

https://dspace.org/

HLA class II associations with HIV-1 control in a black South African population

Murray, Lyle W

2020
https://hdl.handle.net/10539/31690

Downloaded from a DSpace repository. https://dspace.org/

HLA CLASS II ASSOCIATIONS WITH HIV-1 CONTROL IN A BLACK SOUTH AFRICAN POPULATION

Lyle W. Murray

A Research Report submitted to the Faculty of Health Sciences, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Medicine.



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Johannesburg, 2020

CANDIDATE'S DECLARATION

I, Lyle William Murray, declare that this Research Report is my own, unaided work. It is being submitted for the Degree of Master of Medicine (MMed) at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.

_____

(Signature of Dr LW Murray)

On this 6th day of October 2020 in Johannesburg

For Katie

ABSTRACT

HLA class I alleles such as *HLA-B*27* and *HLA-B*57* have reliably been shown to associate with differential rates of HIV-1 disease progression. Little is known, however, of the association of HLA class II molecules and HIV-1 control. The increasingly important role that has been demonstrated for CD4+ T cells in HIV-1 pathogenesis suggests that HLA class II molecules may indeed play a significant role in HIV-1 control.

We hypothesized that certain HLA class II alleles would be found at different frequencies in individuals who spontaneously control HIV-1 viraemia in the absence of antiretroviral therapy (Controllers) and in individuals with progressive disease (Progressors). We therefore determined HLA class II *DPB1*, *-DQB1* and *-DRB1* allele frequencies using sequence-based typing in a group of 26 black African Controllers and compared them to the allele frequencies that had previously been determined in 74 black African Progressors recruited from the same community in Johannesburg.

The HLA class II alleles *-DQB1*03:02:01* (P= 0.011, OR 0.16, CI 0.04 – 0.66) and *-DRB1*04:01:01* (P=0.041, OR 0.16, CI 0.03 – 0.93) were found at significantly higher frequencies amongst Controllers than Progressors. In addition, the two-locus HLA class II haplotypes *-DPB1*01:01:01/-DQB1*04:02:01* (P=0.015, OR 0.22, CI 0.07 – 0.74) and *-DPB1*01:01:01/-DQB1*06:02:01* (P=0.029, OR 0.19, CI 0.05 – 0.85) were more frequent amongst Controllers than Progressors.

These particular allele associations with HIV-1 control have not been shown before. These alleles may play a "protective" role in HIV-1 pathogenesis in black Africans. Our findings suggest an important role for HLA class II molecules and CD4+ T cells in HIV-1 control, however, due to the small number of participants in this study these findings highlight the need for better powered and more comprehensive studies on the role of HLA class II molecules on HIV-1 pathogenesis.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the study participants for enrolling in and engaging with our research study. This research would not have been possible without their generous participation. I would like to extend a special thanks to Prof. Neil Martinson (Perinatal HIV Research Unit, Faculty of Health Sciences, University of the Witwatersrand) and Dr. Osman Ebrahim (Brenthurst Clinic; School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand) for facilitating the recruitment of patients.

This work is based on the research supported by grant awards from the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council, a grantee of the Bill & Melinda Gates Foundation, and the South African Research Chairs Initiative of the Department of Science and Innovation and National Research Foundation of South Africa (Prof. C. Tiemessen). The Study of HIV-Associated Lung Infections in Soweto Cohort (Progressor group) was funded by the National Institutes of Health, USA (R01HL090312 and P30AI094189: R. E. Chaisson). I am grateful to the funders for making this important work possible.

Prof. Caroline Tiemessen conceived and designed this study. I would like to thank her for personally supervising this MMed research project. She has been extremely generous with her time, knowledge and expertise as well as in providing the facilities within her laboratory to carry out this research. Most of all, she has been a pleasure to work with and I appreciate the friendly and supportive environment of her research group. Dr Shayne Loubser also deserves special mention, as he mentored me through the laboratory methods as well as co-supervised this project. His valuable experience with HLA genotyping is unparalleled, and he made it easy to learn and understand the techniques that we used. He also provided valuable input in analysing, interpreting and presenting the data, for which I am very grateful.

Finally, I would like to thank my wife Kate for her unwavering support throughout my Internal Medicine specialty training and her constant interest in this MMed project. My parents and siblings have always been supportive, and I am also incredibly grateful to them.

CONTENTS

CANDIDATE'S DECLARATION ..... **ii**

ABSTRACT ..... **iv**

ACKNOWLEDGEMENTS ..... **v**

CONTENTS ..... **vi**

ABBREVIATIONS ..... **vii**

LIST OF FIGURES ..... **viii**

LIST OF TABLES ..... **viii**

1 INTRODUCTION **1**

1.1 General introduction **1**

1.2 Aim, hypothesis and objectives **2**

1.2.1 Aim 2

1.2.2 Hypothesis 2

1.2.3 Objectives 2

1.3 Literature Review **3**

1.3.1 HLA class I and II molecules 3

1.3.2 Designation of HLA alleles, allele frequencies and haplotypes 4

1.3.3 The role of CD8+ T cells in HIV-1 infection 5

1.3.4 HLA class I associations with HIV-1 disease progression 5

1.3.5 The role of CD4+ T cells in HIV-1 infection 6

1.3.6 HLA class II associations with HIV-1 disease progression 7

1.3.7 HLA class I and class II linkage disequilibrium 9

1.3.8 Summary 9

2 Materials and Methods **10**

2.1 Study design **10**

2.2 Study participants **10**

2.3 Funding sources **11**

2.4 HLA class II genotyping **11**

2.4.1 SBT method 11

2.4.2 Resolving of ambiguous typing 13

2.5 Data analysis **15**

3 Results **16**

3.1 Cohort characteristics **16**

3.2 HLA class II allele frequencies in total cohort **17**

3.2.1 Hardy-Weinberg equilibrium 17

3.2.2 Allele frequencies 17

3.3 Comparison of HLA class II allele frequencies between HIV Controllers and Progressors **20**

3.3.1 HLA-DPB1 locus 20

3.3.2 HLA-DQB1 locus 23

3.3.3 HLA-DRB1 locus 25

3.3.4 Two- and three-locus HLA class II haplotypes 28

3.4 Comparison of HLA class II allele frequencies between Elite controllers and Progressors **32**

4 Discussion **34**

5 References **39**

6 Appendix **45**

ABBREVIATIONS

α	Alpha
AIDS	Acquired immune deficiency syndrome
AFND	Allele Frequency Net Database
ART	Antiretroviral therapy
β	Beta
CD	Cluster of differentiation
CI	Confidence interval
CWD	Common and well-documented
DNA	Deoxyribonucleic acid
EC	Elite controller
Env	Envelope gene
Gag	Group-specific antigen gene
GWAS	Genome-wide association study
HIV-1	Human immunodeficiency virus type 1
HLA	Human leucocyte antigen
HW	Hardy-Weinberg
IFNγ	Interferon gamma
LD	Linkage disequilibrium
LTNP	Long-term non-progressor
MHC	Major histocompatibility complex
MSM	Men who have sex with men
NK	Natural killer
OR	Odds ratio
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
SBT	Sequence-based typing
SIV	Simian immunodeficiency virus
SNP	Single-nucleotide polymorphism
TCR	T cell receptor

LIST OF FIGURES

Figure 1. A schematic representation of HLA class I and class II molecules and the HLA genes associated with them. 4

Figure 2. A representative agarose gel showing PCR amplification products at different HLA loci. 12

Figure 3. A screenshot image of a representative *HLA-DPB1* allele analysis taken from the ASSIGN-SBT software program version 4.7. 13

Figure 4. A flow diagram summarising the process undertaken to resolve allele ambiguities. 14

Figure 5. Frequencies of HLA class II -*DPB1, -DQB1* and -*DRB1* alleles in the total cohort 19

Figure 6. Frequencies of HLA class II *DPB1* alleles in Controllers and Progressors 21

Figure 7. Frequencies of HLA class II *DQB1* alleles in Controllers and Progressors 24

Figure 8. Frequencies of HLA class II *DRB1* alleles in Controllers and Progressors 26

LIST OF TABLES

Table 1. Characteristics of study cohort 16

Table 2. Frequencies of *HLA-DPB1* alleles in Controllers and Progressors. 22

Table 3. Frequencies of *HLA-DQB1* alleles in Controllers and Progressors. 24

Table 4. Frequencies of *HLA-DRB1* alleles in Controllers and Progressors. 27

Table 5. Estimated two-locus (*HLA-DPB1/-DQB1* and *HLA-DQB1/-DRB1*) haplotype frequencies in Controllers and Progressors. 30

Table 6. Estimated three-locus *(HLA-DPB1/-DQB1/-DRB1*) haplotype frequencies in Controllers and Progressors. 31

Table 7. Frequencies of *HLA-DPB1, HLA-DQB1* and *HLA-DRB1* alleles in Elite Controllers and Progressors. 33

1 INTRODUCTION

1.1 General introduction

In the absence of antiretroviral therapy (ART), infection with human immunodeficiency virus type 1 (HIV-1) inevitably leads to progressive CD4+ T cell decline and to susceptibility to opportunistic infections and cancer. The tempo of this progression varies markedly, however, with the average duration being 8-10 years before the onset of the Acquired Immune Deficiency Syndrome (AIDS) (Goulder and Walker, 2012).

A minority of individuals are able to persistently control HIV-1 replication to a level less than 2000 RNA copies/mL without ART and are termed "HIV-1 controllers (Controllers)" (Pereyra *et al.*, 2010). A further subgroup of Controllers are able to control viraemia below the level of detection of most common virological assays (<50 RNA copies/mL) and are termed "elite controllers (ECs)" (Deeks and Walker, 2007). These rare individuals typically maintain high CD4+ T cell counts and have a significantly delayed progression to AIDS. Through studying aspects of immunity in individuals who control HIV-1 infection in the absence of ART, we are able to better understand the correlates of protective immune responses in the setting of natural HIV-1 infection. This knowledge has implications for HIV-1 vaccine design, immunological therapies as well as for HIV-1 eradication strategies.

Both innate and adaptive immunity play important roles in controlling HIV-1 replication, however, the predominant role is played by the adaptive immune system. T cells, natural killer (NK) cells and broadly neutralising antibodies exert inhibitory effects on viral replication, resulting in escape mutations in viral epitopes that are subject to immune selection pressure. The role of CD8+ T cells and human leucocyte antigen (HLA) class I molecules in HIV-1 virological control is well documented (Goulder and Walker, 2012). Despite the CD4+ T cell being the primary target of infection by the virus, CD4+ T cells and HLA class II molecules are increasingly being shown to also play an important role in the pathogenesis of HIV-1 infection.

In this study we aim to explore the association of HLA class II genes with HIV-1 control in a black South African cohort, through the comparison of HLA class II allele frequencies between groups of HIV-1 Controllers and Progressors.

1.2 Aim, hypothesis and objectives

1.2.1 Aim

The aim of this study is to investigate the association of HLA class II alleles with HIV-1 control through the comparative analysis of a cohort of black South African HIV-1 Controllers and a cohort of black South African individuals with progressive HIV-1 disease (Progressors).

1.2.2 Hypothesis

Certain HLA Class II alleles are found at different frequencies in HIV-1 Controllers and HIV-infected individuals with progressive disease (Progressors).

1.2.3 Objectives

1. To determine the frequencies of HLA class II *DPB1*, -*DQB1* and -*DRB1* alleles in 26 Controllers.

2. To compare the frequencies of HLA class II *DPB1*, -*DQB1* and -*DRB1* alleles in 26 Controllers to those previously determined in a group of 74 Progressors from the Longitudinal Study of HIV-Associated Lung Infections in Soweto cohort.

3. To compare the frequencies of two- and three-locus HLA class II *DPB1*, -*DQB1* and -*DRB1* haplotypes in 26 Controllers to those previously determined in a group of 74 Progressors from the Longitudinal Study of HIV-Associated Lung Infections in Soweto cohort.

4. To compare the frequencies of HLA class II *DPB1*, -*DQB1* and -*DRB1* alleles in 17 ECs (a sub-group of the 26 HIV-1 Controllers) to those previously determined in a group of 74 Progressors from the Longitudinal Study of HIV-Associated Lung Infections in Soweto cohort.

1.3 Literature Review

1.3.1 HLA class I and II molecules

The HLA complex is located on chromosome 6 and encodes the genes for HLA class I and class II molecules. An alternative name for the HLA complex is the major histocompatibility complex (MHC). This region of the human genome is highly polymorphic and the products of these genes are essential for immunologic specificity and transplantation histocompatibility (Delves *et al.*, 2006). Consequently, they also play an important role in autoimmune disease (Dendrou *et al.*, 2018).

HLA class I molecules consist of a heavy polypeptide chain (also called the α-chain) non-covalently linked to β2-microglobulin (Figure 1). Two extended helices are formed by the α1 and α2 domains of the heavy chain forming a peptide groove above a floor created by a β-pleated sheet (Delves *et al.*, 2006). The shape of the peptide-binding groove dictates which peptides can be presented by each HLA molecule for recognition by the T cell receptor (TCR). There are three different classical HLA class I α-chain genes namely *HLA-A*, *HLA-B* and *HLA-C*, the products of which are expressed on most nucleated cells.

In the case of HLA class II, the molecule is made up of an α and a β polypeptide chain (Figure 1). The α1 and β1 domains form a peptide-binding groove bound by two α helices and a β-pleated sheet floor. There are also three different class II α-chain and β-chain genes namely *HLA-DP*, *HLA-DQ* and *HLA-DR*. HLA class II molecules are expressed predominantly on antigen presenting cells of the immune system (Jameson *et al.*, 2018).

The shape of the peptide-binding groove differs between HLA class I and II molecules with the groove of HLA class II molecules being open at both ends and HLA class I being closed. As a result, the length of peptide bound within the HLA class II complex can vary significantly whilst the length of peptides bound by HLA class I is restricted and is typically 8-11 amino acids in length (Jameson *et al.*, 2018). In addition, HLA molecules have structural contact sites for the T cell molecules CD4 and CD8. This ensures that predominantly CD8+ T cells respond to peptides presented by HLA class I molecules and that CD4+ T cells respond to peptides presented by HLA class II molecules (Jameson *et al.*, 2018). HLA class I molecules present intracellular peptides whilst HLA class II present extracellular peptides.

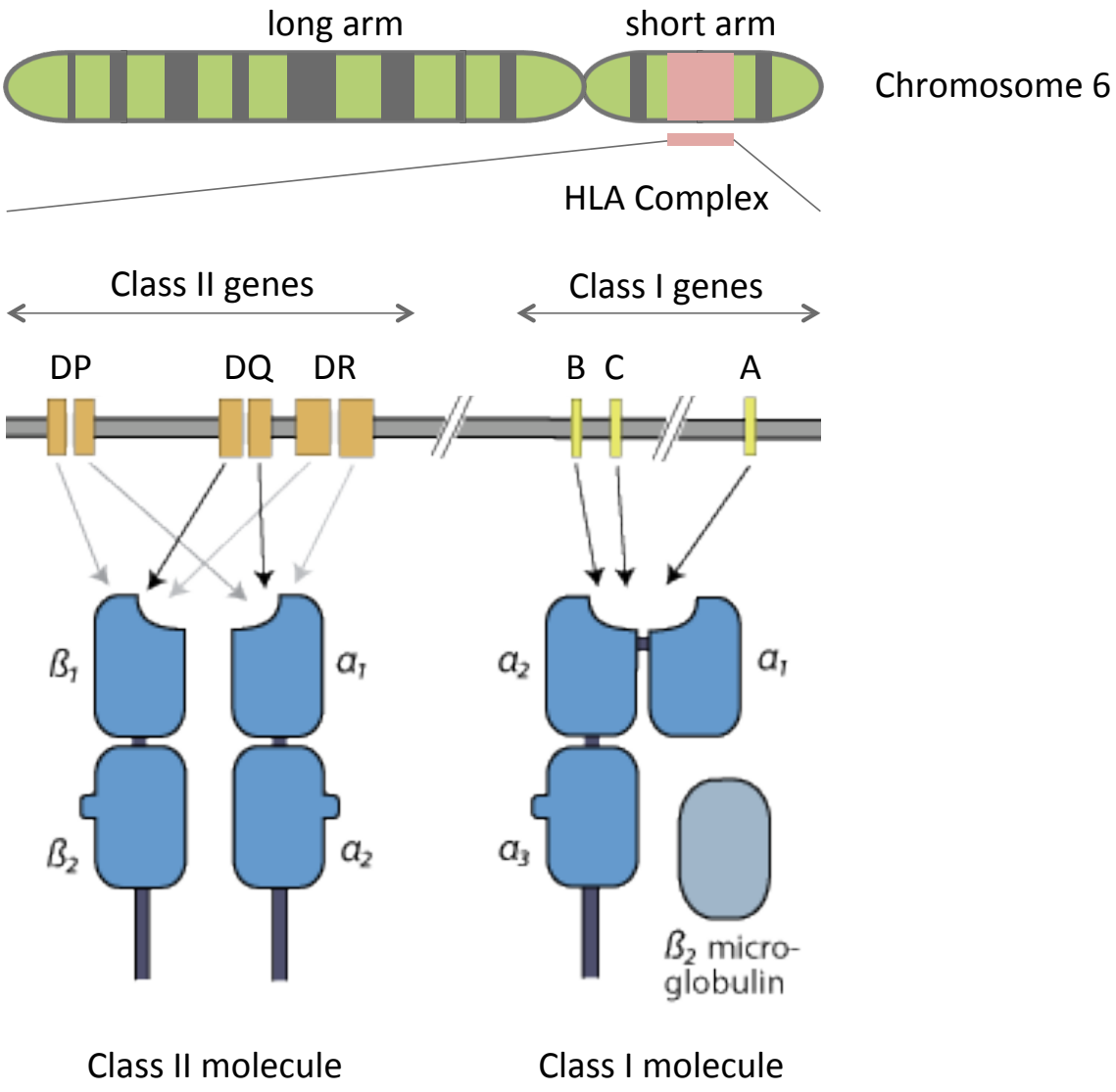

Figure 1. A schematic representation of HLA class I and II molecules and the HLA genes associated with them.

(Image is adapted from Faculte de medicine de Geneve (Faculte de medicine de Geneve, 2015))

1.3.2 Designation of HLA alleles, allele frequencies and haplotypes

HLA alleles are variants of HLA genes found at the same HLA loci. Each individual, therefore, has two HLA alleles for each HLA gene. Each HLA allele is described by a unique number comprising four sets of digits separated by colons. The digits before the first colon signify the antigen type that mostly corresponds to the serological antigen. The second set of digits list subtypes that differ by nucleotide substitutions that translate into amino acid sequence differences. Alleles that differ by nucleotide substitutions that do not change the amino acid sequence are designated by the third set of digits. The last set of digits distinguish alleles

differing by nucleotide substitutions in the introns or untranslated regions of the gene concerned (Marsh *et al.*, 2010).

"Allele frequency" refers to the relative frequency of an allele at a particular locus in a particular population. It is therefore defined as the number of alleles present within a population of "n" individuals, divided by the total number of alleles within that population (2n). It can be expressed as a fraction or as a percentage. A "haplotype" is a combination of alleles that tend to be inherited together on the same chromosome. For the purposes of this study two- and three-HLA class II locus haplotypes were considered.

1.3.3 The role of CD8+ T cells in HIV-1 infection

HIV-specific CD8+ T cells are believed to be the critical component of the initial protective immune response to HIV-1. During acute HIV-1 infection, the resolution of peak viraemia is associated with the appearance of the HIV-specific CD8+ T cell response (Koup *et al.*, 1994). Furthermore, in animal models, CD8+ T cell depletion has been reliably linked to loss of viral control and disease progression (Schmitz *et al.*, 1999). Possibly the strongest evidence of the impact of the CD8+ T cell response on HIV replication, is the rapid development of viral escape mutations within epitopes specifically targeted by CD8+ T cells, as initially demonstrated by Phillips *et al.* (Phillips *et al.*, 1991). CD8+ T cells mediate this viral inhibition by directly inhibiting viral replication and spread through the lysis of HIV-infected cells as well as through the secretion of soluble factors that inhibit viral transcription and proviral gene expression (McBrien *et al.*, 2018).

1.3.4 HLA class I associations with HIV-1 disease progression

HLA class I molecules are responsible for the ability of CD8+ T cells to recognise and kill virally-infected cells (Zinkernagel and Doherty, 1974). Cells display host- or pathogen-derived processed proteins in the form of peptide antigens bound to HLA class I molecules on their cell surface. CD8+ T cells with antigen-specific TCRs recognise these peptides bound to HLA class I molecules and are able to lyse the cells or secrete soluble factors.

The only clear genetic associations with HIV-1 virological control revealed through Genome-wide association studies (GWAS) are single-nucleotide polymorphisms (SNPs) within the HLA class I region linked to *HLA-B*57* and *HLA-C* expression (Fellay *et al.*, 2007; Pereyra *et al.*,

2010). Studies of ECs reveal strong associations with certain HLA Class I alleles such as *HLA-B*57* and *-B*27* alleles in Caucasians and *HLA-B*58:01* and *-B*81:01* in black southern Africans (Carrington and O'Brien, 2003; Klepiela *et al.*, 2004).

The association of HLA class I alleles with differential HIV-1 disease progression is thought to be due to specific structural properties of the HLA class I complex as well as viral epitope-specific factors. Favourable HLA class I alleles are associated with specific characteristics affecting the peptide-binding groove as well as well as factors enhancing CD8+ T cell function (Goulder and Walker, 2012). In individuals with protective HLA class I alleles, CD8+ T cells have been shown to have increased proliferative capacity, effector function and broader TCR epitope reactivity (Horton *et al.*, 2006; Hersperger *et al.*, 2011; Chen *et al.*, 2012). Immune control of HIV-1 infection has also been linked to the targeting of particular dominant viral epitopes associated with protective HLA class I alleles. Most of these epitopes are located in the HIV-1 Gag protein resulting in escape mutations with significant viral replicative costs (Kiepiela *et al.*, 2007; Crawford *et al.*, 2009).

1.3.5 The role of CD4+ T cells in HIV-1 infection

The role of CD4+ T cells in the control of other chronic viral infections is well established, however the role of HIV-specific CD4+ T cells in the control of HIV-1 infection is poorly defined (Matloubian *et al.*, 1994). Current evidence suggests the potential for CD4+ T cells to control HIV-1 viraemia through providing activation signals ("help") for CD8+ T cells and B cells, or via direct antiviral effects including cytolysis (Norris *et al.*, 2004; Soghoian and Streeck, 2010; Chevalier *et al.*, 2011). HIV-specific CD4+ T cells have been shown to correlate with viral control (Rosenberg *et al.*, 1997). A recent study attributed the maintenance of a population of cytolytic CD57+ CD4+ T cells in ECs to potentially playing an important role in the control of HIV viraemia in these individuals (Phetsouphanh *et al.*, 2019). Studies in acute and chronic HIV-1 infection have associated robust virus-specific CD4+ T cell responses with lower viral load set points and delayed disease progression (Ramduth *et al.*, 2009; Soghoian *et al.*, 2012). Further studies have identified HIV-specific CD4+ T cell responses specifically targeted to Gag to be associated with viral control (Ranasinghe *et al.*, 2012; Laher *et al.*, 2017). Laher *et al.* also showed that these Gag-specific CD4+ T cells in HIV-1 controllers showed higher levels of cytolytic proteins granzyme A and B than in HIV-1 progressors,

thereby suggesting a significant role for CD4+ T cell-mediated lysis in controlling viraemia (Laher *et al.*, 2017).

The RV144 HIV-1 vaccine trial showed a very modest protective effect associated with vaccine-induced Env-specific CD4+ T cells and neutralising antibodies, thereby raising the possibility that the CD4+ T cell response contributed to the modest vaccine effect that was observed (de Souza *et al.*, 2012). Further evidence of the role that CD4+ T cells may play in controlling HIV-1 infection was suggested by the simian immunodeficiency virus (SIV) model where a CD4+ T cell-driven escape mutation led to loss of viral control in an elite controller rhesus macaque (Burwitz *et al.*, 2012). This is also supported by indirect evidence in a human study that used computational methods to demonstrate CD4+ T cell-driven viral escape in longitudinal samples of chronically infected individuals following acute HIV-1 infection (Erdmann *et al.*, 2015).

1.3.6 HLA class II associations with HIV-1 disease progression

HLA class II molecules are required for presentation of peptide antigens to CD4+ T cells and are therefore necessary for the generation of HIV-specific CD4+ T cell responses. Considering the evidence for a significant role of HIV-specific CD4+ T cells in the control of viraemia, it is plausible that HLA class II alleles may exert differential immunomodulatory effects on HIV-1 disease progression.

An early study into the influence of HLA class II alleles in a Caucasian paediatric cohort that was vertically infected with HIV-1, showed an association with several *HLA-DRB1*13* alleles and long-term survival (Chen *et al.*, 1997). This finding was supported by another study that aggregated findings from three cohorts of Caucasian men who have sex with men (MSM) to demonstrate the consistent associations of HLA class II alleles on rates of progression of HIV-1 disease (Keet *et al.*, 2002). A small study in individuals treated with ART in acute HIV-1 infection showed a link between the HLA class II *DRB1*13/-DQB1*06* haplotype and HIV-1 suppression after ART interruption. In addition lymphocytes from the individuals with the *HLA-DRB1*13/-DQB1*06* haplotype had higher mean lympho-proliferation and interferon gamma (IFNγ) secretion in response to HIV-1 p24 antigen stimulation than individuals with other haplotypes (Malhotra *et al.*, 2001). *HLA-DRB1*13* and *-DQB1*06* (alone and/or together) were also shown to be enriched in ECs and to be associated with polyfunctional

HIV-specific CD4+ T cell responses in peripheral blood and rectal mucosa (Ferre *et al.*, 2010). In an earlier study of the same cohort, *HLA-DRB1*13* and *-DQB1*06*, when present in conjunction with protective HLA class I alleles (*HLA-B*13, -B*27, -B*57, -B*58,* and *–B*81*), were associated with a greater magnitude of mucosal HIV-specific CD8+ T cell responses than individuals with protective HLA class I alleles alone (Ferre *et al.*, 2009). This suggests that HLA class II alleles may also play a role in HIV-1 viral control via the augmentation of CD8+ T cell responses.

Other studies, also in Caucasian populations, have shown associations with HLA class II allele frequency and rates of disease progression. Vyakarnum *et al.* demonstrated that *HLA-DQB1*06* was represented at a higher frequency amongst a group of long-term non-progressors (LTNPs) when compared to a group of intermediate progressors (Vyakarnam *et al.*, 2004). Ranasinghe *et al.* studied a large cohort of Caucasian Americans chronically infected with HIV-1 and found *HLA-DRB1*15:02* to be significantly associated with low viraemia and *HLA-DRB1*03:01* to be significantly associated with high viraemia. This study demonstrated that *HLA-DRB1* allele variants that were associated with low viraemia were able to present a larger breadth of peptides with lower functional avidity compared to the *HLA-DRB1* allele variants associated with high viraemia (Ranasinghe *et al.*, 2013).

Whilst most studies investigating HLA class II associations with HIV-1 disease progression, have been carried out in Caucasian paediatric or MSM populations, predominantly infected with HIV-1 subtype B virus, two studies have been performed in southern Africa where the HIV-1 epidemic is most severe and is characterised mainly by HIV-1 subtype C infection and heterosexual routes of transmission. Ndung'u *et al.* showed that the predominant HLA class II alleles prevalent in black Africans in Botswana differed significantly from the most frequent alleles in Caucasian populations of Europe and North America. They also found an association amongst HIV-1 infected individuals expressing *HLA-DRB1*08* with lower viral loads as well as a higher frequency of expression of *HLA-DRB1*01* amongst HIV-uninfected individuals within the same community (Ndung'u *et al.*, 2005).

Amongst black African individuals in KwaZulu-Natal who were chronically infected with subtype C HIV-1, the HLA class II allele *DRB1*13:03* was independently associated with lower viral loads (Julg *et al.*, 2011). Interestingly, in the same paper, this finding was confirmed in a separate North American cohort of HIV-1 Clade B-infected Caucasian males, demonstrating

that the *HLA-DRB1*13:03* mediated protection was independent of ethnicity, sex and viral subtype (Julg *et al.*, 2011).

1.3.7 HLA class I and class II linkage disequilibrium

HLA linkage disequilibrium (LD) is the non-random association of alleles at two or more HLA loci (Single and Thomson, 2016). Considering the well-established associations between certain HLA class I alleles and HIV-1 control, the potential exists for confounding of HLA class II allele associations with HIV-1 control if high levels of LD exist between class I and class II loci. As an example, this is explored by Julg et al. who showed that the "protective" HLA class II allele *HLA-DRB1*13:03* was in LD with the HLA class I allele *HLA-B*57:03* that has previously been shown to be associated with lower HIV-1 viral loads (Julg *et al.*, 2011). It is therefore prudent to consider potential LD when assessing significant associations between HLA class II alleles and HIV-1 control. However, for the purposes of this particular analysis, HLA class I data were not available

1.3.8 Summary

Considering the increasing evidence for the important role of HIV-specific CD4+ T cells in HIV-1 control, HLA class II alleles are likely to play a significant role in HIV-1 disease pathogenesis. Studies in highly selective and mostly Caucasian populations suggest a potential role for particular HLA class II alleles in HIV-1 control, however, these associations remain to be confirmed in a cohort of black African patients with subtype C HIV-1 infection.

2 Materials and Methods

2.1 Study design

This is a cross-sectional study using genomic DNA samples obtained from cohorts of HIV-1 Controllers and Progressors recruited in Johannesburg.

2.2 Study participants

This study included HIV-1 infected individuals who were ART-naïve at the time of enrolment. All study participants were recruited in Johannesburg and were "black African" as self-reported in terms of South Africa's four major population groups (black African, white (Caucasian), Indian/Asian and coloured (mixed ancestry))(Statistics South Africa, 2019). The "black African" population in South Africa is "a mixture of Bantu-originating ethnic and linguistic groups" described previously (Paximadis *et al.*, 2012).

Genomic DNA samples from 26 HIV-1 Controllers and 74 Progressors, were selected for this study. The HIV-1 Controllers were recruited as part of a greater study of HIV-1 ECs, LTNPs and progressors (Principal Investigator: Prof. C. T. Tiemessen). HIV-1 Controllers suppress HIV-1 viral replication to <2000 RNA copies/mL whilst ECs suppress viral replication to <50 RNA copies/mL in the absence of ART.

Progressors were recruited as part of the Longitudinal Study of HIV-Associated Lung Infections in Soweto (Martinson *et al.*, 2014). Progressors were defined as individuals with CD4+ T cell counts below 350 cells/µL and viral loads above 10000 RNA copies/mL. HLA class II *DPB1*, *-DQB1* and *-DRB1* genotyping data were available for the progressors (courtesy of Prof. C.T. Tiemessen) as a comparison group.

All study participants provided written informed consent and were at least 18 years of age. Ethical clearance for this study was obtained from the University of the Witwatersrand Human Research Ethics Committee – Clearance Certificate no. M190995 (PI: Prof. C. T. Tiemessen) (Appendix 1).

2.3 Funding sources

This study was funded by grant awards from the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council (a grantee of the Bill and Melinda Gates Foundation), and the South African Research Chairs Initiative of the Department of Science and Innovation and National Research Foundation of South Africa (Prof. C.T. Tiemessen). The Longitudinal Study of HIV-Associated Lung Infections in Soweto was funded by the National Institutes of Health, USA (R01HL090312 and P30AI094189: R.E. Chaisson).

2.4 HLA class II genotyping

High-resolution (6-digit allele level) HLA class II genotyping of *DPB1*, *-DQB1* and *-DRB1* loci was carried out on DNA samples from the Controllers using sequence-based typing (SBT) protocols.

2.4.1 SBT method

1. Genomic DNA was extracted from stored buffy coat samples using the QIAamp DNA blood mini kit (Qiagen, Hilden, Germany).
2. Genomic DNA was quantified using a NanoDrop Spectrophotometer (Thermo Scientific, Waltham, USA).
3. HLA class II *DPB1*, *-DQB1* and *-DRB1* genes were amplified from extracted DNA by polymerase chain reaction (PCR) using locus-specific primers from the SBT Resolver kit (Conexio Genomics, Fremantle, Australia). A representative agarose gel depicting PCR amplification products at different HLA loci is shown in Figure 2.
4. The Agencourt Ampure XP magnetic bead separation method was used to purify the PCR products (Becton-Dickinson, Franklin Lakes, USA).
5. The BigDye terminator v3.1 cycle sequencing kit was used to generate sequence products (Life Technologies, Carlsbad, USA).
6. Sequence products were purified using ethanol-sodium acetate precipitation.
7. Sequence products were resolved on an ABI3100 PRISM Genetic Analyser (Life Technologies, Carlsbad, USA).
8. Sequence results were interpreted using the ASSIGN-SBT software (version 4.7; Conexio Genomics, Fremantle, Australia) and version 3.33 of the IPD-IMGT/HLA

reference database released in July 2018, based on sequence data for exons 1-5 for HLA-DPB1 and exons 2/3 for HLA-DQB1 and -DRB1 (Figure 3).

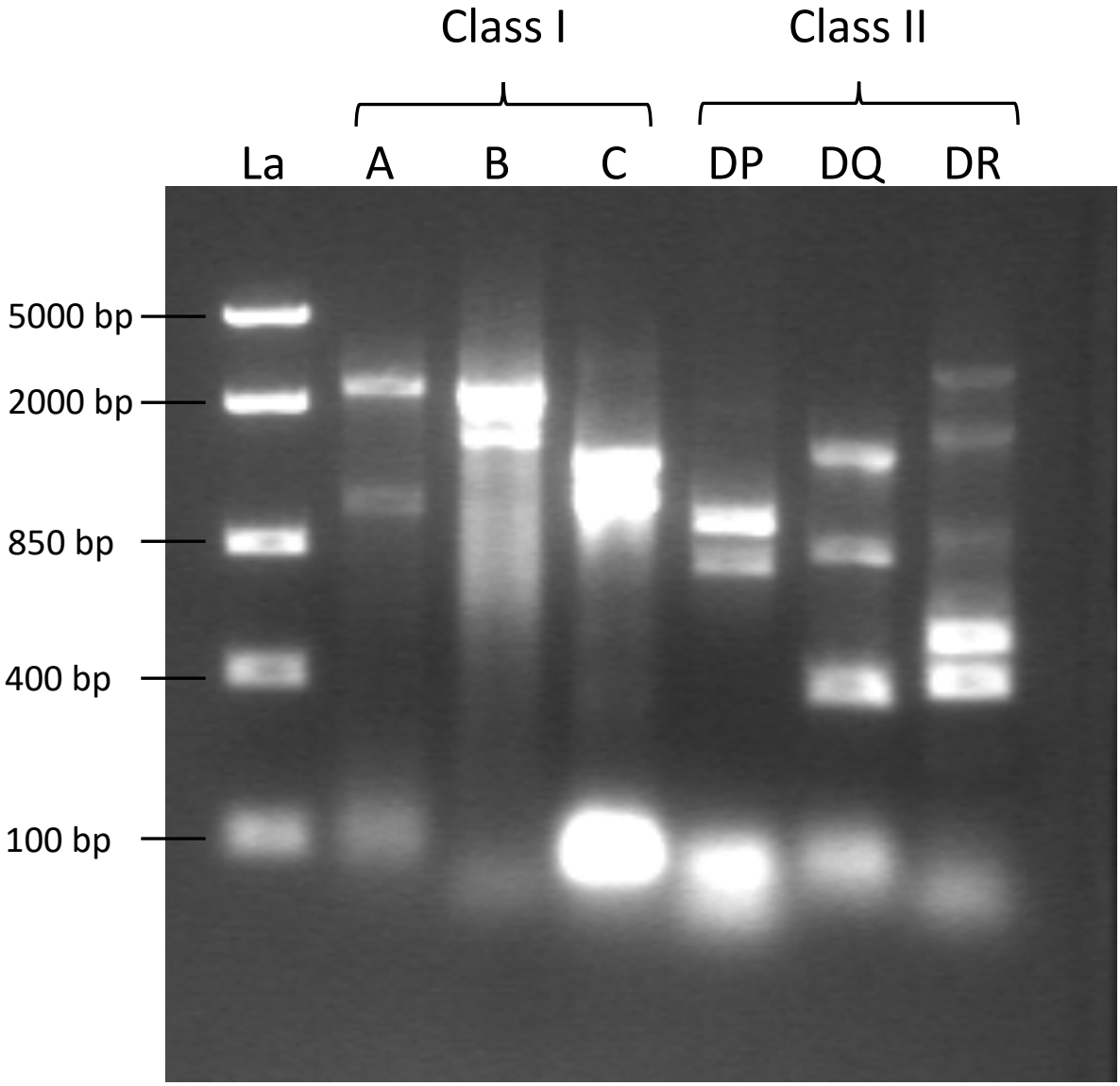

Figure 2. A representative agarose gel showing PCR amplification products at different HLA loci. La = DNA ladder (FastRuler Middle Range DNA ladder - Thermo Scientific, Waltham, USA), bp = base pairs.

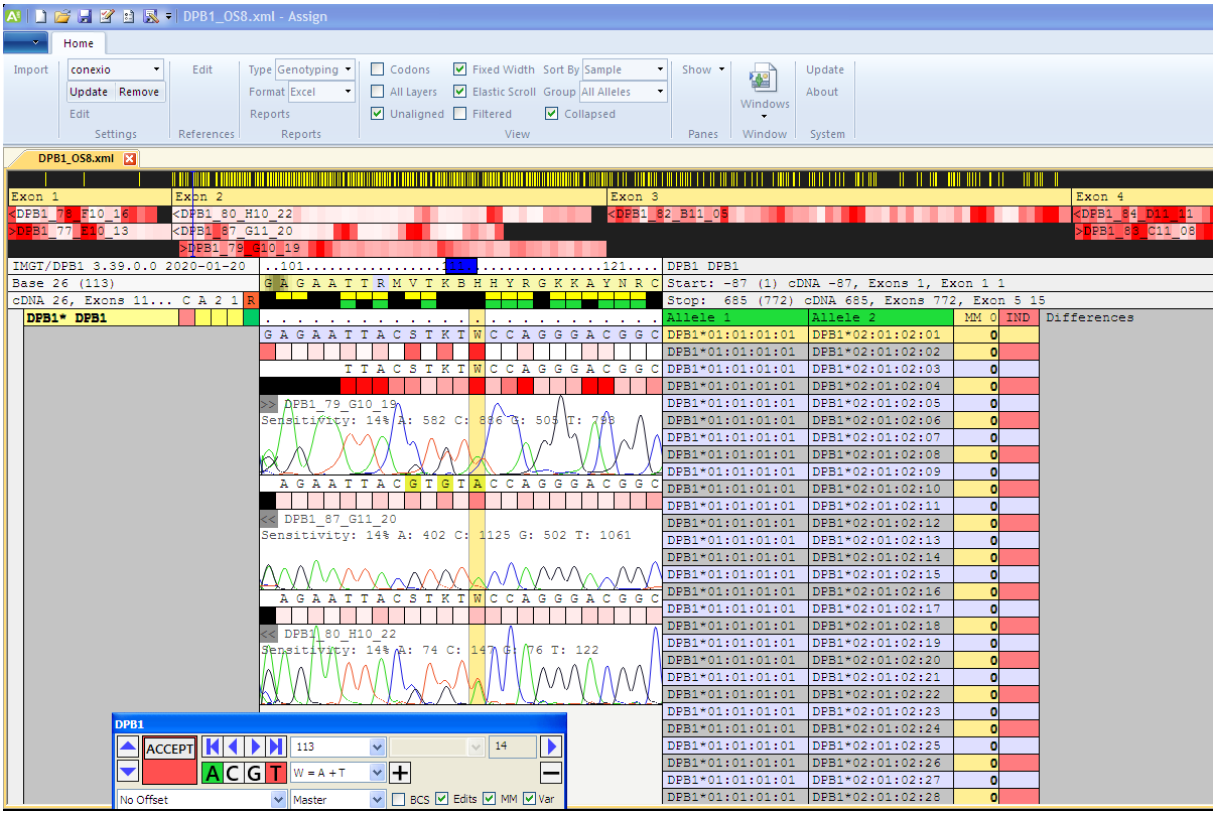

Figure 3. A screenshot image of a representative *HLA-DPB1* allele analysis taken from the ASSIGN-SBT software program version 4.7. The image shows likely allele pairs based on single nucleotide sequence differences (highlighted in the yellow column).

2.4.2 Resolving of ambiguous typing

Due to the fact that the polymorphic region, in some cases, lay outside of the region sequenced and the general heterogenous nature of such SBT analyses, the typing occasionally yielded an ambiguous result. As no further molecular-based methods were used to resolve possible ambiguities, a structured, stepwise, process was followed to exclude unlikely HLA allele combinations:

1. The ASSIGN-SBT version 4.7 software (Conexio Genomics) and version 3.33 of the IPD-IMGT/HLA reference database (July 2018) were used to interrogate SNPs in exons 1-5 for *HLA-DPB1* and exons 2 and 3 for *HLA-DQB1* and *-DRB1*.

2. Other published studies that investigated similar black African populations were consulted to further refine likely allele pairs (Thorstenson *et al.*, 2018; Chambuso *et al.*, 2019; Loubser *et al.*, 2020).

3. The Allele Frequency Net Database (AFND) was then consulted to review HLA class II allele frequencies that have been determined previously in similar populations (www.allelefrequencies.net) (González-Galarza *et al.*, 2015).

4. Finally, likely allele combinations were compared to "Common" and "Well-Documented" alleles listed in the "Common and Well-Documented (CWD) HLA alleles catalogue" version 2.0.0, updated in 2012 (Mack *et al.*, 2013).

The process followed for the resolving of ambiguous typings is summarised in Figure 4.

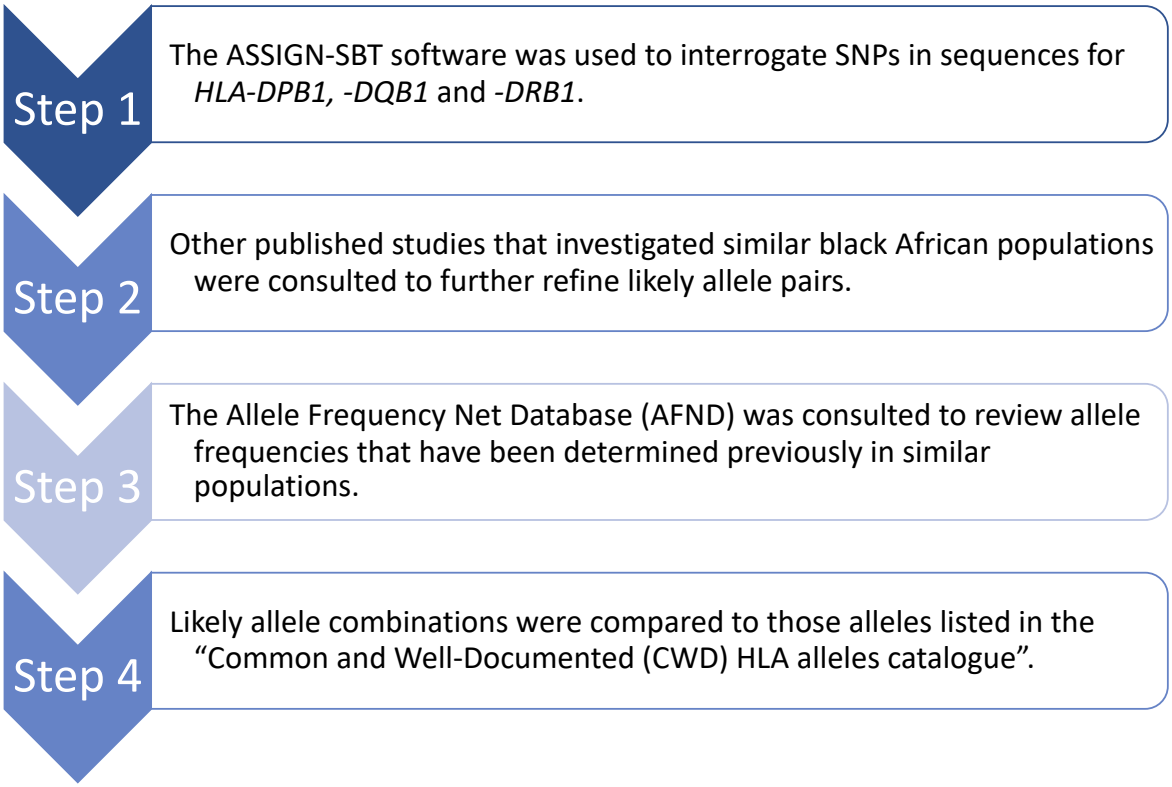

Figure 4. A flow diagram summarising the process undertaken to resolve allele ambiguities.

2.5 Data analysis

Analyses were carried out using GraphPad Prism 5 software (for Mac OS X version 5.0a) (Graphpad Software, San Diego, CA, USA). Frequencies of the alleles genotyped for all the cohort individuals were determined by direct counting. To compare representation of the individual alleles between Controllers and Progressors, Fisher's Exact tests were used, and odds ratios (OR) and 95% confidence intervals (CI) determined. Two-sided tests were used and the level of statistical significance for all analyses was set at $P< 0.05$. Correction for multiple comparisons was not performed due to the exploratory nature of this analysis and the presence of multiple alleles per locus tested.

Haplotype frequencies and calculations for deviations from Hardy-Weinberg equilibrium were performed using PyPop-Win32-0.7.0 software (Lancaster *et al.*, 2007). Comparisons of haplotype frequencies between Controllers and Progressors were carried out using Fisher's Exact tests as described above.

3 Results

3.1 Cohort characteristics

The study cohort included a total of 100 ART-naïve HIV-infected black South African individuals comprising 26 Controllers and 74 Progressors. Of the 26 Controllers, 17 individuals were classified as ECs with HIV-1 viral loads <50 RNA copies/mL.

The majority of participants in the total cohort were female (81%) (Table 1). The gender proportions between the groups of Controllers and Progressors were not significantly different (% female: Controllers 73% vs Progressors 84%, P=0.253, Table 1). The median age of participants within the total cohort was 38 years and the median ages did not significantly differ between Controllers and Progressors (Median age: Controllers 39 years vs Progressors 38 years, P=0.320, Table 1).

Median CD4+ T cell counts and HIV-1 viral loads differed significantly between the two groups, as would be expected due to the clear selection criteria used to define the different groups and as explained in the Methods section 2.2 (Controllers vs Progressors; CD4+ T cell counts and HIV-1 viral load RNA copies/mL: P <0.001 for both, Table 1). Detailed cohort characteristics are shown in Table 1.

Table 1. Characteristics of study cohort

Characteristic	HIV-1 Controllers	HIV-1 Progressors	P value	Total Cohort
	n= 26	n= 74		n= 100
Gender				
Female - n (%)	19 (73)	62 (84)	0.253[b]	81 (81)
Male - n (%)	7 (27)	12 (16)		19 (19)
Age (years) [a]	39 (33 – 48)	38 (34 – 41)	0.320[c]	38 (34 – 42)
CD4 count (cells/µL) [a]	634 (548 – 827)	176 (146 – 211)	< 0.001[c]	NA
Viral load (RNA copies/mL) [a]	20 (20 – 181)	38444 (19709 – 103042)	< 0.001[c]	NA

[a] Data expressed as median (interquartile range). NA = not applicable. [b] P value is shown for Fisher's Exact test. [c] P value is shown for Mann-Whitney test.

3.2 HLA class II allele frequencies in total cohort

We characterised a total of 600 HLA class II alleles (200 alleles each for *HLA-DPB1*, *-DQB1* and *-DRB1* loci) including 156 alleles from Controllers and 444 alleles from Progressors.

3.2.1 Hardy-Weinberg equilibrium

Analyses for Hardy-Weinberg (HW) equilibrium in both Controllers and Progressors revealed no deviations from HW proportions for *HLA-DPB1* (Controllers: P=0.85 ; Progressors: P=0.08)*, -DQB1*(Controllers: P=0.45 ; Progressors: P=0.79) or –*DRB1* (Controllers: P=0.59 ; Progressors: P=0.43).

3.2.2 Allele frequencies

A total of 61 unique HLA class II alleles were identified including 28 *HLA-DPB1* alleles, 12 *-DQB1* alleles and 21 *-DRB1* alleles. The most frequent alleles at each locus within the total cohort were *HLA-DPB1*01:01:01* (34.5%), *-DQB1*06:02:01* (17.5%) and *-DRB1*13:01:01* (15.5%). All of the alleles found in the total cohort were listed in the Common and Well-Documented (CWD) alleles catalogue, on the AFND (www.allelefrequencies.net) or in previous studies of similar populations in South Africa (Mack *et al.*, 2013; González-Galarza *et al.*, 2015; Thorstenson *et al.*, 2018; Chambuso *et al.*, 2019; Loubser *et al.*, 2020). The alleles and their frequencies within the total cohort are shown in Figure 5.

3.2.2.1 HLA-DPB1 locus

HLA-DPB1 was the most polymorphic of the HLA class II alleles determined in this study with 28 alleles detected at this locus - more than twice the number of alleles determined for *HLA-DQB1*. However, only two alleles were detected at a frequency greater than 10% in the total cohort and 15 alleles at low frequencies of 1% or less (Figure 5). The most common *HLA-DPB1* alleles included *-DPB1*01:01:01* and *-DPB1* 105:01:01* at frequencies of 34.5% and 18%, respectively, followed by *HLA-DPB1* 02:01:02* with a frequency of 7.5%. Thirteen individuals were homozygous for *HLA-DPB1*01:01:01*, and one individual each for *HLA-DPB1*02:01:02*, *-DPB1*02:01:02* and *-DPB1*18:01*.

3.2.2.2 HLA-DQB1 locus

HLA-DQB1 was the least polymorphic locus with only 12 unique alleles detected in the total cohort at this locus. Five alleles were present at frequencies greater than 10% with *HLA-DQB1*06:02:01*, *-DQB1*03:19:01* and *-DQB1*04:02:01* the most predominant of those at frequencies of 17.5%, 13.5% and 12% respectively (Figure 5). In contrast to *HLA-DPB1*, only a single *HLA-DQB1* allele (*HLA-DQB1*06:04:01*) was present at a frequency of 1% or less. Five individuals were homozygous for *HLA-DQB1*06:02:01* whilst two were homozygous for *-DQB1*06:09:01*. A single individual was homozygous for *HLA-DQB1*03:19:01*, *-DQB1*04:02:01* and *-DQB1*06:03:01*.

3.2.2.3 HLA-DRB1 locus

Twenty-one *HLA-DRB1* alleles were detected in the total cohort. The most frequent alleles were *HLA-DRB1*13:01:01*, *-DRB1*11:01:02* and *DRB1*07:01:01* with frequencies of 15.5%, 12.5% and 10% respectively (Figure 5). The two most frequent *HLA-DRB1* alleles were consistent with the two most frequent alleles found in a study of black African HIV-uninfected individuals with similar geographic and ethnic backgrounds (Paximadis *et al.*, 2012).

Three alleles were present at frequencies of 10% or greater and three alleles at frequencies of 1% or less. Four individuals in the total cohort were homozygous for *HLA-DRB1*13:01:01* and one each for *-DRB1*03:02:01, -DRB1*11:01:02, -DRB1*13:02:01* and *-DRB1*15:03:01* alleles.

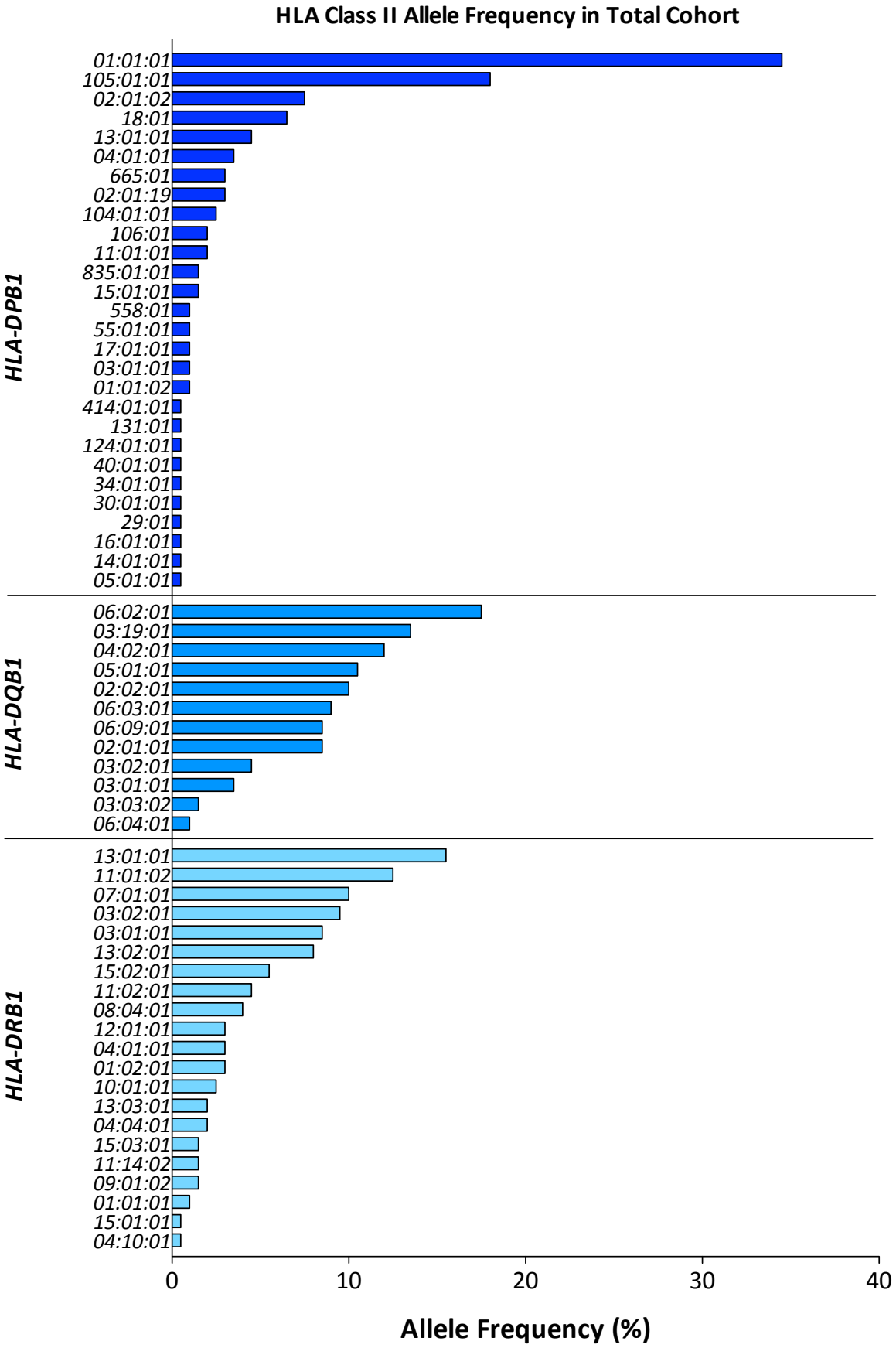

Figure 5. Frequencies of HLA class II *-DPB1, -DQB1* and *-DRB1* alleles in the total cohort (2n=200).

3.3 Comparison of HLA class II allele frequencies between HIV Controllers and Progressors

HLA class II *DPB1, -DQB1* and *-DRB1* allele frequencies were compared between the Controllers and Progressors in order to determine allelic associations with HIV-1 control or with progression of HIV-1 disease. The Progressor group was more polymorphic in general with 56 unique alleles across the three HLA class II loci compared to 44 unique alleles within the Controller group. This should, however, be seen in the context of the greater number of study participants in the Progressor group. Of the 61 unique HLA Class II *DPB1, -DQB1* and *-DRB1* alleles found in the total cohort, 15 were represented exclusively in the Progressor group and five represented exclusively in the Controller group. Forty-one alleles were represented in both groups.

3.3.1 HLA-DPB1 locus

Twenty-eight unique *HLA-DPB1* alleles were determined in the total cohort with 12 alleles exclusively found amongst Progressors and four exclusively amongst Controllers (Figure 6). The frequencies of all *HLA-DPB1* alleles amongst Controllers and Progressors are shown in Table 2 and Figure 6. *HLA-DPB1*01:01:01, -DPB1*105:01:01* and *-DPB1*02:01:02* were the predominant alleles in the Progressors with frequencies of 32.4%, 19.6% and 8.1% respectively. By contrast, although *HLA-DRB1*01:01:01* and *-DPB1*105:01:01* were also the most common alleles in the Controller group with frequencies of 40.4% and 13.5% respectively, *HLA-DPB1*18:01* was the third most predominant allele in that group with a frequency of 9.6%.

Although the frequencies of *HLA-DPB1*01:01:01* (Controllers 40.4% vs Progressors 32.4%) and *-DPB1*18:01* (Controllers 9.6% vs Progressors 5.4%, Table 2, Figure 6) alleles were higher in the Controller group and the frequency of *DPB1*105:01:01* lower (Controllers 13.5% vs Progressors 19.6%, Table 2, Figure 6) than in the Progressor group, these differences were not significant ($P>0.05$ for all, Table 2). Overall there were no significant differences in the frequencies of any single *HLA-DPB1* allele between the groups (Table 2). *HLA-DPB1*01:01:01* was homozygous in nine Progressors compared to four Controllers, however, this difference was also not significant ($P = 0.737$).

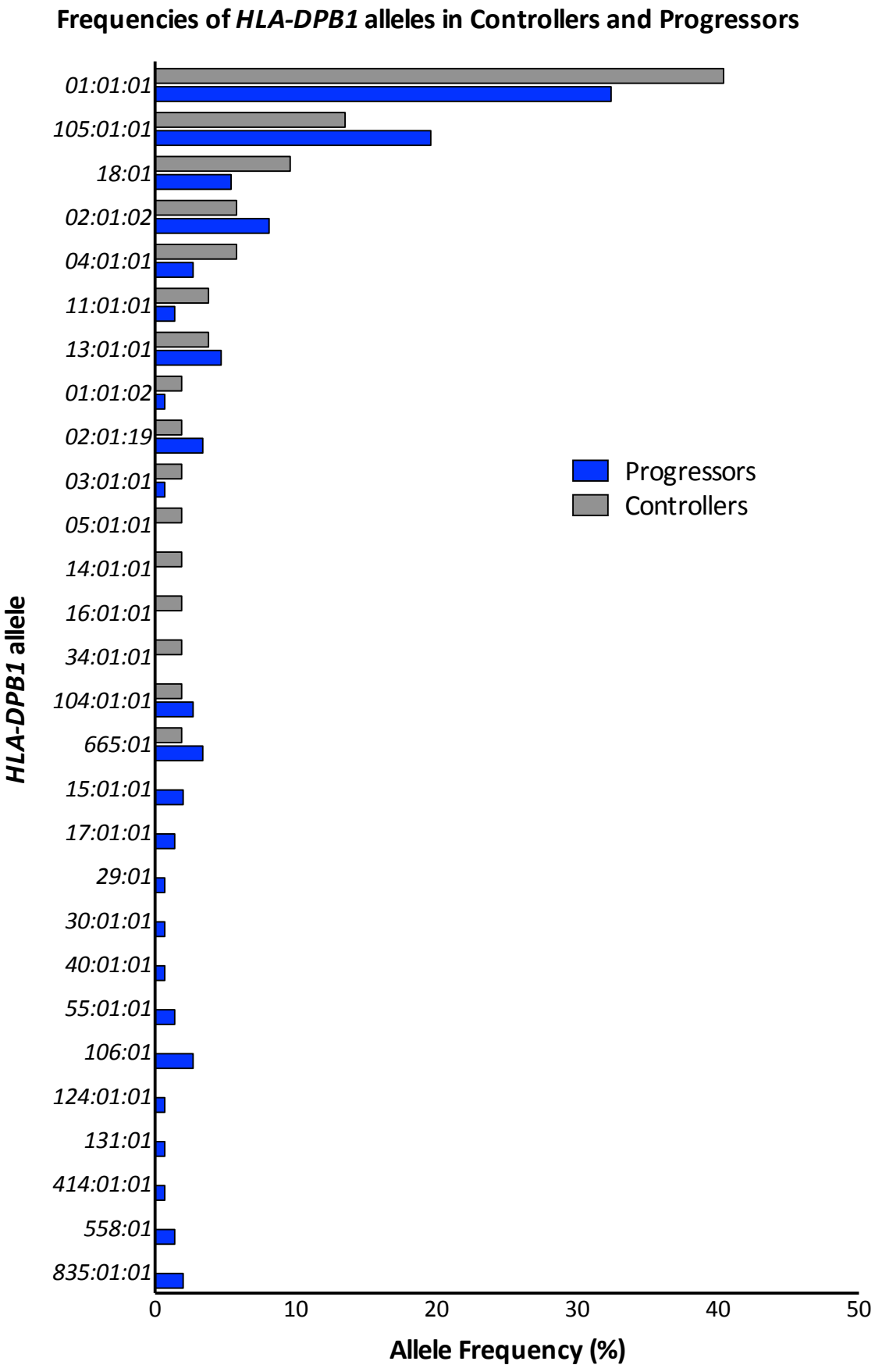

Figure 6. Frequencies of HLA class II *DPB1* alleles in Controllers (2n=52) and Progressors (2n=148). Significant results (P<0.05) are highlighted on the figure.

Table 2. Frequencies of *HLA-DPB1* alleles in Controllers and Progressors.

HLA-DPB1	**HIV Controllers (2n=52)**		**Progressors (2n=148)**			
	Allele frequency %	(n)	Allele frequency %	(n)	OR (95%CI)	P
01:01:01	40.4	21	32.4	48	0.71 (0.37 – 1.4)	0.313
01:01:02	1.9	1	0.7	1	0.35 (0.02 – 5.7)	0.453
02:01:02	5.8	3	8.1	12	1.40 (0.39 – 5.3)	0.734
02:01:19	1.9	1	3.4	5	1.80 (0.20 – 16)	1.000
03:01:01	1.9	1	0.7	1	0.35 (0.02 – 5.7)	0.453
04:01:01	5.8	3	2.7	4	0.45 (0.10 – 2.1)	0.379
05:01:01	1.9	1	0.0	0	0.12 (0.01 – 2.9)	0.260
11:01:01	3.8	2	1.4	2	0.34 (0.05 – 2.5)	0.278
13:01:01	3.8	2	4.7	7	1.20 (0.25 – 6.2)	1.000
14:01:01	1.9	1	0.0	0	0.12 (0.01 – 2.9)	0.260
15:01:01	0.0	0	2.0	3	2.50 (0.13 – 50)	0.569
16:01:01	1.9	1	0.0	0	0.12 (0.01 – 2.9)	0.260
17:01:01	0.0	0	1.4	2	1.80 (0.09 – 38)	1.000
18:01	9.6	5	5.4	8	0.54 (0.17 – 1.7)	0.329
29:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
30:01:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
34:01:01	1.9	1	0.0	0	0.12 (0.01 – 2.9)	0.260
40:01:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
55:01:01	0.0	0	1.4	2	1.80 (0.09 – 38)	1.000
104:01:01	1.9	1	2.7	4	1.40 (0.15 – 13)	1.000
105:01:01	13.5	7	19.6	29	1.60 (0.64 – 3.8)	0.404
106:01	0.0	0	2.7	4	3.30 (0.17 – 62)	0.574
124:01:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
131:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
414:01:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
558:01	0.0	0	1.4	2	1.80 (0.09 – 38)	1.000
665:01	1.9	1	3.4	5	1.80 (0.20 – 16)	1.000
835:01:01	0.0	0	2.0	3	2.50 (0.13 – 50)	0.570

P values, odds ratios (OR) and 95% confidence Intervals (95%CI) are shown for Fisher's Exact tests. Significant results (P<0.05) are highlighted in bold print.

3.3.2 HLA-DQB1 locus

All 12 unique *HLA-DQB1* alleles were present in both the Controller and Progressor groups. The frequencies of all *HLA-DQB1* alleles in both groups are shown in Table 3 and Figure 7. The most predominant alleles in the Progressors were *HLA-DQB1*06:02:01*, *-DQB1*03:19:01* and *-DQB1*05:01:01* with frequencies of 18.2%, 14.9% and 11.5%, respectively. The two most frequent alleles in the Controllers, however, were *HLA-DQB1*06:02:01* and *-DQB1*04:02:01* with identical frequencies of 15.4%. The next most predominant allele in this group was *HLA-DQB1*06:09:01* with a frequency of 13.5%.

*HLA-DQB1*02:01:01* (Controllers 3.8% vs Progressors 10.1%, Table 3, Figure 7) and *-DQB1*03:19:01* (Controllers 9.6% vs Progressors 14.9%, Table 3, Figure 7) were more than 5% more frequent amongst Progressors than Controllers, however these differences were not significant (*HLA-DQB1*02:01:01*: P=0.308; *HLA-DQB1*03:19:01*, P=0.480, Table 3). *HLA-DQB1*03:02:01* was, however, significantly more frequent amongst Controllers than Progressors suggesting that this HLA class II allele may be associated with control in the setting of chronic HIV-1 infection (Controllers 11.5% vs Progressors 2.0%, P=0.011, OR 0.16, CI 0.04 – 0.66, Table 3, Figure 7). *HLA-DQB1*06:09:01* was also more than 5% more frequent amongst Controllers, however, this difference did not reach statistical significance (Controllers 13.5% vs Progressors 6.8%, P=0.153, Table 3).

Whilst four Progressors were homozygous for *HLA-DQB1*06:02:01* only one Controller was homozygous for this allele. By contrast, two Controllers were homozygous for *HLA-DQB1*06:09:01* with none of the Progressors homozygous for the same allele. These differences were, however, not significant (P>0.05 for all).

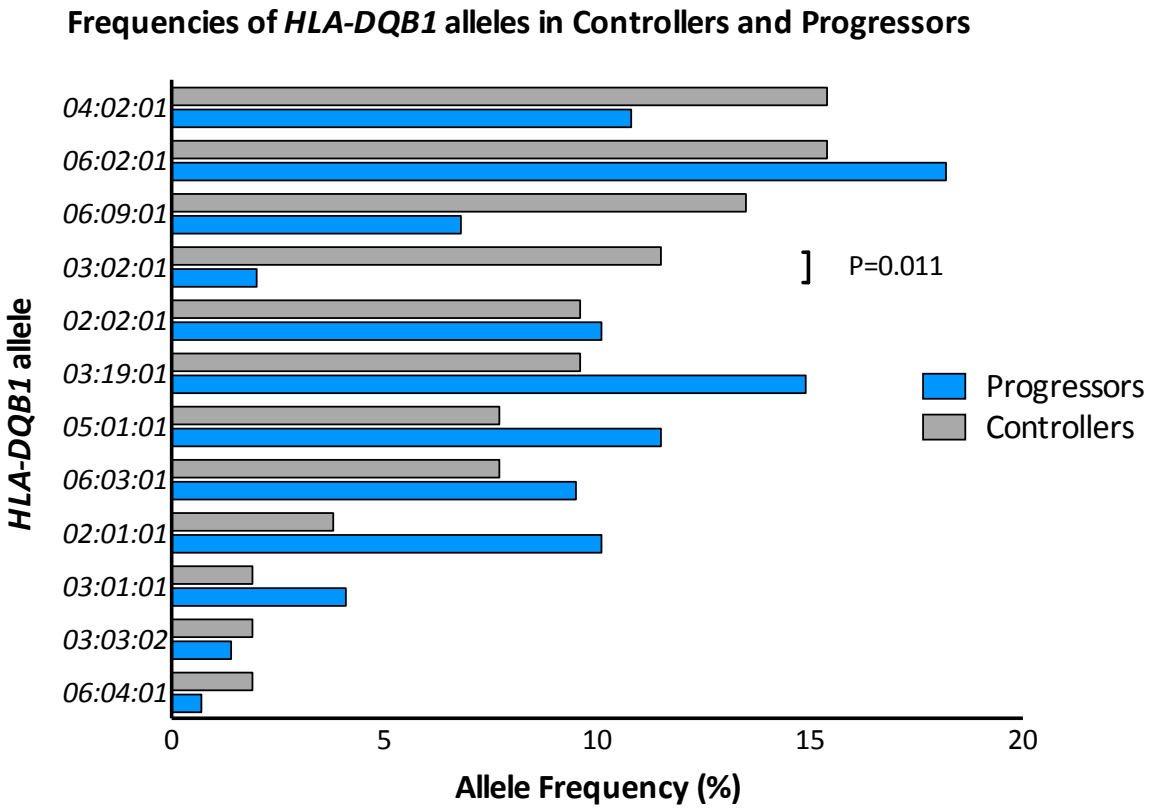

Figure 7. Frequencies of HLA class II *DQB1* alleles in Controllers (2n=52) and Progressors (2n=148). Significant results (P<0.05) are highlighted on the figure.

Table 3. Frequencies of *HLA-DQB1* alleles in Controllers and Progressors.

HLA-DQB1	**Controllers (2n=52)**		**Progressors (2n=148)**			
	Allele frequency %	(n)	Allele frequency %	(n)	OR (95%CI)	P
02:01:01	3.8	2	10.1	15	2.82 (0.62 – 12.8)	0.248
02:02:01	9.6	5	10.1	15	1.06 (0.37 – 3.08)	1.000
03:01:01	1.9	1	4.1	6	2.20 (0.25 – 18)	0.679
03:02:01	**11.5**	**6**	**2.0**	**3**	**0.16 (0.04 – 0.66)**	**0.011**
03:03:02	1.9	1	1.4	2	0.70 (0.06 – 7.9)	1.000
03:19:01	9.6	5	14.9	22	1.60 (0.59 – 4.6)	0.480
04:02:01	15.4	8	10.8	16	0.67 (0.27 – 1.7)	0.457
05:01:01	7.7	4	11.5	17	1.60 (0.50 – 4.9)	0.601
06:02:01	15.4	8	18.2	27	1.20 (0.52 – 2.9)	0.832
06:03:01	7.7	4	9.5	14	1.30 (0.39 – 4.0)	1.000
06:04:01	1.9	1	0.7	1	0.35 (0.02 – 5.7)	0.453
06:09:01	13.5	7	6.8	10	0.47 (0.17 – 1.3)	0.153

P values, odds ratios (OR) and 95% confidence Intervals (95%CI) are shown for Fisher's Exact tests. Significant results (P<0.05) are highlighted in bold print.

3.3.3 HLA-DRB1 locus

Progressors and Controllers shared 15 of the 21 unique *HLA-DRB1* alleles. Five alleles were represented exclusively within the Progressor group whilst only one allele was found exclusively in the Controller group. The frequencies of all *HLA-DRB1* alleles in Controllers and Progressors are shown in Table 4 and Figure 8. *HLA-DRB1*13:01:01, -DRB1*11:02:01* and -*DRB1*07:01:01* were the most frequent alleles in the Progressor group with frequencies of 14.2%, 14.2% and 10.1%, respectively. *HLA-DRB1*13:01:01* was also the most predominant allele in the Controller group with a higher (although not statistically significantly different) frequency of 19.2% (Controllers 19.2% vs Progressors 14.2%, P=0.381, Table 4, Figure 8). *HLA-DRB1*03:02:01* was the next most predominant allele within the Controller group at a frequency of 11.5% followed by -*DRB1 07:01:01* at a frequency of 9.6%.

A comparison of the frequencies of *HLA-DRB1* alleles between the Controllers and Progressors showed a 5% or greater frequency of allele *HLA-DRB1*11:01:02* in the Progressor group, although this did not reach statistical significance (Controllers 7.7% vs Progressors 14.2%, P=0.329, Table 4, Figure 8). *HLA-DRB1*04:01:01* was, however, significantly more represented within the Controller group suggesting a possible role for protection for this allele (Controllers 7.7% vs Progressors 1.4%, P=0.041, OR 0.16, CI 0.03 – 0.93, Table 4, Figure 8). Although *HLA-DRB1*13:01:01* was also 5% more frequent in the Controllers than in the Progressors, this difference was not statistically significant (Controllers 19.2% vs Progressors 14.2%, P=0.381, Table 4.)

Homozygosity for *HLA-DRB1*13:01:01* was present in two individuals from both the Controller and Progressor groups. Due to the low numbers of homozygous individuals, there were no significant differences in the frequencies of homozygosity for any of the *HLA-DRB1* alleles.

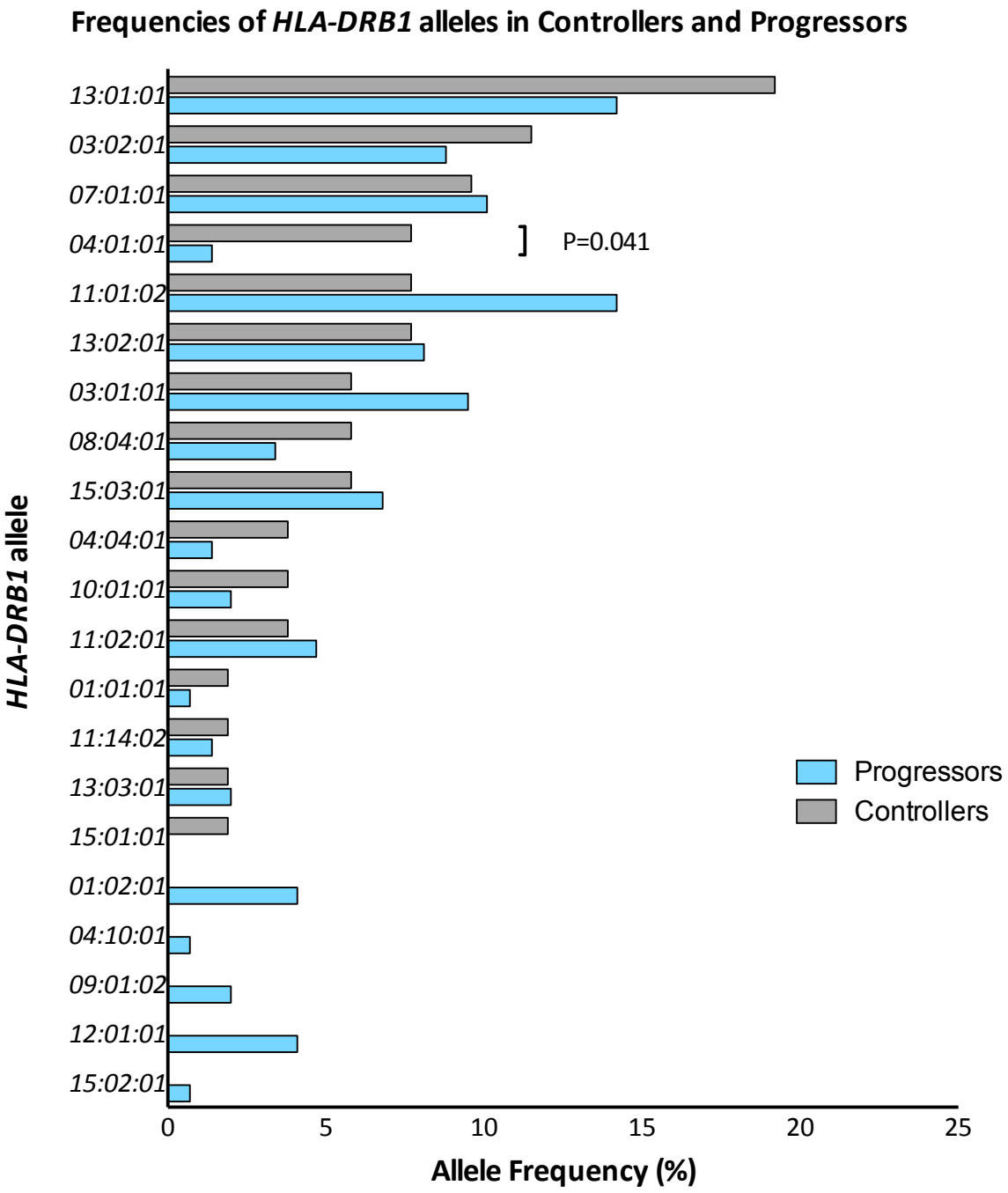

Figure 8. Frequencies of HLA class II *DRB1* alleles in Controllers (2n=52) and Progressors (2n=148). Significant results (P<0.05) are highlighted on the figure.

Table 4. Frequencies of *HLA-DRB1* alleles in Controllers and Progressors.

HLA-DRB1	**Controllers (2n=52)**		**Progressors (2n=148)**			
	Allele frequency %	(n)	Allele frequency %	(n)	OR (95%CI)	P
01:01:01	1.9	1	0.7	1	0.35 (0.02 – 5.7)	0.453
01:02:01	0.0	0	4.1	6	4.80 (0.26 – 87)	0.343
03:01:01	5.8	3	9.5	14	1.70 (0.47 – 6.2)	0.567
03:02:01	11.5	6	8.8	13	0.74 (0.27 – 2.1)	0.586
04:01:01	**7.7**	**4**	**1.4**	**2**	**0.16 (0.03 – 0.93)**	**0.041**
04:04:01	3.8	2	1.4	2	0.34 (0.05 – 2.5)	0.278
04:10:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
07:01:01	9.6	5	10.1	15	1.10 (0.37 – 3.1)	1.000
08:04:01	5.8	3	3.4	5	0.57 (0.13 – 2.5)	0.431
09:01:02	0.0	0	2.0	3	2.50 (0.13 – 50)	0.569
10:01:01	3.8	2	2.0	3	0.52 (0.08 – 3.2)	0.606
11:01:02	7.7	4	14.2	21	2.00 (0.65 – 6.1)	0.329
11:02:01	3.8	2	4.7	7	1.20 (0.25 – 6.2)	1.000
11:14:02	1.9	1	1.4	2	0.70 (0.06 – 7.9)	1.000
12:01:01	0.0	0	4.1	6	4.80 (0.26 – 87)	0.343
13:01:01	19.2	10	14.2	21	0.69 (0.30 – 1.6)	0.381
13:02:01	7.7	4	8.1	12	1.10 (0.33 – 3.4)	1.000
13:03:01	1.9	1	2.0	3	1.10 (0.11 – 10)	1.000
15:01:01	1.9	1	0.0	0	0.12 (0.01 – 2.9)	0.260
15:02:01	0.0	0	0.7	1	1.10 (0.04 – 27)	1.000
15:03:01	5.8	3	6.8	10	1.20 (0.31 – 4.5)	1.000

P values, odds ratios (OR) and 95% confidence Intervals (95%CI) are shown for Fisher's Exact tests. Significant results (P<0.05) are highlighted in bold print.

3.3.4 Two- and three-locus HLA class II haplotypes

Haplotype frequencies were estimated for two-locus and three-locus combinations of *HLA-DPB1, -DQB1* and *-DRB1* alleles using PyPop-Win32-0.7.0 software and compared between the Controller and Progressor groups (Lancaster *et al.*, 2007). Two-locus frequencies were estimated for *HLA-DPB1/DQB1* and *HLA-DQB1/DRB1* combinations and are shown in Table 5. Three-locus haplotype frequencies are shown in Table 6.

3.3.4.1 HLA-DPB1/-DQB1 haplotypes

Table 5 lists 11 *HLA-DPB1/DQB1* haplotypes at a frequency of 2% or greater amongst Controllers. By contrast, 21 *HLA-DPB1/DQB1* haplotypes were present in Progressors at frequencies of 2% or greater. *HLA-DPB1*01:01:01/DQB1*04:02:01* was the most frequent haplotype in the Controller group present at a frequency of 13.5% followed by *HLA-DPB1*01:01:01/DQB1*06:02:01* at a frequency of 9.6%. The *HLA-DPB1*01:01:01/DQB1*04:02:01* haplotype was significantly more frequent amongst Controllers than Progressors with representation in only 3.0% in Progressors (13.5% vs 3.0%, P=0.015, OR 0.22, CI 0.07 – 0.74, Table 5). The *HLA-DPB1*01:01:01/DQB1*06:02:01* haplotype was also significantly more frequent amongst controllers than progressors (9.6% vs 2.0%, P=0.029, OR 0.19, CI 0.05 – 0.85, Table 5). In the Progressor group *HLA-DPB1*105:01:01/DQB1*03:19:01* was most frequent at a frequency of 5.7% with *HLA-DPB1*01:01:01/DQB1*05:01:01* and *-DPB1*01:01:01/DQB1*06:03:01* the next most predominant haplotypes at frequencies of 5.4% and 5.3%, respectively. There were no other statistically significant differences in *HLA-DPB1/DQB1* haplotype frequencies between the two groups.

3.3.4.2 HLA-DQB1/-DRB1 haplotypes

A total of 12 unique *HLA-DQB1/DRB1* haplotypes were represented amongst Controllers, at a frequency of 2% or greater, and 14 amongst Progressors. The *HLA-DQB1*04:02:01/DRB1*03:02:01* haplotype was the most predominant in the Controller group with a frequency of 11.5% followed by four unique haplotypes with identical frequencies of 7.7% (Table 5). In Progressors, *HLA-DQB1*02:01:01/DRB1*03:01:01* was most common with a frequency of 8.8% followed by three haplotypes with an identical frequency of 8.1% (Table 5). There were no statistically significant differences in *HLA-DQB1/DRB1* haplotype frequencies between the two groups.

3.3.4.3 Three-locus HLA class II DPB1/-DQB1/-DRB1 haplotypes

The most predominant three-locus haplotype in Controllers was *HLA-DPB1*01:01:01/DQB1*04:02:01/DRB1*03:03:01* at a frequency of 9.6%. This haplotype was also represented within the Progressors, however, it was the fifth most frequent three-locus haplotype in that group at a lower frequency of 3.2% (Table 6). The frequency of this haplotype did not differ significantly between the groups (P=0.13). In total, Controllers had representation of eight unique three-locus haplotypes at frequencies of 2% or greater in comparison with Progressors with 13. The most frequent haplotype amongst Progressors was *HLA-DPB1*01:01:01/DQB1*06:03:01/DRB1*13:01:01* at 7.3%. There were no statistically significant differences in three-locus haplotype frequencies between the two groups.

Table 5. Estimated two-locus (*HLA-DPB1/-DQB1* and *HLA-DQB1/-DRB1*) haplotype frequencies in Controllers and Progressors.

***HLA-DPB1/HLA-DQB1* haplotypes**						***HLA-DQB1/HLA-DRB1* haplotypes**					
Controllers			**Progressors**			**Controllers**			**Progressors**		
Haplotype alleles		Freq %	Haplotype alleles		Freq %	Haplotype alleles		Freq %	Haplotype alleles		Freq %
HLA-DPB1	***HLA-DQB1***		***HLA-DPB1***	***HLA-DQB1***		***HLA-DQB1***	***HLA-DRB1***		***HLA-DQB1***	***HLA-DRB1***	
01:01:01	***04:02:01***	**13.5**	*105:01:01*	*03:19:01*	5.7	*04:02:01*	*03:02:01*	11.5	*02:01:01*	*03:01:01*	8.8
01:01:01	***06:02:01***	**9.6**	*01:01:01*	*05:01:01*	5.4	*03:02:01*	*04:01:01*	7.7	*06:03:01*	*13:01:01*	8.1
01:01:01	*06:09:01*	5.8	*01:01:01*	*06:03:01*	5.3	*02:02:01*	*07:01:01*	7.7	*02:02:01*	*07:01:01*	8.1
01:01:01	*03:02:01*	5.8	*105:01:01*	*02:02:01*	4.9	*06:09:01*	*13:02:01*	7.7	*04:02:01*	*03:02:01*	8.1
105:01:01	*02:02:01*	5.8	*105:01:01*	*06:02:01*	4.1	*06:03:01*	*13:01:01*	7.7	*03:19:01*	*11:01:02*	6.8
105:01:01	*06:03:01*	3.9	*105:01:01*	*02:01:01*	3.5	*06:02:01*	*15:03:01*	5.8	*06:02:01*	*11:01:02*	6.8
13:01:01	*05:01:01*	3.9	*01:01:01*	*03:19:01*	3.3	*06:09:01*	*13:01:01*	5.8	*06:02:01*	*15:03:01*	6.1
04:01:01	*03:02:01*	3.9	*01:01:01*	*02:01:01*	3.3	*06:02:01*	*11:01:02*	3.9	*06:09:01*	*13:02:01*	5.4
01:01:01	*06:03:01*	3.9	*01:01:01*	*02:02:01*	3.1	*02:01:01*	*03:01:01*	3.9	*05:01:01*	*12:01:01*	4.1
18:01	*06:09:01*	3.9	***01:01:01***	***04:02:01***	**3.0**	*03:19:01*	*11:01:02*	3.9	*05:01:01*	*01:02:01*	4.1
18:01	*06:02:01*	3.9	*02:01:19*	*06:09:01*	2.7	*05:01:01*	*10:01:01*	3.9	*03:19:01*	*08:04:01*	3.4
			01:01:01	*03:01:01*	2.7	*03:19:01*	*11:02:01*	3.9	*06:02:01*	*13:01:01*	2.7
			13:01:01	*03:19:01*	2.5				*03:19:01*	*11:02:01*	2.7
			02:01:02	*05:01:01*	2.0				*03:01:01*	*11:02:01*	2.0
			02:01:02	*06:09:01*	2.0						
			01:01:01	***06:02:01***	**2.0**						
			01:01:01	*03:02:01*	2.0						
			01:06:01	*06:02:01*	2.0						
			02:01:02	*06:02:01*	2.0						
			665:01	*04:02:01*	2.0						
			18:01	*05:01:01*	2.0						

The haplotypes with significant differences in frequencies between Controllers and Progressors, calculated using Fisher's Exact tests ($P<0.05$), are highlighted in bold print.

Table 6. Estimated three-locus (*HLA-DPB1/-DQB1/-DRB1*) haplotype frequencies in Controllers and Progressors.

***HLA-DPB1/-DQB1/-DRB1* haplotypes**							
HIV-1 Controllers 2n = 52				**HIV-1 Progressors 2n = 148**			
Haplotype alleles			Frequency %	Haplotype alleles			Frequency %
HLA-DPB1	***HLA-DQB1***	***HLA-DRB1***		***HLA-DPB1***	***HLA-DQB1***	***HLA-DRB1***	
01:01:01	*04:02:01*	*03:02:01*	9.6	*01:01:01*	*06:03:01*	*13:01:01*	7.3
105:01:01	*02:02:01*	*07:01:01*	5.8	*01:01:01*	*03:19:01*	*11:01:02*	5.4
01:01:01	*03:02:01*	*04:01:01*	3.9	*105:01:01*	*02:02:01*	*07:01:01*	5.4
105:01:01	*06:03:01*	*13:01:01*	3.9	*01:01:01*	*02:01:01*	*03:01:01*	3.6
18:01	*06:09:01*	*13:01:01*	3.9	*01:01:01*	*04:02:01*	*03:02:01*	3.2
18:01	*06:02:01*	*15:03:01*	3.9	*105:01:01*	*02:01:01*	*03:01:01*	3.1
01:01:01	*06:03:01*	*13:01:01*	3.9	*02:01:19*	*06:09:01*	*13:02:01*	2.7
01:01:01	*06:02:01*	*11:01:02*	3.9	*105:01:01*	*06:02:01*	*11:01:02*	2.0
				105:01:01	*03:19:01*	*08:04:01*	2.0
				105:01:01	*05:01:01*	*12:01:01*	2.0
				02:01:02	*06:02:01*	*15:03:01*	2.0
				665:01	*04:02:01*	*03:02:01*	2.0
				18:01	*05:01:01*	*01:02:01*	2.0

3.4 Comparison of HLA class II allele frequencies between Elite controllers and Progressors

Seventeen of the 26 HIV-1 Controllers suppressed HIV-1 viral loads to <50 copies/mL and can be termed Elite Controllers (ECs) (Deeks and Walker, 2007). A sub-study was conducted to determine if this sub-group of Controllers was particularly enriched for certain HLA class II -*DPB1, -DQB1* and *-DRB1* alleles. Allele frequencies were compared between the Progressor and EC groups (Table 7). The alleles tested for significant differences in frequency between the groups were the alleles that had frequencies differing by at least 5% between Controllers and Progressors in the previous analysis (Tables 2, 3 and 4) and that had frequencies differing by at least 5% between ECs and Progressors in the current analysis (Table 7).

The frequency of allele *HLA-DPB1*01:01:01* was higher amongst ECs than Controllers (ECs 44.1% vs Controllers 40.4%) but not significantly different to the frequency in the Progressor group (ECs 44.1% vs Progressors 32.4%, P=0.232, Table 7). *HLA-DQB1*03:02:01* was previously shown to be more frequent within the Controller group, however when tested between ECs and Progressors this difference was not significant (ECs 8.8% vs Progressors 2.0%, P=0.080, Table 7).

There was a statistically non-significant trend towards a higher frequency of *HLA-DQB1*06:09:01* within the Controller group in the previous analysis (P=0.153, Table 3). Although the frequency of *HLA-DQB1*06:09:01* was higher within the EC sub-group compared to the Controller group (ECs 17.6% vs Controllers 13.5%), the difference was still not significant when compared to frequency within the Progressor group (ECs 17.6% vs Progressors 6.8%, P=0.084, Table 7). However, the significance value showed a stronger trend than in the comparison of the frequencies of this allele between the Controllers and Progressors (EC vs Progressors P=0.084 (Table 7); Controllers vs Progressors P=0.153 (Table 3).

The *HLA-DRB1*04:01:01* allele was significantly enriched within the Controller group when compared to the Progressor group (Table 4). However, the frequency of this allele within the EC sub-group of Controllers was lower than that of the total Controller group (ECs 5.9% vs Controllers 7.7%). As a result, the frequency of the *HLA-DRB1*04:01:01* allele was not significantly different between ECs and Progressors (ECs 5.9% vs Progressors 1.4%, P=0.159, Table 7).

In summary, although some of the alleles within the EC group differed in frequency when compared to the total Controller group, none of the allele frequencies was significantly different when compared between the EC sub-group of HIV-1 Controllers and the group of Progressors (Table 7).

Table 7. Frequencies of *HLA-DPB1, HLA-DQB1 and HLA-DRB1* alleles in Elite Controllers and Progressors.

HLA class II allele	Elite controllers (2n=34)		Progressors (2n=148)			
	Allele frequency %	(n)	Allele frequency %	(n)	OR (95%CI)	P
HLA-DPB1						
01:01:01	44.1	15	32.4	48	0.61 (0.28 – 1.3)	0.232
105:01:01	17.6	6	19.6	29	1.10 (0.43 – 3.0)	1.000
HLA-DQB1						
02:01:01	5.9	2	10.1	15	1.80 (0.39 – 8.3)	0.743
03:02:01	8.8	3	2.0	3	0.21 (0.04 – 1.1)	0.080
03:19:01	5.9	2	14.9	22	2.80 (0.62 – 13)	0.259
06:09:01	17.6	6	6.8	10	0.34 (0.11 – 1.0)	0.084
***HLA*-DRB1**						
03:02:01	14.7	5	8.8	13	0.56 (0.18 – 1.7)	0.338
04:01:01	5.9	2	1.4	2	0.22 (0.03 – 1.62)	0.159
11:01:02	8.8	3	14.2	21	1.70 (0.48 – 6.1)	0.576
13:01:01	17.6	6	14.2	21	0.77 (0.29 – 2.1)	0.598

P values, odds ratios (OR) and 95% confidence Intervals (95%CI) are shown for Fisher's Exact tests. Significant results (P<0.05) are highlighted in bold print.

4 Discussion

HIV-specific CD4+ T cells have increasingly been shown to play an important role in HIV-1 disease pathogenesis, suggesting a possible role for HLA class II molecules in HIV-1 control. Despite the vast amount of work that has investigated the impact of HLA class I on HIV-1 disease progression, very little is known about the impact of HLA class II in HIV-1 disease. Studies that have investigated the role of HLA class II genes in HIV-1 control have done so in predominantly male Caucasian populations infected with subtype B HIV-1. However, an understanding of the role of HLA class II genes in a black African population infected with subtype C HIV-1 is needed, as this population group suffers the greatest burden of HIV-1 disease worldwide. The development of effective vaccines and immunological therapeutics in this group would be particularly impactful on the global epidemic (Hemelaar, 2012).

We hypothesized that certain HLA class II alleles may be found at different frequencies in individuals who control HIV-1 disease and those who develop progressive disease. We, therefore, investigated a unique group of individuals who spontaneously control HIV-1 viraemia to very low levels in the absence of ART and are known as HIV-1 Controllers. This group provides the most appropriate natural model for the study of genetic determinants that are associated with delayed progression of HIV-1 disease. Outside of this study, HLA class II allele frequencies have not been investigated in a black African Controller population before. Using SBT, we determined *HLA-DPB1, -DQB1* and *-DRB1* alleles to 6 digit-resolution for 26 Controllers. These allele frequencies were then compared to those in a group of 74 black African individuals with progressive HIV-1 disease from the same community in order to determine any allele associations with HIV-1 control or progression.

A significantly higher frequency of the allele *HLA-DQB1*03:02:01* was found amongst Controllers, suggesting that this allele may be associated with HIV-1 control. No studies have revealed this particular genetic association before, although studies in autoimmune disease have shown an enrichment of *HLA-DQB1*03:02* in individuals with type 1 diabetes, autoimmune thyroid disease and coeliac disease (Kahaly *et al.*, 2018). Previous studies have, however, suggested a role for the allele *HLA-DQB1*06* in HIV-1 control in Caucasians, albeit mostly in association with *HLA-DRB1*13* (Malhotra *et al.*, 2001; Vyakarnam *et al.*, 2004; Ferre *et al.*, 2010). *HLA-DQB1*06* alleles have been associated with more robust CD4+ and CD8+ T cell responses in previous studies (Ferre *et al.*, 2009). *HLA-DQB1*06:09:01*, which

had a frequency of 13.5% in the Controller group, was postulated to play a role in long-term virological control after ART cessation in a South African child in a paper that was recently published by our research group (Violari *et al.*, 2019). Although the frequency of *HLA-DQB1*06:09:01* was increased amongst Controllers in our study, this difference was not statistically significant. This trend may prove to be more significant in a larger study population and warrants further investigation. Despite being associated with control in other Caucasian studies, we did not find significant enrichment of *HLA-DQB1*06*, or any variants thereof, in our Controller group. This difference may be due to viral subtype differences, different population study sizes as well as different HLA class II allele frequencies that are commonly found in Caucasian and black African populations.

Amongst *HLA-DRB1* alleles, *HLA-DRB1*04:01:01* was significantly more frequent amongst Controllers, suggesting that this allele may also be associated with HIV-1 control in the black African population studied here. This association has also not been shown before. In contrast to our findings, a study carried out in chronic HIV-1 infection in a similar black African population in KwaZulu-Natal did not show a significant association between the frequency of the *HLA-DRB1*04:01* allele and CD4+ T cell count or viral load. There was, however, an association evident between the frequency of the *HLA-DRB1*04:04* allele and higher CD4+ T cell counts (Julg *et al.*, 2011). *HLA-DRB1*04:04:01* was present at a very low frequency in our study population and the frequencies of this allele did not differ significantly between the Controller and Progressor groups.

The study by Julg *et al.* also showed that the *HLA-DRB1*13:03* allele was associated with higher CD4+ T cell counts and lower viral loads (Julg *et al.*, 2011). This finding was confirmed in another separate, predominantly male, Caucasian American population in the same study, suggesting that the *HLA-DRB1*13:03*-mediated protection described was independent of sex, ethnicity, and HIV-1 viral clade (Julg *et al.*, 2011). Other studies have also suggested a protective role for *HLA-DRB1*13* alleles in HIV-1 disease (Chen *et al.*, 1997; Keet *et al.*, 2002; Ferre *et al.*, 2010). None of the *HLA-DRB1*13* alleles in our cohort *(DRB1*13:01:01, DRB1*13:02:01* or *DRB1*13:03:01*) were significantly enriched within the Controller group. Although *HLA-DRB1*13* alleles were represented in roughly 25% of our total cohort, *HLA-DRB1*13:03:01* was found at a low frequency of 2% of the total cohort. This allele was present amongst 3.7% of the KwaZulu-Natal cohort (Julg *et al.*, 2011). The black African populations of KwaZulu-Natal and Johannesburg would be expected to be genetically similar,

however, the possibility exists that the Johannesburg cohort may be more heterogeneous due to the fact that the city is the economic hub of South Africa with a greater admixture of cultures than the province of KwaZulu-Natal. These disparities may account for some of the differences in allele frequencies that are found when comparing our results with those of Julg *et al*. In addition, they studied a larger cohort of individuals that were chronically infected with HIV-1 and were not particularly enriched for "Controller" or "Progressor" status. Therefore, HLA class II alleles that have a weak association with HIV-1 control and do not confer "controller" ability may be more likely to be detected in a larger cohort such as theirs. By contrast, our study compared two extreme HIV-1 disease phenotypes and was more suited to detect HLA class II alleles with more powerful associations with HIV-1 control. There is also significant evidence that HIV-1 has adapted to human HLA-mediated selection pressure at a population level, possibly resulting in the development of an attenuated virus over time (Kawashima *et al.*, 2009; Payne *et al.*, 2014; Kløverpris *et al.*, 2016). When considering that the cohorts of Julg *et al.* and ours were recruited more than 10 years apart, this viral adaptation may contribute to different allele frequency associations with virological control found at different time points within the South African HIV-1 epidemic.

The *HLA-DRB1*04:01* allele has also been associated with autoimmune disease. In a South African study, *HLA-DRB1*04:01* was shown to be associated with rheumatoid arthritis in black African and Caucasian patients (Meyer *et al.*, 2004). Other studies have linked this allele to type 1 diabetes as well as multiple sclerosis (Lobnig *et al.*, 2002). Common to both alleles that were found to be significantly more frequent amongst Controllers, is the association with autoimmune disease in other cohorts. As has been postulated in previous studies, the possible mechanism conferring protection in HIV-1 disease, through highly effective antigen presentation, may inadvertently result in autoimmunity in some individuals through the presentation of self-antigens (Fainboim *et al.*, 2001; Julg *et al.*, 2011). The protective HLA class I allele *B*27:05* has similarly been associated with HIV-1 control and autoimmune disease in the case of ankylosing spondylitis (Thomas and Brown, 2010; Goulder and Walker, 2012).

Determining the mechanism of protection conferred by *HLA-DQB1*03:02:01* and *HLA-DRB1*04:01:01* was beyond the scope of this study as this would require extensive further functional analyses. A previous non-human primate study in a single SIV-infected rhesus macaque "elite controller" showed that disease progression in that monkey was associated

with the development of viral mutations in both CD4+ and CD8+ T cell epitopes (Burwitz *et al.*, 2012). They further demonstrated that the development of a mutation in a specific CD4+ T cell epitope, resulted in decreased CD4+ T cell recognition of that epitope and resultant loss of immune control (Burwitz *et al.*, 2012). Another study using a computational approach in humans to determine predicted viral adaptation to HLA class II-mediated CD4+ T cell pressure, showed higher CD4+ T cell responses to predicted adapted epitopes amongst controllers than in non-controllers (Erdmann *et al.*, 2015). HIV-specific CD4+ T cell responses in individuals with "protective" HLA class II alleles have been shown to be more polyfunctional and of greater magnitude and increased breadth than in individuals without these alleles (Ferre *et al.*, 2010; Ranasinghe *et al.*, 2013). Targeting of immunodominant Gag epitopes appears to also mediate some of the effect of these alleles (Ranasinghe *et al.*, 2012, 2013; Laher *et al.*, 2017). Non-CD4+ T cell-mediated mechanisms such as augmentation of HIV-specific CD8+ T cell responses may also play a role in the protection associated with some HLA class II alleles (Ferre *et al.*, 2009). Therefore, we may postulate that the protection conferred by *HLA-DQB1*03:02:01* and *HLA-DRB1*04:01:01* may be associated with a more robust HIV-specific CD4+ T cell response and possibly the associated CD8+ T cell response too, amongst other possible mechanisms. This remains to be elucidated in further studies.

The power of this study to detect significant differences in allele frequency was unfortunately limited by the relatively small numbers of HIV-1 Controllers that could be recruited. This limitation was due to the rarity of the HIV-1 controller phenotype, as well as the fact that, once screened, HIV-infected individuals are currently rapidly initiated on ART as part of the "test and treat" public health response in South Africa. Moreover, due to the large number of HLA class II alleles present at each locus, the multiple comparisons that are performed reduces the statistical power of the findings. Due to the exploratory nature of the study, correction for multiple comparisons was not carried out in this study. An attempt was made to further enrich the "Controller" phenotype through comparing individuals within the broader Controller group that are considered ECs to the Progressors. The result was a loss of significant associations with *HLA-DQB1*03:02:01* and *HLA-DRB1*04:01:01* and HIV-1 control, likely due to the even smaller EC group numbers.

It is important to note that these findings are specific to the particular population studied and need to be confirmed in further studies of larger populations, including other races and ethnic groups, in order to extend the significance of the gene associations uncovered here.

The fact that the overwhelming majority of our participants were women also reduces the ability to apply these findings beyond our specific study population. Only HLA class II data for our study participants were available for the purposes of this study. As a result, analyses to exclude linkage disequilibrium (LD) between HLA class II alleles shown in this study to be associated with HIV-1 control and HLA class I alleles that have previously been shown to be protective within the same population, were not possible. Whilst neither *HLA-DQB1*03:02:01* or *HLA-DRB1*04:01:01* are known to be in LD with known protective HLA class I alleles in this particular population, this would need to be explored further and excluded.

Our findings suggest that further study is warranted to investigate associations between HLA class II frequencies and HIV-1 control. Functional immunological experiments exploring the mechanisms behind HLA class II-mediated protection in HIV-1 disease would add to the ability to apply these findings to current vaccine and HIV-1 eradication approaches. These findings also support the further study of the role that CD4+ T cells play in HIV-1 pathogenesis, suggesting that future HIV-1 preventative and therapeutic vaccine efforts may benefit from a renewed focus on HIV-specific CD4+ T cell responses in addition to the frequently studied HIV-specific CD8+ T cell and antibody responses.

5 References

Burwitz, B. J., Giraldo-Vela, J. P., Reed, J., *et al.* (2012) CD8+ and CD4+ cytotoxic T cell escape mutations precede breakthrough SIVmac239 viremia in an elite controller, *Retrovirology*, 9(91), pp. 1–11.

Carrington, M. and O'Brien, S. J. (2003) The influence of HLA genotype on AIDS, *Annual Review of Medicine*, 54, pp. 535–51.

Chambuso, R., Ramesar, R., Kaambo, E., *et al.* (2019) Human Leukocyte Antigen (HLA) class II -DRB1 and -DQB1 Alleles and the Association with Cervical Cancer in HIV/HPV Co-infected Women in South Africa, *Journal of Cancer*, 10(10), pp. 2145–2152.

Chen, H., Ndhlovu, Z. M., Liu, D., *et al.* (2012) TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection, *Nature Immunology*, 13(7), pp. 691–700.

Chen, Y., Winchester, R., Korber, B., *et al.* (1997) 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, *Human Immunology*, 55(2), pp. 154–162.

Chevalier, M. F., Julg, B., Pyo, A., *et al.* (2011) HIV-1-Specific Interleukin-21+ CD4+ T Cell Responses Contribute to Durable Viral Control through the Modulation of HIV-Specific CD8+ T Cell Function, *Journal of Virology*, 85(2), pp. 733–741.

Crawford, H., Lumm, W., Leslie, A., *et al.* (2009) Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703–positive individuals and their transmission recipients, *The Journal of Experimental Medicine*, 206(4), pp. 909–921.

Deeks, S. G. and Walker, B. D. (2007) Human Immunodeficiency Virus Controllers: Mechanisms of Durable Virus Control in the Absence of Antiretroviral Therapy, *Immunity*, pp. 406–416.

Delves, P. J., Martin, S. J., Burton, D. R., *et al.* (2006) *Roitt's Essential Immunology*. 11th edn. Oxford: Blackwell Publishing.

Dendrou, C. A., Petersen, J., Rossjohn, J., *et al.* (2018) HLA variation and disease, *Nature Reviews Immunology*, pp. 325–339.

Erdmann, N., Du, V. Y., Carlson, J., *et al.* (2015) HLA Class-II Associated HIV Polymorphisms Predict Escape from CD4+ T Cell Responses, *PLoS Pathogens*, 11(8), p. e1005111.

Faculte de medicine de Geneve (2015) *HLA genes and molecules*.

Fainboim, L., Cañero Velasco, M. C., Marcos, C. Y., *et al.* (2001) Protracted, but not acute, hepatitis A virus infection is strongly associated with HLA-DRB1*1301, a marker for pediatric autoimmune hepatitis, *Hepatology*, 33(6), pp. 1512–1517.

Fellay, J., Shianna, K. V, Ge, D., *et al.* (2007) A whole-genome association study of major determinants for host control of HIV-1, *Science*, 317(5840), pp. 944–7.

Ferre, A. L., Hunt, P. W., Critchfield, J. W., *et al.* (2009) Mucosal immune responses to HIV-1 in elite controllers: A potential correlate of immune control, *Blood*, 113(17), pp. 3978–3989.

Ferre, A. L., Hunt, P. W., McConnell, D. H., *et al.* (2010) HIV Controllers with HLA-DRB1*13 and HLA-DQB1*06 Alleles Have Strong, Polyfunctional Mucosal CD4+ T-Cell Responses, *Journal of Virology*, 84(21), pp. 11020–11029.

González-Galarza, F. F., Takeshita, L. Y. C., Santos, E. J. M., *et al.* (2015) Allele frequency net 2015 update: New features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations, *Nucleic Acids Research*, 43(D1), pp. D784–D788.

Goulder, P. J. R. and Walker, B. D. (2012) HIV and HLA class I: an evolving relationship, *Immunity*, 37(3), pp. 426–40.

Hemelaar, J. (2012) The origin and diversity of the HIV-1 pandemic, *Trends in Molecular Medicine*, pp. 182–192.

Hersperger, A. R., Martin, J. N., Shin, L. Y., *et al.* (2011) Increased HIV-specific CD8+ T-cell cytotoxic potential in HIV elite controllers is associated with T-bet expression, *Blood*, 117(14), pp. 3799–3808.

Horton, H., Frank, I., Baydo, R., *et al.* (2006) Preservation of T Cell Proliferation Restricted by Protective HLA Alleles Is Critical for Immune Control of HIV-1 Infection, *The Journal of Immunology*, 177(10), pp. 7406–7415.

Jameson, J. L., Kasper, D. L., Longo, D. L., *et al.* (2018) *Harrison's Principles of Internal Medicine*. 20th edn. Edited by J. L. Jameson. New York: McGraw-Hill.

Julg, B., Moodley, E. S., Qi, Y., *et al.* (2011) Possession of HLA class II DRB1*1303 associates with reduced viral loads in chronic HIV-1 clade C and B infection, *Journal of Infectious Diseases*, 203(6), pp. 803–809.

Kahaly, G. J., Frommer, L. and Schuppan, D. (2018) Celiac disease and endocrine autoimmunity – the genetic link, *Autoimmunity Reviews*, pp. 1169–1175.

Kawashima, Y., Pfafferott, K., Frater, J., *et al.* (2009) Adaptation of HIV-1 to human leukocyte antigen class I., *Nature*, 458(7238), pp. 641–5.

Keet, I. P. M., Tang, J., Klein, M. R., *et al.* (2002) Consistent Associations of HLA Class I and II and Transporter Gene Products with Progression of Human Immunodeficiency Virus Type 1 Infection in Homosexual Men, *Journal of Infectious Diseases*, 180(2), pp. 299–309.

Kiepiela, P., Ngumbela, K., Thobakgale, C., *et al.* (2007) CD8+ T-cell responses to different HIV proteins have discordant associations with viral load, *Nature Medicine*, 13(1), pp. 46–53.

Klepiela, P., Leslie, A. J., Honeyborne, I., *et al.* (2004) Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA, *Nature*, 432(7018), pp. 769–774.

Kløverpris, H. N., Leslie, A. and Goulder, P. (2016) Role of HLA adaptation in HIV evolution, *Frontiers in Immunology*, 6(665), pp. 1–15.

Koup, R. A., Safrit, J. T., Cao, Y., *et al.* (1994) Temporal Association of Cellular Immune Responses with the Initial Control of Viremia in Primary Human Immunodeficiency Virus Type 1 Syndrome, *Journal of Virology*, 68(7), pp. 4650–4655.

Laher, F., Mewalal, N., Pretorius, K., *et al.* (2017) HIV Controllers Exhibit Enhanced Frequencies of Major Histocompatibility Complex Class II Tetramer+ Gag-Specific CD4+ T cells in Chronic Clade C HIV-1 Infection, *Journal of Virology*, 91(7), pp. e02477-16.

Lancaster, A. K., Single, R. M., Solberg, O. D., *et al.* (2007) PyPop update - A software pipeline for large-scale multilocus population genomics, *Tissue Antigens*, 69(SUPPL. 1), pp. 192–197.

Lobnig, B. M., Chantelau, E., Vidgrén, G., *et al.* (2002) HLA-patterns in patients with multiple sclerosis and type I diabetes mellitus: Evidence for possible mutual exclusion of both

diseases, *Diabetes and Metabolism*, 28(3), pp. 217–221.

Loubser, S., Paximadis, M., Gentle, N. L., *et al.* (2020) Human leukocyte antigen class I (A, B, C) and class II (DPB1, DQB1, DRB1) allele and haplotype variation in Black South African individuals, *Human Immunology*. Elsevier, 81(1), pp. 6–7.

Mack, S. J., Cano, P., Hollenbach, J. A., *et al.* (2013) Common and well-documented HLA alleles: 2012 update to the CWD catalogue, *Tissue Antigens*, 81(4), pp. 194–203.

Malhotra, U., Holte, S., Dutta, S., *et al.* (2001) Role for HLA class II molecules in HIV-1 suppression and cellular immunity following antiretroviral treatment, *Journal of Clinical Investigation*, 107(4), pp. 505–517.

Marsh, S. G. E., Albert, E. D., Bodmer, W. F., *et al.* (2010) Nomenclature for factors of the HLA system, 2010, *Tissue Antigens*, 75(4), pp. 291–455.

Martinson, N. A., Gupte, N., Msandiwa, R., *et al.* (2014) CD4 and viral load dynamics in antiretroviral-naïve HIV-infected adults from Soweto, South Africa: A prospective cohort, *PLoS ONE*, 9(5).

Matloubian, M., Concepcion, R. J. and Ahmed, R. (1994) CD4+ T cells are required to sustain CD8+ cytotoxic T-cell responses during chronic viral infection, *Journal of Virology*, 68(12), pp. 8056–63.

McBrien, J. B., Kumar, N. A. and Silvestri, G. (2018) Mechanisms of CD8+ T cell-mediated suppression of HIV/SIV replication, *European Journal of Immunology*, 48(6), pp. 898–914.

Meyer, P. W. A., Brighton, S. W. and Anderson, R. (2004) Association of the HLA-DRB1*04 allele and its subtypes with rheumatoid arthritis in South Africa : research letter, *South African Journal of Science*, 100(5–6), pp. 305–306.

Ndung'u, T., Gaseitsiwe, S., Sepako, E., *et al.* (2005) Major Histocompatibility Complex Class II (HLA-DRB and -DQB) Allele Frequencies in Botswana: Association with Human Immunodeficiency Virus Type 1 Infection, *Clinical and Vaccine Immunology*, 12(9), pp. 1020–1028.

Norris, P. J., Moffett, H. F., Yang, O. O., *et al.* (2004) Beyond Help: Direct Effector Functions of Human Immunodeficiency Virus Type 1-Specific CD4+ T Cells, *Journal of Virology*, 78(16), pp. 8844–8851.

Paximadis, M., Mathebula, T. Y., Gentle, N. L., *et al.* (2012) Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population, *Human Immunology*, 73(1), pp. 80–92.

Payne, R., Muenchhoff, M., Mann, J., *et al.* (2014) Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence, *Proceedings of the National Academy of Sciences of the United States of America*, 111(50), pp. E5393–E5400.

Pereyra, F., Jia, X., McLaren, P. J., *et al.* (2010) The Major Genetic Determinants of HIV-1 Control Affect HLA class I Peptide Presentation, *Science*, 330(6010), pp. 1551–7.

Phetsouphanh, C., Aldridge, D., Marchi, E., *et al.* (2019) Maintenance of Functional CD57+ Cytolytic CD4+ T Cells in HIV+ Elite Controllers, *Frontiers in Immunology*. Frontiers, 10(1844).

Phillips, R. E., Rowland-Jones, S., Nixon, D. F., *et al.* (1991) Human Immunodeficiency Virus Genetic Variation That Can Escape Cytotoxic T Cell Recognition, *Nature*, 354(6353), pp. 453–459.

Ramduth, D., Day, C. L., Thobakgale, C. F., *et al.* (2009) Immunodominant HIV-1 Cd4+ T cell epitopes in chronic untreated clade C HIV-1 infection, *PLoS ONE*, 4(4), p. e5013.

Ranasinghe, S., Cutler, S., Davis, I., *et al.* (2013) Association of HLA-DRB1-restricted CD4+ T cell responses with HIV immune control, *Nature Medicine*, 19(7), pp. 930–933.

Ranasinghe, S., Flanders, M., Cutler, S., *et al.* (2012) HIV-Specific CD4 T Cell Responses to Different Viral Proteins Have Discordant Associations with Viral Load and Clinical Outcome, *Journal of Virology*, 86(1), pp. 277–283.

Rosenberg, E. S., Billingsley, J. M., Caliendo, A. M., *et al.* (1997) Vigorous HIV-1-specific CD4 + T cell responses associated with control of viremia, *Science*, 278(5342), pp. 1447–1450.

Schmitz, J. E., Kuroda, M. J., Santra, S., *et al.* (1999) Control of Viremia in Simian Immunodeficiency Virus Infection by CD8+ Lymphocytes, *Science*, 283(5403), pp. 857–860.

Single, R. M. and Thomson, G. (2016) Linkage Disequilibrium: Population Genetics of Multiple Loci, in *Encyclopedia of Evolutionary Biology*. Elsevier Inc., pp. 400–404.

Soghoian, D. Z., Jessen, H., Flanders, M., *et al.* (2012) HIV-Specific Cytolytic CD4 T Cell Responses During Acute HIV Infection Predict Disease Outcome, *Science Translational Medicine*, 4(123), p. 123ra25.

Soghoian, D. Z. and Streeck, H. (2010) Cytolytic CD4+ T cells in viral immunity, *Expert Review of Vaccines*, pp. 1453–1463.

de Souza, M. S., Ratto-Kim, S., Chuenarom, W., *et al.* (2012) The Thai Phase III Trial (RV144) Vaccine Regimen Induces T Cell Responses That Preferentially Target Epitopes within the V2 Region of HIV-1 Envelope, *The Journal of Immunology*, 188(10), pp. 5166–5176.

Statistics South Africa (2019) *Mid-year population estimates 2019*. Pretoria.

Thomas, G. P. and Brown, M. A. (2010) Genetics and genomics of ankylosing spondylitis, *Immunological Reviews*, pp. 162–180.

Thorstenson, Y. R., Creary, L. E., Huang, H., *et al.* (2018) Allelic resolution NGS HLA typing of Class I and Class II loci and haplotypes in Cape Town, South Africa, *Human Immunology*, 79(12), pp. 839–847.

Violari, A., Cotton, M. F., Kuhn, L., *et al.* (2019) A child with perinatal HIV infection and long-term sustained virological control following antiretroviral treatment cessation, *Nature Communications*, 10(1).

Vyakarnam, A., Sidebottom, D., Murad, S., *et al.* (2004) Possession of human leucocyte antigen DQ6 alleles and the rate of CD4 T-cell decline in human immunodeficiency virus-1 infection, *Immunology*, 112(1), pp. 136–142.

Zinkernagel, R. M. and Doherty, P. C. (1974) Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system, *Nature*, 248(5450), pp. 701–702.

6 Appendix

Appendix 1.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Professor CT Tiemessen, et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M190995

NAME: (Principal Investigator)	Professor CT Tiemessen, et al
DEPARTMENT:	National Institute for Communicable Diseases Centre for HIV and Sexually-Transmitted Infections Sandringham
PROJECT TITLE:	HIV-1 positive South African Elite and Long-term Controllers: viral and host targets for functional cure strategies
DATE CONSIDERED:	Ad hoc
DECISION:	Aproved unconditionally
CONDITIONS:	Renewal of M140926
SUPERVISOR:	Not applicable
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2019/10/04

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the 3rd Floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **September** and will therefore reports and re-certification will be due early in the month of **September** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date

PLEASE QUOTE THE CLEARANCE CERTIFICATE NUMBER IN ALL ENQUIRIES