domains
4	α3 domain
5	Transmembrane domain
6 and 7	Cytoplasmic tail
8	Contributes last two nucleotides to the C-terminal and the 3' untranslated region (exon 8 is completely untranslated in HLA-B)

Class I molecules consist of two chains, the α chain and non-covalently bound β_2-microglobulin (Figure 6). The α1 and α2 chains are the variable regions within the class I genes. These variable regions form the peptide-binding groove designed to bind endogenously-derived peptides. Class I HLA molecules present peptides within the binding groove as a complex with the α1 and α2 domains, which are then spatially recognized by the T-cell receptor (TCR) of $CD8^+$ T lymphocytes. Engagement of the $CD8^+$ TCR, along with adhesion molecules, creates an immunological synapse between the two cell types. The class I endogenous pathway is associated with defence against intracellular pathogens such as viruses.

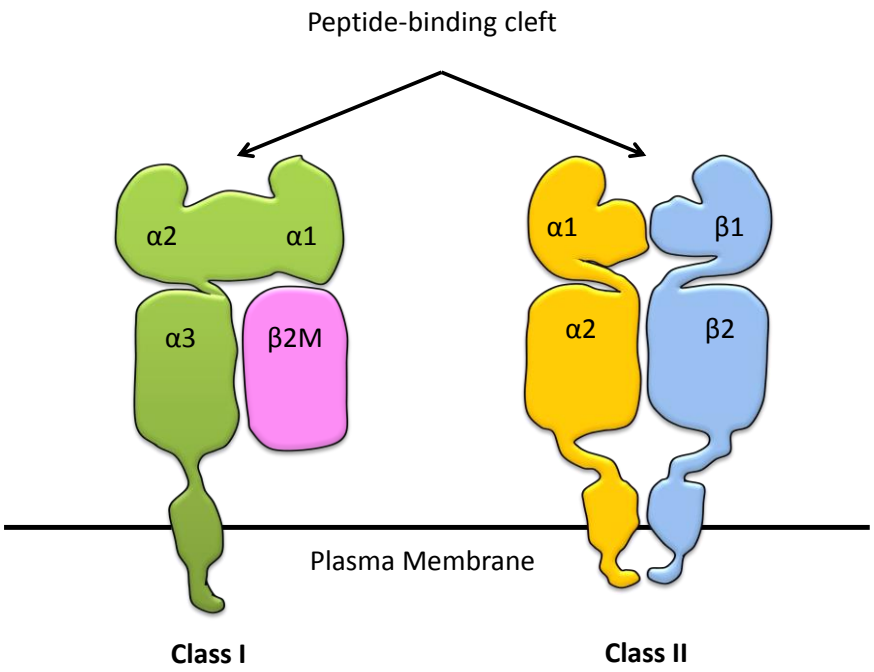


Figure 6: Structure of the class I and II heterodimers. The green and pink structures represent the α domain and β2M for class I molecules, respectively. The yellow and blue structures represent the α and β domains for class II molecules, respectively (Image created by Juanita Mellet, adapted from Figure 2 in Klein and Sato, 2000).

2.4.2. The Endogenous Pathway

The endogenous pathway includes the degradation of intracellular proteins by a proteolytic pathway present in all cells (Figure 7). The proteasome is a large protein complex involved in the degradation of misfolded or unwanted proteins (including pathogens) into smaller peptides. Most proteins that are targeted for degradation have a small protein, called ubiquitin, attached to them. The immune system utilises this proteolytic pathway to produce small peptides for presentation with class I HLA molecules.

Peptides that are generated by the proteasome are transported into the rough endoplasmic reticulum (RER) by a transporter associated with antigen processing (TAP), a membrane-bound heterodimer with an affinity for peptides consisting of eight to 14 amino acids. The class I molecules generally present peptides of approximately nine amino acids. The preferred peptide length is achieved by aminopeptidases in the endoplasmic reticulum (ER). It is optimal for TAP to favour peptides for class I MHC molecules. The α domain and the β2-microglobulin of the class I MHC molecules is synthesised in polysomes. The assembly of

a stable class I molecule requires a peptide in the binding groove of the HLA molecule. These molecules then proceed to the cell surface via the Golgi complex.

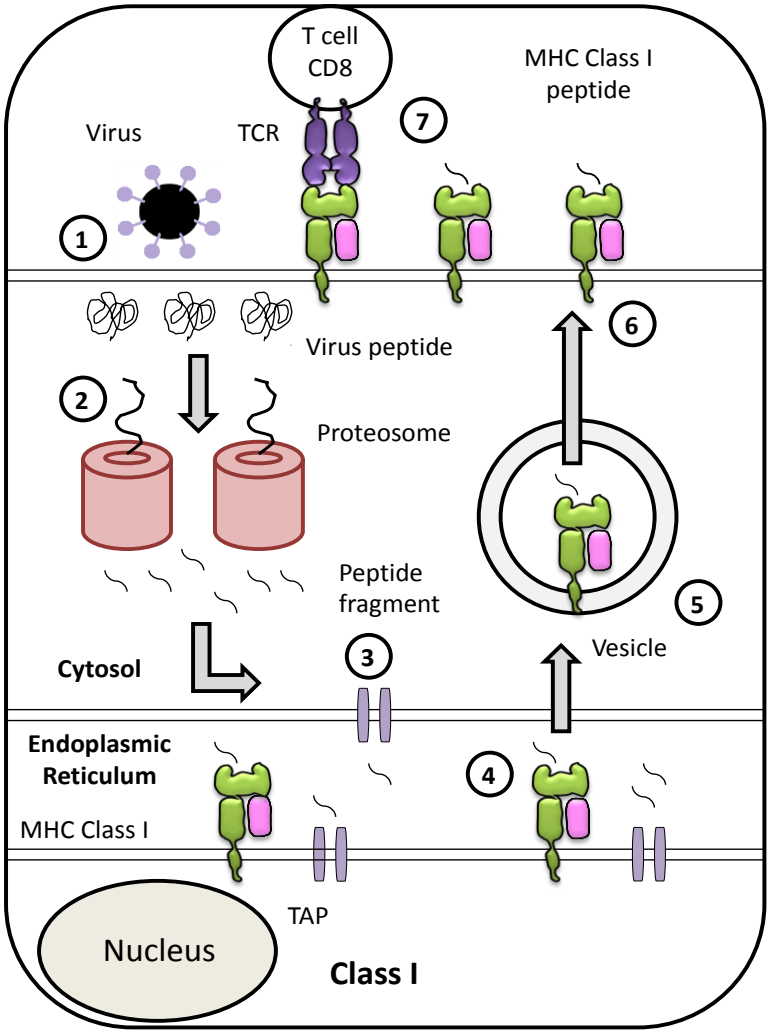


Figure 7: The endogenous pathway. (1) Viruses enter cells through various methods; following which viral peptides are tagged for degradation by ubiquitin. (2) Ubiquitin is recognised by the proteosome and degrades peptides into small peptide fragments. (3) The peptide fragments are transported via TAP molecules from the cytosol to the RER. (4) In the RER the peptide and the class I molecule are assembled. (5) The assembled molecule is transported from the RER to the cell surface via the Golgi complex. (6) Once it reaches the cell membrane it is displayed on the cell surface. (7) Class I HLA molecules are recognised by TCR on $CD8^+$ T lymphocytes (Image created by Juanita Mellet, adapted from Figure 8-23 in Kindt et al., 2006).

2.4.3. Class II HLA molecules

Class II molecules include the HLA-D genes, that are subdivided into DQ, DP, and DR, and are restricted to immune competent cells (B-lymphocytes, macrophages, and endothelial cells

of T lymphocytes). The class II genes encode proteins that are expressed on the cell surface of antigen-presenting cells (APCs), where they present to the helper T cells. All class II molecules consist of two chains, an α chain and a β chain (Figure 8). Recent studies have identified the presence of multiple forms of β, as well as α chain genes in humans. The DR gene only contains one α domain, while DP and DQ both contain two α domains. DM and DO have been identified as non-classical genes within the class II HLA region. These HLA molecules have slightly different functions compared to other genes within this region. The DM gene encodes a class II-like molecule that assists in loading the antigenic peptide into the HLA molecule. The DO gene encodes molecules expressed solely in the thymus and on the surface of mature B cells, where it acts as a class II regulator by negatively modulating HLA-DM. Within the class II genes, exon 2 is the most variable region and also forms the peptide-binding groove (β1).

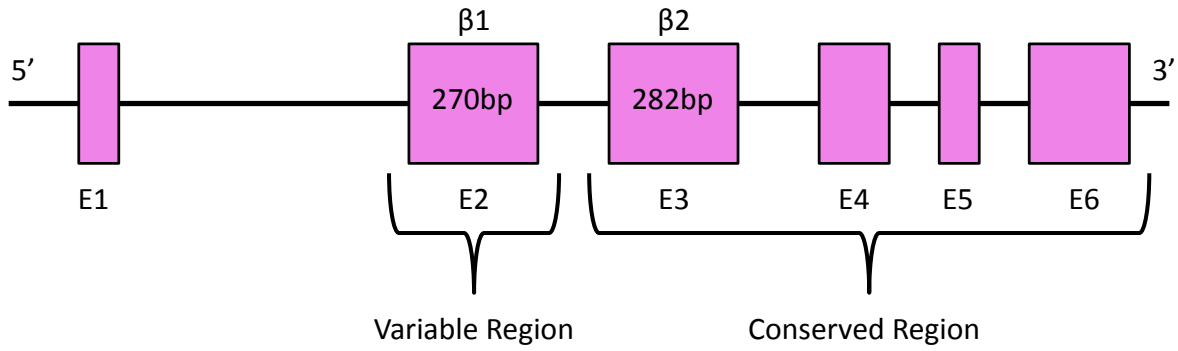


Figure 8: HLA class II, DRB gene structure (Image created by Juanita Mellet, adapted from Figure 3 in Blasczyk, 2003).

Class II molecules are involved in the exogenous pathway and are associated with defence against extracellular pathogens such as bacteria. The TCR, which binds to the exogenously derived peptide-class II HLA complex, is found on helper T cells (CD4 cells). The helper function of these cells involves the activation of the general immune response. This includes cytokines and the cellular and humoral immune responses, which is why these molecules are only present on immunologically active cells (Browning and McMichael, 1996).

2.4.4. The Exogenous Pathway

Antigen-presenting cells have the ability to internalise antigens by means of endocytosis or phagocytosis. Once these antigens are inside the cell they are degraded by various chemical pathways, which include the early and late endosome and finally the lysosome, called the exogenous pathway (Figure 9). The different stages of this pathway occur at a range of pH concentrations and are believed to be transported by small transport vesicles between compartments. The lysosome has acid-dependant hydrolases, such as proteases, nucleases, and various others. The hydrolases degrade the antigen into oligopeptides (13-18 amino acids), which then binds to the class II HLA molecule and acquires protection against any further degradation.

Due to the presence of both class I and II molecules on the surface of a cell, it is essential that these molecules do not bind to the same antigenic peptides within a cell. During the process of HLA class II synthesis, an invariant protein chain is assembled with the α and β domains of the class II HLA molecule. This chain interacts with the peptide-binding groove of the molecule and also aids in folding and exiting from the RER. The invariant chain is degraded in successive stages, however, a small segment of the chain, called CLIP (class II-associated invariant-chain peptide) remains bound to the molecule. The CLIP prevents binding of any premature peptides to the binding-groove of the molecule. HLA-DM is a non-classical class II HLA molecule that assists in the exchange of CLIP with an antigenic peptide, after which the HLA molecule is transported to the plasma membrane and displayed on the cell surface.

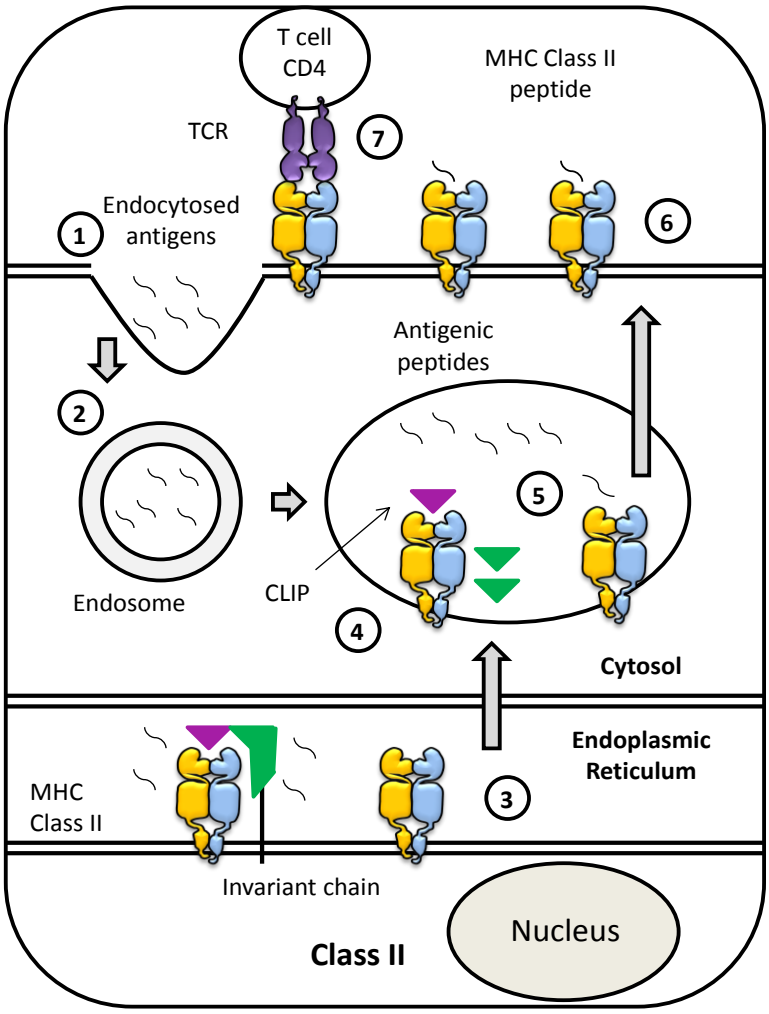


Figure 9: The exogenous pathway. (1) Extracellular antigens enter the cell by endocytosis, after which (2) antigens are degraded within an endosome, which later becomes a lysosome. (3) In the RER, an invariant chain binds to the binding groove of the MHC class II molecule, which inhibits binding to peptides and also aids in exiting from the RER into the cytosol. The MHC class II molecule leaves the RER in a vesicle. (4) The vesicle fuses with the late endosome that still contains the endocytosed peptide. (5) The invariant chain is degraded until only a small piece (CLIP) still blocks the peptide-binding groove. HLA-DM facilitates the removal of CLIP and replaces it with a peptide. (6) A stable MHC molecule is displayed on the cell surface for (7) presentation to TCR on $CD4^+$ T lymphocytes (Image created by Juanita Mellet, adapted from Figure 8-24 in Kindt et al., 2006).

2.5. HLA Gene Expression and Regulation

The HLA molecules are co-dominantly expressed, which implies that heterozygous individuals will express the gene products of both alleles on the surface of their cells. The inheritance of HLA molecules was first demonstrated in 10 infants by Payne and Rolfs in 1958. They observed three successive generations and determined that these molecules are generally inherited in a heterozygous state (Payne and Rolfs, 1958). The HLA alleles within a

family are inherited in a Mendelian fashion, known as an HLA haplotype (Figure 10). A haplotype is defined as a set of alleles inherited on homologous chromosomes. These alleles are closely linked to one another with recombination occurring at low frequencies, and therefore, they are inherited as a block. Children have a one in eight (12.5%) chance of inheriting an identical haplotype to their parents', while siblings have a one in four (25%) chance of inheriting the identical haplotype to their siblings' (Choo, 2007). In a family where parents have more than four children, it is expected that at least two children will be HLA identical, called haplo-identical siblings. Haplo-identical siblings can also be referred to as an 'HLA identical match'.

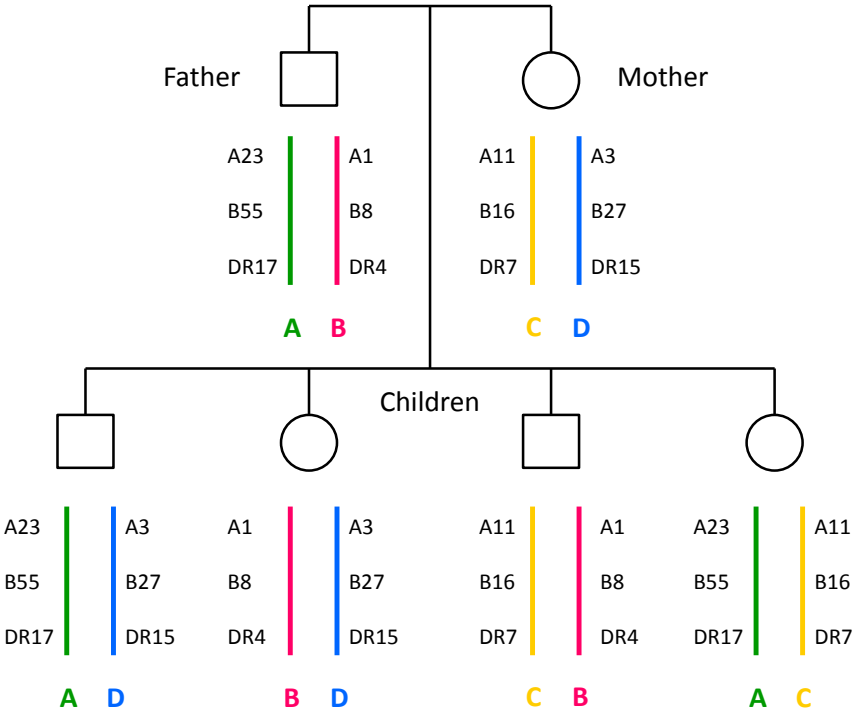


Figure 10: HLA inheritance. This figure illustrates the inheritance of HLA alleles from parents to their offspring. A (green) and B (pink) represents the father's haplotypes, while C (yellow) and D (blue) represents the mother's haplotypes. Children inherit one haplotype from each parent (Image created by Juanita Mellet, adapted from Figure 2 in Choo, 2007).

Normal healthy cells display self peptides in their HLA binding grooves as a result of normal turnover of self proteins. The high degree of variability and diversity within populations enables binding of different peptides due to differences within the peptide-binding groove. Over the last decade it has been determined that peptides that are bound by certain HLA

molecules share specific amino acids (or ones with similar properties) at a few positions, called anchor residues (Falk *et al*., 1991). The anchor region aids in specificity of binding, where it interacts with the surface of the HLA binding groove. Therefore, a specific HLA molecule can bind various peptides that possess specific amino acids in the anchor positions even though the remainder of the amino acid sequence may differ.

As previously mentioned, the classical class I molecules are more commonly expressed on nucleated cells even though the level of expression differs between different cell types. Lymphocytes express the highest level of HLA molecules compared to all other cells in the human body and constitute about 1% (5 x 10^5 molecules) of the plasma membrane proteins per cell. The level of HLA expression varies among different cell types in the body. Unlike class I molecules, class II molecules are constitutively expressed on the surfaces of APCs. Class II expression depends on the differentiation stage of a cell. This is confirmed by the expression of class II molecules on mature B cells, but the lack thereof on pre-B cells. This is also observed on monocytes and macrophages that express low levels of class II molecules until they are activated by interaction with antigens, after which the level of expression increases significantly (Kindt *et al.,* 2006). There are three classical class II genes (DP, DQ, and DR) and each has two forms. Therefore, an individual expresses six class II molecules from either parent. This number can further increase with the presence of multiple β chains. The diversity generated by the presence of multiple chains increases the number of antigenic peptides to which the molecules can bind, and is therefore, considered to be a great advantage to an organism.

All HLA class I and II genes have promoter sequences at the 5’ end where sequence-specific transcription factors are able to bind. Various transcription factors and motifs have already been identified for the various HLA genes. Two transcription factors, class II transcription activator (CIITA) and regulatory factor X (RFX), have been identified for class II genes. In the case where these transcription factors are unable to bind, individuals suffer from severe immunodeficiency due to the lack of HLA class II on their surfaces. The expression of HLA molecules is also largely controlled by cytokines. The interferons and tumour necrosis factor (TNF) have been shown to play a significant role in increased expression of HLA class I molecules on the surfaces of cells. Interferon gamma (IFN-γ) binds to a specific transcription

factor that binds to the promoter sequence which up-regulates the expression of the class I genes. IFN-γ also seems to be involved in inducing CIITA, which in turn also increases the expression of class II molecules. There are also other cell-type specific cytokines that play a role. The expression of these molecules can also be down-regulated by certain factors, such as prostaglandins and corticosteroids, which are known to decrease the expression of class II molecules. IFN-γ also decreases the expression of class II molecules on the surfaces of B cells. The expression of HLA molecules can also be decreased by certain viral infections, which include hepatitis B virus (HBV), human cytomegalovirus (CMV), and adenovirus 12 (Ad12).

2.6. HLA Diversity

The HLA region is the most polymorphic in the human genome (International HapMap Consortium, 2005) and accounts for more than 10% of human diversity, with heterozygosity ranging from between 80-90% (Hughes and Nei, 1988). These allelic variants mostly arise within the nine classical genes of the HLA region. Since the role of HLA is to present peptides from invading organisms that may enter the body, it is likely that HLA has evolved to manage a vast number of peptides. There are currently 8 794 HLA alleles listed in the IMGT/HLA database, of which 6 919 are class I alleles and 1 875 are class II alleles (Robinson *et al.*, 2013). Figure 11 below represents the increase in the number of alleles identified to date. This high number of HLA alleles can result in a synonymous or a non-synonymous change of the amino acid sequence. The most dramatic is a non-synonymous change, which causes a slight difference in the amino acid sequence of the HLA molecules, which can also explain the elevated number of HLA antigens.

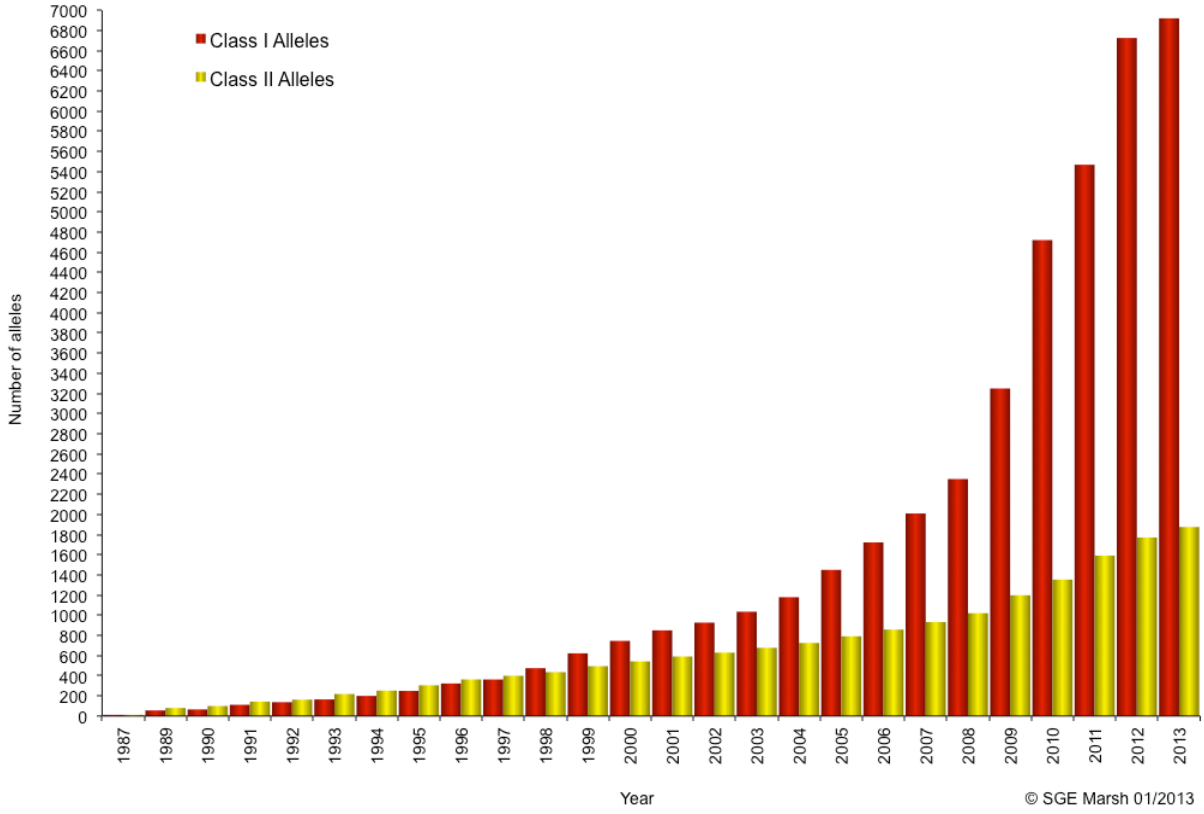


Figure 11: The number of alleles named each year from 1987 to January 2013. The red bars represent the number of class I alleles, while the yellow bars represent the number of class II alleles from 1987 to present. Image courtesy of the HLA Informatics Group, Anthony Nolan Research Institute, London, UK (Robinson et al., 2013).

The reason for the diversity of this highly complex system is the fact that these molecules need to present an enormous array of antigenic peptides to T cells in order to elicit a unique immune response to a wide variety of peptides. The sequence variation of these molecules is not randomly distributed across the gene, instead it is denser in short stretches of the gene (exons 2 and 3) that code for the peptide-binding groove. These genes evolve by processes that take place over many years, which include the accumulation of mutations, gene conversion/interlocus genetic exchange, over-dominant balancing selection, and frequency-dependant selection. The accumulation of mutations can be attributed to the euchromatic state of the DNA over the HLA region, since these genes are frequently transcribed and translated into protein molecules expressed on the cell surfaces of almost all cells. Therefore, the HLA genes are exposed to many mutagens that cause the build-up of mutations across this particular region. The mutations within these genes mostly arise in the

form of single-nucleotide polymorphisms (SNPs), while the SNPs within the coding region are the most informative, and have directed the allelic diversity observed today. The diversity of these molecules has occurred due to the presence of different alleles at a specific locus within a species. The alleles can differ from one another by an alteration at a minimum of one SNP. Several thousand allelic variants of the HLA system have already been described. The presence of alleles differs from one individual to another by 5-10%. The diversity can also be the result of the presence of duplicated genes, with similar or overlapping functions. It is plausible to make this assumption, since the HLA complex consists of genes with similar but not completely identical structure and function (HLA-A, -B, and -C).

The enormous variation observed is highly specific and accounts for the diversity amongst populations (Jin and Wang, 2003). The divergence rate of the HLA genes is due to the long history of independent haplotype evolution, where Africans are genetically more diverse compared to other populations (Wainscoat *et al.,* 1986). This suggests, together with other data, that Africans were the founder population of *Homo sapiens* (Wainscoat *et al*., 1986; Cann *et al*., 1987). Smaller isolated populations will have a decreased number of alleles compared to larger admixed populations, which once again indicates that the high degree of diversity within the HLA region is the result of population demographics as well as selection pressures.

2.7. Evolution, Selection, and Linkage Disequilibrium across the HLA region

Modern humans are thought to have replaced all the archaic human species since their evolution in Africa approximately 200 000 years ago. The ‘Out of Africa’ theory suggests that humans originated in East Africa and evolved into modern humans that migrated from Africa, approximately 60 000 years ago, to the rest of the world (Figure 12). The theory further suggests that the migration event caused modern humans to replace ancient humans, such as Neanderthals.

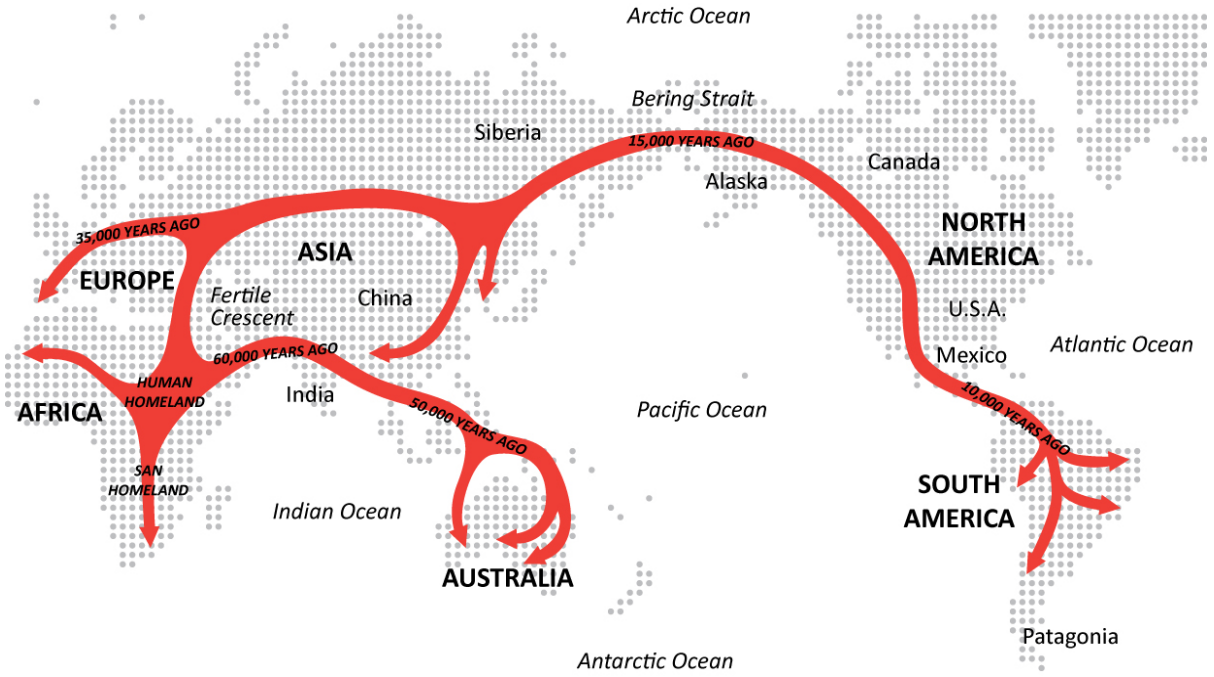


Figure 12: Human migration out of Africa. This figure depicts the 'Out of Africa' theory and the migration of humans from Africa to the rest of the world (Image created by Juanita Mellet, adapted from Figure 2 in Fackenthal and Olopade, 2007).

In 2010, the first Neanderthal genome was sequenced and has shed some light on ancient genomics. This has aided to some extent to answer various questions on human evolution and history. Neanderthals lived in western Eurasia from about 400 000 years ago until 30 000 years ago and are classified as a subspecies of *Homo sapiens*. In December 2010, another population of ancient humans was discovered in a cave in Denisova (Southern Siberia) and are said to be more diverse than Neanderthals (Gibbons, 2011). The Denisovans and the Neanderthals are referred to as archaic humans, since their evolutionary story began to split from those of humans about 500 000 years ago.

The human distribution patterns observed today are due to the migration of populations that started as a result of fluctuating climate changes, which has led to the distribution of populations across continents. There is an alleged theory that interbreeding between modern and archaic humans occurred early after migration from Africa, which has led to an enhanced immune system in modern humans. According to Ferrer-Admetlla and co-workers, these archaic humans had a better adapted immune system to local pathogens (Ferrer-Admetlla *et al*., 2008). This was confirmed by a study completed by Professor John

Hawks and co-workers at the University of Wisconsin, where they discovered that these groups lacked particular forms of genes, which made it possible to fend off certain diseases (John Hawks Weblog, 2012). There are ongoing studies to determine whether these genes are linked to autoimmune diseases.

Professor Peter Parham's team at Stanford University compared the diverse immune genes amongst Neanderthals and Denisovans, and reported at a Royal Society Symposium in London (2011) that they carried HLA genes that are abundant in modern humans from Europe and Asia, respectively (Callaway, 2011). This supports the notion that Denisovans had once lived across Asia. It is estimated that 50% of Europeans owe the variant of a specific HLA gene to Neanderthals. The variants observed in these two groups are either at low frequency in Africans, or entirely absent. It is believed that the most diverse populations are those still residing in Africa, since populations become less diverse the further they have migrated from their origin. An article published in 2004, by Cao and co-workers, suggests that the genetic diversity between all African populations is greater than the diversity observed between European populations (Cao *et al.,* 2004). This is also observed for the HLA regions between distinct populations, which could partly be due to interbreeding with archaic humans, but may be the result of selection pressures that have shaped and increased the presently observed diversity.

2.7.1. Evolution and Balancing Selection

The strong balancing selection at the HLA genes together with recombination, instils the human population with a multitude of HLA alleles and haplotypes. 'Balancing selection' refers to the maintenance of multiple alleles in the gene pool of a given population at a frequency higher than that of gene mutations. Immune defence is strongly dependant on the pivotal role of the HLA genes; therefore, it is of utmost importance to maintain a variety of surface molecules for the long-term survival of humans. An article published in 2011 by Abi-Rached and co-workers (Abi-Rached *et al.*, 2011), compared modern and archaic humans by investigating the HLA genes, since these genes are generally conserved across generations. They discovered similarities, which suggest that interbreeding took place between these groups and modern humans.

Natural selection acts on several aspects, such as morphology, physiology, and also life history traits. These traits exhibit heritable genetic variation, which could explain the vast number of alleles observed within the HLA region (Takahata and Nei, 1990). Various studies have investigated the impact of natural selection on the HLA genes and its critical role in HLA frequencies observed throughout populations. Although previous studies have reported that HLA genes are subjected to selection, it is still a mystery as to how these genes maintain their immense degree of variation. A study by Ferrer-Admetlla and co-workers concluded that there is a significantly higher level of balancing selection in Europeans than in African individuals (Ferrer-Admetlla *et al.,* 2008). There are several selection pressures that are believed to have contributed significantly to the diversity observed within the HLA region today. These include heterozygous advantage (over-dominant selection and frequency-dependant selection). It is well-known that multiple HLA alleles confer resistance or susceptibility to various infectious diseases. However, this is not constant across populations and has led researchers to believe that the selection on these genes is also pathogen-driven. These selection pressures are called 'pathogen driven selection', which operates on specific alleles and are favoured due to their ability to provide protection against certain pathogens.

2.7.1.1. Heterozygous Advantage (over-dominant selection)

In 1975, Doherty and Zinkernagel suggested that heterozygotes have a higher fitness compared to homozygotes (Doherty and Zinkernagel, 1975). Heterozygotes are capable of fending off an increased range of pathogens and will, therefore, survive for prolonged periods of time, which increases their fitness and also the chance to reproduce. This suggestion has been proven in various studies, which indicates that a single heterozygote advantage has an increased effect on the polymorphisms observed throughout this region. A heterozygous individual has an increased number of antigens displayed on the cell surface ,and will therefore, recognise additional pathogenic peptides and respond more efficiently when compared to a homozygous individual, which is a result of heterozygous advantage (over-dominant selection). Over-dominant selection increases the proportion of heterozygous individuals within a given population.

Sanchez-Mazas and co-workers investigated the HLA region and observed lower homozygosity for several loci than would have been expected under normal circumstances (Sanchez-Mazas *et al.,* 2000). An additional study, published in 2012 observed a positive correlation between genetic diversity in HLA-B and pathogen-driven selection, while a negative correlation was observed for DQB1. This has also shown to correlate with genetics and the geographical migration of populations (Sanchez-Mazas *et al.,* 2012).

2.7.1.2. Frequency-dependant Selection

Frequency-dependant selection was first proposed prior to the knowledge on MHC functioning (Bodmer, 1972), and suggested that fitness can vary depending on the allele frequencies present in a population. Pathogens also evolve and acquire new mutations that enable them to increase their own fitness. When organisms evolve they adapt according to their environment. In the case where certain HLA molecules are frequent within a population and capable of destroying a specific organism, organisms will evolve to avoid destruction by such molecules, therefore, rendering themselves resistant to binding by such molecules. A rare or novel HLA molecule, however, will still have the ability to bind to such a pathogen due to minimal exposure of pathogens to such a molecule. Once such a molecule becomes frequent, pathogens will again start adapting and will reduce the fitness of the molecules. Therefore, alleles fluctuate in frequency driven by pathogen adaptation.

2.7.2. Non-synonymous *versus* Synonymous mutations

It is generally believed that synonymous changes occur more frequently within the genome than non-synonymous changes. Synonymous changes can be explained as a substitution that does not alter the amino acid sequence, whereas a non-synonymous change refers to a substitution that alters the amino acid sequence. The occurrence of non-synonymous changes has been determined to be favoured by selection in the binding groove of the HLA molecules. A study performed by Hughes and Nei established that the codons that form the peptide-binding groove of these molecules accumulate a higher rate of non-synonymous changes compared with synonymous changes, whereas regions outside the binding groove showed a higher rate of synonymous changes (Hughes and Nei, 1988; 1989). This has led scientists to believe that selection plays a strong role in maintaining these non-synonymous

substitutions within the HLA region in order to increase the binding capacity to antigenic peptides.

Mutations typically occur randomly throughout the genome. This, however, is not the case within the HLA region. The high number of alleles observed is predominantly across the classical HLA genes. Studies have determined that as the distance from the classical genes increases there is a decrease in the number of polymorphisms within genes located in the HLA region. The genes that are further away have decreased variation, since recombination removes those genes from the selected regions, while it is maintained within the closely linked genes. This suggests and confirms that the classical genes are under strong selection. The selection is, therefore, stronger on genes that are closely linked to the classical genes.

2.7.3. Linkage Disequilibrium

Closely linked genes are referred to as being in linkage disequilibrium (LD). This is a notion that has been studied for many years and has also been a topic of interest within the HLA region. The term linkage disequilibrium describes a setting in which two or more alleles occur together more frequently than would be expected under normal circumstances, and is termed a haplotype. Haplotypes are population-specific, and therefore, their frequencies may differ between populations. Scientists believe that these patterns are conserved over generations, which has led to the concept of the 'ancestral haplotype'. There was always believed to be a relationship between LD and genetic distance, which a study in 2000 failed to prove due to the strong linkage between class I loci considering the increased distance between them (Sanchez-Mazas *et al.,* 2000).

Investigations across the HLA region have discovered strong LD patterns between the class I and class II genes. It was indicated that HLA-B and -C are in strong LD, and HLA-DQ and -DR are also in strong LD with one another. The presence of recombination hotspots between DPB1 and DQB1 loci could be the reason for the weak linkage observed across these genes. Linkage across DP seems to be weak in numerous European populations. Linkage disequilibrium may vary between populations due to the differences in the creation, maintenance and decay in LD patterns. These patterns of LD are dependant on various contributing factors that include migration, admixture, and genetic drift. According to

Sanchez-Mazas and co-workers, the HLA region is subject to a range of selection pressures and strong linkage, which accounts for the high degree of variation observed within this region.

2.8. HLA Nomenclature

HLA typing is performed to determine an individual's molecular genotype for the various HLA loci; however, it is not always as straightforward as it sounds. The continually increasing number of HLA alleles makes it a priority to have a structured way of designating names, while also taking into account the ambiguities in dealing with these genes.

The HLA nomenclature system utilises sets of digits to designate HLA names. Each allele present in the IMGT/HLA database has a unique name consisting of four sets of digits (Figure 13). The name is dependant on the sequence of the allele, and can therefore, be longer than four sets of digits when necessary. The first set of digits indicates the type, which refers to the specific allele group. The next set of digits indicates the subtype; this number is assigned in order of the discovery of the sequences and refers to the specific protein. In the case where the two sets of digits differ, it is an indication that there are one or more nucleotide substitutions that alter the amino acid sequence. Synonymous changes within the coding region of the genes are indicated by the third set of digits. Sequence polymorphisms in intronic, 5'- and 3'-untranslated, and flanking regions are indicated by the fourth set of digits. There are also various suffixes that are used for HLA nomenclature. These suffixes indicate the expression of a particular antigen/allele (Table 2).

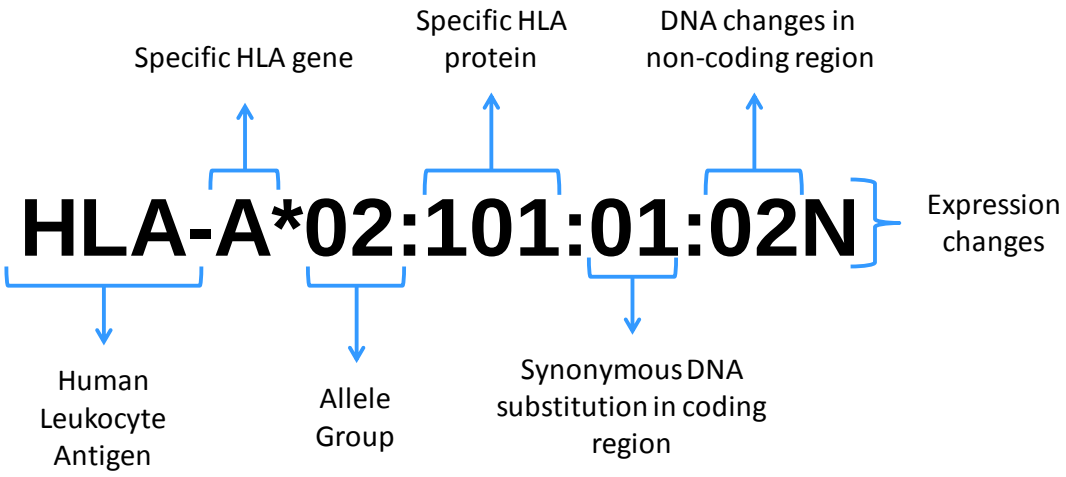


Figure 13: HLA nomenclature. This figure is a schematic representation of the HLA nomenclature. This is important to consider with low, medium, and high resolution typing (Image created by Juanita Mellet, adapted from Marsh et al., 2010).

HLA typing is performed at different resolutions depending on the level of information required for the various applications. The resolution of HLA typing refers to the amount of information that one obtains from the different typing methods, where low resolution is the minimal amount of information needed, while high resolution is the most information that one is able to obtain. Low resolution typing refers to HLA typing at the antigen level and includes information on the first two digits, which can be determined by serological methods. Medium resolution typing makes use of the PCR, as well as DNA-based methods for determination of the HLA types of individuals. However, medium resolution does not determine the exact sequence of the HLA genes. It rather makes use of a more general approach of using DNA probes to determine the HLA type. Medium resolution typing is faster and more cost effective than sequencing, and is therefore, used to screen individuals as potential donors. DNA of donors that match at medium resolution is further sequenced to determine the exact sequence of the specific HLA genes. High resolution molecular HLA typing refers to allelic typing. DNA- or sequence-based methods are generally performed for high resolution typing, which determines the exact sequence of the genes of interest. Allelic typing is the process of using typing methods that provide information on four or more digits. Depending on the method used for typing, it is possible to observe ambiguous results.

Table 2: Suffixes in HLA nomenclature. Summary of what the different possible suffixes mean when used in HLA nomenclature.

HLA Nomenclature	Indication
HLA-A*23:01**N**	Null allele
HLA-A*30:14**L**	An allele encoding a protein that expresses 'low' surface antigens
HLA-A*24:02:01:02**L**	An allele encoding a protein that expresses 'low' surface antigens due to a mutation outside of the coding region
HLA-B*44:02:01:02**S**	An allele outside the coding region that expresses a protein as a 'secreted' molecule only
HLA-A*32:11**Q**	An allele with a mutation that has previously been shown to have a significant effect on cell surface expression, but this was not confirmed and remains 'questionable'

Information reproduced with permission from the HLA Informatics Group, Anthony Nolan Research Institute, London, UK (Robinson et al., 2013).

The HLA nomenclature is important in obtaining and conveying the correct information regarding HLA types. It has taken many years and continuous HLA workshops to come up with the WHO nomenclature system, currently in use. The HLA nomenclature is updated on a monthly basis to include newly discovered alleles and also to correct previously encountered errors.

2.9. HLA Ambiguities

The vast number of HLA alleles that have already been identified, together with the escalating discovery of alleles in this highly polymorphic region, has led to the problem of HLA ambiguity, which is still a matter of concern in HLA typing. Ambiguous typing refers to more than one interpretation of the results, which arises since the majority of the typing techniques focus on exons 2 and 3 to assign a genotype. The reason for this is due to the fact that these two exons are the most polymorphic within these genes. However, it is possible that the diversity is located within a region that is not sequenced, and therefore, the diversity cannot be determined due to identical sequences over exons 2 and 3. Although the change outside the sequenced region may alter one's genotype, it might be that these regions are less important due to the fact that the variation is not within the binding groove of the molecules. Even though this could be a possibility, it is not known whether a change

in another part of the molecule will alter the whole molecule. New typing techniques involve the sequencing of additional exons, which will resolve some of the ambiguity. Previous studies have mentioned that in order to resolve the ambiguities, it is necessary to sequence the entire gene and not just particular regions of it. Like that, it will be possible to determine the complete genotype/haplotype of an individual at a high resolution for various applications.

2.10. HLA Typing Applications

The purpose of HLA typing is to obtain detailed information on the diverse HLA genes for various applications. Various disciplines require specific information regarding these genes. The most important and most frequently used application is typing for transplantation, which requires information on the HLA loci in donors and recipients. Forensic, anthropological, and disease association studies also make use of HLA typing in order to identify criminals, determine migration patterns, ancestry, and also identifying predisposing HLA alleles to specific diseases, e.g. those that are immune associated.

2.10.1. Transplantations

A primary application of HLA typing is the matching between donors and recipients for transplantation, which has been a routine procedure for many years. The increased mismatching between individuals results in a higher risk of rejection and occurrence of GVHD in recipients. Graft rejection is the process in which the recipient's immune system attacks the transplanted tissue (Porth, 2010). This occurs due to the natural function of the immune system to detect 'foreign' material present in the body, and to effectively destroy it, as is the case with viruses and bacteria. Transplant rejections can be avoided by serotyping or DNA-based methods to determine the most appropriate donor-recipient HLA match, or through the administration of immunosuppressive drugs. In the case of HLA mismatching, GVHD can occur when the donor cells recognise the host cells as foreign and start attacking them. HLA mismatching between individuals can promote chimaerism in recipients. Chimaerism refers to the fusion of two genetically distinct cell lines, which could occur in the case of transplants. HLA matching has contributed in a large extent to the success of engraftment in transplantations. There are different transplantations and the

degree of matching is different for each one. These include the transplantation of solid organs as well as stem cells (bone marrow transplantation).

2.10.1.1. Solid Organ Transplantation

Solid organ transplantation was inaugurated in the 1950's, after the performance of the first successful kidney transplant between identical twins in 1954 (Merrill *et al.,* 1956). Since then, numerous solid organ transplants have been performed on a regular basis, that include transplants of the liver, kidney, heart, lung, cornea, and many more. The outcome of organ transplants has improved significantly since the introduction of HLA matching and immunosuppressive drugs in 1979 (Calne *et al.,* 1979), the latter of which limits immune reactions by suppression of the immune system. The degree of HLA typing for solid organ transplants is different when considering the organ for transplantation. It has been shown that HLA typing has had a remarkable impact in kidney transplants and still remains a priority, while it is less considered in heart transplants. Although studies have indicated the benefit of HLA matching in heart and lung transplantations, there are some contradictions. A study performed in 2005 did not demonstrate a positive correlation with a higher degree of matching at HLA-A, -B, and -DR in heart transplants (Almenar *et al.,* 2005). Various other risk factors have rendered HLA compatibility less concerning in heart transplants, while age, body size, and blood group compatibility between donors and recipients are primarily considered. HLA compatibility is also not the primary consideration for lung transplants. The only reason for HLA typing is in the case of sensitisation.

There are, however, concerns regarding HLA compatibility in liver transplants. A study performed in the 1990's indicated that complete class I matching for liver transplants may have a deleterious effect, even though some matching is desirable for graft survival (Donaldson *et al.,* 1993). Cornea transplants do not require an HLA matched donor, since the cornea is not transplanted to an immunologically privileged site. Mismatches for solid organ transplants are tolerated depending on the organ for transplant; however, cross-matching is a requirement, since it could lead to organ rejection. Cross-matching includes testing for already existing antibodies that could cause an immune reaction to a transplanted organ. If an individual is positive for a cross-match, it indicates a response was generated and that the organ cannot be transplanted, while a negative result indicates the

absence of an immune response. This, however, is not the situation in stem cell transplantations, since a higher degree of matching is essential for engraftment and the prevention of GVHD.

2.10.1.2. Stem Cell Transplantation

The high prevalence of various blood disorders globally necessitates thousands of stem cell transplantations yearly. The fact is that only 25% of these individuals have an adequate HLA-matching donor. There are two types of stem cell transplants, autologous and allogeneic. Autologous transplantations involve the transfer of an individual's own stem cells back into the individual, while allogeneic involves the use of donor stem cells, which has to be a genetic HLA match to the recipient. A study by Jean Dausset showed that grafts between siblings were more successful than grafts between unrelated individuals (Dausset, 1958). The preferred donor, is therefore, a matched HLA relative which is not always possible since siblings are not necessarily HLA-identical. As has been mentioned earlier, the possibility of finding an HLA-identical match between relatives is 25%. It has, therefore, become more common to make use of unrelated donors. The mortality rate is higher in unrelated transplants that mismatch at one or two alleles compared to a fully matched transplant (Flomenberg *et al.,* 2004; Petersdorf *et al.,* 2004; Lee *et al.,* 2007). The difficulty in finding a compatible stem cell donor is attributable to the fact that stem cell transplantations require allelic matching between certain loci, which decreases the possibility of obtaining a fully matched donor. Two distinct classes of stem cells are currently utilised in transplantation for the purpose of assorted therapeutic applications. These include bone marrow and umbilical cord blood (UCB) stem cells.

<u>Transplantations utilising Stem Cells from Bone Marrow</u>

The bone marrow is located within the marrow of long and flat bones and is the site in which virtually all blood stem cells reside, constituting what is defined as the stem cell niche. All blood cells are derived from hematopoietic stem cells (HSCs) and transplantation of these cells involves the transfer of immunocompetent cells from donors to recipients for immune restoration. The ultimate result is for these stem cells to migrate to the bone marrow, a process known as stem cell homing, for complete reconstitution of the damaged/destroyed bone marrow. A new technique enables the collection of peripheral

blood stem cells (PBSCs) through a process known as apheresis. The growth factor, granulocyte colony-stimulating factor (G-CSF), facilitates the mobilisation of stem cells into the bloodstream. The mobilised stem cells are obtained from peripheral blood, which is a less invasive procedure when compared to acquiring it from within the bone. Bone marrow stem cell transplants can be used for myeloablative and non-myeloablative treatment. Myeloablative treatment involves the administration of high doses of chemotherapy in order to deplete the recipient's entire bone marrow of cells prior to transplantation, while non-myeloablative treatment does not involve chemotherapy. When donating bone marrow stem cells, HLA typing is performed for 10 HLA loci, five from each parent (HLA-A, -B, -C, -DRB1, and -DQB1), and a 9/10 or 10/10 match between donors and recipients is a prerequisite. Unrelated donors may be found through a national bone marrow registry. In 1990, the SABMR was introduced and established as an organisation that assists in finding an adequate HLA match for those individuals not fortunate enough to possess an HLA-identical relative. As a result of the high degree of matching necessary for bone marrow transplantation, in order to prevent the rejection of the cells, umbilical cord blood has become an alternative source of stem cells, due to the immature nature of these cells.

Umbilical Cord Blood Transplantation

The use of UCB is a recent development in bone marrow transplantation and is obtained from a newborn baby's placenta through the umbilical vein. The blood from the umbilical cord/placenta is a rich source of stem cells (Rocha *et al*., 2004) and due to the immaturity of the immune cells in cord blood, it is only necessary to perform HLA typing for HLA-A, -B, and -DRB1. The outcome of various studies has suggested that a 4/6 to a 6/6 match is adequate for unrelated donors (Wagner *et al*., 2002; Eapen *et al*., 2007; Barker *et al*., 2010). However, a recent study by Eapen and co-workers suggests that typing of HLA-C should be performed in addition to the three loci that are presently being typed for UCB to minimise the risk of mortality after UCB transplantations (Eapen *et al.,* 2011). A previous study by Petersdorf and co-workers have indicated that matching of the class I alleles is vital, since mismatching at these alleles can cause graft failure (Petersdorf *et al.,* 2001). It has previously been reported that the time required to search for a matching cord blood donor is less than the time it takes to find a matching bone marrow donor (Rocha *et al.*, 2000; Barker *et al.*, 2002), which makes UCB a good option for transplantation.

The potential use of UCB was first proposed in 1982 by Edward Boyse, whereafter the first successful HLA-identical UCB transplant was performed in 1988 (Gluckman *et al*., 1989) by Gluckman and co-workers on a 5 year-old Fanconi's Anaemia patient. The first unrelated UCB transplant was performed in 1993 by Krutzberg and Wagner (Kurtzberg *et al.,* 1996; Wagner *et al.,* 1996*)*. UCB is important in transplantations, since it has enough hematopoietic progenitor cells to support engraftment (Broxmeyer *et al.*, 1989). The rapid proliferative capacity of these cells makes it possible to reconstitute the entire bone marrow (Gluckman *et al*., 1997). UCB is valuable for the reconstitution of hematopoiesis in children with malignant, as well as non-malignant disorders. There are sufficient cells within cord blood to assure long term engraftment and the risk of GVHD is low even in HLA mismatched transplants. Clinical observations have shown that the GVHD is decreased following transplantation, in patients who receive cord blood compared to individuals receiving bone marrow (Cairo and Wagner, 1997). Cord blood stem cells differ from bone marrow stem cells, since cord blood stem cells are "immune naive" due to the minimal previous exposure to antigens (Chalmers *et al*., 1998) and also containing fewer helper T cells (Loetscher *et al.,* 1998). According to two papers published by Bensussan *et al.* (1994) and Berthou *et al.* (1995), cord blood contains increased numbers of NK cells and less cytotoxic T cell activity. Stem cells from cord blood mature into anti-inflammatory interleukin-10 producing cells, which induce tolerance, and could therefore, be responsible for reduced occurrence of GVHD in cord blood transplantations (Bacchetta *et al.,* 1994; Eskdale *et al.*, 1998).

There are several advantages of using cord blood instead of bone marrow. These include: (1) cord blood is readily available following birth, and does not require an invasive procedure for harvesting; (2) it is possible to cater for a greater diversity due to the ability of cord blood stem cells to tolerate a greater degree of HLA disparity; (3) cells have a high proliferative capacity; and (4) there is a decreased rate of acute GVHD following transplantation. Although these cells have a number of advantages and are excellent for transplants due to their immunological characteristics, there are also some disadvantages that need to be considered in making use of cord blood for transplantations: (1) there are no additional donor cells available, whereas in the case of bone marrow, it is possible to obtain more blood from the donor; (2) the low number of cells available due to the small volumes of cord blood readily available, makes it best suited for young individuals under the

age of 30, with a body weight of less the 50 kg; (3) higher risk of infection, since cord blood cells are immune-naive; and (4) in the case where more than one unit is required, which is referred to as pooled or sequential UCB transplantation, all of the units need to be matched to the recipient as well as to each other. The characteristics of umbilical cord blood make it an alternative source to bone marrow, which has led to the establishment of public and private cord blood banks across the globe.

Commercial entities provide an opportunity to store a newborn baby's umbilical cord blood in a private bank for a prolonged period of time in the case of disease development later in life. The blood is only available for personal and family use. The storing of these units is a matter of concern at the moment, since it is still unsure whether viable cells will be retained from the UCB units after storing samples for many years. A major concern is the facilities in which these units are stored and whether they are stored at the satisfactory conditions. It has been estimated that the chances of a child needing their own cord blood ranges from 1 in 10 000 to 1 in 200 000 (Kline, 2001). A 2009 study reported that only nine autologous transplantations were performed compared to 41 allogeneic transplantations from privately banked UCB. The participants of this study do not recommend private storage of UCB for individuals of Northern European descent. However, 11% of the participants recommend private storage in the case where parents have different ethnic minorities (Thornley *et al*., 2009). There is also the option of donation to a public bank facility. Public banks are generally owned by the government and involve allogeneic donations, where the blood is taken from the placenta and stored in a general facility for the use of unrelated cord blood transplantations. In the already existing public cord blood banks across the globe, there is no association between the cord blood samples and a donor or the family. As soon as the cord blood is taken it is government property, and can therefore, be used by any matching individual in need of transplantation. There is, however, a risk which involves infection of the donor's cells that may only become apparent months or years after the cord blood was donated. This underlies the importance of comprehensive screening methods at public cord blood facilities for the detection of genetic and/or infectious diseases. Hybrid models are another possibility for storing umbilical cord blood, which entails merging of private and public cord blood banks. This model provides a facility for the preservation of public, as well as private, umbilical cord blood unit storage.

2.10.2. Forensic and Anthropological studies

Forensic science makes use of various techniques to distinguish one individual from another by particularly targeting the hypervariable regions within the genome. This is due to the fact that it is believed that the genomes between individuals vary in 0.1% (The International SNP Map Working Group, 2001). This analysis can be performed by analysing restriction fragment length polymorphisms (RFLP), mitochondrial DNA (mtDNA), gender typing, Y-chromosome typing, HLA typing, and a few others. Y-chromosome and HLA typing can also be performed for the purpose of paternity exclusion testing, a major aspect in forensic medicine. Paternity testing typically makes use of classical blood tests and HLA typing, due to the high degree of variability of the HLA loci. By determining the HLA genotype, 80% of the male population will be eliminated from being an alleged father. In the case where the father possesses an identical haplotype to that of the child, he could be considered as being a potential father. Although paternity testing has been performed for many years in South Africa, the emergence of DNA evidence in crime investigations has only recently become a reality. The enormous frequency of crime offences is a continuous concern and due to lack of education and training, DNA analysis has not yet been implemented to fulfil its full potential in the area of criminal investigations. A DNA database with the genetic profile of each individual will assist in identifying potential suspects in criminal investigations. However, some regions might be identical between different individuals by pure chance. For ethical, as well as practical reasons, only non-coding regions will be used for identification, since genes are not likely to contain such a high degree of variability to distinguish one individual from another. These non-coding regions contain long stretches of DNA, which makes it the preferred choice in detecting differences between individuals.

Anthropology is the study of human behaviour, origin, physical, and cultural development to determine the extent of diversity between populations. Anthropological studies also analyse variable DNA regions within the genome, such as short tandem repeats (STR), mtDNA, and HLA. A topic of great interest in anthropology, is the discovery of the Neanderthal and Denisovan remains and the ongoing process of identifying how they are genetically related to modern humans. It has been indicated that HLA may not be the best choice for comparison with modern humans, since natural selection causes changes within the HLA genes. However, they could provide valuable information about the short-term history of

population events. As has been mentioned earlier in this chapter, the HLA haplotypes present in Africans are different from the haplotypes present in Caucasians and Asians, which is due to strong natural selection that influences the frequency of HLA alleles observed in populations over time.

2.10.3. Disease Associations

The genetic diversity within the HLA region has been identified to be associated with diseases more than any other region within the human genome. Several HLA alleles are present at an increased frequency in individuals with specific diseases compared to the general population. These HLA alleles can either have a positive or a negative genetic association with a particular disease. This includes autoimmune diseases, infectious diseases, and drug-induced hypersensitivity. Various studies are investigating the HLA antigens and alleles for susceptibility to diseases.

Susceptibility to autoimmune diseases include rheumatoid arthritis (RA), multiple sclerosis (MS), serum lupus erythromatosus, diabetes mellitus, and various other diseases (Table 3), while infectious diseases include human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS), hepatitis, tuberculosis (TB), malaria, and many more (Table 4). Studies have indicated that individuals heterozygous at the class II classical HLA genes are more likely to clear hepatitis B viruses than homozygous individuals (Thursz *et al.,* 1997). A similar concept was observed in HIV-infected individuals. HLA heterozygosity delays the onset of AIDS in HIV-infected individuals, while the presence of the HLA-B*35-C*04 haplotype is associated with rapid progressionn to AIDS in Caucasians (Carrington *et al.,* 1999). Even though heterozygosity and certain HLA alleles confer protection, there are other alleles that increase susceptibility to these diseases, such as HLA-B*35, that is associated with increased progression to AIDS. An awareness of increased susceptibility to certain diseases might encourage lifestyle changes or early treatment that will afford an enhanced quality of life.

Table 3: HLA class I and II allele susceptibility and protection to various autoimmune diseases.

		HLA Class I Effect		HLA Class II Effect		
		Susceptive	**Protective**	**Susceptive**	**Protective**	**References**
Autoimmune Diseases	**Addison's Disease**	-	B15	DQ2 DR3	-	(Gombos *et al.*, 2007; Baker *et al.,* 2010; 2011)
	Ankylosing Spondylitis	B*27:01 B*27:04 B*27:05	B*27:06 B*27:09	-	-	(Gonzalez-Roces *et al.*, 1997)
	Celiac Disease	-	-	DQ2 DQ8	-	(Karell *et al.*, 2003)
	Grave's Disease	-	-	DRB1*03 DRB1*08	DRB1*07	(Chen *et al.*, 1999)
	Multiple Sclerosis	-	-	DRB1*15:01 DRB1*13:03	-	(Lang *et al.*, 2002; IMSGC and Wellcome Trust Case Control Consortium 2, 2011)
	Rheumatoid Arthritis	-	-	DRB1*01 DRB1*04 DRB1*10 DRB1*14	-	(Fries *et al.*, 2002; Hughes *et al.*, 2008)
	Systemic Lupus Erythromatosus	-	-	DRB1*03 DRB1*08 DRB1*15	DRB1*12	(Graham *et al.*, 2002; Shankarkumar *et al.*, 2003; Pan *et al.,* 2009; Farouk *et al.*, 2011)
	Type I Diabetes	A*24 B*18 B*39	A*01 A*11 A*31 B*27	DR3 DR4	DR15 DR14	(Nejentsev *et al.*, 2007)

Table 4: HLA class I and II allele susceptibility and protection to various infectious diseases.

		HLA Class I Effect		HLA Class II Effect		
		Susceptive	**Protective**	**Susceptive**	**Susceptive**	**References**
Infectious Diseases	**Hepatitis B**	-	-	DR*03 DR*07	DR*04 DR*13	(Yan *et al.*, 2012)
	Hepatitis C	B*08 B*18	A*03 B*07 B*27 C*01	DRB1*03	DQB1*03 DRB1*01 DRB1*11	(Tripathy *et al.*, 2009; Fanning *et al.*, 2000 Harris *et al.*, 2008; Thio *et al.*, 2001)
	HIV	B*35	B*27 B*57	-	-	(Gao *et al.*, 2005)
	Leprosy	-	-	DR2 DR3	DRB1*04	(reviewed in Mira, 2006; da Silva *et al.*, 2009)
	Malaria	-	B*53	-	DRB1*13:02 DQB1*05:01	(Hill *et al.*, 1991)
	Tuberculosis (TB)	B*51	B*52	DR2	-	(Vijaya Lakshmi *et al.*, 2006; Kettaneh *et al.*, 2006)

Drug induced hypersensitivity reactions are also a major concern that adds to the already existing healthcare burden, due to the often severe and sometimes fatal adverse effects. Different HLA alleles have been identified to be associated with drug hypersensitivity, which include antiretroviral (ARV) agents, such as Abacavir, Nevirapine, and various other drugs (Table 5). HLA typing is valuable in determining whether an individual possesses certain alleles that might predispose them to having an adverse reaction to various drugs. Knowledge of predisposing alleles will result in alternative drug administration to avoid deleterious effects. This concept ties in with personalised medicine, which attempts to predict the outcome of certain genetic variants. The ultimate purpose of personalised medicine is to prescribe the correct medicine at the correct dosage, depending on an individual's genotype. However, accurate diagnostic tests are required for identification of patients that will be able to benefit from such new therapies.

Table 5: The HLA allele associations with adverse drug reactions.

	Drugs	HLA Class I	HLA Class II	References
Adverse Drug Reactions	**Abacavir**	B*57:01	-	(Mallal *et al*., 2002)
	Nevirapine	B*35:05 C*04 C*08	DRB1*01:01 DQB1	(Gatanaga *et al*., 2007; Chantarangsu *et al*., 2009; Carr *et al*., 2013)
	Carbamazepine	A*31:01 B*15:02	-	(Mehta *et al.,* 2009; McCormack *et al*., 2011)
	Allopurinol	B*58:01	-	(Somkrua *et al*., 2011)
	Ximelagatran	-	DRB1*07:01	(Kindmark *et al*., 2008)

Although many studies have indicated an association of certain HLA antigens and alleles with various diseases, there is still a possibility that multiple alleles combined are responsible for the results observed in these studies. This is not just an indication of the involvement of HLA alleles to disease susceptibility, but could also include the association with other genes that, together, create these deleterious effects.

DEMOGRAPHICS, LINGUISTICS, AND HLA DIVERSITY OF THE SOUTH AFRICAN POPULATION

The African continent is the homeland for all modern humans and consists of immense cultural, linguistic, and genetic diversity compared to other continents, which makes it an excellent region for population variation studies. The fluctuations in climate, diet, and exposure to different pathogens over time have contributed significantly to the genotypic and phenotypic diversity observed in African populations. Ancient humans originated in Africa and were subjected to the largest migration events (Tishkoff and Williams, 2002). Africa is the second largest continent of which the majority of the countries are underdeveloped, and therefore, also under sampled. The majority of studies have focused on the more developed countries and the study subjects, were therefore, not a true representation of the African population. A study conducted by Jorde and co-workers have indicated, together with previous studies, that the majority of variation is observed within populations and not necessarily between different populations. This study also observed the effect of migration on variation, since less differentiation was observed between individuals residing on the same continent (Jorde *et al*., 2000). Common genetic variants often do not differ significantly between populations, especially within the same continent. Rare genetic variants, however, are often population specific (Casals and Bertranpetit, 2012). Migration out of Africa has led to the emergence of various founder populations. Population bottlenecks reduce the amount of diversity of haplotypes in founder populations, and therefore, the diversity observed in other populations is less when compared to African individuals.

The South African population is known as ‘the rainbow nation’, which is appropriate for a country with a cultural diversity emphasized by 11 official languages, mostly indigenous to South Africa. The large number of population groups with diverse gene pools currently residing in South Africa contributes to the diversity observed. The SABMR is mostly comprised of Caucasian individuals, and is therefore, not representative of the entire South African population (SABMR, 2008). Due to the high degree of diversity, and the small number of South African black individuals belonging to the SABMR, it is challenging to

obtain donors for these individuals. It is, therefore, believed that a public cord blood bank could contribute to the resolution of this problem in the short-to-medium term in South Africa. The bank will need to be representative of the South African population and may possibly provide black African individuals, locally and in the global diaspora, with a better chance in obtaining an HLA matching donor. Umbilical cord blood only requires matching at six HLA alleles (three from each parent) and a 4/6 match between donors and recipients is considered to be acceptable for transplantation. Bone marrow transplantations require matching at 10 HLA alleles, with a 9/10 match between donors and recipients. The greater degree of HLA disparity allowed for UCB further increases the chances of obtaining an HLA matching donor, while it might be more difficult for bone marrow.

The diversity of the South African population can be considered at three levels, which include: racial/ethnic groups, linguistic groups, and HLA diversity. In order for a cord blood bank to accommodate the diversity of the South Africans, the aforementioned levels require an in-depth study of the literature to determine the basis on which this bank should be constituted.

3.1. Demographics and Linguistics

South Africa is a multiracial as well as multicultural country that comprises of White, Black, Indian/Asian, and Mixed ancestry individuals that together make up this diverse country of residence. The Census conducted in 2011 indicated that South Africa is presently populated by more than 50 million individuals; 40 million Black (79.5%); 5 million White (8.9%); 4 million Mixed ancestry (8.9%); and 1 million Indian/Asian (2.5%) South African individuals (Figure 14) (Statistics South Africa, 2011).

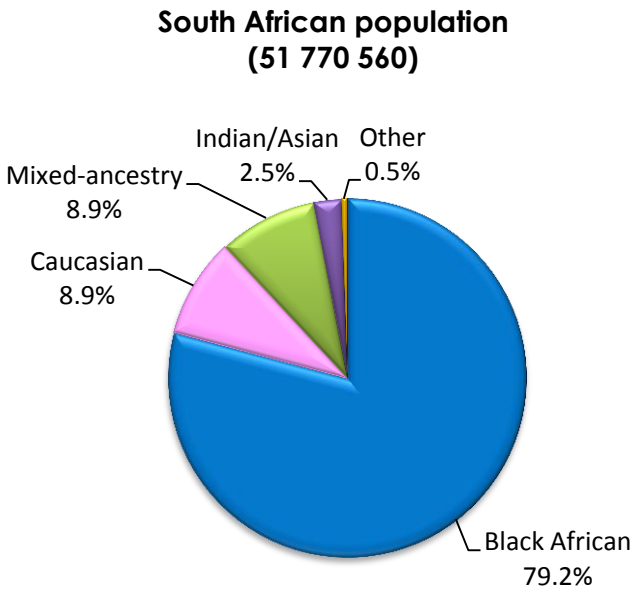
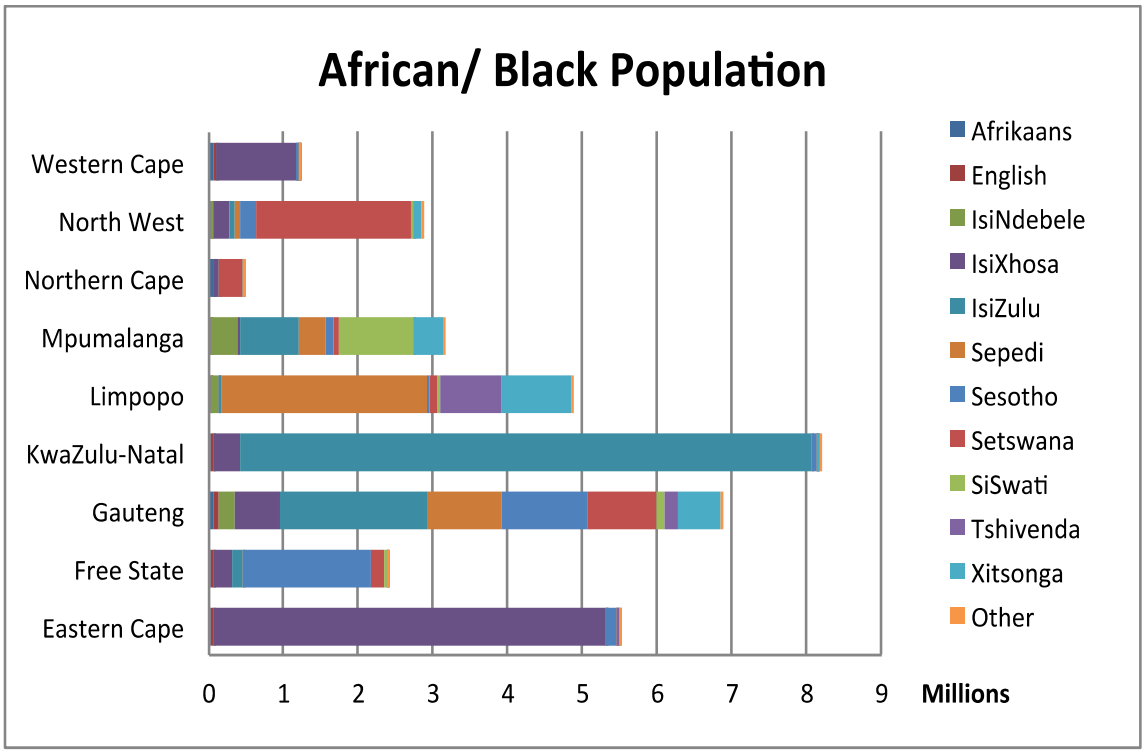


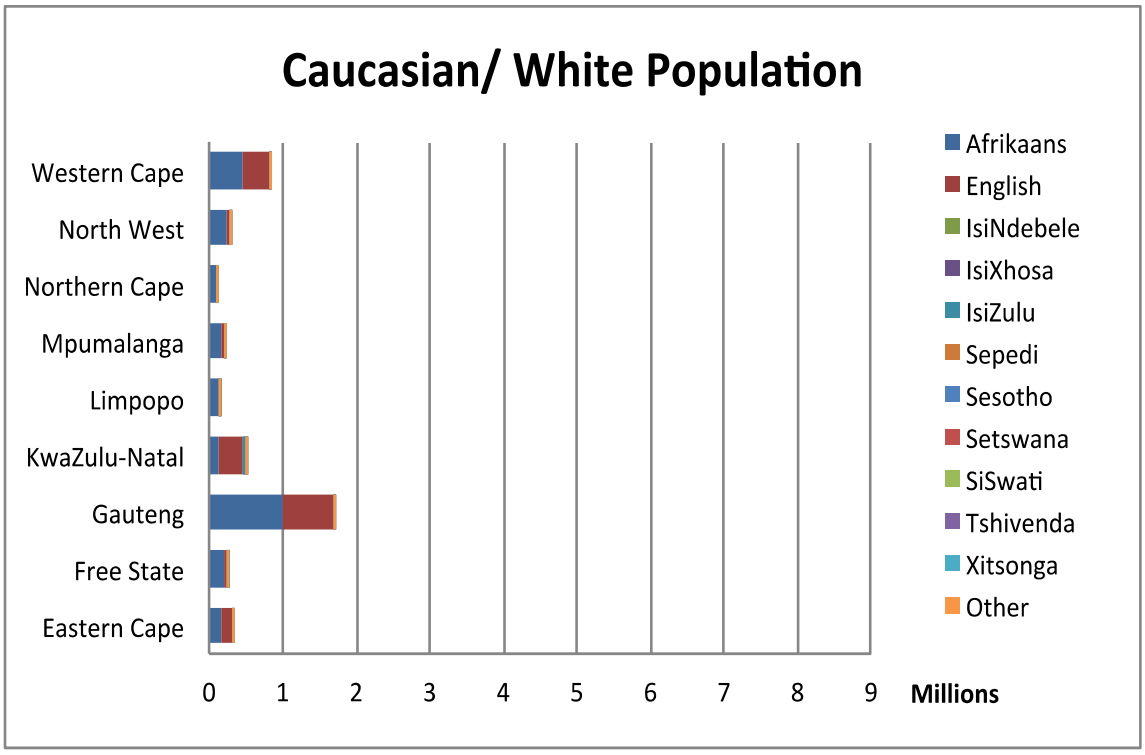
Figure 14: The South African population. The figure depicts the percentages of the different racial groups in South Africa (Data Source: Statistics South Africa, 2011).

South Africa consists of nine provinces that is each comprised of distinct population groups residing within each province. The different population groups are distributed across all of the nine provinces, although the frequency of the linguistic groups differs between the different provinces (Figure 15). The Black South African population consists of four broad groupings, namely: Nguni (Includes Zulu, Xhosa, Ndebele, and Swazi); Sotho-Tswana (Southern, Northern, and Western Sotho population); Tsonga; and Venda; which results in them being the most diverse population group in South Africa (Figure 16). A study by Sonja Bosch and co-workers has indicated the linguistic similarities of the Nguni languages (Bosch *et al*., 2008). A study by Lane and co-workers also indicated that individuals belonging to the Nguni group are genetically more related, which could suggest expansion from a common ancestor. Another cluster is formed by the Sotho-speaking individuals and it has also been shown that the Venda group shows close similarity to the Tsonga group (Lane *et al.*, 2001). Xhosa and Zulu tribes represent the majority of the South African black population. Xhosa and Zulu are classified as Bantu (Nguni) languages. South Africa is known to be the native land of Xhosa-speaking individuals, where they reside mainly in the Eastern Cape.

The language that one speaks is often an indication of the ethnicity of an individual, which could reflect cultural groupings and cross-cultural marriages. The strong religious backgrounds of the majority of South Africans contribute to a large extent in selecting a partner. Populations belonging to the same linguistic group generally tend to cluster together. Even though there are distinct ethnic and cultural groups residing in South Africa, there is a high degree of intermixing between individuals from different linguistic groups (Coetzee *et al.*, 2009). The lowest rate of intermixing is amongst Southern Sotho and Tswana individuals, which is 20% of women and 13% of men; while the highest rate is amongst Xhosa, Northern Sotho, and Tsonga individuals, which is 35-40% of women and 20-23% of men (Statistics South Africa, 2001).

The majority of the White South African population is of European descent and mainly speak Afrikaans and English. Dutch and English were the first official languages of South Africa, since 1910 to 1925, which originated from founder individuals of the colonising population. These individuals came from different nations and different ethnical backgrounds. In 1961, Dutch was replaced by Afrikaans when South Africa became a republic. Nowadays, Afrikaans is spoken by 60% of the White population as a first language, while the other populations speak Afrikaans as a second or a third language. The Mixed ancestry South African population is comprised of individuals of mixed ancestral lineages, of which the majority speak Afrikaans and are distributed mostly in the Western Cape. They are also observed in high numbers in the Eastern and Northern Cape provinces. The Indian and Asian South Africans speak English and mainly live in KwaZulu Natal and Gauteng provinces. There are also a significant number of Asian individuals residing in South Africa, although the majority is due to recent immigrations.





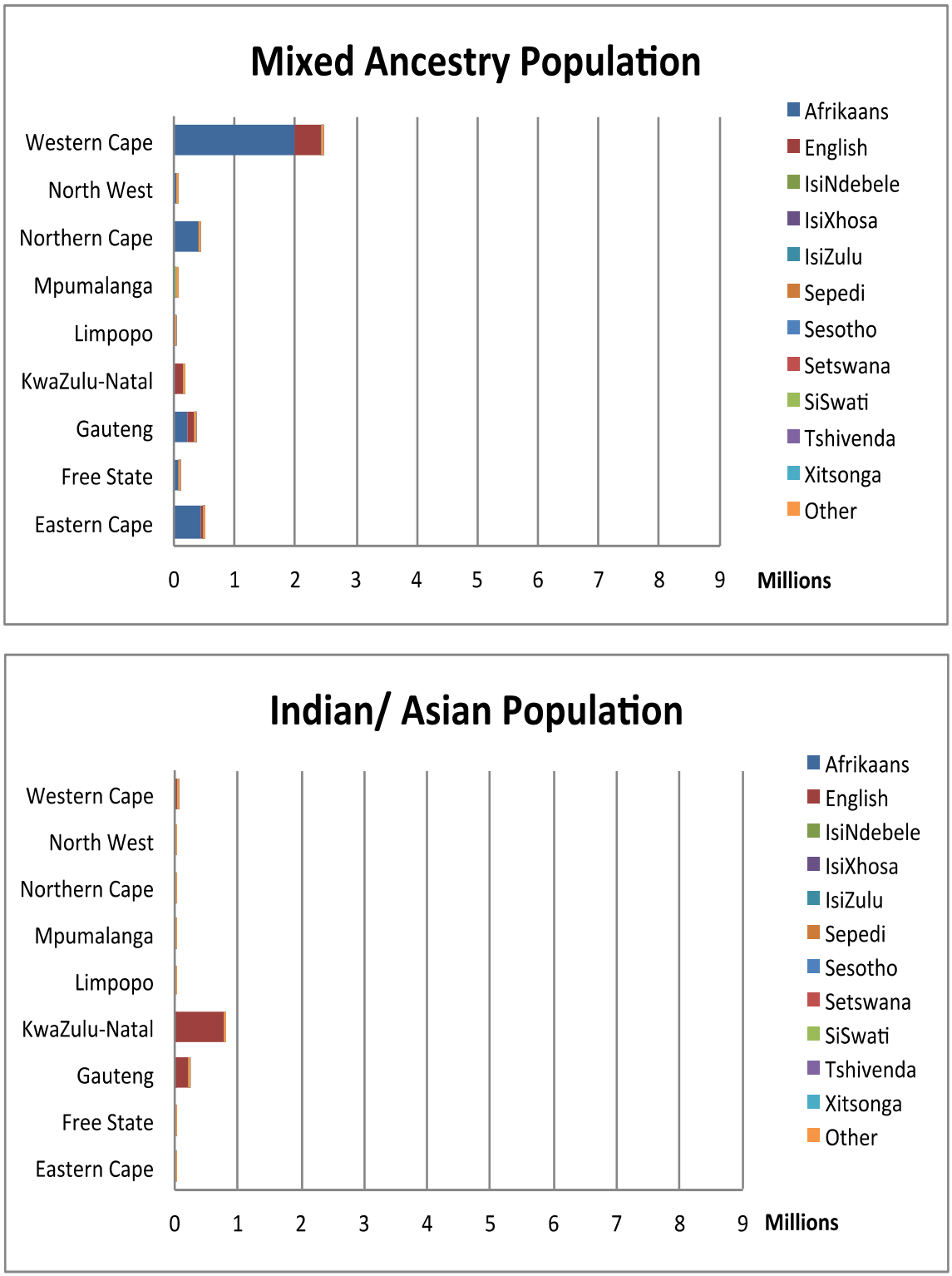


Figure 15: The demographics and linguistics of the South African population. The graphs depict the demographics and the distribution of linguistic groups across the nine different provinces of South Africa (Statistics South Africa, 2001).

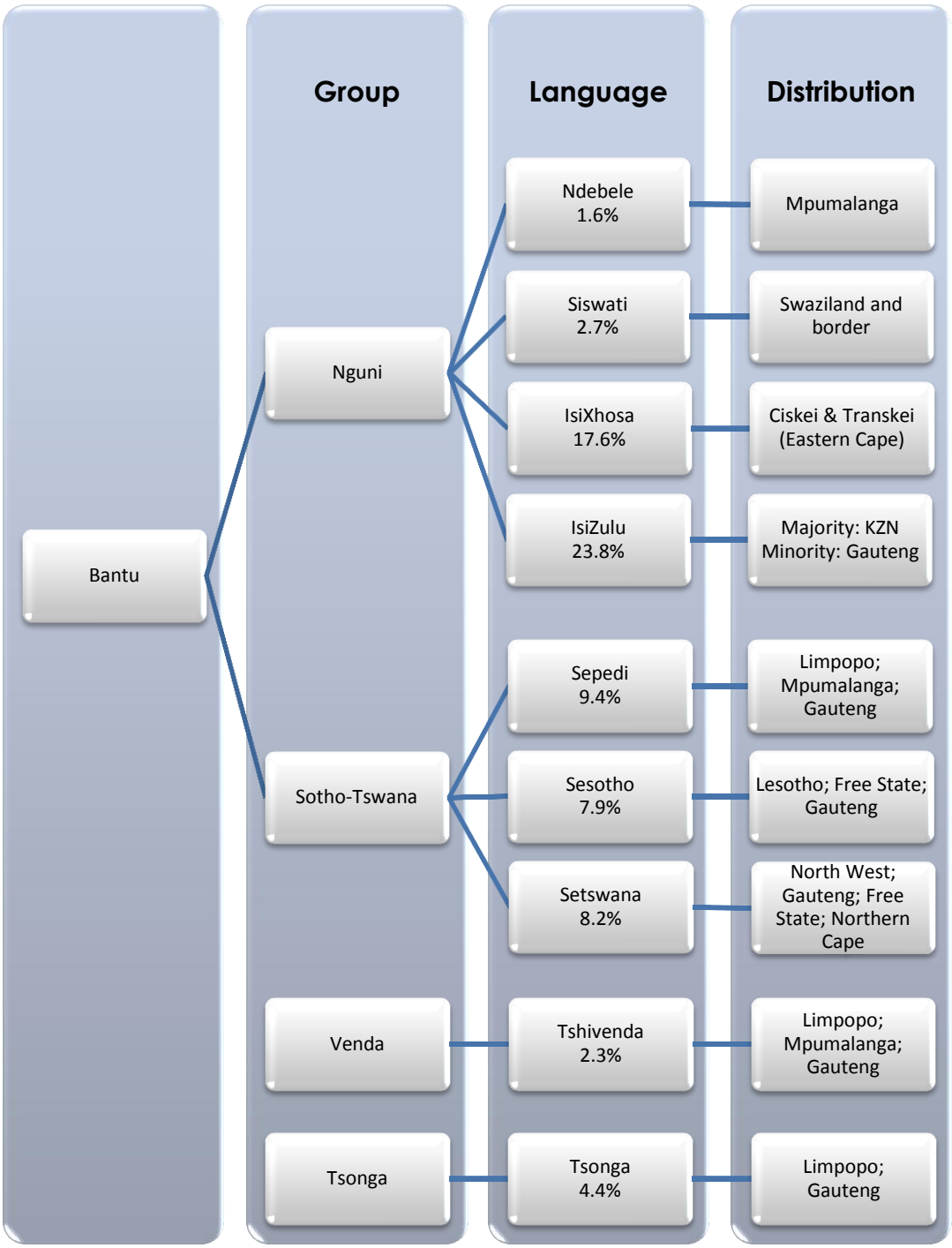
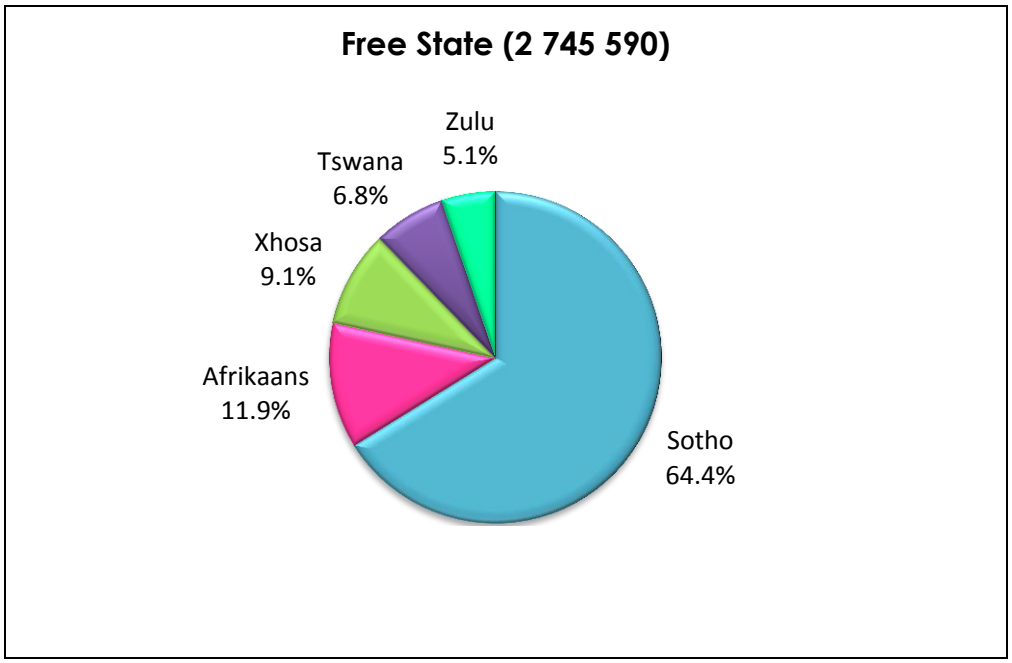
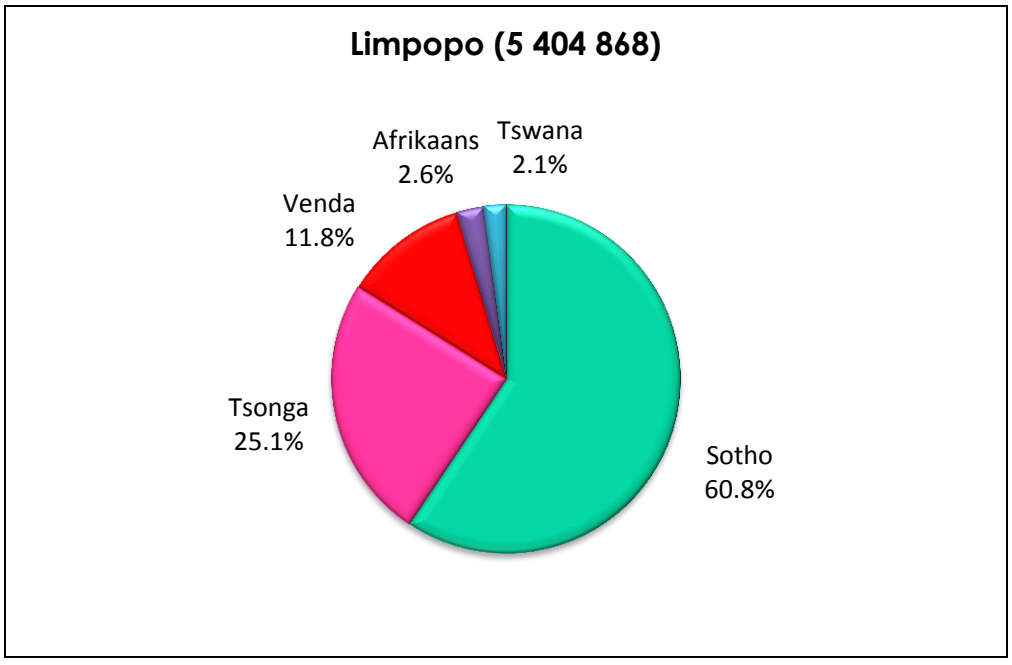


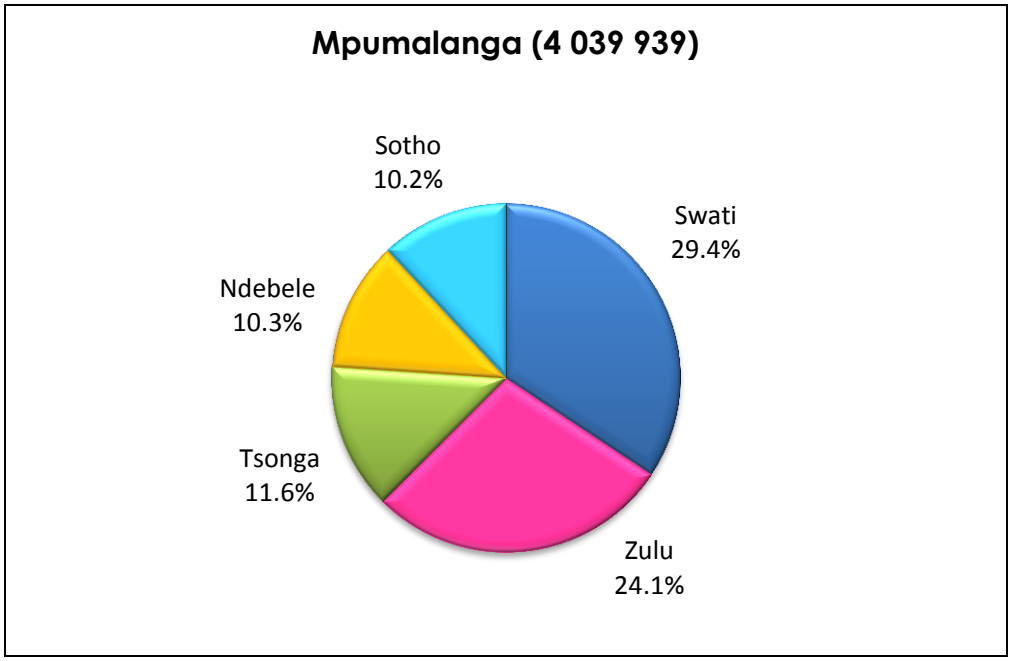
Figure 16: The Black South African languages. The different ethnic groups, languages and the frequency of the language being spoken, and the distribution across provinces (Image created by Juanita Mellet) (Data Source: Statistics South Africa, 2001).

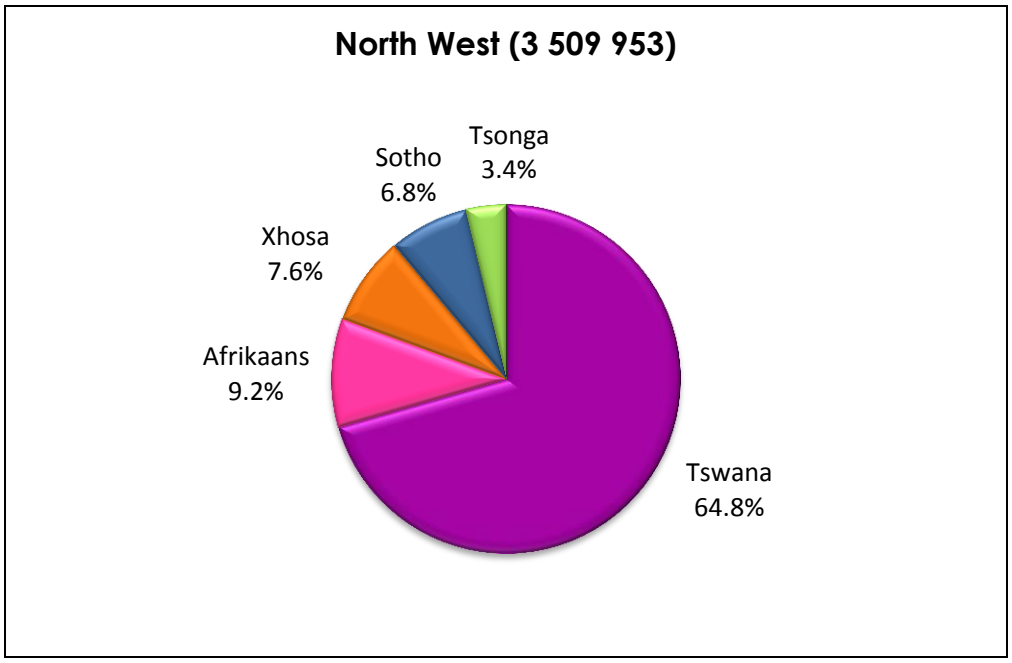
The nine provinces are quite diverse in size, landscape, climate, and economy. Therefore, it is not surprising that it is also diverse in terms of distribution of the different population

groups. Figure 17 illustrates the five most diverse provinces in the country (Free State, Limpopo, Mpumalanga, and North West), where individuals of at least five linguistic groups reside (Statistics South Africa, 2011). Gauteng, however, is far more diverse than the other four provinces in Figure 17. Gauteng is the residence for individuals from all the linguistic groups present in South Africa. This also makes Gauteng best suitable for setting up the first public umbilical cord blood bank, since it will be possible to collect blood from all the different linguistic groups until other units can be set-up in some of the other provinces.









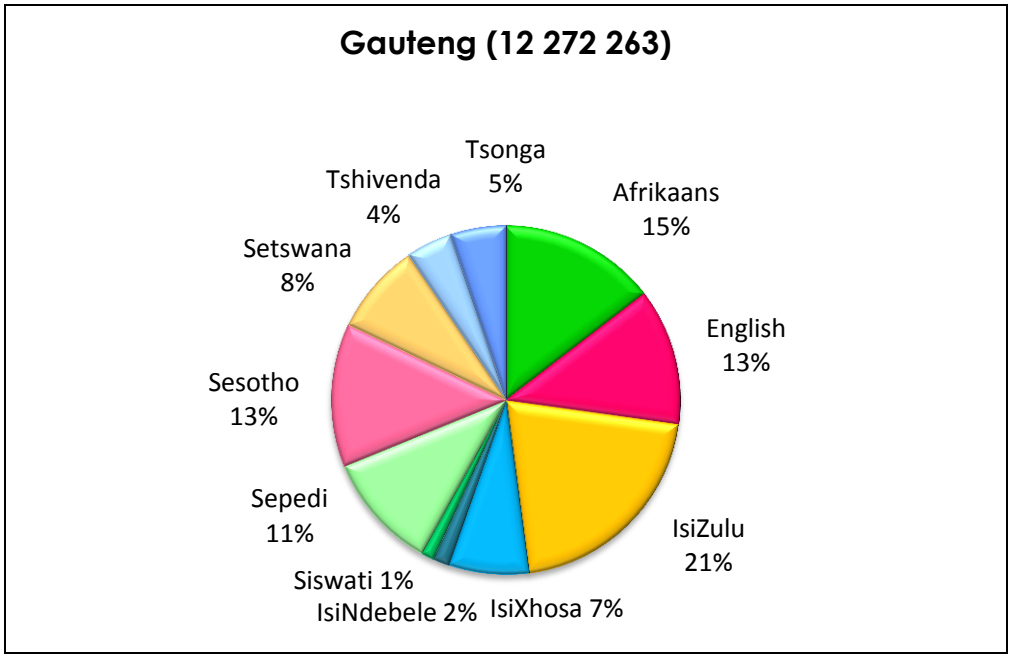


Figure 17: The five most diverse provinces in South Africa. These figures indicate the number of residents and the five most frequently spoken languages in each province (Image created by Juanita Mellet) (Data Source: Statistics South Africa, 2007; 2011).

3.2. HLA Diversity

Although various studies have investigated the distribution of HLA on the African continent, very few have looked at HLA diversity in the South African population. Professor Ernette du Toit from UCT performed several studies in the 1980's, which indicated the distribution and diversity of the HLA alleles in South African populations and especially Black Africans. These studies have focused mainly on the Xhosa, Mixed ancestry, and White/Caucasian populations and HLA typing was performed at low resolution. Table 6 displays the combined results from four different studies performed from 1985-1990 by Ernette du Toit and co-workers (du Toit *et al*., 1988; du Toit *et al*., 1988, p. 42; du Toit *et al.,* 1990a; 1990b).

Table 6: Frequent HLA alleles in the different South African population groups by low resolution typing techniques.

	HLA-A	HLA-B	HLA-C	HLA-D
Black South Africans (Xhosa)	A2[a] A28[a] A29 A30[a]	B7 B42[a] B45 B58[a] B70[a]	C2[a] C6[a] C4 C7	DR2 DR3[a] (DR18; DR17) DR5 DR6[a]
Caucasian South Africans	A1[a] A2[a] A3[a] A11[b]	B7 B8	C3[a] C6 C7	DR1[a] DR2 DR4 DR5 DR7
Mixed ancestry South Africans	A2[a] A3 A30 A43 (Khoisan)	B7 B48 (Mongoloid)	C4 C6 C7	DR2 DR3 DR4 DR5 DR6 DR7
Nigerians[b]	A2 A28 A30 A33	B53 B70	C4 C6 C7	DR2 DR5 DR6

[a]*Observed in all four studies for a specific population.*

[b]*Okoye et al., 1985.*

The combined results indicate the antigens observed at a frequency of more than 10% in the different population groups that were studied. The HLA typing performed for all of these studies were serological in nature and able to distinguish major structural and functional variation between antigens. South African Blacks had an increased number of antigens compared to the other population groups. Various antigens have been shown to be frequent in Black South Africans, while other antigens, such as B*27 is either rare or absent in this group. The Mixed ancestry South Africans share antigens with Black South Africans as well as with White South Africans. They also possess A*43 that is commonly

observed in the Khoisan population and B*48 that has only previously been observed in Mongoloids and could indicate possible admixture.

The results obtained for the Black South African population (Xhosa) were also compared with results from a study performed by Okoye and co-workers (Okoye *et al.*, 1985). Nigerians and Black South Africans share some common antigens, while other antigens were particularly observed in each group. By making use of newly developed high resolution techniques, it would be possible to determine the in depth variation that might go undetected in serological or low resolution typing methods.

A recently performed study by Paximadis and co-workers has also investigated HLA diversity in the South African population (Table 7) (Paximadis *et al.,* 2011). This study included individuals from all the linguistic groups in South Africa, of which the highest numbers were Zulu and Xhosa individuals. This study made use of more recent techniques that performed HLA typing at high resolution (four-digit level).

Table 7: Frequent HLA alleles observed in the South African population by high resolution typing techniques.

	HLA-A	HLA-B	HLA-C	HLA-D
Black South Africans	A*30:01 A*30:02	B*58:02 B*42:01	C*04:01 C*06:02 C*17:01	DRB1*03:01 DRB1*03:02 DRB1*11:01 DRB1*13:01
Caucasian South Africans	A*01:01:01 A*02:01:01 A*03:01:01	B*07:02:01 B*08:01	C*07:01 C*07:02:01	DRB1*03:01 DRB1*04:01 DRB1*15:01

The results from Paximadis and co-workers were consistent with the results from the previous studies and various new alleles were discovered in the different population groups. Even though these alleles were not present at high frequencies, this highlights the diversity of the HLA genes and the increasing discovery of new alleles. It is difficult to compare the different studies that have been performed on the South Africa populations due to differences in population groups, sample sizes, and the different typing techniques used.

There are many challenges that have arisen in attempting to establish the basis on which the public cord blood bank would need to be constituted. These include (1) insufficient data on the HLA frequencies in the South African population. This makes it difficult to determine what alleles are frequent in the different population/linguistic groups; (2) studies that have been performed are not representative of the entire South African population. The 1980's studies focused mainly on the Xhosa, White, and Mixed ancestry South African populations, while the 2011 study included individuals from all the different linguistic groups, although sample sizes remained relatively small; (3) the difficulty in comparing the results from the nineteenth century studies with the more recent study due to the differences in the typing methods that were used, since the resolution (two and four-digit level) differs; and (4) the great degree of HLA diversity observed within and between the South African population groups.

The information that has been gathered and studied makes it possible to conclude to some extent that it is less suitable to constitute the first public umbilical cord blood bank in South Africa based on HLA diversity, due to the vast HLA diversity within the South African population. Intermixing between individuals belonging to distinct linguistic groups is frequently observed and due to extreme patterns of diversity between the different linguistic groups, it will not be possible to determine the ethnicity of an individual based on the language that one speaks. Hence, it is not recommended to constitute the bank on the basis of language.

It is, therefore, recommended to constitute the bank based on social race or major ethnic groupings (Nguni, Sotho-Tswana, Tshivenda, and Xitsonga). The bank would have to be representative of the entire South African population. It has been estimated that a minimum number of 10 000 cord blood units would be needed to initiate the bank, of which 8 000 (80%) would have to be representative of Black South Africans, 900 (9%) of White South Africans, 800 (9%) of Mixed ancestry South Africans, and 300 (3%) of Indian/Asian South Africans.

MATERIALS AND METHODOLOGY

This study was conducted in the Department of Immunology at the University of Pretoria (UP). Ethical approval was granted in 2010 by the Faculty of Health Sciences Research Ethics Committee (*Protocol Number 131/2010*) for the project entitled "Feasibility study for a public cord blood stem cell bank in South Africa", of which the currently proposed project forms part. The proposal for this dissertation was submitted at the end of the year 2011, followed by ethical approval, which was granted by the Ethics Committee of UP (*Protocol number 219/2011*) (*see Appendix A: University of Pretoria Ethics Approval Certificate*). Separate ethical approval was granted by UCT (*Protocol number 523/2011*) for the use of 20 samples obtained from UCT (*see Appendix B: University of Cape Town Ethics Approval Certificate*). Twenty DNA samples were obtained from the Department of Immunology at UCT, courtesy of Professor Clive Gray.

Each DNA sample had already been HLA typed by the Laboratory of Tissue Immunology (LTI) at the time of commencement of the present study. The HLA typing was performed by low or high resolution DNA typing techniques. Various typing techniques were used for HLA typing of these samples (Table 8). Serological techniques provide information on the antigen group (two-digit level resolution). DNA-based methods were initially performed as a supplement to serology. However, they are now used routinely in many laboratories as the 'gold standard'. Sequence specific oligonucleotide probe (SSOP) kits (Gen-Probe, California, USA) were used for HLA class I and II low resolution typing. The second set of digits corresponds to the specific protein present in an individual, which is determined by DNA-based techniques and provides information on the DNA sequence. Sequence-specific primers (SSP) from Invitrogen (Invitrogen, California, USA) and Olerup (Olerup, Oslo, Norway) were used for class II high resolution (four-digit level resolution) HLA typing.

Table 8: Low to high resolution HLA typing techniques.

	Summary	Advantages	Disadvantages
Serology	Serology (antibody-based) makes use of viable lymphocytes to determine the molecular markers (antigens) present on the surface of these cells. B-lymphocytes are required for HLA Class II typing. Serology provides information on the antigen family.	- Quick and simple method - Cost effective - Identifies molecules on the surface of the cells	- Cross-matching - Requires viable lymphocytes - Difficult to obtain antisera for rare antigens - Not suitable for bone marrow transplants
DNA-Based Typing	The DNA-based typing techniques make use of sequence-specific oligonucleotide probes (SSOP) and primers (SSP) to detect genetic variants. SSP products are viewed by gel electrophoresis.	- Gold Standard - Accurate, robust, and time efficient - Low to high resolution - Identify genetic differences	- Not advantageous for large number of samples due to gel electrophoresis. - Dependant on previously identified polymorphisms - Mistyping can occur in the case where an individual possesses a novel allele
SBT	Sequence-based typing techniques make use of specific primers to amplify the polymorphic regions of the gene, after which the base sequence is determined.	- Accurate - High resolution typing - High throughput - Identifies nucleotide sequence of an allele - Able to identify novel alleles	- Requires expensive equipment - Time consuming - Labour intensive

4.1. Sample Selection and Quantification

This study made use of 20 genomic DNA samples isolated from peripheral blood mononuclear cells (PBMCs). DNA extractions were performed using the QIAamp® Blood Mini kit (Qiagen, Hilden, Germany). The procedure required at least 160 ng of DNA for each sample with a concentration of 5 ng/µl (According to manufacturer's guidelines).

The 20 samples from UCT were quantified using the NanoDrop® ND-1000 version 3.3 spectrophotometer (Thermo Scientific, Massachusetts, USA) upon arrival at UP. The NanoDrop calculates DNA concentrations by absorbance of Ultraviolet (UV) light at 260 nm. The ratio of absorbance between DNA and protein (260/280) was used to assess the purity of the DNA. An aliquot of 1 µl of each sample was used to assess the concentration. AE buffer (the substance that DNA was suspended in) (Appendix C) was used as a blank. The peak was at 260 nm and the 260/280 ratio for all the samples was between 1.8 and 1.9.

Quantification with the NanoDrop does not accurately assess DNA concentrations. Therefore, an alternative method of quantification was used to verify that the concentrations obtained from the NanoDrop were correct. This was performed using a Quant-iT™ PicoGreen® double-stranded DNA assay kit (Invitrogen), a fluorescent nucleic acid stain used to quantify double-stranded DNA in solution. The measurements were taken in triplicate by using the NanoDrop® ND-3300 fluorospectrometer (Thermo Scientific).

A standard curve was created with Lambda DNA that accompanied the PicoGreen assay kit (Figure 18). The standard curve consists of serial dilutions that create known DNA concentrations (Table 9). The R^2 value of the standard curve had to be at least 0.98. The DNA standard curve and sample reading values were used to determine the concentrations of all the sample amplicons. The NanoDrop provided sample concentrations in ng/ml, which were converted to ng/µl using the following equation: Sample concentration (ng/µl) = sample concentration (ng/ml)/1000*dilution factor*amount of PicoGreen (µl).

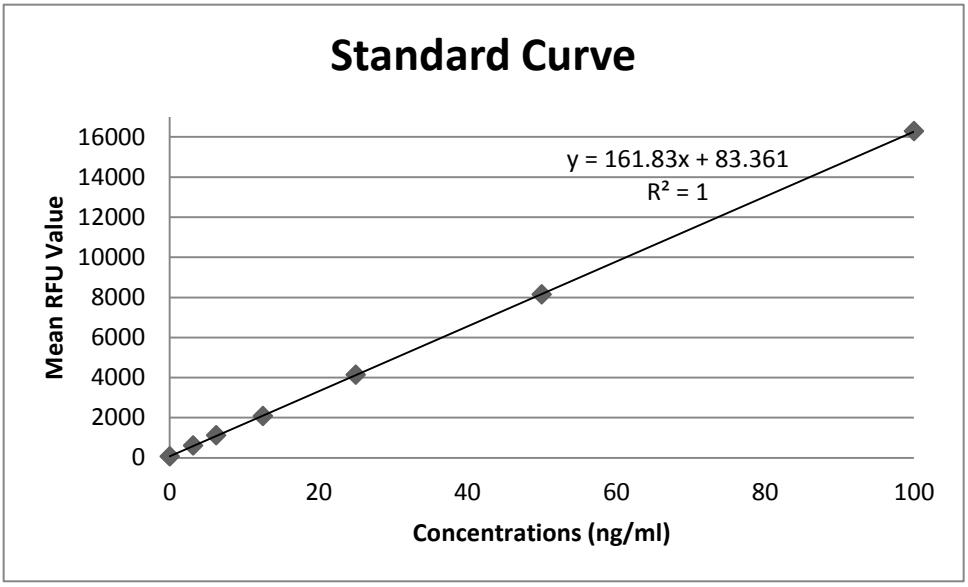


Figure 18: Standard Curve.

Table 9: DNA concentrations of the 8-point standard curve.

Tube #	Well	DNA Amount
Tube 1	A12	200.00 ng/well
Tube 2	B12	100.00 ng/well
Tube 3	C12	50.00 ng/well
Tube 4	D12	25.00 ng/well
Tube 5	E12	12.50 ng/well
Tube 6	F12	6.25 ng/well
Tube 7	G12	3.125 ng/well
Tube 8	G12	0.00 ng/well

4.2. Integrity Check by Agarose Gel Electrophoresis

Agarose gel (1%) electrophoresis was used to determine whether samples were of genomic nature and to ensure that no degradation had occurred (Figure 19).

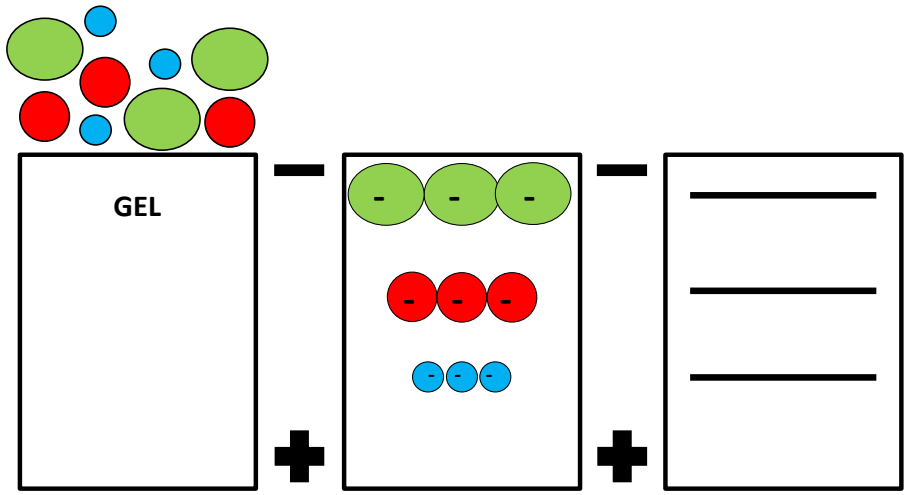


Figure 19: Electrophoretic separation of DNA. The movement of charged particles through an electric field (Image created by Juanita Mellet) (Data Source: Russel, 2006).

The agarose gel was prepared as follows:

- Seakem® LE Agarose powder (Lonza, Basel, Switzerland) was weighed in a heat resistant Schott bottle.
- Tris Borate EDTA (10x) electrophoresis buffer (Fermentas Life Sciences, Burlington, Canada) was diluted with deionised water to create a 1x electrophoresis buffer.
- The buffer was added to the agarose powder and heated to dissolve the powder.
- Once the solution was clear, it was left to cool and 5 µl of Gel Red™ nucleic acid gel stain (Biotium, San Francisco, USA) was added for visualisation under UV light.
- The solution was poured into a gel plate and left to set for a minimum of 30 minutes. A gel comb was inserted to create the wells.

Gel electrophoresis was performed with 2 µl of sample DNA and 3 µl of 6x Mass Ruler™ loading dye (Fermentas Life Sciences). The loading dye and DNA were mixed prior to loading into the wells. The DNA migrated at a low velocity due to the size of genomic DNA. The gel was visualised using a VersaDoc™ imaging system (BioRad, California, USA) together with the Quantity One software program. A solid band was observed for each sample.

4.3. GS GType HLA Typing Kit and Primers

A Roche designed GS GType HLA typing kit was used for typing of the samples. This is a SBT assay, which made use of the 454 NGS instrument (GS Junior Instrument) (Roche Applied Science, Penzberg, Germany). The typing test targets the most hypervariable and mostly studied exons of the MHC region. The GS GType HLA Primer sets were made available as two sets of kits, medium resolution (MR) and high resolution (HR). The two kits contained different dried down primers for the amplification of various regions of different HLA genes. The GS GType HLA MR Primer set (blue plate) contained eight different primer sets for the MHC class I and II exons. The GS GType HLA HR Primer set (yellow plate) contained six additional primer sets for the MHC class I and II exons. The HR set was used in combination with the MR set to resolve some ambiguity. All primers for samples and negative controls were labelled with multiplex identifiers (MIDs). This was the same across MR and HR plates (Figure 20).

GS GType HLA MR Primer Set

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10	Neg. Control	Blank
	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8	MID9	MID10	MID11	
A	A-2	A-2	A-2	A-2	A-2	A-2	A-2	A-2	A-2	A-2	A-2	
B	A-3	A-3	A-3	A-3	A-3	A-3	A-3	A-3	A-3	A-3	A-3	
C	B-2	B-2	B-2	B-2	B-2	B-2	B-2	B-2	B-2	B-2	B-2	
D	B-3	B-3	B-3	B-3	B-3	B-3	B-3	B-3	B-3	B-3	B-3	
E	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	
F	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	
G	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	DQB1-2	
H	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	DRBx-2	

GS GType HLA HR Primer Set

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10	Neg. Control	Blank
	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8	MID9	MID10	MID11	
A	A-4	A-4	A-4	A-4	A-4	A-4	A-4	A-4	A-4	A-4	A-4	
B	B-4	B-4	B-4	B-4	B-4	B-4	B-4	B-4	B-4	B-4	B-4	
C	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	
D	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	DPB1-2	
E	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	DQA1-2	
F	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	DQB1-3	
G												
H												

Figure 20: Medium and high resolution plate layout. The disposition of the exon specific primers and samples on the 96-well plates for MR as well as HR primer sets. The letters indicate the MHC target gene, while the digits indicate the exon of interest. Samples are organised by column and the MID's (grey area) are used to identify each sample and the negative control.

4.4. Polymerase Chain Reaction

Polymerase chain reaction is a fast and inexpensive method used to amplify small segments of DNA. This method is used to amplify the DNA region in question and to ensure that a sufficient amount of DNA is present for molecular and genetic analysis.

The kits for this study contained 14 primer sets (two primers for a specific region of the MHC gene) with unique sequences complementary to a specific region of the MHC (Table 10). There were eight primer sets for the MR plate and six for the HR plate. Each one of the 10 samples also had a unique MID tag sequence (Table 11) for automated software identification after pooling and sequencing. Figure 21 indicates the position of the MID tags relative to the sequence of interest and sequencing primers.

Table 10: The primer sequences and amplicons' fragment length (post-PCR).

Plate	Amplicon	Gene Specific Primer Sequence (5'-3')	Amplicon Length (bp)
MR	HLA-A2 5'	GAAACGGCCTCTGTGGGGAGAAGCAA	514
	HLA-A2 3'	GGTGGATCTCGGACCCGGAGACTGT	
	HLA-A3 5'	GACTGGGCTGACCGCGGGGT	479
	HLA-A3 3'	GAGGGTGATATTCTAGTGTTGGTCCCAA	
	HLA-B2 5'	AGAGCTCGGGAGGAGCGAGGGGACCGCAG	446
	HLA-B2 3'	ACTCGAGGCCTCGCTCTGGTTGTAGTA	
	HLA-B3 5'	AGAGCTCGGGCCAGGGTCTCACA	425
	HLA-B3 3'	ACTCGAGGGAGGCCATCCCCGGCGACCTAT	
	HLA-C2 5'	AGTCGACGAAGCGGCCTCTGCGGA	481
	HLA-C2 3'	ACTCGAGGGGCCGGGGTCACTCA	
	HLA-C3 5'	ACGTCGACGGGCCAGGGTCTCACA	653
	HLA-C3 3'	ACCTCGAGGTCAGCAGCCTGACCACA	
	DQB1-2 5'	AGGATCCCCGCAGAGGATTTCGTGTACCA	370
	DQB1-2 3'	TCCTGCAGGACGCTCACCTCTCCGCTGCA	
	DRB1-x 5'	CCGGATCCTTCGTGTCCCCACAGCACG	366
	DRB1-x 3'	CCGAATTCCGCTGCACTGTGAAGCTCTC	
HR	HLA-A4 5'	GGTTCTGTGCTCTCTTCCCCAT	746
	HLA-A4 3'	GGGCTTGGAACCCTCAGTGA	
	HLA-B4 5'	CTGGTCACATGGGTGGTCC	448
	HLA-B4 3'	AGATATGACCCCTCATCCC	
	HLA-C4 5'	CAAAGTGTCTGAATTTTCTGACTCTTCCC	446
	HLA-C4 3'	TGAAGGGCTCCAGAAGGACTT, TGAAGGGCTCCAGGACTT	
	DPB1-2 5'	GCTGCAGGAGAGTGGCGCCTCCGCTCAT	411
	DPB1-2 3'	CGGATCCGGCCCAAAGCCCTCACTC	
	DQA1-2 5'	GTTTCTTCCATCATTTTGTGTATTAAGGT	490
	DQA1-2 3'	CCATGAGAAGATCTGGGGACCTCT	
	DQB1-3 5'	TGGAGCCCACAGTGACCATCTCC	451
	DQB1-3 3'	AGTGACATCAGGGATAAGAGATGGGAA	

Table 11: The MID sequence tags for the GS GType HLA typing kit. This table indicates the position of the tag on the plate, MID name, and sequence.

Column on plate	MID name	MID sequence
1	MID1	ACGAGTGCGT
2	MID3	AGACGCACTC
3	MID4	AGCACTGTAG
4	MID6	ATATCGCGAG
5	MID7	CGTGTCTCTA
6	MID8	CTCGCGTGTC
7	MID9	TAGTATCAGC
8	MID10	TCTCTATGCG
9	MID13	CATAGTAGTG
10	MID16	TCACGTACTA
11 (negative control)	MID11	TGATACGTCT

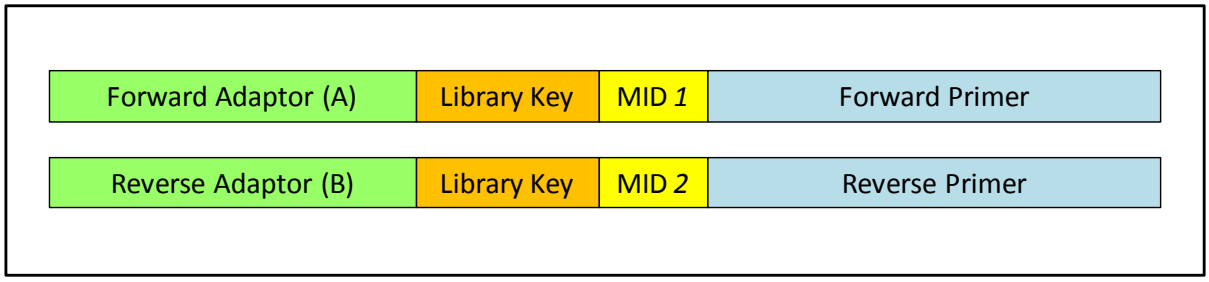


Figure 21: The position of the MID tags and primers, relative to the DNA sequence of interest (Image created by Juanita Mellet, adapted from Figure 1 in Bentley et al., 2009).

The PCR step included the preparation of a master mix consisting of the reagents listed in Table 12. Each PCR step included the processing of two plates, one MR and one HR plate and the volumes in the table were calculated accordingly.

Table 12: The reagents for making up the PCR master mix.

PCR Master Mix for HR Setup	Stock Concentration	1 well (µl)	2 PCR plates (µl)
Molecular Biology Grade Water (Sigma Aldrich, Missouri, USA)	-	14.80	2516
Glycerol (Merck, New Jersey, USA)	80%, w/v	3.125	531.25
PCR Buffer II (Applied Biosystems, California, USA)	10x	2.50	425
$MgCl_2$ (Applied Biosystems)	25 mM	1.80	306
PCR Nucleotide Mix (Roche Applied Science)	10 mM	0.375	63.75
AmpliTaq Gold DNA Polymerase (Applied Biosystems)	5 U/µl	0.40	68

PCR was carried out on the GeneAmp® 9700 PCR System (Applied Biosystems) using the PCR conditions, as indicated in Table 13.

Table 13: PCR protocol

Temperature (˚C)	Duration	
95.0	10 min	
95.0	15 sec	x35 cycles
62.0	30 sec	
72.0	5 min	
4.0	∞	

4.5. Amplicon Purification

The purification step for this study was performed using AMPure XP Beads® (Beckman Coulter, California, USA). This highly efficient method is primarily used for purifying amplicons exceeding 100 bp. This process is based on a magnetic bead-based solid phase reversible immobilisation (SPRI) technology. The magnetic microparticles target size-specific

nucleic acids and immobilise them on these particles by making use of specific buffer conditions. The advantage of this technology is that it has the ability to recover approximately 80% of the input DNA after purification and the yield of high quality data.

The Ampure Beads were added to the amplified DNA amplicons. These beads are magnetic, and therefore, attached to the DNA amplicons after PCR. By making use of the Ambion® 96-well magnetic-ring stand (Applied Biosystems), DNA was separated from the contaminants present in the sample liquid (Figure 22). Samples were washed subsequently with 70% Ethanol (Appendix C) in order to rid the samples of contaminants. Such contaminant could interfere with the downstream processes and reduce the number of usable reads obtained from sequencing. TE buffer (Appendix C) was used to elute the DNA from the magnetic beads.

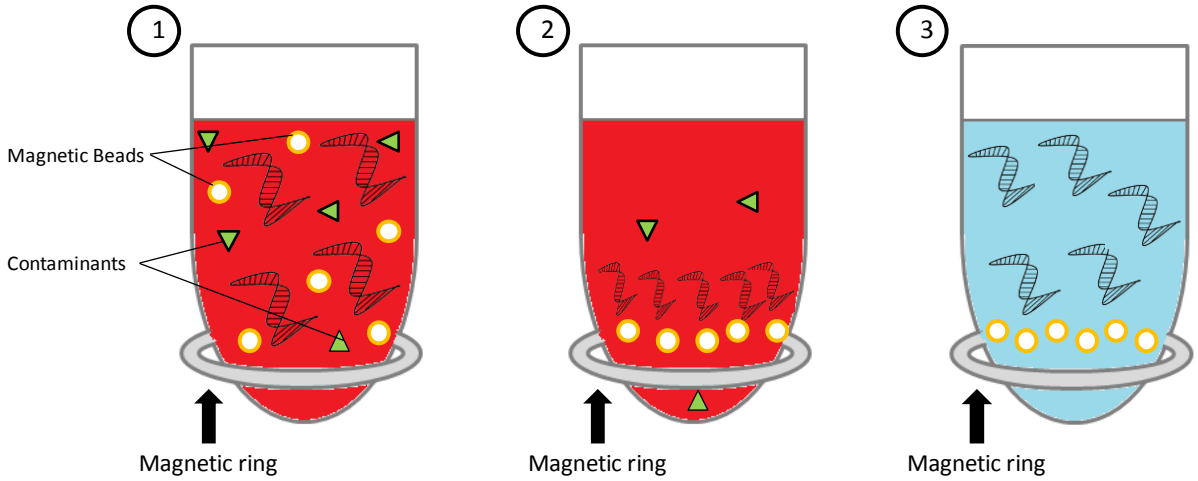


Figure 22: Amplicon purification after PCR. (1) Ampure beads were added to the sample and attached to the nucleic acids for immobilisation, while contaminants remain in solution. (2) A magnetic ring stand is used to produce a magnetic field that pulls the beads out of solution. This separates contaminants and subsequent washing of nucleic acids produces high quality DNA. (3) The addition of TE Buffer to the sample, eluted the nucleic acids from the magnetic beads (Image created by Juanita Mellet, adapted from an image on the Beckman Coulter website: www.beckmancoulter.com).

4.6. Amplicon Quantitation

Quantitation of the amplicons was also performed using the Quant-iT™ PicoGreen® double-stranded DNA assay kit (Invitrogen). This was performed in the same manner as previously described in section 4.2 of this chapter. The measurements were taken in duplicate due to the small volumes available. In the case where sample concentrations were less than 5 ng/µl, samples were further analysed on the Agilent 2100 Bioanalyzer (Agilent Technologies, California, USA).

4.7. Amplicon Normalisation and Sample Pooling

The Emulsion PCR (emPCR) amplification step required that all amplicons be present at equimolar ratios. The HLA Amplicon Dilution Calculator (available on www.454.com/my454) was used for computation of dilutions required for normalisation (Figure 23). The calculator was pre-populated with the names (blue column) and sizes (grey column) of amplicons. The spreadsheet automatically calculated the amount of 1x TE buffer and DNA (dark green column) necessary for each sample. The normalisation of the samples produced a concentration of 1 x 10^9 molecules/µl. The calculator also indicated the amount of each sample that was needed to create the equimolar pool for emPCR (light green column).

454 SEQUENCING **HLA Amplicon Dilution Calculator** Roche

Date/Time:	
Run Name:	
Other Information:	

MR Medium Resolution Plate — paste amplicon conc. values below ▼ — prepare these dilutions to normalize samples to 1 x 10^9 — volume for pools

sample #	plate well position	amplicon	band size	conc (ng/µL)	µL sample	◄	µL TE	µL for pool
Sample 1	A1	HLA-A exon 2	541	2.47	2.0	◄	6.4	5
	B1	HLA-A exon 3	479	5.01	1.0		8.6	5
	C1	HLA-B exon 2	446	2.71	2.0	◄	9.1	5
	D1	HLA-B exon 3	425	6.61	1.0		13.3	5
	E1	HLA-C exon 2	481	1.58	2.0	◄	4.0	5
	F1	HLA-C exon 3	653	0.30	USE 5µl NEAT	◄	X	5
	G1	DQB1 exon 2	370	13.30	1.0		32.0	5
	H1	DRBx exon 2	366	21.47	1.0		52.8	5

454 SEQUENCING **HLA Amplicon Dilution Calculator** Roche

Date/Time:	
Run Name:	
Other Information:	

HR High Resolution Plate — paste amplicon conc. values below ▼ — prepare these dilutions to normalize samples to 1 x 10^9 — volume for pools

sample #	plate well position	amplicon	band size	conc (ng/µL)	µL sample	◄	µL TE	µL for pool
Sample 1	A1	HLA-A exon 4	746	0.14	USE 5µl NEAT	◄	X	5
	B1	HLA-B exon 4	448	3.72	1.0		6.6	5
	C1	HLA-C exon 4	446	1.46	2.0	◄	4.0	5
	D1	DPB1 exon 2	411	25.43	1.0		55.8	5
	E1	DQA1 exon 2	490	1.71	2.0	◄	4.4	5
	F1	DQB1 exon 3	451	13.48	1.0		26.4	5

Figure 23: A representation of the GS Gtype HLA Assay Amplicon Dilution Calculator for MR and HR runs (GS Gtype HLA Assay Manual, March 2011).

There are two 454 sequencing systems that can be used for the sequencing of these samples, the GS FLX and the GS Junior systems. This study made use of the GS Junior system, which is able to run five samples of 14 amplicons (MR and HR) each on one full run. The experiments performed included the processing of two full plates (MR and HR) at a time, and therefore, each emPCR step consisted of 10 samples, which made up two complete runs on the GS Junior sequencer.

Amplicon and sample pooling were performed for 10 samples (two plates) of 14 amplicons each. The amount of 1x TE buffer and DNA were added based on the volumes from the dilution calculator. Pooling of 10 samples for the GS Junior system consisted of two pools

(A and B) of five samples each, while the negative control were added to pool B. The samples had to be at a final concentration of 2 x 10^6 molecules/µl for emPCR.

4.8. Emulsion PCR and Sequencing

The emPCR and sequencing steps were performed by Inqaba Biotec™ (Pretoria). Emulsion PCR and bead recovery were carried out as described in the emPCR amplification method manual (Roche Applied Sciences, emPCR Amplification Method Manual, March 2012), while the GS Junior sequencing run was carried out as described in the GS Junior sequencing method manual (Roche Applied Sciences, Sequencing Method Manual, GS Junior Series, March 2012). Emulsion PCR is the process of isolating DNA to beads by small adaptors attached to the end of the DNA fragments. These adaptors provide primer sequences for amplification and sequencing of a specific region. One adaptor contains a 5'-biotin tag that immobilises the DNA to the streptavidin-coated beads. The immobilised DNA on the beads was emulsified with the amplification reagents to create water-in-oil microreactors (Figure 24a). DNA attaches to capture-beads by adaptors. Each bead carries a unique single-stranded DNA fragment (Figure 24b). Each bead was captured in a microreactor where amplification occurred. After amplification, beads with attached amplicons were loaded into a PicoTiterPlate (PTP) (Figure 24c and e). These plates were designed to load a single bead into each well of the plate, which allows the sequencing of one bead at a time and produces data accordingly. The well is made compact by adding tiny beads and sequencing enzymes to the plate (Figure 24d and f).

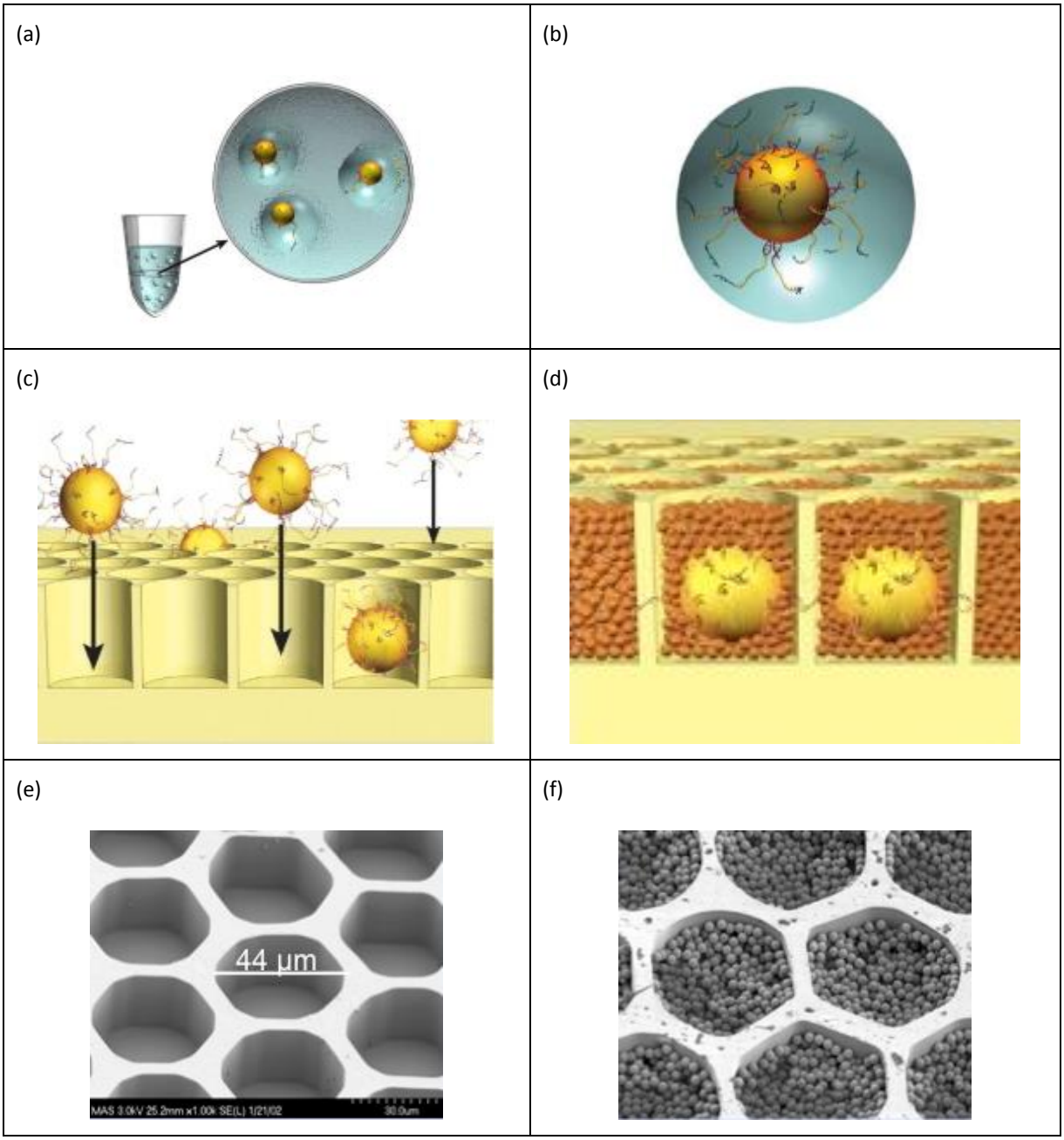


Figure 24: DNA is mixed with capture-beads. (a) Beads and PCR reagents get emulsified in water-in-oil amplification microreactors. (b) DNA attaches to capture-beads by adaptors. Each bead carries a unique single-stranded DNA fragment. (c) Beads were loaded into the picotiterplate and (d) the wells were filled with beads and sequencing enzymes. (e) A Scanning Electron micrograph of the picotiterplate. (f) A micrograph of the tiny beads that were added to the wells (Reproduced with permission from Roche).

A particular sequencing primer was annealed to the DNA strand as an adaptor for the DNA strand to be sequenced, whereafter DNA polymerase synthesized the complementary strand, illustrated in Figure 25. Each deoxynucleotide triphosphate (dNTP) that is being

added at a time is detected and removed after the reaction by the enzyme Apyrase. The enzymes for this reaction are subsequently degraded by heat-inactivation. The incorporation of a single dNTP causes pyrophosphate (PPi) to be released. In the presence of adenosine 5' phosphosulphate, adenosine triphosphate (ATP) sulphurylase converts PPi to ATP. The conversion of luciferin to oxyluciferin requires ATP. This conversion generates visible light to the amount of ATP being produced (Ahmadian *et al*., 2006). The light is detected and represented in a pyrogram, where the heights of the peaks indicate the amount of nucleotides at a particular position.

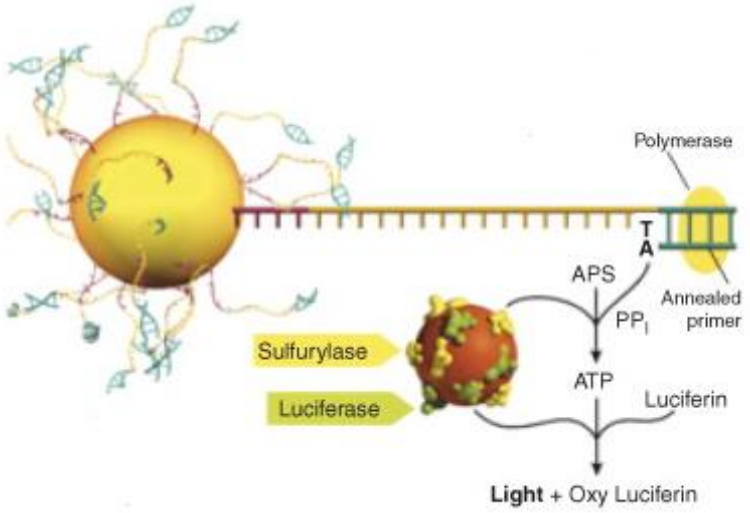


Figure 25: A schematic representation of the process of 454 sequencing (Reproduced with permission from Roche).

4.9. Software Analysis

The raw sequencing data were analysed by Roche (USA) on *Conexio Genomics ATF 454* software (version 3.2.0) (Conexio Genomics, Fremantle, Australia) and by Juanita Mellet on JSI SeqHLA 454 software (version 3.11.0) (JSI Medical Systems, Kippenheim, Germany). The software was designed for sequence reads according to the primer sequences. Sample sequences were compared to the sequences currently in the IMGT/HLA database (Robinson *et al.,* 2013). The sequence reads were sorted to individual samples according to the MID tags and to a specific locus by primer sequences. A genotype was assigned to each sample based on the sequences present in the database. In some cases a mismatch was present at

one or more bases, which indicates the presence of a relatively rare sequence that have not yet been documented in the database.

4.10. Statistical Analysis

The statistical analysis for this study was performed in consultation with Professor Piet Becker (Biostatistician) from the Biostatistics unit at the Medical Research Council (MRC), Pretoria. This was an exploratory study that assessed the agreement between the results from different HLA typing methods in order to determine whether NGS HLA typing is able to determine the HLA types that have been determined by conventional typing methods.

CONVENTIONAL TECHNIQUES *VERSUS* NEXT GENERATION SEQUENCING FOR HLA TYPING

5.1. Introduction

The success of the first twin transplantation in 1954 (Merrill *et al.,* 1956) and the various transplantations that followed thereafter, highlighted the importance of HLA matching between donors and recipients. Only a tiny percentage of the entire population possesses a twin or an HLA identical sibling. Registries and public cord blood banks across the globe provide unrelated HLA donors for patients in need of transplantation. All units that enter a registry or public bank need to undergo HLA typing for the various loci of interest. There are currently many techniques available for HLA typing, although not all are accurate in determining the exact HLA genotypes of individuals. It is, therefore, essential to determine which technique will best suit the HLA typing needs for the establishment of a public cord blood bank in South Africa.

In the past 60 years, HLA typing has moved from serological to cellular and currently the most recent molecular typing techniques. During the 1960's, serological techniques used to be the 'gold standard' for tissue typing, since this was the only method of determining individuals' HLA surface markers at the time. Serological techniques detect antigenic differences on the surfaces of the cells by antibodies (antisera). Even though this method is still being used in several laboratories, there are numerous limitations, that include cross-reactivity and the inability to accurately assign antigens. Cross-reactivity is the process in which the same antigen is recognised by multiple antibodies, that could lead to false positive results. In the mid 1990's, DNA-based techniques became more popular and were used in addition to serological techniques due to the inability of serological techniques to accurately and reliably assign antigens (Bozón *et al*., 1997). Many laboratories have discontinued the use of serological techniques and are simply making use of DNA and SBT techniques. Low to high resolution DNA-based typing methods that are frequently used in various laboratories are PCR-SSOP and PCR-SSP. The low to high resolution techniques have the ability to accurately genotype individuals but due to the ever increasing number of new alleles, it becomes challenging to accurately determine genotypes, since the assignment of

genotypes relies on previously identified alleles. Low resolution typing techniques have proven sufficient for solid organ transplants. In contrast, the transplantation of stem cells requires accurate high resolution genotyping of the HLA alleles for matching between donors and recipients. Inaccurate typing could lead to inadequate HLA matching between donors and recipients, that could ultimately increase the chances of rejection of transplanted tissue and the possibility of developing GVHD.

DNA and SBT techniques have contributed significantly to the current knowledge available on the HLA genes and their immense allelic diversity. DNA-based techniques are able to accurately assign HLA genotypes. In some cases complementary experiments are required to resolve ambiguous results. Sequence-based typing methods provide the highest possible resolution.

Since the completion of the human genome project (HGP) in 2003, and the discovery of the diversity that exists between individuals, various high-throughput NGS technologies have emerged. The initial goals of these technologies were to (1) increase the number of bases per run; (2) increase reads per run; (3) decrease the cost per base; (4) achieve higher accuracy; (5) achieve higher coverage; and (6) increased speed. The 454 platform was the first NGS system to be launched in 2005. Since then several other platforms have emerged. These platforms were initially only used for genomic sequencing but also has the potential for research and diagnostic purposes in various fields. Even though these newly developed techniques have already shown potential in identifying novel alleles, the more conventional techniques are still preferred for routine procedures. The high degree of diversity present in the genes of the HLA region has resulted in the development of a kit that would assign HLA genotypes in a single run. The first potential NGS diagnostics kit for HLA typing was launched in 2011 by Roche. This kit enables multiplex typing of the various polymorphic loci of the HLA region. The ambiguous nature of these genes makes it challenging to assign accurate genotypes. Until recently, only exons 2 and 3 were routinely sequenced in determining the HLA genotypes of individuals. This newly developed HLA typing technique attempted to reduce the ambiguous typing results by sequencing additional exons (Table 14). A study by

Lank and co-workers has identified that the sequencing of exons 2, 3, and 4 enables differentiation between 85% of all documented HLA alleles (Lank *et al*., 2012).

Table 14: GS Gtype HLA primer sets for medium and high resolution typing.

High Resolution (MR + HR)			
Medium Resolution			
GS Gtype MR Primer Set		**GS Gtype HR Primer Set**	
HLA-A	exons 2, 3	HLA-A	exon 4
HLA-B	exons 2, 3	HLA-B	exon 4
HLA-C	exons 2, 3	HLA-C	exon 4
DQB1	exon 2	DQB1	exon 3
DRB1 1,3,4,5	exon 2	DPB1	exon 2
		DQA1	exon 2

(Available at: www.454.com website)

It is well known that the HLA genes are some of the most diverse genes in the human body. The purpose of this study was to determine (1) whether NGS HLA typing is efficient in determining the HLA alleles in South African individuals, (2) the degree of complexity that this technique reveals, and (3) the value of NGS HLA typing in establishing a cord blood bank.

5.2. Results and Discussion

This study performed HLA genotyping by 454 NGS on 20 samples that had previously been HLA typed by conventional methods at the LTI.

5.2.1. Software Analysis

The raw data generated by sequencing was analysed by two different software programs, Conexio 454 ATF and JSI medical (SeqHLA 454). The Conexio analysis was performed by another laboratory in the USA. The JSI medical software for 454 HLA sequencing displayed the combined, sorted, and aligned results, which will be described in more detail below

(Figure 26). Sequences generated from 454 sequencing were compared with sequences already present in the HLA/IMGT database. The HLA database version 3.11.0 was installed prior to commencement of the analysis. The HLA database version consists of all the current allele sequences present in the database. Genotypes were assigned when full sequence alignment occurred with a sequence present in the database. The consensus sequence is displayed at the top of the view panel and displays the sequence of all the possible alleles for a specific gene. The combined sequence is the sequence combined for haplotype 1 and 2 for the data. The sequencing data is visualised as a pseudo-electropherogram, representing the coverage for each base and the minimum coverage as a red dotted line. The forward and reverse reads for exon 2 is also indicated and the coverage for each sequence at the left, with a (*) indicating a perfect match. The perfect match table (left of the sequences) shows perfectly matched alleles for selected exons. Heterozygous positions are highlighted in royal blue and indicate the differences between haplotype 1 and 2 for selected exons. In cases where an allele cannot be accurately assigned, a list of possible alleles will be listed in the results table (Figure 26a). The fragment sequences are the sequences highlighted with light blue colour. Italicised letters indicate reverse strands. The matching table at the bottom of the view panel indicates the number of matches to certain sequences, sorted by probability. The most probable allele is listed at the top, while the least probable allele is listed at the bottom. Alleles with mismatches (highlighted in red) are also displayed in the match table and will indicate in which exon the mismatch occurs. In the case where there are no mismatches for a given sequence, the alleles will be listed in the results window.

(a)

(b)

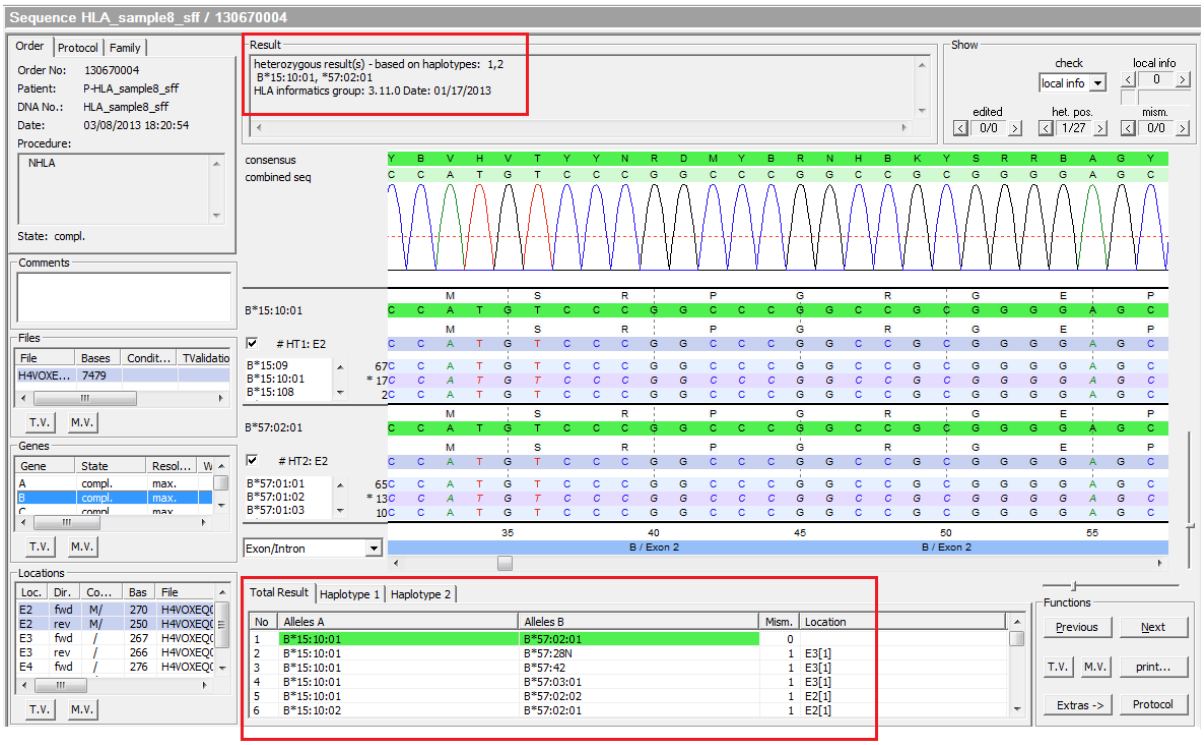


*Figure 26: JSI SeqHLA 454 software results for HLA-B. The screenshots (cropped) display the typing results at the B locus for samples 1 and 8. (a) The result window (red rectangular boxes) for sample 1 indicates a heterozygote with a genotype ambiguity string. (b) The result window for sample 8 indicates a heterozygote with unambiguous typing results, genotype assignment B*15:10:01/B*57:02:01.*

5.2.2. Allelic and Genotypic Agreement

The results from this study are displayed in Tables 15 (samples 1-10) and 16 (samples 11-20) below. The first 10 samples (samples 1-10) have previously been typed by high resolution SBT at class I and SSP at class II loci, while the second 10 samples (samples 11-20) were typed by low resolution Luminex at class I and high resolution SSP at class II loci. These tables display the results from the previously identified genotypes by conventional techniques (column 1), and the results obtained from this study by NGS MR (column 2) and HR (column 3). The conventional techniques display results at a two- to four-digit level of resolution, depending on the technique that was used. The 454 sequencing technique displays results at a four- to eight-digit level of resolution, although only the four-digit level is indicated here for comparison purposes. Alleles were said to be in agreement when the most probable allele assigned was identical to the conventional typing result at the four-digit level. Disagreement between the different techniques is indicated in blue.

Side-by-side comparison of the results obtained from conventional and 454 NGS typing for samples 1-10 showed 96% and 99% agreement with conventional typing techniques for MR and HR, respectively. The results for samples 11-20 showed 95% and 98% agreement with conventional typing techniques for MR and HR, respectively. It was possible to assign accurate genotypes to 95.5% of the loci of interest for the total number of 20 samples by MR, compared to 98.5% for HR. In some cases, manual editing of the sequences was required due to the technical limitations of 454 NGS.

Table 15: HLA typing results for samples 1-10 by conventional and NGS typing techniques.

	ID	Ethnicity	Conventional Techniques (SBT and SSP)		454 NGS (MR)		454 NGS (HR)	
HLA-A	1	Mixed ancestry	02:01	03:01	02:01	03:01	02:01	03:01
	2	Tanzanian	30:02	68:02	30:02	68:02	30:02	68:02
	3	SA Black	68:02	74:01	68:02	74:01	68:02	74:01
	4	Mixed ancestry	02:01	66:01	02:01	66:01	02:01	66:01
	5	SA Black	02:01	29:02	02:01	29:02	02:01	29:02
	6	Mixed ancestry	03:01	11:01	03:01	11:01	03:01	11:01
	7	Kenyan	02:01	02:02	02:01	02:02	02:01	02:02
	8	SA Black	68:02	68:02	68:02	68:02	68:02	68:02
	9	SA Black	68:02	74:01	68:02	74:01	68:02	74:01
	10	Mixed ancestry	30:01	43:01	30:01	43:01	30:01	43:01
HLA-B	1	Mixed ancestry	07:02	08:01	07:02	08:01	07:02	08:01
	2	Tanzanian	08:01	44:03	08:01	44:03	08:01	44:03
	3	SA Black	07:02	15:03	07:02	15:03	07:02	15:03
	4	Mixed ancestry	13:02	35:02	13:02	35:02	13:02	35:02
	5	SA Black	45:01	**45:07**	45:01	**45:01***	45:01	45:07
	6	Mixed ancestry	07:02	**07:06**	07:02	**07:05***	07:02	**07:05***
	7	Kenyan	45:01	51:01	45:01	51:01	45:01	51:01
	8	SA Black	15:10	57:02	15:10	57:02	15:10	57:02
	9	SA Black	15:03	15:10	15:03	15:10	15:03	15:10
	10	Mixed ancestry	15:10	42:01	15:10	42:01	15:10	42:01
HLA-C	1	Mixed ancestry	07:01	07:02	07:01	07:02	07:01	07:02
	2	Tanzanian	07:01	14:03	07:01	14:03	07:01	14:03
	3	SA Black	02:10	07:02	02:10	07:02	02:10	07:02
	4	Mixed ancestry	04:01	06:02	04:01	06:02	04:01	06:02
	5	SA Black	06:02	16:01	06:02	16:01	06:02	16:01
	6	Mixed ancestry	07:02	07:02	07:02	07:02	07:02	07:02
	7	Kenyan	16:01	16:01	16:01	16:01	16:01	16:01
	8	SA Black	03:04	18:01	03:04	18:01	03:04	18:01
	9	SA Black	02:10	08:04	02:10	08:04	02:10	08:04
	10	Mixed ancestry	04:01	17:01	04:01	17:01	04:01	17:01
HLA-DRB1	1	Mixed ancestry	01:01	03:01	01:01	03:01	01:01	03:01
	2	Tanzanian	03:01	13:02	03:01	13:02	03:01	13:02
	3	SA Black	11:01	13:02	11:01	13:02	11:01	13:02
	4	Mixed ancestry	07:01	11:04	07:01	11:04	07:01	11:04
	5	SA Black	11:02	13:01	11:02	13:01	11:02	13:01
	6	Mixed ancestry	15:01	15:01	15:01	15:01	15:01	15:01
	7	Kenyan	03:01	15:03	03:01	15:03	03:01	15:03
	8	SA Black	03:01	13:02	03:01	13:02	03:01	13:02
	9	SA Black	11:01	13:02	11:01	13:02	11:01	13:02
	10	Mixed ancestry	03:02	04:01	03:02	04:01	03:02	04:01
HLA-DQB1	1	Mixed ancestry	02:01	05:01	02:01	05:01	02:01	05:01
	2	Tanzanian	02:01	06:04	02:01	06:04	02:01	06:04
	3	SA Black	**03:19**	06:09	**03:01***	06:09	03:19	06:09
	4	Mixed ancestry	02:02	03:01	02:02	03:01	02:02	03:01
	5	SA Black	03:01	06:03	03:01	06:03	03:01	06:03
	6	Mixed ancestry	05:02	06:02	05:02	06:02	05:02	06:02
	7	Kenyan	02:01	06:02	02:01	06:02	02:01	06:02
	8	SA Black	02:01	06:09	02:01	06:09	02:01	06:09
	9	SA Black	06:02	06:09	06:02	06:09	06:02	06:09
	10	Mixed ancestry	03:02	04:02	03:02	04:02	03:02	04:02

Table 16: HLA typing results for samples 11-20 by conventional and NGS typing techniques.

	ID	Ethnicity	Conventional Techniques (Luminex and SSP)		454 NGS (MR)		454 NGS (HR)	
HLA-A	11	SA Black	29:XX	36:01	29:02	36:01	29:02	36:01
	12	SA Black	23:XX	43:XX	23:01	43:01	23:01	43:01
	13	SA Black	23:XX	66:XX	23:01	66:01	23:01	66:01
	14	SA Black	26:XX	80:XX	26:01	80:01	26:01	80:01
	15	Mixed ancestry	02:XX	68:XX	02:03	68:02	02:03	68:02
	16	SA Black	23:XX	34:XX	23:01	34:02	23:01	34:02
	17	SA Black	24:XX	68:XX	24:02	68:01	24:02	68:01
	18	Mixed ancestry	02:XX	29:XX	02:01	29:01	02:01	29:01
	19	SA Black	03:XX	34:XX	03:01	34:02	03:01	34:02
	20	SA Black	30:01	30:01	30:01	30:01	30:01	30:01
HLA-B	11	SA Black	44:XX	53:XX	44:03	53:01	44:03	53:01
	12	SA Black	**15:03**	15:03	**15:01**	15:03	**15:01**	15:03
	13	SA Black	45:XX	58:XX	45:01	58:02	45:01	58:02
	14	SA Black	15:01	18:XX	15:01	18:01	15:01	18:01
	15	Mixed ancestry	51:XX	53:XX	51:01	53:01	51:01	53:01
	16	SA Black	07:XX	44:XX	07:05	44:03	07:05	44:03
	17	SA Black	08:XX	58:XX	08:01	58:02	08:01	58:02
	18	Mixed ancestry	40:XX	15:03	40:01	15:03	40:01	15:03
	19	SA Black	44:XX	44:XX	44:03	44:03	44:03	44:03
	20	SA Black	15:03	58:XX	15:03	58:01	15:03	58:01
HLA-C	11	SA Black	04:XX	07:XX	04:01	07:01	04:01	07:01
	12	SA Black	04:XX	18:XX	04:01	18:01	04:01	18:01
	13	SA Black	06:XX	16:XX	06:02	16:01	06:02	16:01
	14	SA Black	02:XX	04:XX	02:02	04:01	02:02	04:01
	15	Mixed ancestry	04:XX	14:XX	04:01	14:02	04:01	14:02
	16	SA Black	04:XX	07:XX	04:01	07:02	04:01	07:02
	17	SA Black	06:XX	07:XX	06:02	07:02	06:02	07:02
	18	Mixed ancestry	03:XX	04:XX	03:04	04:01	03:04	04:01
	19	SA Black	02:XX	04:XX	02:10	04:01	02:10	04:01
	20	SA Black	02:XX	06:XX	02:10	06:02	02:10	06:02
HLA-DRB1	11	SA Black	11:01	11:01	11:01	11:01	11:01	11:01
	12	SA Black	03:01	15:01	03:01	15:01	03:01	15:01
	13	SA Black	12:01	13:01	12:01	13:01	12:01	13:01
	14	SA Black	04:05	07:01	04:05	07:01	04:05	07:01
	15	Mixed ancestry	01:02	14:04	01:02	14:04	01:02	14:04
	16	SA Black	03:01	13:01	03:01	13:01	03:01	13:01
	17	SA Black	12:01	13:01	12:01	13:01	12:01	13:01
	18	Mixed ancestry	07:01	13:01	07:01	13:01	07:01	13:01
	19	SA Black	13:01	**15:02**	13:01	**15:01**	13:01	**15:01**
	20	SA Black	04:04	08:04	04:04	08:04	04:04	08:04
HLA-DQB1	11	SA Black	06:02	06:02	06:02	06:02	06:02	06:02
	12	SA Black	03:01	06:02	03:01	06:02	03:01	06:02
	13	SA Black	03:01	06:02	03:01	06:02	03:01	06:02
	14	SA Black	02:02	03:02	02:02	03:02	02:02	03:02
	15	Mixed ancestry	05:01	05:03	05:01	05:03	05:01	05:03
	16	SA Black	03:01	06:03	03:01	06:03	03:01	06:03
	17	SA Black	05:01	06:03	05:01	06:03	05:01	06:03
	18	Mixed ancestry	**02:02**	06:03	**02:01***	06:03	02:02	06:03
	19	SA Black	06:02	06:03	06:02	06:03	06:02	06:03
	20	SA Black	**03:19**	08:04	**03:01***	08:04	03:19	08:04

The genotypic disagreement between the conventional techniques and 454 NGS MR typing was mainly due to ambiguous typing results, indicated by (*). Many alleles are identical across exons 2 and 3 of the HLA genes, since countless polymorphisms are located outside the sequenced region (Robinson *et al*., 2013). Therefore, it is challenging to accurately assign genotypes by only sequencing these two exons. Several ambiguous typing results were obtained from the results produced by only the MR kit. Sequencing of additional exons with the HR kit resolved several ambiguities. Disagreement between results obtained from conventional and 454 NGS HR sequencing was observed at three loci. The disagreement was observed for samples 6 (HLA-B in an individual of Mixed ancestry), 12 (HLA-B in a Black South African), and 19 (HLA-DRB1 in a Black South African). The disagreement observed for sample 6 is due to the sequences of B*07:05 and B*07:06 being identical across the sequenced regions (Robinson *et al*., 2013). The nucleotide change that distinguishes B*07:05 from B*07:06 is present in exon 5 of the HLA-B gene, which is outside of the region sequenced. Even though disagreement was observed for sample 6, when compared to the previous typing results, the correct genotype was still present in the ambiguity string for this study. A study by Paximadis and co-workers failed to identify HLA-B*07:05 in the Black South African populations, while B*07:06 was found to be relatively frequent in the Black South African population (Paximadis *et al*., 2011). The allele frequency database, on the contrary, indicates a higher frequency of B*07:05 in the Zulu population compared to B*07:06 (Gonzalez-Galarza *et al*., 2011).

Class I alleles for samples 11-20 were typed by Luminex, which is a low resolution genotyping technique, that types up to a two-digit level of resolution. The high number of B*15 alleles that exist requires additional kits to assign accurate genotypes. Therefore, all the B*15 alleles are typed to a four-digit level compared to the two-digit level for the other class I alleles. The disagreement observed for sample 12 could have occurred due to the limitations of the Luminex technology to accurately determine heterozygous alleles. HLA-B*15:01 is observed frequently in all populations, whereas B*15:03 is present at a relatively high frequency in Black South African individuals, while observed at a very low frequency in South African Caucasians (Paximadis *et al*., 2011; Gonzalez-Galarza *et al*., 2011). The technique that was used, could therefore, account for the incorrectly assigned homozygous genotype and the disagreement observed. The 454 NGS DRB1 locus

assignment for sample 19 also showed disagreement with the conventional typing results. DRB1*15:02 differs from DRB1*15:01 by a non-synonymous change in the DNA sequence that causes Valine (Val) to be substituted with Glycine (Gly) at position 86 of the peptide-binding groove (Marsh and Bodmer, 1993). Several studies have indicated the high frequency of the DRB1*15:01 allele in Caucasian individuals; this allele has also been observed in Black African individuals at a lower frequency (Paximadis *et al.,* 2011; Fernando *et al*., 2012). The DRB1*15:02 allele has previously been observed in Chinese individuals (Fernando *et al*., 2012). According to recent papers that investigated the diversity in the South African population, the DRB1*15:02 allele was not detected in Black South African individuals (Paximadis *et al.,* 2011; Gonzalez-Galarza *et al*., 2011). Therefore, this could indicate the presence of either a rare allele present in this particular African individual or the mistyping of the DRB1 locus by PCR-SSP.

It is well known that the allele frequencies of variants differ between individuals with different geographical ancestries. African individuals are known to have the most ancient genomes. A recently published study has indicated that Africans carry three times more low-frequency alleles compared with Europeans and Asians (The 1000 Genomes Project Consortium, 2012). The limited studies that have targeted the African population, and the extreme diversity of these individuals, affect the certainty with which a genotype is assigned. In many instances genotypes are assigned based on the frequency of the genotype in a given population. This could affect the assignment of rare alleles, especially in the African population, in which many have not yet been identified.

5.2.3. Unambiguous and Ambiguous Genotype Assignment

Several loci were assigned a unique genotype, while others produced a string of ambiguities. Although sequencing additional exons has resolved some ambiguity, it is not completely resolved, since ambiguous typing results were still produced for several loci. Table 17 illustrates the different loci and the unambiguous and ambiguous genotypes assigned at a four-digit level of resolution for the different MR and HR typing kits.

Table 17: Unambiguous and ambiguous typing results by 454 sequencing.

Locus	Samples with unambiguous genotypes (Four-digit level of resolution)		Ambiguous alleles (Four-digit level of resolution)	
	MR	HR	MR	HR
HLA-A	0	4	34	17
HLA-B	4	7	28	19
HLA-C	3	2	26	27
HLA-DPB1	5	5	24	24
HLA-DQA1	1	1	30	30
HLA-DQB1	0	3	35	25
HLA-DRB1	3	3	29	29
DRB3/4/5/6/7	3	3	25	25

The table compares the unambiguous and ambiguous typing results between the MR and HR kits. The additional primer sets incorporated through the HR kit resolved a great degree of ambiguity for most of the loci, except for HLA-C, where the ambiguity increased by sequencing additional exons. The sequencing of additional exons did not entirely resolve the issue of ambiguity, with HLA-C, -DQA1, and -DRB1 having the most ambiguous allelic assignments. The ambiguous typing results obtained from this study are inconsistent with the results from a paper published in 2011 by Holcomb and co-workers. The unambiguous assignment of genotypes is lower, while the ambiguity observed is greater compared with the results of Holcomb and co-workers (Holcomb *et al.,* 2011). The degree of ambiguity observed could be as a result of little knowledge on the polymorphisms present in Black African individuals at these loci. Ambiguous typing results occur as a result of phase ambiguity or partial sequences generated from sequencing. Sequence-based typing techniques only sequence certain exons of interest (exons 2, 3, and 4), which generate ambiguous typing results. This results from genetic polymorphisms situated outside the routinely sequenced regions. A study by Wang and co-workers has been able to resolve some ambiguity by sequencing several additional class I and II exons (Wang *et al*., 2012). Another way to resolve ambiguity according to Erlich and co-workers is to sequence the

entire gene of interest (Erlich *et al*., 2011). Although sequencing the entire gene would be lucrative, it is not yet cost effective. The reduction in sequencing costs might enable sequencing of the essential HLA genes as a possible solution to resolve ambiguity.

Differences were also observed between the results from the two different software programs that were used to analyse the data (Table 18). This is in all likelihood as a result of the different HLA database versions that were installed prior to analysis. The HLA database is regularly updated to include newly discovered alleles in order to assist in assigning accurate genotypes. The Conexio 454 ATF software used the HLA database version 3.2.0 (updated 15/10/2010) for the data analysis, whereas JSI SeqHLA 454 software used version 3.11.0 (updated 01/17/2013). Since 2010, more than 2 000 novel alleles have been discovered and submitted to the database (Figure 11, Chapter 2), thereby increasing the number of ambiguities due to sequencing only certain exons. Therefore, the software version that was installed for the Conexio software is outdated and not a true representation of the alleles currently present in the HLA database.

Table 18: Unambiguous and ambiguous typing results produced by different software programs.

Locus	Samples with unambiguous genotype (four-digit level of resolution)		Ambiguous alleles (four-digit level of resolution)	
	Conexio	JSI	Conexio	JSI
HLA-A	6	4	12	17
HLA-B	8	7	20	19
HLA-C	6	2	21	27
HLA-DPB1	6	5	20	24
HLA-DQA1	6	1	17	30
HLA-DQB1	7	3	21	25
HLA-DRB1	11	3	9	29
DRB3/4/5/6/7	14	3	5	25

5.2.4. Low to High Resolution HLA Typing

The resolution of HLA typing has fluctuated over the years with the emergence of various typing techniques. Conventional techniques such as serology and DNA based techniques are not reliable in distinguishing the ever increasing number of alleles being discovered. Sequence-specific primers and SSOP are commonly used in clinical laboratories for low and high resolution HLA typing, respectively. Sequence-specific oligonucleotide typing is able to determine alleles based on known variants, however, rare variants cannot be identified by this technique. This makes it challenging to assign HLA genotypes accurately. The class II alleles for all 20 samples were genotyped by PCR-SSP (described in Chapter 4) at the LTI. The SSP method has high specificity and is able to genotype at a high resolution. This method of HLA typing has a high success rate and has proven to be accurate. Techniques such as SSOP and SSP require updated probes and primers in order to detect the allelic diversity present at these loci. Alternative methods are used to complement these techniques to resolve ambiguous typing results. The high number of unknown variants present in African individuals make it challenging to assign genotypes accurately. It is, therefore, essential that a reliable high resolution typing technique is used for accurate determination of the HLA alleles and genotypes. True high resolution typing provides information up to the four-digit level of resolution. The 454 NGS technology (MR and HR kits) has enabled a higher resolution of HLA genotyping up to the eight-digit level. The highest level of resolution might be of significance for research purposes, which involve the discovery of possible vaccines for HIV. However, the four-digit level of resolution is sufficient for HLA genotyping. It is important to note that accurate four-digit typing is not necessarily obtainable by only sequencing exons 2 and 3. Therefore, additional exons might need to be sequenced to resolve ambiguous typing results at the allelic level. It is also possible to adjust the software to display the results to the four-digit level. The equipment used and the resolution of typing for these genes will depend on the laboratory needs. It is necessary for each institution to consider the value of higher resolution typing and whether it is in fact relevant and cost-effective.

5.2.5. Technical Limitations

The library preparation for this technique is time consuming, labour intensive, and expensive (Erlich *et al*., 2011). The cost of sequencing is also relatively high compared to conventional techniques. However, a study by Holcomb and co-workers has indicated that sequencing 20 samples by Sanger sequencing would be more time consuming and the costs for reagents would be more or less equivalent to reagents for NGS (Holcomb *et al*., 2011). An individual would require a certain level of expertise to perform the various procedures involved due to its complexity. DNA quality and quantity has been shown to play a significant role in obtaining accurate results and high quality DNA is not always obtainable. A quality assessment was performed for all amplicons before proceeding to emPCR, and although HLA-A and -C exons were of the lowest quantities prior to emPCR, genotypes were still accurately assigned for these loci. This provides a clear indication of the robust nature of the 454 NGS system. Manual editing was required for several samples, which necessitated knowledge on HLA analysis.

Bi-directional coverage is the process of sequencing both the forward and reverse strands of the DNA. For several samples, this technique was unable to accurately sequence the reverse strand for certain exons of interest and also had difficulty in accurately calling the bases at the ends of sequences. The 454 NGS technique relies on emitted light intensities to differentiate between the different nucleotides, which resulted in inaccurate base-calling at times. Another known limitation is the inability to detect the length of homopolymeric regions. The emitted light intensities make it difficult to determine the number of bases present for a specific nucleotide. Numerous short reads were also generated from the sequencing run, which were filtered out prior to the analysis by specific software tools. This study made use of two different kits that enabled the sequencing of additional exons. Some exonic regions failed to be sequenced for several samples. This was particularly observed for HLA-A, exon 4. Although the presence of exon 4 sequences would have assisted in resolving some ambiguous results, its absence did not affect the accuracy at which the genotypes were called for HLA-A. In some cases sequences of exon 4 were likewise not present in the HLA database. This is due to the fact that exon 4 has only recently been included in HLA

typing. In the case where a reference sequence of exon 4 was absent, genotypes were assigned based on the sequences across exons 2 and 3.

The technical limitations mentioned have been addressed and compensated for in the GS FLX Titanium system (Roche). A new set of HLA primers is also being developed, which will produce longer amplicons and additional exonic and intronic sequences of these genes. The current trials encountered by the current 454 NGS technique will facilitate in improving third- and fourth-generation sequencing in the future.

5.3. Concluding Remarks

HLA genotyping is performed on a regular basis for various applications. Hematopoietic stem cell transplantation is the most frequent application and has been effective in treating many fatal diseases worldwide. Donors and recipients require allele level matching at five HLA loci (10 alleles) for bone marrow transplantation, while UCB only requires matching at three HLA loci (six alleles). Mismatching between donors and recipients could lead to severe adverse reactions. Therefore, accurate genotyping of the HLA genes is crucial for the outcome of transplantation. The ever increasing number of HLA alleles, being discovered daily, creates a challenge to accurately genotype individuals for transplantation purposes. The strong LD, known to exist between genes within the MHC region, suggests that other relevant genes may also play a critical role. Haplotype matching has shown to improve the outcome of transplantation, and could therefore, also be due to the function of other important genes.

Next generation sequencing technologies have developed in the last decade and are continuously improving. The NGS (Roche/454 pyrosequencing) technique is quicker and more efficient than conventional techniques in determining the HLA genotypes of individuals. The results from this study indicate the importance of sequencing additional exons in order to resolve ambiguity for transplantation. Therefore, it is possible to conclude that the 454 NGS HR HLA typing kit has the potential as an alternative method to provide accurate genotyping in routine clinical and diagnostic laboratories. This technique also has the potential to be used as an HLA typing tool, in a public cord blood bank in South Africa, to

ultimately improve HLA matching between donors and recipients. However, the limitation of cost will possibly prevent this technique from being implemented in clinical laboratories in South Africa.

CONCLUDING REMARKS

The discovery of the first HLA molecule in 1952 led to further investigation of the function of these molecules and the pivotal role they play in immune responses. The HLA genes are the most diverse genes in humans. The true diversity of the HLA genes has only become apparent in the last decade with the emergence of DNA- and sequence-based typing methods. HLA typing has moved from cellular- to molecular-based techniques over the years and several typing applications are still performed routinely, with typing for transplantation being the most frequent. The extreme diversity in the Black South African population poses a challenge in obtaining matching donors for individuals in need of a transplant. The SABMR does not reflect the demographics of South Africa, since the majority of donors are Caucasian. Cord blood stem cell transplantation is an effective treatment alternative for many fatal diseases and requires less strict HLA matching criteria when compared to bone marrow. For that reason there is a continuous need for a public umbilical cord blood bank in order to cater for the diverse South African population. Through accurate HLA genotyping, it will be possible to increase the degree on HLA matching between individuals and most likely decrease the chances of developing GVHD and graft rejection.

An objective of the present study was to consider the racial/ethnic groups, linguistic groups, and the HLA diversity of the South African population and thereby determine how a public umbilical cord blood bank would need to be constituted (Chapter 3). South Africa is a multiracial, as well as a multicultural country, presently populated by more than 50 million individuals. It is comprised of White, Black, Indian/Asian, and Mixed ancestry individuals that together make up this diverse population. The Black population is the most diverse population group in South Africa and consists of four broad groupings, namely: Nguni (Includes Zulu, Xhosa, Ndebele, and Swazi), Sotho-Tswana (Southern, Northern, and Western Sotho population), Tsonga, and Venda. Individuals belonging to the Nguni group have shown increased linguistic and genetic similarities. Another cluster is formed by the Sotho-speaking individuals and it has also been indicated that the Venda group show close similarity to the Tsonga group. The White South African population is mainly of European descent and the majority speak Afrikaans and English. The Mixed ancestry South African population is comprised of individuals of mixed ancestral lineage and mainly speaks

Afrikaans. The Indian and Asian South Africans are largely English-speaking. The language that one speaks is often an indication of one's ethnicity, which could reflect cultural groupings as well as cross-cultural marriages. A great degree of intermixing occurs between Black South African individuals, which therefore, indicate that language is not a true representation of an individual's ethnicity. Certain population groups tend to reside in certain provinces, while others are evenly distributed across all of the nine provinces. Gauteng is the smallest of the nine provinces and also the most populous. Individuals from all the different linguistic groups currently reside in Gauteng, making it the most diverse province in South Africa.

Black African individuals are known to possess the most ancient and diverse genomes compared with other population groups across the globe. The vast diversity of these genes and population specific genetic variants decreases the probability of obtaining an HLA matching donor. It is, therefore, not recommended to constitute the cord blood bank on the basis of HLA diversity. Intermixing between individuals belonging to distinct linguistic groups is frequently observed. The extreme patterns of diversity that exist between the different linguistic groups render it difficult to determine the ethnicity of an individual based on the language spoken. Therefore, it is likewise not recommended to constitute the bank on the basis of language.

Through the knowledge obtained from the literature, it is recommended to constitute the bank based on social race or major ethnic groupings (Nguni, Sotho-Tswana, Tshivenda, and Xitsonga), since genetic similarities have been shown to exist between individuals belonging to the same major ethnic groups. Gauteng would be the most suitable province for the establishment of the first public umbilical cord blood bank in South Africa. The bank would have to be representative of the entire South African population. It has been estimated that a minimum number of 10 000 cord blood units would be needed to initiate the bank, of which 8 000 (80%) would have to be representative of Black South Africans, 900 (9%) of White South Africans, 900 (9%) of Mixed ancestry South Africans, and 200 (2%) of Indian/Asian South Africans.

A further objective included the validation of a 454 NGS HLA typing kit on South African samples obtained from the SABMR. This part of the study was performed to determine whether the NGS HLA typing kit would be a feasible option for HLA genotyping in a public cord blood bank or whether the information would be too detailed for the needs of a public bank (Chapter 5).

The HLA genotyping results for this part of the study indicated accurate HLA genotyping for the majority of the South African individuals and also for two individuals from other African countries. The disagreement observed with conventional typing results could have occurred due to the limitations of conventional techniques in detecting the diversity present in South African individuals. Sequencing of additional exons was able to resolve some ambiguities. Various challenges would still need to be addressed before NGS can be used as a suitable HLA typing method. This technique displays potential as an alternative HLA typing method for cord blood banks and clinical laboratories. However, the high cost of sequencing and labour intensity will most likely prevent this technique from being implemented in South Africa.

In conclusion, the main findings of the present study include:

- The diversity of the HLA genes in the South African population is extremely vast
- There is a great degree of intermixing between the different linguistic groups of Black South Africans
- A public umbilical cord blood bank in South Africa should be constituted on the basis of either race or major ethnic groups (Nguni, Sotho-Tswana, Tshivenda, and Xitsonga)
- The cord blood bank would need to be representative of the South African population
- The 454 NGS MR typing kit is not sufficient in accurately distinguishing between the different HLA alleles
- The 454 NGS HR HLA typing kit was able to accurately detect the majority of the HLA genotypes for selected African and South African individuals
- The sequencing of additional exons resolves some ambiguities
- The 454 NGS technique for HLA typing has the potential to be used as a diagnostic tool in a public bank as well as clinical laboratories in South Africa

Limitations of the present study:

- There is little information available on the HLA diversity within the South African population
- Previously conducted studies are not representative of the South African population
- There is limited high resolution HLA data available on the South African population
- HLA ambiguity remains a challenge due to sequencing of only the variable regions

Future studies implemented should focus on the following:

- Investigate the HLA diversity in South Africans by using DNA- or sequence-based methods on a large cohort
- Consider the possibility that various other genes could be found to be significant in matching between donors and recipients

REFERENCES

Abi-Rached, L. Jobin, M.J. Kulkarni, S. *et al.* (2011) The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science*, 334, p. 89-94.

Ahmadian, A. Ehn, M. Hober, S. (2006) Pyrosequencing: history, biochemistry and future. *Clinica Chimica Acta*, 363 (1-2), p. 83-94.

Almenar, L. Maeso, M.L. Martinez-Dolz, L. *et al*. (2005) Influence of HLA matching on survival in heart transplants. *Transplant Proc,* 37 (9), p. 4001-4005.

Tripathy, A.S. Shankarkumar, U. Chadha, M.S. *et al.* (2009) Association of HLA alleles with hepatitis C infection in Maharashtra, western India. *Indian J Med Res*, 130, p. 550-555.

Bacchetta, R. Bigler, M. Touraine, J.L. *et al*. (1994) High levels of interleukin 10 production in vivo are associated with tolerance in SCID patients transplanted with HLA mismatched hematopoietic stem cells. *J Exp Med,* 179, p. 493-502.

Baker, P.R. Baschal, E.E. Fain, P.R. *et al.* (2010) Haplotype analysis discriminates genetic risk for DR3-associated endocrine autoimmunity and helps define extreme risk for Addison's disease. *J Clin Endocrinol Metab*, 95 (10), E263-270.

Baker, P.R. Baschal, E.E. Fain, P.R. *et al.* (2011) Dominant suppression of addison's disease associated with HLA-B15. *J Clin Endocrinol Metab*, 96 (7), p. 2154-2162.

Barker, J.N. Krepski, T.P. DeFor, T.E. *et al.* (2002) Searching for unrelated donor for hematopoietic stem cells. Availability and speed of umbilical cord blood versus bone marrow. *Biol Blood Marrow Tr,* 8, p. 257-260.

Barker, J.N. Scaradavou, A. Stevens, C.E. (2010) Combined effect of total number of nucleated cell dose and HLA match on transplantation outcome in 1061 cord blood recipients with hematologic malignancies. *Blood,* 115 (9), p. 1843-1849.

Beatty, P.G. Boucher, K.M. Mori, M. *et al*. (2000) Probability of finding HLA-mismatched related or unrelated marrow of cord blood donors. *Hum Immunol*, 61 (8), p.834-840.

Beckman Coulter (2012) Agencourt AMPure XP – PCR purification. Available from: http://www.beckmancoulter.com [Accessed 08 October 2012].

Belov, K. Deakin, J.E. Papenfuss, A.T. *et al.* (2006) Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biol*, 4 (3), p. 317-328.

Bensussan, A. Gluckman, E. El-Marsafy, S. *et al.* (1994) BY55 monoclonal delineates within human cord blood and bone marrow lymphocytes distinct cell subsets mediating cytotoxic activity. *Proc Natl Acad Sci USA,* 91, p. 9136-9140.

Bentley, G. Higuchi, R. Hoglund, B. *et al*. (2009) High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens*, 74, p. 393-403.

Berthou, C. Legros-maîda, S. Soulié, A. *et al.* (1995) Cord blood T lymphocytes lack constitutive perforin expression in contrast to adult peripheral blood T-lymphocytes. *Blood*, 85, p. 1540-1546.

Blasczyk, R. (2003) DNA diagnostic sequencing - conception, application and automation. *J Lab Med*, 27 (9/10), p. 359-368.

Bodmer, W.F. (1972) Evolutionary significance of the HLA system. *Nature,* 237, p. 139-145.

Bosch, S. Pretorius, L. Fleisch, A. (2008) Experimental bootstrapping of morphological analysers for Nguni languages. *Nordie J Afr Stud,* 17 (2), p. 66-88.

Bozón, M.V. Delgado, J.C. Selvakumar, A. *et al.* (1997) Error rate for HLA-B antigen assignment by serology: implications for proficiency testing and utilization of DNA-based typing methods. *Tissue Antigens*, 50, p. 387-394.

Browning, M. McMichael, A. (eds.) (1996). *HLA and MHC: Genes, molecules and function.* Oxford: Bios Scientific Publishers Limited, p. 97-109.

Broxmeyer, H.E. Douglas, G.W. Hangoc, G. *et al.* (1989) Human umbilical cord blood as a potential source of transplantable hematopoietic stem/progenitor cells. *Proc Natl Acad Sci USA,* 86, p. 3828-3832.

Cairo, M.S. Wagner, J.E. (1997) Placental and/or umbilical cord blood: An alternative source of hematopoietic stem cells for transplantation. *Blood*, 90, p. 4665-4678.

Callaway, E. (2011) Ancient DNA reveals secrets of human history. *Nature*, 476, p. 136-137.

Calne, R.Y. Rolles, K. White, D.J. *et al.* (1979) Cyclosporine A initially as the only immunosuppressant in 34 recipients of cadaveric organs; 32 kidneys, 2 pancreases, and 2 livers. *Lancet*, 2 (8151), p. 1033-1036.

Campbell, N.A. Reece, J.B. (2005) *Biology*. 7th Ed. San Francisco: Pearson Benjamin Cummings. p. 909.

Cann, R.L. Stoneking, M. Wilson, A.C. (1987) Mitochondrial DNA and human evolution. *Nature*, 325 (6099), p. 31-36.

Cao, K. Moormann, A.M. Lyke, K.E. *et al.* (2004) Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens,* 63 (4), p. 293-325.

Carr, D.F. Chaponda, M. Jorgensen, A.L. *et al.* (2013) Association of human leukocyte antigen alleles and nevirapine hypersensitivity in a Malawian HIV-infected population. *Clin Infect Dis*, 56 (9), p. 1330-1339.

Carrington, M. Nelson, G.W. Martin, M.P. *et al*. (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science*, 283, p. 1748-1752.

Casals, F. Bertranpetit, J. (2012) Human genetic variation, shared and private. *Science*, 337, p. 39-40.

Chalmers, I.M.H. Janossy, G. Contreras, M. *et al*. (1998) Intracellular cytokine profile of cord and adult blood lymphocytes. *Blood*, 92, p. 11-18.

Chantarangsu, S. Mushiroda, T. Mahasirimongkol, S. *et al.* (2009) HLA-B*3505 allele is a strong predictor for nevirapine-induced skin adverse drug reactions in HIV-infected Thai patients. *Pharmacogenet Genomics,* 19 (2), p. 139-146.

Chen, Q.Y. Huang, W. She, J.X. *et al*. (1999) HLA-DRB1*08, DRB1*03/DRB3*0101, and DRB3*0202 are susceptibility genes for graves' disease in North American Caucasians, whereas DRB1*07 is protective. *J Clin Endocrinol Metab*, 84 (9), 3182-3186.

Choo, S.Y. (2007) The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J,* 2007, 48 (1), p. 11-23.

Coetzee, V. Greeff, J.M. Barrett, L. *et al.* (2009) Facial-based ethnic recognition: insights from two closely related but ethnically distinct groups. *S Afr J Sci*, 105 (11-12), p. 464-466.

da Silva, S.A. Mazini, P.S. Reis, P.G. *et al*. (2009) HLA-DR and HLA-DQ alleles in patients from the south of Brazil: markers for leprosy susceptibility and resistance. *BMC Infect Dis*, 9, p. 134-140.

Dausset, J. (1958) Iso-leuco-anticorps. *Acta Haematol*, 20, p. 156–166.

Doherty, P.C. Zinkernagel, R.M. (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 complex. *Nature*, 256, p. 50-52.

Donaldson, P. Underhill, J. Doherty, D. *et al*. (1993) Influence of human leukocyte antigen matching in liver allograft survival and rejection: "the dualistic effect". *Hepatology,* 17 (6), p. 1008-1015.

du Toit, E.D. MacGregor, K.J. Taljaard, D.G. *et al.* (1988) HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African Negroes, Cape Coloureds and South African Caucasoid. *Tissue Antigens,* 31, p. 109-125.

du Toit, E.D. (1988) 'HLA polymorphisms in Southern African population' in WR Mayr (ed.), *Adv Foren H 2.* Springer-Verlag Berlin Heidelberg, p. 42.

du Toit, E.D. Emmanuel, J.C. West, G. *et al.* (1990) HLA frequencies in Black Zimbabweans. *Tissue Antigens*, 36, p. 122-124.

du Toit, E.D. Taljaard, D.G. Marshall, J. *et al.* (1990) The HLA genetic constitution of the Bushmen (San). *Hum Immunol,* 28, p. 406-415.

Eapen, M. Rubinstein, P. Zhang, M.J. *et al*. (2007) Comparison of outcomes after transplantation of unrelated donor umbilical cord blood and bone marrow in children with acute leukemia. *Lancet*, 369 (9577), p. 1947-1954.

Eapen, M. Klein, J.P. Sanz, G.F. *et al*. (2011) Effect of donor-recipient HLA matching at HLA A, B, C, and DRB1 on outcomes after umbilical-cord blood transplantation for leukaemia and myelodysplastic syndrome: a retrospective analysis. *Lancet Oncol*, 12, p. 1214-1221.

Erlich, R.L. Jia, X. Anderson, S. *et al*. (2011) Next-generation sequencing for HLA typing of class I loci. *BMC Genomics*, 12, p. 42-54.

Eskdale, J. Gallagher, G. Verweij, C.L. *et al*. (1998) Interleukin 10 secretion in relation to human IL-10 locus haplotypes. *Proc Natl Acad Sci USA,* 95, p. 9465-9470.

Fackenthal, J.D. Olapade, O.I. (2007) Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer*, 7, p. 937-948.

Falk, K. Rotzscehke, O. Stevanovic, S. *et al*. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature,* 351, p. 290-296.

Fanning, L.J. Levis, J. Kenny-Walsh, E. *et al*. (2000) Viral clearance in hepatitis C (1b) infection: relationship with human leukocyte antigen class II in a homogeneous population. *Hepatology*, 31 (6), p. 1334-1337.

Farouk, H.M. Gabal, M.M.A. Ali, A.M. *et al*. (2011) Contribution of Human Leucocyte Antigen Hla-dr in Systemic Lupus Erythematosus: effect on disease susceptibility and outcome. *Res J Med Med Sci*, 6 (1), p. 43-50.

Fernando, M.M.A. Freudenberg, J. Lee, A. *et al*. (2012) Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G. *Ann Rheum Dis*, 71, p. 777-784.

Ferrer-Admetlla, A. Bosch, E. Sikora, M. *et al.* (2008) Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol,* 181 (2), p. 1315-1322.

Flomenberg, N. Baxter-Lowe, L.A. Confer, D. *et al.* (2004) Impact of HLA class I and class II high resolution matching on outcomes of unrelated donor bone marrow transplantation. *Blood*, 104, p. 1923-1930.

Fries, J.F. Wolfe, F. Apple, R. *et al*. (2002) HLA–DRB1 genotype associations in 793 white patients from a rheumatoid arthritis inception cohort: frequency, severity, and treatment bias. *Arthritis Rheum*, 46 (9), p. 2320-2329.

Gao, X. Bashirova, A. Iversen A.K.N. *et al.* (2005) AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nat Med*, 11 (12), p. 1290-1292.

Gatanaga, H. Yazaki, H. Tanuma, J. *et al*. (2007) HLA-Cw8 primarily associated with hypersensitivity to nevirapine. *AIDS*, 21 (2), p. 264-265.

Gibbons, A. (2011) Who were the Denisovans? *Science,* 333, p. 1084-1087.

Gluckman, E.G. Broxmeyer, H.E. Auerbach, A. *et al*. (1989) Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical-cord blood from an HLA-identical sibling. *N Engl J Med,* 321 (17), p. 1174-1178.

Gluckman, E.G. Rocha, V. Chastang, C. (1997) Use of cord blood cells for banking and transplant. *Oncologist*, 2 (5), p. 340-343.

Gombos, Z. Hermann, R. Kiviniemi, M. *et al.* (2007) Analysis of extended human leukocyte antigen haplotype association with addison's disease in three populations. *Eur J Endocrinol*, 157, p. 757-761.

Gonzalez-Galarza, F.F. Christmas, S. Middleton, D. *et al.* (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res,* 39, D913-D919.

Gonzalez-Roces, S. Alvarez, M.V. Gonzalez, S. *et al*. (1997) HLA-B27 polymorphism and worldwide susceptibility to ankylosing spondylitis. *Tissue Antigens*, 49, p. 116-123.

Gorer, P. (1937) The genetic and antigenic basis of tumour transplantation. *J Pathol Bacteriol,* 44, p. 691-697.

Graham, R.R. Ortmann, W.A. Langefeld, C.D. *et al.* (2002) Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet*, 71, p. 543-553.

Harris, R.A. Sugimoto, K. Kaplan, D.E. *et al*. (2008) Human leukocyte antigen class II associations with Hepatitis C virus clearance and virus-specific CD4 T cell response among Caucasians and African Americans. *Hepatology*, 48 (1), p. 70-79.

Hill, A.V.S. Allsopp, C.E.M. Kwiatkowski, D. *et al.* (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature,* 352, p. 595-600.

Holcomb, C.L. Hoglund, B. Anderson, M.W. *et al.* (2011) A multi-site study using high-resolution HLA genotyping by next generation sequencing. *Tissue Antigens,* 77, p. 206-217.

Hughes, A.L. Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature,* 335 (6186), p. 167-170.

Hughes, A.L. Nei, M. (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA*, 86 (3), p. 958-962.

Hughes, L.B. Morrison, D. Kelley, J.M. *et al*. (2008) The HLA–DRB1 shared epitope is associated with susceptibility to rheumatoid arthritis in African Americans through European genetic admixture. *Arthritis Rheum*, 58 (2), p. 349-358.

International Multiple Sclerosis Genetics Consortium and Wellcome Trust Case Control Consortium 2 (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, 476 (7359), 214-219.

Jin, P. Wang, E. (2003) Polymorphism in clinical immunology|From HLA typing to immunogenetic profiling. *J Transl Med*, 1, p. 1-11.

John Hawks Weblog. Paleoanthropology, genetics and evolution, http://johnhawks.net/weblog [Accessed 18 July 2012].

Jorde, L.B. Watkins, W.S. Bamshad, M.J. *et al*. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and y-chromosome data. *Am J Hum Genet*, 66, p. 979-988.

Karell, K. Louka, A.S. Moodie, S.J. *et al*. (2003) HLA types in celiac disease patients not carrying the *DQA1*05-DQB1*02* (DQ2) heterodimer: Results from the European genetics cluster on celiac disease. *Hum Immunol*, 64, p. 469-477.

Kasahara, M. Vazquez, M. Sato, K. *et al.* (1992) Evolution of the major histocompatibility complex: isolation of class II A cDNA clones from the cartilaginous fish. *Proc Natl Acad Sci*, 89, p. 6688-6692.

Kelley, J. Walter, L. Trowsdale, J. (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, 56, p. 683-695.

Kettaneh, A. Seng, L. Tiev, K.P. *et al*. (2006) Human leukocyte antigens and susceptibility to tuberculosis: a meta-analysis of case-control studies. *Int J Tuberc Lung Dis*, 10 (7), p. 717-725.

Kindmark, A. Jawaid, A. Harbron, C.G. *et al*. (2008) Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J*, 8, p. 186-195.

Kindt T.J. Goldsby, R.A. Osborne, B.A. (2006) *Kuby's textbook of Immunology*, 6th Ed. New York: WH Freeman and Company. p. 210-217.

Klein, J. Sato, A. (2000) The HLA system part 1. Erratum in: *N Engl J Med*, 343 (11), p. 702-709.

Kline, R.M. (2001) Whose blood is it anyway? *Sci Am,* 284, p. 42-49.

Kurtzberg, J. Laughlin, M. Graham, M.L. *et al.* (1996) Placental blood as a source of hematopoietic stem cells for transplantation into unrelated recipients. *New Engl J Med,* 335, p. 157-166.

Lane, A.B. Soodyall, H. Arndt, S. *et al*. (2001) Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. *Am J Phys Anthropol,* 119 (2), p. 175-185.

Lang, H.L.E. Jacobsen, H. Ikemizu, S. *et al*. (2002) A functional and structural basis for TCR cross-reactivity in multiple sclerosis. *Nat Immunol*, 3 (10), p. 940-943.

Lank, S.M. Golbach, B.A. Creager, H.M. *et al.* (2012) Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed cDNA amplicon pyrosequencing. *BMC Genomics*, 13, p. 387-399.

Lee, S.J. Klein, J. Haagenson, M. *et al.* (2007) High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood,* 110, p. 4576-4583.

Loetscher, P. Uguccioni, M. Bordoli, L. *et al*. (1998) CCR5 is characteristic of Th1 lymphocytes. *Nature,* 391, p. 344-345.

Mallal, S. Nolan, D. Witt, C. *et al.* (2002) Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet*, 359 (9308), p. 727-732.

Marsh, S.G.E. Bodmer, J.G. (1993) HLA class II nucleotide sequences. *Immunogenetics*, 37 (2), p. 79-94.

Marsh, S.G.E. Albert, E.D. Bodmer, W.F. *et al*. (2010) Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, 75, p. 291-455.

McCormack, M. Alfirevic, A. Bourgeois, S. *et al*. (2011) HLA-A*3101 and carbamezapine-induced hypersensitivity reactions in Europeans. *N Engl J Med*, 364, p. 1134-1143.

Mehta, T.Y. Prajapati, L.M. Mittal, B. *et al*. (2009) Association of HLA-B*1502 allele and carbamazepine-induced Stevens-Johnson syndrome among Indians. *Indian J Dermatol Venereol Leprol*, 75 (6), p. 579-582.

Mehra, N.K. Kaur, G. (2003) MHC-based vaccination approaches: progress and perspectives. *Expert Rev Mol Med*, 5, p. 1-17.

Merrill, J.P. Murray, J.E. Harrison, J.H. *et al*. (1956) Successful homotransplantation of the human kidney between identical twins. *J Am Med Assoc,* 160, p. 277-282.

Migliaccio, A.R. Adamson, J.W. Stevens, C.E. *et al.* (2000) Cell dose and speed of engraftment in placental/umbilical cord blood transplantation: graft progenitor cell content is a better predictor than nucleated cell quantity. *Blood*, 96, p. 2717-2722.

Mira, M.T. (2006) Genetic host resistance and susceptibility to leprosy. *Microbes Infect*, 8, p. 1124-1131.

Mungall, A.J. Palmer, S.A. Sims, S.K. *et al*. (2003) The DNA sequence and analysis of human chromosome 6. *Nature*, 425, p. 889-899.

Nejentsev, S. Howson, J.M.M. Walker, N.M. *et al*. (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A*. Nature*, 450 (7171), p. 887-892.

Okoye, R.C. Williams, E. Alonso, A. *et al.* (1985) HLA polymorphisms in Nigerians. *Tissue Antigens*, 25, p. 142-155.

Pan, C.F. Wu, C.J. Chen, H.H. *et al.* (2009) Molecular analysis of HLA-DRB1 allelic associations with systemic lupus erythematous and lupus nephritis in Taiwan. *Lupus*, 18, p. 698-704.

Paximadis, M. Mathebula, T.Y. Gentle, N.L. *et al*. (2011) Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. *Hum Immunol*, 73 (1), p. 80-92.

Payne, R. Rolfs, M.R. (1958) Fetomaternal leukocyte incompatibility. *J Clin Invest,* 37, p. 1756-1762.

Petersdorf, E.W. Hansen, J.A. Martin, P.J. *et al.* (2001) Major-histocompatibility-complex class I alleles and antigens in hematopoietic-cell transplantation. *N Engl J Med,* 345, p. 1794-1800.

Petersdorf, E.W. Anasetti, C. Martin, P.J. *et al.* (2004) Limits of HLA mismatching in unrelated hematopoietic cell transplantation. *Blood*, 104, p. 2976-2980.

Porth, C.M. (2010). *Essentials of Pathophysiology.* 3rd Ed. Philadelphia: Wolters Kluwer/Lippincott Williams and Wilkins. p. 354.

Robinson, J. Halliwell, JA. Marsh, GE. (2013) 'Exon identity and ambiguous typing combinations'. *Anthony Nolan Research Institute*.

Robinson, J. Halliwell, J.A. McWilliam, H. *et al*. (2013) The IMGT/HLA database. *Nucleic Acids Res,* 41 (Suppl 1), D1222-1227.

Rocha, V. Wagner, J.E. Sobocinski, K.A. *et al*. (2000) Graft-versus-host disease in children who have received a cord-blood or bone marrow transplant from an HLA-identical sibling. *N Engl J Med*, 342 (25), p. 1846-1854.

Rocha, V. Labopin, M. Sanz, G. *et al.* (2004) Transplants of umbilical cord blood or bone marrow from unrelated donors in adults with acute leukemia. *New Engl J Med,* 351 (22), p. 2276-2285.

Roche Products (Pty) Ltd, Diagnostic Division. GS GType Assays: HLA. Available from: http://www.454.com [Accessed 04 February 2013].

Russel, P.J. (2006) *iGenetics: A molecular approach.* 2nd Ed. San Francisco: Pearson Benjamin Cummings. p. 192-194.

Sanchez-Mazas, A. Djoulah, S. Busson, M. *et al.* (2000) A linkage disequilibrium map of the MHC region based on the analysis of 14 loci haplotypes in 50 French families. *Euro J Hum Genet,* 8, p. 33-41.

Sanchez-Mazas, A. Lemaitre, J.F. Currat, M. (2012) Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. *Phil Trans R Soc B*, 367, p. 830-839.

Shankarkumar, U. Ghosh, K. Badakere, S.S. *et al*. (2003) HLA-DRB1*03 and DQB1*0302 associations in a subset of patients severely affected with systemic lupus erythematosus from western India. *Ann Rheum Dis*, 62, p. 92-93.

Snell, G.D. (1986) Some recollections of Peter Gorer and his work on this fiftieth anniversary of his discovery of H-2. *Immunogenetics,* 24, p. 339-340.

Somkrua, R. Eickman, E.E. Saokaew, S. *et al*. (2011) Association of HLA-B*5801 allele and allopurinolinduced stevens johnson syndrome and toxic epidermal necrolysis: a systematic review and meta-analysis. *BMC Med Genet*, 12, p. 118-127.

South African Bone Marrow Registry (2008) Chairman's Report. http://www.sabmr.co.za/chairmanreport.htm [Accessed 13 February 2013].

Statistics South Africa (2001) *Census in Brief*. Statistical Release 03-02-03, Pretoria. Available from: http://www.statssa.gov.za. [Accessed 20 September 2012].

Statistics South Africa (2007) *Community Survey*. Statistical Release 03-01-27. Available from: http://www.statssa.gov.za [Accessed 11 November 2012].

Statistics South Africa (2011) *Mid-year population estimates.* Statistical Release P0302. Available from: http://www.statssa.gov.za [Accessed 05 December 2012].

Takahata, N. Nei, M. (1990) Allele genealogy under over-dominant and frequency dependant selection and polymorphism of major histocompatibility complex loci. *Genetics,* 124, p. 967-978.

The 1000 Genome Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, p. 56-65.

The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature,* 437 (7063), p. 1299-1320.

The International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409, p. 928-933.

The MHC Sequencing Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. *Nature,* 401, p. 921-923.

The Nobel Prize in Physiology or Medicine (1980) Nobelprize.org. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1980/ [Accessed 18 February 2013].

Thio, C.L. Thomas, D.L. Goedert, J.J. *et al.* (2001) Racial differences in HLA class II associations with hepatitis C virus outcomes. *J Infect Dis*, 184 (1), p. 16-21.

Thornley, I. Eapen, M. Sung, L. *et al.* (2009) Private cord blood banking: Experiences and views of pediatric hematopoietic cell transplantation physicians. *Pediatrics*, 123 (3), p. 1011-1017.

Thursz, M.R. Thomas, H.C. Greenwood, B.M. *et al.* (1997) Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat Genet*, 17, p. 11-12.

Tishkoff, S.A. Williams, S.M. (2002) Genetic analysis of African populations: Human evolution and complex disease. *Nat Rev Genet*, 3 (8), p. 611-621.

Townsend, C.M. *et al*. (2007) *Sabiston textbook of surgery*: *the biological basis of modern surgical practice*. 18th Ed. Philadelphia: Elsevier Saunders. p. 660.

van Rood, J.J. Eernisse, J.G. van Leeuwen, A. (1958) Leucocyte antibodies in sera from pregnant women. *Nature,* 181, p. 1735-1736.

Vijaya Lakshmi, V. Rakh, S.S. Anu, R.B. *et al.* (2006) Role of HLA-B51 and HLA-B52 in susceptibility to pulmonary tuberculosis. *Infect Genet Evol,* 6 (6), p. 436-439.

Wagner, J.E. Rosenthal, J. Sweetman, R. *et al.* (1996) Successful transplantation of HLA matched and HLA-mismatched umbilical cord blood from unrelated donors: analysis of engraftment and acute graft-versus-host disease. *Blood,* 88, p. 795-802.

Wagner, J.E. Barker, J.N. DeFor, T.E. *et al.* (2002) Transplantation of unrelated donor umbilical cord blood in 102 patients with malignant and nonmalignant diseases: influence of CD34 cell dose and HLA disparity on treatment-related mortality and survival. *Blood,* 100, p. 1611-1618.

Wainscoat, J.S. Hill, A.V.S. Boyce, A.L. *et al*. (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature*, 319, p. 491-493.

Wang, C. Krishnakumar, S. Wilhelmy, J. *et al.* (2012) High-throughput, high-fidelity HLA genotyping with deep sequencing. *P Natl Acad Sci USA*, 109 (22), p. 2-7.

Yan, Z.H. Fan, Y. Wang, X.H. *et al.* (2012) Relationship between HLA-DR gene polymorphisms and outcomes of hepatitis B viral infections: a meta-analysis. *World J Gastroenterol*, 18 (24), p. 3119-3128.

Zhang, M. Davies, S.M. Camitta, B.M. *et al*. (2012) Comparison of outcomes after HLA matched sibling and unrelated donor transplantation for children with high-risk acute lymphoblastic leukemia. *Biol Blood Marrow Tr*, 18, p. 1204-1210.

Zinkernagel, R.M. Doherty, P.C. (1974) Restriction of in-vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature,* 248, p.701-702.

APPENDIX A

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

* **FWA** 00002567, Approved dd 22 May 2002 and Expires 20 Oct 2016.
* **IRB** 0000 2235 IORG0001762 Approved dd 13/04/2011 and Expires 13/04/2014.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences Research Ethics Committee
Fakulteit Gesondheidswetenskappe Navorsingsetiekkomitee

DATE: 24/11/2011

NUMBER	**219/2011**
TITLE OF THE PROTOCOL	Factors determining the composition of a public cord blood stem cell bank including HLA diversity
PRINCIPAL INVESTIGATOR	**Student Name & Surname:** Miss Juanita Mellet **Dept:** Department of Immunology, University of Pretoria. **Cell:** 0795236401 **E-Mail:** juanitamellet@yahoo.co.uk
SUB INVESTIGATOR	None
STUDY COORDINATOR	Prof Michael S Pepper
SUPERVISOR (ONLY STUDENTS)	**Name & Surname:** Prof Michael S Pepper **E-Mail:** michael.pepper@up.ac.za
STUDY DEGREE	MSc
SPONSOR COMPANY	None
MEETING DATE	23/11/2011

The Protocol was approved on 23/11/2011 by a properly constituted meeting of the Ethics Committee subject to the following conditions:

1. The approval is valid for 2 years period [till the end of December 2013], and
2. The approval is conditional on the receipt of 6 monthly written Progress Reports, and
3. The approval is conditional on the research being conducted as stipulated by the details of the documents submitted to and approved by the Committee. In the event that a need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

Members of the Research Ethics Committee:

Prof M J Bester	(female)BSc (Chemistry and Biochemistry); BSc (Hons)(Biochemistry); MSc(Biochemistry); PhD (Medical Biochemistry)
Prof R Delport	(female)BA et Scien, B Curationis (Hons) (Intensive care Nursing), M Sc (Physiology), PhD (Medicine), M Ed Computer Assisted Education
Prof JA Ker	MBChB; MMed(Int); MD – Vice-Dean (ex officio)
Dr NK Likibi	MBB HM – Representing Gauteng Department of Health) MPH
Dr MP Mathebula	(female)Deputy CEO: Steve Biko Academic Hospital; MBCHB, PDM, HM
Prof A Nienaber	(female) BA(Hons)(Wits); LLB; LLM; LLD(UP); PhD; Dipl.Datametrics(UNISA) – Legal advisor
Mrs MC Nzeku	(female) BSc(NUL); MSc(Biochem)(UCL, UK) – Community representative
Prof L M Ntlhe	MbChB (Natal) FCS (SA)
Snr Sr J Phatoli	(female) BCur(Eet.A); BTec(Oncology Nursing Science) – Nursing representative
Dr R Reynders	MBChB (Prêt), FCPaed (CMSA) MRCPCH (Lon) Cert Med. Onc (CMSA)
Dr T Rossouw	(female) MBChB (cum laude); M.Phil (Applied Ethics) (cum laude), MPH (Biostatistics and Epidemiology (cum laude), D.Phil

2011/11/25MS: dd 2011/11/25: C:\Documents and Settings\user\My Documents\Protokolle\Grade briewe\Letters 2011\219.2011.DOC

Dr L Schoeman	(female) B.Pharm, BA(Hons)(Psych), PhD – Chairperson: Subcommittee for students' research
Mr Y Sikweyiya	MPH; SARETI Fellowship in Research Ethics; SARETI ERCTP; BSc(Health Promotion)Postgraduate Dip (Health Promotion) – Community representative
Dr R Sommers	(female) MBChB; MMed(Int); MPharmMed – **Deputy Chairperson**
Prof TJP Swart	BChD, MSc (Odont), MChD (Oral Path), PGCHE – School of Dentistry representative
Prof C W van Staden	MBChB; MMed (Psych); MD; FCPsych; FTCL; UPLM - **Chairperson**

DR R SOMMERS; MBChB; MMed(Int); MPharmMed.
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

♦Tel:012-3541330 ♦Fax:012-3541367 / 0866515924 ♦E-Mail: manda@med.up.ac.za
♦Web: //www.healthethics-up.co.za ♦H W Snyman Bld (South) Level 2-34 ♦ Private Bag x 323, Arcadia, Pta, S.A., 0007

2011/11/25MS: dd 2011/11/25: C:\Documents and Settings\user\My Documents\Protokolle\Grade briewe\Letters 2011\219.2011.DOC

APPENDIX B

UNIVERSITY OF CAPE TOWN

Faculty of Health Sciences
Human Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Ms S Ariefdien - Tel: [021]4066492 • Fax: [021]4066411
email: sumayah.ariefdien@uct.ac.za

03 November 2011

HREC REF: 523/2011

Prof CL Gray,
Immunology Department
IIDMM

Dear Prof Gray,

PROJECT TITLE: FACTORS DETERMINING THE COMPOSITION OF PUBLIC CORD BLOOD STEM CELL BANK INCLUDING RACE, ETHNICITY AND HLA DIVERSITY

Thank you for submitting your new study to the Faculty of Health Sciences Human Research Ethics Committee

It is a pleasure to inform you that the Ethics Committee has formally approved the above-mentioned study.

Approval is granted until 28 November 2012

Please submit an annual progress report (FHS016) if the research continues beyond the expiry date. Please submit a brief summary of findings if you complete the study within the approval period so that we can close our file (FHS010).

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the HREC. REF in all your correspondence.

Yours sincerely

PROFESSOR MARC BLOCKMAN
pp **CHAIRPERSON, FHS HUMAN RESEARCH ETHICS**
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.

The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

APPENDIX C

Solutions:

AE Buffer

10 mM Tris-Cl

0.5 mM EDTA

pH 9.0

1% Agarose gel - Research Tray

0.6 g agarose

60 ml 1x TBE

3 µl Gel Red

Dissolve the Agarose and the TBE in the microwave until a clear solution is formed. Add Gel Red to the clear solution and mix. Pour cool liquid into the gel tray with combs to form the wells. Allow to cool for 30-45 min.

2% Agarose gel - Diagnostic Tray

4 g agarose

200 ml 1x TBE

20 µl Gel Red

Dissolve the Agarose and the TBE in the microwave until a clear solution is formed. Add Gel Red to the clear solution and mix. Pour cool liquid into the gel tray with combs to form the wells. Allow to cool for 30-45 min.

70% Ethanol

70 ml Ethanol (100%)

30 ml Molecular Grade Water

80% Glycerol (w/v)

Pre-weigh a 50 ml tube. Add 30 ml of 100% glycerol. Determine the weight of the glycerol by subtracting the weight of the tube from the total weight. Calculate the total volume of solution (in ml) that will make 80% w/v by multiplying the weight of glycerol by 1.25. Determine the amount of molecular biology grade water to add to the glycerol by subtracting the weight of glycerol (in grams) from the total volume of solution needed (in ml). Add the amount of molecular biology grade water calculated to the glycerol and mix thoroughly.

1x Tris EDTA (TE) Buffer

10mM Tris-HCl (pH 7.6-8.0)

1 mM EDTA

Make up to 100 ml by adding ddH_2O and store at room temperature