Human Leukocyte Antigen (HLA) and Killer Immunoglobulin-Like Receptors (KIR) in HIV-2 Infection

Thesis

Human Leukocyte Antigen (HLA) and Killer Immunoglobulin-like Receptors (KIR) in HIV-2 Infection

Louis Marie Yindom

Medical Research Council Laboratories

Banjul, The Gambia

A thesis submitted for the degree of Doctor of Philosophy to the Open University in the United Kingdom.

Supervisors: **Prof Robert Walton**
Centre for Health Sciences, Barts and the London Medical School
Abernethy Building, London E1 2AT

Prof Sarah Rowland-Jones
MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine

Dr Mary Carrington
Director, Basic Sciences, NCI, SAIC-Frederick
Miller Drive, Frederick, MD 21702

(January 2010)

Date of submission: 21 January 2010
Date of award: 21 April 2010

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

DEDICATION

To God, my parents, brothers and sisters and the Yindoms

Human Leukocyte Antigen (HLA) and Killer Immunoglobulin-like Receptors (KIR) in HIV-2 Infection

Louis Marie Yindom, Open University

D.Phil. Thesis, January 2010

ABSTRACT

In West Africa, the Acquired Immunodeficiency Syndrome (AIDS) is caused by both types of Human Immunodeficiency Virus HIV-1 and HIV-2 and the majority of people infected with HIV-2 remain healthy for over 15 years.

The major goal of this work was to determine and describe genetic variants associated with phenotypic changes at the population level and study their role in pathogenesis and immune response. Studies presented in this thesis were carried out in two well established HIV-2 cohorts in Guinea Bissau and Gambia. Using recent tools in molecular medicine i.e. sequence-based techniques (SBT) (Chapter 2), we determined and comprehensively described variations at the two most polymorphic regions of the human genome: *HLA* and *KIR* gene complexes located on chromosomes 6 and 19, respectively, in these cohorts (Chapters 3 and 6). The data showed high heterogeneity in allele and genotype frequencies between the studied populations. Furthermore, we related the presence of gene variants to HIV-2 antibody status and study their effect on markers of disease progression, notably $CD4^+$ T cell count and viral load (Chapter 4). Here we showed for the first time that *HLA-B*1503* associates with poor prognosis after HIV-2 infection and *HLA-B*0801* with susceptibility to HIV-2 while the compound genotypes *KIR2DL2+HLA-C1* and *KIR2DS2+HLA-C1* protect against HIV-2 acquisition in the Manjako ethnic group. None of the HLA class I alleles/haplotypes and *KIR* gene profiles was found to influence HIV-2 infection in the second cohort, which was of mixed ethnic origin (Chapter 5).

In general, we observed that alleles previously shown to be associated with HIV-1 disease in western populations showed no effect in HIV-2 infection. This emphasizes the need to study *HLA* and *HLA/KIR* combinations in different populations in order to better inform subsequent vaccine design and evaluation in target populations.

ACKNOWLEDGEMENTS

This project was funded by the European and Developing Countries Clinical Trials Partnership (EDCTP) through a "Training Awards PhD Scholarships" programme awarded to me in 2006. I remain grateful to EDCTP executive management for making this opportunity available to young graduates in developing countries.

I wish to thanks the management of MRC (UK) Fajara Laboratories in The Gambia for their full support and for providing me with a conducive environment to carry out my research.

I am most grateful to my supervisory team for their constant support, encouragement, and inspiration. My Director of Studies (Prof. Robert Walton) operated an "open door" policy whereby his office and phone lines were open to me at all times. Professor Sarah Rowland-Jones was always available for guidance and discussion. She has been a model of strong courage and determination and a mother to me. Dr. Mary Carrington was the back bone of this project. My first visit in her laboratory in 2005 changed my life for good. She incorporated the spirit of hard work in my soul and trained me on most of the typing techniques and data analysis skills used in this thesis. She ensured that data generated were of the highest possible standard and that the analyses were done accordingly. Many thanks again to Mary and her wonderful team of scientists.

Many thanks to Prof Hilton Whittle for his special interest in my work and academic progress; Hilton was always available for advice and to decrypt some of my unknowns. He made me to believe that nothing is impossible under the sun and that there are no such things like "the sky is your limit" provided you are hard

working and inspirational. I thank Dr Assan Jaye – the advisor for this project, Dr Matthew Cotten – the third party monitor, and Dr David Conway – my line manager, for their continuous encouragement. Thank you to all my co-workers at the MRC Fajara and Caio fields station, especially past and present members of the Viral Disease Programme and Genetics Department for their technical and moral support.

Many thanks to all study participants in Caio, Guinea Bissau and Fajara, The Gambia for consenting to take part in this study. I am very grateful to all those who helped in any way in the realisation of this work especially, Alessandra Leligdowicz, Peter Aka, Carla Van Tienen, Tim Vincent, Akum Aveika, Archibald Worwui, Mathurin Diatta, Mamadi Njie, Bakary Sanneh, Bankole Ahadzie, and many others.

Special thanks to Mrs Fozeu Kwekam Yindom, Georges Junior Yindom Kenko, Claire Yindom Sussussi, Nathan Gaddiel Yindom Fozeu for allowing me to spend most of my time in the laboratory till very late at night.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ADCC	-	antibody-dependent cellular cytotoxicity
AF	-	Allele frequency
AIDS	-	Acquired Immunodeficiency Syndrome
APC	-	Antigen presenting cell
β_2m	-	Beta-2 microglobulin
CD	-	Cluster of differentiation
CLIP	-	Class II associated invariant chain peptide
CMV	-	Cytomegalovirus
CSW	-	Commercial sex workers
CTL or T_c	-	Cytotoxic T cells
DC	-	Dendritic cell
DM	-	HLA class II isoform M
DMA	-	Gene encoding the alpha chain of DM
DMB	-	Gene encoding the beta chain of DM
DNA	-	Deoxyribonucleic acid
dNTP	-	Deoxyribonucleotide triphosphate
DO	-	HLA class II isoform O
DOA	-	Gene encoding the alpha chain of DO
DOB	-	Gene encoding the beta chain of DO
DP	-	HLA class II isoform P
DPA	-	Gene encoding the alpha chain of DP
DPB	-	Gene encoding the beta chain of DP
DQ	-	HLA class II isoform Q
DQA	-	Gene encoding the alpha chain of DQ
DQB	-	Gene encoding the beta chain of DQ
DQB1	-	Gene encoding the beta chain number 1 of DQ
DR	-	HLA class II isoform R
DRA	-	Gene encoding the alpha chain of DR
DRB1	-	Gene encoding the beta chain number 1 of DR
DRB4	-	Gene encoding the beta chain number 4 of DR
DRB9	-	Gene encoding the beta chain number 9 of DR
EDTA	-	Ethylenediaminetetraacetic acid
ER	-	Endoplasmic reticulum
GF	-	Genotype frequency
gp	-	Group of proteins
GUM	-	Genito Urinary Medicine
H-2	-	Murine histocompatibility antigens 2
HBV	-	hepatitis B virus
HGNC	-	HUGO Genome Nomenclature Committee
HIV-1	-	Human immunodeficiency virus type 1
HIV-2	-	Human immunodeficiency virus type 2
HLA	-	Human leukocyte antigens
HPLC	-	High performance liquid chromatography
H-W	-	Hardy-Weinberg equilibrium
IDDM	-	Insulin-dependent diabetes mellitus
Ii	-	Invariant chain
IPD	-	Immuno Polymorphism Database
IQR	-	Inter Quarter Range
ITIM	-	Immuno-tyrosine-based inhibitory motif
kb	-	Kilobase

KCl	-	Potassium chloride
kDa	-	Kilo Dalton
KIR	-	Killer Immunoglobulin-like Receptor
LAIR	-	leukocyte-associated inhibitory receptors
LD	-	Linkage disequilibrium
LILR	-	leukocyte immunoglobulin-like receptors
LIR	-	leukocyte immunoglobulin-like receptors
LRC	-	leukocyte receptor complex
LTNP	-	Long-term non-progressor
LTR	-	Long terminal repeats
$MgCl_2$	-	Magnesium chloride
MHC	-	Major histocompatibility complex
MIIC	-	MHC class II compartment
MRC	-	Medical Research Council
NK	-	Natural Killer cells
NKCR	-	natural killer cell receptors
PBMC	-	Peripheral blood mononuclear cells
PCR	-	polymerase chain reaction
RBC	-	Red blood cells
RNA	-	Ribonucleic acid
SDS	-	Sodium dedocyl sulphate
SIV	-	Simian Immunodeficiency Virus
SIV_{CPZ}	-	Simian Immunodeficiency Virus of Chimpanzees
SIV_{SM}	-	Simian Immunodeficiency Virus of Sooty Mangabeys
Sqrt	-	Square root
SSP	-	Sequence specific primers
STD	-	Sexually transmitted diseases
TAP	-	Transporter protein associated with antigen processing
TCR	-	T cell receptor
TCR	-	T cell receptors
TE	-	Triz EDTA
T_h	-	Helper T cells
TM	-	Transmembrane region
Tris-HCl	-	Trizmabase hydrochloric acid
UNAIDS	-	Joint United Nations Programme on HIV/AIDS
USA	-	United States of America
UTR	-	Untranslated region
VL	-	Viral load
WBC	-	White blood cells
WHO	-	World Heath Organisation

CHAPTER 1 : INTRODUCTION

This chapter describes the epidemiology of the human immunodeficiency virus type 1 (HIV-1) and type 2 (HIV-2) with particular reference to the spread of HIV-2 in West Africa and other parts of the world. The structure and genomic organisation and life cycle of both viruses are discussed. A review of host immunity in HIV infection is presented with information on the role of Human Leukocyte Antigens (HLA) and Killer Immunoglobulin-like Receptors (KIR) genes in the control of HIV from previous epidemiological studies. The structure and functions of the major histocompatibility complex (MHC) in humans and those of the genes encoding KIR molecules are also described. The interaction between KIR and HLA, which governs to an extent the activity of NK cells, is an important part of the innate immune response to pathogens.

HIV epidemiology

The Acquired Immunodeficiency Syndrome (AIDS) is the most devastating pandemic of our times (**Figure 1.1**). It is caused by two closely related retroviruses of the family retroviridae namely: Human Immunodeficiency Virus (HIV) type 1 (HIV-1) and HIV type 2 (HIV-2). HIV-1 was first isolated in 1983 in France and United States of America (USA)[1 2], and three years after, the second retrovirus (HIV-2) was isolated from two West African patients[3] one from the Cape Verdes Islands and the other from Guinea Bissau. By 2007 it was estimated that more than 33 million people have been infected with both viruses[4]. While HIV-1 is present in nearly every country, HIV-2 remains geographically confined primarily to West Africa since its discovery in 1986. Both viruses have caused substantial morbidity and mortality worldwide.

Figure 1.1: The global distribution of HIV infection in adults
(Reproduced from Report on the Global AIDS Epidemic, UNAIDS/WHO, 2008)

HIV epidemic in Africa

Africa remains the epicentre of HIV/AIDS and is the continent most affected since the start of the pandemic in the 1980's. Sub Saharan Africa harbours more than 2/3rd of all people affected by this scourge and accounts for three quarters of all AIDS related deaths[4]. Although the global proportion of people living with the disease has levelled off in most parts of the world since the year 2000, the proportion of African living with HIV/AIDS in Sub Saharan Africa is continually rising as well as are the number of new infections (**Figure 1.2**). Of the 2.7 million new HIV infections recorded in 2007, 1.9 million were from Sub Saharan Africa. There are, however, regional differences with regards to the prevalence and incidence of HIV in this part of the world. For example, in most West African countries the prevalence of HIV infection is less than 2%, whereas it is greater than 20% in many Southern African countries[4]. Similarly, the epidemic varies significantly from one country to another even within the same region probably due to many complex reasons including but not limited to population genetic diversity,

social behaviour and environmental factors, and differences in the circulating strains/type of virus predominating.

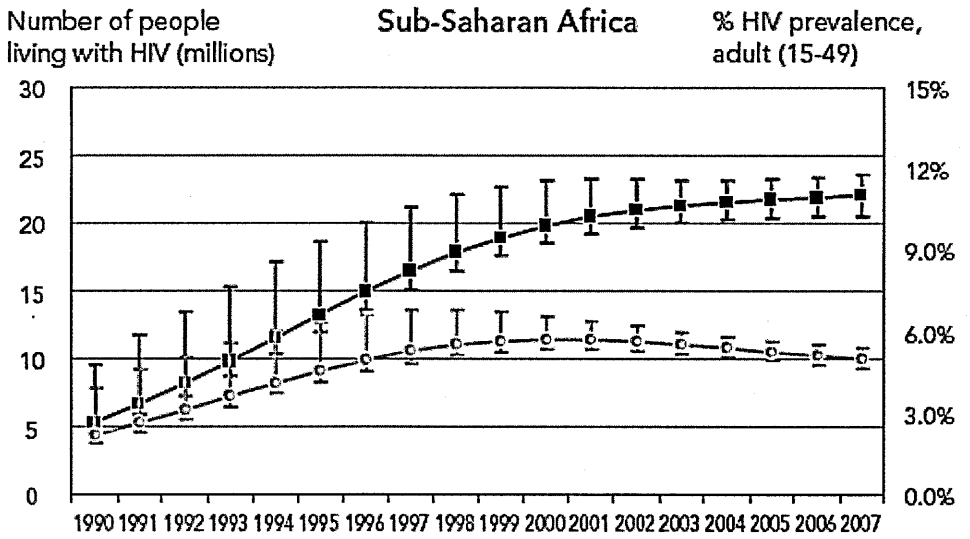


Figure 1.2 An estimate of the number of people living with HIV from the adult populations in Sub Saharan Africa for a 17 years period (1990-2007).
Black squares represent the number of people living with HIV; Red circles represent the percent HIV prevalence among adults aged 15-49 years; Bars indicate the range around the estimate (Reproduced from Report on the Global AIDS Epidemic, UNAIDS/WHO, 2008)

HIV epidemic in West Africa

AIDS in West Africa is caused by two closely related retroviruses HIV-1 and HIV-2, which differ greatly in pathogenesis[5]. Clustering analyses of sequences of Simian Immunodeficiency Virus (SIV) from chimpanzees (Pan troglodytes troglodytes) (SIV_{CPZ}) and sooty mangabeys (Cercocebus torquatus atys) (SIV_{SM}) and those of HIV isolated from people living in close proximity to these animals in Sierra Leone, Liberia and Ivory Coast suggest that HIV-1 and HIV-2 entered the humans in Africa through separate cross-species transmission events[6 7]. HIV-2 is believed to have resulted from at least eight distinct cross-species transmissions of SIV_{SM} in West Africa[8-12] and is thought to have been in the region prior to the war of independence in Guinea Bissau in the 1960's[13 14]. However, it was first described to have infected healthy commercial sex workers in Dakar Senegal[15] in

1985 and isolated the following year from two West African patients: one from Guinea Bissau and the other from the Cape Verdes Islands[3]. Since then, HIV-2 remained confined to the West African sub region. Only a few incident cases of HIV-2 infections have been reported outside West Africa. Interestingly, most of these cases have been linked to history of previous contacts with West Africa or countries with economic ties with the sub region (**Figure 1.3**). Examples of countries where occasional cases of HIV-2 infection have been reported include Mozambique, Angola, France and Portugal. Guinea Bissau has reported one of the highest frequencies of HIV-2 infection in the world, reaching 20% amongst adults aged 40 years and above living in urban areas[16 17]. For the past decade or so, the prevalence of HIV-2 in West Africa has been stable or declining in most countries while that of HIV-1, which was recently introduced in many of these countries, is steadily rising and overtaking HIV-2.

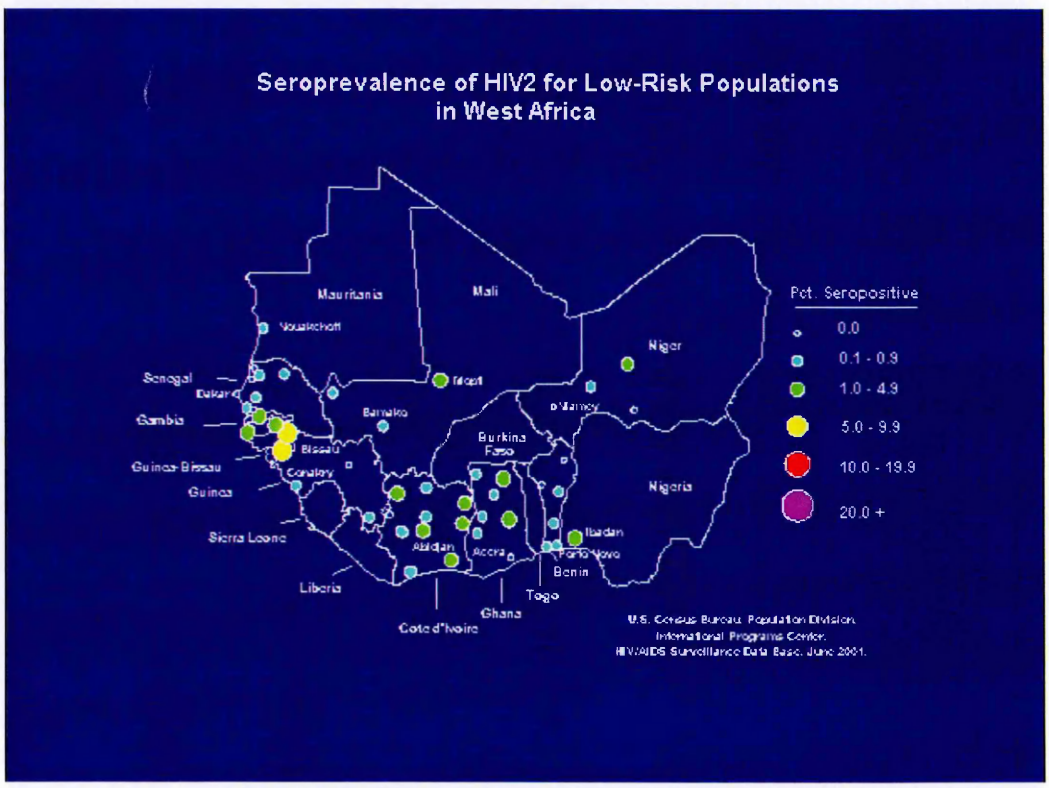


Figure 1.3: Map of West Africa
Showing the distribution of HIV-2 in the sub region (Source: The U.S. Census Bureau Population Division)

The Genetic structure and organisation of HIV

Morphology

The mature HIV virion has 2 strands of RNA (**Figure 1.4**), measures on average between 100-120nm in diameter and is roughly circular in shape. HIV particles surround themselves with a coat of fatty material known as the viral envelope. Projecting from the envelope are the gp120 and gp41 (collectively known as gp160) proteins. Just below the viral envelope is a layer of matrix proteins (p17). The viral core (or capsid) is found within the cytoplasm and constitutes the p24 proteins. Three vital enzymes are a prerequisite for HIV replication: reverse transcriptase, integrase and protease. They are located inside the viral capsid as are the viral genetic materials (RNA).

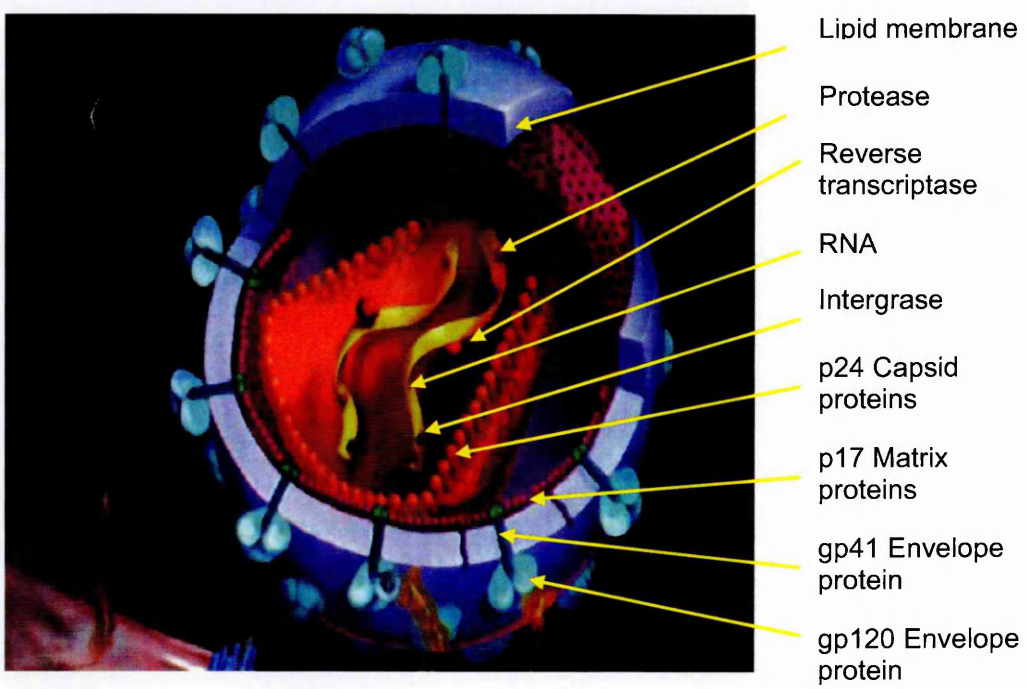


Figure 1.4: Structure of an HIV virion
(Adapted from Structural Biology of HIV (Felix Voigts Hoffmann), http://spider.iwr.uni-heidelberg.de/teaching/wes/2005/03_Structural_Biology_of_HIV_(Felix_Voigts-Hoffmann).pdf)

Genomic organization

HIV-2 genome sequences are much closer to that of SIV_{SM} than those of HIV-1. HIV-2 genome is made up of about 9700 nucleotide compared to 9500 nucleotides in HIV-1[18]. The difference arises from the difference in the length of long terminal repeats (LTR), which are slightly longer in HIV-2. But in general, the genetic layout of HIV-2 genome is similar to that of HIV-1: 5'LTR – *gag* – *pol* – central region – *env* – *orf*F – 3'LTR (**Figure 1.5**).

HIV has nine genes: three of them (*gag, pol* and *env)* contain the information needed to make structural proteins for new viral particles, the other six genes, known as accessory/regulatory genes: *tat, rev, nef, vif, vpr* and *vpu* (or *vpx* in HIV-2), code for proteins that control the ability of the virus to successfully infect a cell, replicate, assemble and bud-off new virions. The function of *tat, rev* and *nef* is to regulate viral gene expression. The *tat* gene is very important in maintaining the infectivity of the virus. Studies have shown that any mutation or deletion in the *tat* gene abolish viral infectivity and that the control of HIV gene expression depends on the interaction between the *tat* and *rev* genes[19 20]. On the other hand, the *nef* gene mediates the enhancement of viral infectivity and replication, downregulates CD4 and MHC class I molecules from the cell surface of the infected cell, and also modulates cellular activation pathways.

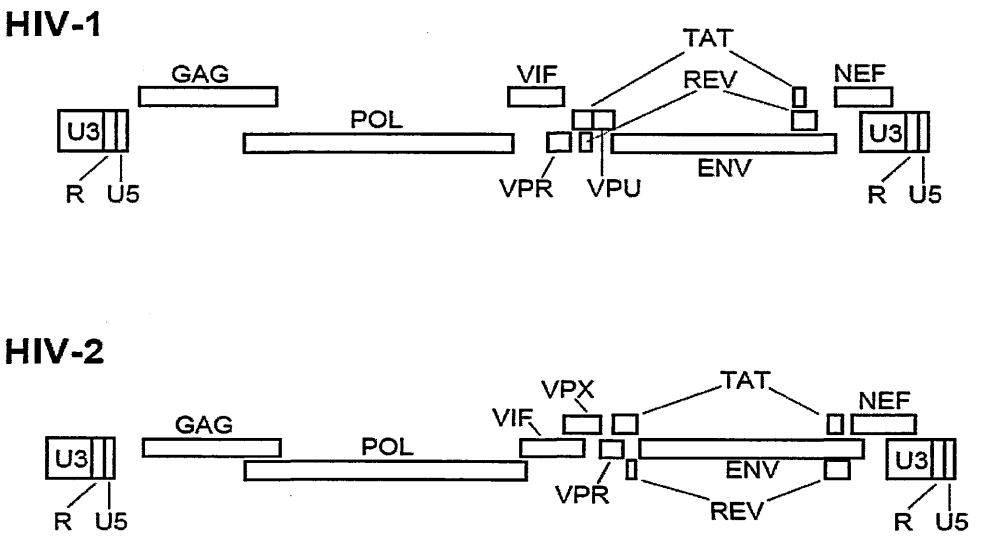


Figure 1.5: HIV-1 and HIV-2 genomic organization
(from http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/1995/INTRO.pdf)

The HIV replication cycle

The HIV life cycle can be divided into two phases: early phase and late phase (**Figure 1.6**). The former begins when a virus particle attaches to a CD4 receptor and one of two co-receptors (CCR5 or CXCR4) on the surface of a $CD4^+$ T lymphocyte. This interaction subsequently allows the viral envelope to fuse with the host cell membrane. The contents of the HIV particle containing the genetic material (RNA) are then released into the cell, leaving the envelope behind. Upon entry into the cell, an HIV enzyme (reverse transcriptase) converts the single stranded viral RNA into a double stranded HIV DNA, which is then transported into the cell's nucleus. Once in the nucleus, the enzyme integrase splices and integrates the viral DNA into the host DNA. The integrated HIV DNA is known as a provirus and may lie dormant within a cell for a long time (several years) awaiting cell activation. When the cell is activated the provirus uses the host enzyme (RNA polymerase) to make copies of itself. The human DNA (including the proviral DNA) is converted into messenger RNA, which is transported outside

the nucleus, and used as a blueprint for producing long chains of HIV proteins and enzymes. Among the strands of messenger RNA produced by the cell are complete copies of HIV genetic material. Another HIV enzyme (protease) cuts the long chain of HIV proteins into smaller individual HIV proteins. These gather together with the newly made copies of HIV RNA and enzymes to form new viral particles which eventually get out of the host cell, develop further into mature viruses and attach to other CD4 molecules and co-receptors to infect other $CD4^+$ cells.

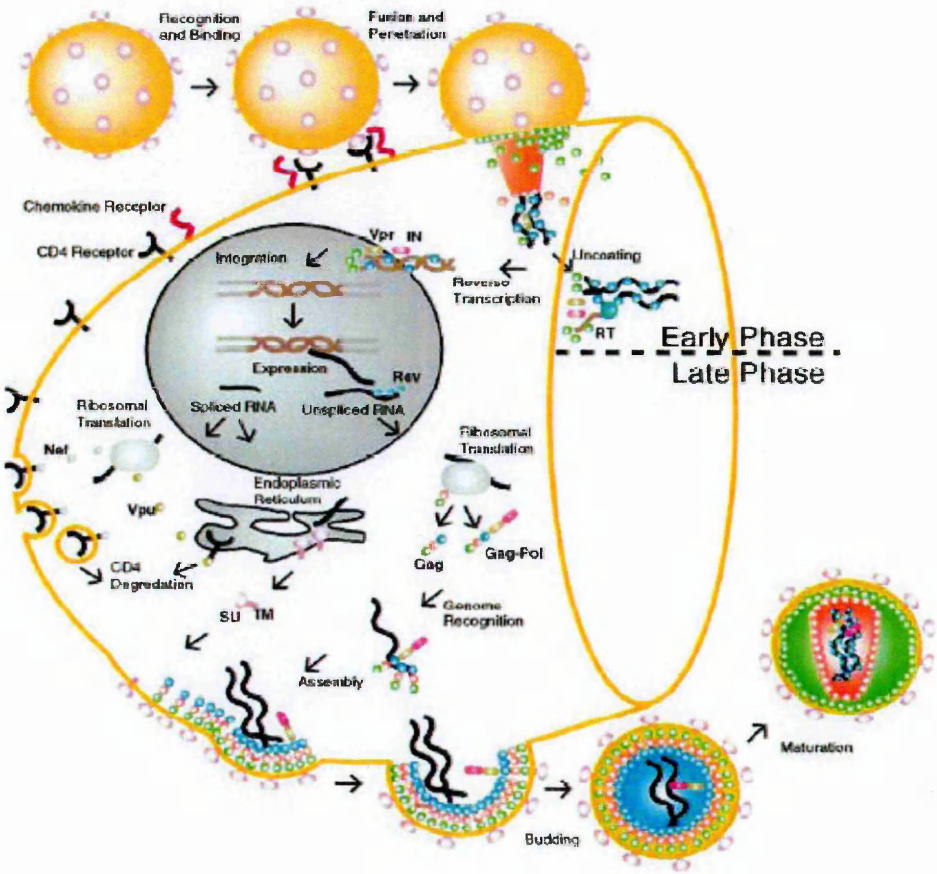


Figure 1.6: Features of HIV replication cycle.
The early phase begins with CD4 recognition and ends with integration of the HIV DNA into the host genome; the late phase includes all other events from transcription of the integrated DNA to virus budding and maturation (Reproduced from: Turner et al. 1999 Structural Biology of HIV, JMB 285:1-32).

Comparison of HIV-1 and HIV-2 infection

HIV-1 and HIV-2 share between 30-60% similarities in most genes. However the viruses differ in many aspects. The HIV-2 genome contains the *vpx* gene that takes the place of the *vpu* gene in the HIV-1 genome[18]. The ease of transmission and pathogenicity of HIV-2 are reduced compared to HIV-1[21]. HIV-2 uses a broad range of chemokine receptors such as CCR5, CXCR4, CCR1, CCR3, CCR2B, BOB and BONZO[22 23]. The proviral genomes of HIV-1 and HIV-2 are approximately 10 kb in length. The error rate of the virus reverse transcriptase is likely to be higher in HIV-1 than in HIV-2. A number of studies have estimated the rate of base substitution errors to range between 1:2000 to 1:4000, suggesting a value between 5 and 10 misincorporation per virus genome in each round of reverse transcriptase and as a result, no two virus isolates are genetically identical[24 25].

Natural history of HIV infection

HIV can be transmitted through blood, blood products, bodily fluids, breast milk, sexual contact with an infected person, and vertically from a mother to her infant *in utero* or during delivery[26]. A successful HIV infection is largely dependent on the interaction between the host and viral factors. A balance between the host factors and those of the virus is also a major determinant of the outcome of the infection. HIV targets are specific host cells carrying receptors and co-receptors necessary for binding and entry into the host cell. HIV infection activates the host immune system causing a wide pool of cell death (apoptosis). As a result, there is depletion of $CD4^+$ T lymphocytes and a contest for power between the virus and the immune system. Some people, however, are able to mount a better immune response to control the rate at which the virus multiplies. Such individuals, commonly known as long-term non-progressors (LTNP), may remain chronically

infected for many years with minimal loss of $CD4^+$ T cells and very low or undetectable plasma viral load. In contrast, other people succumb to the disease within a very short time following infection when their immune systems are unable to keep the virus under check by failing to control its replication successfully. Following infection, the clinical course of HIV can be divided into three phases: acute/primary, asymptomatic/latent and AIDS (**Figure 1.7**).

Acute/primary phase of HIV infection

The initial/acute infection with HIV is subclinical and may be clinically undetected. Within 2-3 weeks the infected individual may develop a rash followed by fever and lymphadenopathy. Opportunistic infections are generally not seen at this stage and a large majority of patients remain asymptomatic.

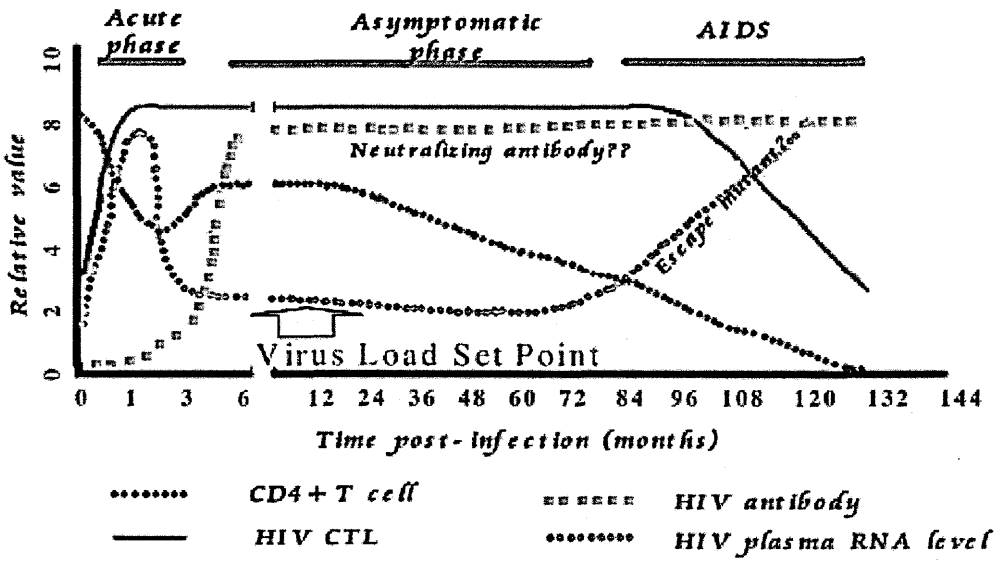


Figure 1.7: Phases of HIV infection and $CD4^+$ T cell and viral load dynamics
(Reproduced from Paranjape RS, Indian J Med Res 2005 121:240-55)

The primary phase of HIV infection begins soon after seroconversion – the time when the HIV antibody test becomes positive for the first time. Seroconversion usually takes place four to eight weeks after primary infection with

HIV[27 28]. Symptoms are rare during the acute phase of HIV infection but some 15% of individuals at this stage experience some syndromes including sore throat, headache, mild fever, fatigue, muscle and joint pains, swelling of the lymph nodes, rash, and (occasionally) oral ulcers, lasting between one and two weeks[27].

The initial burst of HIV, releasing up to 10^7 HIV RNA copies per ml of plasma[27] takes place during the acute phase of infection and because of the high level of circulating virus after the burst, there is a rapid systemic dissemination of the virus throughout the body. This stimulates the immune system to respond vigorously to the attack, with both humoral and cell mediated immune responses coming into play to curtail the spread of the virus[29 30]. Consequently, a dramatic downregulation of viraemia and virus expression in peripheral blood mononuclear cells (PBMC) follows the vigorous immune pressure mounted against the virus[31 32]. Unfortunately, this reduction in viral load in the peripheral blood does not completely eliminate viruses from the body but only drives them to an equilibrium level termed the 'viral set point' (**Figure 1.7**). This is an indication that the control of virus expression by the immune response is ineffective and can not completely arrest viral replication during acute/primary infection. However successful downregulation of the amount of circulating virus to a low level that can be tolerated by the body (viral set point) is usually associated with resolution of the acute viral syndrome. In the majority of HIV-1 infected individuals this occurs six to eight weeks after the onset of the symptoms and coincides with the first appearance of HIV-specific immune responses[27 32].

The immunologic events that take place during the course of acute/primary HIV infection are complex and the role played by the various arms of the immune system immediately after primary infection to control the spread of the virus is currently unclear. Several studies have demonstrated that vigorous HIV-specific humoral and cell-mediated immune responses can be detected very early during

the primary phase of HIV infection in the large majority of HIV-1 infected individuals[30 32-34], and they are likely to contribute to the suppression of the initial burst of viral replication[27]. High titres of anti-HIV antibodies are found at the time of peak viraemia and at the point of downregulation of viraemia[32] but these antibodies produced during the primary phase of infection are non-protective and lack neutralising activity needed to reduce the amount of free living viruses from the periphery[30 33]. It is surprising and still not well understood why neutralising antibodies are not produced during the acute/primary phase of HIV infection but appear rather late during the course of the disease.

The $CD8^+$ T lymphocyte count increases dramatically during acute/primary phase of HIV infection (up to 20-fold above normal in many patients) while the $CD4^+$ T cell count drops significantly. The appearance of HIV-specific $CD8^+$ T cell-mediated cytotoxicity also known as the cytotoxic T lymphocyte (CTL) function that are readily detected in peripheral blood mononuclear cells (PBMC) correlate well with the downregulation of the high viraemia observed following primary HIV infection[30 34]. It is proposed that $CD8^+$ T cells contribute to the downregulation of viraemia from the initial viral burst by directly killing cells in which active virus replication is taking place (a cytotoxic mechanism) or by secreting cytokines and other soluble factors to mediate the suppression effect[27 34].

A panel of cytokine expressions have been detected during primary infection by many investigators[35-37]. IFN-γ expression predominates with the level remaining constant throughout the primary phase of infection in certain individuals while in others, its peaks very early post infection[35]. It is expressed predominantly by the $CD8^+$ T cell population. In contrast, the levels of TNF-α and IL-10 peak late during the course of primary infection[35]. The levels of other cytokines including IL-2, IL-4, and IL-6, measured during the acute/primary phase of HIV infection are usually very low or barely detectable in many subjects[35].

The reason(s) associated with lack of clearance of HIV from the system by the vigorous and specific immune responses observed in most HIV infected subjects during the acute/primary phase of HIV infection remain speculative. Some have suggested that this could be due to the formation of a large pool of latently infected cells that succeeded in avoiding the HIV-specific CTL responses[36 38] while others proposed that the trapping of infected cells and viral particles in different body compartments not readily accessible by specific CTLs serves as reservoir for de novo infection[39 40].

It is however well documented that high viral turnover leading to complete changes in virus phenotypes and genotypes helps HIV to escape from host immune responses[38].

Asymptomatic/latent/chronic phase of HIV infection

This phase begins immediately after the reduction of viraemia and the resolution of the primary syndrome. In the majority of HIV infected individuals, the asymptomatic phase spans over many years, on average 10 and 15 years for HIV-1 and HIV-2 infected individuals, respectively. The virus load is at its set point but active viral replication persists at much slower rate[41] compared to that observed following primary infection.

Variegated and broad HIV-specific CTL responses are continually detected over the latent period in a good proportion of HIV infected individuals targeting envelope, gag, pol and some of the regulatory proteins including nef and tat[27 42]. In addition, $CD4^+$ T cell responses to a variety of HIV proteins have also been observed in many subjects[27], but their role in any attempt to control viral replication is not clear.

Neutralising antibodies are active in reducing the numbers of free virions but are less effective in blocking cell-to-cell transmission[43-45]. The activities of

neutralising antibodies are detected for a very long time during the latent phase of infection, but rarely neutralise the currently circulating strain as viral escape through envelope mutation occurs rapidly. In addition, a variety of other antibodies against HIV proteins are also detected during this phase[27].

AIDS phase of HIV infection

AIDS defining illnesses and a significant drop of $CD4^+$ T cell count (< 200 cells per µl of blood) characterise this stage of infection. The virus load increases dramatically and continuously as will all other virologic parameters in the peripheral blood and lymphatic system[27 40].

A profound immunosuppressive state with decline of HIV-specific cytotoxic activity is observed at this last stage of the disease[42]. Antibody titres against most of the HIV proteins significantly decrease and neutralising antibodies are rarely detected. In contrast, the level of production of certain cytokines including IFN-γ and IL-10 mainly by $CD8^+$ T lymphocytes remains high suggesting that $CD8^+$ T cell activity remains the only protective mechanism at the AIDS stage, since most of the $CD4^+$ T cells are either infected or destroyed[27].

Why should HIV target $CD4^+$ T cells: the function of T helper lymphocytes in the immune system

HIV (both type 1 and type 2) primarily targets a subgroup of lymphocytes (a type of white blood cells) that carry the CD4 molecules on their cell surface for destruction. These cells are collectively known as $CD4^+$ T cells and are not known to have significant cytotoxic activity[28]. The destruction of a particular category of $CD4^+$ T cells known as T helper (T_h) cells, brings about the immunodeficiency syndrome observed in all HIV infected individuals at the last stage of disease. Not all T cells with CD4 molecules on their cell surface are helper cells. In line with this, some natural killer T cells, and regulatory T cells express CD4 but do not

perform the same function as T_h cells. Helper T cells are particularly important to the immune system as they help in maximising the ability of other immune cells to deal with invading pathogens. The role of T_h cells is to activate and direct other immune cells in carrying out their function and preserve the integrity of the body.

During the course of HIV infection, $CD4^+$ T cells are massively infected after the initial burst of the virus at the primary phase of infection and a good proportion of them are destroyed leading to a sudden drop in $CD4^+$ T cell counts in the peripheral blood. But this event is short lived since the body quickly accommodates, probably by producing more T helper cells from the thymus to help other immune cells bring the situation to normal. It has been proposed that during the latent phase of HIV infection, the viruses present have low affinity for T cells (but high affinity for other $CD4^+$ antigen presenting cells (APCs), such as dendritic cells (DCs) and macrophages) and as such, the rate at which $CD4^+$ T cells are destroyed is minimised while the body compensates for loss of T_h cells by generating more cells from the thymus. Towards the end of the latency period, viruses gain tropism for T lymphocytes, probably by switching co-receptor usage. Early in infection, the virus tropism is predominantly for CCR5, which is expressed on APCs and memory T_h cells, but later in infection the virus usually switches to use the CXCR4 receptor, which is widely expressed in the naïve T-cell pool. Once these viruses become lymphotropic, they begin to infect $CD4^+$ T cells more effectively such that the body's replenishment mechanism becomes exhausted and overwhelmed. This coincides with the time that the much-needed help from T_h cells gradually becomes unavailable, probably because the $CD4^+$ T cell population pool is diminishing progressively. Other immune cells such as $CD8^+$ T cells and B cells that depend on getting help from T_h cells to carry out their immunologic functions[28] become inert and no longer able to protect the body from invading pathogens, including newly generated HIV virions. Various pathogens

escaping the immune surveillance then cause opportunistic infections that gradually increase in severity, leading to the collapse of the immune system and death.

T_h cells produce a wide range of cytokines. The pro-regulatory (type 1) cytokines they produce include IFN-γ and IL-12; and the pro-inflammatory cytokines (type 2) include IL-4, IL-5, IL-10, and IL-13. These cytokines are very important in activating and maintaining the effector function of other immune cells such as phagocytes (e.g. macrophages). The profile of cytokines produced by T_h cells during HIV infection and the fine balance between these cytokines in terms of qualities and quantities have been shown to correlate with the stage of infection[46] [47]. For example, a strong $CD4^+$ T cell response during the primary phase of HIV-1 infection was found to correlate with slow disease progression and better control of viral replication[46]. Recently, Ostrowski et al. and Imami and colleagues independently demonstrated that impaired production of cytokines predicted mortality in HIV-1 infected individuals while preserved capacity to produce the correct cytokine profile was associated with prolonged survival and LTNP[48] [49].

In HIV-2 infection, unlike HIV-1, the large majority of patients maintain a healthy $CD4^+$ T cell population with a strong HIV-2-specific T_h cell response that is highly polyfunctional (producing both IFN-γ and IL-2) with high proliferative capacity for a long time post infection[50]. Duvall and colleagues proposed that the maintenance of a well-preserved and functionally heterogeneous population of HIV-specific T_h cell response can significantly delay disease progression; a phenomenon frequently observed in the majority of HIV-2 infected subjects in West Africa.

The role of cytotoxic T lymphocytes in HIV infection and disease

The cytotoxic T (Tc) cells are a subgroup of leukocytes that are vested with the ability to defend the body against intracellular pathogens (e.g. certain bacteria and viruses). Tc cells express a wide variety of T cell receptors (TCR) that interact with target cells presenting specific antigenic peptides bound to MHC class I molecules at their cell surfaces.

Following HIV infection, $CD8^+$ T cells are exposed to virally infected cells. Tc cells recognise HIV infected cells when they express HIV antigens on the cell surface bound to a specific HLA class I molecule. Tc activity is dependent on cell-to-cell contact. During the cell-to-cell interaction, the Tc enzyme known as perforin, perforates the cell membrane of the infected cell and allows granzyme (another enzyme) to penetrate into the infected cell. The presence of granzyme in the cytoplasm of the infected cell results in the activation of a series of enzymes that eventually lead to apoptosis (programmed cell death)[28].

During the acute phase of HIV infection, the HIV-specific $CD8^+$ T cell response is detected after the initial viral burst and is thought to contribute significantly to the decline in viraemia[29 30], a process that influences disease progression[51 52]. HIV-1 Gag- and Nef-specific CTLs have been reported to appear very early post HIV infection but during the course of the disease HIV-specific CTL responses to most of HIV antigens would normally be detected. In HIV-2 infection, Gag-specific CTL responses are readily detected at a higher magnitude in people who control the viraemia better[53]. However, it is not yet well understood why these CTL responses are incapable of preventing further infection of $CD4^+$ T cells or completely stop viral replication in the same individual[54 55]. The progressive loss of CTL activity observed as the disease progresses has been attributed to the gradual loss of $CD4^+$ T cell population and particularly to the decrease in the

number of T helper cells[46 56]. Some defects in enzyme expression and cytokine production have been reported to also affect CTL activities[57 58].

Immunogenetic factors that influence susceptibility to infectious diseases in Africa.

Infectious diseases are a huge health problem worldwide and account for most of the deaths in developing countries particularly those in Sub Saharan Africa. In this part of the world, a combination of pathogen and human genetic factors, complexes with environmental factors to determine the susceptibility to and the outcome of any infectious disease. For many years, great efforts have been made by researchers to identify and characterise relevant genes that influence susceptibility to bacterial, viral and parasitic diseases[59-64].

Given the complex nature of infectious diseases that prevail in Sub Saharan Africa, the high inter-population heterogeneity, and the complex immune responses that follow exposure to an infectious agent, it seems likely that several human genes may collectively and/or individually play a role in determining susceptibility to infection. In addition, the outcome of the infection might also depend on the genetic profiles of the infected individuals[65 66].

Genes of the HLA loci are particularly important in the initiation and regulation of immune responses to diverse range of pathogens. Their role in the susceptibility and progression to several infectious diseases have been thoroughly investigated in cohorts and large population studies[59 60 62 67-71]. In Africa, one of the largest HLA association studies in an infectious disease was done in malaria in The Gambia[60]. The results of that study showed for the first time that there were strong HLA associations with the outcome of infection by demonstrating that *HLA-B*5301* and *HLA-DRB1*1302* independently associated with protection against severe malaria in young African children[59 72].

Genetic factors that influence susceptibility to HIV infection and progression to AIDS.

Since the onset of HIV pandemic, a number of approaches have been used to estimate the role of host genetic polymorphisms, with particular interest accorded to genes of the MHC region, in the susceptibility to HIV infection and disease progression. A few cohort studies have clearly demonstrated the differential susceptibility to HIV acquisition in "at risk" groups such as the commercial sex workers (CSW) in The Gambia and Kenya[73-75]. The control of $CD4^+$ T cell depletion by virus specific cytotoxic T cells (CTL) is an important immunogenetic response protecting individuals from both infection and rapid disease progression[76 77]. The most frequently studied phenotypes in large population and cohort studies have been "time to AIDS" and "time to death" following HIV infection[65].

An increasing number of chemokine and cytokine genes and their receptors have also been identified from association studies to have some impact on HIV outcomes (reviewed in ref 78[78]). Of note is the association of a variant of CCR5 (CCR5Δ32) with acquisition of HIV-1 and progression to AIDS. People who carry two copies (homozygotes) of this variant of CCR5 gene are protected against heterosexual acquisition of HIV-1 while HIV-1 infected individuals carrying only one copy of CCR5Δ32 (heterozygotes) progress very slowly to AIDS[79 80]. This mutation has not yet been found in people of Sub Saharan Africa descent[81] implying that the slow progression to AIDS observed in the majority of HIV-2 infected individuals in West Africa can not be attributed to the presence of this mutant in West African patients. One study showed that a single nucleotide polymorphism (SNP) in the CCR5 ligand RANTES has been linked to more rapid disease progression[82]. Others have demonstrated that an amino acid change in

CCR2 gene from Valine to Isoleucine (CCR2V64I) disrupts the interaction with CXCR4 and reduces its expression, leading to a delay in the onset of AIDS in HIV-1 infected individuals[83]. Winkler and colleagues found in a large multi-cohorts study that a single mutation in stromal derived factor 1 (SDF-1), the ligand for CXCR4, considerably delays the onset of AIDS in HIV-1 patients[84]. Polymorphisms in other cytokine genes such as IL-2, IL-4, IL-10, NRAMP1, IFN-γ, and TNF-α have been inconsistently associated with HIV outcomes in different populations[85-87]. A summary of some of these polymorphic genes with specific mutations that affect the outcome of HIV infection is presented in **Table 1.1**

Table 1.1: Human Genes that affect HIV/AIDS

Gene	Allele	Mode	Effect	Mechanism of action
HIV entry				
CCR5	Δ32	Recessive	Prevent infection	Knockout CCR5 expression
	Δ32	Dominant	Prevent lymphoma (L)	Decrease available CCR5
	Δ32	Dominant	Delay AIDS	Decrease available CCR5
CCR5	P1	Recessive	Accelerate AIDS (E)	Increase CCR5 expression
CCR2	I64	Dominant	Delay AIDS	Interact with CXCR4
CCL5	In1.1c	Dominant	Accelerate AIDS	Decrease RANTES expression
CXCL12	3'A	Recessive	Delay AIDS (L)	ImpedeCCR5-CXCR4 transition
CXCR6	E3K	Dominant	Accelerate PCP (L)	Altered T-cell activations (?)
CCL2-CCL7-CCL11	H7	Dominant	Enhance infection	Stimulate immune response
Cytokine anti-HIV				
IL10	5'A	Dominant	Limit infection	Decrease IL10 expression
	5'A	Dominant	Accelerate AIDS	Decrease IL10 expression
IFNG	–179T	Dominant	Accelerate AIDS (E)	
Acquired immunity, cell mediated				
HLA	A,B,C	Homozygous	Accelerate AIDS	Decrease breadth of HLA class I epitope recognition
	*B*27*	Codominant	Delay AIDS	Delay HIV-1 escape
	*B*57*	Codominant	Delay AIDS	Delay HIV-1 escape
	*B*35-Px*	Codominant	Accelerate AIDS	Not clear
Acquired immunity, innate				
KIR3DS1	3DS1	Epistatic with HLA-Bw4	Delay AIDS	Not clear

E, acts early in AIDS progression; L, acts late in AIDS progression; PCP, *Pneumocystis carinii* pneumonia; ?,plausible mechanism of action with no direct empirical support. All other mechanisms have quantitative expression differences among alternative alleles that suggest a mechanism. (Adapted from: O"Brien SJ and Nelson GW, Nature genetics 2004 36(6):565)

A number of HLA alleles have been identified that may influence HIV outcomes but consistent associations are rare probably due to lack of power to detect particularly strong associations. One other reason could be that of inadequate sample sizes and inappropriate selection of controls in multiethnic groups of subjects participating in these association studies.

It is important to note that innate antiviral immunity is an ancient defense mechanism, reflecting the struggle of multicellular organisms such as humans to survive. A wide variety of other innate defences against HIV other than HLA and KIR have been reported for ecample APOBEC-3G and Trim5 alpha.

The HLA system

Genes of the major histocompatibility complex (MHC) in humans govern the expression of proteins on the surface of cells, which were originally described in the recognition of 'self' and 'non-self' in tissue transplantation and are key elements of the immune response to infection. This complex in man is called the "Human Leukocyte Antigen" (HLA) system. It comprises a set of highly polymorphic genes located on chromosome 6 (6p21.31) that encode cell surface molecules, vested with the ability to present short peptides to T cells for recognition, resulting in the initiation of an adaptive type of immune response.

The HLA system was described in the early 1960's[88-90], following earlier work in murine transplantation models on histocompatibility antigens 2 (H-2)[91]. This discovery was facilitated by observations made at the start of the last century by scientists interested in knowing why inbred mice could accept tumour grafts that were usually rejected. A group of genes from a number of loci that were contributing to graft/tissue rejection were later identified. These loci were collectively called histocompatibility loci. The locus that contributed most to graft compatibility was named the major histocompatibility complex while the remainder was called the minor histocompatibility complex.

Further research demonstrated that genes of the MHC encode polymorphic molecules that differ between strains and were the main targets for alloantibodies made against them by the recipient immune system. In this respect, these polymorphic molecules were considered to be 'alloantigens' (proteins occurring in a member of a species, which when introduced into members of that species who lack it, will induce an immune response by stimulating the production of alloantibodies). In the past, serological techniques were used to define distinct classes of MHC antigens in both human and mouse, but with the advent of

molecular (DNA-based) techniques, previous classifications have been refined and the discovery of new antigens has been simplified.

In mice, alloantigens are present in abundance on the surfaces of both red and white blood cells (RBC and WBC, respectively) but in humans, these antigens are expressed only on the leukocyte cell surface and absent on RBC, thus the name "human leukocyte antigens" (HLA). So far, three distinct classes or regions of HLA have been defined mapping to genes on the short arm of chromosome 6 in humans (**Figure 1.8**). These regions contain several genes that have repeatedly been implicated in immune responses to "non-self" including pathogens that have successfully breached the host innate barriers and gained entrance into the body.

Because of its great diversity and pivotal role in modulating host immune responses to pathogens, the HLA system has been frequently investigated both in biological and anthropological studies. The constant probing of this region of the human genome, coupled with the advent of molecular techniques in the 1980's led to a rapid rate of discovery of new HLA antigens in different populations and ethnic groups around the world. The number of known alleles at these loci continues to grow exponentially (**Figure 1.9**) as more and more populations are investigated aided by the availability and affordability of newer typing techniques to researchers in low-income countries.

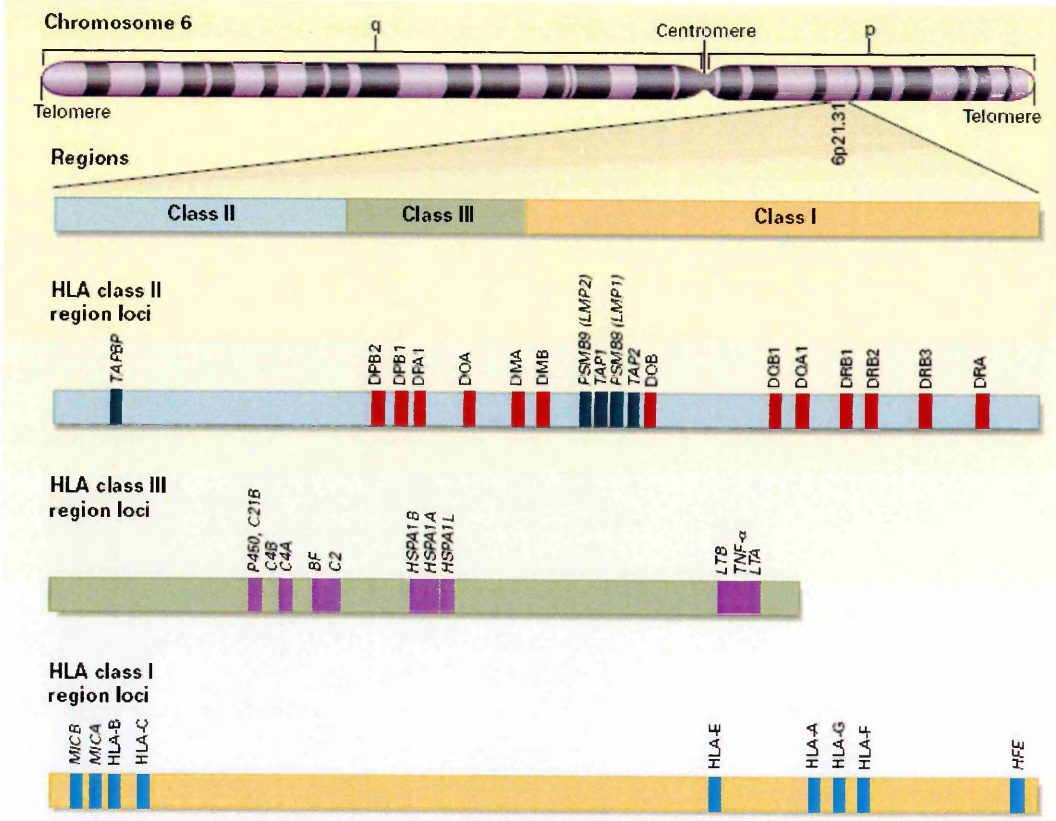


Figure 1.8: The HLA region on human chromosome 6
The class I region is nearest the telomere while class II genes are closer to the centromere. The third region comprises genes encoding the complement factors (C2, C4 and factor B (Bf)) and is sandwiched between classes I and II. (Reproduced from ref. 92)

More than 200 genes have been mapped to the HLA region spanning about 4 Mb of the human genome but only 20% of them code for leukocyte antigens [92 93] that are involved in regulating host immunity. Those directly involved in immune regulation are called the "classical *HLA* genes", which can broadly be divided into two main classes: *HLA* class I (*A*, *B*, and *C*) on one hand and *HLA* class II including *HLA-DR*, *-DQ*, *-DP*, *-DO*, and *-DM* among others.

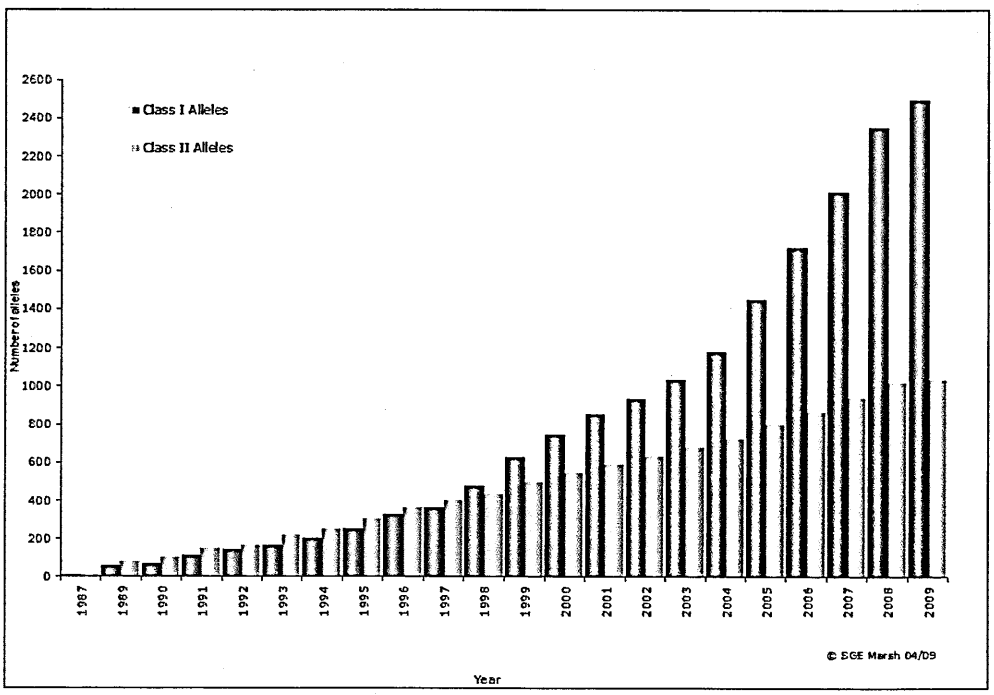


Figure 1.9: Classical HLA class I and class II antigen discoveries
(Reproduced from The IMGT/HLA database[94])

HLA class I: molecule, structure and gene organisation

HLA class I antigens are proteins encoded by the class I genes and consist of two polypeptides chains: a heavy chain called alpha (α) and a non-covalently linked light chain known as beta-2 microglobulin (β_2m).

The α-chain is a glycoprotein measuring about 45 kDa with five distinct domains namely the extracellular domains (α1, α2, and α3), the transmembrane region and the cytoplasmic tail (**Figure 1.10**). The two membrane-distal domains (α1 and α2) fold in a special way to form the peptide-binding groove, which accommodates and holds a short pathogen- or self-derived fragment (8-10 amino acid long) for presentation to cytotoxic ($CD8^+$) T lymphocytes. Changes in the amino acid sequence of the binding groove determine the binding efficacy and stability of peptides delivered from the processing cascade with major impact on T cell recognition. Most of the polymorphisms in the class I genes lie in specific exons coding for proteins that form the peptide-binding groove.

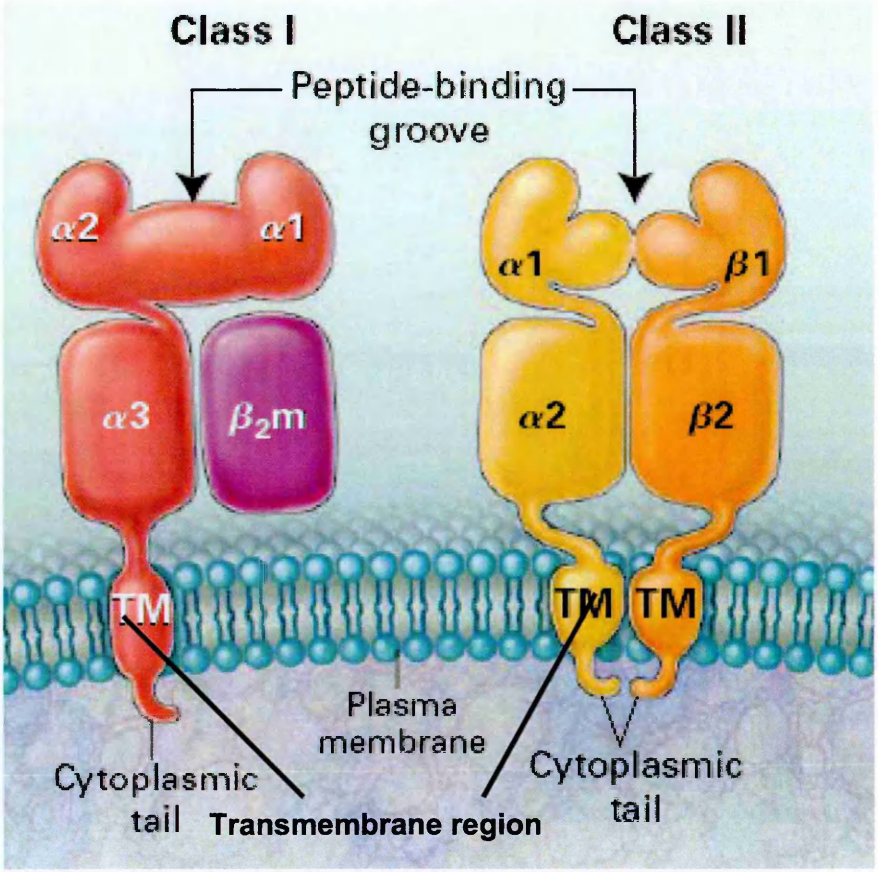


Figure 1.10: HLA class I and II molecules on the cell surface
(Reproduced from ref. 92)

The light chain (β_2m) is a water-soluble protein measuring 12 kDa in weight that constitutes an integral part of all HLA class I molecules and some class I-like molecules. Beta-2 microglobulin is encoded by a monomorphic gene located on chromosome 15 and is thus not classified as part of the HLA family because it is a product of a gene that is not located within the MHC. Although β_2m is not polymorphic, it constitutes an important component of all HLA class I molecules. Several reports have demonstrated that soluble β_2m in the serum or plasma is a good marker of immune activation. Its levels are abnormally high in people infected with persistent viruses such as HIV, cytomegalovirus (CMV), and hepatitis B virus (HBV)[95-97] to name a few.

There are over 20 genes within the class I region that are ubiquitously expressed but only three are considered to be highly polymorphic (*HLA-A*, *HLA-B*,

and *HLA-C*). Each class I gene encodes a heavy chain (α-chain) and each domain is coded by a separate exon (**Figure 1.11**). The leader sequence is encoded by exon 1; the extracellular domains are encoded by exons 2, 3 and 4 respectively; the transmembrane region by exon 5; and the cytoplasmic tail by exons 6 and 7. Variations in exons 2 and 3 are the basis of the antigenic polymorphisms observed in the classical class I genes (*HLA-A, -B,* and *-C*).

The number of *HLA* class I alleles currently described in the immunogenetics database is 3007 alleles[94]. *HLA-B* is by far the most polymorphic with 1431 known alleles (**Table 1.2**) followed by *HLA-A* and *HLA-C* (893 and 569 alleles, respectively). The other non-classical class I genes (including *HLA-E, -F,* and *-G*) are less polymorphic.

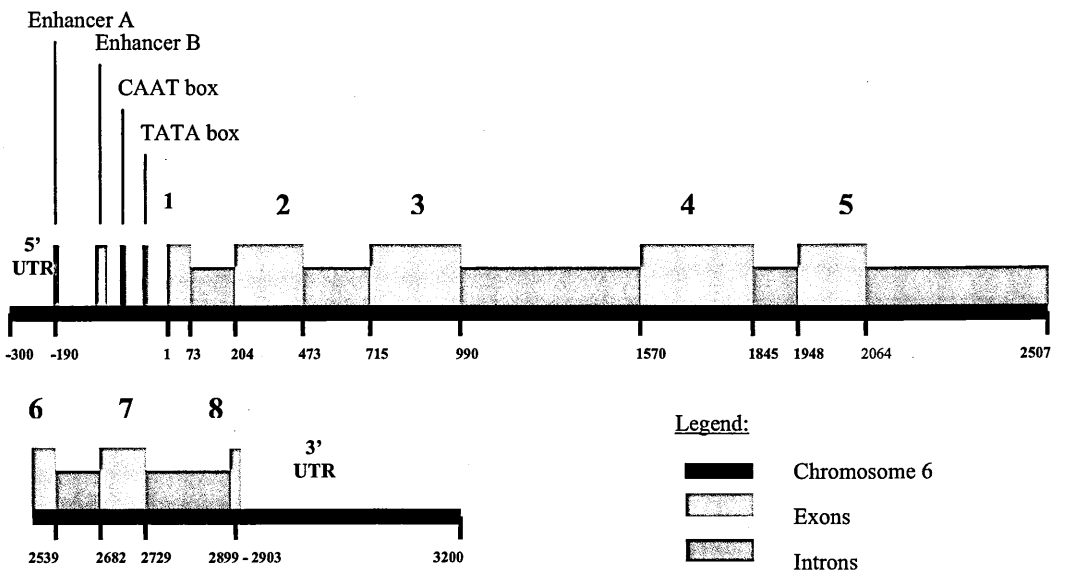


Figure 1.11: The HLA class I gene organisation

HLA class II: molecule, structure and gene organisation

Class II molecules are heterodimers consisting of two heavy chains: a 35 kDa α-chain and a β-chain measuring 28 kDa[98]. Each of these heavy chains anchors on the cell membrane and has four domains - two extracellular domains (α1, α2; β1, and β2 for alpha and beta chains, respectively), a transmembrane

region and a cytoplasmic tail (**Figure 1.10**). The two membrane-distal extracellular domains (α1 and β1) are referred to as the peptide-binding domains because they fold in an interactive manner to form the peptide-binding groove with similar function as described above for HLA class I molecules except that they present their fragments (peptides) to $CD4^+$ T cells and have the ability to accommodate longer peptides (up to 14 amino acid long).

In contrast to HLA class I molecules that are ubiquitously expressed on nucleated cells, the products of *HLA* class II genes are expressed only on subpopulations of immune cells including B cells, macrophages, activated T cells, DCs and other APCs[99]. Their levels of expression vary according to the tissue location (**Table 1.3**).

Table 1.2: Number of alleles at the HLA class I and HLA class II loci

Class	Genes	Number of alleles
I	A	893
	B	1431
	C	569
	E	9
	F	21
	G	45
II	DRA	3
	DRB	814
	DQA1	35
	DQB1	106
	DPA1	28
	DPB1	136
	DMA	4
	DMB	7
	DOA	12
	DOB	9

There are five different class II isoforms: *HLA-DM, -DO, -DP, -DQ* and *-DR* (**Figure 1.8**). Each of these types has a gene encoding an α-chain designated A (e. g. *DQA1, DRA*) and one or more corresponding genes encoding for β-chains designated B (e.g. *DQB1, DRB1, DRB4*, etc).

Table 1.3: The differential expression of classical HLA class I and HLA class II molecules on human cell surfaces

Cell type	HLA class I	HLA class II
B cells	+++	+++
T cells	+++	+/-
Macrophages	+++	++
Other APCs	+++	+++
Thymic epithelial cells	+	+++
Neutrophils	+++	-
Brain cells	+	-
Hepatocytes	+	-
Kidney cells	+	-
Red blood cells	-	-

The structural organisation of the genes encoding each class II heavy chains is similar to that of class I heavy chains, i.e. exon-intron organisation (**Figure 1.12**). With the exception of the transmembrane region of the α-chains, each domain of a class II protein is encoded by a separate exon[100 101]. The leader sequence in both α and β chains is encoded by exon 1; exons 2 and 3 encode for α1 and α2 domains (α-chains), and β1 and β2 domains (β-chains); the transmembrane region and the cytoplasmic tail in all class II α-chains are coded

by exon 4; in contrast, the transmembrane region and the cytoplasmic tail of the β-chains are coded by exons 4 and 5, respectively.

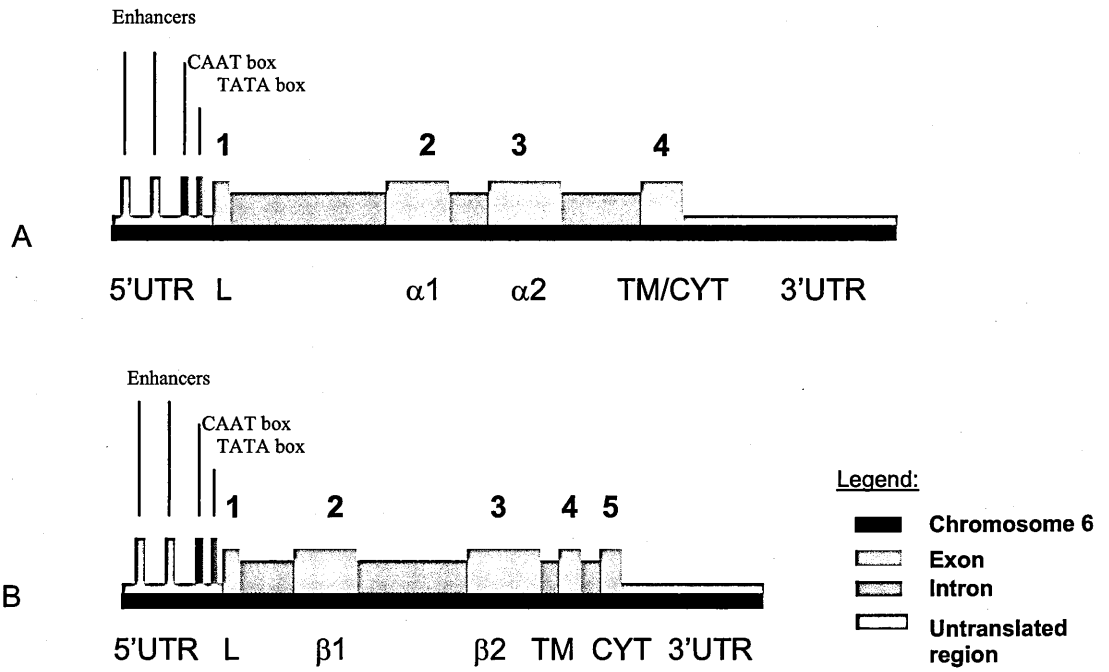


Figure 1.12: The HLA class II gene organisation

Most of the polymorphisms observed in HLA class II molecules accumulate around the peptide-binding groove especially on the membrane-distal extracellular domains (α1 and β1). The α-chain genes of certain class II isoforms are monomorphic or oligomorphic (with only a few variants), e.g. *HLA-DRA, -DPA1* and *-DQA1*, while in most cases the β-chain genes are highly polymorphic.

The number of known *HLA* class II alleles in the HLA database is 1154 (**Table 1.3**), of which 814 are alleles of *DRB* genes[94]. There are 9 different *DRB* genes (*DRB1-DRB9*) but *DRB1* is the most polymorphic of all with 722 alleles representing 89% of alleles encoded by the *DRB* family.

HLA nomenclature

The WHO Nomenclature Committee for factors of the HLA system is responsible for assigning and publishing names of new *HLA* alleles[102]. Since 1987, each *HLA* allele has been uniquely indentified using a four-, six- or eight-digit number. The length of the allele designation depends on the sequence of the allele and that of its closest relative. The first two digits represent the type (serological antigen); the third and fourth digits correspond to the subtypes and are given serially based on the order of discovery. Alleles whose numbers differ in the first four digits have one or more nucleotide substitutions that change the amino acid sequence of the protein they encode. The fifth and sixth digits are used to distinguish between alleles that differ only by synonymous nucleotide substitutions (silent or non-coding substitutions) within the coding sequence. Furthermore, alleles that only differ by sequence polymorphisms in the intronic regions or in the 5' untranslated region (5' UTR) or in the 3' UTR are distinguished by the addition of seventh and eight digits. In addition to these unique allele identification numbers, suffixes are optionally added after the last digit to indicate the levels of expression of the protein encoded by that allele (see **Table 1.4**).

Given that the number of alleles for some *HLA* genes was fast approaching the maximum possible number that could be allocated with the current naming convention; a new naming convention has been proposed by the Nomenclature Committee for implementation in April 2010. All current *HLA* allele names will have to accommodate colons (:) to serve as field delimiters and alleles that were previously rolled over would be renamed. Examples of allele names under the new naming convention as of April 2010 will be as follows: A*01010101 will become A*01:01:01:01; A*02010102L will become A*02:01:01:02L; all other fields description remain as described above and shown on **Table 1.4** below.

Table 1.4: Nomenclature of the factors of HLA system

Nomenclature	Indicates
HLA	the HLA region and prefix for an HLA gene
HLA-DRB1	a particular HLA locus i.e. DRB1
*HLA-DRB1*13*	a group of alleles which encode the DR13 antigen
*HLA-DRB1*1301*	a specific HLA allele
*HLA-DRB1*1301N*	a null allele
*HLA-DRB1*130102*	an allele which differs by a synonymous mutation
*HLA-DRB1*13010102*	an allele which contains a mutation outside the coding region
*HLA-A*3014L*	an allele encoding a protein with significantly reduced or 'Low' cell surface expression
*HLA-A*24020102L*	an allele encoding a protein with significantly reduced or 'Low' cell surface expression, where the mutation is found outside the coding region
*HLA-B*44020102S*	an allele encoding a protein which is expressed as a 'Secreted' molecule only
*HLA-A*3211Q*	an allele which has a mutation that has previously been shown to have a significant effect on cell surface expression, but where this has not been confirmed and its expression remains 'Questionable'

HLA restriction and T cell maturation

T cell development takes place in the thymus gland where naïve T lymphocytes are educated and trained to interact with autologous (self) antigens such as HLA class I and II molecules. By the end of their educational period, each mature T cell will express either an αβ receptor made up of one α and one β chain, or the γδ receptor comprising one γ and one δ chain[103]. Only those T cells expressing the αβ-TCRs can effectively interact with the HLA-peptide complex.

A fundamental requirement for successful T cell development is the need to rearrange some of the genes necessary for the immune response in order to become functional. This process starts with the rearrangement of the β-chain genes, which when successful, initiates the rearrangement of the α-chain genes. The next step entails a positive or negative selection, which eliminates those cells that bind with high affinity to autologous class I or II allotypes (negative selection), while retaining those that bind with intermediate affinity to one class I or class II molecule (positive selection). Failure to bind or poor binding will result in termination of the maturation process and death[104].

Each αβ-T cell receptor only recognises peptides presented by one HLA class I or II allotype (the one that it was positively selected for in the thymus) and is said to be restricted to that HLA allotype. By virtue of their maturation, a given T cell clone is therefore functionally monoclonal. In addition, T cell restriction to a particular class of HLA is in part determined by other cell-surface proteins namely CD4 and CD8 molecules, which like TCRs also interact with HLA-peptide complex but at a membrane-proximal level (α3 domain on class I chain and β2 domain on class II chain, respectively). These cell surface proteins have been given the name "co-receptors". CD4 and CD8 expression on T cells is generally mutually exclusive and T cells expressing the CD4 co-receptor primarily recognise peptides

bound to HLA class II molecules while those positive for CD8 glycoprotein interact with peptides presented by HLA class I molecules[105].

HLA and antigen presentation

HLA class I and II molecules are key components of the adaptive immune system. They present short fragments derived from endogenous and exogenous sources to cytotoxic ($CD8^+$) and helper ($CD4^+$) T lymphocytes. A successful presentation of peptides to T cells occurs when the T cells are able to recognise and interact with the peptide loaded onto the HLA molecule via T cell receptors (TCRs). For this to happen, T cells must have been primed or educated during their maturation in the thymus to restrict their scope of MHC interaction to a particular class I or II molecule. The interaction with T cells could lead to the initiation of a specific cell-mediated immune response against the source of the antigenic peptide or the organism from which the peptide was derived.

Presentation through HLA class I pathway

For the class I pathway, antigen processing begins with the assembly of the heavy chain (α) with the β_2m within the lumen of the endoplasmic reticulum (ER), while protein degradation is taking place in the cytosol by proteasomes to generate antigenic peptides. The transporter protein associated with antigen processing (TAP) then moves the peptides from the cytosol to the lumen of ER where they are loaded onto class I heavy chains with the help of tapasin. Two chaperonins (calnexin and calreticulin) residing in the ER intervene with the stabilisation of newly assembled but very unstable class-I-heavy-chain/β_2m complex. Tapasin then helps with the off-loading of peptides from TAP and loading them onto the class I binding groove (one peptide per groove). When the complex (α-chain + β_2m + peptide) is stable enough, it leaves the lumen of the ER

and passes through the Golgi apparatus to get to the plasma membrane where the peptide is presented to $CD8^+$ T cells to initiate an appropriate immune response.

Presentation through HLA class II pathway

The heavy chains of HLA class II (α and β chains) molecules are also assembled in the lumen of the ER with help from chaperonins (calnexin and calreticulin) but they are not permitted to bind a peptide in this cellular compartment. During the assembly process, another polypeptide called the invariant chain (Ii) attaches itself to and occupies the class II binding groove thereby preventing any peptide loading within the lumen of the ER. This trimeric structure is stable enough to go through the Golgi body into endocytic vesicles where endocytosed or phagocytosed proteins are being degraded into small fragments (peptides). Once in the vesicles, the invariant chain is also degraded to free the binding groove but in some cases a small part of the invariant chain still remains. This small piece is called the class II associated invariant chain peptide (CLIP). When this happens, HLA-DM catalyses the release of CLIP and replaces it with a single peptide resulting from the degradation of extracellular proteins from endocytosed or phagocytosed organisms or non-self cells. The stable complex (class II molecule + peptide) leaves the endocytic vesicle also known as the MHC class II compartment (MIIC) to the cell membrane where the antigenic peptide is presented to $CD4^+$ T cells to stimulate and initiate a cascade of host immune response.

HLA polymorphism

The HLA system remains the most polymorphic genetic system ever studied, even though some of its non-classical loci are monomorphic or oligomorphic. Recent technologies (DNA-typing methods) such as sequence-based techniques and crystallography have provided fine details of the *HLA* genes

and molecules, and made the comparison of allele sequences from polymorphic loci possible. Furthermore, they have helped to demonstrate that nucleotide substitutions in polymorphic class I and II genes are concentrated in exons that encode the peptide binding groove, and sites that interact directly with TCRs on the membrane-proximal domains of the α-chain (class I) and β2-chain (class II).

The mechanisms that give rise to polymorphic variants in the HLA genes have progressively become clearer over the years. Natural selection combined with point mutation, gene conversion and recombination has contributed to the high level of diversity observed in some *HLA* class I and class II genes[106]. It is also believed that past exposures to infectious diseases by our ancestors have helped in selecting for variants conferring a survival advantage. A number of important scientific observations suggest that HLA allotypes are associated with functional differences that can be explained by differential natural selection[107].

Generation and pattern of MHC diversity

Maximum overall heterozygosity for HLA determines the number of antigens an individual's T cell population can respond to and is likely to give greatest selective advantage[108]. One of the characteristics of the HLA complex is the lack of a "wild type" phenotype probably because of the extensive number of known alleles resulting from polymorphisms at the MHC region[106]. It has been demonstrated that *HLA* genes generate their diversity through positive selection of alleles with coding mutations in key areas resulting in differences in the amino acid sequences of the peptide-binding site[109].

Phylogenetic analyses of HLA sequences from different populations around the world have shown that the HLA region exhibits greater sequence divergence between loci than is observed in any other part of the human genome[109]. In keeping with this, the number of non-synonymous substitutions in the exons of

HLA-B locus is by far the highest compared to those observed in HLA-A and -C loci, respectively. The mechanisms for the generation of distinct allelic forms within and between HLA loci are complex and have been a subject of debate in the scientific community for many years.

Gene duplication

Gene duplication is a hallmark of the MHC region and results in the formation of gene clusters. In higher primates, the MHC class I and class II genes undergo periodic expansion and contraction presumably in response to the changing range of pathogens. It is believed that gene duplication has played a significant role in the generation of polymorphisms in the HLA regions leading to the multiplicity of both functional and non-functional *HLA* genes. This is evident by the fact that groups of related genes and some times remnants of genes are found throughout the MHC region particularly in the class I region where a number of pseudogenes are found[109].

It has been shown that about 5% of human genome can be attributed to segmental duplication - a common feature of the human genome that is not limited only to multi-gene families such as those of the human MHC that are involved in immune regulation[110].

Point mutation

Point mutation is the introduction of a unique substitution within an allele. It may involve the loss of a nucleotide or the insertion of an additional nucleotide, or the substitution of one nucleotide for another. In combination with other forms of substitutions, point mutation has greatly contributed in HLA diversity[109].

Gene conversion and recombination

Gene conversion is another form of genetic recombination during which there is a non-reciprocal transfer of sequence motif from one allele to another at the same locus during which the transferred motif completely replaces the corresponding motif in the recipient allele[109]. Studies on germ-line mutations have suggested that inter-allelic gene conversions do contribute immensely to MHC polymorphism[111]. However, sequence comparison analyses have indicated that inter-locus conversion does not contribute substantially to the overall generation of diversity in the HLA region[109]. Multiple recombination or crossing-over events can occur at different sites between homologous chromosomes resulting in the exchange of DNA segments of varying length.

HLA haplotypes

A haplotype is a combination of alleles encoded by different loci of the same chromosome and within a certain chromosomal region. The resultant effect of HLA polymorphisms on an individual's ability to elicit an effective immune response against a given antigen is determined by the individual's phenotype, which is the combination of gene products contributed by two human chromosomes 6. The mechanisms for the generation of different haplotypes in HLA class I and HLA class II are similar and are likely to have been by gene duplication and recombination[109]. The presence of multiple *Alu* sequences (a marker for the origin of different genes) particularly within the class II genes indicates that most recombination events occurred through unequal crossing-over with gene misalignment leading to increased variability within the region[112]. Certain HLA haplotypes may predominate in particular populations due to advantageous recombination of alleles on those specific haplotypes.

HLA and HIV infection and disease progression

Data demonstrating *in vivo* T cell-induced protective immunity to HIV are accumulating from cohort and large population-based studies around the world[73] [113 114]. This function depends on the ability of the individual's HLA molecules to present antigenic peptides more effectively to $CD8^+$ and $CD4^+$ T cells, a biological function that can be altered by genetic polymorphisms capable of generating structural changes in peptide-binding grooves.

HLA association with HIV transmission

Many large studies from Western populations and cohorts focusing on identification of the genetic determinants of susceptibility to HIV infection have implicated certain *HLA* class I genes with favourable or unfavourable HIV-1 disease outcomes[68-70 74 115-118]. There is also considerable evidence, from a small number of studies targeting individuals who have been exposed to HIV but did not seroconvert after a long period of time, suggesting that natural immunity may protect against HIV acquisition and that HLA-restricted CTLs may be the mediator of such protection[119]. Some of the alleles that have been shown to influence HIV-1 transmissibility include *HLA-A*02*[114 120] with low transmissibility and *HLA-A*2301, -A*29,-Cw*07,* and - *Cw*08* with high transmission[114 120].

HLA concordance between mother-child pair has also been shown to be a risk factor for HIV-1 transmission[121 122]. In another study, Winchester and co-workers demonstrated that protection against transmission of HIV-1 to infants born to HIV-1 positive mothers was associated with the class II genes *HLA-DRB1*1501* and *HLA-DRB1*13*[123].

Comparative analyses using data from West Africa have shown that transmissibility is slow in HIV-2 compared to HIV-1[124]. Some have suggested that HIV-2 protects against HIV-1 acquisition but this issue remains controversial as

data from Senegal, The Gambia and elsewhere in West Africa are not consistent[125-127].

A number of *HLA* alleles have been found to increase substantially the susceptibility to HIV-1 in different epidemiological studies in Africa[114], Argentina[128], America[129], and Australia[130]; while others were associated with resistance to HIV infection. For example in The Gambia, *HLA-B*35* has been linked to resistance to HIV in a cohort of HIV-exposed but uninfected commercial sex workers (CSW) who demonstrated a strong B35-restricted CTL response to both HIV-1 and HIV-2 peptide epitopes[73]. In a similar risk group in Nairobi, *HLA class II* (*HLA-DRB1*01*) was found to mediate resistance to seroconversion observed in exposed but uninfected CSW in that part of East Africa.

HLA association with HIV disease progression

Homozygosity at any of the HLA loci decreases the number of epitopes that can be seen or targeted by host CTLs. This appears to be a selective advantage to a virus like HIV with a high replicative capacity. HLA homozygosity has been associated in a dose dependent manner with progression to AIDS following HIV-1 infection in several epidemiological studies[69 118 131]. Specifically, homozygosity for *HLA-B Bw4* is associated with slower rate of decline in $CD4^+$ T cells[132]. In other cohort studies, *HLA-B*35* has consistently been implicated with rapid HIV disease progression in a variety of risk groups[69 115 116 133 134]. The effect of *B*35* seems to be co-dominant and subtype dependent[69 133]. Other *HLA* class I and II genes that have been marginally linked to rapid progression to AIDS include *A*23, A*24, B*08, DRB3,* and *DRB5* genes and the ancestral haplotype *A1-B8-DR3*[116 130 131 135]

On the other hand, several alleles have been found to associate with protection against progression to AIDS and include *A*25, A*26, A*68, B*27, B*57,* and *DRB6*. However, the most consistent findings have been those associating

*B*27* and *B*57* with protection against rapid HIV-1 disease progression. *HLA-B*27* and *-B*57* are rare alleles in most populations but their effect on slowing HIV-1 disease progression has been consistently proven[116 131 133-136]. Heterozygosity at several of HLA class I loci is advantageous against rapid development of AIDS[69 118] and high HIV viral load[137] in Caucasians.

In the past 20 years, most of the HLA associations with HIV disease outcomes resulted from candidate gene studies in which, suspected gene variants were analysed to discover their role in either the pathogenesis or immune response to HIV[78 79 84 138]. In recent past, new genome-wide approaches have been used in several studies (GWAS) to globally assess the influence of the host genome on HIV disease outcomes including viral load set point and progression to AIDS[139 140]. These studies have provided strong functional data supporting the central role of *HLA* genes in restricting HIV replication and have open up new perspectives in the understanding of the effect of human genetic variation on the individual response to HIV infection and disease progression.

Together, these data have provided valuable insights into the mechanisms of protective immunity against HIV infection, which still remain poorly understood. More recent insights have come from studies associating HIV disease outcome with genotypes of KIRs.

Killer cell immunoglobulin-like receptors (KIR)

Given that both the humoral and adaptive immune responses are relatively ineffective against HIV when compared to other pathogens, innate immune mechanisms may play a particularly important role in governing susceptibility to infection and rate of disease progression. This may also result from the ability of several HIV genes to down regulate HLA expression, thereby reducing T-cell recognition of infected cells. HIV-1 disease tends to progress rapidly and very few people are able to control the infection for any length of time. In contrast HIV-2 infection is characterised by stable viral load over the course of many years. Thus HIV-2 provides a good model for examining the mechanisms by which the immune system is able to control these viruses, potentially identifying better methods of influencing the natural history of HIV infection and informing vaccine development.

Overview of NK cells

Natural killer (NK) cells are essential components of the innate immune system, which serve as the first line of defence against invading pathogens including viruses. Morphologically, NK cells are a heterogeneous population of large granular lymphocytes representing 10-15% of circulating lymphocytes that contribute to the immune response against infection[141]. Derived from the bone marrow, NK cells share a common progenitor with T cells (key effector cells of the adaptive immune system), but unlike the latter they have the spontaneous ability to kill "abnormal" cells, which are recognised as 'non-self' without prior sensitisation – hence the name "natural killer". They further differ from T and B cells in that they are thought not to express antigen-specific cell surface receptors, although recent evidence suggests that at least in some cases the interaction between NK cells and target cells may be peptide specific[142]. Genes encoding NK

cell surface receptors are thought not to undergo clonal germ-line rearrangement as is common in other lymphocytes.

Phenotypically, mature NK cells do not express CD3 or CD4 molecules on their cell membranes. Some, however, do express the α/α form of CD8 molecules in addition to the common CD16 and CD56 markers that are expressed by all NK cells. Functionally, NK cells destroy their targets using a range of different mechanisms including: (1) secretion of regulatory cytokines including IFN-γ, TNF-α, TGF-β, IL-1, IL-5, IL-10, and chemokines such as RANTES, MIP-1α and MIP-1β[143-145] to recruit other cells to the site of infection for concerted action against the pathogen; and (2) direct killing through antibody-dependent cellular cytotoxicity (ADCC). These activities are tightly regulated by a group of receptors collectively known as the "natural killer cell receptors" (NKCR) encoded by specific genes located on human chromosome 12 and within the leukocyte receptor complex (LRC) on human chromosome 19.

NK cell recognition and the "missing self hypothesis"

By nature of their education, NK cells are programmed to recognise and kill specific target cells (unhealthy non-self cells) and to produce a wide range of chemokines and cytokines. NK cell recognition differs from that of other immune cells in that NK cells must make no mistake in differentiating between "self" and "non-self" (altered self). A normal and healthy cell (self) is usually not attacked by NK cells but as soon as that cell gets infected or starts to undergo some transformation, NK cells immediately sense the change from self to non-self and destroy it. NK cells must therefore have the required receptors on their cell surface to enable them distinguish between self and non-self.

Unlike T cells, which are activated through the recognition of antigenic peptides presented by MHC proteins on cell surfaces, NK cells become functional

by sensing the lost of MHC class I molecules on cell surfaces; a process termed the "missing self hypothesis". NK cells are constantly probing other cell surfaces for the presence of their ligands. If the inhibitory KIR receptor on the NK cell surface comes in contact with its corresponding ligand (HLA class I molecules), NK cell cytolytic activity is inhibited but if the expression of the ligand is absence or reduced, NK cells become activated and destroy the target[146]. Their cytolytic activity is tightly controlled by an equilibrium between the activating and inhibitory signals mediated by their respective receptors expressed on the NK cell surface (**Figure 1.13**).

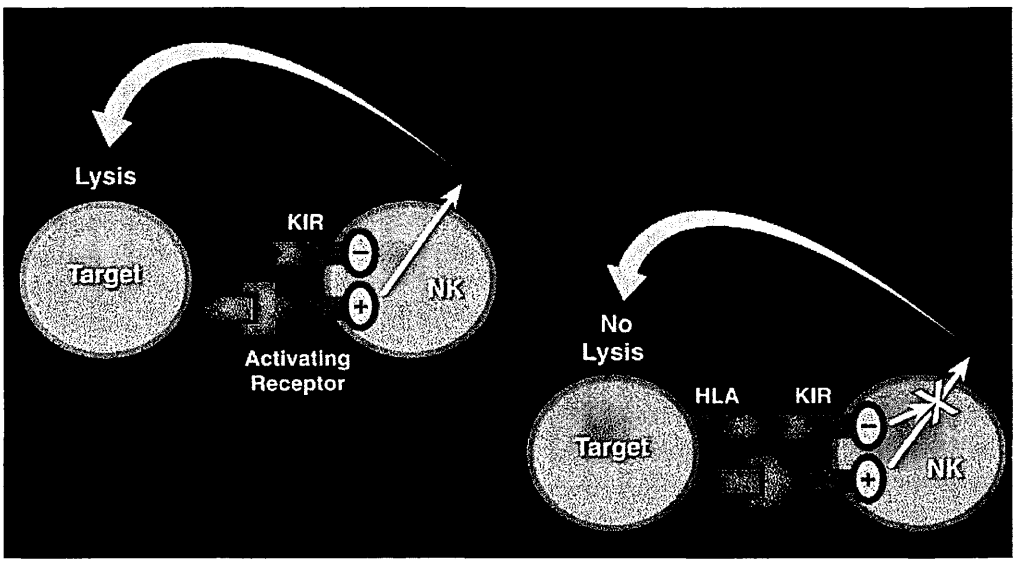


Figure 1.13: KIR-HLA interaction and modulation of NK cell cytolytic activity.
NK cells are licensed to kill targets with altered MHC class I expression (Carrington M, personal communication).

NK cell receptors in humans

A variety of membrane-bound receptors are found on NK cell surfaces belonging to two main families; namely, the Immunoglobulin Superfamily and the C-type lectin family. Genes encoding the former include the *leukocyte immunoglobulin-like receptors (LILR)*, *leukocyte-associated inhibitory receptors (LAIR)*, *leukocyte immunoglobulin-like receptors (LIR)*, *Killer cell immunoglobulin-like receptors (KIR)*, *FcαR,* and *NKp46* among others mapping to the Leukocyte

Receptor Complex (LRC) on human chromosome 19 (19q13.4) (**Figure 1.14**). Genes encoding for the C-type lectin family are located at the centromeric end of human chromosome 12 and include the *CD94* (also known as *NKG2*) gene.

KIR gene family

In humans, KIR molecules are encoded by a diverse and compact set of genes located within the LRC on chromosome 19. The family comprises 15 functional genes (*KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DL5A, KIR2DL5B, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DL1, KIR3DL2, KIR3DL3* and *KIR3DS1*) and 2 pseudogenes (*KIR2DP1* and *KIR3DP1*). The expression and function of each of these genes influences the expression and function of other members of the gene family[147]. KIR molecules are glycoproteins that belong to the immunoglobulin superfamily and expressed as receptors on the surfaces of a subset of lymphoid cells (NK cells), a subpopulation of γδ T cells and some memory αβ T cells[148 149].

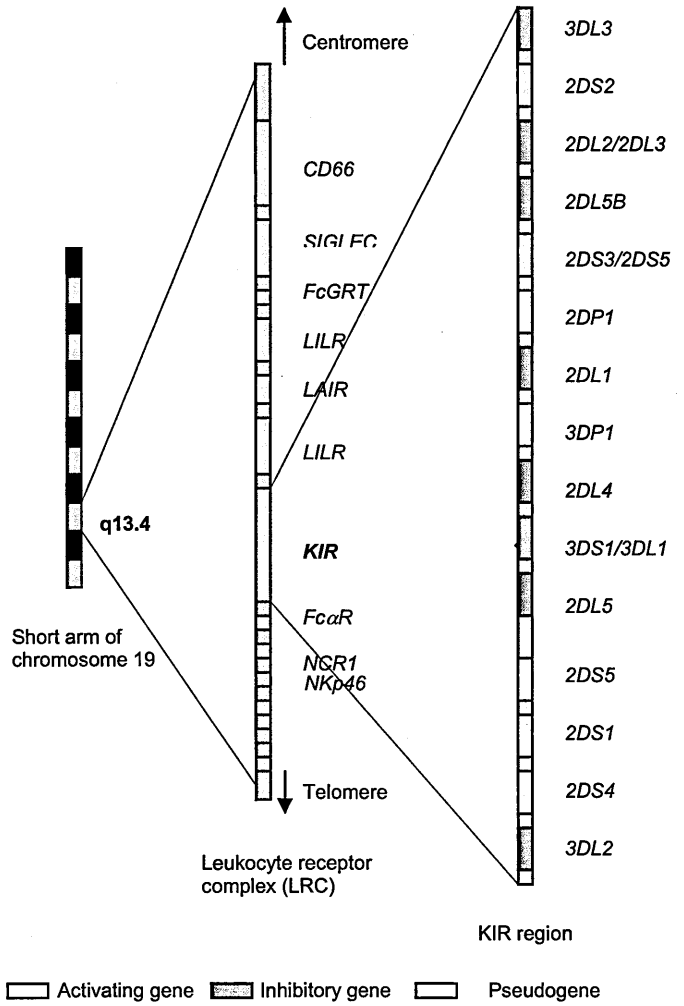


Figure 1.14: KIR genes within the Leukocyte Receptor Complex.

KIR molecules are structurally similar but can be divided on functional grounds into activating and inhibitory receptors based on the number of extracellular domains, the length of the cytoplasmic tail and the composition of the transmembrane region (**Figure 1.15**). Activating KIR molecules have two or three extracellular domains, a short (S) cytoplasmic tail with a positively charged residue in the transmembrane region, while the inhibitory KIR molecules have two or three extracellular domains, a long (L) cytoplasmic tail containing at least one immuno-tyrosine-based inhibitory motifs (ITIM)[150]. Those with two extracellular domains can further be classified as type I if they possess the D1 domain or type II if they

instead have the membrane distal domain (D0) in addition to the common D2 possessed by all KIR molecules.

The role the human KIR is to survey surfaces of other cells for the expression of appropriate human leukocyte antigens (HLA) class I molecules, which serve as their ligands. In the presence of intact ligands for the inhibitory receptors, NK cell-mediated cytolysis is inhibited whereas their absence or down regulation is associated with NK cell activation[151].

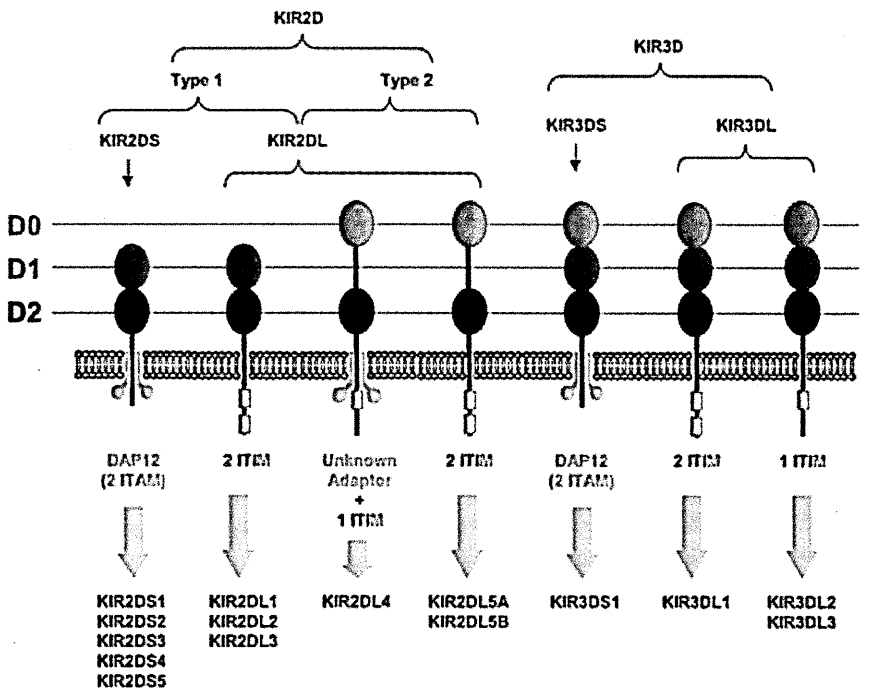


Figure 1.15: Structural characteristics of KIR gene products
Activating molecules have short cytoplasmic tails (S) and a positively charged residue that allows them to interact with an adapter molecule (DAP12), while the inhibitory receptors have long cytoplasmic tails (L) with one or two immuno tyrosine-based inhibitory motifs (ITIM). (Reproduced from ref[152])

KIR Nomenclature

Since 2002, naming of KIR genes is the responsibility of the HUGO Genome Nomenclature Committee (HGNC). KIR genes are named taking into accounts the structures of the molecules they encode. Each name then has five major divisions: the acronym KIR stands for "Killer-cell Immunoglobulin-like Receptor"; the first digit following the acronym corresponds to the number of Ig-like domains in the molecule encoded by that gene; the "D" is an abbreviation for "domain"; the D is either followed by an 'L' indicating that the protein encoded by that gene has a "Long" cytoplasmic tail, or an "S" indicating that the protein encoded by that gene has a "Short" cytoplasmic tail, or a "P" an abbreviation for pseudogenes. The last digit indicates the order in which the genes were discovered. For example KIR2DL2 is a KIR gene that encodes receptors having two extracellular Ig-like domains and a long cytoplasmic tail and the second of this type of 2DL family. In addition, when two or more genes have very similar structures and their sequences are not that much different, they are given the same last digit number but distinguished by a final letter (e.g. KIR2DL5A and KIR2DL5B genes).

KIR alleles are currently named in an analogous fashion as described for HLA above. Briefly, after the gene name (e.g. KIR2DL2) an asterisk is used as a name delimiter before a numerical allele designation is added. The first three digits following the asterisk are used to indicate alleles that differ in the sequences of their encoded proteins. Two other digits are added to distinguish alleles that only differ by synonymous (non-coding) differences within the coding region (exon). Two other digits are added to distinguish alleles that only differ by nucleotide substitutions in non-coding regions (e.g. intron, promoter, and others). An example of a KIR allele is KIR2DL2*0030101.

KIR gene organization

All KIR genes are located within a 150 kb segment of human DNA on chromosome19q13.4. Although the region is extremely variable and has gone though several episodes of expansion and contraction due to gene duplication and unequal crossing-over[153], KIR genes are arranged in a tightly organised manner within the region in a tail-to-tail fashion[154]. Each of the 16 KIR genes is about 10-16 kb long and each pair of genes is separated by about 2 kb except for the pair KIR2DL4 – KIR3DP1, which is separated by a 14 kb space.

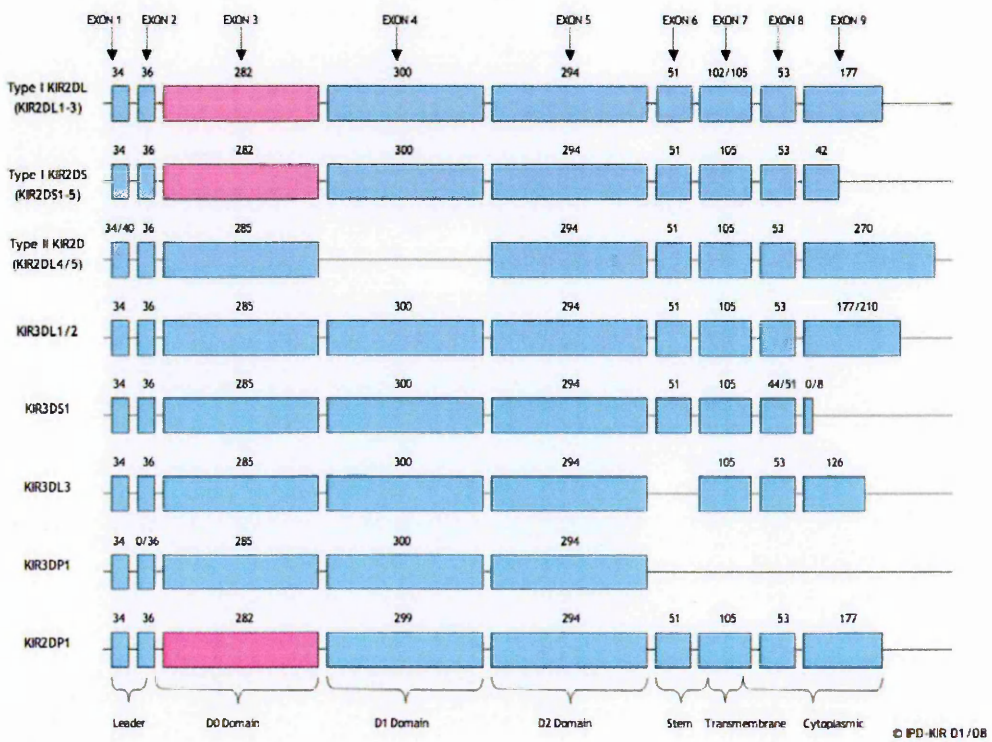


Figure 1.16: KIR genomic organisation
(Reproduced from ref. 152)

KIR genomic organisation is remarkably consistent across all genes. The basic arrangement is shown in **Figure 1.16**. Briefly the signal/leader sequence is encoded by exons 1 and 2; the 3 domains are encoded by exons 3, 4, and 5, respectively; the linker/stem and the transmembrane regions by exons 6 and 7, respectively; and finally the cytoplasmic tail is encoded by exons 8 and 9. There is

however, a pseudoexon 3 in all type I 2D *KIR* genes that may have arisen as a result of a three-base-pair deletion[155] and as a result, all type I 2D genes encode for proteins lacking the D0 domain. Similarly, the lack of D1 in all type II 2D molecules resulted from the complete absence of exon 4 in type II KIR2D genes[156]. These abnormalities (possessing a pseudoexon or missing an exon) are not limited to type II KIRs and are also found in other KIR genes and pseudogenes.

KIR polymorphism and diversity

Within the LRC, the 150 kb region that accommodates KIR genes is one of the most polymorphic regions of the human genome[155]. The latest release from the Immuno Polymorphism Database – a centralised repository for human KIR sequences (IPD/KIR)[157], shows 335 KIR alleles with KIR3DL3, KIR3DL1, KIR3DL2, KIR2DL1 and KIR2DL4 being the most polymorphic (55, 52, 45, 25 and 25 alleles, respectively).

KIR haplotypes can be classified into 2 broad categories (A and B) based on the type and number of genes (activating or inhibitory) that constitute the haplotype (**Figure 1.17**). The group A haplotypes have a fixed gene content of seven expressed genes and two pseudogenes. These haplotypes strongly contribute to the polymorphic nature of the *KIR* gene complex as their genes are highly diversified and polymorphic. In contrast, the group B haplotypes are less polymorphic but highly polygenic. Group B gene contents vary greatly and contain several other genes and allelic forms that are absent in group A haplotypes. Another structural difference between A and B haplotypes is that the latter comprise a mixture of both activating and inhibitory *KIRs* while the former lack all activating genes with the exception of *KIR2DS4,* which is ubiquitously expressed by every NK cell.

The frequency distribution of group A and B haplotypes varies from one population to another. For example, they are evenly distributed in Caucasians[158], however in Japanese group A haplotypes predominate[159] while the predominant group in Aboriginal Australians is B[160]. Overall, the group A haplotype is by far the most common haplotype in most studied populations[154].

The number and type of haplotypes present in an individual constitute his/her KIR gene profile. Profile A (i.e. two copies of "A" haplotype in the same individual) is relatively homogeneous in terms of gene content and comprises mainly genes encoding inhibitory receptors (except *KIR2DL2)*[161] with *KIR2DS4*. On the other hand, Profile B (which can either be AB or BB haplotypes) exhibits substantial variation in gene content and contains several activating and inhibitory genes[155 162]. Some *KIR* genes are common to both profiles, for example *KIR3DL3*, *KIR3DP1, KIR2DL4*, and *KIR3DL2* - the so called "framework genes".

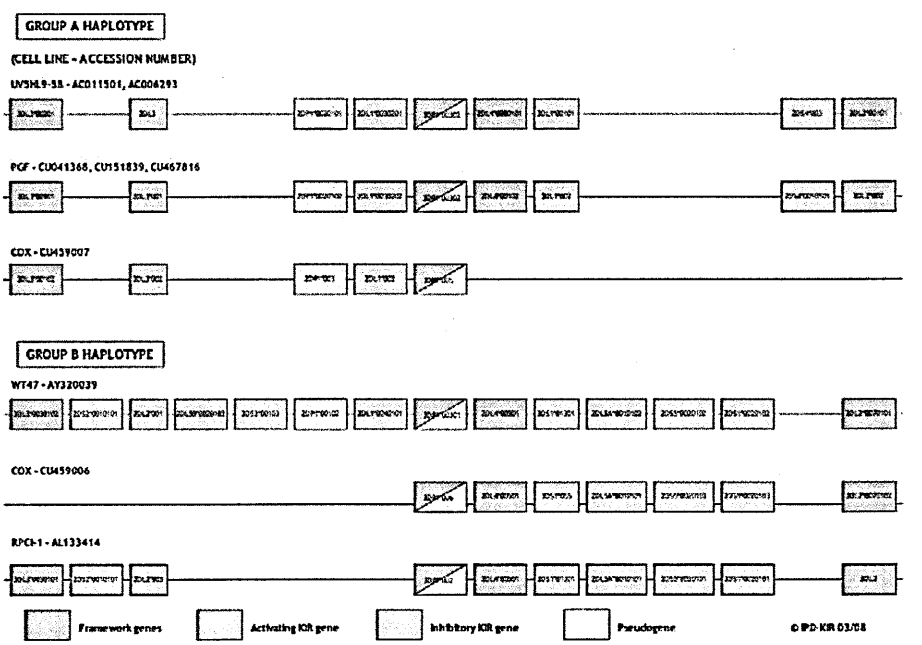


Figure 1.17: KIR haplotypes A and B
(Reproduce from ref. 152)

KIR evolution and rapid expansion

The high level of homology in both coding and non-coding regions of KIR genes is a consequence of their recent evolution and rapid expansion in numbers. This is evidenced by the observation that most of the KIR-associated *Alu* sequences are of the S subclass (a more recent evolutionary subclass)[163]. Furthermore, *KIR* orthologues are yet to be found in rodents but the receptors have been found in Chimpanzee[164]. *KIR* and *Ly49* in humans and mice, respectively, are structurally unrelated and have independently evolved to be used as variable NK cell receptors for polymorphic MHC class I[147].

Several investigators have demonstrated that the human *KIR* gene complex and mouse *Ly49* gene complex are functional analogues[163 165]. Although genes of these complexes encode distinct protein families - *Ly49* encodes C-type lectin-related proteins while *KIR* encodes Ig-related proteins, they have many features in common. For examples, some *KIR* as well as *Ly49* genes code for inhibitory receptors that use MHC class I molecules as ligands. These inhibitory receptors all have different numbers of ITIMs within their cytoplasmic domain to help them mediate their inhibitory activities. Both *KIR* and *Ly49* families contain pairs of inhibitory and stimulatory receptors with high degree of nucleotide homology in their extracellular domains.

Fahlen and colleagues have demonstrated that the acquisition of Ly49 receptors by mouse NK cells is informed by interactions with cognate MHC class I ligands expressed on bone marrow cells, which reduces the frequency of receptor expressing cells[166]. Comparative studies between different populations and other species including mice have shown that some interactions between NK cell receptors and MHC class I are highly conserved while others are highly diverse and rapidly evolving under natural selection[147]. It is believed that over millions of years, natural selection, imposed by adaptations to new and short-lived pathologic

challenges from microorganisms with high mutation rates, has provided some selective advantages leading to rapid evolution of modern day polymorphic KIR and MHC genes[167].

KIR genes and diseases

HLA and *KIR* interactions have recently been documented to have a major effect in modulating the host immune responses to pathogens. A number of recent studies have demonstrated that different *HLA–KIR* combinations, involving activating and inhibitory *KIRs,* impact on many diseases outcomes[168-175]. For example, the *KIR2DS2* gene associates with diabetes in people with a subset of *HLA-C* having asparagine at position 80 (C1)[170]. The combination between *HLA-C* and *KIR2DL1* in the absence of the corresponding activating gene is associated with pre-eclampsia[173]. In hepatitis C research, homozygosity for both *HLA-C1* and *KIR2DL3* is associated with resolution of HCV infection[171 176]. *KIR2DS1* and *KIR2DS2* genes in the absence of *HLA-C* gene that encodes for a ligand for the corresponding inhibitory receptors confer susceptibility to psoriatic arthritis[177]. In HIV/AIDS research, an epistatic interaction between *KIR3DS1* and a subset of *HLA-B Bw4* alleles has been shown to delay progression to AIDS in HIV-1 infected individuals[169].

A number of viruses down regulate HLA class I expression from the surfaces of cells that they infect as soon as they gain entrance into and cease the host cell's machinery. However, this leaves the cells open to attacks by NK cells. In trying to avoid NK cell attacks, some viruses target only specific HLA alleles for down regulation. For example, it has been reported that HIV selectively downregulates HLA-A and HLA-B but not HLA-C or HLA-E, which are major ligands for a significant number of NK cell receptors including KIR[178].

Rationale

The burden of HIV epidemic rests on developing countries, most of which are in sub Saharan Africa. Since its first report in 1986, HIV-2 remains largely confined to West Africa[3] where it coexists with HIV-1 and other pathogens. HIV-2 prevalence is high, reaching 20% among adults in Guinea Bissau. Since most people who are infected with HIV-2 control the virus well for many years this provides an excellent opportunity to investigate the mechanisms by which the immune system prevents disease progression. This information is likely to shed light on potential mechanisms to combat the more aggressive HIV-1 infection. A preventive vaccine that would curtail the spread of the HIV virus remains elusive. The main barrier to the development of an effective vaccine against HIV infection is a lack of understanding of the mechanisms that would confer protective immunity.

Immunogenetic studies in Western cohorts have provided considerable insights into potential mechanisms, for example by highlighting associations of particular HLA molecules and/or KIR genotypes with good or bad outcome in HIV-1 infection. These studies have strongly implicated aspects of the $CD8^+$ and $CD4^+$ T cells and NK cell responses as important in the control of HIV replication. Most current candidate HIV vaccines aim to stimulate a strong cellular immune response against the virus, for which knowledge of the most common HLA molecules and the epitopes they present is essential for both vaccine design and immune monitoring in vaccine trials. However, there is a paucity of information about the distribution of HLA molecules in vaccine target populations, particularly in Africa.

This study was designed to determine the distribution and nature of HLA and KIR alleles, genotypes and gene profiles in two West African populations in

The Gambia and Guinea Bissau. The effects of different KIR/HLA combinations on susceptibility to HIV-2 infection and disease progression using recognised markers such as the rate of $CD4^+$ T cell decline and viral load were also studied.

Hypotheses

Main hypothesis

Specific *HLA* and *KIR* genes/alleles and/or *KIR* gene profiles associate with susceptibility or protection to HIV-2 infection and progression to AIDS assessed by decrease in $CD4^+$ T cell counts and/or increase in HIV-2 viral load.

Secondary hypotheses

1. *KIR* and *HLA* frequencies in our study populations are similar to other populations in West Africa and other parts of the world.
2. *KIR-HLA* combinations associated with NK cell activation rather than inhibition are related to reduced susceptibility to HIV-2 and slower disease progression.
3. Homozygosity at one or more HLA class I loci is associated with susceptibility to HIV-2 infection and rapid progression in a dose dependent manner with increasing homozygosity.

CHAPTER 2 : MATERIALS AND METHODS

This chapter describes the cohorts and techniques used for these studies. The Caio cohort is a community cohort established 20 years ago, which has had repeated follow up surveys. The cohort comprises an homogenous population (>95% Manjako) derived from a collection of 10 small isolated villages. The original aim of founders of this cohort was to study the natural history of HIV-2 infection in a community with one of the highest prevalences of HIV-2 infection in the world. However a few years later, HIV-1 was introduced in that community and is now overtaking HIV-2 in prevalence. Anti-retroviral therapy (ART) was not available in Caio until April 2007.

The Fajara HIV cohort is a clinical cohort established by the Medical Research Council (MRC) (UK) Laboratories, The Gambia more than 24 years ago, which is located in Fajara on the The Gambian coast in the Greater Banjul Area. Members of the cohort belong to a variety of ethnic groupings including Fula, Mandinka, Wolof, Serahuli, Aku, Jola, Serer, and Manjako. The two cohorts are among the best well characterised HIV cohorts in West Africa.

This project is one in the series of projects designed at the MRC Laboratories in The Gambia to understand futher the natural history of HIV-2 infection. In September 2005, EDCTP (European and Developing Countries Clinical Trial Parthnership) launched a call for application for two PhD training fellowships which I applied for. In March 2006, I was one of the two winners of that prestigious award and was awarded the grant in August 2006. With the help of my mentors, I then wrote three proposals for approval by the MRC SCC (Scientific Coordinating Committee) which approved the scientific merit of the project; the Joint MRC/Gambian Government Ethics Committee which approved

that there were no ethical issues with the conduct of this project; and the Life and Biological Science Board of the Open University, UK which approved that the project was worthy of consideration for an award of a doctorate degree.

Caio community cohort

This is a unique, isolated community in the Cacheu region on the coast of Northwest Guinea Bissau (**Figure 2.1**), which is geographically characterised by lowlands of mangroves, rice fields, and palm groves. The inhabitants, numbering about 10,000, belong to the Manjako tribe and have a traditional way of life and strong beliefs in earth spirits and ancestral cults[179]. In the late 1980's an observation was made by health workers in one sexually transmitted diseases (STD) Clinic in neighbouring Ziguinchor (Senegal), that the majority of the commercial sex workers visiting their clinic for routine check-ups were from the Caio sector in Guinea Bissau. At that time, the rate of HIV-2 infection among prostitutes of Bissau origin seen in clinics in Ziguinchor was 36%[180].

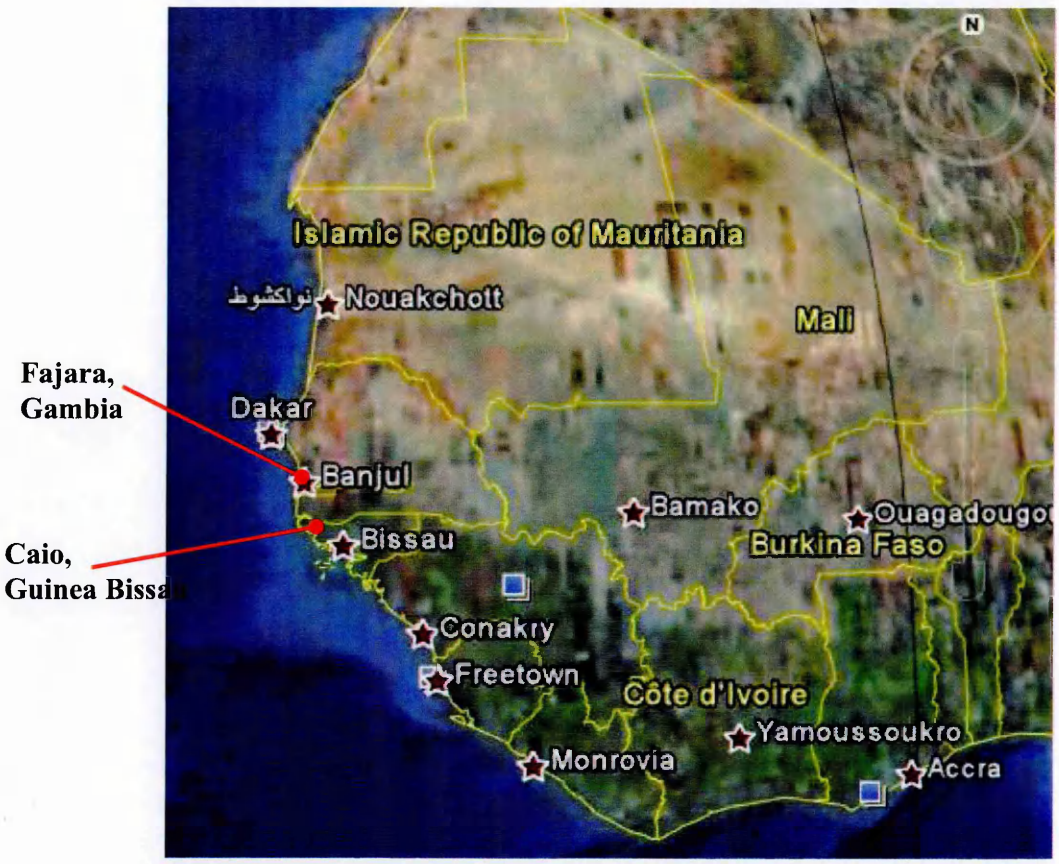


Figure 2.1 Map of West Africa showing the study sites
(Source: Google Earth)

The first survey conducted in the Caio sector revealed an HIV-2 prevalence of 8.5% among sexually active adults[181]. Since 1989, the MRC (UK) Laboratories in the Gambia in collaboration with Professor Peter Aaby's group at the "Projecto Saude Bandim" in Bissau have conducted three serological surveys in Caio. Many other studies have been done including case-control studies and extensive clinical, epidemiological and laboratory data have been gathered from this cohort to study the natural history of HIV.

Fajara HIV clinical cohort

The HIV cohort in Fajara recruits its members through the clinic and hospital of the MRC (UK) Laboratories in Fajara, The Gambia (**Figure 2.1**). Patients aged 15 years and above attending the outpatient department and

suspected of having contracted a sexually transmitted disease are referred to the Genito Urinary Medicine (GUM) clinic for counselling, and a blood sample is collected for HIV antibody status if the patient consents to testing. Those confirmed to be infected with HIV are invited to join the cohort. For the purposes of this study, those testing negative were invited to participate as controls. The patient information and consent form used for recruitment into the Fajara HIV cohort is reproduced in Appendix 1. Social history information was randomly collected on every other patient who newly joined the cohort (see Appendices 2 and 3 for male and female social history forms respectively).

Samples

I was responsible for coordinating the sample collection between 2004 and 2005 in Caio and the collection of controls from the Genitory Urinary Medicine Clinic (GUM) in Fajara. This entailed designing appropriate forms including the patient information form, consent form and sample form that accompanied each sample to the laboratory. I made more than seven trips to and from Caio to provide the site with laboratory consumables and bring back the samples to Fajara for processing. At the Caio field station blood samples were collected by the field site medical doctor who provides free medical care for the research participants and their family members or relatives. In Fajara, I was responsible for providing the GUM (Genito Urinary Medicine) clinic with the materials needed for sample collection which was perfomed by trained phlebotomists. The samples were then taken to the Human Genetic laboratory for processing and DNA extraction as described below.

A total of 1113 DNA samples extracted from HIV-2 infected and uninfected individuals were used for this study: 513 came from the Caio community cohort and 600 from the Fajara HIV clinical cohort. Of these, 750 were archival samples

from the MRC Laboratories Biobank and the remaining (363) were prospectively collected from HIV negative individuals through the same GUM clinic in Fajara from 30th October 2006 through 17th September 2007. This part of the study was performed prospectively because the Fajara clinical cohort did not initially recruit HIV negative individuals. The characteristics of the study participants are shown in **Table 2.1** and **Table 2.2** for Caio and Fajara cohorts, respectively. Individuals infected with both HIV-1 and HIV-2 viruses (dual status) were excluded from analyses because of their limited number and the potential for HIV-1 to confound our analyses.

Table 2.1: Characteristics of the study population (n=513), Caio, 2003-7

	N (%)	Median age (IQR)
HIV negative		
Male	114 (35)	36.2 (25.8-52.5)
Female	213 (65)	46.2 (30.8-62.8)
HIV-2 infected		
Male	49 (32)	53.7 (43.0-62.9)
Female	102 (68)	59.5 (49.5-69.4)
HIV-1&-2 dually infected		
Male	6 (17)	37.0 (34.7-49.3)
Female	29 (83)	52.8 (42.8-61.7)
Ethnic groups		
Manjako	479 (93.4)	
Pepel	15 (3.0)	
Mandinka	11 (2.1)	
Others	8 (1.5)	

Table 2.2: Characteristics of the Fajara study population (n= 600)

	N (%)	Median age (IQR)
HIV negative		
Male	61 (17)	31 (25-38)
Female	302 (83)	27 (22-35)
HIV-2 infected		
Male	71 (30)	44 (35-53)
Female	166 (70)	38 (30-44)
Ethnic groups (n = 493)		
Mandinka	229 (46.5)	
Fula	66 (13.4)	
Jola	67 (13.6)	
Wolof	52 (10.6)	
Serer	18 (3.6)	
Manjako	18 (3.6)	
Serahuli	13 (2.6)	
Aku	3 (0.6)	
Others	27 (5.5)	

The participant information form and the consent form used for the recruitment of the control individuals in Fajara are shown in Appendices 4 and 5, respectively. The subject information/consent form used in Caio to recruit study participants into the Caio community cohort is reproduced in Appendix 6. The social history information was collected on every participant that consented to join the study as a control. The male and female social history forms were the same as in Appendices 2 and 3 mentioned above.

HIV antibody testing by serological techniques

Each of the participants was pre-test counselled and a blood samples collected from consented individuals were placed into two containers: (1) a plain tube that allowed the blood to clot and (2) an EDTA (ethylenediaminetetraacetic acid) tube that prevented the blood from clotting. The plain tube was sent to the MRC HIV diagnostic laboratory (accompanied with a sample request form) where the serum was collected and screened for the presence or absence of HIV antibodies using a Murex ICE HIV-1.2.0 capture enzyme immunoassay - a commercial kit from Murex (Murex Diagnostics). Suspected positive or reactive sera were confirmed using a rapid differential Hexagon test (Human GmbH) - an immunochromatographic test capable of differentiating between HIV-1 and HIV-2. Weakly reactive HIV-1 or HIV-2 or suspected dually positive sera were further tested using Pepti-Lav 1-2 (Sanofi Diagnostics Pasteur) - a synthetic peptide-based strip that also discriminates between HIV-1 and HIV-2. Samples with positive HIV-1 or HIV-2 or dual (HIV-1&2) status were excluded from this round of prospective recruitment of controls and only those with negative HIV antibody test results were retained. DNA was extracted as described below. Recruitment of HIV infected individuals into the Fajara cohort is routine and achieved using the same study sites (GUM clinics in Fajara) but using a different protocol where by patients are to be pre and post-test counselled before their consent is sought to willingly join the cohort. The sample request form used for this study is reproduced in Appendix 7.

Sample collection and DNA extraction

Two millilitres of venous blood were collected into an EDTA container and sent to the MRC human genetics laboratory for DNA extraction. The samples were stored in -20 ^{o}C freezer awaiting the HIV status from the HIV diagnostic

laboratory. Only samples confirmed to be HIV seronegative were processed. Genomic DNA was extracted from each blood samples using an in-house salting-out technique. Briefly, red blood cells (RBC) were lysed using a solution of Tris-EDTA buffer (TE-20-5) pH 8.0 and 20% sodium dedocyl sulphate (SDS) solution was used to lyse white blood cells (WBC) and release the nuclear material into solution. Unwanted proteins were digested using 10 mg/ml proteinase K and 7.5 M ammonium acetate added to bind onto DNA. This complex was subsequently precipitated using ice cold absolute ethanol. TE 20-5-N_aCl solution containing 0.2 M sodium chloride was added into the mixture to replace the ammonium salt with sodium salt. The DNA-salt complex was again precipitated using ice cold absolute ethanol. The DNA was redissolved in TE 20-1, quantified and stored at -20 °C for future use. The DNA extraction protocol used for this study is described in Appendix 8.

KIR typing by PCR-SSP techniques

The techniques involved the use of 60 KIR specific primers (see list of KIR primers in Appendix 9) to amplify conserved regions of the gene of interest by PCR-SSP (sequence specific priming)[182]. Primers were synthesised by Metabion (Metabion International, Germany) as desalted and HPLC purified oligonucleotides. They were reconstituted to 100 µM in TE buffer pH 8.0 and used to prepare primer mixes as indicated on the worksheet in Appendix 10.

This technique detects 14 functional *KIR* genes (*KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DL5, KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DS5, KIR3DL1, KIR3DL2, KIR3DL3,* and *KIR3DS1*) and 1 pseudogene (*KIR2DP1*). The template was designed to contain 30 PCR reactions per sample. Two pairs of primers annealing at two different positions (exons) of the *KIR* gene of interest were used to reduce the chance of missing the presence of a gene due

to mutation at one of the primer binding sites. A pair of internal control primers was introduced into each PCR reaction (except for lane 6) to amplify a 796 bp fragment from the third intron of *HLA-DRB1* gene to check the effectiveness of the PCR reactions. The master mix contained 200 µM dNTP, 500 nM primer, 1.5 mM $MgCl_2$, 20 mM Tris-HCl, pH 8.4, 50 mM KCl, and 0.5 U of Taq polymerase (Bioline Ltd, London, UK). Amplifications were performed using a PTC-200 (tetrad) Thermal Cycler (MJ Research, Inc., MA, USA) with the following conditions: 94 °C for 2 minutes followed by 5 cycles of 94 °C for 5 seconds, 65 °C for 15 seconds, and 72 °C for 30 seconds; then 21 cycles of 94 °C for 5 seconds, 60 °C for 15 seconds, and 72 °C for 30 seconds; followed by 4 cycles of 94 °C for 5 seconds, 55 °C for 1 minute, and 72 °C for 2 minutes; and finally 1 cycle of 72 °C for 7 minutes. The PCR products were separated in a 3% agarose gel containing ethidium bromide at 220 volts for 30 minutes and gel picture taken for interpretation and analysis.

The protocol (SOPL2) used for *KIR* genotyping is reproduced in Appendix 11. In January 2008, I visited Dr. Mary Carrington's laboratory in Frederick, MD, USA for 3 months for training on gel scoring techniques. All gel pictures generated during the course of this study were independently scored by two people. The second scoring was done by a senior scientist in Dr. Carrington's laboratory at the National Cancer Institute, NCI-Frederick, USA. The scoring results were compared and any sample with discordant results re-typed. Most of the repeats were done in Dr. Carrignton's laboratory particularly to confirm that the frequency of *KIR3DS1* was correct since it has been previously reported to be low in people of African descent. A sample of the gel interpretation sheet is shown in appendix 12. The results were double entered into an Access database for analysis.

HLA class I typing by sequence-based techniques

HLA typing was done following the sequence-based method as described fully elsewhere[183]. Briefly, genomic DNA from each individual was amplified using locus-specific primers that flanked exons 2 and 3 of HLA-A, -B, and -C loci. Sequence-specific primers were then used to sequence the amplified products in both directions as illustrated in **Figure 2.2**. The list of HLA primers used in this study is reproduced in Appendix 13. Primers were synthesised, desalted and HPLC purified by Metabion (Metabion International, Germany). They were reconstituted to 100 µM in TE buffer pH 8.0.

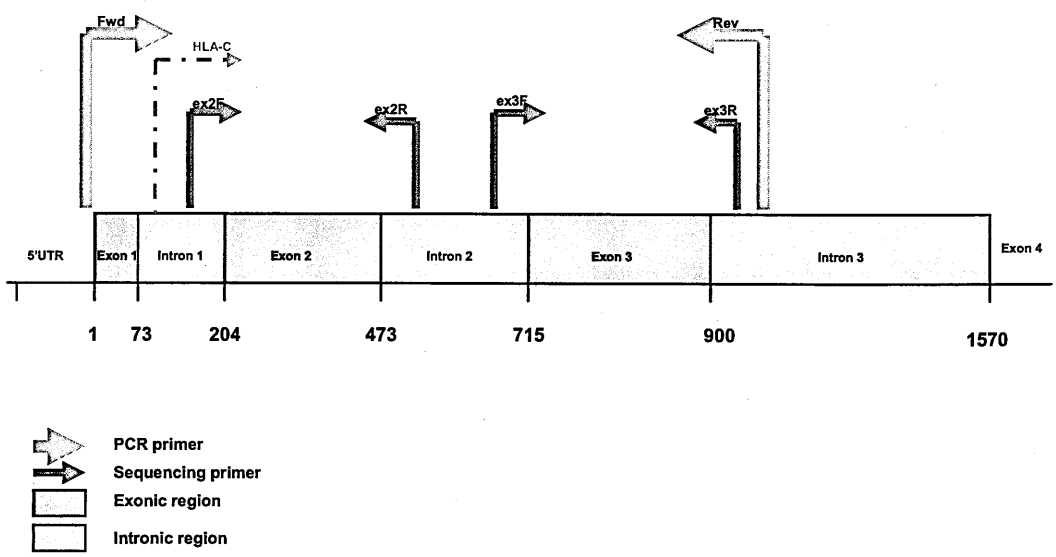


Figure 2.2 HLA class I sequencing map with primer positions

Polymerase chain reaction (PCR)

Each locus was amplified by polymerase chain reaction (PCR) in a 20 µl reaction mixture consisting of 2 µl of 10X PCR buffer, 0.4 µl of 10 mM dNTPs, 0.8 µl of 50mM $MgCl_2$, 1.0 Unit of Taq polymerase (Bioline Ltd, London, UK), 0.4 µl of

10 µM of each primer and 150 ng DNA. Amplifications were performed using a PTC-200 (tetrad) Thermal Cycler (MJ Research, Inc., MA, USA) programmed as follows: for HLA-A, 1 cycle of 96 °C for 2 minutes, 30 cycles of 96 °C for 25 seconds, 60 °C for 25 seconds, and 72 °C for 1 minute, followed by 1 cycle of 72 °C for 7 minutes and 4 °C for ever; HLA-B, 1 cycle of 96 °C for 2 minutes, 40 cycles of 96 °C for 15 seconds, 62 °C for 15 seconds, and 72 °C for 1 minute, followed by 1 cycle of 72 °C for 7 minutes and 4 °C for ever; and HLA-C, 1 cycle of 96 °C for 2 minutes, 30 cycles of 96 °C for 25 seconds, 70 °C for 25 seconds, and 72 °C for 1 minute, followed by 1 cycle of 72 °C for 7 minutes and 4 °C for ever. The efficiency of the PCR reaction and the sizes of the amplified products were checked by electrophoresing 5 µl of the products in 2% agarose gel and staining with ethidium bromide. Gels were visualized using a UV transilluminator and a gel picture taken with a Kodak gel image documentation camera (DC120). The remaining 15 µl products were purified using either Exonuclease I and Shrimp Alkaline Phosphatase (GE Healthcare UK Ltd, Buckinghamshire, England) or AMPure kit (Agentcourt Bioscience Corporation, Beverly, USA).

Sequencing reaction and interpretation

The BigDye Terminator v3.1 Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA, USA) was used to sequence the purified products in both directions following the manufacturer's instructions. Briefly, the mix contained (per reaction) 2.07 µl of 5X buffer, 0.25 µl of BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA), 5.36 µl of water, 0.32 µl of primer, and 2.0 µl of purified PCR products. The sequencing conditions were 1 cycle of 96 °C for 1 minute; followed by 30 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds, and 60 °C for 4 minutes; and finally one cycle of 4 °C on hold.

The sequencing products were purified using either Sephadex G-50 superfine (Sigma-Aldrich, Dorset, UK) in MultiScreen HV 0.45 µm plates (Millipore Corporation, Watford, UK) or CleanSEQ 96 kit (Agentcourt Bioscience Corporation, Beverly, USA). Cleaned products were analysed using either ABI 3130xl or 3730xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). In June 2008, I re-visited Dr. Mary Carrington's laboratory at the NCI-Frederick, MD, USA for another 3 months period to learn about sequence trace interpretation and analysis techniques. All sequence traces generated from this study were analysed by two independent readers using the "Assign 400" software (Conexio Genomics, Western Australia). The second reading was performed by a senior scientist in Dr Carrington's laboratory at the National Cancer Institute, NCI-Frederick, USA. Results were compared and repeats done for samples with discordant results. The HLA typing protocol is reproduced in Appendix 14.

Statistical Analysis

Data were managed using Microsoft Access. *HLA-A, HLA-B,* and *HLA-C* alleles and genotype frequencies were calculated using SPSS v16 (SPSS Inc, Chicago, USA) and STATA v9.2 (Stata Corporation, Texas, USA). SAS 9.1 (SAS Institute) was used for statistical analyses. The χ^2 test was used to examine differences in allele and genotype frequencies between Caio, Fajara and other West African populations. PROC FREQ was used to compute frequencies on individual variables. PROC LOGISTIC was used for categorical analyses to obtain odds ratios and 95% confidence intervals. PROC GLM was used for analyzing continuous variables. Both logistic and linear regression models were adjusted for age and gender.

Allele frequency was calculated as a percentage of the total number of alleles in the locus of interest and the genotype frequency was estimated based on

the number of individuals carrying the corresponding allele. The Hardy-Weinberg (HW) test of equilibrium and the two- and three-locus haplotypes estimates were performed using Arlequin v 3.11 based on the genotype data. All p-values are uncorrected for multiple comparisons and were considered significant when p was less than 0.05.

Although several comparisons (~90) were performed on these highly polymorphic loci, we only have power to detect significant changes in 32 comparisons because of the small numbers in each of the HLA groups. However, since this is the first time a thorough examination of HLA and KIR has been done in a sizeable cohort of HIV-2 infected people, we considered it best to present uncorrected p-values at this preliminary stage. It is well known that for diseases that are being investigated for the first time with regards to HLA and KIR, it is generally best not to apply corrections that can result in false negative interpretations (i.e. an allele that actually show an effect may appear not to be significant). We recommend that these preliminary findings be taken cautiously pending replication in other HIV-2 cohorts. Subsequent HLA and KIR studies on these and other HIV-2 cohorts will need to adjust for multiple comparisons and some of the methods that can be used to correct for multiplicity testing include Bonferroni test and test for false discovery rate.

The Bonferroni test is one of the simplest and most conservative approaches to correct for multiple testing when several dependent or independent statistical tests are being performed simultaneously on a set of data. This is based on the assumption that an alpha level of 0.05 might not be appropriate for a set of multiple comparisons. The Bonferroni correction then reduces the alpha value for the comparisons in such a way that the probability of having a significant p-value by chance is much more reduced. This is achieved by multiplying 0.05 by the total number of comparisons.

CHAPTER 3 : HLA CLASS I ALLELE, KIR GENE FREQUENCIES AND HIV-2 ANTIBODY STATUS IN CAIO, GUINEA BISSAU

This chapter describes the frequencies of individual genes/alleles at the *HLA-A, -B* and *-C* loci and the common HLA haplotypes found in the Caio population in Guinea Bissau. The relationship between H*LA* alleles and HIV-2 status defined by antibody levels was examined with antibody positivity implying HIV infection. A well-validated diagnostic algorithm exists and is currently in routine use at MRC laboratories, The Gambia to distinguish between HIV-1 and HIV-2 and avoid false positive results from cross-reactivity as described in **Chapter 2**. Frequencies of *KIR* genes and gene profiles are given together with their relationship to HIV-2 status. The effect of HLA/KIR combinations was also investigated where *KIR* genes were paired with their respective HLA ligands.

HLA allele frequencies in the Caio population

Although 513 individuals consented to take part in the study, a small number were excluded from allele frequency (AF) analyses because their HLA results were non-conclusive. A sample was classified as indeterminate if the sequence trace deviated from a known *HLA* allele consensus by at least one mismatch. These indeterminate samples are likely to contain new *HLA* alleles that are yet to be described. Our future plan is to characterise these alleles further by designing new primers to sequence extensively the remaining exons. *HLA-A* data were available for 492 subjects, *HLA-B* and *HLA-C* for 406 and 462 individuals, respectively.

Table 3.1 summarises the allele frequency of classical *HLA* class I (*A, B,* and *C*) loci found in this sample of Manjako population from Caio. A total of 24

HLA-A, *32 HLA-B* and 21 *HLA-C* alleles were detected. Homozygosity was low in all three loci (*HLA-A*: 0.134; *HLA-B*: 0.140; and *HLA-C*: 0.134) and decreased with increasing number of loci (HLA-AB: 0.038; HLA-AC: 0.027; HLA-BC: 0.043; and HLA-ABC: 0.019) (data not shown). Overall, *HLA* allele frequencies were similar between males and females across all three class I loci (data not shown).

Table 3.1: HLA-A, -B, and -C allele frequencies in the Caio population

HLA-A alleles (n = 984)	**Freq.**	**HLA-B alleles (n = 812)**	**Freq.**	**HLA-C alleles (n = 924)**	**Freq.**
*A*010101*	0.051	*B*0702*	0.033	*Cw*0102*	0.043
*A*0102*	0.019	***B*080101***	**0.108**	*Cw*0203*	0.002
*A*0201*	0.066	*B*130201*	0.032	***Cw*0210***	**0.095**
*A*0202*	0.050	*B*1401*	0.003	*Cw*0302*	0.079
*A*0205*	0.008	*B*140201*	0.043	*Cw*030301*	0.013
*A*030101*	0.026	***B*1503***	**0.127**	***Cw*0304***	**0.144**
A*2301	**0.173**	***B*1510***	**0.099**	***Cw*040101***	**0.154**
*A*260101*	0.078	*B*1516*	0.011	*Cw*050101*	0.030
*A*290201*	0.017	*B*1801*	0.028	*Cw*060201*	0.048
*A*3001*	0.026	***B*3501***	**0.095**	***Cw*0701***	**0.167**
*A*3002*	0.071	*B*3910*	0.004	*Cw*070201*	0.030
*A*310102*	0.002	*B*4001*	0.001	*Cw*0704*	0.001
*A*3201*	0.011	*B*4016*	0.001	*Cw*0802*	0.042
*A*3301*	0.024	*B*4102*	0.007	*Cw*0804*	0.038
A*330301	**0.186**	*B*4103*	0.001	*Cw*120301*	0.008
*A*340101*	0.001	*B*4201*	0.009	*Cw*140201*	0.014
*A*3402*	0.021	*B*4403*	0.026	*Cw*1502*	0.004
*A*3601*	0.002	*B*4410*	0.001	*Cw*1505*	0.017
*A*6601*	0.013	*B*4501*	0.001	*Cw*160101*	0.035
*A*6602*	0.002	*B*470101*	0.001	*Cw*17*[a]	0.015

*A*6801*	0.037	*B*4901*	0.067	*Cw*18*[a]	0.021
*A*680201*	0.050	*B*5001*	0.010		
*A*7401*	0.054	*B*5101*	0.016		
*A*8001*	0.010	*B*5201*	0.015		
		B*530101	**0.113**		
		*B*5601*	0.007		
		*B*5702*	0.001		
		*B*570301*	0.015		
		B*5801	**0.113**		
		*B*7801*	0.001		
		*B*7802*	0.001		
		*B*8201*	0.009		

n = total number of alleles for the locus of interest; [a]: alleles that could not be discriminated with exons 2 and 3 sequences (*Cw*17* = *Cw*1701/Cw*1702,* and *Cw*18* = *Cw*1801/Cw*1802*). In bold are common *HLA* class I alleles with a population frequency ≥ 9%.

HLA-A alleles and susceptibility to HIV-2 infection

Twenty four alleles were identified at the A locus in 492 individuals. The most frequent *HLA-A* allele was *A*330301* (18.6%), followed by *A*2301* (17.3%). Seven other alleles of this locus had their frequencies ranging from 5.0 to 7.8% (*A*010101, A*0201, A*0202, A*260101, A*3002, A*680201*, and *A*74*) (**Table 3.1**). *A*7401* and *A*7402* could not be fully discriminated with sequences accrued from exons 2 and 3 since their nucleotide and amino acid compositions are similar at these exons but differ in exon 1 by only one nucleotide at position 67 (codon 23). A*7401 has an "A" at that position while A*7402 has a "T", changing the amino acid at codon 23 from Arginine (R) in A*7401 to Tryptophan (W) in A*7402. We re-sequenced all A*74+ individuals to include exon 1. Analysis of their traces

showed that they were all A*7401. The nine most frequent alleles with a population frequency of at least 5.0% made up 78.0% of all *HLA-A* cumulative frequencies in our study population. The most diverse group was *HLA-A*02* with three different alleles (*A*0201, A*0202,* and *A*0205*). Their group cumulative frequency (12.4%) was shared almost equally between *A*0201* (5.3%) and *A*0202* (5.0%).

We also determined the population genotype frequencies (GF) of all four-digit HLA alleles found in the study population (**Figure 3.1**). A univariate analysis was performed to compare HLA-A genotype frequencies between cases (HIV-2 infected) and controls (HIV antibody negative) (**Table 3.2**).

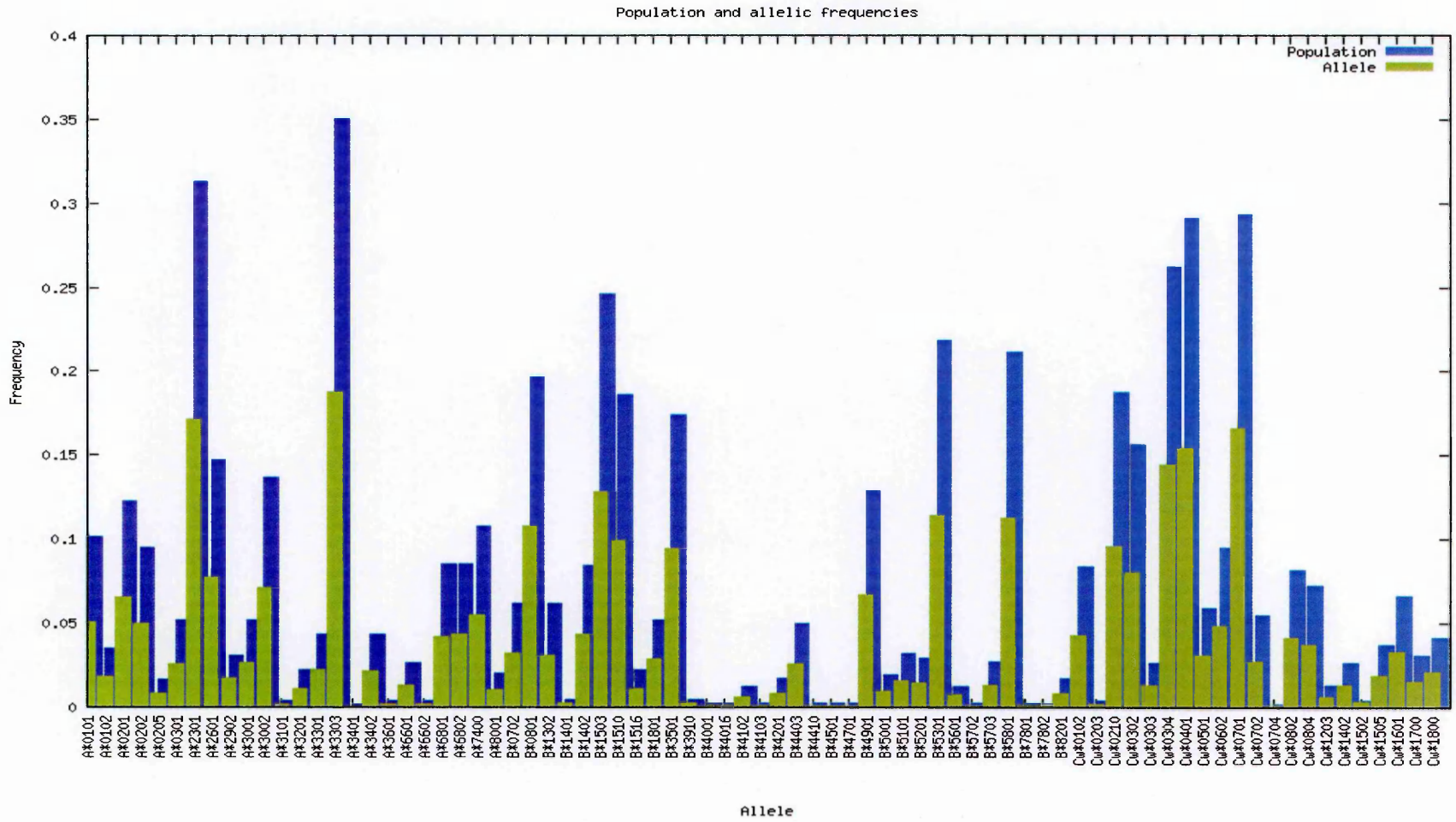


Figure 3.1: HLA class I (A, B, and C) genotype and allele frequencies in Caio

In most instances, frequencies were similar between cases (n = 149) and controls (n = 309). However, the frequency of *HLA-A*8001*, though a rare allele in this population (2.0% GF in the general population), was significantly higher in the infected group than controls (0.054 vs. 0.006, OR: 8.90 (1.86-42.53), p = 0.01). Eight out of ten (8/10) individuals carrying this allele were infected with HIV-2 and all of them were heterozygous for *A*8001*; the other allele being either *A*010101, A*260101, A*3001, A*3002, A*330301*, or *A*680201*. This suggests that individuals with *A*8001* might have an increased risk of acquiring HIV-2 compared to those without this allele.

Table 3.2: HLA-A genotypes and HIV-2 antibody status

Allele	HIV-2 (n = 149)	HIV-neg (n = 309)	OR (95% CI)	p
*A*01*	0.121 (18)	0.146 (45)	0.80 (0.44 - 1.44)	0.45
*A*02*	0.188 (28)	0.243 (75)	0.73 (0.45 - 1.18)	0.20
*A*03*	0.047 (7)	0.048 (15)	0.95 (0.38 - 2.40)	0.92
*A*23*	0.336 (50)	0.314 (97)	1.11 (0.73 - 1.68)	0.63
*A*26*	0.168 (25)	0.136 (42)	1.28 (0.75 - 2.20)	0.37
*A*29*	0.013 (2)	0.029 (9)	0.46 (0.10 - 2.15)	0.32
*A*30*	0.242 (36)	0.172 (53)	1.54 (0.95 - 2.48)	0.08
*A*32*	0.013 (2)	0.023 (7)	0.59 (0.12 - 2.86)	0.51
*A*33*	0.383 (57)	0.392 (121)	0.96 (0.64 - 1.43)	0.84
*A*34*	0.047 (7)	0.039 (12)	1.21 (0.47 - 3.15)	0.69
*A*36*	0.007 (1)	0.003 (1)	2.11 (0.13 - 34.03)	0.60
*A*66*	0.020 (3)	0.036 (11)	0.55 (0.15 - 1.99)	0.36
*A*68*	0.174 (26)	0.146 (45)	1.24 (0.73 - 2.10)	0.43

*A*74*	0.067 (10)	0.120 (37)	0.53 (0.26 - 1.10)	0.09
A*80	0.054 (8)	0.006 (2)	8.90 (1.86 - 42.53)	**0.01**

n = number of individuals, p: values after adjusting for gender and ethnicity

HLA-B alleles and HIV-2 acquisition

This predominantly Manjako population showed extensive diversity at the B locus with 32 different alleles out of the total 77 class I alleles detected (**Table 3.1**). The most common group of alleles (two-digit alleles) was *HLA-B*15* with a cumulative frequency of 23.7% distributed mostly between *B*1503* (12.7%) and *B*1510* (9.9%). *HLA-B*15* was also the group with the highest number of four-digit alleles (*B*1503, B*1510*, and *B*1516*) at the B locus. Six alleles were present in our samples at relatively high frequencies ranging from 9.5% to 12.7%. They include H*LA-B*3501, -B*1510, -B*080101, -B*530101, -B*5801,* and *-B*1503.* These alleles together with *B*140201* and *B*4901* had a cumulative frequency of 76.5% amongst all observed HLA-B frequencies in the study population.

A comparative analysis was done to see whether there was a difference in genotype frequencies between males (n = 131) and females (n = 275) and between HIV-2 infected participants (n = 143) and uninfected controls (n = 234). Individuals dually infected with HIV-1 and -2 (n = 29) were excluded from analyses because of their limited number and the potential for HIV-1 to confound our analyses. With the exception of *HLA-B*080101*, no significant difference in *HLA-B* frequencies was observed between cases and controls nor between males and females. However, univariate analysis showed that the common *HLA-B*080101* (10.8% AF and 19.6% GF at the population level (**Figure 3.1**)), was more likely to be present in the HIV-2

infected group than the control group (0.273 (n = 39) vs. 0.145 (n = 34), OR: 2.20 (1.31-3.06), p = 0.003) (**Table 3.3**).

Table 3.3: HLA-B genotypes influence on HIV-2 acquisition

Allele	HIV-2 (n = 143)	HIV-neg (n = 234)	OR (95% CI)	p
*B*07*	0.049 (7)	0.073 (17)	0.66 (0.27 - 1.63)	0.37
B*08	**0.273 (39)**	**0.145 (34)**	**2.20 (1.31 - 3.70)**	**0.003**
*B*13*	0.056 (8)	0.060 (14)	0.93 (0.38 - 2.28)	0.88
*B*14*	0.098 (14)	0.081 (19)	1.23 (0.60 - 2.54)	0.57
*B*15*	0.385 (55)	0.449 (105)	0.77 (0.50 - 1.17)	0.22
*B*18*	0.063 (9)	0.043 (10)	1.51 (0.60 - 3.80)	0.39
*B*35*	0.140 (20)	0.214 (50)	0.60 (0.34 - 1.06)	0.08
*B*39*	0.014 (2)	0.004 (1)	3.30 (0.30 - 36.78)	0.33
*B*41*	0.014 (2)	0.013 (3)	1.10 (0.18 - 6.64)	0.92
*B*42*	0.028 (4)	0.004 (1)	6.73 (0.74 - 60.82)	0.09
*B*44*	0.056 (8)	0.051 (12)	1.09 (0.44 - 2.75)	0.85
*B*49*	0.140 (20)	0.132 (31)	1.07 (0.58 - 1.95)	0.84
*B*50*	0.021 (3)	0.021 (5)	0.98 (0.23 - 4.16)	0.98
*B*51*	0.035 (5)	0.030 (7)	1.17 (0.36 - 3.76)	0.80
*B*52*	0.014 (2)	0.038 (9)	0.36 (0.08 - 1.67)	0.19
*B*53*	0.161 (23)	0.235 (55)	0.62 (0.36 - 1.07)	0.08
*B*56*	0.014 (2)	0.004 (1)	3.30 (0.30 - 36.78)	0.33
*B*57*	0.049 (7)	0.017 (4)	2.95 (0.85 - 10.28)	0.09
*B*58*	0.224 (32)	0.192 (45)	1.22 (0.73 - 2.03)	0.46
*B*82*	0.021 (3)	0.017 (4)	1.23 (0.27 - 5.58)	0.79

n = number of individuals, p: values after adjusting for gender and ethnicity

Almost all of these individuals in both the HIV-2 infected group (38/39) and the control group (30/34) were heterozygous for this allele. This indicates that *HLA-B*080101* could be directly or indirectly (if in strong LD with another locus) mediating the susceptibility effect and that individuals carrying this allele could be more susceptible to HIV-2 infection than those negative for *B*080101* in this community.

We then grouped all the *HLA-B* alleles into two mutually exclusive groups as Bw4 or Bw6 based on their amino acid composition at positions 77-83. Those within the Bw4 group were further subdivided into Bw4-80I or Bw4-80T depending on whether the amino acid at position 80 was an isoleucine (I) or a threonine (T). It has been shown that *HLA-B Bw4* alleles and in particular those with isoleucine at position 80 (Bw4-80I) serve as ligands for some KIR molecules with 3 extracellular domains (KIR3D)[184-186]. However, Bw6 alleles predominated in our samples compared to Bw4 alleles (58.9% vs. 41.1% respectively, $p < 10^{-7}$). The majority of individuals (65.3%) carried at least one copy of the Bw4 allele but only 17% of them were homozygous for Bw4 (Bw4/Bw4) while 34.7% were homozygous for Bw6 alleles. This difference between homozygosity for Bw4 and Bw6 was significant ($p < 10^{-7}$) within the cohort as a whole, and within groups (HIV-2 infected) and control (HIV negative).

Overall, there were more Bw4 alleles with isoleucine at position 80 than Bw4-80T in the entire cohort (57.8% vs. 42.2%, $p = 7.93 \times 10^{-5}$). A comparison of the proportion of individuals carrying at least one copy of Bw4-80I to those with Bw4-80T using the Bw6 homozygous as a reference group (**Table 3.4**) showed no difference between HIV-2 positive and HIV negative groups, suggesting that there is no influence of these groups of *HLA-B* alleles on HIV-2 acquisition.

Table 3.4: HLA-B Bw4 alleles and susceptibility to HIV-2 infection

Genotypes	HIV-2 (n = 143)	HIV-neg (n = 234)	OR (95% CI)	p
Bw4/x	0.643 (92)	0.645 (151)	0.99 (0.64 - 1.53)	0.97
Bw4-80I/x	0.378 (54)	0.440 (103)	0.85 (0.53 - 1.38)	0.52
Bw4-80T/x	0.336 (48)	0.295 (69)	1.15 (0.69 - 1.91)	0.60
Bw6/Bw6	0.357 (51)	0.355 (83)		

n = number of individuals, p: values after adjusting for gender and ethnicity

HLA-C alleles and susceptibility to HIV-2 infection in Caio

HLA-C locus had the lowest number of alleles (21) compared to A and B loci (24 and 32, respectively). Four hundred and sixty two samples were successfully assigned a *HLA-C* genotype. Those with one or more mismatches were excluded from analyses. The most common allele was *Cw*0701* (16.7%) followed by *Cw*040101* (15.4%) and *Cw*0304* (14.4%) (**Table 3.1**). The eight most common alleles (*Cw*0802, Cw*0102, Cw*060201, Cw*0302, Cw*0210, Cw*0304, Cw*040101,* and *Cw*0701*) had frequencies ranging from 4.2% to 16.7%. A comparative analysis of *HLA-C* allele and genotype frequencies between HIV-2 infected (n = 147 individuals) and uninfected (n = 283) revealed no difference between groups (**Table 3.5**).

Table 3.5: The effect of HLA-C genotypes on HIV-2 infection

Allele	HIV-2 (n = 147)	HIV-neg (n = 283)	OR (95% CI)	p
*Cw*01*	0.088 (13)	0.074 (21)	1.21 (0.59 - 2.50)	0.60
*Cw*02*	0.150 (22)	0.212 (60)	0.66 (0.38 - 1.12)	0.12
*Cw*03*	0.401 (59)	0.438 (124)	0.86 (0.57 - 1.28)	0.45
*Cw*04*	0.293 (43)	0.300 (85)	0.96 (0.62 - 1.49)	0.87
*Cw*05*	0.082 (12)	0.042 (12)	2.02 (0.88 - 4.61)	0.10
*Cw*06*	0.075 (11)	0.095 (27)	0.75 (0.36 - 1.57)	0.45
*Cw*07*	0.381 (56)	0.336 (95)	1.22 (0.81 - 1.85)	0.35
*Cw*08*	0.136 (20)	0.155 (44)	0.85 (0.48 - 1.51)	0.59
*Cw*12*	0.014 (2)	0.018 (5)	0.77 (0.15 - 4.04)	0.76
*Cw*14*	0.014 (2)	0.039 (11)	0.34 (0.07 - 1.55)	0.16
*Cw*15*	0.054 (8)	0.039 (11)	1.45 (0.57 - 3.69)	0.44
*Cw*16*	0.054 (8)	0.074 (21)	0.72 (0.31 - 1.67)	0.44
*Cw*17*	0.034 (5)	0.021 (6)	1.62 (0.48 - 5.39)	0.43
*Cw*18*	0.041 (6)	0.035 (10)	1.16 (0.41 - 3.26)	0.78

n = number of individuals, p: values after adjusting for gender and ethnicity

We next grouped all *HLA-C* alleles into one of two mutually exclusive categories C1 or C2 based on their amino acid composition at position 80. If the amino acid at position 80 is asparagine, the allele is classified as group 1 (C1) whereas if lysine is at that position it is grouped as C2. This classification is important because certain killer immunoglobulin-like receptors on the surfaces of NK cells use HLA-C group 1 molecules while others use group 2 molecules as their ligands in order to modulate NK cell cytotoxic activities. We then determined the distribution of

C1 and C2 in Caio and tested for their potential effect on HIV-2 infection (**Table 3.6**). We hypothesised that HIV-2 infection would be less likely in the group of individuals with functional activating KIR molecules (activating KIR + ligand) since they can readily detect non-self cells and more effectively activate NK cells to lyse the targets sooner than later post infection. It is known that the presence of one without the other leads to a null phenotype.

Table 3.6: The effect of HLA-C groups on HIV-2 infection

Genotypes	HIV-2 (n = 147)	HIV-neg (n = 283)	p
C1/C1	0.338 (57)	0.392 (111)	ns
C1/C2	0.463 (68)	0.438 (124)	
C2/C2	0.150 (22)	0.167 (48)	

n: number of individuals, p: p-values at 5% level of significance

Overall, HLA-C group 1 was over-represented in our cohort (61.3% vs. 38.7%, $p < 10^{-7}$ for C1 and C2, respectively). More than a third of the cohort (38.5%) was homozygous for C1 (C1/C1) and only 16.0% were homozygous for C2 (C2/C2). Although C1/C1 was over-represented in both the HIV-2 infected and the uninfected groups separately, there was no significant difference in the various categories (C1/C1, C1/C2, and C2/C2) between the two groups (**Table 3.6**). This is an indication that individual *HLA-C* alleles or groups of alleles did not influence susceptibility to HIV-2 infection in the study population.

KIR genes and HIV-2 infection in the Caio population

DNA samples from the entire cohort were successfully typed for the presence or absence of 15 KIR genes including *2DL1, 2DL2, 2DL3, 2DL4, 2DL5, 2DS1, 2DS2, 2DS3, 2DS4, 2DS5, 3DL1, 3DL2, 3DL3, 3DS1,* and *2DP1*. The technique (PCR-SSP) involved the use of two pairs of *KIR* gene-specific primers targeting conserved regions and annealing at 2 different exons of the same *KIR* gene in order to avoid false negative results, which could occur if there is a point mutation at one of the primer binding sites. All 14 functional *KIR* genes and the pseudogene *2DP1* were detected in our samples. As has been observed in other populations, most of the inhibitory *KIR* genes were present in almost all DNA samples (**Table 3.7**), with the exception of *2DL2* and *2DL5*, which were each present at a frequency of about 50%. The ubiquitous *2DS4* gene was the only activating *KIR* gene that was found in nearly all samples (98.8%). All other activating *KIR* gene frequencies were near 30% on average.

A comparison of *KIR* gene frequencies between cases (HIV-2 infected) and controls (HIV negative) revealed that *KIR2DS2* and the corresponding inhibitory gene *KIR2DL2* (both of which are in strong LD with each other), were more likely to be present in the control groups than in cases (**Table 3.8**). The differences were statistically significant: OR = 0.67 (0.45-0.98), p = 0.04 for *2DS2* and OR = 0.63 (0.43-0.93), p = 0.02 for *2DL2*. This finding indicates that individuals with KIR2DS2 and/or KIR2DL2 genes are more likely to be protected against HIV-2 acquisition in this predominantly Manjako community than those without any of these genes.

Table 3.7 KIR gene frequency in the Caio population

Type	Gene	Number of individuals	Frequency
Inhibitory	*2DL1*	508	0.99
	2DL2	287	0.56
	2DL3	459	0.90
	2DL4	512	1.00
	2DL5	246	0.48
	3DL1	508	1.00
	3DL2	513	1.00
	3DL3	511	1.00
Activating	*2DS1*	101	0.20
	2DS2	262	0.51
	2DS3	139	0.27
	2DS4	507	0.99
	2DS5	131	0.26
	3DS1	94	0.18
Pseudogene	*2DP1*	505	0.98

KIR molecules are primarily expressed on NK cell surfaces and some T-cells. They are key regulators of NK cell activities and are of two types: the inhibitory and the activating KIRs. NK cells survey other cells in order to detect abnormalities in levels of HLA class I expression on the target cell surface. Both the KIR molecule and its corresponding ligand (s) need to be present in the same individuals for the KIR to be functional. **Table 3.9** shows some KIR and their known or putative ligands. The absence of one renders the other null.

Table 3.8: KIR genes and susceptibility/resistance to HIV-2 infection

Genes	HIV-2 (n = 150)	HIV-neg (n = 328)	OR (95% CI)	p
2DL1	0.980 (147)	0.994 (326)	0.31 (0.05 - 1.85)	0.20
2DL3	0.860 (129)	0.909 (298)	0.62 (0.34 - 1.13)	0.12
2DL5	0.507 (76)	0.466 (153)	1.17 (0.80 - 1.72)	0.42
2DS1	0.227 (34)	0.189 (62)	1.25 (0.78 - 2.01)	0.35
2DS3	0.247 (37)	0.277 (91)	0.85 (0.55 - 1.33)	0.48
2DS4	0.993 (149)	0.985 (323)	2.37 (0.27 - 20.47)	0.43
2DS5	0.293 (44)	0.238 (78)	1.34 (0.87 - 2.06)	0.19
2DP1	0.973 (146)	0.988 (324)	0.45 (0.11 – 182)	0.26
2DS2	**0.440 (66)**	**0.543 (178)**	**0.67 (0.45 - 0.98)**	**0.04**
2DL2	**0.480 (72)**	**0.595 (195)**	**0.63 (0.43 - 0.93)**	**0.02**
3DS1	0.167 (25)	0.204 (67)	0.78 (0.47 - 1.29)	0.34
3DL1	0.987 (148)	0.991 (325)	0.70 (0.11 - 4.22)	0.69

n = number of individuals, p: values after adjusting for gender and ethnicity

To look at the effect of HLA-KIR compound genotypes on HIV-2 infection, we studied the group of individuals that have the corresponding ligands for their KIR molecules. **Table 3.10** shows some of the HLA-KIR combinations found in our study population. A comparison of HLA-KIR compound genotype frequencies between HIV-2 infected and uninfected groups showed that individuals expressing KIR2DS2 or KIR2DL2 along with at least one copy of HLA-C group 1 were less likely to be HIV-2 positive (OR: 0.63 (0.41-0.95), p = 0.03; and OR: 0.66 (0.44-0.99), p = 0.04, for KIR2DS2 and 2DL2, respectively).

Table 3.9: KIR and their corresponding ligands

KIR	Ligands	specificity	HLA alleles
KIR2DL2, KIR2DL3 (KIR2DS2)[d]	C1	77S - 80N	Cw1, Cw3, Cw7, Cw8, Cw12, Cw13, Cw14, CW16
KIR2DL1 (KIR2DS1)[d]	C2	77N - 80K	Cw2, Cw4, CW5, CW6, Cw15, Cw17, CW18
KIR3DL1 (KIR3DS1)[d]	Bw4	80I	B*1516, B*4901/03, B*51, B*52, B*53, B*57, B*58, B*59
KIR3DL1		80T	B*13, B*27, B*37, B*44, B*4902, B*5303/09
KIR2DL4	HLA-G		
KIR3DL2	HLA-A3		
	HLA-A11		

[d]: The activating receptors KIR2DS1, KIR2DS2, and KIR3DS1 are very similar to their corresponding inhibitory counterparts in nucleotide and amino acid composition (>90% similarity) and are thought to be using similar ligands but with much weaker interaction for the activating receptors.

In brief, the protective effect described earlier with individual *2DS2* and *2DL2* genes was linked to functional *KIR* genes i.e. *KIR* genes present in the same individual together with their corresponding ligands. No difference was observed in individuals having KIR2DS2 or 2DL2 molecules without C1 and *vice-versa*.

Table 3.10: The effect of KIR-HLA combinations on HIV-2 infection

Genotypes	HIV-2 (n)	HIV negative (n)	OR (95% CI)	p
3DS1 : BW4-80I/x	0.035 (5)	0.074 (17)	0.48 (0.17 - 1.40)	0.18
Bw6/Bw6	0.352 (50)	0.355 (82)		
3DS1 : BW4-80I/x	0.035 (5)	0.074 (17)	0.46 (0.16 - 1.27)	0.13
Others	0.965 (137)	0.926 (214)		
2DS2 : C1/x	0.338 (49)	0.449 (124)	0.63 (0.41 - 0.95)	**0.03**
Others	0.662 (96)	0.551 (152)		
2DL2 : C1/x	0.386 (56)	0.489 (135)	0.66 (0.44 - 0.99)	**0.04**
Others	0.614 (89)	0.511 (141)		
2DS1 : C2/x	0.138 (20)	0.120 (33)	1.18 (0.65 - 2.13)	0.60
Others	0.862 (125)	0.880 (243)		
2DL1 : C1/x	0.834 (121)	0.826 (228)	1.06 (0.62 - 1.82)	0.83
Others	0.165 (24)	0.174 (48)		

n = number of individuals, p: values after adjusting for gender and ethnicity.

We also compared the frequency of individual *KIR* genes in Caio with those reported in other populations in Africa, Europe and Asia and to the best of our knowledge this is the first study examining KIR/HLA combinations in Guineans and Gambians. In 2005, Denis et al. reported KIR frequencies in 118 subjects from Senegal. In our study we found the frequency of all activating genes (except KIR2DS5) to be higher compared to data from the neighbouring Senegalese population (**Table 3.11**). The most striking difference was observed among carriers of *KIR3DS1* (19% vs. 4 % respectively). Our future plan is to sequence *KIR3DS1* along with those genes that showed significant differences between cases and controls (*KIR2DS2* and *KIR2DL2*) to confirm our present observations. The *KIR3DS1*

gene has previously been reported to be low in most African populations and some experts believe, based on sequences derived to date, that the *KIR3DS1* gene is monomorphic. However, the statistics from the international KIR database indicates 14 different allelic forms of this gene. Our observation (high frequency of *KIR3DS1* compared to other African populations), if confirmed by sequencing, will further strengthen our beliefs that this group of Manjako population from the Caio sector in north-western coast of Guinea Bissau is distinct. Moreover we may be able to identify genuine polymorphisms in the *KIR3DS1* gene. Most if not all the inhibitory genes frequencies were similar to that reported by Denis and his group[187].

In a more genetically distant South African population, most of the activating *KIR* genes frequencies were higher compared to what we obtained in this study. *KIR3DS1* was not reported which could either mean that the authors did not look for this gene or that the gene was completely absent. The frequencies of framework genes are consistent in most populations reported in **Table 3.11**.

Table 3.11: KIR genes frequencies in Caio and other populations

Genes	Manjago	Senegalese*	South African *	English¤	French*	Indian§	South Korean¥
	(n=513)	(n= 118)	(n= 50)	(n= 136)	(n= 108)	(n= 72)	(n= 154)
2DL1	99	100	96	91	97	88	99
2DL2	57	55	72	49	50	79	14
2DL3	90	90	64	92	91	65	99
2DL4	100	100	100	100	100	100	100
2DL5	48	52	82	N/A	47	79	38
2DS1	21	13	40	45	36	54	38
2DS2	52	42	64	51	51	62	17
2DS3	27	24	38	24	31	43	16
2DS4	99	100	100	96	96	81	94
2DS5	24	30	62	32	27	47	27
3DL1	99	99	100	97	96	88	94
3DL2	100	100	100	100	100	100	100
3DL3	100	100	100	N/A	100	100	100
3DS1	19	4	N/A	N/A	44	N/A	N/A
2DP1	99	100	98	N/A	97	N/A	N/A

*Denis L et al. 2005 Tissue Antigens 66, 267-76; *Williams F et al. 2004 Human Immunology 65(9-10), 1084-1085; ¤Norman P et al. 2004 Immunogenetics; §Rajalingam R et al. 2002 Immunogenetics 53, 1009-19; ¥Whang, D. H. 2005 Human Immunology 66,146; N/A: data not available

Discussion

One of the objectives of this study was to determine the type of *HLA* alleles and *KIR* genes present in two well characterised West African cohorts: one in Caio, Guinea Bissau and the other in Fajara, The Gambia. Also we aimed to study the relationship between the presence or absence of these highly polymorphic genes with susceptibility or resistance to HIV-2 infection and disease progression assessed using markers of progression to AIDS such as decline in $CD4^+$ T cell count and increase in viral load over time.

*HLA-A*8001* weakly associates with susceptibility to HIV-2 infection

We detected 15 different *A* serological groups (two-digit alleles) and 24 *HLA-A* subtypes (four-digit alleles) (**Figure 3.1**) in Caio. Although, the allele and genotype frequencies did not differ between groups (HIV-2 infected and HIV negative controls), some of them differed considerably from those observed in neighbouring populations.

*HLA-A*01* genotype frequency in this predominantly Manjako community (13.8%) was comparable to that seen in the Fulani population from Burkina Faso (13.3%)[188] and Fula and Mandinka in The Gambia[189] (13.3% and 13.2%, respectively), but differed from that reported in a Malian population (7.3%)[190].

*HLA-A*02* is one of the most widely studied class I allele in different populations. We found its genotype frequency in Caio to be 22.5% (**Table 3.1**), which is slightly lower compared to other populations. Allsopp and colleagues reported that 35.8% of Wolof and 44.0% of Serere in The Gambia are *A2* positive[189]. In Mali, Kalidi et al. reported *A2* frequency to be 29.2%[190]. A low *A2* frequency has been described also in the Mossi tribe (3.7%) in Burkina Faso[188]. In many Caucasoid and Oriental populations, the frequency of *A2* varies between 25% and 42% average[98]. *HLA-A*02*

is one of the most polymorphic group of alleles at the A locus with currently 244 closely related subtypes described in the Immunogenetic database[94]. In Caio, we found that *A*02* was the most diverse group of allele at the A locus with 3 subtypes: *A*0201, A*0202* and *A*0205*. The former being the most common as is customary in other populations worldwide but its allele frequency (6.6%) was relatively low compared to most populations with high resolution typing data. For example, in Ugandans, *A*0201* allele frequency was 18.4%[191] while it varied between 12.1% and 14.9% in different regions of the Cape Verdes Islands[192].

In the present study, a rare allele *HLA-A*8001* showed a weak association with HIV-2 acquisition. Individuals positive for *HLA-A*8001* were more likely to be HIV-2 infected than those without. Eighty percent of individuals with this genotype were HIV-2 infected (**Table 3.2**). This is an indication that despite being a rare allele in this population, *HLA-A*8001* could be directly or indirectly mediating HIV-2 acquisition. To the best of our knowledge, this trend has not been reported before in HIV-2 and warrants further investigation in other well established HIV-2 cohorts. It is however in line with the suggestions from Trachtenberg and colleagues[137] that rare *HLA* alleles may have a selective advantage over more frequent alleles in populations where HIV-1 rapidly adapts to the more frequent *HLA* alleles. If confirmed that *A*8001* leads to increased susceptibility to HIV-2 infection in HIV-2 endemic areas, investigations into the mechanisms by which *A*8001* affects susceptibility to HIV-2 may provide new insight into how HIV infects cells and may also inform future vaccine design to block HIV infection.

HLA-B alleles associate with HIV-2 infection

Our data showed that B locus was the most diverse of all class I loci in Caio with 32 different alleles. We demonstrated in this study for the first time that a strong association exists between *HLA-B*080101* and susceptibility to HIV-2 infection in this sample of the West African study populations. Individuals with this phenotype were two times more likely to be HIV-2 infected than those without. *B*08* is a common allele in Caio, its AF (10.8%) and GF (19.6%) are relatively high compared to other West African populations. It was almost absent in three ethnic groups in Burkina Faso[188]. Our data are in accordance with those of Andrien et al. from a study conducted in a population in Dakar Senegal (GF= 20.5%)[193]. In The Gambia, the phenotype frequency of B*08 varied from 13.3% to 23.9% in different ethnic groups[189]. HLA-B*08 forms part of the 8.1 ancestral haplotype, which is carried by most Caucasians with the serological B8 type and has been associated with accelerated HIV-1 disease, susceptibility to insulin-dependent diabetes mellitus (IDDM), and systemic lupus erythematosus[194] among others.

HLA-C alleles did not affect HIV-2 infection

None of the individual *HLA-C* alleles were associated with susceptibility or resistance to HIV-2 infection in Caio, even after grouping them into C1 and C2. However, the classification of C alleles into two mutually exclusive groups (C-group 1 (C1) and C-group 2 (C2)) was important because certain HLA-C molecules have been shown to serve as ligands for KIR molecules and both the KIRs and their corresponding ligands need to be present in the same individual for the KIRs to become functional.

Activating KIR frequencies are higher in Caio compared to neighbouring populations

In this study we used KIR specific primers (SSP) to investigate the presence or absence of 15 *KIR* genes in an isolated community cohort of HIV-2 infected and uninfected individuals belonging mostly to the Manjako tribe in North-Western Guinea Bissau. We found all 15 *KIR* genes in this sample of West African population at various frequencies. A comparative analysis of the *KIR* gene frequencies found in Caio with published data from neighbouring populations and elsewhere revealed that activating *KIR* genes frequencies were significantly higher in Caio compared to the neighbouring populations e.g. Senegalese, while those of the inhibitory *KIRs* were generally similar. The frequencies of activating *KIRs* in Caio were similar to those in Southern Africa, Europe and Asia.

Literature search indicates that *KIR3DS1* is rare in people of African ancestry and common in Caucasoid populations. Indeed, most published KIR data from Africa rarely include *KIR3DS1* in their tables or analyses due its low frequency in those populations. Recent studies employing new molecular tools to detect *KIR* genes have reported that *KIR3DS1* frequencies vary greatly from one population to another even within geographically small areas. For examples, Single and colleagues have reported the frequency of *3DS1* in three major ethnic groups in Nigeria to be 12% in Yuroba, 6.3% in Ibo, and 3.4% in Hausa[195], all within the same country. In Ivory Coast, Jennes et al. found *3DS1* frequency to be as high as 16% in a cohort of commercial sex workers in Abidjan[196].

Genetic diversity is generally high in Africa and there is a wide range of ethnic groupings in Sub Saharan Africa. The reason for the differences observed between *KIR* gene frequencies in Manjakos from Caio and the Senegalese is uncertain. One

of the reasons why these two neighbouring populations have different KIR gene frequencies could be due to sample sizes. Denis et al. analysed KIR data generated from 90 Senegalese while we used 513 individuals in our analysis. One other reason could be the ethnic diversity. In Caio, more than 95% of the study participants were from a single tribe (Manjako) and anthropological studies have shown that they are sexually isolated from neighbouring populations. There are however, considerable inter-marriages between Manjakos and Europeans especially French and Portuguese. Thus it is possible that the high frequencies reflect admixture with populations with high frequencies of activating KIR genes. In Caio, almost every family sponsors at least one of their children to go to Europe for work or education. It may therefore not be surprising that the frequencies of activating *KIR*s in Caio are very similar to those of the French. For example *KIR2DS2* frequency is 51% in Manjakos and 51% in French[187] while it is 42% in Senegalese[187]; that of *2DS5* is 26% in Caio and 27% in French while it is 30% in Senegalese and 62% in South African[197].

KIR-HLA compound genotypes in HIV-2 infected Guineans

The interaction between KIR and HLA molecules at the cellular level strictly governs the activity of natural killer (NK) cell in an individual, a very important step in the innate immune system. KIR molecules on the surface of NK cells continually interact with other cell surface molecules looking for their corresponding ligands, which are HLA class I molecules. The nature and degree of interaction between these molecules determines the type of signal to be sent to the NK cell. In the presence of the corresponding ligands for the inhibitory KIRs, an inhibitory signal is sent to the NK cell not to kill the target, which is recognised as a "self" cell. But in the absence of the ligands for the inhibitory receptor, the signal sent to the NK cell will

instead be a stimulatory one. This in turns activates NK cells to kill the target, which must have been recognised as a "non-self" cell (e.g. a virally infected cell or a cell undergoing transformation to become a tumour). In Caio, we observed that most of the inhibitory *KIR* genes were present in almost every sample while the activating *KIR* genes were relatively low compared to their corresponding inhibitory counterparts.

But epidemiological evidence is accumulating from large population studies in Western cohorts indicating that *KIR3DS1* in individuals carrying at least one copy of *HLA-B Bw4-80I* (one of the Bw4 serological group of alleles with Isoleucine at position 80), is associated with rapid progression to AIDS following HIV-1 infection. Although, the frequency of *KIR3DS1* was several folds higher in our study population compared to other African populations, it was not as high as in most Caucasoid populations. The reason we could not find any association between *KIR3DS1* and HIV-2 disease in this community cohort could relate to several factors: (1) the fact that HIV-2 has a different course of disease presentation compared to that of HIV-1, (2) population genetic differences, and (3) the differential frequency of *KIR3DS1* in the studied populations.

In this study however, we report some new associations between certain *KIR* genes and HIV-2 infection. We discovered in the course of our analyses that the large majority of Caio individuals carrying the *KIR2DS2* or *KIR2DL2* genes in the presence of their corresponding ligands HLA-C group 1 alleles are more likely to be free from HIV-2 infection. This is an indication that these functional *KIR* genes in people with the compound genotype (KIR2DS2/C1 or KIR2DL2/C1) may be associated with protection from HIV-2 acquisition at least in this community. At this stage we are unable to distinguish which of these KIR2D genes is mediating the observed effect since both 2DS2 and 2DL2 segregate from the same locus. Plans

are on the way to sequence these genes in order to describe this association in future publications. There were no data on KIR from an HIV-2 cohort prior to our work, which made it impossible to compare our findings with previous findings in the context of HIV-2 infection and/or disease. We have replicated this study in a multi-ethnic cohort in Fajara, The Gambia and a comparative analysis of both data sets is presented in the concluding chapter (**Chapter 7**) of this thesis.

In summary, we found that individuals from this predominantly Manjako community in Caio with certain HLA genotypes including *HLA-A*8001* or *HLA-B*080101* were more likely to be at a greater risk of acquiring HIV-2 infection than those without any of these alleles. None of the *HLA-C* alleles individually or collectively influenced susceptibility or resistance to HIV-2. However, those individuals expressing KIR2DL2 or KIR2DS2 molecule together with at least one copy of their corresponding ligands (C1) were protected against HIV-2 infection.

CHAPTER 4 : THE EFFECT OF HLA CLASS I AND KIR GENE FREQUENCIES ON MARKERS OF DISEASE PROGRESSION

In this chapter we discussed the HLA and KIR data collected from Caio participants who were singly infected with HIV-2. A substantial proportion of this group had participated in the majority of studies conducted in the Caio field station over the years since the first serosurveys in 1989. This group of people also included a subset of 65 individuals who were infected before 1989, of whom 64 were still alive 19 years later. The MRC unit in The Gambia has conducted three population serosurveys in Caio since the establishment of the field station.

The first serosurveys was conducted in 1989-90, in which 7.9% of the adult population was infected with HIV-2[181]. The second survey was conducted in 1996-8 and the third took place in 2007-8. Several other studies with specific research questions were conducted between the surveys. These studies focussed on gaining a better understanding of mechanisms driving HIV-2 infection into long-term non-progression (LTNP) – a phenomenon observed in the majority of HIV-2 infected individuals in West Africa. During these studies and surveys, different immunological and virological parameters were recorded including $CD4^+$ T cell counts (absolute and percentage), $CD8^+$ T cell counts, CD38, β2m, neopterin, and HIV-2 viral load measurements. In the present study, the effects of individual classical *HLA* class I gene variants on HIV-2 outcomes were explored using internationally accepted markers of HIV disease progression (decline in $CD4^+$ T cell count and increase in HIV viral load over time).

The effect of HLA-A gene variants on HIV-2 viral load

A total of 138 individuals had between two and four HIV-2 viral load measurements (median = 3) documented at various time points between 1991 and 2007 and 135 of them also had a similar number of measurements of absolute and/or percent $CD4^+$ T cell counts over the same period of time, in addition to their HLA-A data. The dates of seroconversion are unknown for most infected subjects.

To study the effect of allelic variations at the HLA-A locus on HIV-2 disease and rate of progression, we computed the mean square root (sqrt) of absolute $CD4^+$ T cell count measurements recorded in 1991, 1996, 2003 and 2006 for each person and compared the values between groups of individuals positive for a particular *HLA-A* allele versus those negative for the allele of interest (**Table 4.1**). The average mean sqrt $CD4^+$ T cell count for the entire cohort was 24.70. Similarly, we log-transformed HIV-2 viral load (VL) measurements collected over the same period of time, calculated their means, and compared them between groups. The average mean $\log_{10}$HIV-2 viral load was 3.10 for the entire cohort. These comparisons showed that none of the *HLA-A* variants had a significant influence on $CD4^+$ T cell dynamics or the rate of HIV-2 replication (**Table 4.1**). There was however a trend that HIV-2 patients who had the *A*7401* genotype might do better in controlling viral replication than those negative for this genotype since they had an overall low HIV-2 VL (mean $\log_{10}$HIV-2 VL = 2.41, $p = 0.09$).

Table 4.1: The effect of HLA-A variants on mean square root of absolute $CD4^+$ T cell count and mean log_{10}HIV-2 viral load (VL)

Allele	Mean sqrt CD4 (n = 135)	R	SE	P	Mean log HIV-2 VL (n = 138)	R	SE	p
*A*01+*	23.71 (16)	0.98	1.61	0.76	2.97 (16)	1.02	0.25	0.87
*A*01-*	24.23 (119)				2.92 (122)			
*A*02+*	25.20 (23)	1.05	1.32	0.39	2.98 (24)	1.02	0.20	0.80
*A*02-*	23.96 (112)				2.92 (114)			
*A*03+*	24.45 (7)	1.02	2.41	0.90	2.74 (7)	0.93	0.38	0.61
*A*03-*	24.15 (128)				2.94 (131)			
*A*23+*	23.38 (46)	0.95	0.93	0.30	2.91 (48)	0.99	0.14	0.89
*A*23-*	24.57 (89)				2.94 (90)			
*A*26+*	24.47 (23)	1.01	1.33	0.80	2.85 (24)	0.97	0.20	0.68
*A*26-*	24.11 (112)				2.95 (114)			
*A*29+*	19.15 (1)	0.79	6.38	0.43	4.01 (1)	1.37	1.00	0.28
*A*29-*	24.21 (134)				2.92 (137)			
*A*30+*	23.18 (33)	0.95	1.10	0.31	3.12 (34)	1.09	0.17	0.19
*A*30-*	24.49 (102)				2.87 (104)			
*A*32+*	29.22 (2)	1.19	4.47	0.26	3.15 (2)	1.08	0.71	0.76
*A*32-*	24.09 (133)				2.93 (136)			
*A*33+*	24.90 (53)	1.05	0.87	0.28	2.90 (54)	0.98	0.14	0.78
*A*33-*	23.69 (82)				2.95 (84)			
*A*34+*	26.31 (7)	1.09	2.39	0.36	2.95 (7)	1.01	0.38	0.96
*A*34-*	24.05 (128)				2.93 (131)			
*A*36+*	15.05 (1)	0.62	6.31	0.15	2.33 (1)	0.80	1.00	0.55
*A*36-*	24.24 (135)				2.93 (137)			
*A*66+*	26.27 (3)	1.09	3.68	0.56	2.06 (3)	0.70	0.57	0.13
*A*66-*	24.12 (132)				2.95 (135)			
*A*68+*	23.53 (24)	0.97	1.31	0.59	3.01 (24)	1.03	0.21	0.69
*A*68-*	24.31 (111)				2.91 (114)			
*A*74+*	24.77 (10)	1.03	2.02	0.76	2.41 (10)	0.81	0.31	0.09
*A*74-*	24.12 (125)				2.97 (128)			
*A*80+*	23.30 (8)	0.96	2.25	0.69	3.19 (8)	1.10	0.35	0.45
*A*80-*	24.22 (127)				2.91 (130)			

n = number of individuals; +: individuals with the allele of interest; -: individuals without the allele of interest; R: ratio of with/without; p = uncorrected p-values, adjusted for age and gender and considered significant when ≤ 0.05.

HLA-B variants and markers of HIV disease progression

The number of HIV-2 infected individuals with both HLA-B data and serial VL measurements (median = 3) was 138 and 135 of these also had serial CD4 measurements (two to four) collected over time as described above. The mean square root (sqrt) absolute $CD4^+$ T cell counts was calculated as well as the mean log_{10}HIV-2 VL. A comparison of these means was done between groups of individuals with and without the *HLA-B* variant of interest as shown in **Table 4.2** below. Individuals positive for both HIV-1 and HIV-2 (dually infected) were excluded from these analyses.

Table 4.2: The effects of HLA-B alleles on CD4+ T cell count and HIV-2 viral load measurements

Allele	Mean sqrt CD4 (n = 135)	R	SE	P	Mean log_{10} HIV-2 VL (n = 138)	R	SE	p
*B*07+*	22.38 (6)	0.91	2.54	0.42	2.78 (6)	0.95	0.40	0.73
*B*07-*	24.47 (129)				2.92 (132)			
*B*08+*	24.43 (37)	1.00	1.02	0.95	2.92 (37)	1.00	0.16	0.97
*B*08-*	24.36 (98)				2.91 (101)			
*B*13+*	22.59 (8)	0.92	2.17	0.40	2.88 (8)	0.99	0.34	0.92
*B*13-*	24.49 (127)				2.91 (130)			
B*14+	26.72 (14)	1.11	1.65	0.14	2.39 (14)	0.80	0.26	**0.04**
*B*14-*	24.11 (121)				2.97 (124)			
B*15+	22.39 (52)	0.87	0.83	**0.003**	3.24 (53)	1.20	0.13	**0.001**
*B*15-*	25.62 (83)				2.70 (85)			
*B*18+*	22.67 (9)	0.93	2.05	0.39	2.97 (9)	1.02	0.33	0.84
*B*18-*	24.50 (126)				2.91 (129)			
*B*35+*	26.19 (19)	1.09	1.41	0.17	2.55 (19)	0.86	0.22	0.08
*B*35-*	24.08 (116)				2.97 (119)			
*B*39+*	27.87 (1)	1.14	6.20	0.57	2.56 (1)	0.88	0.98	0.72
*B*39-*	24.35 (134)				2.91 (137)			
*B*41+*	19.38 (1)	0.79	6.16	0.42	3.77 (2)	1.30	0.69	0.21

*B*41-*	24.42 (134)				2.90 (136)			
*B*42+*	20.38 (4)	0.83	3.07	0.19	2.99 (4)	1.03	0.49	0.88
*B*42-*	24.50 (131)				2.91 (134)			
B*44+	26.58 (7)	1.10	2.33	0.33	2.16 (8)	0.73	0.34	**0.03**
*B*44-*	24.26 (128)				2.96 (130)			
*B*45+*	25.26 (1)	1.04	6.18	0.89	2.53 (1)	0.87	0.98	0.70
*B*45-*	24.37 (134)				2.91 (137)			
*B*47+*	31.18 (1)	1.28	6.15	0.27	2.91 (1)	1.00	0.98	1.00
*B*47-*	24.33 (134)				2.91 (137)			
B*49+	21.59 (19)	0.87	1.40	**0.03**	3.28 (20)	1.15	0.22	0.07
*B*49-*	24.83 (116)				2.85 (118)			
*B*50+*	28.54 (3)	1.18	3.56	0.24	2.27 (3)	0.78	0.56	0.25
*B*50-*	24.28 (132)				2.92 (135)			
*B*51+*	26.63 (5)	1.10	2.75	0.41	2.38 (5)	0.81	0.43	0.21
*B*51-*	24.29 (130)				2.93 (133)			
*B*52+*	23.33 (2)	0.96	4.37	0.81	2.76 (2)	0.95	0.69	0.83
*B*52-*	24.39 (133)				2.91 (136)			
*B*53+*	26.71 (21)	1.12	1.33	0.06	2.68 (23)	0.91	0.20	0.21
*B*53-*	23.95 (114)				2.96 (115)			
*B*56+*	19.55 (2)	0.80	4.35	0.27	2.59 (2)	0.89	0.69	0.64
*B*56-*	24.45 (133)				2.92 (136)			
*B*57+*	28.22 (7)	1.17	2.30	0.09	3.43 (7)	1.19	0.37	0.15
*B*57-*	24.17 (128)				2.88 (131)			
*B*58+*	25.16 (31)	1.04	1.11	0.42	3.10 (31)	1.09	0.17	0.21
*B*58-*	24.14 (104)				2.85 (107)			
*B*82+*	26.48 (3)	1.09	3.55	0.55	1.86 (3)	0.63	0.56	0.06
*B*82-*	24.33 (132)				2.93 (135)			

n = number of individuals; +: individuals with the allele of interest; -: individuals without the allele of interest; R: ratio of with/without; p = uncorrected p-values adjusted for age and gender and considered significant when ≤ 0.05

The results revealed that *HLA-B*15* was a strong predictor of HIV-2 disease progression assessed using well correlated markers of progression to AIDS ($CD4^+$ T cell count and viral load). Individuals positive for *HLA-B*15* had their overall mean sqrt $CD4^+$ T cell counts significantly lower (22.39, SE: 0.83, p = 0.003,) and their overall mean log_{10}HIV-2 VL considerably higher (3.24, SE: 0.13, p = 0.001) compared

to those without *B*15* (mean sqrt $CD4^+$ T cell counts = 25.62, SE: 0.65; and mean log_{10}HIV-2 VL = 2.70, SE: 0.10, respectively). Three different *HLA-B*15* subtypes were present in the study population: *B*1503, B*1510,* and *B*1516. B*1503* (AF=12.8%, and GF=24.6%) and *B*1510* (AF=9.9%, and GF=18.6%) were the predominant subtypes at the population level. A univariate analysis comparing CD4 count and viral load among groups of infected individuals positive for individual *B*15* subtypes revealed that those participants expressing *B*1503* had even lower CD4 count (mean sqrt $CD4^+$ T cell count = 21.51, SE: 1.5, p = 0.006) and higher HIV-2 viral load (mean log_{10}HIV-2 VL = 3.36, SE: 0.18, p = 0.006) compared to *B*1503*-negative individuals (mean sqrt $CD4^+$ T cell count = 24.85, SE: 0.55; and mean log_{10}HIV-2 VL = 2.92, SE: 0.08) (**Table 4.3**). This effect was not seen with the other common *B*15* allele (*B*1510*) neither was it seen with the minor *B*1516* allele (AF=1.1%, and GF=2.2%). This suggests that *B*1503* accounts for the overall poor prognosis seen with *HLA-B*15.*

Table 4.3: Effects of HLA-B*15 subtypes on CD4 count and HIV-2 viral load

	Mean sqrt CD4 (n)	R	SE	p	Mean log HIV-2 VL (n)	R	SE	p
*B*15+*	22.39 (52)	0.87	0.83	**0.003**	3.24 (53)	1.20	0.13	**0.001**
*B*15-*	25.62 (83)				2.70 (85)			
*B*1503+*	21.51 (27)	0.86	1.15	**0.006**	3.36 (28)	1.20	0.18	**0.006**
*B*1503-*	25.10 (108)				2.80 (110)			
*B*1510+*	24.39 (25)	1.00	1.23	0.99	3.07 (25)	1.07	0.19	0.36
*B*1510-*	24.38 (110)				2.87 (113)			
*B*1516+*	20.89 (3)	0.85	3.57	0.32	2.96 (3)	1.02	0.57	0.94
*B*1516-*	24.46 (132)				2.91 (135)			

n: number of $HIV\text{-}2^+$ individuals with serial $CD4^+$ T cell counts or viral load measurements; +: individuals with the allele of interest; -: individuals without the allele of interest; R: ratio of with/without; p: uncorrected p-values adjusted for age and gender and considered significant when ≤ 0.05; The total number of HIV-2 infected individuals analysed was 135.

Some of the *B* alleles showed a marginal association with control of viral replication including *HLA-B*14* and *HLA-B*44*. HIV-2 infected people with either of these genotypes had a significantly lower HIV-2 VL compared to those without any *B*14* or *B*44* in their genotypes (**Table 4.2**). On the other hand, *HLA-B*49* individuals who are infected with HIV-2, tended to have a lower $CD4^+$ T-cell count and a trend towards higher viral load (**Table 4.2**), behaving almost like their *B*1503*-positive counterparts.

Furthermore, we checked the effect of HLA-B molecular grouping into Bw4, Bw6; Bw4-80I, and Bw4-80T, on $CD4^+$ T cell count and their impact on HIV-2 viral load measurements over time, but no significant difference was noted between groups (data not shown).

The effect of HLA-C alleles on HIV-2 disease

We performed similar analyses as described above for *HLA-A* and *-B* using 138 HIV-2 infected participants who had their *HLA-C* genotype assigned and also had two to four time points measurements recorded for $CD4^+$ T cell count and HIV-2 VL (**Table 4.4**). *HLA-Cw*02*, which is in LD with *HLA-B*15*, showed similar trends as observed with *B*15*-positive individuals. HIV-2 infected people with *Cw*02* genotype (carriers of at least one copy of *Cw*02* alleles) had significantly lower $CD4^+$ T cell count (mean sqrt $CD4^+$ T cell count = 21.26, SE: 1.39, $p = 0.02$). Cw*02 carriers also showed a trend towards having high HIV-2 VL (mean $\log_{10}$HIV-2 VL = 3.29, SE: 0.21, $p = 0.10$). Meanwhile their counterparts (HIV-2 infected individuals who were negative for *Cw*02* allele) had a better preserved immune system (mean sqrt $CD4^+$ T cell count = 24.74, SE: 0.58; and mean $\log_{10}$HIV-2 VL = 2.90, SE: 0.09, respectively).

Table 4.4: The effect of HLA-C on CD4 count and HIV-2 viral load

Allele	Mean sqrt CD4 (n = 135)	R	SE	p	Mean log_{10}HIV-2 VL (n = 138)	R	SE	p
*Cw*01+*	23.04 (13)	0.95	1.76	0.48	3.14 (13)	1.07	0.28	0.48
*Cw*01-*	24.35 (122)				2.94 (125)			
Cw*02+	21.26 (20)	0.86	1.39	**0.02**	3.29 (21)	1.13	0.21	0.10
*Cw*02-*	24.74 (115)				2.90 (117)			
*Cw*03+*	24.44 (56)	1.02	0.84	0.74	2.83 (57)	0.93	0.13	0.19
*Cw*03-*	24.07 (79)				3.05 (81)			
*Cw*04+*	24.91 (41)	1.04	0.98	0.40	2.86 (41)	0.95	0.15	0.43
*Cw*04-*	23.92 (94)				3.00 (97)			
*Cw*05+*	24.50 (12)	1.01	1.83	0.87	2.90 (12)	0.98	0.29	0.83
*Cw*05-*	24.19 (123)				2.96 (126)			
*Cw*06+*	26.02 (9)	1.08	2.13	0.38	2.56 (10)	0.86	0.32	0.19
*Cw*06-*	24.09 (126)				2.99 (128)			
Cw*07+	23.76 (49)	0.97	0.90	0.53	3.19 (51)	1.13	0.14	**0.04**
*Cw*07-*	24.48 (86)				2.82 (87)			
*Cw*08+*	24.70 (20)	1.02	1.41	0.71	2.67 (20)	0.89	0.22	0.16
*Cw*08-*	24.14 (115)				3.01 (118)			
*Cw*12+*	28.12 (1)	1.16	6.35	0.54	2.55 (1)	0.86	1.00	0.68
*Cw*12-*	24.19 (134)				2.96 (137)			
*Cw*14+*	19.46 (1)	0.80	6.35	0.45	3.48 (1)	1.18	1.00	0.60
*Cw*14-*	24.26 (134)				2.95 (137)			
*Cw*15+*	22.52 (6)	0.93	2.57	0.50	3.18 (7)	1.08	0.37	0.54
*Cw*15-*	24.30 (129)				2.95 (131)			
*Cw*16+*	28.47 (7)	1.19	2.36	0.07	2.32 (7)	0.78	0.37	0.08
*Cw*16-*	23.99 (128)				2.99 (131)			
*Cw*17+*	20.22 (5)	0.83	2.80	0.15	2.96 (5)	1.00	0.44	0.99
*Cw*17-*	24.38 (130)				2.96 (133)			
*Cw*18+*	24.45 (6)	1.01	2.58	0.93	3.22 (6)	1.09	0.41	0.50
*Cw*18-*	24.21 (129)				2.95 (132)			

n: number of HIV-2^+ individuals with serial CD4 count or viral load measurements; +: individuals with the allele of interest; -: individuals without the allele of interest; R: ratio of with/without; p: uncorrected p-values adjusted for age and gender and considered significant when ≤ 0.05

Two *HLA-Cw*02* subtypes were present in our samples (*Cw*0203* and *Cw*0210*), the most frequent of which was *Cw*0210* (AF=9.6% and GF=18.7%).

*Cw*0203* was rare in this population (AF=0.2% and GF=0.4%). A comparative analysis revealed that individuals with *HLA-Cw*0210* had a much lower CD4 count (mean sqrt $CD4^+$ T cell count = 20.64, SE: 1.53, $p < 0.001$) and a much higher HIV-2 VL (mean log_{10}HIV-2 VL = 3.45, SE: 0.24, $p < 0.001$) compared to those without this genotype (mean sqrt $CD4^+$ T cell count = 24.51, SE: 0.54 and mean log_{10}HIV-2 VL = 3.02, SE: 0.08, respectively). The other *HLA-C* allele with a significant susceptibility effect on VL was *Cw*07*. There were 3 *HLA-Cw*07* subtypes (*Cw*0701, Cw*070201,* and *Cw*0704*), the most frequent being *Cw*0701* (AF=16.7% and GF=29.4%). HIV-2 infected individuals expressing *Cw*0701* had higher HIV-2 viral load compared to those without this allele (mean log_{10}HIV-2 VL = 3.34, SE: 0.15 vs. 2.97, SE: 0.09, p = 0.03), but the $CD4^+$ T cell count did not differ between those positive for *Cw*0701* and those negative for that allele. This effect was neither seen with individuals expressing *Cw*070201* nor those with the *Cw*0704* genotype.

We then grouped individuals into two mutually exclusive groups: HLA-C group 1 (C1) or HLA-C group 2 (C2) based on the amino acid composition of their *HLA-C* alleles at position 80 as described above. A comparative analysis was performed on overall CD4 count and viral load. No significant difference was found between groups indicating that these groups had no influence on $CD4^+$ T cell count or HIV-2 VL in this community cohort.

The influence of KIR genes on markers of HIV-2 disease

All samples were successfully genotyped for the 15 KIR genes. A substantial number of HIV-2 infected subjects (n = 142) had serial HIV-2 VL measurements and 139 of these also had serial $CD4^+$ T cell measurements over the same period of time

as described earlier. Their CD4 counts and HIV-2 viral loads were analysed as for HLA class I above. We compared the mean sqrt $CD4^+$ T cell count as well as the mean log_{10}HIV-2 VL between groups of individuals positive for a particular KIR gene versus others (individuals who are HIV-2 positive but are negative for the gene of interest). The analysis showed that none of the individual *KIR* genes independently affected the course of HIV-2 disease once infection was established (**Table 4.5**). No significant difference was noted in the mean sqrt $CD4^+$ T cell count of people with a particular *KIR* gene and those without. The mean log_{10}HIV-2 VL was similar between individuals with a *KIR* gene of interest and those without.

Table 4.5: The effect of KIR genes on CD4 count and HIV-2 viral load

Genes	Mean sqrt CD4 (n = 139)	R	SE	p	Mean log HIV-2 VL (n = 142)	R	SE	p
2DL1+	24.20 (136)	0.91	0.54	0.51	2.95 (139)	1.15	0.08	0.51
2DL1-	26.64 (3)				2.56 (3)			
2DL2+	23.78 (65)	0.96	0.78	0.41	2.92 (65)	0.99	0.12	0.87
2DL2-	24.66 (74)				2.95 (77)			
2DL3+	24.25 (120)	1.00	0.57	1.00	2.96 (123)	1.06	0.09	0.50
2DL3-	24.24 (19)				2.79 (19)			
2DL4+	24.25 (139)	-	-	-	2.94 (142)	-	0.08	-
2DL4-	0.00 (0)				0.00 (0)			
2DL5+	23.69 (68)	0.96	0.76	0.30	2.86 (69)	0.95	0.12	0.37
2DL5-	24.79 (71)				3.01 (73)			
2DS1+	25.42 (31)	1.06	1.12	0.24	2.86 (31)	0.97	0.18	0.64
2DS1-	23.91 (108)				2.96 (111)			
2DS2+	23.63 (59)	0.96	0.81	0.32	2.96 (59)	1.01	0.13	0.78
2DS2-	24.70 (80)				2.92 (83)			
2DS3+	23.66 (32)	0.97	1.11	0.54	2.80 (33)	0.94	0.17	0.37
2DS3-	24.43 (107)				2.98 (109)			
2DS4+	24.29 (138)	1.31	0.53	0.37	2.94 (141)	1.51	0.08	0.32
2DS4-	18.60 (1)				1.95 (1)			
2DS5+	23.98 (40)	0.98	0.99	0.75	2.94 (40)	1.00	0.16	0.95

2DS5-	24.36 (99)				2.93 (102)			
3DL1+	24.29 (138)	1.31	0.53	0.37	2.94 (141)	1.51	0.08	0.32
3DL1-	18.60 (1)				1.95 (1)			
3DL2+	24.25 (139)	-	0.53	-	2.94 (142)	-	0.08	-
3DL2-	0.00 (0)				0.00 (0)			
3DL3+	24.25 (139)	-	0.53	-	2.94 (142)	-	0.08	-
3DL3-	0.00 (0)				0.00 (0)			
3DS1+	25.70 (22)	1.07	1.33	0.24	2.70 (22)	0.91	0.21	0.22
3DS1-	23.98 (117)				2.98 (120)			
2DP1+	24.18 (135)	0.91	0.54	0.45	2.96 (138)	1.33	0.08	0.14
2DP1-	26.57 (4)				2.22 (4)			

n: number of HIV-2$^+$ individuals; +: individuals with the allele of interest; -: individuals without the allele of interest; R: ratio of with/without; p: uncorrected p-values adjusted for age and gender and considered significant when ≤ 0.05.

Next, we grouped individuals based on whether or not they have a *KIR* gene together with its corresponding ligand(s) since we know that *KIR* is non-functional in the absence of the *HLA* class I gene(s) that encode for its ligand(s). We performed similar comparative analyses as above between people with a particular KIR-HLA compound genotype and those without. $CD4^+$ T cell count and HIV-2 VL measurements were similar between individuals with a compound genotype of interest and those lacking that genotype as illustrated in **Table 4.6** below. This suggests that the compound genotypes under investigation do not influence disease progression assessed using $CD4^+$ T cell count and viral load measurements.

Table 4.6: The effect of KIR-HLA compound genotypes on HIV-2 disease

Genotypes	Mean sqrt CD4 (n)	R	SE	p	Mean log VL (n)	R	SE	p

3DS1 : Bw4-80I/x	27.88 (5)	1.17	2.91	0.20	3.07 (5)	1.07	0.43	0.67
Bw6/Bw6	23.93 (47)		0.94		2.87 (47)		0.14	
3DS1 : Bw4-80I/x	27.88 (5)	1.15	2.74	0.20	3.07 (5)	1.05	0.44	0.76
Others	24.24 (130)		0.54		2.91 (133)		0.08	
2DS2 : C1/x	22.92 (43)	0.92	0.95	0.10	3.01 (43)	1.02	0.15	0.69
Others	24.83 (92)		0.65		2.94 (95)		0.10	
2DL2 : C1/x	23.07 (50)	0.93	0.89	0.10	2.95 (50)	1.00	0.14	0.92
Others	24.90 (85)		0.68		2.96 (88)		0.11	
2DS1 : C2/x	24.16 (18)	1.00	1.49	0.97	2.94 (18)	0.99	0.23	0.91
Others	24.23 (117)		0.58		2.96 (120)		0.09	
2DL1 : C2/x	24.15 (112)	0.98	0.60	0.76	2.97 (115)	1.02	0.09	0.83
Others	24.59 (23)		1.32		2.92 (23)		0.21	

n: number of HIV-2^+ individuals; R: ratio of individuals with the gene of interest / Others; p: uncorrected p-values adjusted for age and gender and considered significant when ≤ 0.05.

Discussion

HLA and *KIR* genes are located on separate chromosomes indicating that at meiosis both blocks of genes segregate independently into newly formed daughter cells. Differential susceptibility to progression to AIDS following HIV infection is well documented and understanding why some people progress rapidly to overt immunodeficiency shortly after infection with HIV-1 while others go on to become LTNPs (> 50% of HIV-2 patients) is a major goal in the field of HIV research. It is however becoming progressively clearer that viral characteristics, individual behaviour, and several other factors related to the environment and genetics of the host are key determinants of susceptibility to infection and disease progression[70 79 83 84 198].

In this study, we aimed to investigate the relative contribution of two very important host genetic factors - *HLA* and *KIR* genes - in HIV-2 disease in West African populations. Killer cell immunoglobulin-like receptors are important regulators of NK cell activity and determine the outcome of the first encounter between immune cells and the viral particles, which succeeded in breaching the first layer of the body's defence mechanisms and gain entrance into the body. Together with HLA alleles, KIR molecules determine the molecular targets of the cellular immune responses of a given host once infection is established. However, genetic polymorphisms in the MHC and KIR regions on chromosomes 6 and 19, respectively, which lead to variability in peptide-epitope binding, can effectively compromise the presentation of peptide-epitope complexes to T cells thereby interfering with the appropriate host immune responses against specific pathogens.

Ideally, genetic diversity at the HLA and KIR loci should provide protection against epidemic infections while the heterogeneity in immune response to a pathogen observed at the individual level ensures that escape mechanisms of an infectious agent are encountered and dealt with accordingly. This has clearly been demonstrated in several infectious diseases[61 66 67] and is likely to be true also in HIV infection. A growing number of associations from independent epidemiological studies have been reported between *HLA* and *KIR* genes and differential HIV-1 disease progression but very little is known of their role in HIV-2 disease manifestations. Here we report our findings on the possible association between *HLA* and *KIR* genes and HIV-2 disease using internationally accepted markers of disease progression in a group of individuals singly infected with HIV-2 in a community cohort in Caio, Guinea Bissau.

HLA-A genotypes did not influence HIV-2 disease progression

We generated HLA class I and KIR genotype data from 151 individuals singly infected with HIV-2. A significant proportion of these individuals were long-term non-progressors that have been living with the virus for more than 19 years with no AIDS-related symptoms. This is an indication that the majority of HIV-2 infected individuals are controlling the virus better than their HIV-1 counterparts. Analysis of individual *HLA* alleles and *KIR* genes comparing known indicators of HIV disease progression ($CD4^+$ T cell count and HIV-2 viral load) between those with the allele or *KIR* gene of interest and those without was also performed.

None of the 24 *HLA-A* alleles (four-digit alleles) and none of the 15 *HLA-A* group of alleles (two-digit alleles) was strongly associated with markers of disease progression in this sample of the West African population. The exception, however,

was a trend towards better control of HIV-2 viral replication observed with people having at least one copy of *HLA-A*7401* in their genotypes. HIV-2 infected individuals with this genotype had an overall higher $CD4^+$ T cell count than their counterparts who were negative for *A*7401* (although the difference did not reach significant level) and their overall mean log_{10}HIV-2 VL was also slightly lower compared to those without this genotype ($p = 0.09$). All *A*7401*-positive individuals in Caio ($n = 11$) were heterozygotes with a majority of them pairing with *A*2301* (6 out of 11). We did not see any influence of the only *A*23* allele found in this population (*A*2301*) on $CD4^+$ T cell count or HIV-2 VL. However, in two cohorts of 241 HIV-1 infected Caucasians, Kaslow and colleagues found that A*23 was associated with rapid progression to AIDS[116]. In another study, Chen and co-workers while working on a small paediatric cohort of 36 LTNP and 14 rapid progressors, showed that *A*2301* also associates with rapid disease progression following HIV-1 infection[199]. The other alleles in these *A*7401*-positive individuals included *A*3303, A*6802* and *A*2601.* These other alleles individually did not influence the rate of $CD4^+$ T cell decline or increase in HIV-2 viral load.

HLA-B alleles showed opposite effects with HIV-2 disease outcomes

We analysed individual *HLA-B* alleles as for *HLA-A* above, comparing $CD4^+$ T cell count and HIV-2 viral load in a group of HIV-2 patients positive for a particular HLA-B allele to those of the remainder (HIV-2 infected individuals negative for the allele of interest). Out of twenty two HLA-B groups of alleles (two-digit alleles), four of them showed significant differences between those possessing the allele of interest and those without. Two of these (*HLA-B*15* and *-B*49*) were associated with bad

prognosis (a sign for rapid disease progression) while the other two (*HLA-B*14* and -*B*44*) were associated with protection against rapid progression to AIDS.

The most striking finding was that of *HLA-B*15,* which clearly showed that individuals with this genotype had an overall low $CD4^+$ T cell count ($p = 0.003$) and an overall high HIV-2 viral load ($P = 0.001$) compared to their *HLA-B*15*-negative counterparts. This effect was attributed to the common *B*1503* subtype and not seen with the other two *B*15* subtypes *B*1510* and *B*1516.* The other *B* allele that tended to behave like *B*15* was *B*49.* HIV-2 infected individuals with this genotype had a significantly lower $CD4^+$ T cell count ($p = 0.03$) and a trend towards a high viral load ($p = 0.07$). This is in accordance with Kaslow findings that *B*49* was associated with rapid disease progression in homosexual Caucasians infected with HIV-1[116].

We found *B*14* and *B*44* carriers in our study population to be associated with long-term non-progression and protection against disease progression. In two independent studies of mixed high risk groups of Caucasians, Magierowska and Hendel showed that *B*14* was associated with non-progression[134 200]. Our finding of *B*14* being associated with non-progression also supports a more recent finding by Leligdowicz et al. who reported while working on the same Caio population that *B*14+* individuals chronically infected with HIV-2 elicited a stronger, highly avid and polyfunctional cell mediated immune responses of higher magnitude and broader specificity to HIV-2 *gag* peptides designed from p26 region of the HIV-2 proteome than people with other *HLA-B* alleles[201]. The association of *HLA-B*44* with control of viremia in HIV-2 infected patients is an interesting observation. Flores-Villanueva et al. reported in 2001 a similar finding in a cohort of thirty-nine HIV-1 seropositive subjects with known date of seroconversion[132]. They found that a significant proportion of subjects with HLA-*B*44* alleles were able to control viral replication and

thus progressed more slowly to full blown AIDS. On our part we found that amongst HIV-2 infected individuals, the group of people with at least one copy of *B*44* alleles (either *B*4403* or *B*4410*) were better controllers of HIV-2 replication as their overall viral load was significantly lower compared to the group without any of the *B*44* alleles. Their overall $CD4^+$ T cell count was however not different from that of the *B*44*-negative HIV-2 group.

A number of *HLA-B* alleles have been associated with progression to AIDS following HIV-1 infection in different populations and "at risk" groups with opposite or conflicting findings. But the most consistent findings have been with *B*27, B*57*, and a subset of *B*35* alleles known as *B*35*-Px including *B*3502, B*3503*, and *B*3504*. Alleles of *HLA-B*27* and *-B*57* offer a protective effect against rapid progression to AIDS in HIV-1 infected subjects. In Caio, *B*27* was completely absent while *B*57* was rare (AF = 1.6% in the general population). On the other hand, HLA alleles that have been categorised as *B*35*-Px have been consistently associated with rapid progression to AIDS. By definition, *B*35*-Px group of alleles comprises those *B*35* alleles that can accommodate peptides with proline (P) at position 2 and any other amino acid apart from tyrosine (Y) at position 9. In Caio, all the *B*35*-positive individuals were *B*3501,* which is not a member of the *B*35*-Px complex and we did not see any significant difference comparing their $CD4^+$ T cell count and HIV-2 viral load to those of *B*35*-negative HIV-2-positive subjects. This implies that most of the alleles found associated with HIV-1 disease by other investigators did not show similar effects in HIV-2 disease. This might be due to a number of factors ranging from the viral factors to environmental and/or population diversity. Most of the HIV-1 association studies have been conducted in Caucasian cohorts with known dates of seroconversion. Our study subjects were blacks from a seroprevalent cohort.

HLA-C and KIR-HLA compound genotypes and HIV-2 outcomes

We carried out similar analysis as described above for each of the *HLA-C* alleles detected in the study population. Two out of fourteen *HLA-C* groups of alleles (two-digit alleles) were associated with HIV-2 disease. *HLA-Cw*02,* which is in strong LD with *HLA-B*15* was associated with low $CD4^+$ T cell count but not significantly so with HIV-2 viral load. HIV-2 infected individuals who carried any of the *Cw*02* alleles (*Cw*0203* and *Cw*0210)* had a significantly low $CD4^+$ T cell count compared to those without the *Cw*02* genotype but their overall HIV-2 viral load, although high, was not statistically different from that obtained from *Cw*02*-negative HIV-2 infected counterparts. However, further analyses discriminating between *Cw*02* subtypes revealed clearly that *HLA-Cw*0210* was strongly associated with bad prognosis. Individuals who were HIV-2 positive and carrying at least one copy of *Cw*0210* were more likely to progress with the disease faster than those negative for that allele. Their $CD4^+$ T cell count was much lower ($p < 0.001$) and their HIV-2 $\log_{10}$ VL much higher ($p < 0.001$) compared to others (HIV-2 positive *HLA-Cw*0210*-negative). This effect was not seen with the other *Cw*02* allele (*HLA-Cw*0203).* The second *HLA-C* allele, which was found to associate with HIV-2 disease in this predominantly Manjako population was *HLA-Cw*07.* The group of individuals positive for *Cw*07* had an overall higher HIV-2 viral load than those negative for *Cw*07,* indicating that *Cw*07* could be associated with rapid disease progression. This demonstrates that this group of individuals might not be able to control viral replication even though their $CD4^+$ T cell count was similar to that of the *Cw*07*-negative group. All other *HLA-C* alleles did not affect the $CD4^+$ T cell count and HIV-2 viral load dynamics. Grouping all *HLA-C* alleles into C1 and C2 did not change the nature of the association.

Certain HLA-B and -C molecules are used as ligands by some KIR molecules. The interaction between a KIR and its corresponding ligand(s) determines the type of signal to be delivered to the NK cell harbouring that particular KIR on its cell surface. An epistatic interaction has been described between KIR3DS1-HLA-Bw4-80I and HIV-1 disease[169]. In the present study, we also grouped individuals based on whether or not they have a KIR and its corresponding ligand and studied the effect of the identified compound genotypes on the rate of $CD4^+$ T cell count decline and increase in HIV-2 viral load. We observed that none of the KIR-HLA combinations was associated with HIV-2 disease.

CHAPTER 5 : KIR GENE PROFILES AND HLA CLASS I HAPLOTYPES IN CAIO, GUINEA BISSAU

In order to determine the relative contribution of individual *KIR* gene profiles in HIV-2 acquisition and/or disease progression, we reconstructed the *KIR* gene profile of each study participant in the order in which the KIR genes are arranged on chromosome 19. We subsequently grouped them into the conventionally accepted *KIR* profiles "A" and "B" and then analysed the profile pool stratified by HIV status. Similarly, we carried out *HLA* class I haplotypes reconstruction using Arlequin v3.11[202] and estimated the frequencies of two- and three-locus haplotypes in our study population.

The categorisation of KIR genes into profiles "A" or "B" depends on the number of activating and inhibitory *KIR* genes present in an individual without taking into consideration whether they appear on only one or both haplotypes. Haplotype A comprises mainly genes coding for inhibitory receptors (KIR2DL and KIR3DL) - in the absence of *KIR2DL2* - with or without KIR2DS4, which is expressed by almost every NK cell. Haplotype B is a mixture of genes encoding both activating (KIR2DS and KIR3DS) and inhibitory KIR molecules. *KIR* gene profile A (two copies of the A haplotypes) lacks all activating *KIR* genes except *2DS4* and also lacks the inhibitory gene *2DL2*. *KIR* gene profile B, has a variable number of activating and inhibitory *KIR* genes; when a B profile is present, it is difficult to determine whether the KIR haplotypes present are BB or AB except the genes are sequenced and their nucleotides composition analysed. A substantial number of genes on A haplotypes are highly polymorphic, as are a smaller subset of genes on B haplotypes. The

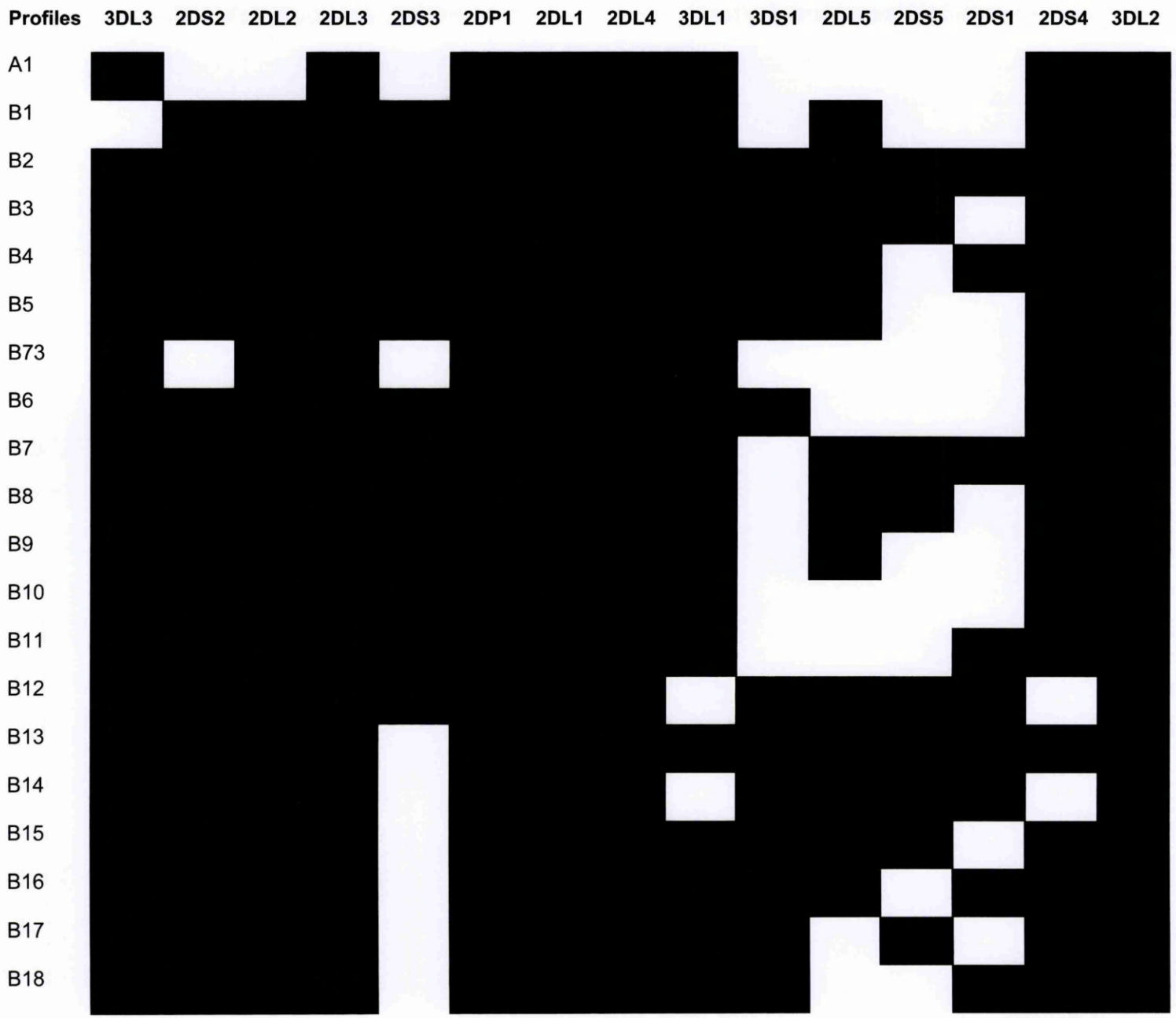
majority of B haplotypes exhibit extensive variation in gene content and are said to be polygenic in nature.

KIR gene profiles in Caio

A *KIR* gene profile represents a combination of inhibitory, activating, as well as pseudogenes that are present on at least one haplotype in a given individual. They are classified into two groups (A and B) depending on the number and type of individual haplotypes making the profile. KIR profile A comprises two copies of A haplotypes while profile B can either be AB or BB haplotypes. Analysis of our *KIR* gene pool revealed 75 different profiles (2 A and 73 B profiles) with 26 of them being present in at least 1% of the study population (**Table 5.1**). The most frequent *KIR* gene profile was one composed of two A haplotypes (A1), which was present in 32% of our samples. A1 was slightly, but not significantly over-represented in the infected group (**Figure 5.1**) (38% vs. 29%, p = 0.062). The second "A" haplotype (A2) was seen in only one sample and was very similar to A1 except that it lacked *2DL3*.

Of the 73 B profiles, B1 was the most common; 12% in the general population, 10% in the HIV-2 infected group, and 12% in the control group. B2 was the second most common B haplotype: 7% in the general population, 4% and 9% in the HIV-2 infected and HIV negative groups, respectively. None of the differences in *KIR* gene profile frequencies between cases and controls reached significant level. One of the B profiles closely resembled A1, the only difference being the presence of *KIR2DL2* (B73 profile).

Table 5.1: The KIR gene profiles in the Caio population



Profiles	3DL3	2DS2	2DL2	2DL3	2DS3	2DP1	2DL1	2DL4	3DL1	3DS1	2DL5	2DS5	2DS1	2DS4	3DL2	n	Freq
A1	1	0	0	1	0	1	1	1	1	0	0	0	0	1	1	163	0.32
B1	0	1	1	1	1	1	1	1	1	0	1	0	0	1	1	60	0.12
B2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	37	0.07
B3	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	15	0.03
B4	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	15	0.03
B5	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	14	0.03
B73	1	0	1	1	0	1	1	1	1	0	0	0	0	1	1	14	0.03
B6	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	12	0.02
B7	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	12	0.02
B8	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	11	0.02
B9	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	10	0.02
B10	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	10	0.02
B11	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	9	0.02
B12	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	8	0.02
B13	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	8	0.02
B14	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	7	0.01
B15	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	7	0.01
B16	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	6	0.01
B17	1	1	1	1	0	1	1	1	1	0	1	1	0	1	1	6	0.01
B18	1	1	1	1	0	1	1	1	1	0	0	0	1	1	1	5	0.01

Open box: gene is absent; closed box: gene is present; n: number of individuals with the profile of interest. Only profiles that are present in at least 1% of the study population are shown.

Univariate analyses were performed to see if any of the *KIR* profiles detected in this study could influence HIV-2 acquisition or predict progression to AIDS, but none of them impacted on susceptibility/resistance to HIV-2 as their frequencies were similar between infected and uninfected groups. There was also no significant association with disease progression assessed using markers of HIV disease progression (absolute $CD4^+$ T cell count and HIV-2 viral load). KIR profile frequencies were very similar between cases and controls and also between males and females.

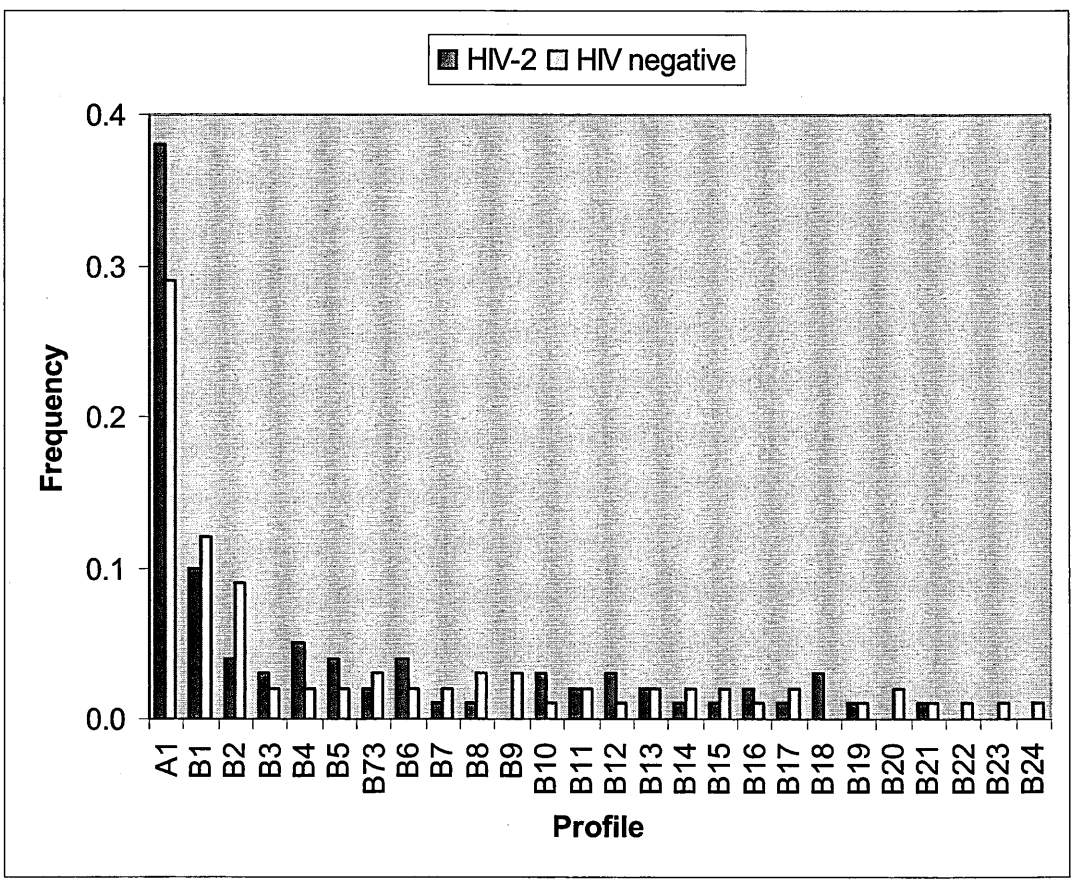


Figure 5.1: Most frequent KIR gene profiles found in Caio

HLA class I haplotypes in the Caio population

The test for Hardy-Weinberg (H-W) equilibrium was performed with Arlequin v3.11. Using a Fisher's Exact test, at the intra-population level, all three loci were in equilibrium (p-values = 0.18517, 0.25241, and 0.10134 for HLA-A, -B, and -C loci, respectively). **Table 5.2** shows the frequency of two-locus HLA class I haplotypes detected in the Caio community. Haplotypes with frequencies less than one percent are not shown in this table. *HLA B*1503-Cw*0210* was the most frequent two-locus haplotype with a frequency of 9.8% in the general population. Seven other haplotypes had frequencies above 5% and together made up 48.3% of all B-C haplotypes. Only two A-B haplotypes (*A*330301-B*1510* and *A*2301-B*1503*), and two A-C haplotypes (*A*330301-Cw*0304* and *A*2301-Cw*0210*) had frequencies greater than or equal to 5%. Analysis of the three-locus *HLA* class I data showed that *HLA-A*330301-B*1510-Cw*0304* (6.2%) and *A*2301-B*1503-Cw*0210* (5.3%) were by far the most frequent three-locus haplotypes in our study population. Other three-locus haplotypes with more than 1% population frequency are shown below in **Table 5.3**. Haplotype frequencies were not statistically different comparing males and females and were similar between cases and controls in this community.

Table 5.2: Common HLA class I two-locus haplotypes in the Manjako population

HLA-A-B		*HLA-A-C*		*HLA-B-C*	
Haplotype	**Frequency[#]**	**Haplotype**	**Frequency[#]**	**Haplotype**	**Frequency[#]**
*A*330301-B*1510*	0.072	*A*330301-Cw*0304*	0.080	*B*1503-Cw*0210*	0.098
*A*2301-B*1503*	0.057	*A*2301-Cw*0210*	0.048	*B*1510-Cw*0304*	0.079
*A*260101-B*080101*	0.039	*A*260101-Cw*0304*	0.041	*B*350101-Cw*040101*	0.075
*A*2301-B*4901*	0.035	*A*2301-Cw*0701*	0.033	*B*530101-Cw*040101*	0.065
*A*260101-B*5801*	0.030	*A*260101-Cw*0701*	0.029	*B*4901-Cw*0701*	0.062
*A*010101-B*080101*	0.025	*A*3002-Cw*040101*	0.029	*B*5801-Cw*0701*	0.053
*A*330301-B*130201*	0.025	*A*7400-Cw*040101*	0.028	*B*5801-Cw*0302*	0.051
*A*330301-B*5801*	0.025	*A*330301-Cw*0302*	0.026	*B*080101-Cw*0304*	0.044
*A*7400-B*530101*	0.024	*A*330301-Cw*0804*	0.026	*B*140201-Cw*0802*	0.039
*A*6801-B*1503*	0.021	*A*010101-Cw*0102*	0.025	*B*080101-Cw*0102*	0.035
*A*2301-B*350101*	0.018	*A*2301-Cw*0302*	0.024	*B*130201-Cw*0804*	0.028
*A*010101-B*530101*	0.017	*A*3002-Cw*050101*	0.019	*B*530101-Cw*060201*	0.026
*A*3002-B*1801*	0.017	*A*6801-Cw*0210*	0.019	*B*1801-Cw*050101*	0.025
*A*2301-B*5801*	0.014	*A*2301-Cw*060201*	0.018	*B*1503-Cw*040101*	0.014
*A*7400-B*140201*	0.014	*A*3002-Cw*0701*	0.018	*B*0702-Cw*1505*	0.012
*A*2301-B*530101*	0.012	*A*2301-Cw*040101*	0.017	*B*4403-Cw*030301*	0.012

*A*3002-B*350101*	0.012	*A*680201-Cw*040101*	0.017	*B*0702-Cw*070201*	0.011
*A*330301-B*350101*	0.012	*A*010101-Cw*040101*	0.015	*B*350101-Cw*0701*	0.011
*A*330301-B*4403*	0.012	*A*680201-Cw*0701*	0.015	*B*4403-Cw*0304*	0.011
*A*0201-B*5201*	0.011	*A*0201-Cw*040101*	0.014	*B*5101-Cw*160101*	0.011
*A*3002-B*1503*	0.011	*A*0201-Cw*160101*	0.014	*B*5201-Cw*160101*	0.011
*A*3002-B*530101*	0.011	*A*7400-Cw*0802*	0.014	*B*080101-Cw*0701*	0.010
*A*6801-B*350101*	0.011	*A*0102-Cw*0701*	0.012		
*A*680201-B*530101*	0.011	*A*0202-Cw*0701*	0.011		
*A*0102-B*4901*	0.010	*A*330301-Cw*0701*	0.011		
*A*0201-B*350101*	0.010	*A*0202-Cw*0210*	0.010		
*A*0202-B*4901*	0.010	*A*3301-Cw*0701*	0.010		
*A*330301-B*0702*	0.010	*A*330301-Cw*040101*	0.010		
*A*7400-B*1503*	0.010				

#: haplotypes with frequencies ≥ 0.01

Table 5.3: Common HLA class I three-locus haplotypes in Caio

Haplotype	Frequency[#]
*A*330301-B*1510-Cw*0304*	0.062
*A*2301-B*1503-Cw*0210*	0.053
*A*2301-B*4901-Cw*0701*	0.030
*A*260101-B*080101-Cw*0304*	0.030
*A*010101-B*080101-Cw*0102*	0.025
*A*260101-B*5801-Cw*0701*	0.024
*A*7400-B*530101-Cw*040101*	0.024
*A*330301-B*130201-Cw*0804*	0.021
*A*2301-B*5801-Cw*0302*	0.018
*A*330301-B*5801-Cw*0302*	0.018
*A*6801-B*1503-Cw*0210*	0.018
*A*3002-B*1801-Cw*050101*	0.017
*A*010101-B*530101-Cw*040101*	0.014
*A*2301-B*350101-Cw*040101*	0.014
*A*3002-B*350101-Cw*040101*	0.014
*A*7400-B*140201-Cw*0802*	0.012
*A*0201-B*5201-Cw*160101*	0.011
*A*3002-B*1503-Cw*040101*	0.011
*A*330301-B*350101-Cw*040101*	0.011
*A*0102-B*4901-Cw*0701*	0.010
*A*0201-B*350101-Cw*040101*	0.010
*A*0202-B*4901-Cw*0701*	0.010
*A*680201-B*530101-Cw*040101*	0.010

[#]: haplotypes with frequencies ≥ 0.01

Discussion:

In humans, the *KIR* gene complex is located within a region that undergoes rapid expansion and contraction over time, resulting in extensive variability in gene contents across many haplotypes. The extreme polymorphism at the KIR loci is thought to be beneficial since they provide protection against a wide range of pathogens and diseases. Two KIR haplotypes have been defined to regroup and harmonise the huge amount of haplotypic data continually being generated by various investigators. Certain *KIR* genes have consistently been found on virtually all haplotypes and have been called "framework genes" while others have been reported on only a few haplotypes[163]. The so called framework genes include *KIR2DL4, 3DL2, 3DL3*, and *3DP1*.

KIR gene profile "A" predominates in the Caio population

KIR gene profile frequencies in Caio were determined by direct counting after we individually assigned a profile to each and every one of the study participants. As is usual in populations worldwide, B profiles were highly diversified on the basis of gene content with 73 distinct B profiles (B1-B73) while A profiles were very much conserved with only two A profiles being present (A1 and A2). The only difference between the two A profiles was the absence of the inhibitory gene *KIR2DL3* in profile A2. In contrast to Caucasoid populations, *KIR* gene profile A1 was by far the commonest in this Manjako population with 163 (32%) subjects carrying this profile comprising six inhibitory *KIR* genes: *3DL3, 2DL3, 2DL1, 2DL4, 3DL1,* and *3DL2* in addition to the single activating gene *2DS4,* and the pseudogene *2DP1.* In most populations worldwide, the frequencies of A and B profiles are evenly distributed. The second most common profile was a B profile (B1), present in 12% of the study

population. The difference in gene content between A1 and B1 profiles was the presence of two genes encoding the activating receptors 2DS2 and 2DS3, two others encoding the inhibitory receptors 2DL2 and 2DL5, and the absence of the gene encoding 3DL3 in profile B1. Thirty-seven individuals (7%) had all 15 *KIR* genes in a B profile named B2. Overall, twenty-six (26) *KIR* gene profiles (25 B and 1 A, respectively) were present in at least 1% of the study population.

KIR profiles and HLA class I haplotypes did not influence HIV-2 infection or disease

In an attempt to understand the relative contribution of the *KIR* gene profiles on HIV-2 acquisition, we stratified our analysis of *KIR* gene profile pool by HIV status and compared the frequencies of individual gene profiles between HIV-2 infected and uninfected control groups. We observed that the frequency of A1 was slightly higher in the infected group than in controls but the difference was not statistically significant. As expected, some of the B profiles were slightly over-represented in controls while others were more frequent in cases. Overall, none of the *KIR* gene profiles found in this community cohort significantly influenced susceptibility or resistance to HIV-2 infection.

Similarly, we looked at the potential role of *KIR* gene profiles in the course of HIV-2 disease using universal markers of HIV disease progression. We compared the $CD4^+$ T cell count and viral load data between groups of individuals whom are carriers of a *KIR* gene profile of interest and those negative for that profile. For all profiles, we observed no significant influence on $CD4^+$ T cell dynamics after comparing the overall $CD4^+$ T cell count between individuals with a profile of interest to those without. Similarly, none of the *KIR* gene profiles could influence viral

replication since no significant difference was found comparing viral load measurements between groups with any of the *KIR* gene profile and those without the profile of interest. This is an indication that none of the *KIR* gene profiles detected in this predominantly Manjako population can influence disease progression following HIV-2 infection.

Similarly, we determined the two- and three-loci HLA class I haplotypes in this population and carried out similar analyses as for the *KIR* gene profiles above. None of the two-loci and none of the three-loci haplotypes were associated with HIV-2 acquisition. There was also no indication of them influencing disease progression after HIV-2 infection has been established.

CHAPTER 6 : HLA CLASS I AND KIR GENE FREQUENCIES AND HIV-2 ANTIBODY STATUS IN FAJARA, THE GAMBIA

This chapter describes the frequencies of classical *HLA* class I genes (*HLA-A, -B* and *–C)* as well as those of 15 *KIR* genes in 600 adults Gambians living in and around the Greater Banjul area and attending the MRC Genito-Urinary Medicine (GUM) clinics in Fajara. The relationship between *HLA* alleles and HIV-2 status defined by antibody levels was examined. Analysis of *KIR* genes profiles in relation to susceptibility or resistance to HIV-2 infection was performed. The effect of *HLA-KIR* combinations on HIV-2 acquisition was also explored.

HLA alleles and genotype frequencies in Fajara, The Gambia

A total of 600 samples collected from consented adult Gambians were typed for *HLA* class I and 15 *KIR* genes. HLA typing was done by SBT using the BigDye Terminator v3.1 (Applied Biosystems, USA) while KIR typing was done by the PCR-SSP method. Analysis of HLA sequences was done with "Assign 400" software (Conexio Genomic, Western Australia) and verified both manually and with other conventional sequence analysis software such as "Sequence Analysis v5.2" (Applied Biosystems, USA), and CLC Main Workbench v5 (CLC BIO A/S, Aarhus C, Denmark). A few samples could not be assigned a genotype because their sequence traces did not precisely match known HLA consensus sequences, which are available in the public HLA and immunogenetics databases[94]. A deviation from a known consensus sequence by at least one nucleotide was considered to be a mismatch and the result labelled indeterminate. Such samples were repeated and if the

mismatch persisted after two repeats, the samples were classified as indeterminate. We suspect that these are most likely new *HLA* class I alleles that are yet to be described in other populations and therefore have not made it yet to the public databases. Due to limitation of funds, we could not further characterise these supposedly new alleles in the present study. Our HLA data analysis therefore was based on 589 HLA-A, 583 HLA-B, and 595 HLA-C genotypes.

Table 6.1 summarises the allele frequencies of classical HLA class I (A, B, and C) loci (4- to 6-digit alleles) found in the Fajara clinical cohort. Individual allele frequencies were calculated as a proportion of the total number of alleles determined for the locus of interest. The genotype frequency of a particular allele was estimated as the proportion of individuals positive for the allele of interest compared to the total number of individuals with a genotype result assigned for that locus. Overall, 28 *HLA-A*, *50 HLA-B* and 23 *HLA-C* alleles (≥4-digits) were detected in this sample of the Gambian population. **Figure 6.1** shows the frequency of two-digit HLA class I results. The number of individuals with two copies of the same allele (homozygosity) was highest at the C locus: *A-A* = 5.60% (n=33); *B-B* = 2.92% (n=17); and *C-C* = 9.41% (n=56). Individuals homozygous at two loci were rare - only 2 individuals were homozygotes A-B (0.35%) and 2 others homozygotes B-C (0.35%), while just a single person was homozygote A-C (0.17%). No three-locus homozygosity was seen in our samples (**Table 6.2**).

Table 6.1: HLA-A, -B, and -C allele frequencies in the Fajara population

HLA-A Alleles	Frequency (n=1178)	HLA-B Alleles	Frequency (n=1166)	HLA-C Alleles	Frequency (n=1190)
A*010101	0.048 (56)	B*070201	0.047 (55)	Cw*0102	0.021 (25)
A*0102	0.017 (20)	B*070501	0.011 (13)	Cw*0201	0.001 (1)
A*0103	0.001 (1)	**B*080101**	**0.087 (102)**	Cw*020202	0.021 (25)
A*0105	0.001 (1)	B*1301	0.001 (1)	**Cw*0210**	**0.096 (114)**
A*020101	**0.080 (94)**	B*130201	0.004 (5)	**Cw*0302**	**0.061 (72)**
A*0202	0.045 (53)	B*1401	0.001 (1)	Cw*030301	0.010 (12)
A*0205	0.015 (18)	B*140201	0.029 (34)	**Cw*0304**	**0.097 (116)**
A*030101	0.037 (44)	B*1403	0.001 (1)	**Cw*040101**	**0.187 (222)**
A*2301	**0.171 (201)**	**B*1503**	**0.096 (112)**	Cw*0403	0.001 (1)
A*240201	0.013 (15)	B*1510	0.032 (37)	Cw*050101	0.022 (26)
A*260101	**0.056 (66)**	B*1516	0.015 (17)	**Cw*060201**	**0.064 (76)**
A*290101	0.002 (2)	B*1518	0.001 (1)	**Cw*0701**	**0.109 (130)**
A*290201	0.017 (20)	B*1801	0.021 (25)	Cw*070201	0.030 (36)
A*3001	**0.065 (76)**	B*2703	0.010 (12)	Cw*0802	0.034 (40)
A*3002	**0.069 (81)**	B*2705	0.005 (6)	Cw*0804	0.008 (9)
A*3004	0.003 (3)	**B*3501**	**0.130 (151)**	Cw*120301	0.010 (12)
A*310102	0.006 (6)	B*350801	0.001 (1)	Cw*140201	0.014 (17)
A*3201	0.022 (26)	B*3510	0.001 (1)	Cw*1403	0.002 (2)
A*3301	**0.062 (73)**	B*370101	0.007 (8)	Cw*150201	0.001 (1)
A*330301	**0.074 (87)**	B*3901	0.002 (2)	Cw*1505	0.023 (27)
A*3402	0.036 (42)	B*3910	0.013 (15)	**Cw*160101**	**0.121 (144)**
A*3601	0.002 (2)	B*400201	0.005 (6)	**Cw*17[a]**	**0.056 (67)**
A*6601	0.013 (15)	B*4016	0.001 (1)	Cw*18[a]	0.013 (15)
A*6603	0.003 (3)	B*4102	0.010 (12)		

A*6801	0.041 (48)	B*4103	0.002 (2)
A*680201	**0.058 (68)**	B*4201	0.018 (21)
A*7401	0.036 (42)	B*4202	0.022 (26)
A*8001	0.013 (15)	B*440201	0.001 (1)
		B*4403	0.015 (18)
		B*4410	0.004 (5)
		B*45[a]	0.021 (25)
		B*470101	0.001 (1)
		B*4901	0.039 (46)
		B*5001	0.017 (20)
		B*5002	0.001 (1)
		B*5101	0.023 (27)
		B*5104	0.001 (1)
		B*5109	0.002 (2)
		B*5201	0.016 (19)
		B*530101	**0.119 (139)**
		B*5501	0.002 (2)
		B*5601	0.008 (9)
		B*5605	0.001 (1)
		B*5702	0.005 (6)
		B*570301	0.006 (7)
		B*5704	0.001 (1)
		B*5801	**0.076 (89)**
		B*7801	**0.055 (64)**
		B*780202	0.001 (1)
		B*8201	0.011 (13)

n: total number of alleles for the locus of interest; [a]: alleles that could not be discriminated with exons 2 and 3 sequences (B*45 = *B*4501/07*, *Cw*17* = *Cw*1701/Cw*1702/Cw*1703,* and *Cw*18* = *Cw*1801/Cw*1802*).

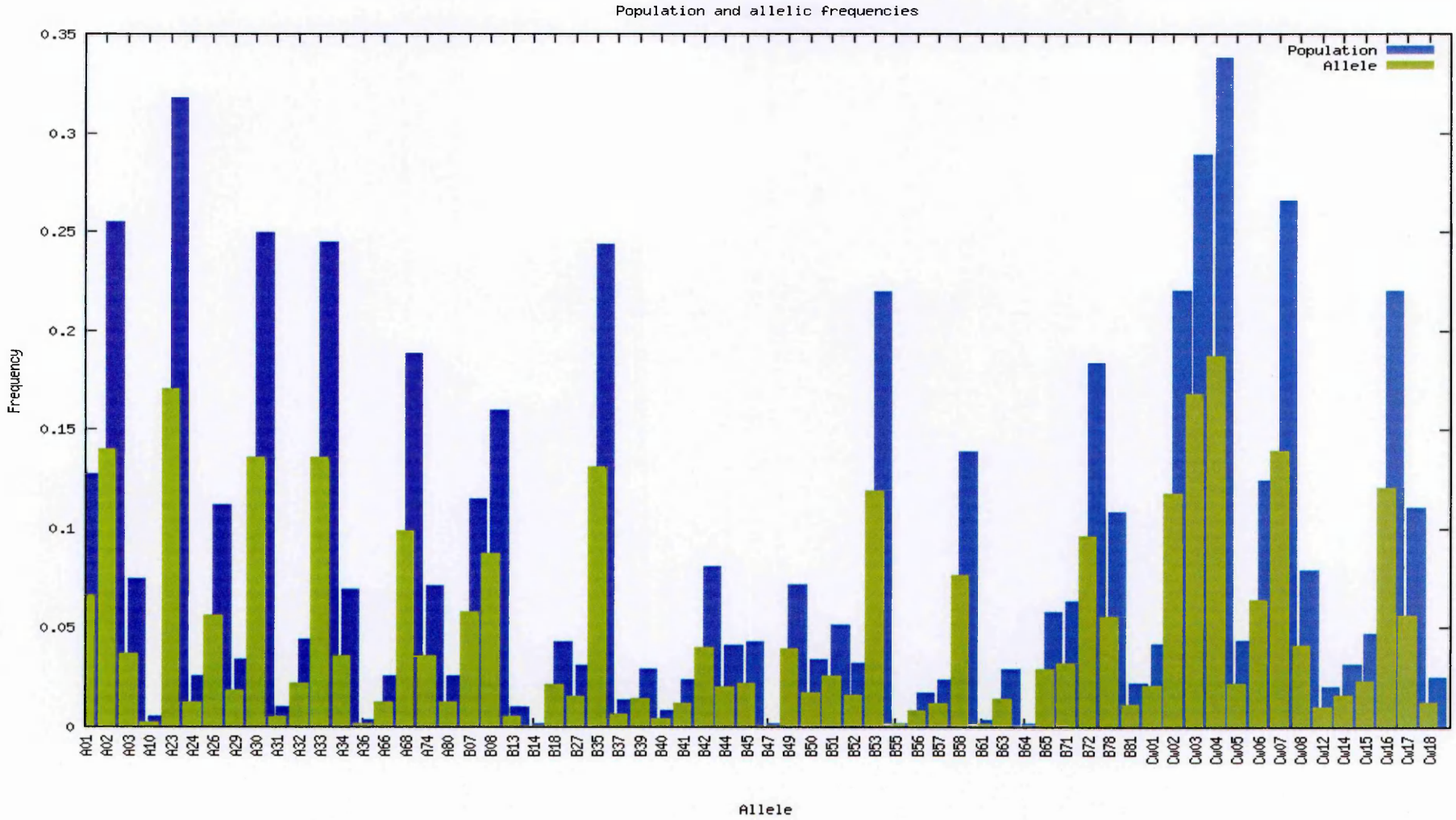


Figure 6.1: Allele and genotype frequencies of classical HLA class I in Fajara

Table 6.2: Zygosity at HLA-A, -B, and -C loci

Locus	Percent homozygous (n)	Percent heterozygous (n)
A (n=589)	5.60 (33)	94.40 (556)
B (n=583)	2.92 (17)	97.08 (566)
C (n=595)	9.41 (56)	90.59 (539)
A-B (n=564)	0.35 (2)	99.65 (562)
A-C (n=578)	0.17 (1)	99.83 (577)
B-C (n=572)	0.35 (2)	99.65 (570)
A-B-C (n=556)	0 (0)	100.00 (556)

n: number of genotypes

HLA-A alleles and susceptibility to HIV-2 infection in Fajara

Twenty eight different *HLA-A* alleles (4- to 6-digits) were detected from 589 individuals (**Table 6.1**). The most frequent *HLA-A* allele was *A*2301* (17.1%) followed by *A*020101* (8.0%) and *A*330301* (7.4%). Five other *A* alleles were also relatively frequent with frequencies ranging from 5.6% to 6.9% (*A*260101, A*3001, A*3002, A*3301,* and *A*680201*). The eight most frequent alleles at the A locus with at least 5% allele frequency at the population level made up 63.5% of all *HLA-A* frequencies in this sample of the Gambian population. We re-sequenced all samples containing A*74 (n = 40) to include exon 1 since we were unable to discriminate between *A*7401* and *A*7402,* which have similar nucleotide compositions at exons 2 and 3. The analysis showed that they were all A*7401 with the amino acid in codon 23 being R (Arginine). The most diverse group of alleles (2-digits) was *HLA-A*01*

with four different subtypes: *A*010101, A*0102, A*0103,* and *A*0105.* Their overall population frequency was 6.7%, with *A*010101* (4.8%) predominating.

Univariate analysis was performed to compare HLA-A genotype frequencies between 189 HIV-2 infected and 250 uninfected controls (**Table 6.3**). Comparisons between cases and controls were adjusted for gender, ethnicity and age at date of sample collection. In all instances, genotype frequencies were similar between cases and controls.

Table 6.3: HLA-A and susceptibility to HIV-2 infection

HLA-A Alleles	HIV-2 (n = 189)	HIV negative (n = 250)	OR (95% CI)	p^{Φ}
A*01	0.106 (20)	0.120 (30)	0.86 (0.43-1.70)	0.663
A*02	0.243 (46)	0.260 (65)	0.84 (0.51-1.39)	0.503
A*03	0.079 (15)	0.072 (18)	1.19 (0.54-2.62)	0.662
A*23	0.360 (68)	0.304 (76)	1.27 (0.81-2.01)	0.303
A*24	0.021 (4)	0.024 (6)	1.11 (0.26-4.76)	0.893
A*26	0.138 (26)	0.096 (24)	1,59 (0.81-3.13)	0.180
A*29	0.016 (3)	0.032 (8)	0.78 (0.19-3.26)	0.738
A*30	0.222 (42)	0.284 (71)	0.67 (0.40-1.11)	0.117
A*31	0.010 (2)	0.012 (3)	2.05 (0.31-13.50)	0.455
A*32	0.063 (12)	0.032 (8)	2.13 (0.76-6.00)	0.152
A*33	0.217(41)	0.276 (69)	0.67 (0.40-1.11)	0.119
A*34	0.058 (11)	0.080 (20)	1.02 (0.43-2.39)	0.968
A*36	0.005 (1)	0.004 (1)	1.36 (0.05-37.75)	0.858
A*66	0.048 (9)	0.024 (6)	1.85 (0.54-6.29)	0.325
A*68	0.169 (32)	0.176 (44)	0.96 (0.55-1.70)	0.901

A*74	0.095 (18)	0.060 (15)	1.41 (0.64-3.12)	0.397
A*80	0.026 (5)	0.024 (6)	0.66 (0.15-2.89)	0.580

$^{\Phi}$: p values adjusted for gender, ethnicity and age but not for multiple comparisons

HLA-B alleles and HIV-2 acquisition in Fajara

The B locus was the most diverse in this population with 32 different serologically defined groups (2-digit alleles) and 50 alleles (4- to 6-digits) detected in 583 adults (**Table 6.1**). The most frequent allele at the B locus was *B*3501* (AF=12.6%, GF=23.7%), closely followed by *B*530101* (AF=11.8%, GF=21.9%) and *B*1503* (AF=9.9%, GF=18.8%). Other frequent alleles include *B*080101, B*5801*, and *B*7801* with allele frequencies ranging from 5.4% to 8.9%. Only six alleles at the B locus had an allele frequency above 5% that cumulated to 56.3% of all *B* alleles. The most common and diverse group of alleles was *HLA-B*15* with four subtypes (*B*1503, B*1510, B*1516*, and *B*1518*) and a cumulative allele frequency of 14.7% while their combined genotype frequency was 28.4%. *B*1503* represented 67% of all *B*15* subtypes. The second most diverse group of alleles was *HLA-B*35* (AF=12.8%, GF=24.1%). Three *B*35* subtypes were found in our samples: *B*3501, B*350801,* and *B*3510*. The predominant subtype was *B*3501* representing 98.6% of all *B*35* subtypes.

A comparative analysis was done to explore differences between HIV-2 infected (n = 203) and uninfected control subjects (n = 247). Of note, individuals infected with HIV-1 or dually infected with both HIV-1 and HIV-2 were excluded from this study. In a multivariate model adjusting for gender, ethnicity, and age, we found that two *HLA-B* alleles (B*14 and B*57) were associated with protection against HIV-2 acquisition (**Table 6.4**). Individuals positive for *HLA-B*14* were less likely to be HIV-2

infected (OR = 0.31, 95% CI: 0.11-0.90, p = 0.031). Similarly, the number of individuals with at least one copy of *HLA-B*57* in their genotypes was more in the uninfected group than in the HIV-2 infected group (OR = 0.13, 95% CI: 0.02-0.80, p = 0.029).

Table 6.4: HLA-B alleles and susceptibility to HIV-2 infection

HLA-B Alleles	HIV-2 (n = 203)	HIV negative (n = 247)	OR (95% CI)	p*
B*07	0.128 (26)	0.089 (22)	1.36 (0.69-2.66)	0.370
B*08	0.177 (36)	0.158 (39)	1.21 (0.68-2.14)	0.518
B*14	**0.034 (7)**	**0.077 (19)**	**0.31 (0.11-0.90)**	**0.031**
B*15	0.300 (61)	0.279 (69)	1.22 (0.76-1.94)	0.409
B*18	0.069 (14)	0.036 (9)	1.54 (0.56-4.23)	0.398
B*27	0.030 (6)	0.032 (8)	1.20 (0.35-4.15)	0.777
B*35	0.271 (55)	0.215 (53)	1.42 (0.86-2.34)	0.165
B*37	0.010 (2)	0.020 (5)	0.79 (0.14-4.60)	0.796
B*39	0.020 (4)	0.032 (8)	0.27 (0.07-1.08)	0.065
B*41	0.025 (5)	0.028 (7)	0.73 (0.21-2.60)	0.631
B*42	0.089 (18)	0.077 (19)	1.05 (0.49-2.26)	0.900
B*44	0.030 (6)	0.049 (12)	0.81 (0.28-2.32)	0.691
B*45	0.049 (10)	0.049 (12)	1.19 (0.45-3.16)	0.731
B*49	0.059 (12)	0.073 (18)	0.83 (0.35-1.98)	0.676
B*50	0.025 (5)	0.036 (9)	0.89 (0.26-3.03)	0.855
B*51	0.059 (12)	0.046 (12)	1.21 (0.46-3.18)	0.692
B*52	0.030 (6)	0.028 (7)	1.11 (0.29-4.15)	0.881
B*53	0.212 (43)	0.206 (51)	0.93 (0.56-1.56)	0.790

B*56	0.025 (5)	0.016 (4)	2.14 (0.47-9.67)	0.323
B*57	**0.010 (2)**	**0.040 (10)**	**0.13 (0.02-0.80)**	**0.028**
B*58	0.118 (24)	0.121 (30)	1.20 (0.62-2.32)	0.594
B*78	0.084 (17)	0.130 (32)	0.60 (0.30-1.20)	0.148
B*82	0.020 (4)	0.032 (8)	0.81 (0.22-3.04)	0.756

*: p values adjusted for gender, ethnicity and age at the time of sample collection; in bold are alleles with significantly different frequencies in cases and controls

We next grouped all *B* alleles into the two mutually exclusive categories Bw4 and Bw6 based on their amino acid composition at positions 77-83. The Bw4 group was further subdivided into Bw4-80I and Bw4-80T depending on whether the amino acid at position 80 was an isoleucine (I) or a threonine (T). *HLA-B Bw4-80I* alleles serve as ligands for KIR3DL1 and possibly KIR3DS1 molecules although the latter is yet to be proven experimentally[184-186]. Our analysis showed that none of these categories (Bw4, Bw4-80I, and Bw4-80T) could significantly influence susceptibility or resistance to HIV-2 (**Table 6.5**).

Table 6.5: The effect of Bw4-80I and Bw4-80T on HIV-2 infection

Alleles	**HIV-2 (n = 203)**	**HIV negative (n = 247)**	**OR (95% CI)**	**p***
Bw4+	0.537 (109)	0.580 (143)	0.88 (0.57-1.35)	0.550
Bw6/Bw6	0.463 (94)	0.421 (104)		
Bw4-80I+	0.473 (96)	0.502 (124)	0.89 (0.57-1.39)	0.610
Bw6/Bw6	0.463 (94)	0.421 (104)		
Bw4-80T+	0.084 (17)	0.113 (28)	0.82 (0.39-1.73)	0.596
Bw6/Bw6	0.463 (94)	0.421 (104)		

*: p values adjusted for gender, ethnicity and age

HLA-C and susceptibility to HIV-2 infection in Fajara

A total of 23 *HLA-C* alleles (4- to 6-digits) were found in 595 adult Gambians that participated in this study. Compared to A and B loci, this locus had the lowest number of alleles. The number of people with potentially new alleles (one or more mismatches) was also small (n = 5). They were all excluded from analyses and we plan to re-sequence them at a later date in order to characterise them better. The most frequent *HLA-C* allele was *Cw*040101* (AF=18.7%, GF=33.6%) followed by *Cw*160101* (AF=12.1%, GF=22.0%) and *Cw*0701* (AF=10.9%, GF=20.8%) (**Table 6.1**). Five other alleles had their allele frequencies ranging from 5.6% to 9.7% including *Cw*0210, Cw*0302, Cw*0304, Cw*060201,* and *Cw*1700.*

Comparative analyses of *HLA-C* allele and genotype frequencies between HIV-2 infected (n = 203) and uninfected (n = 266) (**Table 6.6**) and between males (n = 131) and females (n = 457) was performed. Frequencies were similar between groups in a multivariate model adjusting for gender, ethnicity and age.

We also grouped all *HLA-C* alleles into one of two mutually exclusive categories C1 (group 1), if the amino acid at position 80 is asparagine or C2 (group 2), if it is lysine. We then analysed the distribution of these categories to see if they had any influence on HIV-2 acquisition. None of the categories had any influence on HIV-2 infection since frequencies of C1 and C2 were similar between groups. Individuals with these genotypes: C1/C1, C1/C2, and C2/C2 were also equally distributed between the infected and the uninfected groups.

Table 6.6: HLA-C and susceptibility to HIV-2 infection

HLA-C Alleles	HIV-2 (n = 203)	HIV negative (n = 266)	OR (95% CI)	p*
Cw*01	0.059 (12)	0.034 (9)	2.31 (0.84-6.36)	0.106
Cw*02	0.232 (47)	0.241 (64)	1.00 (0.61-1.64)	0.993
Cw*03	0.261 (53)	0.301 (80)	0.81 (0.51-1.29)	0.382
Cw*04	0.360 (73)	0.320 (85)	1.15 (0.74-1.78)	0.536
Cw*05	0.064 (13)	0.038 (10)	1.40 (0.53-3.73)	0.496
Cw*06	0.094(19)	0.120 (32)	0.99 (0.51-1.91)	0.967
Cw*07	0.296 (60)	0.233 (62)	1.48 (0.92-2.39)	0.104
Cw*08	0.059 (12)	0.090 (24)	0.46 (0.20-1.08)	0.075
Cw*12	0.010 (2)	0.026 (7)	0.17 (0.03-1.03)	0.054
Cw*14	0.030 (6)	0.038 (10)	1.12 (0.37-3.44)	0.842
Cw*15	0.034 (7)	0.049 (13)	0.44 (0.15-1.28)	0.132
Cw*16	0.227 (46)	0.222 (59)	1.11 (0.68-1.84)	0.671
Cw*17	0.128 (26)	0.105 (28)	1.01 (0.53-1.92)	0.983
Cw*18	0.020 (4)	0.038 (10)	0.33 (0.09-1.21)	0.094

*: p-values adjusted for gender, ethnicity and age, significant if $p \leq 0.05$

HLA class I haplotypes in Fajara

Each HLA locus was checked for Hardy-Weinberg (H-W) equilibrium using Arlequin software[202]. All three loci were in equilibrium (p values: 0.12752, 0.33722, and 0.16409 for HLA-A, -B, and -C loci, respectively). **Table 6.7** and **Table 6.8** show the two- and three-locus HLA class I haplotype frequencies in Fajara (excluding those haplotypes with frequencies less than 1% in the general population).

The total number of possible A-B haplotypes was 504. Of these, only 17 had a population frequency greater than or equal to one. *A*2301-B*3501* was the most frequent A-B haplotype with a frequency of 3.3% followed by *A*2301-B*1503* (2.2%). The estimated number of possible A-C haplotypes was 351, of which 27 had a frequency of 1% or higher. *A*2301-Cw*0401* was the most frequent A-C haplotype (3.3%). Similarly, only 25 out of 390 estimated B-C haplotypes were relatively frequent and seven of these had frequencies ranging from 3.9% to 9.4% (**Table 6.7**). The most frequent was *B*3501-Cw*0401* (9.4%) followed by *B*1503-Cw*0210* (8.8%) and *B*5301-Cw*0401* (7.5%).

Table 6.7: Two-locus HLA class I haplotypes in Fajara

A-B	Freq$^{\Phi}$	A-C	Freq$^{\Phi}$	B-C	Freq$^{\Phi}$
A*2301 B*3501	0.033	A*2301 Cw*0401	0.033	B*3501 Cw*0401	0.094
A*2301 B*1503	0.022	A*2301 Cw*0701	0.027	B*1503 Cw*0210	0.088
A*3301 B*7801	0.019	A*3002 Cw*0401	0.023	B*5301 Cw*0401	0.075
A*2301 B*5301	0.018	A*0202 Cw*0401	0.022	B*0801 Cw*0304	0.061
A*0201 B*1503	0.018	A*2301 Cw*0210	0.022	B*7801 Cw*1601	0.050
A*2301 B*4901	0.017	A*3301 Cw*1601	0.020	B*5801 Cw*0302	0.045
A*0101 B*0801	0.016	A*3001 Cw*1700	0.020	B*4901 Cw*0701	0.039
A*0202 B*3501	0.015	A*3002 Cw*0701	0.016	B*5301 Cw*0602	0.034
A*2601 B*0801	0.015	A*6802 Cw*0401	0.016	B*1510 Cw*0304	0.029
A*2301 B*0801	0.015	A*0201 Cw*0210	0.016	B*1402 Cw*0802	0.028
A*6802 B*5301	0.014	A*2301 Cw*0304	0.016	B*5801 Cw*0701	0.027
A*3002 B*3501	0.013	A*2301 Cw*1601	0.015	B*4202 Cw*1700	0.020
A*3303 B*3501	0.011	A*3303 Cw*0304	0.015	B*5101 Cw*1601	0.019
A*2301 B*0702	0.011	A*2601 Cw*0304	0.015	B*4500 Cw*1601	0.018

A*0301	B*5301	0.011	A*2301	Cw*0602	0.015	B*1801	Cw*0501	0.018
A*0101	B*5301	0.010	A*2301	Cw*0302	0.015	B*4201	Cw*1700	0.017
A*2301	B*5801	0.010	A*0201	Cw*1601	0.014	B*5001	Cw*0602	0.017
			A*3301	Cw*0304	0.013	B*0702	Cw*1505	0.017
			A*0101	Cw*0401	0.013	B*0702	Cw*0702	0.016
			A*0301	Cw*0401	0.013	B*5201	Cw*1601	0.016
			A*3303	Cw*0210	0.012	B*3501	Cw*0701	0.015
			A*3001	Cw*0304	0.012	B*1516	Cw*1402	0.013
			A*3002	Cw*1601	0.011	B*8201	Cw*0302	0.012
			A*0201	Cw*0401	0.010	B*0801	Cw*0102	0.010
			A*3402	Cw*0401	0.010	B*4102	Cw*1700	0.010
			A*2301	Cw*0702	0.010			
			A*3303	Cw*0401	0.010			

$^{\Phi}$: Only haplotypes with at least 1% population frequency are listed, Hap: haplotype, Freq: frequency

Analysis of the three-locus *HLA* class I haplotype data showed that *HLA-A*2301-B*3501-Cw*0401* (2.9%) and A*2301-B*1503-Cw*0210 (2.3%) were the most frequent three-locus haplotypes in our samples (**Table 6.8**). Only 13 out of 1311 estimated three-locus haplotypes were relatively frequent with frequencies ranging from 1.0% to 2.9%.

Table 6.8: Three-locus HLA haplotypes in the Fajara cohort

A-B-C			Freq[Φ]
A*2301	B*3501	Cw*0401	0.029
A*2301	B*1503	Cw*0210	0.023
A*3301	B*7801	Cw*1601	0.020
A*2301	B*4901	Cw*0701	0.018
A*0202	B*3501	Cw*0401	0.013
A*0201	B*1503	Cw*0210	0.013
A*2601	B*0801	Cw*0304	0.013
A*3002	B*3501	Cw*0401	0.013
A*2301	B*5801	Cw*0302	0.013
A*6802	B*5301	Cw*0401	0.013
A*2301	B*5301	Cw*0602	0.012
A*2301	B*5301	Cw*0401	0.010
A*3303	B*1503	Cw*0210	0.010

[Φ]: Only haplotypes with a frequency of 1% or greater are listed.

KIR genes/profiles and HIV-2 infection in Fajara

All 600 samples were successfully typed for 15 KIR genes by PCR-SSP as described in **Chapter 2**. We used two pairs of *KIR* gene-specific primers to target conserved regions. These primers annealed to 2 different exons of the same *KIR* gene to reduce the chances of missing the presence of the gene if there happened to be a point mutation at one of the primer binding sites. All 14 *KIR* genes and the pseudogene *2DP1* were detected in our samples. Framework genes (*KIR3DL3, 3DL2* and *2DL4*) were present in almost every sample. Most of the inhibitory genes, with the exceptions of *2DL2* and *2DL5*, had frequencies ranging from 86.7% to 100.0% (**Figure 6.2**). *KIR2DL2* and *KIR2DL5* genes were present in less than half of the study population. Apart from the ubiquitous *2DS4* gene that was present in 596 out of 600 samples, other activating *KIR* gene frequencies ranged between 6.7% to 37.7% with most of them (4/5) being well below 30%.

In a multivariate model comparing *KIR* frequencies between HIV-2 infected cases (n=215) and HIV antibody negative controls (n=272) controlling for sex, ethnic groups and age, none of the *KIR* genes showed a particularly strong association with susceptibility or resistance to HIV-2 infection (**Table 6.9**). However, *KIR2DL4* showed a borderline significant difference between cases and controls ($p = 0.03$) where *KIR2DL4* was common among HIV antibody negative (56.4%) compared to HIV-2 infected individuals (43.6%). *KIR* gene distribution was similar between males (n=132) and females (n=468).

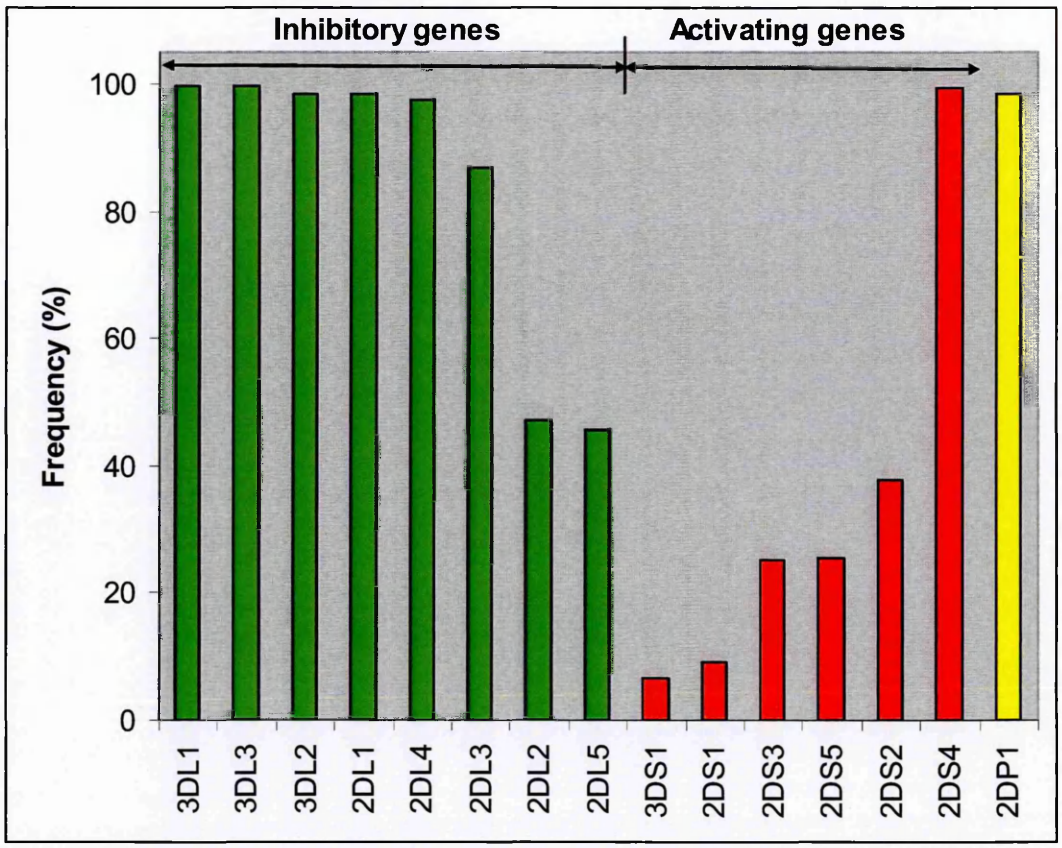


Figure 6.2 KIR gene frequencies in the Fajara clinical cohort

A *KIR* gene without its ligand has a null phenotype. To study the effect of *KIR-HLA* combinations on HIV-2 infection in a clinical cohort in Fajara, we grouped individuals based on whether the ligand(s) for their *KIR* genes were present in the same individual. A comparative analysis of *KIR-HLA* compound genotypes between HIV-2 infected and uninfected groups showed that none of the compound genotypes had a significant influence on HIV-2 acquisition (**Table 6.10**) in this multi-ethnic population. The analysis was again adjusted for gender, ethnic grouping and age at recruitment.

Table 6.9: KIR gene frequencies in cases and controls

	Genes	HIV-2 (n=215)	HIV negative (n=272)	OR (95% CI)	p^{Φ}
Inhibitory	2DL1	0.972 (209)	0.985 (268)	0.35 (0.08-1.43)	0.144
	2DL2	0.428 (92)	0.482 (131)	0.97 (0.64-1.46)	0.871
	2DL3	0.902 (194)	0.849 (231)	1.44 (0.78-2.69)	0.247
	2DL4	0.963 (207)	0.985 (268)	0.23 (0.06-0.84)	**0.026**
	2DL5	0.460 (99)	0.452 (123)	1.18 (0.78-1.78)	0.423
	3DL2	0.977 (210)	0.989 (269)	0.46 (0.10-2.18)	0.329
Activating	2DS1	0.112 (24)	0.085 (23)	1.27 (0.65-2.51)	0.487
	2DS2	0.358 (77)	0.364 (99)	1.09 (0.71-1.66)	0.698
	2DS3	0.219 (47)	0.265 (72)	0.84 (0.52-1.35)	0.468
	2DS4	0.986 (212)	0.996 (271)	0.32 (0.02-4.31)	0.392
	2DS5	0.256 (55)	0.254 (69)	1.13 (0.71-1.78)	0.615
	3DS1	0.060 (13)	0.085 (23)	0.69 (0.31-1.52)	0.356
Pseudo	2DP1	0.986 (212)	0.982 (267)	1.01 (0.21-4.77)	0.991

$^{\Phi}$: P values adjusted for gender, ethnicity, and age at sample collection

The number of activating and inhibitory KIR genes present in an individual constitutes his/her KIR profile, which in turn comprises a number of haplotypes. There are two basic types of KIR haplotypes: the "A" haplotypes comprising mainly inhibitory KIR genes (but without *2DL2*) and no activating KIR other than *2DS4.* The "B" haplotype on the other hand is a combination of both activating and inhibitory KIRs.

Table 6.10: KIR-HLA compound genotypes in cases and controls

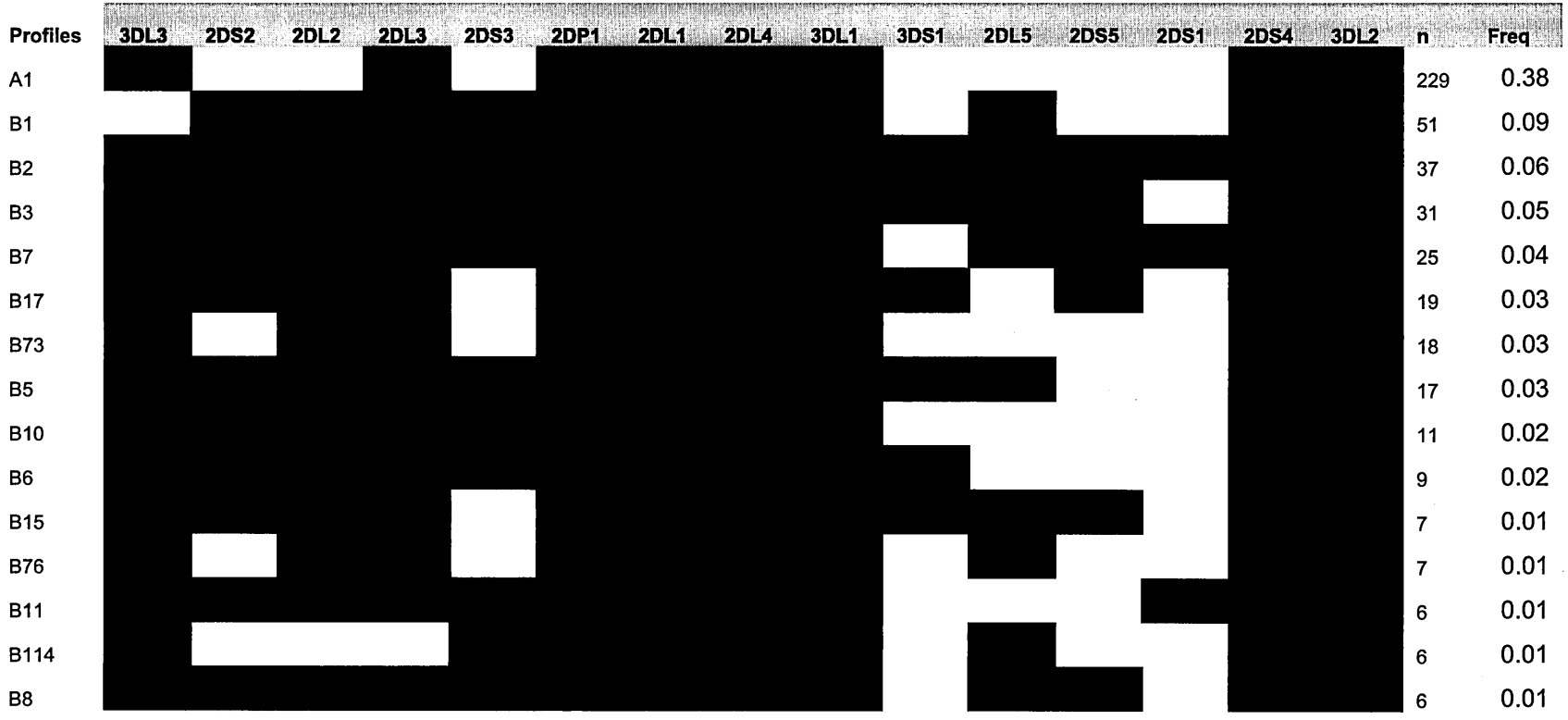
	HIV-2		HIV negative			OR	p^{Φ}
	N	%	N	%		(95% CI)	
KIR3DS1+Bw4-80I	6	0.030	13	0.053	KIR3DS1+Bw4-80I vs. Bw6/Bw6	0.39 (0.11-1.33)	0.133
Bw6/Bw6	94	0.463	104	0.421			
KIR3DL1+Bw4-80I	111	0.547	175	0.709	KIR3DL1+Bw4-80I vs. Bw6/Bw6	0.70 (035-1.41)	0.320
Bw6/Bw6	94	0.463	104	0.421			
KIR3DS1+Bw4-80I	6	0.030	13	0.053	KIR3DS1+Bw4-80I vs. Others	0.50 (0.16-158)	0.237
Others	197	0.970	234	0.947			
KIR3DL1+Bw4-80I	111	0.496	175	0.526	KIR3DL1+Bw4-80I vs. Others	0.75 (0.38-1.48)	0.410
Others	113	0.504	158	0.474			
KIR2DL2 + C1/C1	27	0.170	34	0.167	KIR2DL2 + C1/C1 vs. Others	1.36 (0.73-2.52)	0.328
Others	132	0.830	170	0.833			
KIR2DL3 + C1/C1	46	0.204	79	0.221	KIR2DL3 + C1/C1 vs. Others	0.65 (0.31-1.35)	0.250
Others	179	0.796	278	0.779			
KIR2DS2 + C1/C1	19	0.119	24	0.118	KIR2DS2 + C1/C1 vs. Others	1.32 (0.65-2.70)	0.442
Others	140	0.881	180	0.882			
KIR2DL1 + C2/C2	48	0.213	76	0.213	KIR2DL1 + C2/C2 vs. Others	0.91 (0.37-2.21)	0.840
Others	177	0.787	281	0.787			
KIR2DS1 + C2/C2	5	0.022	2	0.006	KIR2DS1 + C2/C2 vs. Others	3.15 (0.12-78.84)	0.460

Others	220	0.978	355	0.994			
KIR2DL2 + C1/x	57	0.358	80	0.392	KIR2DL2 + C1/x vs. Others	1.03 (0.63-1.67)	0.917
Others	102	0.642	124	0.608			
KIR2DL3 + C1/x	157	0.698	233	0.653	KIR2DL3 + C1/x vs. Others	1.02 (0.49-2.10)	0.960
Others	68	0.302	124	0.347			
KIR2DS2 + C1/x	46	0.289	61	0.299	KIR2DS2 + C1/x vs. Others	1.00 (0.60-1.68)	0.988
Others	113	0.711	143	0.701			
KIR2DL1 + C2/x	134	0.843	167	0.819	KIR2DL1 + C2/x vs. Others	1.46 (0.77-2.78)	0.249
Others	25	0.157	37	0.181			
KIR2DS1 + C2/x	9	0.057	11	0.054	KIR2DS1 + C2/x vs. Others	0.79 (0.27-2.28)	0.659
Others	150	0.943	193	0.946			

$^{\Phi}$: p values adjusted for gender, ethnicity and age at recruitment

The analysis of all *KIR* gene profiles obtained from this Gambian population showed the presence of 91 different profiles: two profiles A (A1 and A2) and 89 B profiles. Profile A1 was by far the most frequent (38.2%), followed by B1 (8.5%) and B2 (6.2%). A2 profile, detected in 5 out of 600 samples, was very similar to A1 except for the absence of *KIR2DL3*. In addition to B1 and B2, twelve other B profiles were present in at least 1% of the study population (**Table 6.11**). A comparative analysis of some of the common *KIR* gene profiles in the HIV-2 infected and control groups showed that none of the profiles was significantly associated with HIV-2 infection (**Figure 6.3**).

Table 6.11: Most frequent KIR profiles in the Fajara population



Open box: gene is absent; closed box: gene is present; n: number of individuals with the profile of interest. Only profiles that are present in at least 1% of the study population are shown.

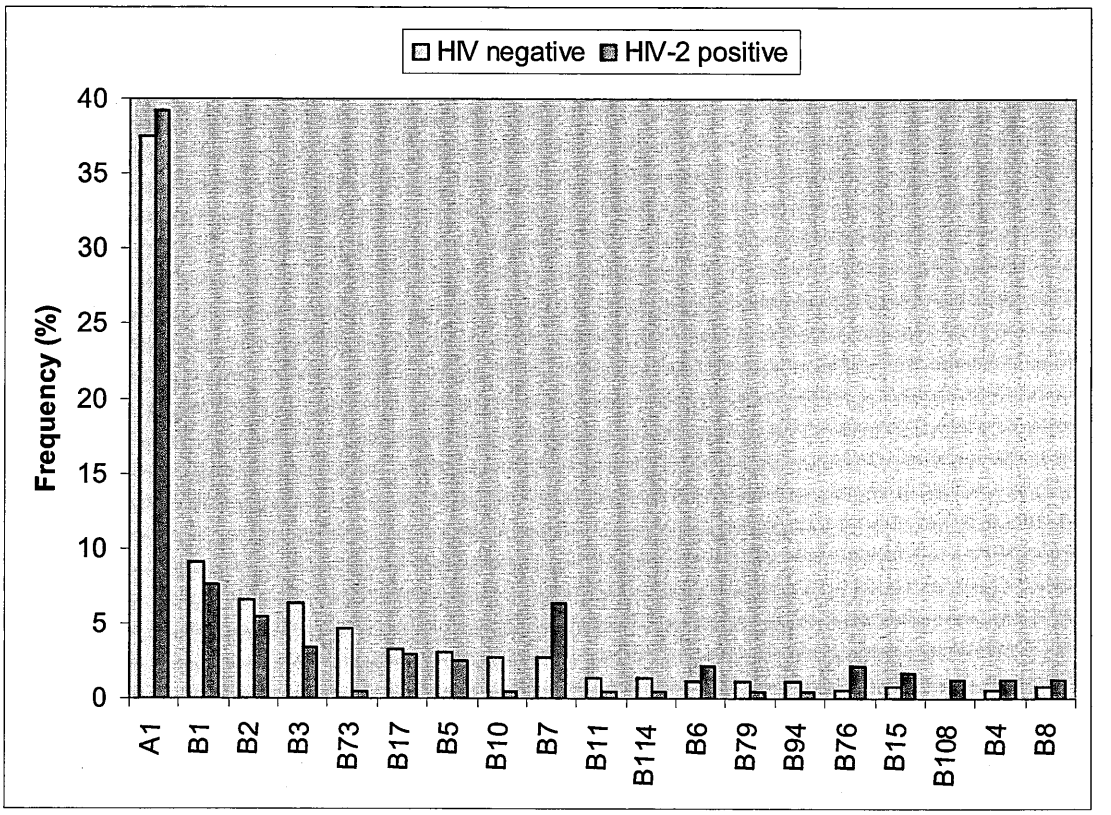


Figure 6.3 Most common KIR gene profiles in cases and controls

Discussion

In this part of the study, we determined and analysed HLA and KIR genetic polymorphisms in the Fajara clinical cohort for two important reasons: (1) to describe the genomic variations at the HLA class I and KIR loci in this other sample of West African populations and (2) to examine whether variation at these selected loci is associated with susceptibility to HIV-2. We also aimed to study the relationship between KIR-HLA compound genotypes and HIV-2 infection.

HLA class I alleles/haplotypes do not influence HIV-2 infection in Fajara

A total of 17 *HLA-A* serological groups (2-digit alleles) and 28 *HLA-A* subtypes (4- to 6-digit alleles) were detected in Fajara. *HLA-A* alleles and genotype frequencies were similar between cases and controls. Comparing our data with other published frequencies from the sub region, we found that *HLA-A*02* allele and genotype frequencies in Fajara (AF=13.2%, GF=23.9%) were lower compared to other populations. For example, in the Bakola Pygmies in Cameroon, *A*02* genotype frequency was 50% while it was 37% in the Mbenzele population in Central Africa Republic[203]. Our *HLA-A* allele frequencies findings were similar to those reported by Allsopp and colleagues[189] who serologically determined the frequencies of classical HLA class I in different tribes in the Gambia in the early 1990's.

Analysis of HLA-B data generated from this HIV-2 clinical cohort in Fajara revealed, as is the case in most of the world populations, that B locus was the most diversified of all class I loci. We found 50 distinct *B* alleles with four to six digits that can be grouped in 28 different serological groups. Two alleles (*B*14* and *B*57*) were more frequent in the control group than in the HIV-2 infected group, suggesting that these groups of alleles could be associated with protection against HIV-2 acquisition.

Although the p values remained significant after adjusting for gender, ethnicity and age at recruitment, the number of subjects positive for any of these alleles was small (26 for *B*14* and 12 for *B*57*, respectively) and the results should be interpreted cautiously. Both *B*14* and *B*57* had three subtypes each but with the exception of *B*140201* (AF=3.0%, GF=5.9%), all the other subtypes were rare (allele frequencies below 1%). In general, authors of most published HLA data from Sub Saharan Africa have detected very low allele and genotype frequencies for *B*14* and *B*57*. For example, Modiano et al. found in random samples collected from three ethnic groups in Burkina Faso, that the allele frequency of *B*14* was 0.0% in Rimaibe, 0.9% in Mossi, and 3.0% in Fulani while that of B*57 was 4.7% in Mossi and 0.0% in Fulani and Rimaibe, respectively. If the protective effect observed in this Gambian cohort is real, it would support the proposition made by Trachtenberg et al. that rare *HLA* alleles more often offer protection against HIV infection and/or outcomes than common alleles in populations where HIV viruses have adapted to the more frequent alleles[137].

In this cohort, we detected 14 *HLA-C* group of alleles (2-digits) comprising 23 distinct alleles (4- to 6-digits). We repeated similar analyses as for *HLA-A* and *-B* above to determine the influence of individual *HLA-C* alleles on HIV-2 infection. None of the *C* alleles or genotypes was associated with susceptibility or resistance to HIV-2 infection. Similarly, none of the C1, C2, C1/C1, C1/C2, or C2/C2 categories could influence susceptibility or resistance to HIV-2 in this sample of West African population.

In order to determine the influence of HLA class I haplotype on HIV-2 infection, we reconstructed and estimated the frequency of all possible HLA class I haplotypes using Arlequin software. Comparing the two-locus and three-locus HLA class I

haplotypes between HIV-2 infected and uninfected groups, we observed that none of the differences in haplotype frequencies between the two groups reached significant level ($p > 0.05$).

Potential new *KIR2DL2* alleles detected in Fajara samples

We detected all 15 *KIR* genes in our Fajara samples. Knowing that genetic diversity is high in African populations, we anticipated that there could be some new polymorphisms not yet described in the literature in some of our samples. We then opted to use two pairs of primers annealing to two different exons of the same gene to detect the presence of each *KIR* gene. This was an attempt to minimise the likelihood of assigning false negative results to samples with new *KIR2DL2* alleles containing some point mutation(s) at the primer binding sites. Rightly so, we observed that for some of our samples (about 10%), there was a specific band on lane 4 but none in lane 3 and this was consistent after several repeats (see **Appendix 12** for a sample of the gel scoring sheet). Lanes 3 and 4 contained a pair of primers each, designed to amplify two segments of *KIR2DL2*: one in exon 4 at position 419 to 591 (a segment of 173 bp) and the other in exon 5 at position 813 to 963 (a segment of 151 bp). Ideally, in the presence of known *KIR2DL2*, there should be a specific band in both lanes 3 and 4 simultaneously.

The pattern described above indicated that there could be new *KIR2DL2* alleles in our study population that have not yet been described or published in public databases since these primers were designed for conserved regions using all traces available in public databases. We repeated the typing 2 to 3 times to confirm the pattern and designed a new set of primers to amplify and sequence some of the

samples for the region encompassing the primer binding sites in exons 4 and 5 of *KIR2DL2* .

Bi directional sequencing was performed and the resultant traces analysed using Mutation Surveyor software version 3.1 (SoftGenetics, USA). The analysis revealed that there were two point mutations at both primer binding sites preventing primers from annealing properly thereby resulting in no amplification in lane 3. Due to limited financial resources, we could not further characterise the suspected new alleles. However plans are in place to fully sequence these samples for *KIR2DL2* genes. We also plan to deposit the resultant traces in public databases for the benefit of the scientific community.

KIR frequencies in Fajara are similar to other West African populations

Comparing our data to the little available data in Sub Saharan Africa and particularly West Africa, we observed that the frequencies of most of the *KIR* genes detected in these Gambian samples were in agreement with data from Senegal[187], Ghana[193], and Nigeria[195]. Of note, however, was the particularly low frequency of activating *KIR* genes in most of these populations. Single et al. recently showed that the frequencies of two activating *KIR* genes (*KIR2DS1* and *KIR3DS1*) were very low in West and East Africa and that the phenotypic frequencies of these two genes increased with increasing geographical distance from East Africa eventually peaking in the Oceania and Americas[195].

In a multivariate analysis comparing the frequency of individual *KIR* genes between cases and controls, adjusting for age, sex and ethnicity, we observed that only *KIR2DL4* was weakly associated with protection against HIV-2 infection. All other individual *KIR* gene frequencies were similar between both groups. Similarly,

the frequencies of "functional" *KIR* genes (those *KIR* genes that were present with their corresponding ligands in the same individuals) were similar between cases and controls.

NK cell activity is fundamental in humans and activating receptors are needed to activate NK cells to carry out their cytolytic activities in response to immune stimulation by a "non-self". The consistent low frequency of activating *KIR* genes in most African population raises important questions: (1) Are there new activating KIR genes in Africans that can not be detected with our current typing techniques? (2) Are there alternative pathways to activate NK cells in the absence of activating KIR molecules? To date, ligands for most activating KIRs are yet to be discovered. Further research needs to be conducted to address some of these pertinent questions.

We detected 91 distinct *KIR* gene profiles in Fajara samples with the most common being the same A1 profile that predominated in Caio. However, none of these profiles was found to influence HIV-2 infection.

In brief, we analysed 600 samples from adults Gambians in Fajara for *HLA-A, HLA-B*, *HLA-C*, and *KIR* genes. We detected 28 *HLA-A*, 50 *HLA-B*, and 23 *HLA-C* alleles (4- to 6-digits). Apart from the high heterogeneity at the B locus, two *HLA-B* alleles (*B*14, B*57*) showed weak protection against HIV-2 acquisition (p=0.031 and p=0.029, respectively). None of the HLA class I haplotypes was associated with susceptibility or resistance to HIV-2 infection in this Gambian population. All 15 KIR genes were detected in our samples but none of these showed a particularly strong association with susceptibility or resistance to HIV-2 infection. HLA-KIR compound genotypes also did not affect risk of HIV-2 acquisition.

KIR profile A1 was the most frequent of the 91 profiles detected in this sample of the Gambian population. The frequencies of all the *KIR* gene profiles were similar between cases and controls and therefore did not significantly influence HIV-2 infection.

CHAPTER 7 : CONCLUSIONS

In nearly three decades since the first identification of the new human retrovirus causing an unusual and debilitating disease with profound immune suppression, AIDS remains the most devastating disease of our time. It is the leading cause of death among hospitalised adults in developing world[4]. The burden of AIDS on most African populations is huge. The rate of new infection is highest in Sub Saharan Africa[4] and prevalence continues to increase alongside improved survival following the introduction and successful implementation of antiretroviral programmes in many countries South of Sahara in the last ten years or so. In this part of the world, most hospital wards are still full of emaciated men and women infected with the virus. More than 50 million Africans have been directly or indirectly affected by the scourge of HIV/AIDS. The demographic impact of AIDS pandemic is complex and not yet fully understood. Sub Saharan Africa remains the epicentre of the pandemic where both HIV-1 and HIV-2 co-exist with many other infectious diseases such as malaria, gastroenteritis, pneumonia and tuberculosis.

The HIV epidemic in West Africa is caused by two closely related retroviruses HIV-1 and HIV-2. The latter, which is the focus of this thesis, was first reported to have infected healthy Senegalese prostitutes in 1985[15] and was isolated the following year from two West African patients with AIDS: one from the Cape Verdes Islands and the other from Guinea Bissau[3]. It is structurally very similar to HIV-1 and shares between 30 to 60% homology at nucleotide level. Although they both target same cell populations for destruction, their rate of infection and the resultant pathogenesis are

quite different. Slow $CD4^+$ T cell decline, slower progression to full-blown AIDS and higher numbers of long-term non-progressors (LTNP) are some of the characteristics of HIV-2. However, at the later stage of the disease, HIV-2 infected patients eventually develop immunodeficiency with low $CD4^+$ T cell count, high HIV-2 viral load and a clinical feature similar to that observed in their HIV-1 infected counterparts at the same stage of disease[204].

The reason(s) behind the attenuated course of HIV-2 infection and disease progression in a substantial proportion of HIV-2 infected individuals are yet to be uncovered. Immunogenetic factors are likely to be key determinants of whether or not an infected person rapidly develops clinically overt immunodeficiency or becomes a LTNP. Epidemiological evidence is accumulating, indicating a range of distinct viral and host genetics and possibly some environmental factors that contribute to the observed differences. Among these are molecules of the MHC and KIR regions that are the most important antiviral innate and adaptive immune response regulators.

A lot is now known about the pathogenesis of HIV infection but not much is known about correlate of protective immunity observed in most patients. The factors that influence the interindividual variability with regards to susceptibility to HIV infection and progression to AIDS are not well understood. Host immunogenetic factors such and *HLA* and more recently *KIR* genes, which are extraordinarily polymorphic, have been repeatedly investigated in many large and well established cohorts around the world since the isolation of HIV-1 in 1981. A good number of these genes and alleles as well as compound genotypes have been implicated in the control of HIV-1. Independent epidemiological studies have demonstrated the effect of *HLA-B*35*-Px subsets, notably *B*3502, B*3503*, and *B*3504*, with rapid progression to AIDS following HIV-1 infection[69 134]; while *HLA-B*27,* and *-B*57,* as

well as *KIR3DS1+Bw4-80I* have been associated with long-term survival. However, very little is known about the role of the above mentioned genes/alleles in HIV-2 infection and disease progression.

The main goal of this study was to determine the type of *HLA* alleles and *KIR* genes present in two well characterised West African cohorts: one in Caio, Guinea Bissau (a country with one of the highest prevalence of HIV-2 infection in the world) and the other in Fajara, The Gambia (a country with one of the lowest prevalence of HIV infection in general). Our second major objective was to relate the presence or absence of these highly polymorphic genes with susceptibility or resistance to HIV-2 infection and disease progression assessed using known markers of progression to AIDS such as decline in $CD4^+$ T cell count and increase in viral load over time.

To the best of our knowledge, there is no information about HLA-linked genetic control of HIV susceptibility in these populations (Caio and Fajara). In general, within the African continent, there is a paucity of information on the importance of *HLA* and *KIR* genes in the susceptibility to infectious diseases including HIV. This is partly due to the difficulties in clearly identifying "at risk" populations for longitudinal studies since the frequency and degree of exposure to HIV is difficult to measure more so in an African context. The fact that there is limited amount of HLA and KIR genotyping information on African cohorts and the heterogeneity in the typing techniques used in generating the few currently available data, makes it difficult for us to compare our data with others in the sub region. We only have to rely on studies reporting HLA genotype distribution in HIV-positive and HIV negative individuals in a single risk group across the continent and elsewhere in the western world.

The exact time of seroconversion is unknown for the infected group and the extend of exposure of those considered HIV negative at the time of testing can not be defined in this study. This is a general phenomenon in most African cohorts.

This study provides and describes frequencies of MHC class I *HLA-A, -B* and -*C* and *KIR* genes in a predominantly Manjako community in Caio, Guinea Bissau and in a multi-ethnic Gambian cohort in Fajara, The Gambia. Comparisons of allele and genotype frequencies were made between these populations and their neighbouring populations in West Africa including Senegal, Mali, Cameroon and other more distant populations in East and South Africa. *HLA* class I alleles with higher frequencies in the Caio community were as follows: *A*3303*, and *A*2301* for *HLA-A*; *B*1503, B*5301, B*5801, B*0801, B*1510,* and *B*3501* for *HLA-B*; and *Cw*0701, Cw*0401, Cw*0304,* and *Cw*0210* for *HLA-C*. In the Fajara cohort the most common alleles (frequency of 9% and above) included: *A*2301 for HLA-A; B*5301, B*3501, and B*1503* for *HLA-B*; and *Cw*0401, Cw*1601, Cw*0701, Cw*0304,* and *Cw*0210* for *HLA-C*. Both populations are separated by a distance of more than 600 Km on the West coast of Africa and yet share many common HLA types albeit at different frequencies.

HLA-A frequencies in Caio and Fajara

We found the genotype frequency of *HLA-A*33* to be significantly higher in Caio compared to Fajara (38.8% vs. 24.7%, $p < 1x10^{-6}$). *A*33* is a common African allele existing in all ethnic groups with an average frequency of about 5% in blacks[98]. Amerindian and Oriental populations share similar frequencies at the population level (about 5%) while it is less frequent in Caucasians. The lowest population frequency of *A*33* was recorded from Australian Aboriginals (0.5%)[98]. In African populations as

well as African Americans, the frequency of *A*33* varies greatly ranging between 0.0 and 16.0%. For example, in West African pygmy populations e.g. Mbenzel and Bakolas, *A*33* was reported to be 0.08% and 0.18%, respectively[203]; it is 2.8% in Ethiopians, 3.1% in Botswana, and 13.3% in Fulani population in Burkina Faso.

*HLA-A*33* is a split of the serologically defined A19 group of alleles[205] at the HLA-A locus. Currently there are 29 subtypes of *A*33* described in the IMGT/HLA database[94] making it the second most polymorphic of the A19 splits after *A*30*. In both the Caio and Fajara subjects studied, we only identified two A*33 subtypes: *A*3301* and *A*330301*. The latter was by far the most frequent subtype in Caio, where its frequency was seven fold higher than that of *A*3301*. In Fajara, however, both subtypes were equally distributed.

Spinola and colleagues recently reported *HLA-A*33* frequency among males from four different ethnic groups (Balanta, Fula, Bijagos, and Papel) within Guinea Bissau[206] and found the highest frequency to be in the Bijagos ethnic group (10.7%). A literature search of published data from Sub Saharan Africa indicates that the Caio population exhibit the highest frequency of *A*33*, notably *A*330301*. The reason(s) for this high frequency of *HLA-A*33* in Caio can only be speculative. Data from a neighbouring Senegalese population reported a frequency of 12.9% in a group of female commercial sex workers (CSW)[115]. It is however difficult to compare our *A*33* genotype frequency to those mentioned above since each of these studies derived their frequency from only a handful of people of a single gender (23 Bijagos males in the case of Spinola et al. and 62 female CSW in the case of Diouf et al.).

In The Gambia, Allsop et al. serologically analysed the frequency of classical HLA class I antigens in various ethnic groups[189] and reported that reactivity to A33 antigen was as high as 36.0% in Serere (n = 25), 24.4% in Jola (n = 123), 23.3% in

Fula (n = 90), and 24.4% in Manjago (n = 41). In our study, although using a different typing technique and a larger sample size (n = 600), we found the overall frequency of *A*33* to be 24.7% in our Gambian cohort. It is worth noting that we studied a multi-ethnic group in Fajara comprising Mandinka (46.5%), Jola (13.6%), Fula (13.4%), Wolof (10.6%), Manjako (3.7%), Serer (3.7%), Serahuli (2.6%), Aku (0.6%), and other minority ethnic groups classified together as "Other" (5.5%).

Analysis of HLA-A types stratified by ethnicity showed that *HLA-A*33* frequency was highest in Manjako (35.3%). In a predominantly Manjako community in Caio, we saw a similar trend that *A*33* frequency was high (38.8%). This is an indication that people of the Manjako tribe may overall exhibit a high frequency of *HLA-A*33* independent of their geographical location since these two Manjako populations (Caio and Fajara) are quite far apart from each other and yet have similar *A*33* genotype frequencies (38.8% vs. 35.3%, respectively, p = 0.768). This supports previous observations that Manjakos are truly a unique set of people with their unique values[207] and now distinct immunogenetic factors that distinguish them from their neighbouring West African populations.

All Manjako subjects from the Fajara cohort carried the *A*330301* subtype as opposed to all other ethnic groups in which the frequencies of *A*330301* and *A*3301* were similar. For example, the Wolof ethnic group in The Gambia also had high genotype frequency of *A*33* (34.0% overall) but the frequency of *A*330301* was not significantly different from that of *A*3301* amongst the Wolof people (52.9% vs. 47.1%, respectively, p = 0.791).

Another report by Cao et al, showed high heterogeneity of *HLA-A*33* frequency in different populations[191] across South, East, Central and West Africa:

Zulu (0.5%), Kenyan Nandi (0.4%), Kenyan Luo (2.3%), Zambians (1.2%), Ugandans (4.6%), Cameroonians (2.8%), and Malians (11.2%).

The genotype frequency for *HLA-A*74* was also high in Caio when compared to Fajara but with a borderline statistical significance (10.6% vs. 7.1%, respectively, p = 0.045). *A*74* is another split of A19 serological group of alleles. It is generally considered to be an African allele. Its frequency varies among studied populations in Sub Saharan Africa: 0.2% in Zimbabwe Harare Shona (n = 230)[193], 6.5% in Zulu (n = 51), 5.0% in Cameroonians (n = 47)[191], 6.6% in Mossi in Burkina Faso (n = 53)[188], and 19.6% in male Bijagos in Guinea Bissau (n = 23)[206]. *A*74* is less polymorphic compared to other members of the A19 splits with only 14 alleles described to date[94].

The sequences of exons 2 and 3 of all of our samples that were positive for *A*74* were analysed using "Assign 400" - a specialised software from Conexio Genomics, Western Australia and they suggested that the genotypes of these individuals could either be *A*7401* or *A*7402*. Both of these alleles have identical nucleotide composition in those exons that we investigated. It is known that *A*7401* differ from *A*7402* by a single point mutation in exon 1 at nucleotide position 67 (codon 23) where the former has an "A" and the latter a "T" in that position[208]. This single non-synonymous mutation results in an amino acid change from Arginine (R) to Tryptophan (W) or *vice versa* at codon 23 of A*74 protein depending on which of these two alleles existed first.

In an attempt to distinguish between these two alleles, we amplified a segment comprising exon 1 from all A*74-positive samples and re-sequenced them. Analysis of the traces showed that the amino acid at codon 23 was R, indicating therefore that the *A*74* allele in both Caio and Fajara was *A*7401*.

*HLA-A*24* allele was completely absent in Caio but present in 2.6% of our Gambian subjects (p = 0.0003) mainly from the Mandinka and Fula tribes. It is one of the most polymorphic alleles at the A locus with 118 subtypes currently described in the public database[94] and it is not a common allele of Africans[98]. The average population frequency of *HLA-A*24* in Africans is 3.1%, ranging from 0.5% to 6.2%. It is therefore not surprising that *A*24* was absent in the predominantly Manjako community in Caio, Guinea Bissau and in many other tribes in our Gambian cohort. It is worth mentioning that this allele was also absent in Manjako of Gambian origin that we studied. Australian Aboriginals harbour the highest frequency of *A*24* in the world (48.6%, ranging from 32.0% to 65.7%)[98].

The only other *HLA-A* allele that we found to be present in Gambians at significantly higher phenotype frequency than in Guineans was *A*30* (25.6% vs. 19.6%, respectively, p = 0.031). This allele is commonly found in Africans with frequencies ranging from 9.3% to 31.6% (average 14.8%)[98]. *A*30* is also a split from the A19 serological lineage. It is considered the most polymorphic of all the A19 splits with 32 known subtypes. In our samples we found three of the 32 subtypes: *A*3001*, A*3002, and *A*3004*. The most frequent subtype in Caio was *A*3002* (14.5%) and none had the *A*3004* genotype, which was also rare among the Fajara subjects (0.6%). All the three subtypes were present in Fajara but the overall *A*30* genotype frequency (25.6%) was equally distributed between individual carrying either *A*3001* or *A*3002* (12.6% and 13.0%, respectively).

In studying the effect of *HLA-A* variation on risk of HIV-2 acquisition we found one of the 24 *HLA-A* alleles (*HLA-A*8001*) to be weakly associated with increased risk of HIV-2 infection. This association was only seen in the Caio community cohort and not in the Fajara clinical cohort. The fact that we did not see this weak effect in

Fajara might be due to the multi ethnic nature of the cohort, since in Caio we had almost an homogenous population (>95% Manjako).

HLA-B frequencies in Caio and Fajara

HLA-B locus was the most diversified of all class I loci in our study populations: 50 and 32 alleles (4- to 6-digits) detected in Fajara and Caio, respectively. Four of them were present at significantly higher allele and genotype frequencies in Caio compared to Fajara: *HLA-B*13, -B*15, -B*49*, and *-B*58*. On the other hand, six others alleles had significantly higher allele and genotype frequencies in Fajara than Caio including *HLA-B*07, -B*35, -B*39, -B*42, -B*45,* and *-B*78*. We also found that three of the 50 B alleles in Fajara were absent in Caio samples: *HLA-B*27, -B*37*, and *-B*55*.

*HLA-B*13* was relatively rare in Fajara compared to Caio (1.0% vs.6.4%, $p = 2.6x10^{-6}$). It was detected in only 6 out of 583 Fajara samples that were successfully assigned a *B* genotype. *B*13* frequency in Caio was comparable to the average frequency found in Oriental populations[98] where it predominates. Shen and colleagues reported a frequency of 11.1% in a subset of Han population[209] from Xi'an in China. Other studies on the same Han populations from the Northern[210] and Southern[211] parts of China found that *B*13* frequencies ranged from 9.0% to 11.2%. In contrast, apart from this high frequency observed in the Manjako population in Caio, most populations in Sub Saharan Africa exhibit low frequencies of *B*13* ranging from 0.0% to about 3.0%. For example, in a study carried out on 52 individuals from the Zulu[191] population in South Africa, *HLA-B*13* frequency was 1.5%, in Cameroon it was 2.7%, while in Ugandans[212] it was 2.0%. In another study in Burkina Faso, none

of the subjects from the three tribes studied (Fulani (n = 49), Mossi (n = 53), and Rimaibe (n = 47)) carried the B*13 allele[188].

Of the 32 subtypes of *HLA-B*13* described in the immunogenetics database, we detected only two: *B*1301* and *B*130201*. The former was very rare and seen only in one Fajara sample and the latter, *B*130201* was quite common in both populations.

A number of studies have implicated *HLA-B*13* with susceptibility to type II psoriasis vulgaris[213], Chronic Myelogenous Leukemia (CLM)[214], and the control of HIV-1 clade C viral replication[215 216]. Many epidemiological studies continue to show that specific *HLA-B* alleles (such as *B*13, B*14, B*15, B*27, B*35, B*53, B*57*, and *B*58*) collectively exert a strong impact on HIV-1 infection and disease progression. For example, in individuals infected with clade C HIV-1 virus in Zambia, *HLA-B*13* was associated with successful viral control[216]. In our samples, the phenotypic distribution of *HLA-B*13* between individuals infected with HIV-2 virus and controls (HIV negative individuals) was similar in both Caio and Fajara.

*HLA-B*15* was the most frequently encountered allele in both Caio and Fajara. Comparative analysis of *HLA-B*15* phenotype frequency between these cohorts showed a statistically significant difference (41.4% vs. 26.8%, respectively, $p = 1.4 \times 10^{-6}$). The phenotypic distribution of *HLA-B*15* varies from one African population to another. In West African Pygmies, *B*15* allele frequency was 8.5% in the Mbenzele tribe (n = 36) and 20.0% in the Bakolas ethnic group (n = 50)[203]; in a cohort of HIV-1 positive and negative individuals[217] in Botswana (n = 161) it was reported as 2.6%; in Burkina Faso, *B*15* genotype frequencies ranged between 6.6% (Mossi) and 17.0% (Rimaibe)[188]. In neighbouring Senegal it was 22.3% in Dakar population (n = 112); and in the Twana group in South Africa it was 26.8% (n = 41)[193].

*HLA-B*15* is one of the most polymorphic alleles in humans. There are currently 181 subtypes in the Immunogenetics database. Of these, we detected three in Caio: *B*1503, B*1510*, and *B*1516*; and four in Fajara *B*1503, B*1510, B*1516,* and *B*1518*. The latter was very rare and present only in a single Fajara sample. The most common subtype was *B*1503* followed by *B*1510* in both cohorts. None of the four *B*15* subtypes was associated with susceptibility to HIV-2 infection. Their allele and genotype frequencies were similar between cases (HIV-2 infected) and control (HIV negative) individuals.

In Caio, where the majority of HIV-2 infected individuals had laboratory data available on $CD4^+$ T cells count and HIV-2 VL measurements overtime, we found a strong association between *HLA-B*15* and disease progression assessed using universal markers of HIV disease progression such as $CD4^+$ T cell count and viral load. This association was entirely attributed to *B*1503* since the other common *B*15* allele *B*1510* frequency was similar between groups with or without the allele. Individuals positive for *B*1503* had a considerably higher HIV-2 VL and lower absolute $CD4^+$ T cell count compared to those negative for *B*1503*. This was not a haplotype effect since none of the HLA class I haplotypes carrying the *B*1503* allele was seen to be mediating a similar effect. The haplotype frequencies were similar between infected and uninfected groups in both cohorts. We hypothesised that this association was ethnic-specific, even though we could not see a similar trend in the few Manjakos in our multi-ethnic Fajara cohort probably because of the small sample size (n=18). Is *B*1503* a surrogate marker for another gene or group of genes that could be the actual mediator(s) of the observed effect? This possibility can only be excluded after further investigations into the functional role of *HLA-B*1503* in HIV infection. Part of these analyses concerning the association between *HLA-B*1503*

and markers of disease progression ($CD4^+$ T cell count and HIV-2 VL in the Caio cohort has been written and submitted for publication in a peer review journal (Journal of Immunology) and is currently under review (see **Appendix 15** for a copy of the submitted manuscript)

The other *B* allele that was present in Caio at a significantly high frequency than in Fajara was *HLA-B*49* (12.8% vs. 7.2%, respectively, p = 0.003). *B*49* has previously been reported as a common allele in blacks and Caucasians[98]. Our frequencies were higher compared to those reported in cohorts in Zambia[216] (2.4%), Botswana[217] (0.3%), Uganda[212] (5.4%), Senegal[193] (7.1%); and significantly lower compared to that found in Ethiopians (31.4%) from a low resolution typing[218]. *HLA-B*49* was found to be associated with rapid progression to AIDS following HIV-1 infection with subtype B virus[116]. The region of the human MHC encompassing HLA-B*49 is one of the least polymorphic within B locus with only six subtypes described in the immunogenetics database. Only one of these was found in our study populations (B*4901) in line with most African studies that employed high resolution (4- to 6-digits) HLA typing techniques.

*HLA-B*58* was over-represented in the Caio community (21.2% vs. 13.9%, p = 0.002, Caio and Fajara genotype frequencies, respectively). This group of alleles has been reported to have the highest population frequency in blacks[98]. There are 23 alleles of *B*58* officially published to date[94]. In this study, we found only one of the subtypes - *B*5801*. This is in line with observations made from recent studies in West Africa where *B*5801* was the only *B*58* allele found at relatively high frequency among 65 male Guineans[192], 138 Malians[191], and 149 individuals from 3 ethnic groups in Burkina Faso[188]. The over-representation of *HLA-B*5801* has been associated with protection against rapid progression to AIDS following HIV-1 infection

in Western populations and in East African cohorts[219] [220]. Since the date of seroconversion was unknown for most of our study participants, we could not determine the exact effect of these genes/alleles on the rate of disease progression with high precision. We, however, used available information on markers of disease progression to assess the potential effect of *HLA* alleles detected in the study population on HIV-2 disease. There was no trend to indicate that any of the markers of disease progression ($CD4^+$ T cell count and HIV-2 VL) was influenced in any direction by *HLA-B*5801*.

We also observed that the phenotypic distribution of *HLA-B*78* was hugely imbalanced between the two populations with a relatively high frequency in Fajara compared to Caio (10.8% vs. 0.5%, respectively, $p < 1x10^{-7}$). The dramatic low frequency seen in the Caio population is in agreement with what is known of other black populations where the average *B*78* allele frequency has been reported to be 0.6%, ranging from 0.0% to 5.2% in Sub Saharan Africa. Conversely, the high frequency observed in the Fajara population both at the allelic and phenotypic levels is similar to that reported in a Senegalese population in Dakar[193] (7.6% allele frequency and 16.1% phenotype frequency). *B*78* is generally rare in most populations world wide and has not been detected in many Caucasoid populations or Australian Aboriginals[98].

*B*78* is an oligomorphic group of alleles with only a few alleles described in the Immunogenetics database – seven alleles. We found one *B*78* subtype in our Caio samples (*B*7801*) and two in Fajara: *B*7801* and *B*780201*. The latter was seen in only 1 out the 583 Fajara samples. We hypothesise that *B*7801* is most likely a West African allele predominant in the Senegambia area of West Africa, which comprises the republic of Senegal and The Gambia. *B*7801* allele frequency increases as you

move from Cameroon towards the tip of West Africa. It is 0.6% in the Beti ethnic group in Cameroon[221], 1.1% in Rimaibe in Burkina Faso[188], 2.3% in Guinea Bissau[192], 2.3% in Ivory Coast[222], and 6.9% in the Bandiagara tribe in Mali[191].

HLA-C frequencies in Caio and Fajara

The C locus had the lowest number of alleles (4- to 6-digits) in both cohorts: 21 in Caio and 23 in Fajara, respectively. None of these *HLA-C* alleles were individually or collectively (C1, C2) associated with susceptibility or resistance to HIV-2 infection in both cohorts. However, some of the *KIR/HLA* compound genotypes showed weak associations with protection against HIV-2 acquisition; in particular *KIR2DL2* and *KIR2DS2* when present with their corresponding/putative ligand – HLA-C group 1 were seen at significantly higher frequency in the uninfected group than in those infected with HIV-2. We also reported in this study using the Caio population that individuals positive for Cw*02 had low $CD4^+$ T cell count with a trend towards higher HIV-2 viral load than those without this genotype. Further analysis attributed this bad prognostic effect entirely to *Cw*0210*, which is in strong LD with *HLA-B*1503* (as demonstrated by the high frequency of *B*1503-Cw*0210* haplotype in this sample of the Manjako population from the Caio sector). HIV-2 infected individuals with the *Cw*0210* genotype had a much lower $CD4^+$ T cell count and a much higher HIV-2 VL than previously observed with the overall Cw*02-positive group. This effect was not seen with the other *Cw*02* allele (*Cw*0203*). We also observed in Caio that *HLA-Cw*07* was weakly associated with high HIV-2 viral load but not CD4 count.

KIR genes and profiles in Caio and Fajara

Inhibitory *KIR* gene frequencies were similar between Caio and Fajara as is the case in other West African populations. But the frequencies of some of the

activating KIR were significantly higher in Caio than elsewhere in West Africa but more similar to those found in Europe particularly France. Anthropological findings have shown that the Manjako communities are more open (sexually) to Europeans, usually from Portugal and France, than they are to their closest neighbours in West Africa. Interestingly, *KIR3DS1*, which has been shown to be rare or absent in people of African ancestry, was present in Caio at a significant frequency. It is reasonable to think therefore, that the *KIR3DS1* gene may have been introduced in Caio from Europe, especially Portugal since Guinea Bissau is a former Portuguese colony.

KIR3DS1 frequency in Fajara was comparable to those reported in other populations in Sub Saharan Africa. At least three epidemiological studies have demonstrated that *KIR3DS1* is associated with protection against rapid progression to AIDS following HIV-1 infection, particularly when the infected individuals have its putative ligand HLA-B Bw4-80I[169]. We did not see that effect in any of our HIV-2 cohorts. Most of the beneficial or deleterious effects seen in HIV-1 cohorts were not seen in this HIV-2 study. Does this suggest different mechanisms of protective immunity between the two virus strains, or is it because the two viruses have different properties? Are our observations related to population diversity? Or the study designs? It is worth noting that we used two prevalent cohorts for our study and the exact dates of seroconversion are unknown for the infected groups, but this is likely to apply to any other HIV-2 cohort.

Summary

In the present study, we determined and analysed the frequencies of classical *HLA* class I and *KIR* genes in 1113 consented adults (513 from a remote community cohort in Caio, Guinea Bissau, and 600 from an urban clinical setting in Fajara, The

Gambia). We detected 24 *HLA-A*, 32 *HLA-B* and 21 *HLA-C* alleles (4- to 6-digits) in Caio while in Fajara we found 28 *HLA-A*, 50 *HLA-B* and 23 *HLA-C* alleles (4- to 6-digits).

Individuals from the predominantly Manjako community in Caio who were positive for *HLA-A*8001* or *HLA-B*080101* were at higher risk of acquiring HIV-2 infection than those without any of these alleles while those expressing either KIR2DL2 or KIR2DS2 molecules together with at least one copy of their corresponding ligands (C1) were protected against HIV-2 infection. In Fajara, two *HLA-B* alleles (*B*14, B*57*) showed weak protection against HIV-2 acquisition (p=0.031 and p=0.029, respectively).

In the community cohort in Caio Guinea Bissau, HIV-2 infected individuals carrying the *B*1503* (but not *B*1510*) allele had significantly higher HIV-2 viral loads and lower CD4 counts compared to those without this allele, suggesting that this allele is associated with poor control of viral replication and more rapid disease progression. The mechanisms of action of *B*1503* in mediating this susceptibility effect is yet to be elucidated.

None of the HLA class I haplotypes and/or KIR gene profiles was associated with neither susceptibility/resistance to HIV-2 infection nor disease progression assessed using universal markers of HIV progression ($CD4^+$ T cell count and viral load) in both of these West African cohorts. KIR profile A1 was the most frequent profiles but its frequency was similar between cases and controls.

We observed a few differences in allele and genotype frequencies between the two study populations. One of the reasons that could help in explaining these differences could be in population diversity. Although both cohorts are located in the coast of West Africa, one is an isolated rural community cohort with the majority of the

participants (>95%) belonging to a single ethnic group (Manjako). The other (Fajara cohort) is located in an urban area with participants drawn from more than eight different ethnic groups. Epidemiological evidences have demonstrated that the Caio population is sexually isolated from neighbouring populations and more open to Europeans particular those from Portugal and France.

Although, the date HIV-2 entered into each of the West African populations is not well known, it seems unlikely that HIV-2 has been around long enough to impose some form of selection pressure on those who survive the infection. If selection pressure can be used to explain the observed differences, it is likely pressure from other retroviruses closely related to HIV-2 such as HTLV-1 and HTLV-2, which are common in older women of Guinea Bissau origin[223 224].

Strengths and weaknesses

Studies described in this thesis are among the very few conducted in Sub Saharan Africa employing state-of-the-art molecular techniques such as SBT and PCR-SSP typing methods to detect the presence and describe the frequency of genes from highly polymorphic regions of the human genome – MHC and LRC regions.

This is the first time that HLA class I, KIR and KIR-HLA compound genotypes are studied in great details in a sizable number of HIV-2 infected individuals in West Africa. This will no doubt form the basis for more advanced immunological studies in these and other HIV-2 cohorts to understand better why HIV-2 is well controlled and less aggressive to the human host than HIV-1.

These studies also confirm that there is high diversity in African populations even in a geographically small region. One lesson learned from this work is that

genetic studies carried out in specific ethnic groups may be more informative than those that study a more heterogeneous population. We have shown in a predominantly homogenous population (>95% Manjako) that people with HLA-B*1503 genotype when infected with HIV-2 are more likely to progress to AIDS defining illnesses faster than those without this genotype. Secondly, HLA alleles that were associated with susceptibility to or protection against HIV-2 in Caio (albeit weakly) did not show any effect in a multi-ethnic Fajara cohort.

An inherent weakness of this study is the lack of knowledge on the date of seroconversion of HIV-2 infected subjects. The infected subjects in Caio were mostly recruited during serosurveys. The length of time between the serosurveys was quite long to see any trend in data collected between different time points. In Fajara, the ethnicity data was not available for the majority of those infected with HIV-2, and we were therefore unable to use the available CD4 count and HIV-2 viral load data to predict disease progression in that cohort.

Although a sample size of 1113 individuals can be considered large compared to previous studies, it is still a small sample size with a weak power to detect small but clinically significant differences between groups. We therefore recommend a merging and consolidation of all HIV-2 cohorts around the globe to form a consortium for future immunogenetics studies.

There is paucity of data generated using SBT on HLA and KIR genes in the West African sub region and we had no choice but to compare our data with those available in the literature and public databases irrespective of the techniques used in generating the data.

Future work

As mentioned earlier, data generated from this study will form the basis for more advanced functional work in the future. Some of the areas that we are keen on investigating include the following:

1. The role of *HLA-B*1503* in the immune response to HIV-2.

   In this study, we plan to use a series of *in vitro* immunological techniques to understand the immunological basis of *HLA-B*1503* association with rapid disease progression in HIV-2 infected individuals. This will be done in two phases: (a) first, we will investigate the functional and phenotypic properties (cell surface markers that can help in their identification) of HLA-B*1503-restricted HIV-2 specific $CD8^+$ T cells to see if they generate a particular profile of soluble factors that promote HIV-2 replication *in vitro;* (b) second, through binding and affinity assays we will determine which of the KIR molecule(s) use HLA-B*1503 as their ligand and then investigate the nature of KIR interactions with HLA-B*1503 in modulating NK cell cytolytic activities. KIR-ligand interaction results in transduction of either an activating or an inhibitory signal to the NK cell. The overall balance of both signals is very important in determining what the action of NK cell will be on the target cell. If the activating signal predominates over the inhibitory one, the NK cell is stimulated to carry out its cytolytic function, but if the opposite is the case, the NK cell function will be inhibited.

2. Determination of new *KIR* and *HLA* class I alleles in two West African populations

   Here we plan to carry out full length sequencing of those KIR genes that showed unusual patterns with PCR-SSP techniques. We shall concentrate on those KIR genes that consistently showed a specific band on one but not the other

of the two primer pairs that were designed to detect the presence of the gene by annealing to two highly conserved regions of the gene. We shall design new primers to amplify the region encompassing both primer pairs, sequence that region and analyse it for the presence of mutation(s) at the primer binding sites. If mutations are detected that have not yet been described in public databases, we shall then design more primers to sequence the entire gene.

For the *HLA* class I genes, we plan to re-sequence all the samples that we could not amplify during the course of this project and those that we excluded from analysis because they had one or more mismatches compared to alleles already described in HLA and Immunogenetics databases.

We also plan to deposit complete sequence traces of all new *KIR* and *HLA* alleles in public databases to benefit the scientific community.

3. Paper write up and publication

We have submitted one paper titled "The influence of *HLA* class I and *HLA-KIR* compound genotypes on HIV-2 infection and markers of disease progression in a Manjako community in West Africa", which is currently under review for publication in the Journal of Immunology (see **Appendix 15**). Two more manuscripts are being drafted to describe the genetic variants in *HLA* class I and *KIR* genes in both Caio and Fajara populations.

REFERENCES

1. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983;220(4599):868-71.
2. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* 1983;220(4599):865-7.
3. Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 1986;233(4761):343-6.
4. UNAIDS. *2008 Report on the global AIDS epidemic.* Geneva: UNAIDS, 2008.
5. Odehouri K, De Cock KM, Krebs JW, Moreau J, Rayfield M, McCormick JB, et al. HIV-1 and HIV-2 infection associated with AIDS in Abidjan, Cote d'Ivoire. *Aids* 1989;3(8):509-12.
6. Chen Z, Telfier P, Gettie A, Reed P, Zhang L, Ho DD, et al. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J Virol* 1996;70(6):3617-27.
7. Santiago ML, Range F, Keele BF, Li Y, Bailes E, Bibollet-Ruche F, et al. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J Virol* 2005;79(19):12515-27.
8. Van Heuverswyn F, Peeters M. The Origins of HIV and Implications for the Global Epidemic. *Curr Infect Dis Rep* 2007;9(4):338-346.
9. Lockett SF, Robertson JR, Brettle RP, Yap PL, Middleton D, Leigh Brown AJ. Mismatched human leukocyte antigen alleles protect against heterosexual HIV transmission. *J Acquir Immune Defic Syndr* 2001;27(3):277-80.
10. Liegeois F, Lafay B, Formenty P, Locatelli S, Courgnaud V, Delaporte E, et al. Full-length genome characterization of a novel simian immunodeficiency virus lineage (SIVolc) from olive Colobus (Procolobus verus) and new SIVwrcPbb strains from Western Red Colobus (Piliocolobus badius badius) from the Tai Forest in Ivory Coast. *J Virol* 2009;83(1):428-39.
11. Locatelli S, Lafay B, Liegeois F, Ting N, Delaporte E, Peeters M. Full molecular characterization of a simian immunodeficiency virus, SIVwrcpbt from Temminck's red colobus (Piliocolobus badius temminckii) from Abuko Nature Reserve, The Gambia. *Virology* 2008;376(1):90-100.
12. Locatelli S, Liegeois F, Lafay B, Roeder AD, Bruford MW, Formenty P, et al. Prevalence and genetic diversity of simian immunodeficiency virus infection in wild-living red colobus monkeys (Piliocolobus badius badius) from the Tai forest, Cote d'Ivoire SIVwrc in wild-living western red colobus monkeys. *Infect Genet Evol* 2008;8(1):1-14.
13. Kawamura M, Yamazaki S, Ishikawa K, Kwofie TB, Tsujimoto H, Hayami M. HIV-2 in west Africa in 1966. *Lancet* 1989;1(8634):385.
14. Bryceson A, Tomkins A, Ridley D, Warhurst D, Goldstone A, Bayliss G, et al. HIV-2-associated AIDS in the 1970s. *Lancet* 1988;2(8604):221.

15. Barin F, M'Boup S, Denis F, Kanki P, Allan JS, Lee TH, et al. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of west Africa. *Lancet* 1985;2(8469-70):1387-9.
16. Poulsen AG, Aaby P, Gottschau A, Kvinesdal BB, Dias F, Molbak K, et al. HIV-2 infection in Bissau, West Africa, 1987-1989: incidence, prevalences, and routes of transmission. *J Acquir Immune Defic Syndr* 1993;6(8):941-8.
17. Poulsen AG, Kvinesdal B, Aaby P, Molbak K, Frederiksen K, Dias F, et al. Prevalence of and mortality from human immunodeficiency virus type 2 in Bissau, West Africa. *Lancet* 1989;1(8642):827-31.
18. Guyader M, Emerman M, Sonigo P, Clavel F, Montagnier L, Alizon M. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* 1987;326(6114):662-9.
19. Peterlin BM, Luciw PA. Molecular biology of HIV. *Aids* 1988;2 Suppl 1:S29-40.
20. Cann AJ, Karn J. Molecular biology of HIV: new insights into the virus life-cycle. *Aids* 1989;3 Suppl 1:S19-34.
21. Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* 1994;265(5178):1587-90.
22. McKnight A, Dittmar MT, Moniz-Periera J, Ariyoshi K, Reeves JD, Hibbitts S, et al. A broad range of chemokine receptors are used by primary isolates of human immunodeficiency virus type 2 as coreceptors with CD4. *J Virol* 1998;72(5):4065-71.
23. Owen SM, Ellenberger D, Rayfield M, Wiktor S, Michel P, Grieco MH, et al. Genetically divergent strains of human immunodeficiency virus type 2 use multiple coreceptors for viral entry. *J Virol* 1998;72(7):5425-32.
24. Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. *Science* 1988;242(4882):1168-71.
25. Palaniappan C, Wisniewski M, Wu W, Fay PJ, Bambara RA. Misincorporation by HIV-1 reverse transcriptase promotes recombination via strand transfer synthesis. *J Biol Chem* 1996;271(37):22331-8.
26. Lewthwaite P. Natural history of HIV/AIDS. *Medicine* 2005;33(6):10.
27. Pantaleo G, Fauci AS. Immunopathogenesis of HIV infection. *Annu Rev Microbiol* 1996;50:825-54.
28. Paranjape RS. Immunopathogenesis of HIV infection. *Indian J Med Res* 2005;121(4):240-55.
29. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J Virol* 1994;68(9):6103-10.
30. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol* 1994;68(7):4650-5.
31. Graziosi C, Pantaleo G, Butini L, Demarest JF, Saag MS, Shaw GM, et al. Kinetics of human immunodeficiency virus type 1 (HIV-1) DNA and RNA synthesis during primary HIV-1 infection. *Proc Natl Acad Sci U S A* 1993;90(14):6405-9.
32. Daar ES, Moudgil T, Meyer RD, Ho DD. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. *N Engl J Med* 1991;324(14):961-4.
33. Moore JP, Cao Y, Ho DD, Koup RA. Development of the anti-gp120 antibody response during seroconversion to human immunodeficiency virus type 1. *J Virol* 1994;68(8):5142-55.

34. Pantaleo G, Demarest JF, Soudeyns H, Graziosi C, Denis F, Adelsberger JW, et al. Major expansion of CD8+ T cells with a predominant V beta usage during the primary immune response to HIV. *Nature* 1994;370(6489):463-7.
35. Graziosi C, Gantt KR, Vaccarezza M, Demarest JF, Daucher M, Saag MS, et al. Kinetics of cytokine expression during primary human immunodeficiency virus type 1 infection. *Proc Natl Acad Sci U S A* 1996;93(9):4386-91.
36. Baier M, Werner A, Bannert N, Metzner K, Kurth R. HIV suppression by interleukin-16. *Nature* 1995;378(6557):563.
37. Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* 1995;270(5243):1811-5.
38. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* 1991;354(6353):453-9.
39. Pantaleo G, Graziosi C, Butini L, Pizzo PA, Schnittman SM, Kotler DP, et al. Lymphoid organs function as major reservoirs for human immunodeficiency virus. *Proc Natl Acad Sci U S A* 1991;88(21):9838-42.
40. Pantaleo G, Graziosi C, Demarest JF, Cohen OJ, Vaccarezza M, Gantt K, et al. Role of lymphoid organs in the pathogenesis of human immunodeficiency virus (HIV) infection. *Immunol Rev* 1994;140:105-30.
41. Piatak M, Jr., Saag MS, Yang LC, Clark SJ, Kappes JC, Luk KC, et al. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* 1993;259(5102):1749-54.
42. Haynes BF, Pantaleo G, Fauci AS. Toward an understanding of the correlates of protective immunity to HIV infection. *Science* 1996;271(5247):324-8.
43. Pantaleo G, Demarest JF, Vaccarezza M, Graziosi C, Bansal GP, Koenig S, et al. Effect of anti-V3 antibodies on cell-free and cell-to-cell human immunodeficiency virus transmission. *Eur J Immunol* 1995;25(1):226-31.
44. Berman PW, Gregory TJ, Riddle L, Nakamura GR, Champe MA, Porter JP, et al. Protection of chimpanzees from infection by HIV-1 after vaccination with recombinant glycoprotein gp120 but not gp160. *Nature* 1990;345(6276):622-5.
45. Emini EA, Nara PL, Schleif WA, Lewis JA, Davide JP, Lee DR, et al. Antibody-mediated in vitro neutralization of human immunodeficiency virus type 1 abolishes infectivity for chimpanzees. *J Virol* 1990;64(8):3674-8.
46. Gloster SE, Newton P, Cornforth D, Lifson JD, Williams I, Shaw GM, et al. Association of strong virus-specific CD4 T cell responses with efficient natural control of primary HIV-1 infection. *Aids* 2004;18(5):749-55.
47. McKay PF, Barouch DH, Schmitz JE, Veazey RS, Gorgone DA, Lifton MA, et al. Global dysfunction of CD4 T-lymphocyte cytokine expression in simian-human immunodeficiency virus/SIV-infected monkeys is prevented by vaccination. *J Virol* 2003;77(8):4695-702.
48. Ostrowski SR, Gerstoft J, Pedersen BK, Ullum H. Impaired production of cytokines is an independent predictor of mortality in HIV-1-infected patients. *Aids* 2003;17(4):521-30.
49. Imami N, Pires A, Hardy G, Wilson J, Gazzard B, Gotch F. A balanced type 1/type 2 response is associated with long-term nonprogressive human immunodeficiency virus type 1 infection. *J Virol* 2002;76(18):9011-23.
50. Duvall MG, Jaye A, Dong T, Brenchley JM, Alabi AS, Jeffries DJ, et al. Maintenance of HIV-specific CD4+ T cell help distinguishes HIV-2 from HIV-1 infection. *J Immunol* 2006;176(11):6973-81.

51. Cao Y, Qin L, Zhang L, Safrit J, Ho DD. Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. *N Engl J Med* 1995;332(4):201-8.
52. Harrer T, Harrer E, Kalams SA, Barbosa P, Trocha A, Johnson RP, et al. Cytotoxic T lymphocytes in asymptomatic long-term nonprogressing HIV-1 infection. Breadth and specificity of the response and relation to in vivo viral quasispecies in a person with prolonged infection and low viral load. *J Immunol* 1996;156(7):2616-23.
53. Leligdowicz A, Yindom LM, Onyango C, Sarge-Njie R, Alabi A, Cotten M, et al. Robust Gag-specific T cell responses characterize viremia control in HIV-2 infection. *J Clin Invest* 2007;117(10):3067-74.
54. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med* 1997;3(2):205-11.
55. Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, Troop M, et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A* 1997;94(5):1890-5.
56. Kalams SA, Buchbinder SP, Rosenberg ES, Billingsley JM, Colbert DS, Jones NG, et al. Association between virus-specific cytotoxic T-lymphocyte and helper responses in human immunodeficiency virus type 1 infection. *J Virol* 1999;73(8):6715-20.
57. Goepfert PA, Bansal A, Edwards BH, Ritter GD, Jr., Tellez I, McPherson SA, et al. A significant number of human immunodeficiency virus epitope-specific cytotoxic T lymphocytes detected by tetramer binding do not produce gamma interferon. *J Virol* 2000;74(21):10249-55.
58. Kostense S, Vandenberghe K, Joling J, Van Baarle D, Nanlohy N, Manting E, et al. Persistent numbers of tetramer+ CD8(+) T cells, but loss of interferon-gamma+ HIV-specific T cells during progression to AIDS. *Blood* 2002;99(7):2505-11.
59. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, et al. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 1991;352(6336):595-600.
60. Hill AV, Bennett S, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, et al. HLA, malaria and dominant protective associations. *Parasitol Today* 1992;8(2):57.
61. Hill AV, Elvin J, Willis AC, Aidoo M, Allsopp CE, Gotch FM, et al. Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature* 1992;360(6403):434-9.
62. Bellamy R. Identifying genetic susceptibility factors for tuberculosis in Africans: a combined approach using a candidate gene study and a genome-wide screen. *Clin Sci (Lond)* 2000;98(3):245-50.
63. Mahdi OS. Impact of host genetics on susceptibility to human Chlamydia trachomatis disease. *Br J Biomed Sci* 2002;59(2):128-32.
64. Blackwell JM. Genetics of host resistance and susceptibility to intramacrophage pathogens: a study of multicase families of tuberculosis, leprosy and leishmaniasis in north-eastern Brazil. *Int J Parasitol* 1998;28(1):21-8.
65. Frodsham AJ, Hill AV. Genetics of infectious diseases. *Hum Mol Genet* 2004;13 Spec No 2:R187-94.
66. Hill AV. The immunogenetics of human infectious diseases. *Annu Rev Immunol* 1998;16:593-617.
67. Hill AV, Yates SN, Allsopp CE, Gupta S, Gilbert SC, Lalvani A, et al. Human leukocyte antigens and natural selection by malaria. *Philos Trans R Soc Lond B Biol Sci* 1994;346(1317):379-85.

68. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* 2000;97(6):2709-14.
69. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, et al. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 1999;283(5408):1748-52.
70. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med* 2003;54:535-51.
71. Thio CL, Carrington M, Marti D, O'Brien SJ, Vlahov D, Nelson KE, et al. Class II HLA alleles and hepatitis B virus persistence in African Americans. *J Infect Dis* 1999;179(4):1004-6.
72. Gupta S, Hill AV. Dynamic interactions in malaria: host heterogeneity meets parasite polymorphism. *Proc Biol Sci* 1995;261(1362):271-7.
73. Rowland-Jones S, Sutton J, Ariyoshi K, Dong T, Gotch F, McAdam S, et al. HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women. *Nat Med* 1995;1(1):59-64.
74. Rowland-Jones SL, Dong T, Fowke KR, Kimani J, Krausa P, Newell H, et al. Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. *J Clin Invest* 1998;102(9):1758-65.
75. Fowke KR, Nagelkerke NJ, Kimani J, Simonsen JN, Anzala AO, Bwayo JJ, et al. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet* 1996;348(9038):1347-51.
76. Rowland-Jones S, Tan R, McMichael A. Role of cellular immunity in protection against HIV infection. *Adv Immunol* 1997;65:277-346.
77. Gotch F, Gallimore A, McMichael A. Cytotoxic T cells--protection from disease progression--protection from infection. *Immunol Lett* 1996;51(1-2):125-8.
78. O'Brien SJ, Nelson GW. Human genes that limit AIDS. *Nat Genet* 2004;36(6):565-74.
79. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 1996;273(5283):1856-62.
80. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med* 1996;2(11):1240-3.
81. Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB. Global distribution of the CCR5 gene 32-basepair deletion. *Nat Genet* 1997;16(1):100-3.
82. An P, Nelson GW, Wang L, Donfield S, Goedert JJ, Phair J, et al. Modulating influence on HIV/AIDS by interacting RANTES gene variants. *Proc Natl Acad Sci U S A* 2002;99(15):10002-7.
83. Smith MW, Carrington M, Winkler C, Lomb D, Dean M, Huttley G, et al. CCR2 chemokine receptor and AIDS progression. *Nat Med* 1997;3(10):1052-3.
84. Winkler C, Modi W, Smith MW, Nelson GW, Wu X, Carrington M, et al. Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). *Science* 1998;279(5349):389-93.

85. Smolnikova MV, Konenkov VI. Association of IL2, TNFA, IL4 and IL10 Promoter Gene Polymorphisms with the Rate of Progression of the HIV Infection. *Russ J Immunol* 2002;7(4):349-56.
86. Donninger H, Cashmore TJ, Scriba T, Petersen DC, Janse van Rensburg E, Hayes VM. Functional analysis of novel SLC11A1 (NRAMP1) promoter variants in susceptibility to HIV-1. *J Med Genet* 2004;41(4):e49.
87. Shin HD, Winkler C, Stephens JC, Bream J, Young H, Goedert JJ, et al. Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of IL10. *Proc Natl Acad Sci U S A* 2000;97(26):14467-72.
88. Van R, Van L. Leukocyte Grouping. A Method and Its Application. *J Clin Invest* 1963;42:1382-90.
89. Zinkernagel RM, Doherty PC. The discovery of MHC restriction. *Immunol Today* 1997;18(1):14-7.
90. Doherty PC, Zinkernagel RM. A biological role for the major histocompatibility antigens. *Lancet* 1975;1(7922):1406-9.
91. Klein J. *Biology of the Mouse Histocompatibility Complex*. Berlin: Springer-Verlag, 1975.
92. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med* 2000;343(10):702-9.
93. Forbes SA, Trowsdale J. The MHC quarterly report. *Immunogenetics* 1999;50(3-4):152-9.
94. Robinson J, Waller MJ, Fail SC, McWilliam H, Lopez R, Parham P, et al. The IMGT/HLA database. *Nucleic Acids Res* 2009;37(Database issue):D1013-7.
95. Francioli P, Clement F, Vaudois CH. Beta 2-microglobulin and immunodeficiency in a homosexual man. *N Engl J Med* 1982;307(22):1402-3.
96. Bhalla RB, Safai B, Mertelsmann R, Schwartz MK. Abnormally high concentrations of beta 2 microglobulin in acquired immunodeficiency syndrome (AIDS) patients. *Clin Chem* 1983;29(8):1560.
97. Lambin P, Lefrere JJ, Doinel C, Fine JM, Salmon D, Salmon C. Neopterin and beta 2-microglobulin in serum of HIV-seropositive subjects during a two-year follow-up. *Clin Chem* 1988;34(6):1367-8.
98. Marsh SG, Parham P, Barber LD. *The HLA Facts Book*. London: Academic Press, 2000.
99. Klein J. *Natural history of the major histocompatibility complex*. New York: Wiley, 1986.
100. Trowsdale J. Molecular genetics of HLA class I and class II regions. In: Browning M, McMichael, A., editor. *HLA and MHC: Genes, Molecules and Function*. Oxford: Bios Scientific Publishers, 1996:23-38.
101. Trowsdale J, Young JA, Kelly AP, Austin PJ, Carson S, Meunier H, et al. Structure, sequence and polymorphism in the HLA-D region. *Immunol Rev* 1985;85:5-43.
102. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, et al. Nomenclature for factors of the HLA system, 2004. *Tissue Antigens* 2005;65(4):301-69.
103. Fehling HJ, Gilfillan S, Ceredig R. Alpha beta/gamma delta lineage commitment in the thymus of normal and genetically manipulated mice. *Adv Immunol* 1999;71:1-76.
104. Sebzda E, Mariathasan S, Ohteki T, Jones R, Bachmann MF, Ohashi PS. Selection of the T cell repertoire. *Annu Rev Immunol* 1999;17:829-74.
105. Ellmeier W, Sawada S, Littman DR. The regulation of CD4 and CD8 coreceptor gene expression during T cell development. *Annu Rev Immunol* 1999;17:523-54.
106. Parham P. Virtual reality in the MHC. *Immunol Rev* 1999;167:5-15.
107. Bodmer W. HLA polymorphism: Origin and Maintenance. In: Terasaki PI, Gjertson DW, editors. *HLA 1997*. Los Angeles: The Reagents of the University of California, 1997.
108. Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science* 1996;272:67-73.

109. Little AM, Parham P. Polymorphism and evolution of HLA class I and II genes and molecules. *Rev Immunogenet* 1999;1(1):105-23.
110. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, et al. Recent segmental duplications in the human genome. *Science* 2002;297(5583):1003-7.
111. Zangenberg G, Huang MM, Arnheim N, Erlich H. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nat Genet* 1995;10(4):407-14.
112. Satta Y, Mayer WE, Klein J. HLA-DRB intron 1 sequences: implications for the evolution of HLA-DRB genes and haplotypes. *Hum Immunol* 1996;51(1):1-12.
113. Beyrer C, Artenstein AW, Rugpao S, Stephens H, VanCott TC, Robb ML, et al. Epidemiologic and biologic characterization of a cohort of human immunodeficiency virus type 1 highly exposed, persistently seronegative female sex workers in northern Thailand. Chiang Mai HEPS Working Group. *J Infect Dis* 1999;179(1):59-67.
114. MacDonald KS, Fowke KR, Kimani J, Dunand VA, Nagelkerke NJ, Ball TB, et al. Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 2000;181(5):1581-9.
115. Diouf K, Sarr AD, Eisen G, Popper S, Mboup S, Kanki P. Associations between MHC class I and susceptibility to HIV-2 disease progression. *J Hum Virol* 2002;5(1):1-7.
116. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* 1996;2(4):405-11.
117. Roger M. Influence of host genes on HIV-1 disease progression. *Faseb J* 1998;12(9):625-32.
118. Tang J, Costello C, Keet IP, Rivers C, Leblanc S, Karita E, et al. HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection. *AIDS Res Hum Retroviruses* 1999;15(4):317-24.
119. Shearer GM, Clerici M. Protective immunity against HIV infection: has nature done the experiment for us? *Immunol Today* 1996;17(1):21-4.
120. MacDonald KS, Embree JE, Nagelkerke NJ, Castillo J, Ramhadin S, Njenga S, et al. The HLA A2/6802 supertype is associated with reduced risk of perinatal human immunodeficiency virus type 1 transmission. *J Infect Dis* 2001;183(3):503-6.
121. MacDonald KS, Embree J, Njenga S, Nagelkerke NJ, Ngatia I, Mohammed Z, et al. Mother-child class I HLA concordance increases perinatal human immunodeficiency virus type 1 transmission. *J Infect Dis* 1998;177(3):551-6.
122. Polycarpou A, Ntais C, Korber BT, Elrich HA, Winchester R, Krogstad P, et al. Association between maternal and infant class I and II HLA alleles and of their concordance with the risk of perinatal HIV type 1 transmission. *AIDS Res Hum Retroviruses* 2002;18(11):741-6.
123. Winchester R, Chen Y, Rose S, Selby J, Borkowsky W. Major histocompatibility complex class II DR alleles DRB1*1501 and those encoding HLA-DR13 are preferentially associated with a diminution in maternally transmitted human immunodeficiency virus 1 infection in different ethnic groups: determination by an automated sequence-based typing method. *Proc Natl Acad Sci U S A* 1995;92(26):12374-8.
124. De Cock KM, Adjorlolo G, Ekpini E, Sibailly T, Kouadio J, Maran M, et al. Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. *Jama* 1993;270(17):2083-6.
125. Kanki PJ, Eisen G, Travers KU, Marlink RG, Essex ME, Hsieh CC, et al. Response: HIV-2 and Natural Protection Against HIV-1 Infection. *Science* 1996;272(5270):1959b-1960b.

126. Travers KU, Eisen GE, Marlink RG, Essex ME, Hsieh CC, Mboup S, et al. Protection from HIV-1 infection by HIV-2. *Aids* 1998;12(2):224-5.
127. Schim van der Loeff MF, Aaby P, Aryioshi K, Vincent T, Awasana AA, Da Costa C, et al. HIV-2 does not protect against HIV-1 infection in a rural community in Guinea-Bissau. *Aids* 2001;15(17):2303-10.
128. de Sorrentino AH, Marinic K, Motta P, Sorrentino A, Lopez R, Illiovich E. HLA class I alleles associated with susceptibility or resistance to human immunodeficiency virus type 1 infection among a population in Chaco Province, Argentina. *J Infect Dis* 2000;182(5):1523-6.
129. Roe DL, Lewis RE, Cruse JM. Association of HLA-DQ and -DR alleles with protection from or infection with HIV-1. *Exp Mol Pathol* 2000;68(1):21-8.
130. Geczy AF, Kuipers H, Coolen M, Ashton LJ, Kennedy C, Ng G, et al. HLA and other host factors in transfusion-acquired HIV-1 infection. *Hum Immunol* 2000;61(2):172-6.
131. Keet IP, Tang J, Klein MR, LeBlanc S, Enger C, Rivers C, et al. Consistent associations of HLA class I and II and transporter gene products with progression of human immunodeficiency virus type 1 infection in homosexual men. *J Infect Dis* 1999;180(2):299-309.
132. Flores-Villanueva PO, Yunis EJ, Delgado JC, Vittinghoff E, Buchbinder S, Leung JY, et al. Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proc Natl Acad Sci U S A* 2001;98(9):5140-5.
133. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R, et al. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* 2001;344(22):1668-75.
134. Hendel H, Caillat-Zucman S, Lebuanec H, Carrington M, O'Brien S, Andrieu JM, et al. New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol* 1999;162(11):6942-6.
135. McNeil AJ, Yap PL, Gore SM, Brettle RP, McColl M, Wyld R, et al. Association of HLA types A1-B8-DR3 and B27 with rapid and slow progression of HIV disease. *Qjm* 1996;89(3):177-85.
136. Costello C, Tang J, Rivers C, Karita E, Meizen-Derr J, Allen S, et al. HLA-B*5703 independently associated with slower HIV-1 disease progression in Rwandan women. *Aids* 1999;13(14):1990-1.
137. Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, et al. Advantage of rare HLA supertype in HIV disease progression. *Nat Med* 2003;9(7):928-35.
138. Lama J, Planelles V. Host factors influencing susceptibility to HIV infection and AIDS progression. *Retrovirology* 2007;4:52.
139. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007;317(5840):944-7.
140. Fellay J. Host genetics influences on HIV type-1 disease. *Antivir Ther* 2009;14(6):731-8.
141. Trinchieri G. Biology of natural killer cells. *Adv Immunol* 1989;47:187-376.
142. Hansasuta P, Dong T, Thananchai H, Weekes M, Willberg C, Aldemir H, et al. Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific. *Eur J Immunol* 2004;34(6):1673-9.
143. Lanier LL. NK cell recognition. *Annu Rev Immunol* 2005;23:225-74.
144. Biron CA, Brossay L. NK cells and NKT cells in innate defense against viral infections. *Curr Opin Immunol* 2001;13(4):458-64.
145. Lanier LL, Phillips JH. Natural killer cells. *Curr Opin Immunol* 1992;4(1):38-42.
146. Ljunggren HG, Karre K. In search of the 'missing self': MHC molecules and NK cell recognition. *Immunol Today* 1990;11(7):237-44.

147. Parham P. The genetic and evolutionary balances in human NK cell receptor diversity. *Semin Immunol* 2008;20(6):311-6.
148. Mingari MC, Schiavetti F, Ponte M, Vitale C, Maggi E, Romagnani S, et al. Human CD8+ T lymphocyte subsets that express HLA class I-specific inhibitory receptors represent oligoclonally or monoclonally expanded cell populations. *Proc Natl Acad Sci U S A* 1996;93(22):12433-8.
149. Bauer S, Groh V, Wu J, Steinle A, Phillips JH, Lanier LL, et al. Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* 1999;285(5428):727-9.
150. Long EO. Regulation of immune responses through inhibitory receptors. *Annu Rev Immunol* 1999;17:875-904.
151. Long EO, Rajagopalan S. HLA class I recognition by killer cell Ig-like receptors. *Semin Immunol* 2000;12(2):101-8.
152. Robinson J, Waller MJ, Stoehr P, Marsh SG. IPD--the Immuno Polymorphism Database. *Nucleic Acids Res* 2005;33(Database issue):D523-6.
153. Carrington M, Martin MP. The impact of variation at the KIR gene cluster on human disease. *Curr Top Microbiol Immunol* 2006;298:225-57.
154. Hsu KC, Chida S, Geraghty DE, Dupont B. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunol Rev* 2002;190:40-52.
155. Vilches C, Parham P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu Rev Immunol* 2002;20:217-51.
156. Selvakumar A, Steffens U, Palanisamy N, Chaganti RS, Dupont B. Genomic organization and allelic polymorphism of the human killer cell inhibitory receptor gene KIR103. *Tissue Antigens* 1997;49(6):564-73.
157. IPD-KIR. Release 1.4.0, 2007.
158. Uhrberg M, Valiante NM, Shum BP, Shilling HG, Lienert-Weidenbach K, Corliss B, et al. Human diversity in killer cell inhibitory receptor genes. *Immunity* 1997;7(6):753-63.
159. Yawata M, Yawata N, Draghi M, Little AM, Partheniou F, Parham P. Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function. *J Exp Med* 2006;203(3):633-45.
160. Toneva M, Lepage V, Lafay G, Dulphy N, Busson M, Lester S, et al. Genomic diversity of natural killer cell receptor genes in three populations. *Tissue Antigens* 2001;57(4):358-62.
161. Martin AM, Freitas EM, Witt CS, Christiansen FT. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics* 2000;51(4-5):268-80.
162. Shilling HG, Guethlein LA, Cheng NW, Gardiner CM, Rodriguez R, Tyan D, et al. Allelic polymorphism synergizes with variable gene content to individualize human KIR genotype. *J Immunol* 2002;168(5):2307-15.
163. Wilson MJ, Torkar M, Haude A, Milne S, Jones T, Sheer D, et al. Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci U S A* 2000;97(9):4778-83.
164. Khakoo SI, Rajalingam R, Shum BP, Weidenbach K, Flodin L, Muir DG, et al. Rapid evolution of NK cell receptor systems demonstrated by comparison of chimpanzees and humans. *Immunity* 2000;12(6):687-98.
165. Trowsdale J. Genetic and functional relationships between MHC and NK receptor genes. *Immunity* 2001;15(3):363-74.

166. Fahlen L, Lendahl U, Sentman CL. MHC class I-Ly49 interactions shape the Ly49 repertoire on murine NK cells. *J Immunol* 2001;166(11):6585-92.
167. McDade TW, Worthman CM. Evolutionary process and the ecology of human immune function. *Am J Hum Biol* 1999;11(6):705-717.
168. Carrington M, Wang S, Martin MP, Gao X, Schiffman M, Cheng J, et al. Hierarchy of resistance to cervical neoplasia mediated by combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci. *J Exp Med* 2005;201(7):1069-75.
169. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, Goedert JJ, et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet* 2002;31(4):429-34.
170. van der Slik AR, Koeleman BP, Verduijn W, Bruining GJ, Roep BO, Giphart MJ. KIR in type 1 diabetes: disparate distribution of activating and inhibitory natural killer cell receptors in patients versus HLA-matched control subjects. *Diabetes* 2003;52(10):2639-42.
171. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, Astemborski J, et al. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 2004;305(5685):872-4.
172. Nelson GW, Martin MP, Gladman D, Wade J, Trowsdale J, Carrington M. Cutting edge: heterozygote advantage in autoimmune disease: hierarchy of protection/susceptibility conferred by HLA and killer Ig-like receptor combinations in psoriatic arthritis. *J Immunol* 2004;173(7):4273-6.
173. Hiby SE, Walker JJ, O'Shaughnessy K M, Redman CW, Carrington M, Trowsdale J, et al. Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J Exp Med* 2004;200(8):957-65.
174. Luszczek W, Manczak M, Cislo M, Nockowski P, Wisniewski A, Jasek M, et al. Gene for the activating natural killer cell receptor, KIR2DS1, is associated with susceptibility to psoriasis vulgaris. *Hum Immunol* 2004;65(7):758-66.
175. Naumova E, Mihaylova A, Stoitchkov K, Ivanova M, Quin L, Toneva M. Genetic polymorphism of NK receptors and their ligands in melanoma patients: prevalence of inhibitory over activating signals. *Cancer Immunol Immunother* 2005;54(2):172-8.
176. Rauch A, Laird R, McKinnon E, Telenti A, Furrer H, Weber R, et al. Influence of inhibitory killer immunoglobulin-like receptors and their HLA-C ligands on resolving hepatitis C virus infection. *Tissue Antigens* 2007;69 Suppl 1:237-40.
177. Martin MP, Nelson G, Lee JH, Pellett F, Gao X, Wade J, et al. Cutting edge: susceptibility to psoriatic arthritis: influence of activating killer Ig-like receptor genes in the absence of specific HLA-C alleles. *J Immunol* 2002;169(6):2818-22.
178. Cohen GB, Gandhi RT, Davis DM, Mandelboim O, Chen BK, Strominger JL, et al. The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. *Immunity* 1999;10(6):661-71.
179. Buckner M. Manjako sex and gender [PhD], 2000.
180. Kanki P, M'Boup S, Marlink R, Travers K, Hsieh CC, Gueye A, et al. Prevalence and risk determinants of human immunodeficiency virus type 2 (HIV-2) and human immunodeficiency virus type 1 (HIV-1) in west African female prostitutes. *Am J Epidemiol* 1992;136(7):895-907.
181. Wilkins A, Ricard D, Todd J, Whittle H, Dias F, Paulo Da Silva A. The epidemiology of HIV infection in a rural area of Guinea-Bissau. *Aids* 1993;7(8):1119-22.
182. Martin MP, Carrington M. KIR locus polymorphisms: genotyping and disease association analysis. *Methods Mol Biol* 2008;415:49-64.
183. Prada N, Davis B, Jean-Pierre P, La Roche M, Duh FM, Carrington M, et al. Drug-susceptible HIV-1 infection despite intermittent fixed-dose combination

tenofovir/emtricitabine as prophylaxis is associated with low-level viremia, delayed seroconversion, and an attenuated clinical course. *J Acquir Immune Defic Syndr* 2008;49(2):117-22.
184. Gumperz JE, Barber LD, Valiante NM, Percival L, Phillips JH, Lanier LL, et al. Conserved and variable residues within the Bw4 motif of HLA-B make separable contributions to recognition by the NKB1 killer cell-inhibitory receptor. *J Immunol* 1997;158(11):5237-41.
185. Carr WH, Pando MJ, Parham P. KIR3DL1 polymorphisms that affect NK cell inhibition by HLA-Bw4 ligand. *J Immunol* 2005;175(8):5222-9.
186. Cella M, Longo A, Ferrara GB, Strominger JL, Colonna M. NK3-specific natural killer cells are selectively inhibited by Bw4-positive HLA alleles with isoleucine 80. *J Exp Med* 1994;180(4):1235-42.
187. Denis L, Sivula J, Gourraud PA, Kerdudou N, Chout R, Ricard C, et al. Genetic diversity of KIR natural killer cell markers in populations from France, Guadeloupe, Finland, Senegal and Reunion. *Tissue Antigens* 2005;66(4):267-76.
188. Modiano D, Luoni G, Petrarca V, Sodiomon Sirima B, De Luca M, Simpore J, et al. HLA class I in three West African ethnic groups: genetic distances from sub-Saharan and Caucasoid populations. *Tissue Antigens* 2001;57(2):128-37.
189. Allsopp CE, Harding RM, Taylor C, Bunce M, Kwiatkowski D, Anstey N, et al. Interethnic genetic differentiation in Africa: HLA class I antigens in The Gambia. *Am J Hum Genet* 1992;50(2):411-21.
190. Kalidi I, Fofana Y, Rahly AA, Bochu V, Dehay C, Gony J, et al. Study of HLA antigens in a population of Mali (West Africa). *Tissue Antigens* 1988;31(2):98-102.
191. Cao K, Moormann AM, Lyke KE, Masaberg C, Sumba OP, Doumbo OK, et al. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens* 2004;63(4):293-325.
192. Spinola H, Bruges-Armas J, Middleton D, Brehm A. HLA polymorphisms in Cabo Verde and Guine-Bissau inferred from sequence-based typing. *Hum Immunol* 2005;66(10):1082-92.
193. Middleton D, Menchaca L, Rood H, Komerofsky R. New allele frequency database: http://www.allelefrequencies.net. *Tissue Antigens* 2003;61(5):403-7.
194. Price P, Witt C, Allcock R, Sayer D, Garlepp M, Kok CC, et al. The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol Rev* 1999;167:257-74.
195. Single RM, Martin MP, Gao X, Meyer D, Yeager M, Kidd JR, et al. Global diversity and evidence for coevolution of KIR and HLA. *Nat Genet* 2007;39(9):1114-9.
196. Jennes W, Verheyden S, Demanet C, Adje-Toure CA, Vuylsteke B, Nkengasong JN, et al. Cutting edge: resistance to HIV-1 infection among African female sex workers is associated with inhibitory KIR in the absence of their HLA ligands. *J Immunol* 2006;177(10):6588-92.
197. Middleton D, Meenagh A, Moscoso J, Arnaiz-Villena A. Killer immunoglobulin receptor gene and allele frequencies in Caucasoid, Oriental and Black populations from different continents. *Tissue Antigens* 2008;71(2):105-13.
198. Carrington M, Martin MP, van Bergen J. KIR-HLA intercourse in HIV disease. *Trends Microbiol* 2008;16(12):620-7.
199. Chen Y, Winchester R, Korber B, Gagliano J, Bryson Y, Hutto C, et al. Influence of HLA alleles on the rate of progression of vertically transmitted HIV infection in children: association of several HLA-DR13 alleles with long-term survivorship and the potential association of HLA-A*2301 with rapid progression to AIDS. Long-Term Survivor Study. *Hum Immunol* 1997;55(2):154-62.

200. Magierowska M, Theodorou I, Debre P, Sanson F, Autran B, Riviere Y, et al. Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. *Blood* 1999;93(3):936-41.
201. Leligdowicz AM. Evaluation of T lymphocytes in HIV-2 infection in West Africa:The role of antigen-specific immune responses in disease non-progression [PhD]. Oxford University, 2008.
202. Excoffier LGL, and S. Schneider. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 2005;1:47-50.
203. Bruges Armas J, Destro-Bisol G, Lopez-Vazquez A, Couto AR, Spedini G, Gonzalez S, et al. HLA class I variation in the West African Pygmies and their genetic relationship with other African populations. *Tissue Antigens* 2003;62(3):233-42.
204. De Cock KM, Odehouri K, Colebunders RL, Adjorlolo G, Lafontaine MF, Porter A, et al. A comparison of HIV-1 and HIV-2 infections in hospitalized patients in Abidjan, Cote d'Ivoire. *Aids* 1990;4(5):443-8.
205. Jaini R, Naruse T, Kanga U, Kikkawa E, Kaur G, Inoko H, et al. Molecular diversity of the HLA-A*19 group of alleles in North Indians: possible oriental influence. *Tissue Antigens* 2002;59(6):487-91.
206. Spinola C, Bruges-Armas J, Brehm A, Spinola H. HLA-A polymorphisms in four ethnic groups from Guinea-Bissau (West Africa) inferred from sequence-based typing. *Tissue Antigens* 2008;72(6):593-8.
207. Buckner M. Manjako sex and gender [PhD], 2000.
208. Blasczyk R, Wehling J, Onaldi-Mohr D, Rebmann V, Chandanayingyong D, Grosse-Wilde H. Structural definition of the A*74 group: implications for matching in bone marrow transplantation with alternative donors. *Tissue Antigens* 1996;48(3):205-9.
209. Shen C, Zhu B, Liu M, Li S. Genetic polymorphisms at HLA-A, -B, and -DRB1 loci in Han population of Xi'an city in China. *Croat Med J* 2008;49(4):476-82.
210. Hong W, Fu Y, Chen S, Wang F, Ren X, Xu A. Distributions of HLA class I alleles and haplotypes in Northern Han Chinese. *Tissue Antigens* 2005;66(4):297-304.
211. Li S, Jiao H, Yu X, Strong AJ, Shao Y, Sun Y, et al. Human leukocyte antigen class I and class II allele frequencies and HIV-1 infection associations in a Chinese cohort. *J Acquir Immune Defic Syndr* 2007;44(2):121-31.
212. Kijak GH, Walsh AM, Koehler RN, Moqueet N, Eller LA, Eller M, et al. HLA class I allele and haplotype diversity in Ugandans supports the presence of a major east African genetic cluster. *Tissue Antigens* 2009;73(3):262-9.
213. Atasoy M, Pirim I, Bayrak OF, Ozdemir S, Ikbal M, Erdem T, et al. Association of HLA class I and class II alleles with psoriasis vulgaris in Turkish population. Influence of type I and II psoriasis. *Saudi Med J* 2006;27(3):373-6.
214. Naugler C, Liwski R. HLA risk markers for chronic myelogenous leukemia in Eastern Canada. *Leuk Lymphoma* 2009;50(2):254-9.
215. Honeyborne I, Prendergast A, Pereyra F, Leslie A, Crawford H, Payne R, et al. Control of human immunodeficiency virus type 1 is associated with HLA-B*13 and targeting of multiple gag-specific CD8+ T-cell epitopes. *J Virol* 2007;81(7):3667-72.
216. Tang J, Tang S, Lobashevsky E, Myracle AD, Fideli U, Aldrovandi G, et al. Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. *J Virol* 2002;76(16):8276-84.
217. Novitsky V, Flores-Villanueva PO, Chigwedere P, Gaolekwe S, Bussman H, Sebetso G, et al. Identification of most frequent HLA class I antigen specificities in Botswana: relevance for HIV vaccine design. *Hum Immunol* 2001;62(2):146-56.

218. Ferrari G, Currier JR, Harris ME, Finkelstein S, de Oliveira A, Barkhan D, et al. HLA-A and -B allele expression and ability to develop anti-Gag cross-clade responses in subtype C HIV-1-infected Ethiopians. *Hum Immunol* 2004;65(6):648-59.
219. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 2004;432(7018):769-75.
220. Serwanga J, Shafer LA, Pimego E, Auma B, Watera C, Rowland S, et al. Host HLA B*allele-associated multi-clade Gag T-cell recognition correlates with slow HIV-1 disease progression in antiretroviral therapy-naive Ugandans. *PLoS ONE* 2009;4(1):e4188.
221. Torimiro JN, Carr JK, Wolfe ND, Karacki P, Martin MP, Gao X, et al. HLA class I diversity among rural rainforest inhabitants in Cameroon: identification of A*2612-B*4407 haplotype. *Tissue Antigens* 2006;67(1):30-7.
222. Ellis JM, Hoyer RJ, Costello CN, Mshana RN, Quakyi IA, Mshana MN, et al. HLA-B allele frequencies in Cote d'Ivoire defined by direct DNA sequencing: identification of HLA-B*1405, B*4410, and B*5302. *Tissue Antigens* 2001;57(4):339-43.
223. Holmgren B, Aaby P, Jensen H, Larsen O, da Silva Z, Lisse IM. Increased prevalence of retrovirus infections among older women in Africa. *Scand J Infect Dis* 1999;31(5):459-66.
224. Holmgren B, da Silva Z, Larsen O, Vastrup P, Andersson S, Aaby P. Dual infections with HIV-1, HIV-2 and HTLV-I are more common in older women than in men in Guinea-Bissau. *Aids* 2003;17(2):241-53.
225. Rowland-Jones SL, Whittle HC. Out of Africa: what can we learn from HIV-2 about protective immunity to HIV-1? *Nat Immunol* 2007;8(4):329-31.
226. Bjorling E, Scarlatti G, von Gegerfelt A, Albert J, Biberfeld G, Chiodi F, et al. Autologous neutralizing antibodies prevail in HIV-2 but not in HIV-1 infection. *Virology* 1993;193(1):528-30.
227. Mackelprang RD, John-Stewart G, Carrington M, Richardson B, Rowland-Jones S, Gao X, et al. Maternal HLA homozygosity and mother-child HLA concordance increase the risk of vertical transmission of HIV-1. *J Infect Dis* 2008;197(8):1156-61.
228. Lanier LL. Up on the tightrope: natural killer cell activation and inhibition. *Nat Immunol* 2008;9(5):495-502.
229. Jonsson AH, Yokoyama WM. Natural killer cell tolerance licensing and other mechanisms. *Adv Immunol* 2009;101:27-79.
230. Berry N, Jaffar S, Schim van der Loeff M, Ariyoshi K, Harding E, N'Gom PT, et al. Low level viremia and high CD4% predict normal survival in a cohort of HIV type-2-infected villagers. *AIDS Res Hum Retroviruses* 2002;18(16):1167-73.
231. Schim van der Loeff MF, Aaby P. Towards a better understanding of the epidemiology of HIV-2. *Aids* 1999;13 Suppl A:S69-84.
232. Ricard D, Wilkins A, N'Gum PT, Hayes R, Morgan G, Da Silva AP, et al. The effects of HIV-2 infection in a rural area of Guinea-Bissau. *Aids* 1994;8(7):977-82.
233. Aaby P, Ariyoshi K, Buckner M, Jensen H, Berry N, Wilkins A, et al. Age of wife as a major determinant of male-to-female transmission of HIV-2 infection: a community study from rural West Africa. *Aids* 1996;10(13):1585-90.
234. Schmidt WP, Van Der Loeff MS, Aaby P, Whittle H, Bakker R, Buckner M, et al. Behaviour change and competitive exclusion can explain the diverging HIV-1 and HIV-2 prevalence trends in Guinea-Bissau. *Epidemiol Infect* 2008;136(4):551-61.
235. Sanchez-Mazas A, Steiner QG, Grundschober C, Tiercy JM. The molecular determination of HLA-Cw alleles in the Mandenka (West Africa) reveals a close

genetic relationship between Africans and Europeans. *Tissue Antigens* 2000;56(4):303-12.

236. Stewart CA, Laugier-Anfossi F, Vely F, Saulquin X, Riedmuller J, Tisserant A, et al. Recognition of peptide-MHC class I complexes by activating killer immunoglobulin-like receptors. *Proc Natl Acad Sci U S A* 2005;102(37):13224-9.
237. Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, et al. Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes. *Nat Immunol* 2006;7(2):173-8.

APPENDIX

APPENDIX 1

MRC Clinical Cohort Study OPDNO:|___|___|___|___|___| / |___|___|___|___|___|

Participant Information and Consent Sheet for Adults
(to be read after the patient has been post-test counselled)

AIDS is caused by a small germ, called HIV. Some people become ill within a few years of becoming infected, others may stay healthy for 10 or more years. Currently there is no cure for HIV infection. There are expensive drugs that can slow down the disease - unfortunately these are not routinely available in The Gambia due to their high costs.

The MRC has been studying HIV infections for many years. We are investigating the body's defences against this infection. We do this by using a small amount of blood in tests in the laboratory, which reveal whether the defence system is strong or weak, and over time, whether it is maintaining its strength or is losing power. Because the immune system is a complicated system, there are many different tests than can be done. Also the amount and type of virus in the blood can be measured. New tests are developed every year. We hope that these various studies will lead to the development of a vaccine in the long run. A good vaccine would prevent infection with HIV, or prevent the development of disease, including AIDS.

We invite you to enrol in our ongoing studies. We offer you clinical check-ups every three months (or more often as necessary), and medication as needed, both free of charge. Your travel fees to and from the clinic will be paid back to you. We will ask you twice a year (but sometimes up to 4 to 6 times per year) to provide a blood sample of about 15 mls (3 teaspoons). This blood sample is used to test the condition of your immune system (CD4 count), the strength of your blood (Hb), or to help us detect additional illnesses, which may be treatable. The doctor will give you the results of these blood tests at your next visit. What remains of the blood sample will be either directly used in research tests, or stored in deep-freezers, for later use in scientific tests. Some of these later tests could help to understand why some people might be more susceptible to infection or disease than others. Part of the reason could be that some people are born with a better ability to fight infection than others; such differences are called genetic. Taking this amount of blood is completely safe, and will not make you sick.

In case you haven't come to the clinic for more than three months, a field worker will visit you at home to check whether you are well, and to invite you again to the clinic.
The study and investigations have been approved by the Gambia Government/MRC Joint Ethical Committee. This study is meant to go on for many years, and no ending date is currently foreseen. Participation in this study is voluntary. If you agree to join the study, but change your mind later, and want to withdraw from the study, you may do so at any time.

The doctors and scientists of the study will have access to information about you and your condition, but apart from these people no one else will. Your files are locked away in a safe place and no outsider will have access.

You are not obliged to provide blood samples, and you are completely free to decline this every time when it is asked; this will not affect the usual care you get from the MRC clinicians. If you do not understand this information, or you have any other concerns on this study, you can discuss this with Dr. Stephen Allen.

Do you have any questions?
Do you agree to enter the study?

I, ...(clinician/counsellor) have explained to the patient the above in a language familiar to the patient.

Signature: ..... Date .....

I, ... have read / been explained the above information about the MRC's clinical cohort study and I understand the information. I agree to participate in the study. I understand that I can withdraw from the study at any time, without needing to give reasons, and without this affecting the care I receive at the MRC.

Signature / thumbprint: .....

Date.....

APPENDIX 2

CONFIDENTIAL MRC GENITO-URINARY CLINIC

SOCIAL HISTORY FORM MALE (NSH2m)

version 27th September 2000

PART I KEY INFORMATION

1	OPDNO	|___|___|___|___| / |___|___|___|___|	
2	Sex	Male = 1 Female =2	
3	Last Name		
4	First Name		
5	Date of interview (DD / MM / YYYY)	|___|___| / |___|___| / |___|___|___|___|	
6	What is your age? (years)	|___|___|	
7	Is an exact birth date available?	Yes = 1 No = 2	>>9
8	If yes, what is the exact birth date? (DD / MM / YYYY)	N.A.=88 |___|___|/|___|___|/|___|___|___|___|	
9	Do you agree to participate in this study?	Yes = 1 No = 2	>part V

If patient does not want to participate in the study, OR the patient is less than 15 years of age, stop the interview here, and go to part V.

PART II PERSONAL DETAILS

11	What is your nationality?	Gambian = 1 Senegalese = 2 Other = 3	>>13 >>13
12	Please specify your nationality.	N.A. = 88	
13	What is your ethnic group?	Mandinka=01 Wolof=04 Manjago=07 Fula=02 Serahuli=05 Aku=08 Jola=03 Serer=06 Other=09	
14	What is your religion?	Muslim = 1 Christian = 2 Other = 3	
15	What is your permanent area of residence?	.....	

16	In which Division is this address?	Greater Banjul = 01 URD = 06 WD = 02 Outside Gambia = 07 LRD = 03 Not applicable = 88 NBD = 04 Unknown = 99 MID = CRD = 05	

17	Is this address a rural or an urban area?	Rural = 1 Urban = 2 Unknown = 9 N.A. = 8	

18	What is your marital status?	Single = 1 Married = 2 Divorced = 3 Widowed = 4	>> 31
19	How many spouses have you had in your life?	N.A. = 88 |___|___|	
20	How many spouses did you divorce?	N.A. = 88 |___|___|	
21	How many spouses of yours have died?	N.A. = 88 |___|___|	
22	MARRIED MEN ONLY: How many spouses do you have at present?	N.A. = 88 |___|___|	

PART III OCCUPATION, EDUCATION, TRAVEL

31	What is your occupation?	.....	
32	Have you had any Koranic education?	Yes = 1 No = 2	>> 34
33	How many years did you complete?	N.A. = 88 |___|___|	
34	Have you had any primary education?	Yes = 1 No = 2	>> 38
35	How many years did you complete?	N.A. = 88 |___|___|	
36	Have you had any secondary education?	Yes = 1 No = 2 N.A. = 8	>> 38 >> 38
37	How many years did you complete?	N.A. = 88 |___|___|	
38	Language Skills: Can you read an English newspaper? Can you speak English?	Very well A bit Not at all 1 2 3 1 2 3 1 2 3	

	Can you write English? Can you read a French newspaper? Can you speak French? Can you write French?	1 2 3 1 2 3 1 2 3	
39	Have you travelled outside The Gambia during the last 10 years?	Yes = 1 No = 2	
40	Have you stayed outside The Gambia for longer than one week at one time in the last 10 years?	Yes = 1 No = 2 N.A. = 8	 >> 51 >> 51

OPDNO: |___|___|___|___| / |___|___|___|___|

41. Indicate for all locations outside The Gambia where you stayed for longer than one week at a time, during the last 10 years:

	Country/City	Code	Year of last stay	How long did you stay?	Code
A					
B					
C					
D					
E					
F					
G					
H					
I					
I					

PART IV HEALTH & SEXUAL HISTORY

51	Have you ever had a blood transfusion?	Yes = 1 No = 2 Unknown = 9	 >> 53 >> 53
52	When was the blood transfusion? (state year of last blood transfusion)	N.A.=8888 |___|___|___|___|	
53	Have you been circumcised?	Yes = 1 No = 2 Unknown = 9	 >> 56 >> 56
54	Where were you circumcised?	Bush = 1 Hospital/HC = 2 Other = 3 N.A. = 8	
55	How old were you when you were circumcised? (age in completed years)	N.A. = 88 |___|___|	
56	Have you ever had sexual intercourse?	Yes = 1 No = 2 Unknown = 9	 >> part V >> part V
57	How old were you when you had your first sexual intercourse? (age in years)	N.A. = 88 |___|___|	

58	How many different sexual partners have you had in your whole life?	N.A. = 88 |___|___|	
59	Have you ever seen a condom?	Yes = 1 No = 2 N.A. = 8	 >> 62 >> 62
60	Have you ever used a condom?	Yes = 1	

		No = 2 N.A. = 8 Unknown = 9	>> 62 >> 62 >> 62
61	How often did you use condoms while having sexual intercourse during the last 12 months?	Always = 1 Mostly = 2 Sometimes = 3 Rarely = 4 Never = 5 N.A. = 8 Unknown =9	
62	Have you ever had a sexually transmitted disease?	Yes = 1 No = 2 N.A. = 8 Unknown = 9	
63	Have you ever had sexual intercourse in exchange for presents or money?	Yes = 1 No = 2 Not applicable = 8 Unknown = 9	
64	Have you ever had sex with a prostitute?	Yes = 1 No = 2 Not applicable = 8 Unknown = 9	
65	Have you ever had sexual intercourse with another man?	Yes = 1 No = 2 Not applicable = 8 Unknown = 9	

PART V INTERVIEW DETAILS (do not ask these questions to patient)

Other details:.....

.....

Language of interview: .....

Name of field worker:..... Signature:..... |___|___|

APPENDIX 3

CONFIDENTIAL MRC GENITO-URINARY CLINIC

SOCIAL HISTORY FORM FEMALE (NSH2f)

version 27th September 2000

PART I KEY INFORMATION

1	OPDNO	|___|___|___|___| / |___|___|___|___|	
2	Sex	Male = 1 Female =2	
3	Last Name		
4	First Name		
5	Date of interview (DD / MM / YYYY)	|___|___| / |___|___| / |___|___|___|___|	
6	What is your age? (years)	|___|___|	
7	Is an exact birth date available?	Yes = 1 No = 2	>> 9
8	If yes, what is the exact birth date? (DD / MM / YYYY)	N.A.=88 |___|___|/|___|___|/|___|___|___|___|	
9	Do you agree to participate in this study?	Yes = 1 No = 2	>part V

If patient does not want to participate in the study, OR the patient is less than 15 years of age, stop the interview here, and go to part V.

PART II PERSONAL DETAILS

11	What is your nationality?	Gambian = 1 Senegalese = 2 Other = 3	>> 13 >> 13
12	Please specify your nationality.	N.A. = 88	
13	What is your ethnic group?	Mandinka=01 Wolof=04 Manjago=07 Fula=02 Serahuli=05 Aku=08 Jola=03 Serer=06 Other=09	
14	What is your religion?	Muslim = 1	

		Christian = 2 Other = 3	
15	What is your permanent area of residence?	....	
16	In which Division is this address?	Greater Banjul = 01 URD = 06 WD = 02 Outside Gambia = 07 LRD = 03 Not applicable = 88 NBD = 04 Unknown = 99 MID = CRD = 05	
17	Is this address a rural or an urban area?	Rural = 1 Urban = 2 Unknown = 9 N.A. = 8	

18	What is your marital status?	Single = 1 Married = 2 Divorced = 3 Widowed = 4	>> 31
19	How many spouses have you had in your life?	N.A. = 88 |___|___|	
20	How many spouses did you divorce?	N.A. = 88 |___|___|	
21	How many spouses of yours have died?	N.A. = 88 |___|___|	
23	MARRIED WOMEN ONLY: How many spouses does your husband have at present (including yourself)?	N.A. = 88 |___|___|	
24	MARRIED WOMEN ONLY: What is your rank among the wives?	N.A. = 88 |___|___|	

PART III OCCUPATION, EDUCATION, TRAVEL

31	What is your occupation?	.....	
32	Have you had any Koranic education?	Yes = 1 No = 2	>> 34
33	How many years did you complete?	N.A. = 88 |___|___|	
34	Have you had any primary education?	Yes = 1 No = 2	>> 38
35	How many years did you complete?	N.A. = 88 |___|___|	
36	Have you had any secondary education?	Yes = 1 No = 2 N.A. = 8	>> 38 >> 38

37	How many years did you complete?	N.A. = 88 |___|___|	
38	Language Skills:	Very well A bit Not at all	
	Can you read an English newspaper?	1 2 3	
	Can you speak English?	1 2 3	
	Can you write English?	1 2 3	
	Can you read a French newspaper?	1 2 3	
	Can you speak French?	1 2 3	
	Can you write French?	1 2 3	
39	Have you travelled outside The Gambia during the last 10 years?	Yes = 1 No = 2	
40	Have you stayed outside The Gambia for longer than one week at one time in the last 10 years?	Yes = 1 No = 2 N.A. = 8	>> 42 >> 42

OPDNO: |___|___|___|___|___| / |___|___|___|___|

41. Indicate for all locations outside The Gambia where you stayed for longer than one week at a time, during the last 10 years:

	Country/City	Code	**Year of last stay**	**How long did you stay?**	**Code**
A					
B					
C					
D					
E					
F					
G					
H					
I					
J					

42	MARRIED WOMEN ONLY: What is your husband's occupation?	N.A. = 88	
43	MARRIED WOMEN ONLY: Has your husband travelled or lived outside The Gambia during the last 10 years?	Yes = 1 No = 2 N.A. = 8 Unknown = 9	 >> 51 >> 51
44	MARRIED WOMEN ONLY: Has your husband stayed outside The Gambia for longer than one week at one time, during the last 10 years?	Yes = 1 No = 2 N.A. = 8 Unknown = 9	 >> 51 >> 51 >> 51

45. MARRIED WOMEN ONLY: Indicate for all locations outside The Gambia where your husband stayed for longer than one week at a time during the last 10 years:

	Country/City	**Code**	**Year of last stay**	**How long did he stay?**	**Code**
A					
B					
C					
D					
E					

PART IV HEALTH & SEXUAL HISTORY

51	Have you ever had a blood transfusion?	Yes = 1 No = 2 Unknown = 9	 >> 53 >> 53
52	When was the blood transfusion? (state year of last blood transfusion)	N.A. = 8888 |___|___|___|___|	
53	Have you been circumcised?	Yes = 1 No = 2 Unknown = 9	 >> 56 >> 56

54	Where were you circumcised?	Bush = 1 Hospital/HC = 2 Other = 3 N.A. = 8	
55	How old were you when you were circumcised? (age in completed years)	N.A. = 88 |___|___|	
56	Have you ever had sexual intercourse?	Yes = 1 No = 2 Unknown = 9	>>part V >>part V
57	How old were you when you had your first sexual intercourse? (age in years)	N.A. = 88 |___|___|	
58	How many different sexual partners have you had in your whole life?	N.A. = 88 |___|___|	
59	Have you ever seen a condom?	Yes = 1 No = 2 N.A. = 8	 >> 62 >> 62
60	Have you ever used a condom?	Yes = 1 No = 2 N.A. = 8 Unknown = 9	 >> 62 >> 62 >> 62
61	How often did you use condoms while having sexual intercourse during the last 12 months?	Always = 1 Mostly = 2 Sometimes = 3 Rarely = 4 Never = 5 N.A. = 8 Unknown =9	
62	Have you ever had a sexually transmitted disease?	Yes = 1 No = 2 N.A. = 8 Unknown = 9	
63	Have you ever had sexual intercourse in exchange for presents or money?	Yes = 1 No = 2 Not applicable = 8 Unknown = 9	

PART V INTERVIEW DETAILS (do not ask these questions to patient)

Other details:.....
Language of interview:.....

66. Is this patient a "professional"? Yes = 1 No = 2 Unknown=9 |___|

Name of field worker:..... Signature:..... |___|___|

APPENDIX 4

MRC Study SCC048 OPDNO:|___|___|___|___|/|___|___|___|___|

Project title: Human leukocyte antigen (HLA) and killer immunoglobulin-like receptor (KIR) in HIV-2 infection: a key component to HIV vaccine design and its evaluation in Africa

Participant Information sheet

AIDS is caused by a small germ, called HIV. Some people become ill within a few years of becoming infected, others may stay healthy for 0 or more years. Currently there is no cure for HIV infection. However, there are expensive drugs that can slow down the disease and these are now available in The Gambia.

The MRC has been studying HIV infections for many years. We are investigating the body's defences against this infection. We do this by using a small amount of blood in tests in the laboratory, which reveal whether the defence system is strong or weak, and over time, whether it is maintaining its strength or is losing power. Because the immune system is a complicated system, there are many different tests that can be done. New tests are developed every year. We hope that these various studies will lead to the development of a vaccine in the long run. A good vaccine would prevent infection with HIV, or prevent the development of disease, including AIDS.

We invite you to enrol in one of our studies, which is designed to research on why some people might be more susceptible to HIV-2 infection than others and why some infected individuals progress faster with the disease than others. Part of the reason could be that some people are born with a better ability to fight infection than others; such differences are called genetics. The blood sample you are about to give for your HIV screening will help us to determine your HIV status and if negative, what remains of the blood sample will be used for this study and future genetic studies that could help in understanding why some people might be more susceptible to infection or are at risk of progressing faster than others once infected. We will offer you the usual MRC clinical care when you are ill and provide you with medication as needed, both free of charge. Taking this amount of blood, which is less than teaspoon is completely safe, and will not make you sick.

The study and investigations have been approved by the Gambia Government/MRC Joint Ethical Committee. Participation in this study is voluntary. If you agree to join the study, but change your mind later, and want to withdraw from the study, you are free to do so at any time.

The doctors and scientists of the study will have access to information about you and your condition, but apart from these people no one else will. Your files are locked away in a safe place and no outsider will have access.

You are not obliged to provide the blood sample, and you are completely free to decline this every time when it is asked; this will not affect the usual care you get from the MRC clinicians. If you do not understand this information, or you have any other concerns on this study, you can discuss this with Mr. Louis Marie Yindom.

Do you have any questions? .....

Signature: .....thumbprint: .....

Date.....

APPENDIX 5

MRC Study SCC1048 OPDNO:|___|___|___|___|/ |___|___|___|___|

Project title: Human leukocyte antigen (HLA) and killer immunoglobulin-like receptor (KIR) in HIV-2 infection: a key component to HIV vaccine design and its evaluation in Africa

Consent Form

Do you agree to enter the study? .....

I, ...(clinician/counsellor) have read and explained the information as is in the "Participant Information Sheet" of this study to this participant in a language familiar to him/her.

Signature: ..... Date .....

I, ... have read/been explained the information from the "Participant Information Sheet" of the above mentioned study and I understand the information. I agree to participate in the study. I understand that I can withdraw from the study at any time, without needing to give reasons, and without this affecting the usual care I receive at the MRC.

Signature:..... thumb print:.....

Date.....

APPENDIX 6
Medical Research Council Laboratories and Projecto de Saude Bandim

Invitation to the Caio Case Control cohort

Subject Information Sheet

You have been given this sheet because you are invited to participate in the Caio Cohort. Please take time to read the following information. Please ask if there is anything that is not clear or if you would like to know more. Thank you for your time.

The Caio Field Station

The Bandim Health Project has been situated in Caio since 1989. The project provides health care linked to research and works closely with the Medical Research Council in The Gambia. A small group of fieldworkers, nurses, laboratory technicians, and doctors work at the project. They provide different services: continuous census, follow-up of children and pregnant women, accessibility to a doctor, transport for medical evacuation to Canchungo, malaria tests. In the past, we have been involved in different studies, for example Stepping Stones, Mother and Child Study, Case-Control Studies, Trachoma Study. The information we collect in these studies helps us to improve health care in general and specifically in Caio.

The Caio Case Control Cohort

We would like to invite you to the Caio Case Control cohort. This cohort is a group of about 600 adult people of which some are HIV positive and some are HIV negative. We have followed these people in our clinic and they have participated in different studies. All people in the Cohort receive a Green Card with their name, ID and photo. With this card, the person can come to visit the project and receive free medical care, medication and specialized treatment at referral hospitals if necessary. Last year and this year, a study was done in all the adults in Caio to see how many people are now infected with HIV. Unfortunately, many new people have become infected with HIV and we want to offer health care to these people.We are inviting both HIV positiveand HIV negative people. When you are part of the cohort, you might be asked to participatein future scientific studies. Participation in these studies will be completelyvoluntary. Not participatingwill not affect receiving a green card or free access to free healt care.

The reason that a fieldworker is reading this information sheet to you is because we would like you to become a member of the cohort. The project`s field workers do not know your status. You decide if you want to know your HIV result or not. If you would like to know your status, you can ask one of our conselors to counsel you abou your result.

If you want to participate, we would like you to sign a Consent Form, to know that you have understood all the information. After signing we will invite you to come visit the Project. There you will be registered and you will receive your personal Green Card. Then you will be seen by Dr. Carlos da Costa who will examine you and prescribe medication if necessary. He will ask you for a 5ml of blood donation, some of which will be used for your personal care and some for storage for future studies. The Field station doctor can ask you to donate 5ml of blood in de future for your

clinical follow-up. Agin, a portion of the follow-up samples would be stored for future research. These bood samples are very valuable and could help improve our knowledge of VIH.the aim of this knowledge is to contribute to beter care for people infected with HIV. You are free to participate or not and this will not affect your Green Card and your free access to health care.

HIV

The Human Immunodeficiency Virus (HIV) is a virus that can cause AIDS, a deadly disease for which there is presently no cure. The virus can pass to someone who is not infected through sexual intercourse with an infected person, by direct contact with blood from an infected person, and from infected mothers during childbirth and breastfeeding. People can protect themselves against HIV using condoms during sexual intercourse, by avoiding direct contact with blood of other people, or using unsterilized razor blades or needles that have been used by others.

The most common virus causing AIDS is HIV-1, but a HIV-2, a similar virus, is more common in Caio than HIV-1. HIV-2 can cause AIDS as well, but not in all people. In addition, some people are exposed to HIV-1 and/or HIV-2 but don't get infected with the virus and some people who are infected but do not develop AIDS for a very long time. No one knows why some people are naturally protected from disease or why HIV-1 and HIV-2 are so different.

The Guinean Government is offering drug treatment for people that have an advanced stage of HIV disease. These drugs cannot completely cure AIDS but can offer a productive and an almost normal life. These drugs are now also available in Caio and people are being treated with them by the Project's physician.

Contact for further details: Tim Vincent, Carlos da Costa, Carla van Tienen, Aleksandra Leligdowicz, Thushan de Silva

Medical Research Council Laboratories and Projecto de Saude Bandim

Caio Case Control cohort

Consent Form

I have read the Information Sheet/The Information Sheet has been read to me.

I have understood the Information Sheet and I have had an opportunity to ask questions and have them answered.

I understand what participation in the Caio Case Control cohort means for me.

I understand that both HIV positive and HIV negative people are members of the Caio Cohort.

I understand that I receive a Green Card, with which I have access to free medical care and medication if I get sick.

I understand that I have certain rights as a member of the Caio Cohort, as stated in the "Caio Cohort: Rights" (see Annex)

I give consent to participate in the Caio Cohort

Signature or index finger print of participant: _____

This form has been read by/I have read the above to:

_____

(name of participant) in the language that he/she understands.

I believe that he/she understood what I have explained and that he/she has freely agreed to take part in the study.

Name of counsellor:_____
Signature of counsellor:_____

Name of witness: _____
Signature of witness: _____

Date: |__|__| / |__|__| / |__|__|__|__|

For further information contact:
Carlos da Costa, Tim Vincent, Sarah Rowland-Jones, Carla Van Tienen, Aleksandra Leligdowicz, Thushan de Silva
Address:
Prof Sarah Rowland-Jones
MRC Laboratories, Fajara
Atlantic Road, PO Box 273
The Gambia, West Africa
Tel: +220 449 6188 (the Gambia)

APPENDIX 7

MRC Study SCC1048

Project title: Human leukocyte antigen (HLA) and killer immunoglobulin-like receptor (KIR) in HIV-2 infection: a key component to HIV vaccine design and its evaluation in Africa

SPECIMEN REQUEST FORM

Hospital number: 

Last Name: ----------

First Name: ----------

Sex: ---------- Age: ----------Years

Date of sample collection: ----------

APPENDIX 8

SOPL1 - Salting-out technique

DAY ONE

Sample recording
Record the sample details (out patient department (OPD) number and the lab number) on the record sheet
Label a 50 ml falcon tube for each sample

LYSING OF RBCs

Carefully transfer the sample into the labelled falcon tube
Vortex vigorously (60-90 seconds)
Add TE 20-5 pH 8.0 buffer up to the 40 ml mark
Mix by inverting the tube several times
Allow to stand on ice for 15 minutes
Centrifuge at 3500 rpm for 15 minutes
Carefully remove the supernatant (with 25 ml pipette) without disturbing the white cell pellet
Vortex vigorously (60-90 seconds)
Add more TE 20-5 pH 8.0 buffer up to 40 ml mark
Vortex to mix (20 seconds)
Centrifuge at 3500 rpm for 15 minutes
Repeat washing (from step 10) until the white cell pellet is clean (appearing pinkish or white). NB: if the pellet is still not clean after 5 wash, add 20 ml sterile distilled water + 20 ml TE 20-5 pH 8.0 for an additional wash.

LYSING OF WBCs

After the final wash, vortex vigorously (90-120 seconds) or until no clump is seen in the tube
Add 2 ml of TE 20-5 pH 8.0 buffer
Add 100 µl of 20% SDS (final concentration = 1%)
Mix gently by inverting the tube several times

PROTEIN DIGESTION

Remove the proteinase-K (10 mg/ml solution) from the freezer only when ready to use and allow to thaw (preferably on ice)
Add 20 µl of proteinase K (10 mg/ml solution)
Return the proteinase-K immediately into the freezer
Gently invert the tube several times to mix
Place in a shaking water bath at 42°C
Incubate overnight and/or until no pellet is seen in the tube

DAY TWO

Remove samples from the water bath and allow to cool to room temperature
Add 2.5 ml of 7.5 M ammonium acetate solution

Mix gently by inverting the tube several times and place on ice.
Add 10 ml of ice cold absolute ethanol into each tube
Invert the tube several times to precipitate the DNA-ammonium salt complex out of the solution. A whitish thread will be seen floating in the solution and immediately settles to the bottom of the tube upon standing
If the yield is not satisfactory, keep the tubes in –20°C freezer for 1-2 hours (optional step)
Centrifuge at 5000 rpm for 30 minutes
Remove the supernatant without disturbing the DNA pellet at the bottom of the tube
Drain the remaining solution from the pellet by inverting the tube over a dry absorbent tissue briefly
Add 4 ml of sterile TE 20-5-NaCl 0.2M pH 8.0
Vortex briefly (5-10 seconds)
Re-dissolve the DNA in a shaking water bath at 42°C until fully dissolved
Remove from the water bath and allow to cool on ice
Re-precipitate the DNA in 8 ml of ice cold ethanol (the cooler the ethanol the better)
Centrifuge at 5000 rpm for 30 minutes
Remove the supernatant carefully without disturbing the pellet and gently invert the tube over a dry absorbent tissue while watching your pellet making sure it doesn't float out of the container during the draining process
Allow to air dry until no fluid is seen by the sides of the tube, then turn tubes upright for complete evaporation of ethanol.
Add 150 µl of TE 20-1 buffer in each tube and adjust the volume for bigger pellet
Incubate in a water bath at 58°C overnight to re-dissolve. NB: do not shake.
Label a 1.5 ml eppendorf tube as follows: by the side – study, lab number and date; on the lid – lab number.
Remove tubes with DNA from the water bath, spin briefly at 1000 rpm and transfer the DNA into labelled Eppendorf vials.
Quantify the DNA using the Nanodrop instrument and store at +4 °C for immediate use or -20 °C for future use

APPENDIX 9

List of KIR primers (SCC1048)

PRIMER Name	Primer Sequence	Number of bases	Synthesis scale	Purification Method
K1	5'- gTT ggT CAg ATg TCA TgT TTg AA	23	0.2 µM	HPLC & desalted
K2	5'- ggT CCC TgC CAg gTC TTg Cg	20	0.2 µM	HPLC & desalted
K3	5'- Tgg ACC AAg AgT CTg CAg gA	20	0.2 µM	HPLC & desalted
K4	5'- TgT TgT CTC CCT AgA AgA Cg	20	0.2 µM	HPLC & desalted
K5	5'- CTg gCC CAC CCA ggT Cg	17	0.2 µM	HPLC & desalted
K6	5'- ggA CCg ATg gAg AAg TTg gCT	21	0.2 µM	HPLC & desalted
K7	5'- gAg ggg gAg gCC CAT gAA T	19	0.2 µM	HPLC & desalted
K8	5'- TCg AgT TTg ACC ACT CgT AT	20	0.2 µM	HPLC & desalted
K9	5'- CTT CAT CgC Tgg TgC Tg	17	0.2 µM	HPLC & desalted
K10	5'- Agg CTC TTg gTC CAT TAC AA	20	0.2 µM	HPLC & desalted
K11	5'- TCC TTC ATC gCT ggT gCT g	19	0.2 µM	HPLC & desalted
K12	5'- ggC Agg AgA CAA CTT Tgg ATC A	22	0.2 µM	HPLC & desalted
K13	5'- CAg gAC AAg CCC TTC TgC	18	0.2 µM	HPLC & desalted
K14	5'- CTg ggT gCC gAC CAC T	16	0.2 µM	HPLC & desalted
K15	5'- ACC TTC gCT TAC AgC CCg	18	0.2 µM	HPLC & desalted
K16	5'- CCT CAC CTg TgA CAg AAA CAg	21	0.2 µM	HPLC & desalted
K17	5'- TTC TgC ACA gAg Agg ggA AgT A	22	0.2 µM	HPLC & desalted
K18	5'- ggg TCA CTg ggA gCT gAC AA	20	0.2 µM	HPLC & desalted
K19	5'- Cgg gCC CCA Cgg TTT	15	0.2 µM	HPLC & desalted
K20	5'- ggT CAC TCg AgT TTg ACC ACT CA	23	0.2 µM	HPLC & desalted
K21	5'- Tgg CCC ACC CAg gTC g	16	0.2 µM	HPLC & desalted
K22	5'- TgA AAA CTg ATA ggg ggA gTg Agg	24	0.2 µM	HPLC & desalted
K23	5'- CTA TgA CAT gTA CCA TCT ATC CAC	24	0.2 µM	HPLC & desalted
K24	5'- AAg CAg Tgg gTC ACT TgA C	19	0.2 µM	HPLC & desalted
K25	5'- CTg gCC CTC CCA ggT CA	17	0.2 µM	HPLC & desalted
K26	5'- TCT gTA ggT TCC TgC AAg gAC Ag	23	0.2 µM	HPLC & desalted
K27	5'- gTT CAg gCA ggA gAg AAT	18	0.2 µM	HPLC & desalted
K28	5'- gTT TgA CCA CTC gTA ggg AgC	21	0.2 µM	HPLC & desalted
K29	5'- TgA Tgg ggT CTC CAA ggg	18	0.2 µM	HPLC & desalted
K30	5'- TCC AgA ggg TCA CTg ggC	18	0.2 µM	HPLC & desalted
K31	5'- CTT CTC CAT CAg TCg CAT gAA	21	0.2 µM	HPLC & desalted
K32	5'- CTT CTC CAT CAg TCg CAT gAg	21	0.2 µM	HPLC & desalted
K33	5'- AgA ggg TCA CTg ggA gCT gAC	21	0.2 µM	HPLC & desalted
K34	5'- CgC TgT ggT gCC TCg A	16	0.2 µM	HPLC & desalted
K35	5'- ggT gTg AAC CCC gAC ATg	18	0.2 µM	HPLC & desalted
K36	5'- CCC Tgg TgA AAT CAg gAg AgA g	22	0.2 µM	HPLC & desalted
K37	5'- TgT Agg TCC CTg CAA ggg CAA	21	0.2 µM	HPLC & desalted
K38	5'- CAA ACC CTT CCT gTC TgC CC	20	0.2 µM	HPLC & desalted
K39	5'- gTg CCg ACC ACC CAg TgA	18	0.2 µM	HPLC & desalted
K40	5'- CCC ATg AAC gTA ggC TCC g	19	0.2 µM	HPLC & desalted
K41	5'- CAC ACg CAg ggC Agg g	16	0.2 µM	HPLC & desalted
K42	5'- AgC CTg CAg ggA ACA gAA g	19	0.2 µM	HPLC & desalted
K43	5'- gCC TgA CTg Tgg TgC TCg	18	0.2 µM	HPLC & desalted
K44	5'- CCT ggT gAA ATC Agg AgA gAg	21	0.2 µM	HPLC & desalted
K45	5'- gTC CCT gCA Agg gCA C	16	0.2 µM	HPLC & desalted

K50	5'- gCg CTg Tgg TgC CTC g	16	0.2 µM	HPLC & desalted
K51	5'- gAC CAC TCA ATg ggg gAg C	19	0.2 µM	HPLC & desalted
K52	5'- TgC AgC TCC Agg AgC TCA	18	0.2 µM	HPLC & desalted
K53	5'- ggg TCT gAC CAC TCA TAg ggT	21	0.2 µM	HPLC & desalted
K54	5'- gTC TgC CTg gCC CAg CT	17	0.2 µM	HPLC & desalted
K55	5'- gTg TgA ACC CCg ACA TCT gTA C	22	0.2 µM	HPLC & desalted
K56	5'- CCA TCg gTC CCA TgA Tgg	18	0.2 µM	HPLC & desalted
K57	5'- CAC Tgg gAg CTg ACA ACT gAT g	22	0.2 µM	HPLC & desalted
K58	5'- ACA gAg Agg ggA CgT TTA ACC	21	0.2 µM	HPLC & desalted
K59	5'- ATg TCC AgA ggg TCA CTg gg	20	0.2 µM	HPLC & desalted
K60	5'- gTC Agg ACA AgC CCT TCC TC	20	0.2 µM	HPLC & desalted
K61	5'- gAg TgT ggg TgT gAA CTg CA	20	0.2 µM	HPLC & desalted
K62	5'- TTC TgC ACA gAg Agg ggA TCA	21	0.2 µM	HPLC & desalted
K63	5'- gAg CCg ACA ACT CAT Agg gTA	21	0.2 µM	HPLC & desalted
C1	5'- TgC CAA gTg gAg CAC CCA A	19	0.2 µM	HPLC & desalted
C2	5'- gCA TCT TgC TCT gTg CAg AT	20	0.2 µM	HPLC & desalted

APPENDIX 10

Primer mix worksheet for KIR typing by PCR-SSP (SCC1048)

Lane	Primer mix	F. primer	Vol(µl)	R. primer	Vol(µl)	C1+C2	Water
1	KM1	K1	50	K2	50	50	850
2	KM2	K3	50	K4	50	50	850
3	KM3	K5	50	K6	50	50	850
4	KM4	K7	50	K8	50	50	850
5	KM5	K9	50	K10	50	50	850
6	KM6	K11	50	K12	50	50	850
7	KM7	K13	50	K14	50	50	850
8	KM8	K15	50	K16	50	50	850
9	KM9	K17	50	K18	50	50	850
10	KM10	K19	50	K20	50	50	850
11	KM11	K21	50	K22	50	50	850
12	KM12	K23	50	K24	50	50	850
13	KM13	K25	50	K26	50	50	850
14	KM14	K27	50	K28	50	50	850
15	KM15	K29	50	K30	50	50	850
16	KM16	K58	50	K59	50	50	850
17	KM17	K34	50	K35	50	50	850
18	KM18	K36	50	K37	50	50	850
19	KM19	K38	50	K39	50	50	850
20	KM20	K40	50	K41	50	50	850
21	KM21	K42	50	K43	50	50	850
22	KM22	K44	50	K45	50	50	850
23	KM23	K60	50	K61	50	50	850
24	KM24	K62	50	K63	50	50	850
25	KM25	K50	50	K51	50	50	850
26	KM26	K52	50	K53	50	50	850
27	KM27	K54	50	K55	50	50	850
28	KM28	K56	50	K57	50	50	850
29	KM29	K31	50	K32+K33	100	50	800
30	C	C1	50	C2	50		900

APPENDIX 11

SOPL2: KIR Typing protocol by PCR-SSP (SCC1048)

Master mix

	µl/plate(12 samples)
D/water	1370.33
10X buffer	214.50
$MgCl_2$ (50mM)	64.35
dNTPs (25mM)	17.16
Taq polymerase	10.725

Add 129 µl of this master mix into an 0.5 ml eppendorf tube containing 3 µl of DNA (50 ng/µl)
Add primer mixes into 30 wells of the thermowell plate using a multichannel pipette

Primer mix 1 µl/well

Dispense 4 µl of the mega mix into wells containing the primer mix
Total volume should be 5 µl.
Seal the plate, centrifuge briefly (1000 rpm for 1 minutes)
Load the plate into the PTC-200 thermal cycler (MJ Research, USA)
Run the program labeled "KIR"

Amplification condition

1 cycle	94 °C	3:00				
5 cycles	94 °C	0:15	65 °C	0:15	72 °C	0:30
21 cycles	94 °C	0:15	60 °C	0:15	72 °C	0:30
5 cycles	94 °C	0:15	55 °C	1:00	72 °C	2:00
1 cycle	72 °C	7:00				

Prepare a 3% agarose gel as bellow and allow to solidify

Agarose gel preparation

Agarose powder	12 g
0.5X TBE buffer	400 ml

1. Mix and heat in a microwave at full power for 4.5 minutes
2. Allow to cool rapidly in water
3. Add 8 µl of Ethidium bromide (10gm/ml)
4. Mix and pour onto a casting tray with sealed edges
5. Put the combs in place and allow to set.

Gel electrophoresis

Add 2.5 µl of loading buffer into each well containing the PCR products and load 7 µl of the mixture into the corresponding well of the agarose gel. Connect the gel tank to the power pack and run at 220 V for 30 minutes.

Gel Photography

Cut the gel to size and place on the UV light box.
Position the Kodak Digital Gel Documentation Camera in place and take the gel picture.
Save the gel picture in the designated location in the computer.
Copy and paste each sample picture onto the word document template for subsequent interpretation.

Gel Interpretation

Adjust the gel picture to match the lanes on the template.
Score one (1) if the specific band is present and zero (0) if it is absent. **Note:** The specific band is the second band after the control band and any band in lane 6 should be the specific band (no control introduced here).
Save the word document with the corresponding sample IDs.

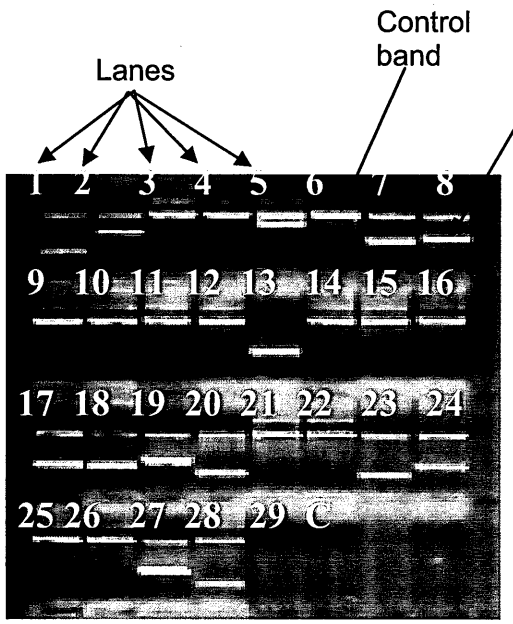
Data entry

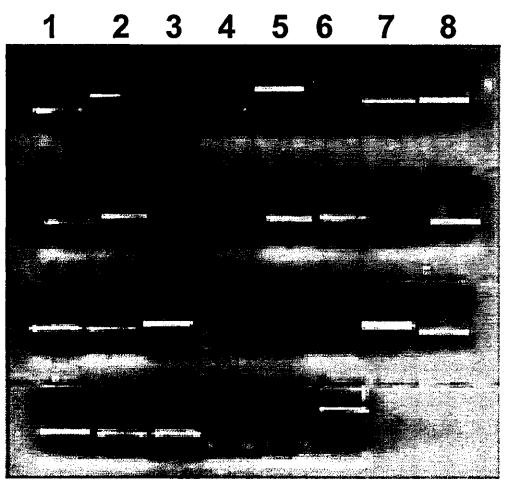
Enter the results into Microsoft Access data base designed for this project and pass the gels to the data office for second entry.

APPENDIX 12

Sample of KIR gel interpretation sheet

Cohort	FAJARA
Study	EDCTP
ID Number	**XXXX**
DNA Number	**YYYYY**
Date	Xx/xx/2006





Gene	Result	Lane	Size
2DL1	**1**	*1*	*146*
	1	*2*	*330*
2DL2	**0**	*3*	*173*
	0	*4*	*151*
2DL3	**1**	*5*	*550*
	1	*6*	*800*
2DL4	**1**	*7*	*254*
	1	*8*	*288*
2DS2	**0**	*9*	*175*
	0	*10*	*240*
2DS3	**0**	*11*	*242*
	0	*12*	*190*
2DS4	**1**	*13*	*204*
	0	*14*	*197/219*
2DS5	**0**	*15*	*126*
	0	*16*	*178*
3DL1	**1**	*17*	*191*
	1	*18*	*186*
3DL2	**1**	*19*	*211*
	1	*20*	*130*
3DS1	**0**	*21*	*300*
	0	*22*	*180*
3DL3	**1**	*23*	*112*
	1	*24*	*190*
2DL5	**0**	*25*	*214*
	0	*26*	*191*
2DP1	**1**	*27*	*205*
	1	*28*	*89*
2DS1	**0**	*29*	*102*
C	**0**	*C*	*796*

APPENDIX 13

List of primers for HLA typing by sequencing

PCR primers:

HLA-A: ccc aga cgc cga gga tgr csg	(Fwd)
HLA-A: gca ggg cgg aac ctc aga gtc act ctc t	(Rev)
HLA-B: tcc cag ttc taa agt ccc cac g	(Fwd)
HLA-B: tcc att caa ggg agg gcg ac	(Rev)
HLA-C: agc gag gkg ccc gcc cgg cga	(Fwd)
HLA-C: gga gat ggg gaa ggc tcc cca ct	(Rev)

Sequencing primers:

HLA-A: gga tct cgg acc cgg ag	exon2 Rev
HLA-A: gtt tag gcc aaa aat ycc cc	exon3 Fwd
HLA-B: gga tct cgg acc yrg ag	exon2 Rev
HLA-B: acc cgg ttt cat ttt cag ttg a	exon3 Fwd
HLA-C: cga ccc ggg ccg tc	exon2 Rev
HLA-C: gcc aaa atc ccc gcg ggt tgg t	exon3 Fwd

APPENDIX 14

HLA sequencing Protocol

HLA class I buffer preparation (10X buffer)

Trizma base	40.568 g
Ammonium Sulphate	10.96 g
Tween 20	5 ml

Dissolve the Trizma base in 400 ml of sterile distilled water and adjust the pH to 8.8 with concentrated HCl
Add the ammonium sulphate into the solution and mix to dissolve
Pass the mixture through a 22 µm filter into a clean sterile bottle
Add Tween 20 and make up the volume to 500 ml with sterile d/water

HLA PCR

Master mix	µl/rxn
Class I buffer	2.0
dNTPs (10 uM)	0.4
$MgCl_2$ (25 mM)	1.6
Primer (F)	0.4
Primer (R)	0.4
100X BSA (optional)	0.2
DMSO (optional)	1.0
Taq polymerase	0.2
Total	6.2
DNA (150 ng) + D/w	13.8
Final volume/rxn	20

Primers

A1: ccc aga cgc cga gga tgr csg	(F)
A4: gca ggg cgg aac ctc aga gtc act ctc t	(R)
B1: tcc cag ttc taa agt ccc cac g	(F)
B4: tcc att caa ggg agg gcg ac	(R)
C1: agc gag gkg ccc gcc cgg cga	(F)
C4: gga gat ggg gaa ggc tcc cca ct	(R)

Amplification conditions

HLA-A locus (HLAASEQ)

1 cycle	96 °C	2:00				
30 cycles	96 °C	0:25	60 °C	0:25	72 °C	1:00
1 cycle	72 °C	7:00				
1 cycle	4 °C	∞				

HLA-B locus (HLABSEQ)

1 cycle	96 °C	2:00				
40 cycles	96 °C	0:15	62 °C	0:15	72 °C	1:00
1 cycle	72 °C	7:00				

1 cycle	4 °C	∞

HLA-C locus (HLACSEQ)

1 cycle	96 °C	2:00			
30 cycles	96 °C	0:25	70 °C	0:25	72 °C 1:00
1 cycle	72 °C	7:00			
1 cycle	4 °C	∞			

EXO-SAP cleaning

	Per reaction
Exo I	0.5 µl
SAP	0.5 µl
Water	4.0 µl
PCR product	5.0 µl (10.0 if band is weak)

Add exonuclease I, SAP and sterile αQH_2O in a sterile vial and mix. Make enough for all the samples plus 10% to care for pipetting errors
Dispense 5 µl of the mixture into wells of a 96-well plate
Add the PCR products, mix by pipetting up-and-down, seal with and adhesive film and centrifuge briefly
Place in the thermal cycler and run the EXOSAP programme

Cycling conditions (EXOSAP programme)

37 °C	1hr 15 mins
80 °C	20 mins

Alternative PCR cleaning protocol (AMPURE)

You will need:
Freshly prepared 70% ethanol
Sterile water or TE
Tips
Multichannel pipette

Procedure

Gently mix the Ampure reagent by inverting several times at room temperature
Add 18 µl of Ampure into 10 µl of PCR products
Pipette mix 10 times or vortex for 30 seconds
Allow to stand at room temperature for 5 minutes
Place on the plate magnet for up to 5 minutes
Completely aspirate the clear solution from the bottom of the well and discard while the plate is still on the magnet

Add 200 µl of 70% ethanol (**freshly prepared**) into each well and wait for 30 seconds (plate should still be on the magnet)
Completely aspirate the alcohol from the bottom of the well and discard while the plate is still on the magnet
Repeat 7-8
Place the plate on the bench to completely air dry (15-20 minutes)
Add 40 µl of sterile water or TE to elute and pipette mix 10 times or seal and vortex for 30 seconds. (if smaller volume is used, make sure you vortex well to contact the ring of beads)
Place the plate on the magnet for 5 minutes and transfer 35 µl into a new plate.

Sequencing reaction

	Vol/rxn (µl)
D/water	5.36
5X buffer	2.07
BigDye v1.1 (or 3.1)	0.25
Primer (10 uM)	0.32
Total	8.0
PCR products	2.0

Seq Primers

A2: gga tct cgg acc cgg ag	ex2
A3: gtt tag gcc aaa aat ycc cc	ex3
B2: gga tct cgg acc yrg ag	ex2
B3: acc cgg ttt cat ttt cag ttg a	ex3
C2: cga ccc ggg ccg tc	ex2
C3: gcc aaa atc ccc gcg ggt tgg t	ex3

Add 2.0 of the PCR products to their respective wells
Add 8.0 µl of the cocktail into each well
Mix briefly and centrifuge briefly
Load onto the thermal cycler and run the SEQ programme

Sequencing cycling conditions (SEQ)

96 °C	1 min	
96 °C	10 sec	} 30 cycles
50 °C	5 sec	
60 °C	4 mins	
4 °C	∞	

Making the G-50 microtiter plate

Add the G-50 powder onto the black plate (MultiScreen Column Loader), spray over to fill in the number of wells needed using the MultiScreen Loader Scraper
Put back the excess powder into its container
Place the MultiScreen plate (MAHVN4510) in position
Turn the black plate up-side-down on the MultiScreen plate
Tap gently on the black plate to transfer G-50 powder into the corresponding wells of the 96-well plate
Remove the black plate and add 300 µl of water into each well containing the G-50 powder
Allow to sit at room temp for at least 2 hours. This can be kept in the fridge for several days (seal to avoid evaporation)

Centrifugation

Place the G-50 plate on a 96-well U or V bottom plate and centrifuge at 2300 rpm for 5 minutes to remove water from the G-50. Use frame alignment between the two plates if needed and make sure you balance the plates in the centrifuge
Wash the column twice to remove particles by adding 150 µl of water and centrifuge as above
Add 5 µl of water into each well containing the PCR product
Label a new sequencing plate with the plate ID (this label should match with Excel spreadsheet to be submitted electronically)
Place the G-50 plate on the labelled sequencing plate
Gently add 11 µl of product into the G-50 plate (avoid touching the G-50 with the pipette tips), cover the plate and hold the two plates together by tapping their sides with brown adhesive tape
Centrifuge at 2300 rpm for 5 minutes (this process filters out the products into the corresponding well of the sequencing plate while holding back excess BDT and primers)

Drying of samples for sequencing

Place the sequencing plate containing the products in the thermal cycler
Cover with a light wipe to avoid anything dropping accidentally into the wells
DO NOT close the lid
Run the programme DRY, which is set at 90°C for 20 minutes

Denaturation

Add 10 µl of HiDi formamide into each well containing the sequencing product
Seal with adhesive film, vortex briefly, place in the thermal cycler and run the programme DENATURE, which is set at 90°C for 2 minutes
Remove, centrifuge briefly and submit for loading into the 3130xl DNA Analyzer
Make a new excel file with the same name as on the sequencing plate by filling the "User's template" in \\Falcon\SequenceAnalyzer
Fill in the sample IDs to match with their corresponding positions on the 96-well sequencing plate and submit for loading into the 3130xl instrument.

Alternative cleaning "Cleanseq" procedure

You will need:
Freshly prepared 85% ethanol
Sterile water or TE
Tips
Multichannel pipette

Procedure

Gently mix the CleanSeq reagent by inverting several times at room temperature
Add 10 µl of CleanSeq into 10 µl of sequencing products
Add 42 µl of 85% ethanol into each well [or V = 2.077 x (sample vol + 10) if different sample volume is used] NB: use **freshly prepared** 85% ethanol

Pipette mix 7 times
Place on the magnet for 5 minutes
Completely aspirate the clear solution from the bottom of the well and discard while the plate is still on the magnet
Add 100 µl of 85% ethanol into each well and wait for 30 seconds (plate should still be on the magnet)
Completely aspirate the alcohol from the bottom of the well and discard while the plate is still on the magnet
Repeat 7-8
Allow to air dry for 10 minutes at room temperature
Add 40 µl of sterile water and allow tostand on the magnet for 5 minutes. (if smaller volume is used, make sure you vortex well to contact the ring of beads)
Transfer 30-35 µl into the sequencing plate, seal, label and submit for analysis on the 3130xl instrument as usual

APPENDIX 15

(Manuscript submitted to the Journal of Virology for consideration and publication as a cutting edge paper)

The influence of *HLA* class I and *HLA-KIR* compound genotypes on HIV-2 infection and markers of disease progression in a Manjako community in West Africa[1]

Louis-Marie Yindom[*], Aleksandra Leligdowicz[*,†], Maureen P. Martin[‡], Xiaojiang Gao[‡], Ying Qi[‡], Syed M.A. Zaman[*,§], Maarten Schim van der Loeff*[,¶], Carla van Tienen[*], Assan Jaye[*], Akum Aveika[*], Archibald Worwui[*], Mathurin Diatta[*], Tim Vincent[*], Hilton C. Whittle[*], Sarah L. Rowland-Jones*[,†], Robert Walton[*,||], Mary Carrington[‡]

* Medical Research Council (UK), The Gambia

† MRC Human Immunology Unit, Oxford, UK

‡ Laboratory of Experimental Immunology, SAIC-Frederick Inc., NCI-Frederick, Frederick, MD 21702

§ Centre for Infections, Health Protection Agency, Colindale, London, NW9 5EQ, UK

¶ GGD Amsterdam & Centre for Infection and Immunity Amsterdam (CINIMA), CE 1000 Amsterdam, The Netherlands

|| Centre for Health Sciences, Barts and the London Medical School, London E1 2AT

Running title: *HLA* and *KIR* in HIV-2 infection

Address correspondence to Dr. Mary Carrington, Laboratory of Experimental Immunology, NCI-Frederick, Frederick, Maryland 21702. E-mail address: carringm@mail.nih.gov

Abstract

Overall, the time to AIDS after HIV-2 infection is longer than with HIV-1 and many individuals infected with HIV-2 virus remain healthy throughout their lives. Multiple *HLA* and *KIR* gene products have been implicated in the control of HIV-1 but the effect of variation at these loci on HIV-2 disease is unknown. Here we show for the first time that *HLA-B*1503* associates significantly with poor prognosis after HIV-2 infection and *HLA-B*0801* associates with susceptibility to infection. Interestingly, previous data indicate that *HLA-B*1503* associates with low viral loads in HIV-1 clade B-infection, but has no significant effect on viral load in clade C infection. In general, alleles strongly associated with HIV-1 disease showed no effect in HIV-2 disease. These data emphasize the unique nature of the effects of *HLA* and *HLA/KIR* combinations on HIV-2 immune responses relative to HIV-1, which could be related to their distinct clinical course.

Word count: 147

Introduction

Since its first report in 1986, HIV-2 remains largely confined to West Africa (1). It shares between 30-60% nucleotide and amino acid homology with HIV-1 but differs greatly in pathogenicity and transmissibility (2). Studies on HIV-2 patients across West Africa have shown that some people remain uninfected despite repeated exposure (3) and a substantial proportion of infected people remain relatively healthy for a very long time with low plasma viral load and normal $CD4^+$ T cell counts, a characteristic of long-term non-progressors (LTNPs) infected with HIV-1 (4). This is perhaps a reflection of an effective immune response mounted against the virus including a vigorous $CD8^+$ T-cell response (5), maintenance of HIV-specific $CD4^+$ T-cell function (6), and the presence of a strong neutralizing antibody response in many subjects (7), features that are highly desirable for a successful HIV-1 vaccine. Thus, HIV-2 disease course provides a natural model for investigating mechanisms that control HIV infection and a better understanding of these mechanisms might inform new strategies for HIV prevention and treatment.

HLA class I molecules present antigenic epitopes to cytotoxic T cells and are central to the acquired immune response. A number of associations between *HLA* class I alleles and HIV disease outcomes have been reported (8), the most consistent being *B*57* and *B*27*, which show strong protection across studies, and certain subtypes of *B*35*, which associate with more rapid progression (9). While several mother-infant studies have implicated sharing of certain *HLA* alleles in transmission of the virus from mother to infant (10, 11), there is no convincing data that particular *HLA* class I alleles protect against HIV infection in general.

HLA class I allotypes also serve as ligands for killer cell immunoglobulin-like receptors (KIR), which modulate natural killer (NK) cell function. KIRs are structurally similar to one another and can be divided into activating and inhibitory receptors. NK cells are key components of the innate immune system and constantly survey host cell surfaces for

appropriate levels of HLA class I molecules through a network of NK cell receptors, including KIR (12). Upon engagement with their ligand, inhibitory KIR suppress NK cell activity, but if the ligand is missing or has been down regulated on target cells, the threshold for NK cell activation is lowered thus allowing for activation signals to dominate (13).

HLA and *KIR* genes are found on chromosomes 6 and 19, respectively, so they segregate independently. As such, the genes/alleles for the corresponding receptor/ligand pair must be present to confer functionality, whereas presence of one without the other results in a null phenotype. A number of *HLA* and *KIR* gene products either individually or collectively has been implicated in the control of HIV-1 (14), but nothing is known of their role in HIV-2.

Epidemiological data from Caio and other cohorts in West Africa (15, 16) indicate that HIV-2 infection in a substantial proportion of infected individuals is compatible with normal survival and without signs of immunodeficiency, suggesting distinct viral pathogenic mechanisms and protective host factors against HIV-2 relative to HIV-1. Here we determined the *HLA* class I and *KIR* gene profiles of the Caio population (>95% Manjako) and investigated their effects on susceptibility to HIV-2 infection and disease progression.

Materials and Methods

Study populations

The Medical Research Council (UK) unit in The Gambia has followed up for two decades an open cohort of HIV-2, HIV-1, and dually infected adults, and a similar number of age- and sex-matched uninfected controls in the rural village of Caio, Guinea-Bissau, a relatively isolated community of the Manjako tribe, an animist tribe in which marriages usually occur within the community. The Caio cohort was established in 1989 and a demographic census and sero-survey of the area was conducted in 1989-90, in which 7.9% of

the adult population was shown to be infected with HIV-2 (17). This was followed by a case-control study in 1991 (18), a re-survey in 1996-8, and re-examination of the cases and controls in 1996 and 2003. Our study used the samples and data collected in the 2003 study. The original case-control cohort was carefully recruited to match cases and controls as far as possible for demographic factors such as age, sex, and location in the village.

A total of 513 samples from HIV-2 infected and uninfected adults, predominantly of the Manjako tribe, were typed for *HLA* class I (*HLA-A, -B,* and *-C*) and *KIR* genes. **Table I** shows the characteristics of all the study participants. There were more females than males in each category and females were also slightly older: this is consistent with previous epidemiological data from Africa, in which older women appear to be more susceptible to HIV infection in general and particularly to HIV-2 (19, 20). The differences in age between categories were not statistically significant. Dually infected (HIV-1&2) individuals were excluded from all other analyses because of the potential for HIV-1 to confound our analyses. Precise dates of seroconversion are unknown for most of the infected participants in our study. Although it is difficult to estimate HIV exposure in the control group accurately, previous anthropological studies in Caio strongly suggest that the unique traditional way of life of the Manjakos in Caio lead to most adults having been exposed to sexually transmitted diseases, including HIV, from multiple partners during their four year initiation period into adulthood (21, 22).

The study was approved by the Gambian Government/MRC joint Ethical Committee, the Open University's Life and Biomolecular Sciences Management Group, UK, and the Office of Human Subjects Research, National Institutes of Health, USA.

KIR and HLA class I genotyping

Genomic DNA was genotyped for presence or absence of the following *KIR* genes: *2DL1, 2DL2, 2DL3, 2DL4, 2DL5, 2DS1, 2DS2, 2DS3, 2DS4, 2DS5, 3DL1, 3DL2, 3DL3,*

3DS1, 2DP1, and *3DP1*. Genotyping was performed using PCR amplification with two pairs of primers specific for each locus (PCR-SSP) as previously described (23). *HLA* class I genotyping was done by sequencing. Briefly, locus-specific primers flanking exons 2 and 3 were used to amplify *HLA-A*, *-B*, and *-C* loci. Purified PCR products were sequenced in both directions using exon-specific primers and BigDye Terminator version 3.1 Cycle Sequencing Kit (Appliedbiosystems, Foster City, USA) in an ABI-3130XL DNA Analyzer (Appliedbiosystems, Foster City, USA). Sequence traces were analysed using "Assign 400" software (Conexio genomics, Western Australia).

Statistical analysis

Allele and genotype frequencies were calculated using SPSS v16.0.1 (SPSS Inc, Chicago, USA) and STATA v9.2 (Stata Corporation, Texas, USA). Haplotype reconstruction, test for Hardy-Weinberg equilibrium, and haplotype frequency estimations were performed using Arlequin version 3.11. χ^2 test was used to examine differences in allele and genotype frequencies between Caio and other West African populations. All other statistical analyses were performed using SAS 9.1 (SAS Institute). PROC FREQ was used to compute frequencies of individual variables. PROC LOGISTIC was used for categorical analyses to obtain odds ratios and 95% confidence intervals. PROC GLM was used for analysing continuous variables. Both logistic and linear regression models were adjusted for age and gender. We did not adjust for ethnicity because more than 95% of the subjects were from the same ethnic group (Manjako).

Results

HLA class I allele frequencies in Caio show significant differences compared to other West African populations

We first compared the frequencies of each *HLA* class I allele with data from neighbouring West African populations in Senegal (24), Mali (25), and The Gambia (LM Yindom, manuscript in preparation) (**Fig. 1**). Frequencies of *HLA-A, -B,* and *-C* alleles in Caio (n = 513) differed significantly for some alleles from those in neighbouring countries, where *HLA-A*3303* was more common and *-A*3001* less common (**Fig. 1a**). *HLA-B*0801, -B*1302, -B*1503, -B*1510,* were also relatively common in Caio as was *-B*5801*, which has previously been shown to be protective against HIV-1 (26) (**Fig. 1b**). *HLA-Cw*1601* and *-Cw*17* allele frequencies were lower, whereas *HLA-Cw*0701, -Cw*0304* and *-Cw*0804* were higher in Caio than elsewhere (**Fig. 1c**).

KIR genotypes in Caio contain more activating KIRs than reported in other African populations

We also determined *KIR* gene frequencies in our samples and compared their frequencies with neighbouring populations (**Fig. 2**). There were no significant differences in frequency between the populations for any of the inhibitory *KIR* genes. On the other hand, activating receptors were present at higher frequencies in Caio compared to neighbouring populations (**Fig. 2**). Interestingly, *KIR3DS1*, which is relatively rare in people of African ancestry, was not uncommon in Caio. *KIR3DS1* has previously been shown to confer protection against rapid progression to AIDS following HIV-1 infection in combination with its putative ligand HLA-B Bw4-80I (27).

HLA is associated with resistance to HIV-2 infection and predicts disease progression

In order to determine whether certain *HLA* class I alleles could predict HIV-2 disease progression, we analyzed the $CD4^+$ T-cell counts and viral load data from 141 HIV-2 infected individuals who provided a blood sample in 2003. The overall mean square root (sqrt) $CD4^+$

T-cell count was 24.68, (SE: 0.55) and mean log HIV-2 viral load was 3.07, (SE: 0.09). Analysis of *HLA-B* genotypes, $CD4^+$ T-cell counts (n=131) and viral load (n=136) data from HIV-2 infected subjects revealed that *HLA-B*15+* individuals had significantly lower mean sqrt $CD4^+$ T-cell counts (22.61, SE: 0.86, p = 0.001) and higher mean log HIV-2 viral load (3.41, SE: 0.14, p = 0.001) compared to those without this genotype (26.27, SE: 0.69; and 2.81, SE: 0.11, respectively) (**Table II**). This association appeared to be due completely to the frequent *B*1503* allele, which showed lower mean sqrt $CD4^+$ T-cell counts (21.18, SE: 1.20, p = 0.001) and higher mean log HIV-2 viral load (3.48, SE: 0.20, p = 0.020) compared to that of *B*1503*-negative individuals (25.75, SE: 0.60; and 2.93, SE: 0.10, respectively). No susceptibility effect was seen with the other common *B*1510* allele. *HLA-B*49* was also associated with lower mean sqrt $CD4^+$ T-cell count (21.51, SE: 1.53, p = 0.02) and a tendency towards a higher mean log HIV-2 viral load (**Table II**). Not surprisingly, *HLA-Cw*02*, which is in linkage disequilibrium with *B*15*, also associated with lower mean sqrt $CD4^+$ T-cell count (20.43, SE: 1.46, p = 0.002) and trended towards high mean log HIV-2 viral load. Other alleles weakly associated with markers of disease progression are shown in **Table II**. Interestingly, most of the protective effects observed in this study are mediated by relatively rare alleles (B*82, B*44, and B*14) as has also been observed for HIV-1 infection (28). However, given the limited numbers and weak effects in most of these analyses, further studies will be required to verify these associations.

Next, we compared two-digit *HLA* class I genotype frequencies for alleles present in at least 2% of the study population, between cases and controls (**Table III**). All three loci were in Hardy-Weinberg (H-W) equilibrium (p-values = 0.18517, 0.25241, and 0.10134 for *HLA-A*, *-B*, and *-C* loci respectively). Two alleles, *HLA*A80* and *HLA-B*08*, were more frequent in HIV-2 infected individuals than controls (5.3% vs. 0.7%, p = 0.006, and 27.1% vs 14.6%, p = 0.003, respectively) suggesting that they are associated with susceptibility to HIV-2 infection

in this population. Only one subtype of *A*80* (*A*8001*) and *B*08* (*B*0801*) was present in Caio. None of the *HLA-C* alleles were associated with susceptibility or resistance to HIV-2 infection.

We also estimated haplotype frequencies using Arlequin version 3.11 (29) and the most frequent two- and three-locus haplotypes were *HLA-B*1503-Cw*0210* (9.8%), *HLA-A*330301-B*1510-Cw*0304* (6.2%), and *HLA-A*2301-B*1503-Cw*0210* (5.3%) (data not shown). None of the haplotypes had any significant influence on HIV-2 acquisition or disease progression.

HLA and KIR compound genotypes and resistance to HIV-2 infection

KIR molecules are known to interact with their HLA class I ligands to modulate NK cell activity. The ligands for KIR2DL are HLA-C alleles which are classified as C-group 1 (C1) if the amino acid at position 80 is asparagine or C-group 2 (C2) if lysine occupies that position. The inhibitory KIR2DL2 and 2DL3 (which are alleles of the same locus) recognize C1 while KIR2DL1 recognizes C2 allotypes. KIR3DL1 recognizes HLA-B Bw4 allotypes, particularly those with isoleucine at position 80 (30). The activating receptors KIR2DS2, 2DS1 and 3DS1 share high sequence similarity in their extracellular domains with the corresponding inhibitory receptors KIR2DL2/3, 2DL1 and 3DL1, respectively. KIR2DS1 and 2DS2 appear to bind the same set of HLA class I ligands as their inhibitory counterparts, although with much lower affinity (31). Epidemiological data are consistent with a receptor-ligand relationship between KIR3DS1 and HLA-B Bw4-80I, but this has never been shown formally.

To determine the effect of *HLA/KIR* genotypes on susceptibility or resistance to HIV-2 infection, we grouped individuals based on whether they had appropriate ligands for their *KIR* genetic profile, and compared the various *HLA/KIR* combinations between HIV-2 infected and

uninfected groups. We also looked at HIV-2 viral load and $CD4^+$ T-cell counts in the infected group. There was no significant effect of individual *KIR* genes on infection or markers of disease progression (data not shown). However, individuals carrying either the activating *KIR2DS2* or the inhibitory *KIR2DL2* (these two *KIR* genes are in very strong linkage disequilibrium (LD) with one another), with at least one copy of their corresponding HLA ligands (C1) were more likely to be HIV-2 negative compared to those without these compound genotypes, raising the possibility that the compound genotypes *2DS2:HLA-C1/x* and/or *2DL2:HLA-C1/x* (where x could be C1 or C2) might protect against HIV-2 acquisition (**Table IV**). Due to the strong LD between *KIR2DL2* and *2DS2*, it is difficult to determine which locus is implicated in the protective effect. Analysis of the various *HLA/KIR* gene profiles and HIV-2 viral load and $CD4^+$ T-cell counts found no trend indicative of an influence on these markers of disease progression.

Discussion

The HIV epidemic in West Africa is caused by two closely related retroviruses HIV-1 and HIV-2. The latter, which is the focus of this study, was first reported to have infected healthy Senegalese prostitutes in 1985 (32) and was isolated the following year from two West-African patients with AIDS: one from Cape Verde Islands and the other from Guinea-Bissau (1). HIV-2 is structurally very similar to HIV-1 and shares between 30 to 60% homology at the nucleotide level.

Although both HIV-1 and HIV-2 use the same repertoire of co-receptors and target similar cell populations, their rate of infection and the resultant pathogenesis are quite different. The majority of HIV-2 infected people in Caio show high rates of long-term non-progression (LTNP). Indeed, recent studies in Caio have shown that the majority of HIV-2 infected subjects that presented in 1991 (when viral load measurements were first made) with

viral load below detection have maintained undetectable viral loads for at least 15 years: these subjects have the same mortality risk as HIV uninfected subjects (Schim van der Loeff, ms submitted). However, in contrast to the ancestral virus (SIVsm) infecting its natural host, the sooty mangabey, a significant proportion of HIV-2-infected subjects later progress to AIDS, with low CD4 count, high HIV-2 viral load and clinical features that are indistinguishable from that observed in HIV-1 infection when CD4 count falls below 200 cells/μl (33). The observation that the majority of HIV-2 infected individuals are LTNPs suggests that HIV-2 infection presents a unique human model of naturally attenuated HIV infection that could provide considerable insights into the mechanism(s) of protective immunity needed for HIV vaccine development. The mechanisms underlying the attenuated course of HIV-2 infection and lack of disease progression in a substantial proportion of HIV-2 infected individuals are yet to be uncovered, but previous studies in Caio have shown a strong correlation between T-cell responses to a conserved region of HIV-2 gag and low plasma viral load (5).

Accumulating epidemiological evidence indicates that a range of distinct viral and host genetics and possibly some environmental factors contribute to the observed differences in clinical outcome after HIV infection. Amongst the host genetic factors are the *HLA* and *KIR* genes that are highly polymorphic and are very important antiviral innate and adaptive immune response regulators in humans. Epidemiological studies have implicated certain *HLA* alleles in rate of progression to AIDS in HIV-1 disease (8) and provided evidence that *HLA* genes are involved in the control of viral replication (34, 35), but their role in HIV-2 infection and disease progression is not clear. A small study on 62 female commercial sex workers infected with HIV-2, using p26 antigen exposure as a surrogate marker for disease progression, showed that *HLA-B*35* was associated with lack of p26 antibodies and higher risk of disease progression (36). However, a detailed analysis on the effects of *HLA* and *KIR* genes on susceptibility to HIV-2 and disease progression is largely missing in the literature. In

this study, we determine for the first time the frequencies of *HLA* class I and *KIR* genes simultaneously from a predominantly Manjako community (>95%) in Caio, Guinea-Bissau and demonstrated that specific *HLA* alleles can influence disease progression and risk of HIV-2 acquisition.

Comparison of *HLA* allele frequencies between Caio and neighbouring populations showed distinct differences in allele frequencies that are consistent with previous anthropological findings suggesting that Manjakos from the Caio sector are a unique set of people forming a close-knit community that tends to be isolated sexually from other ethnic groups in the sub region. Recent studies on sequence analysis of the gag and nef genes in members of the Caio cohort showed uniform infection with closely related strains of clade A HIV-2 (Onyango et al, Vaccine 2009 in press, Feldmann et al, manuscript in preparation). Preliminary results from a similar study looking at the effect of *HLA* and *KIR* genes in a multi-ethnic cohort in Fajara, The Gambia revealed that the Manjakos living in The Gambia exhibit similar *HLA* class I allele frequencies to those found in this study suggesting that there is little admixture between Manjakos and other tribes in the sub region irrespective of their geographical location (Yindom et al. manuscript in preparation).

Analysis of our *KIR* gene pool comparing frequencies of individual *KIR* genes with the scant available data from other West African populations showed that most of the activating *KIR* genes were present at higher frequencies in Caio (**Fig. 2**). *KIR3DS1* which is rarely found in people of African ancestry was significantly higher in the Caio population as compared to other West African populations such as Senegal (37), Nigeria (38), and The Gambia (Yindom et al. manuscript in preparation). *KIR3DS1* has previously been shown to confer protection against rapid progression to AIDS following HIV-1 infection in combination with its putative ligand HLA-B Bw4-80I (27).

Data on specific dates of seroconversion were not available for this cohort, as is usually the case in most African cohorts. We relied on information collected on known markers of HIV disease progression such as CD4 counts and viral load measurements to predict the effect of *HLA* class I and *KIR* gene variants on progression to AIDS following HIV-2 infection. We analysed the *HLA-A, -B* and *-C* allele and genotype data from individuals singly infected with HIV-2 and found that individuals carrying the common *B*1503* allele were more likely to progress to AIDS than those without this allele. No susceptibility effect was seen with the other common *B*15* allotype, *B*1510. B*1503* and *B*1510* differ by only 3 amino acids, all at critical sites in the peptide binding groove: positions 63 and 67, which reside in the alpha helix and contribute to the P2 pocket, and position 116, which resides on the floor of the peptide binding groove and is part of the P9 pocket. A single amino acid change at position 116 in *B*3501* vs. *B*3503* has previously been implicated in differential susceptibility to HIV-1 disease progression (9), emphasizing the importance of this position in distinguishing the functional activity of HLA class I molecules. *HLA-B*1503* has been associated with low viral loads in individuals infected with HIV-1 clade B, but no protective effect of this allele was observed in a C clade cohort from Durban, South Africa (39). *HLA-B*1503* and the haplotype *B*1503-Cw*0210* are unique makers confined to populations of Sub-Sahara African origins (40). Therefore, the effect of *HLA-B*1503* is more likely to be observed in these populations than in populations with a low frequency of this allele. Some minor/rare alleles were found to weakly associate with high CD4 count and a trend towards low viral load, indicating that they may offer a protective effect against rapid progression to AIDS following HIV-2 infection. This supports the notion proposed by Trachtenberg and co workers that rare *HLA* alleles may have a selective advantage in protecting against rapid progression to AIDS in populations where HIV-1 has adapted to the most frequent *HLA* alleles (28).

Concluding remarks

Here we determined *HLA* class I and *KIR* gene profiles of a relatively isolated population in Caio, Guinea-Bissau, with one of the highest prevalence rates of HIV-2 infection in the world. The *HLA-B*15* alleles *B*1503* and *B*1510* were observed at a relatively high frequency in this community compared to neighbouring populations. Furthermore, HIV-2-infected individuals with *B*1503* (but not *B*1510*) had significantly higher HIV-2 viral loads and lower CD4 counts compared to those without this allele, suggesting that this allele might be linked to poor control of viral replication and more rapid disease progression. Notably, none of the strongest *HLA* associations with HIV-1 were observed in our HIV-2 cohort. The frequencies of activating *KIR* genes were higher than reported for other populations in West Africa. The frequency of *KIR3DS1*, which was previously shown to be protective in HIV-1 disease progression, was significantly higher in the Manjako group than that reported in any other African population. *KIR2DS2* and *KIR2DL2* were weakly associated with reduced risk of HIV-2 infection. In general, the strongest associations in our study conferred susceptibility to HIV-2 outcomes, while protective factors were quite weak. This observation is contrary to that observed in HIV-1 disease, where the strongest *HLA* associations confer protection. Perhaps this reflects the less pathogenic nature of HIV-2 and the ability of most HLA class I allotypes to effectively control the virus, as compared to HIV-1, where most allotypes are unable to maintain viral restriction. Our study is the first to provide a detailed analysis of the effects of *HLA* and *KIR* genetic variation on resistance/susceptibility to HIV-2 infection and disease progression. It will be of great interest to determine whether these effects can be validated in the other rare HIV-2 cohorts collected to date.

Acknowledgements

We are very grateful to all participants in this study. We also thank all Caio past investigators, the Caio field staff and all MRC Fajara staff (past or present) who contributed to the success of this work. Special thanks to David Jeffries and Matthew Cotten for useful comments and advice, and to Darlene Marti and Fuh-Mei Duh for technical assistance.

REFERENCES

1. Clavel, F., D. Guetard, F. Brun-Vezinet, S. Chamaret, M. A. Rey, M. O. Santos-Ferreira, A. G. Laurent, C. Dauguet, C. Katlama, C. Rouzioux, and et al. 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science (New York, N.Y* 233:343-346.
2. Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon. 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* 326:662-669.
3. Rowland-Jones, S., J. Sutton, K. Ariyoshi, T. Dong, F. Gotch, S. McAdam, D. Whitby, S. Sabally, A. Gallimore, T. Corrah, and et al. 1995. HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women. *Nat Med* 1:59-64.
4. Rowland-Jones, S. L., and H. C. Whittle. 2007. Out of Africa: what can we learn from HIV-2 about protective immunity to HIV-1? *Nature immunology* 8:329-331.
5. Leligdowicz, A., L. M. Yindom, C. Onyango, R. Sarge-Njie, A. Alabi, M. Cotten, T. Vincent, C. da Costa, P. Aaby, A. Jaye, T. Dong, A. McMichael, H. Whittle, and S. Rowland-Jones. 2007. Robust Gag-specific T cell responses characterize viremia control in HIV-2 infection. *The Journal of clinical investigation* 117:3067-3074.
6. Duvall, M. G., A. Jaye, T. Dong, J. M. Brenchley, A. S. Alabi, D. J. Jeffries, M. van der Sande, T. O. Togun, S. J. McConkey, D. C. Douek, A. J. McMichael, H. C. Whittle, R. A. Koup, and S. L. Rowland-Jones. 2006. Maintenance of HIV-specific CD4+ T cell help distinguishes HIV-2 from HIV-1 infection. *J Immunol* 176:6973-6981.
7. Bjorling, E., G. Scarlatti, A. von Gegerfelt, J. Albert, G. Biberfeld, F. Chiodi, E. Norrby, and E. M. Fenyo. 1993. Autologous neutralizing antibodies prevail in HIV-2 but not in HIV-1 infection. *Virology* 193:528-530.
8. Carrington, M., and S. J. O'Brien. 2003. The influence of HLA genotype on AIDS. *Annu Rev Med* 54:535-551.
9. Gao, X., G. W. Nelson, P. Karacki, M. P. Martin, J. Phair, R. Kaslow, J. J. Goedert, S. Buchbinder, K. Hoots, D. Vlahov, S. J. O'Brien, and M. Carrington. 2001. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* 344:1668-1675.
10. Mackelprang, R. D., G. John-Stewart, M. Carrington, B. Richardson, S. Rowland-Jones, X. Gao, D. Mbori-Ngacha, J. Mabuka, B. Lohman-Payne, and C. Farquhar. 2008. Maternal HLA homozygosity and mother-child HLA concordance increase the risk of vertical transmission of HIV-1. *The Journal of infectious diseases* 197:1156-1161.

11. MacDonald, K. S., J. Embree, S. Njenga, N. J. Nagelkerke, I. Ngatia, Z. Mohammed, B. H. Barber, J. Ndinya-Achola, J. Bwayo, and F. A. Plummer. 1998. Mother-child class I HLA concordance increases perinatal human immunodeficiency virus type 1 transmission. *The Journal of infectious diseases* 177:551-556.
12. Lanier, L. L. 2008. Up on the tightrope: natural killer cell activation and inhibition. *Nature immunology* 9:495-502.
13. Jonsson, A. H., and W. M. Yokoyama. 2009. Natural killer cell tolerance licensing and other mechanisms. *Adv Immunol* 101:27-79.
14. Carrington, M., M. P. Martin, and J. van Bergen. 2008. KIR-HLA intercourse in HIV disease. *Trends in microbiology* 16:620-627.
15. Berry, N., S. Jaffar, M. Schim van der Loeff, K. Ariyoshi, E. Harding, P. T. N'Gom, F. Dias, A. Wilkins, D. Ricard, P. Aaby, R. Tedder, and H. Whittle. 2002. Low level viremia and high CD4% predict normal survival in a cohort of HIV type-2-infected villagers. *AIDS research and human retroviruses* 18:1167-1173.
16. Schim van der Loeff, M. F., and P. Aaby. 1999. Towards a better understanding of the epidemiology of HIV-2. *Aids* 13 Suppl A:S69-84.
17. Wilkins, A., D. Ricard, J. Todd, H. Whittle, F. Dias, and A. Paulo Da Silva. 1993. The epidemiology of HIV infection in a rural area of Guinea-Bissau. *Aids* 7:1119-1122.
18. Ricard, D., A. Wilkins, P. T. N'Gum, R. Hayes, G. Morgan, A. P. Da Silva, and H. Whittle. 1994. The effects of HIV-2 infection in a rural area of Guinea-Bissau. *Aids* 8:977-982.
19. Holmgren, B., P. Aaby, H. Jensen, O. Larsen, Z. da Silva, and I. M. Lisse. 1999. Increased prevalence of retrovirus infections among older women in Africa. *Scandinavian journal of infectious diseases* 31:459-466.
20. Aaby, P., K. Ariyoshi, M. Buckner, H. Jensen, N. Berry, A. Wilkins, D. Richard, O. Larsen, F. Dias, M. Melbye, and H. Whittle. 1996. Age of wife as a major determinant of male-to-female transmission of HIV-2 infection: a community study from rural West Africa. *Aids* 10:1585-1590.
21. Buckner, M. 2000. Manjako sex and gender.
22. Schmidt, W. P., M. S. Van Der Loeff, P. Aaby, H. Whittle, R. Bakker, M. Buckner, F. Dias, and R. G. White. 2008. Behaviour change and competitive exclusion can explain the diverging HIV-1 and HIV-2 prevalence trends in Guinea-Bissau. *Epidemiology and infection* 136:551-561.
23. Martin, M. P., and M. Carrington. 2008. KIR locus polymorphisms: genotyping and disease association analysis. *Methods in molecular biology (Clifton, N.J* 415:49-64.
24. Sanchez-Mazas, A., Q. G. Steiner, C. Grundschober, and J. M. Tiercy. 2000. The molecular determination of HLA-Cw alleles in the Mandenka (West Africa) reveals a close genetic relationship between Africans and Europeans. *Tissue antigens* 56:303-312.
25. Cao, K., A. M. Moormann, K. E. Lyke, C. Masaberg, O. P. Sumba, O. K. Doumbo, D. Koech, A. Lancaster, M. Nelson, D. Meyer, R. Single, R. J. Hartzman, C. V. Plowe, J. Kazura, D. L. Mann, M. B. Sztein, G. Thomson, and M. A. Fernandez-Vina. 2004. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue antigens* 63:293-325.
26. Serwanga, J., L. A. Shafer, E. Pimego, B. Auma, C. Watera, S. Rowland, D. Yirrell, P. Pala, H. Grosskurth, J. Whitworth, F. Gotch, and P. Kaleebu. 2009. Host HLA B*allele-associated multi-clade Gag T-cell recognition correlates with slow HIV-1 disease progression in antiretroviral therapy-naive Ugandans. *PLoS ONE* 4:e4188.
27. Martin, M. P., X. Gao, J. H. Lee, G. W. Nelson, R. Detels, J. J. Goedert, S. Buchbinder, K. Hoots, D. Vlahov, J. Trowsdale, M. Wilson, S. J. O'Brien, and M.

Carrington. 2002. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature genetics* 31:429-434.

28. Trachtenberg, E., B. Korber, C. Sollars, T. B. Kepler, P. T. Hraber, E. Hayes, R. Funkhouser, M. Fugate, J. Theiler, Y. S. Hsu, K. Kunstman, S. Wu, J. Phair, H. Erlich, and S. Wolinsky. 2003. Advantage of rare HLA supertype in HIV disease progression. *Nature medicine* 9:928-935.
29. Excoffier, L. G. L., and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
30. Carr, W. H., M. J. Pando, and P. Parham. 2005. KIR3DL1 polymorphisms that affect NK cell inhibition by HLA-Bw4 ligand. *J Immunol* 175:5222-5229.
31. Stewart, C. A., F. Laugier-Anfossi, F. Vely, X. Saulquin, J. Riedmuller, A. Tisserant, L. Gauthier, F. Romagne, G. Ferracci, F. A. Arosa, A. Moretta, P. D. Sun, S. Ugolini, and E. Vivier. 2005. Recognition of peptide-MHC class I complexes by activating killer immunoglobulin-like receptors. *Proceedings of the National Academy of Sciences of the United States of America* 102:13224-13229.
32. Barin, F., S. M'Boup, F. Denis, P. Kanki, J. S. Allan, T. H. Lee, and M. Essex. 1985. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of west Africa. *Lancet* 2:1387-1389.
33. De Cock, K. M., K. Odehouri, R. L. Colebunders, G. Adjorlolo, M. F. Lafontaine, A. Porter, E. Gnaore, L. Diaby, J. Moreau, W. L. Heyward, and et al. 1990. A comparison of HIV-1 and HIV-2 infections in hospitalized patients in Abidjan, Cote d'Ivoire. *Aids* 4:443-448.
34. Kaslow, R. A., M. Carrington, R. Apple, L. Park, A. Munoz, A. J. Saah, J. J. Goedert, C. Winkler, S. J. O'Brien, C. Rinaldo, R. Detels, W. Blattner, J. Phair, H. Erlich, and D. L. Mann. 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature medicine* 2:405-411.
35. Fellay, J., K. V. Shianna, D. Ge, S. Colombo, B. Ledergerber, M. Weale, K. Zhang, C. Gumbs, A. Castagna, A. Cossarizza, A. Cozzi-Lepri, A. De Luca, P. Easterbrook, P. Francioli, S. Mallal, J. Martinez-Picado, J. M. Miro, N. Obel, J. P. Smith, J. Wyniger, P. Descombes, S. E. Antonarakis, N. L. Letvin, A. J. McMichael, B. F. Haynes, A. Telenti, and D. B. Goldstein. 2007. A whole-genome association study of major determinants for host control of HIV-1. *Science (New York, N.Y* 317:944-947.
36. Diouf, K., A. D. Sarr, G. Eisen, S. Popper, S. Mboup, and P. Kanki. 2002. Associations between MHC class I and susceptibility to HIV-2 disease progression. *Journal of human virology* 5:1-7.
37. Denis, L., J. Sivula, P. A. Gourraud, N. Kerdudou, R. Chout, C. Ricard, J. P. Moisan, K. Gagne, J. Partanen, and J. D. Bignon. 2005. Genetic diversity of KIR natural killer cell markers in populations from France, Guadeloupe, Finland, Senegal and Reunion. *Tissue antigens* 66:267-276.
38. Single, R. M., M. P. Martin, X. Gao, D. Meyer, M. Yeager, J. R. Kidd, K. K. Kidd, and M. Carrington. 2007. Global diversity and evidence for coevolution of KIR and HLA. *Nature genetics* 39:1114-1119.
39. Frahm, N., P. Kiepiela, S. Adams, C. H. Linde, H. S. Hewitt, K. Sango, M. E. Feeney, M. M. Addo, M. Lichterfeld, M. P. Lahaie, E. Pae, A. G. Wurcel, T. Roach, M. A. St John, M. Altfeld, F. M. Marincola, C. Moore, S. Mallal, M. Carrington, D. Heckerman, T. M. Allen, J. I. Mullins, B. T. Korber, P. J. Goulder, B. D. Walker, and C. Brander. 2006. Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes. *Nature immunology* 7:173-178.

40. Middleton, D., L. Menchaca, H. Rood, and R. Komerofsky. 2003. New allele frequency database: http://www.allelefrequencies.net. *Tissue antigens* 61:403-407.

FOOTNOTES

[1]This work was funded by the European and Developing Countries Clinical Trials Partnership (EDCTP) and by the Medical Research Council (UK) The Gambia. This project was funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This Research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Figure legends

Figure 1. *HLA* allele frequencies in Caio, Guinea Bissau compared to frequencies in neighbouring West African Countries. a) *HLA-A*, b) *HLA-B*, c) *HLA-C* allele frequencies in Guinea-Bissau (513 samples) and neighbouring countries: Senegal (165 samples), The Gambia (592 samples), and Mali (138 samples). Alleles with frequencies of <1% in all populations are not shown.*: $p < 0.05$; **: $p < 0.001$.

Figure 2. Frequencies of activating and inhibitory *KIR* genes in Caio, Guinea-Bissau compared to those in The Gambia and Senegal. *: $p<0.05$; **: $p<0.001$.

Table I. Characteristics of the study population (n = 513), Caio, 2003-7

	N (%)	Median age (IQR)
HIV-negative		
Male	114 (35)	36.2 (25.8-52.5)
Female	213 (65)	46.2 (30.8-62.8)
HIV-2 infected		
Male	49 (32)	53.7 (43.0-62.9)
Female	102 (68)	59.5 (49.5-69.4)
HIV-1&-2 dually infected		
Male	6 (17)	37.0 (34.7-49.3)
Female	29 (83)	52.8 (42.8-61.7)

Age difference between males and females was not statistically significant between cases and controls.

Table II. Effects of *HLA-B* and *HLA-C* on $CD4^+$ T lymphocyte counts and viral load

HLA (n=135)	Mean sqrt CD4 (n)	SE	p	Mean logHIV-2 VL (n)	SE	p
*B*14* (P)	26.81 (14)	1.71	0.23	2.39 (14)	0.29	**0.02**
*B*15* (S)	22.61 (51)	0.86	**0.001**	3.41 (52)	0.14	**0.001**
*B*1503* (S)	21.18 (26)	1.20	**0.001**	3.48 (28)	0.20	**0.02**
*B*1510*	24.78 (24)	1.31	0.96	3.22 (25)	0.22	0.36
*B*1516*	19.75 (3)	3.68	0.16	3.55 (2)	0.78	0.51
*B*44 (P)*	27.01 (7)	2.42	0.36	2.27 (8)	0.38	**0.04**
*B*49 (S)*	21.51 (17)	1.53	**0.02**	3.39 (20)	0.24	0.12
*B*57 (P)*	29.82 (7)	2.37	**0.03**	3.30 (7)	3.30	0.51
*B*82 (P)*	27.71 (3)	3.68	0.43	1.70 (3)	0.61	**0.03**
*Cw*02* (S)	20.43 (19)	1.46	**0.002**	3.41 (20)	0.24	0.14
*Cw*07* (S)	24.39 (47)	0.96	0.73	3.35 (50)	0.15	**0.03**

(P): Protective allele with either a mean sqt $CD4^+$ T cell count greater than 24.7 (average mean sqrt $CD4^+$ T cell counts) and/or mean logVL less than 3.1 (average mean HIV-2 VL) and a p-value < 0.05; (S): Susceptible allele with either a mean sqrt $CD4^+$ T cell count less than 24.7 (average mean $CD4^+$ T cell counts) and/or mean logVL greater than 3.1 (average mean HIV-2 VL) and a p-value < 0.05. p-values stated are uncorrected for multiple comparisons.

Table III. *HLA-A, -B,* and *-C* genotypes and HIV-2 antibody status in Caio community

Allele	HIV-2$^+$ (n)	HIV-negative (n)	All (n)	p	OR	95% CI
*A*01*	12.0 (18)	14.6 (45)	13.8 (63)	0.45	0.80	0.44 - 1.44
*A*02*	18.7 (28)	24.4 (75)	22.5 (103)	0.20	0.73	0.45 - 1.18
*A*03*	4.7 (7)	4.9 (15)	4.8 (22)	0.92	0.95	0.38 - 2.40
*A*23*	33.3 (50)	31.5 (97)	32.1 (147)	0.63	1.11	0.73 - 1.68
*A*26*	16.7 (25)	13.6 (42)	14.6 (67)	0.37	1.28	0.75 - 2.20
*A*29*	1.3 (2)	2.9 (9)	2.4 (11)	0.32	0.46	0.10 - 2.15
*A*30*	24.0 (36)	17.2 (53)	19.4 (89)	0.08	1.54	0.95 - 2.48
*A*33*	38.0 (57)	39.3 (121)	38.9 (178)	0.84	0.96	0.64 - 1.43
*A*34*	4.7 (7)	3.9 (12)	4.2 (19)	0.69	1.21	0.47 - 3.15
*A*66*	2.0 (3)	3.6 (11)	3.1 (14)	0.36	0.55	0.15 - 1.99
*A*68*	18.0 (26)	14.3 (45)	15.5 (71)	0.43	1.24	0.73 - 2.10
*A*74*	7.3 (10)	11.7 (37)	10.3 (47)	0.09	0.53	0.26 - 1.10
A*80	5.3 (8)	0.7 (2)	2.2 (10)	**0.006**	8.90	1.86 - 42.53
*B*07*	4.9 (7)	7.3 (17)	6.4 (24)	0.37	0.66	0.27 - 1.63
*B*08*	27.1 (39)	14.6 (34)	19.4 (73)	**0.003**	2.20	1.31 - 3.71
*B*13*	5.6 (8)	6.1 (14)	5.8 (22)	0.88	0.93	0.38 - 2.28
*B*14*	10.4 (14)	7.7 (19)	8.8 (33)	0.57	1.23	0.60 - 2.54
*B*15*	38.2 (55)	45.1 (105)	42.4 (160)	0.22	0.77	0.50 - 1.17
*B*18*	6.3 (9)	4.3 (10)	5.0 (19)	0.39	1.51	0.60 - 3.80
*B*35*	13.9 (20)	21.5 (50)	18.6 (70)	0.08	0.60	0.34 - 1.06
*B*44*	5.6 (8)	5.2 (12)	5.3 (20)	0.85	1.09	0.44 - 2.75
*B*49*	13.9 (20)	13.3 (31)	13.8 (51)	0.84	1.07	0.58 - 1.95
*B*50*	2.1 (3)	2.2 (5)	2.1 (8)	0.98	0.98	0.23 - 4.16
*B*51*	3.5 (5)	3.0 (7)	3.2 (12)	0.87	1.17	0.36 - 3.76
*B*52*	1.4 (2)	3.8 (9)	2.9 (11)	0.19	0.36	0.08 - 1.67
*B*53*	16.7 (23)	23.2 (55)	20.7 (78)	0.08	0.62	0.36 - 1.07

*B*57*	4.9 (7)	1.7 (4)	2.9 (11)	0.09	2.95	0.85 - 10.28
*B*58*	22.2 (32)	19.3 (45)	20.4 (77)	0.46	1.22	0.73 - 2.03
*Cw*01*	8.9 (13)	7.5 (21)	7.9 (34)	0.60	1.21	0.59 - 2.50
*Cw*02*	14.9 (22)	21.3 (60)	19.1 (82)	0.12	0.66	0.38 - 1.12
*Cw*03*	39.9 (59)	44.0 (124)	42.6 (183)	0.45	0.86	0.57 - 1.28
*Cw*04*	29.7 (43)	29.8 (85)	29.8 (128)	0.87	0.96	0.62 - 1.49
*Cw*05*	8.1 (12)	4.3 (12)	5.6 (24)	0.10	2.02	0.88 - 4.61
*Cw*06*	4.7 (11)	3.9 (27)	4.2 (18)	0.45	0.75	0.36 - 1.57
*Cw*07*	37.8 (56)	33.7 (95)	35.1 (151)	0.35	1.22	0.81 - 1.85
*Cw*08*	14.2 (20)	15.3 (44)	14.8 (64)	0.59	0.85	0.48 - 1.51
*Cw*14*	1.4 (2)	3.9 (11)	3.0 (13)	0.16	0.34	0.07 - 1.55
*Cw*15*	5.4 (8)	3.9 (11)	4.4 (19)	0.44	1.45	0.57 - 3.69
*Cw*16*	5.4 (8)	7.5 (21)	6.7 (29)	0.44	0.72	0.31 - 1.67
*Cw*17*	3.4 (5)	2.1 (6)	2.6 (11)	0.43	1.62	0.48 - 5.39
*Cw*18*	4.1 (6)	3.5 (10)	3.7 (16)	0.78	1.16	0.41 - 3.26

All: represents the percentage of individuals positive for the indicated allele in the community; n: number of individuals carrying the allele of interest; p: p-values uncorrected for multiple comparisons; OR: the odds ratio derived by comparing the number of individuals positive for the indicated allele versus those without the allele between cases (HIV-2^{+}) and controls (HIV-negative) groups, p value and OR are calculated by logistic regression with controlling for gender and ethnicity. Only alleles with population frequency of at least 2% are shown.

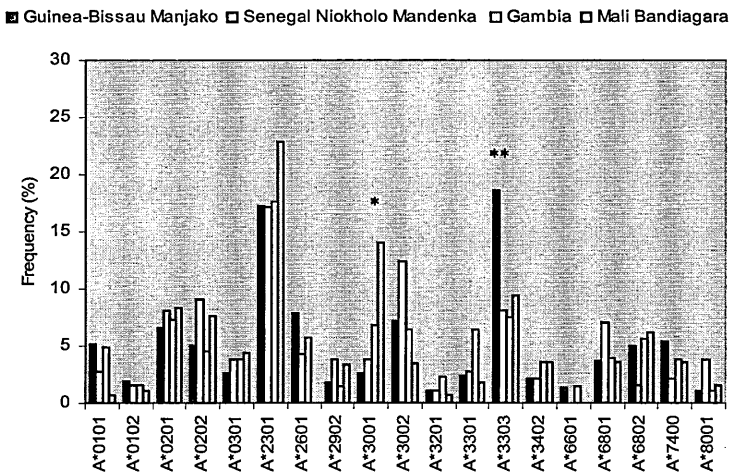
Table IV. The effects of KIR and HLA compound genotypes on HIV-2 antibody status

	HIV-2+		HIV-negative			OR	95% CI	p
	n	%	n	%				
2DS2	66	44.00	178	54.27	*2DS2* vs. others	0.67	0.45-0.98	**0.04**
Others	84	56.00	150	45.73				
2DL2	72	48.00	195	59.45	*2DL2* vs. others	0.63	0.43-0.93	**0.02**
Others	78	52.00	133	40.55				
2DS2+/C1+	49	33.79	124	44.93	*2DS2+ / C1+* vs. others	0.63	0.41-0.95	**0.03**
Others	96	66.21	152	55.07				
2DL2+/C1+	56	38.62	135	48.91	*2DL2+ / C1+* vs. others	0.66	0.44-0.99	**0.04**
Others	89	61.38	141	51.09				

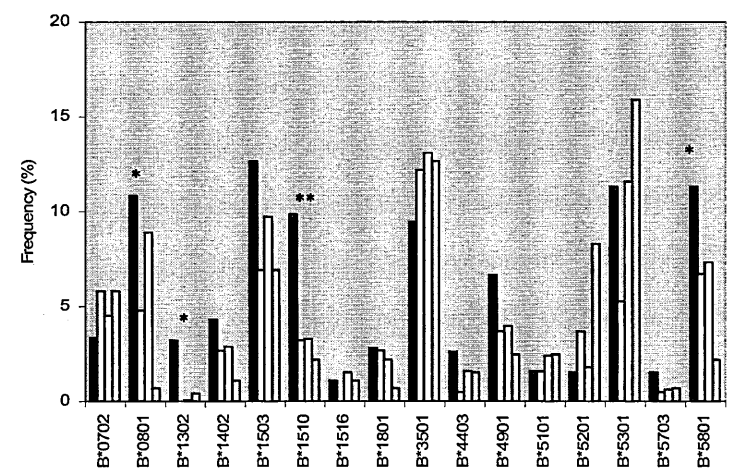
n: number of individuals with the gene/genotype or without the gene/genotype (others); p: p-values uncorrected for multiple comparisons; OR: the odds ratio derived by comparing the number of individuals positive for the indicated allele/genotype versus those without that allele or genotype between cases (HIV-2^{+}) and controls (HIV-negative), p value and OR are calculated by logistic regression with adjusting for gender and ethnicity.

Figure 1

a)



b)



c)

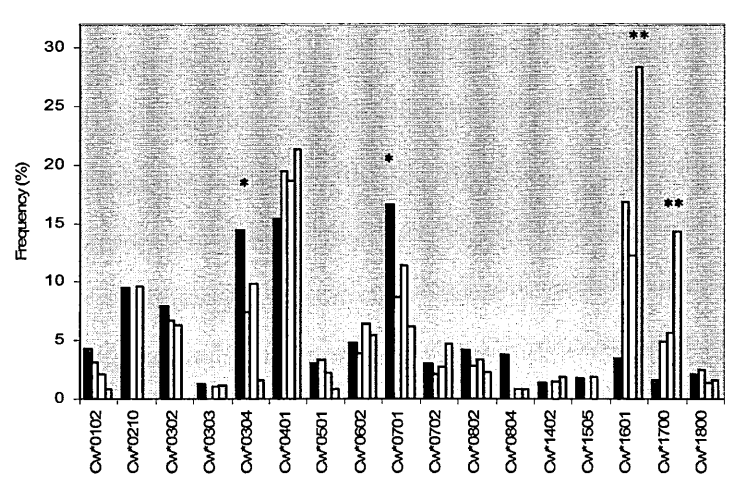
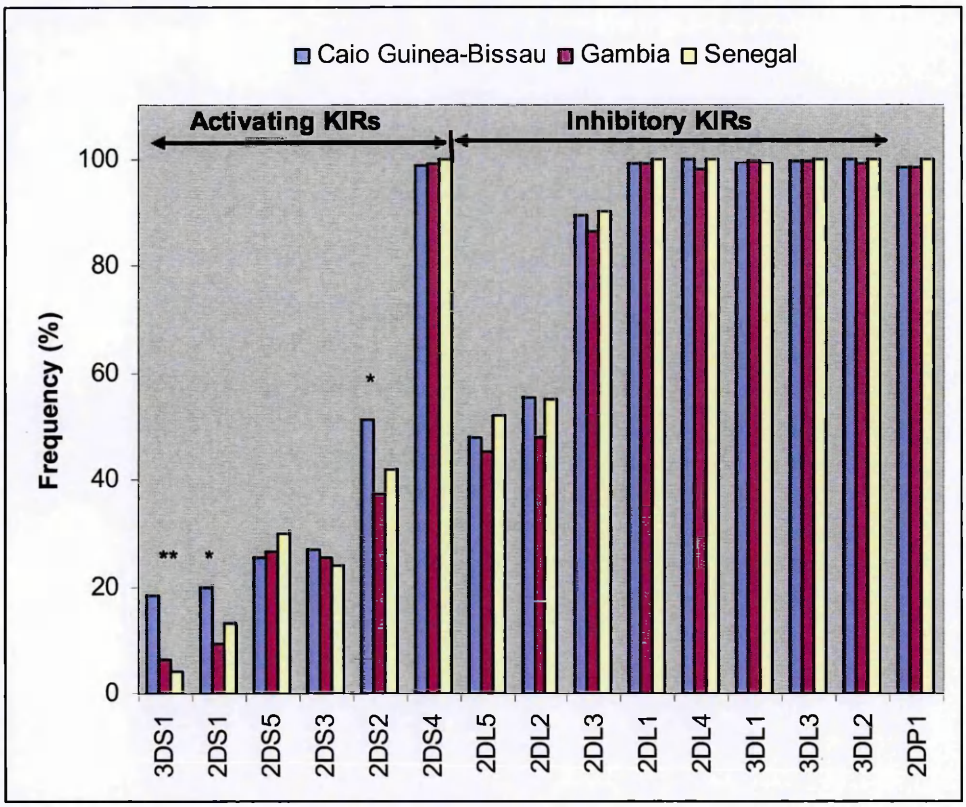


Figure 2


Caio Guinea-Bissau
Gambia
Senegal
Activating KIRs
Inhibitory KIRs
Frequency (%)
100
80
60
40
20
0
**
*
*
3DS1
2DS1
2DS5
2DS3
2DS2
2DS4
2DL5
2DL2
2DL3
2DL1
2DL4
3DL1
3DL3
3DL2
2DP1