University of Cape Town

Department of Clinical Laboratory Sciences

THE ROLE OF EARLY CYTOTOXIC LYMPHOCYTE (CTL) ESCAPE IN THE PATHOGENESIS OF HIV-1 SUBTYPE C INFECTION

ROMAN SABA NTALE

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Division of Medical Virology
Department of Clinical Laboratory Sciences in the Faculty of Health Sciences
and the Institute of Infectious Diseases and Molecular Medicine
UNIVERSITY OF CAPE TOWN

SUPERVISOR

Prof. CAROLYN WILLIAMSON

DECEMBER 2012

CONTENTS

ABSTRACT

HIV-1 rapid escape from immune surveillance can be detrimental to the host. However, if these escape mutations are located in functionally important regions of the viral genome, they can attenuate viral replication and provide some benefit to the host. Understanding the dynamics of viral evolution in early infection and events associated with control of infection will inform vaccine design. This study investigated the frequency and timing of cytotoxic T-lympthocyte (CTL) escape and its pathogenic consequences on HIV-1 subtype C disease progression.

Samples were obtained from 58 individuals recruited within three months of HIV-1 infection into the CAPRISA 002 acute infection cohort. Full-length *gag (n=365)* and *nef (n=289)* sequences were generated from enrolment 0 (median 1.4 months [IQR=1-2]), 3, 6, 12 and 24 months post infection. Additional sequences (n=120) were generated to investigate the fixation of escape mutations. Putative escape was defined as non-synonymous changes within or flanking known epitopes restricted by the participants HLA. Timing of escape was defined as either acute (≤ 3 months), early (>3 to ≤6 months) or late (>6 to 12 month). For the newly identified mutation, replication of wild-type and mutant viruses in PBMCs was compared.

All individuals were infected with subtype C viruses: 51 had single infections and seven individuals were dually infected with two phylogenetically distinct strains. Of these seven, four were co-infected at or close to transmission, and three were superinfected. CTL escape was detected in individuals with single infection and identified in 69% (35/51) of individuals. It was more frequently observed in the acute than early/late infection phases (3.8 versus 1.1 escape/month/100 amino acid; p=0.0231). Participants with viruses escaping CTL responses in Gag and/or Nef had significantly higher CD4+T-cell counts at 3 and 6 months post-infection (PI) compared to those whose virus never escaped (p=0.0018 and 0.0542, respectively). Furthermore, although the decline in CD4+ counts was similar after 3 months, the initial preservation of the CD4+ cells in early infection resulted in an overall higher CD4 counts over the first 12 months of infection in individuals whose

viruses escaped in acute infection, compared to those that never escaped over this time period (p=0.0274). Moreover, HLA-B alleles selected a significantly higher number of viral escape mutations (72.4%, 42/58) compared to HLA-A alleles (17.2%, 10/58) (p=0.0134).

In the second part of this work, the timing and impact of mutations in the TL9 epitope on disease progression in five B*81:01- and two B*39:10-positive subtype C infected individuals was investigated. While cytotoxic T-lymphocyte (CTL) mediated escape mutations in this epitope have been associated with an *in vitro* fitness cost to the virus, the impact on disease progression *in vivo* had never been investigated. Whereas both B*39:10 participants sampled at two months post-infection had viruses with mutations in the TL9 epitope, in three of the five (3/5) B*81:01 participants, TL9 escape mutations were only detected 10 months after infection, taking an additional 10 to 15 months to reach fixation. In the two remaining B*81:01 individuals, one carried a TL9 escape variant at 2 weeks post-infection, whereas no escape mutations were detected in the virus from the other participant up to 33 months post-infection, despite CTL targeting of the epitope. In all participants, escape mutations in TL9 were linked to co-evolving residues in the region of Gag known to be associated with host tropism. Late escape in TL9, together with co-evolution of putative compensatory mutations, coincided with spontaneous increase in viral loads in two individuals who were otherwise controlling infection.

Finally, in an attempt to resolve the conflict in the field over the impact of transmitted CTL escape mutations on disease progression, viral loads and CD4+ counts were compared in participants infected with viruses carrying polymorphisms associated with beneficial HLA-B*57/58:01 and B7 supertype (B*39:10/81:01) allelic selective pressure in HLA mismatched participants. In HLA-B*57/58:01-negative participants, 51% (25/49) were infected with viruses carrying at least one of the B*57/58:01 attenuating escape mutations in ISW9 (A146P:19/49), KF11 (A163G: 10/49) or TW10 (T242N/S: 11/49) epitope. In HLA-B7 negative participants, only 9.4% (3/32) had viruses carrying mutations in the HLA-B7 immunodominant TL9 epitope. A novel mutation in Gag p17, Q65H, was identified which, in PBMCs assays, reduced replication by 22% compared to wild-type. This mutation, thought to be restricted by HLA-B7 and A*30 alleles, was in 5.4% (2/37) of

B7 and A*30 negative participants. Unlike a previous study in this cohort, which analysed a subset of 21 individuals, infection with viruses carrying mutations in HLA-B*57/58:01 restricted epitopes alone did not impact on viral load setpoint in HLA-mismatched participants, although CD4 counts at 3 months post infection were higher in participants infected with viruses carrying greater than any 2 of these mutations (p=0.031). However, HLA-mismatched participants infected with viruses carrying 3 or more of any of the B*57/58:01 or B*39:10/81:01 attenuating escape mutations had significantly lower viral load and higher CD4 counts at 3 months (p=0.017 and p=0.041) and at 12 months (p=0.008 and 0.091) post-infection, respectively.

Overall, this study demonstrates a high frequency of CTL escape in early infection compared to later, and that immune responses selecting escape mutations in acute infection provide benefit within the first year of infection. Furthermore, this study suggests that multiple mutations generated when viruses are passaged through individuals with beneficial HLAs are needed to attenuate the virus. However, B*81:01 mediated viral evolution in chronic infection, in a single Gag p24 epitope, had a detrimental impact on the control of viraemia. Despite this, the overall benefit of escape in acute infection on survival in the first year of infection provides support for vaccination methodologies that aim to render the virus less fit

This study has contributed to the following publications:

R.S. Ntale, D.R. Chopera, N.K. Ngandu, D. Assis de Rosa, M. Mlotshwa, L. Werner, Z. Woodman, K. Mlisana, S. Abdool Karim, C.M. Gray, C. Williamson and the CAPRISA 002 Study Team.2012. **Temporal association of HLA-B*81:01 and B*39:10 mediated HIV-1 p24 sequence evolution with disease progression.** *Journal of Virology*, 86:12013-12024

RS Ntale, DR Chopera, NK Ngandu, M Abrahams, A Debra, M Mlotswa, L Werner, Z Woodman, K Mlisana, S Karim, CM Gray, C Williamson, CAPRISA 002 AI Study Team (13 September 2012). **Beneficial HLA-mediated viral polymorphisms on the transmitted virus additively influence disease progression in HIV-1, subtype C infection.** *Retrovirology* 2012, 9(Suppl 2):O60

R.S. Ntale, D.R. Chopera, N. Ngandu, *et al.* 2011. **Association of Timing of HLA B*8101 and B*3910 Mediated p24 Sequence Evolution and Pathways to CTL Escape with Disease Progression.** ABSTRACT In: AIDS RESEARCH AND HUMAN RETROVIRUSES, Volume: 27 Issue: 10 Pages: A14-A14

R.S. Ntale, D.R. Chopera, D. Assis de Rosa, CM. Gray, K. Mlisana, S. Abdool Karim, C. Williamson and the CAPRISA 002 Study Team**. Association between early HLA-B restricted CTL escape in Nef and Gag with higher CD4+ T-Cell counts but not with viral load:** *Conference Abstracts, HIV Evolution, Genomics and Pathogenesis (X7)*, Whistler, March 20 - 25, 2011

Mandla Mlotshwa, Catherine Riou, Denis Chopera, Debra de Assis Rosa, **Roman Ntale,** Florette Treunicht, Zenda Woodman, Lise Werner, Francois van Loggerenberg, Koleka Mlisana, Salim Abdool Karim, Carolyn Williamson, Clive M. Gray, the CAPRISA 002 Study Team **Fluidity of HIV-1-Specific T-Cell Responses during Acute and Early Subtype C HIV-1 Infection and Associations with Early Disease Progression.** *Journal of Virology,* Nov. 2010; 84:12018-12029.

ABBREVIATIONS

AIDS: Acquired Immune Deficiency Syndrome

APOBEC3G: Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G

CAPRISA: Centre for the AIDS Programme of Research in South Africa

cDNA: copy DNA

CRFs: Circulating recombinant forms

CTL: Cytotoxic T lymphocytes

CXCR4: CXC chemokine receptor 4

EIA: Enzyme immunoassay

ELISA: Enzyme-Linked Immunosorbent Assay

GARD: Genetic Algorithms for Recombination Detection

IFN-γ: Interferon gamma

IL-: Interleukin

IQR: Interquartile range

HIV-1: Human Immunodeficiency Virus type 1

HLA: Human Leucocyte Antigen

HTA: Heteroduplex Tracking Assay

LTNPs: Long-term Nonprogressors

LTRs: Long Terminal Repeats

mRNA: messanger Ribonucleic acid

PBMCs: Peripheral Blood Mononuclear Cells

PCR: Polymerase Chain Reaction

RDP: Recombination Detection Program

RT: Rreverse Transcriptase

RNA: Ribonucleic acid

SI: Superinfection

SIV: Simian Immunodeficiency Virus

TNF-α : Tumour necrosis factor alpha

NK cells: Natural killer cells

μg: micro gram

μl: micro liters

UCT: University of Cape Town

UKZN: University of Kwazulu Natal

UNAIDS: Joint United Nations Programme on HIV/ AIDS

V1V2: Variable regions 1 and 2

DECLARATION

I, Roman Saba Ntale, hereby declare that the work contained in this thesis is my own original work and has not previously been submitted at any university for a degree. Contributions of others or use of published data have been duly acknowledged in the text. This work was done under the supervision of Professor Carolyn Williamson in the Division of Medical Virology, Department of Clinical Laboratory Sciences at the Institute of Infectious Diseases and Molecular Medicine, University of Cape Town.

Signed by candidate

Roman Saba Ntale

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisor Prof Carolyn Williamson for providing me an opportunity to undertake this research project in her department. I also wish to acknowledge the funding she provided, without which I doubt I would have been able to go this far. I am very grateful for the guidance, training and transferable skills I have received from her.

I also wish to acknowledge the contribution of all my fellow students and staff in the division of Medical Virology for their academic support whenever I needed their help. I am particularly grateful to Ms Melissa Abrahams and Dr Denis Chopera for the training on viral replication. I am also very grateful to Ms Debbie Stewart for help not only with administrative issues, but also for the assistance in proof reading this thesis. My thanks also go to Ms Nyameka for her assistance compiling my research references into the Reference Manager programme and Dr Ros Chapman for her training on the Reference Manager Software programme; and in particular, I am grateful for her positive attitude whenever I approached her for assistance.

I would like to take this opportunity to extend my thanks to the Centre for the AIDS Programme of Research in South Africa (CAPRISA) Acute Infection Study and the Study Team, including Professor Salim Abdool Karim, the director of CAPRISA.

Last but not least, I wish to thank my family for their tolerance, continued interest and moral support.

1.0. Background

Human immunodeficiency viruses (HIV) belong to the lentivirus genus of the family Retroviridae (van Regenmortel *et al.*, 2000). As the causative agent of the global AIDS epidemic, the impact of HIV type 1 (HIV-1) on humanity has been immense. At least 25 million people have died of HIV-1-related causes since the beginning of the epidemic, while it is estimated that approximately 34 million people were living with the virus by the end of 2010 (UNAIDS World AIDS Day Report, 2011). South Africa is the country with the largest number of people infected with HIV globally (UNAIDS World AIDS Day Report, 2011). In 2010, approximately 5.6 million people in South Africa were living with HIV, while 390 000 individuals were newly infected in 2010 alone (UNAIDS World AIDS Day Report, 2011)

Increased access to antiretroviral therapy is having a significant impact on HIV mortality rate; however, the only way to effectively control the HIV-1 epidemic is with an effective vaccine. In 2009, the first HIV vaccine trial to show efficacy was the RV144 trial in Thailand. This trial, which tested a canary pox vector expressing the *gp120*/*gag*/protease genes, followed by a gp120 protein boost, was shown to result in a 31.2% reduction in incidence of HIV infection (Rerks-Ngarm *et al.*, 2009). Two correlates of risk were identified: vaccinated volunteers with high levels of plasma IgG binding antibodies to V1/V2 had a 41% reduction in infection, with those with medium to high levels of these antibodies having a 75% reduction in HIV infection rate. However high level of plasma IgA antibody responses had the opposite effect with individuals with these responses having a 54% increased HIV infection rate. The infection rate in the placebo was the same as the IgA responder group suggesting that these IgA antibodies interfered with protection rather than enhancing HIV infection (Haynes *et al.*, 2012).

While the RV144 trial provided some insights, the protection was modest and the precise responses needed to effectively block infection remain unclear. Furthermore, as HIV-infected people are unable to eradicate the virus, it is impossible to use natural immunity to help elucidate these correlates of protection. However, it is still generally accepted that an effective vaccine should elicit both broadly cross-neutralizing antibodies and cytotoxic T-lymphocyte

(CTL) responses (Koup *et al.*, 2011; Richman *et al.*, 2009; Richman *et al.*, 2003). Major challenges in the vaccine development are, firstly, the design of immunogens that would elicit broadly cross-neutralizing antibodies that would be effective against the diversity of viral envelope glycoproteins; and secondly, the quality and breadth of cellular immune responses that correlate with protection are unknown (Burton *et al.*, 2004; Haynes *et al.*, 2012; McElrath and Haynes, 2010; Moore, 1997; Moore *et al.*, 200; Taylor *et al.*, 2008; van Regenmortel, 2011). This thesis will focus on understanding cellular immune responses associated with control of viral replication. Characterisation of how the virus adapts provides a sensitive tool to elucidating immune pressure and is one approach to identifying responses associated with control of infection that may be important for a vaccine to elicit. In this project, in a well-characterised subtype C cohort recruited from very early in infection, we perform a detailed characterisation of viral sequences to better understand early CTL pressure exerted on the virus, and to identify relevant epitopes for inclusion into vaccine immunogens.

This study focused on the Gag and Nef proteins of the virus as these are consistently shown to be the most immunodominant proteins (Gray *et al.*, 2009a; Kiepiela *et al.*, 2007) and as Gag is recognised earlier than any other protein (Payne *et al.*, 2010; Sacha *et al.*, 2007). Furthermore, targeting of Gag has been associated with better disease outcome (Kiepiela *et al.*, 2007; Miura *et al.*, 2009; Rolland *et al.*, 2008), although a study of Gray *et al.* (2009) found that Gag responses (as measured with IFN-γ ELISpot assay) in acute infection (<3 months post-infection) did not predict the subsequent course of HIV-1 subtype C infection.

This chapter will overview the genetic diversity of the global pandemic, the structure and function of Gag and Nef proteins, the natural history of disease, and finally the role of CTL mediated immune escape in influencing disease progression.

1.1 Genetic Diversity

Extensive and dynamic genetic diversity is a hallmark of HIV due to the error prone nature of the reverse transcriptase enzyme (Ho *et al.*, 1995; Keulen *et al.*, 1997; Op de Coul *et al.*, 2001; Roberts *et al.*, 1988), the genomic recombination capacity (Temin, 1993) and the high viral turnover (Ho *et al.*, 1995). A number of factors drive this diversity including the prolonged host immune selective pressure on the virus (Lemey *et al.*, 2007; Michael, 1999). This high genetic diversity impacts on diagnostic efforts, as well as on vaccine and drug development (Buonaguro *et al.*, 2007; Lal *et al.*, 2005; Peeters *et al.*, 2003). It also challenges our understanding of viral transmission and pathogenesis.

HIV has been classified into two types, HIV-1 and HIV-2 but this review will focus on HIV-1 only. HIV-1 strains identified throughout the world have been classified into four major phylogenetic groups, which are M (Major), O (Outlier), N (non-M, non-O) (Peeters and Sharp, 2000; Simon *et al.*, 1998) and a newly identified HIV-1 group P consisting of an O-like variant that is more closely related to SIV-gorilla than to group O (Plantier *et al.*, 2009; Vallari *et al.*, 2011). Group M viruses are responsible for the global pandemic and have been further subdivided into 9 distinct phylogenetic subtypes, namely A, B, C, D, F, G, H, J and K. Subtypes A and F have been classified into sub-subtypes (A1–A2, F1–F2), and there are over 48 circulating recombinant forms (CRFs) (www.hiv.lanl.gov) identified. Recombinant forms are generated when a person is infected with two or more subtypes, which recombine such that the viral genomes comprise of segments originating from different subtypes and are referred to as CRF if they are transmitted and become circulating viruses.

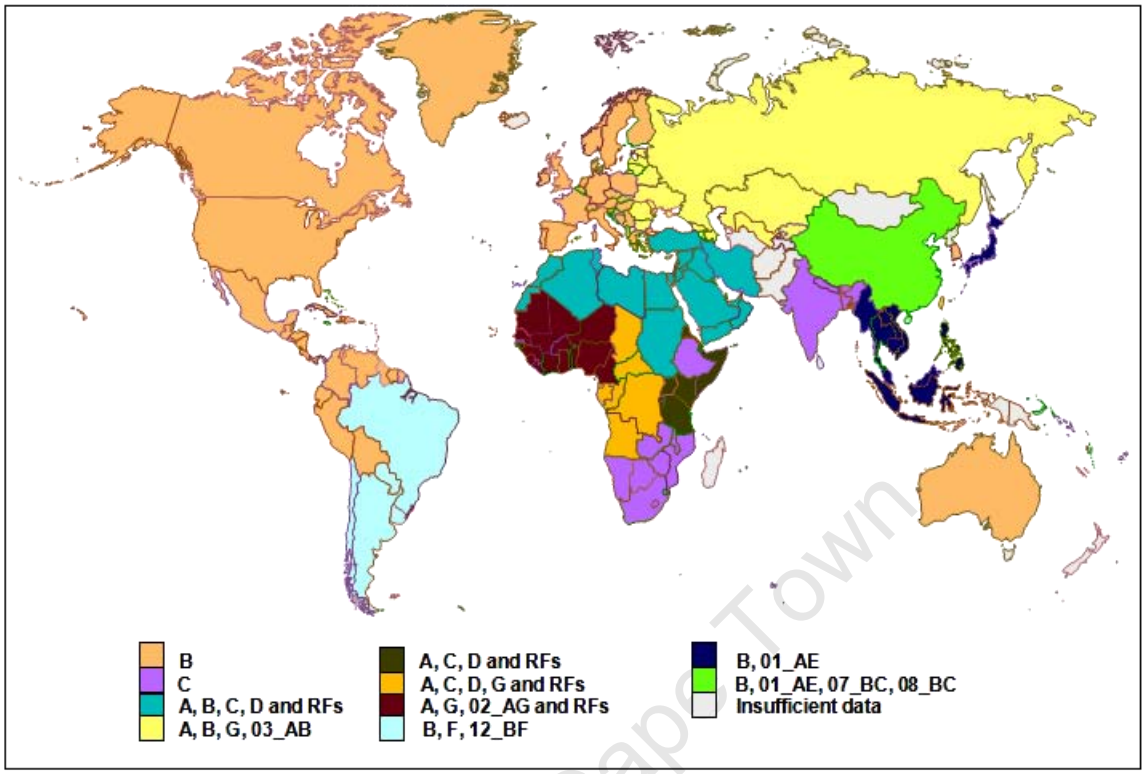


Figure 1.1 An overview map of global distribution of HIV-1 subtypes and recombinants (from Woodman and Williamson, 2009). The colours illustrate the dominant subtypes or recombinant forms circulating in these regions.

Within the group M, the average inter-subtype genetic variability for the *gag* gene is 15%, whereas it is 25-30% for the *env* gene (Janssens *et al.*, 1994; Kuiken *et al.*, 2000). While at the moment all the known HIV-1 subtypes are present in sub-Saharan Africa, their distribution in the region vary remarkably: subtype A and its recombinants predominating in west Africa; subtype A and D in East African countries; and subtype C in the southern and south eastern region of Africa, as well as Ethiopia (Hemelaar *et al.*, 2011). Subtype A is also distributed across Eastern Europe and Central Asia. Subtype B is found in America, Western Europe and East Asia, while Subtype D is most prevalent in North Africa and the Middle East. Subtype G predominates in West Africa and is also commonly found in Central Africa (Hemelaar *et al.*, 2011).

In southern Africa, Ethiopia and India, the epidemics are nearly exclusively caused by subtype C (Abebe *et al.*, 1997; Hemelaar *et al.*, 2011; Hussein *et al.*, 2000). Further, the BC CRFs_07 and _08 have become dominant in China. According to a report of Hemelaar *et al.* (2011), subtype C represents the largest number of HIV-1 infections worldwide accounting for nearly a half (48%) of the global infections.

1.2. Organisation of the HIV Genome

The integrated form of HIV-1 or the provirus is about 9.8kb in length (Muesing *et al.*, 1985) (Figure 1.2). A repeated sequence known as the long terminal repeats (LTRs) flank the central region of the proviral DNA that carries the genes encoding nine proteins (Gallo *et al.*, 1988). The HIV proteins can be divided into three classes, which are the major structural proteins, Gag, Pol, and Env; the regulatory proteins Tat and Rev; and the accessory proteins, Vpu, Vpr, Vif, and Nef. This thesis focuses on HIV-1 escape of cytotoxic lymphocytes (CTL) in Gag and Nef, and since escape has been shown to affect HIV-1 replication capacity (Boutwell *et al.*, 2009; Martinez-Picado *et al.*, 2006; Peyerl *et al.*, 2004; Rolland *et al.*, 2008; Wright *et al.*, 2010; Wright *et al.*, 2011), it has been important to have an understanding of the functional domains of these two proteins.

1.2.1 Structural protein – Gag

The viral Gag protein serves as the major structural component of the HIV-1 virion (Hunter, 1994). The *gag* gene is initially translated as a Gag polyprotein precursor ($Pr55^{gag}$), a gene product of 55-kilodaltons (kD) (Goff, 2007a). This polyprotein is cleaved post-translationally into matrix (MA or p17), capsid (CA or p24), nucleocapsid (NC) (Morellet *et al.*, 1992) and p6 - and two spacer peptides, SP1 (p2) and SP2 (p1) that separate the CA from NC and p6 respectively (Freed, 1998; Gelderblom, 1991; Gottlinger *et al.*, 1989; Kaplan *et al.*, 1994). In the 55 kDa Gag precursor, these proteins are arranged in the order of MA-CA-SP1/p2-NC-SP2/p1-p6 (Freed and Martin, 2007).

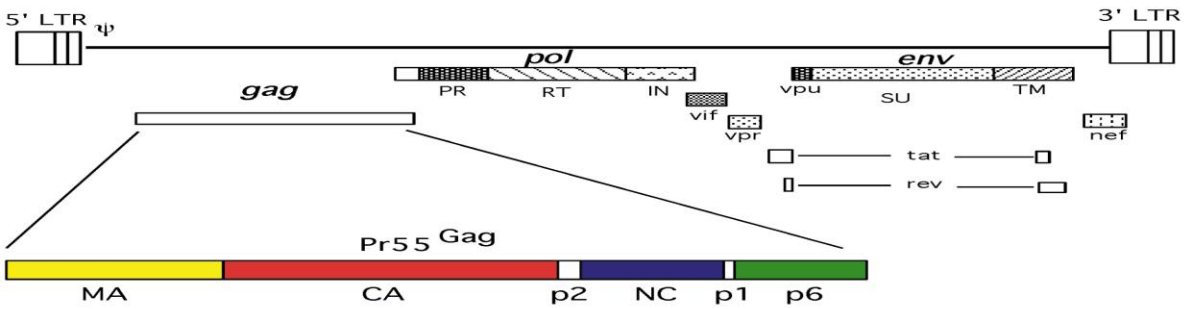


Figure 1.2 Organisation of the HIV genome with the *gag* open reading frame enlarged (from Freed, 1998) and the matrix (MA), capsid (CA), p2, nucleocapsid (NC), p1 and p6 proteins shown.

1.2.2 The Matrix protein (p17 or MA) and its role in the viral life cycle

The MA protein consists of about 132-amino acid residues and is post-translationally myristoylated at the N-terminus (Bryant and Ratner, 1990; Gottlinger *et al.*, 1989). The two major roles of this protein are in viral assembly and stabilization of the lipid envelope.

The role of MA in viral assembly is largely due to its membrane binding properties. Targeting to the plasma membrane depends upon the myristoylation of the MA domain of the Gag precursor at the N terminus (Pal *et al.*, 1990). The membrane-binding (membranotropic) signal is found in the first six amino acids of the protein (Figure 1.3) (Bryant and Ratner, 1990; Freed *et al.*, 1994; Spearman *et al.*, 1994). This portion regulates intracellular localization that targets $Pr55^{gag}$ to the plasma membrane where viral assembly and budding take place (Dong *et al.*, 2005; Jouvenet *et al.,* 2006; Lindwasser and Resh, 2004). This membrane-binding signal works together with a highly basic patch (residues 6–8) and the polybasic region (residues 26 and 32) (KKKYKLK) to promote tight Gag membrane binding (Ono *et al.*, 2000; Ono and Freed, 1999; Paillart and Gottlinger, 1999).

The central region of the MA (residues 41 to 78) has been shown to be essential for viral assembly, with single-amino-acid changes that prevent dimerization and viral like particle (VLP) formation being C57S and L64A (Morikawa *et al.*, 1998). These are located in a single

alpha helix 4 (residues 54 to 68), which forms part of a finger-like projection from one side of the molecule. Furthermore, the MA protein is essential for envelop incorporation into virions through association with the gp41 cytoplasmic tail (Bhattacharya *et al.*, 2006; Dorfman *et al.*, 1994; Murakami and Freed, 2000). Although the mechanism involved in envelope incorporation during viral assembly is not clearly defined, the following multiple mutations in the two groups, (R18G, K20I, R22A) and (K32I, H33L) in the N-terminal globular head region, and L49D were shown to affect Env incorporation (Bhatia *et al.,* 2007; Casella *et al.,* 1997; Davis *et al.,* 2006; Hill *et al.,* 1996). Furthermore, a study recently demonstrated that the MA E99V mutant viruses can also affect Env-incorporation and substitution in MA residue 84 were identified to restore (compensate for) defects due to E99V, suggesting a key role in HIV-1 Env incorporation of the C-terminal hydrophobic pocket of MA (Brandano and Stevenson, 2012).

Unlike other retroviruses, HIV is able to infect non-dividing cells due to the karyophilic signal properties of the MA, which the cellular nuclear import machinery recognizes (Bukrinsky *et al.*, 1993; Lewis *et al.*, 1992). The MA nuclear localization signal (NLS) has been mapped to the exposed basic patch (residues 24–31) on the globular head (Bukrinsky *et al.*, 1993; Zhou *et al.*, 1994). There are two NLS, encompassing residues 24–31 (KKKYKLK) and 110–114, (KSKKK), both of which consist of basic amino acids that introduce positive charge properties, crucial for the nuclear targeting (Bukrinsky *et al.*, 1993; Haffar *et al.,* 2000; von Schwedler *et al.*, 1994). The MA interacts with different nuclear transport receptors (cellular importins) and together with Vpr, integrase and NC, it regulates and contributes to nuclear import of the preintegration complex (PIC) (Bukrinsky, 2004; Butterfield-Gerson *et al.*, 2006). Furthermore, mutations in its nuclear localization signal affect postchromatin-binding events and impair viral infectivity without significantly changing the nuclear transport of PIC, suggesting that MA is involved in viral DNA integration into the host cell chromosome (Mannioui *et al.*, 2005).

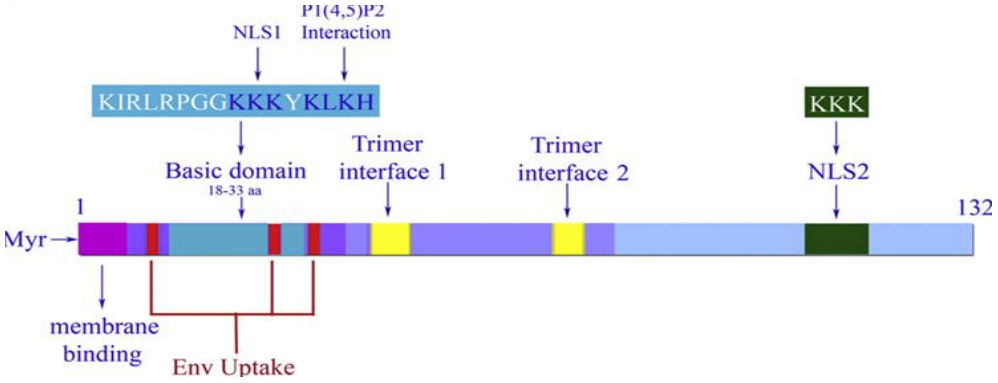


Figure 1.3: Schematic presentation of the MA and known functional sites (from Fiorentini *et al.*, 2010). The N-terminal myristoylation site (N-Myr), nuclear export signal (NES), and nuclear localization sequences (NLS) and Env uptake region are also shown. Interfaces thought to be involved in trimerisation of p17 are shown.

On the other hand however, to ensure that components required for virion assembly are available in the cytoplasm, a nuclear export signal (NES) is present which targets the Gag–RNA complex to the cytosol and is required to counteract the NLS (Dupont *et al.*, 1999). Mutations in MA, K18, K20 and R22 that block its NES result in the abnormal accumulation of Gag and viral genomic RNA in the nucleus (Dupont *et al.*, 1999).

1.2.3 The Capsid protein (p24)

The HIV-1 capsid (CA) forms a shell or a distinctive conical core surrounding the viral RNA genome and core-associated proteins in the mature virion (Briggs *et al.*, 2004; Carlson *et al.*, 2008; Tang *et al.*, 2001). During viral assembly and budding, the CA incorporates the Gag-Pol precursor into capsid of virions (Chien *et al.*, 2006; Smith *et al.*, 1993), which is an important event for the recruitment of reverse transcriptase, integrase and protease into viral particles. The N-terminal region of the CA functions in virion maturation and incorporation of the cellular protein cyclophilin A (CypA), while the C-terminal region contributes to Gag–Gag interactions.

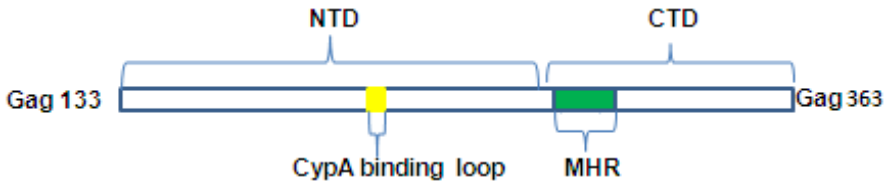


Figure 1.4 Schematic representation of the Capsid protein (CA), showing the position of cyclophilin A binding loop (Yoo *et al.*, 1997) and the major homology region (Chang *et al.*, 2007) (re-drawn from Adamson and Freed, 2007). The N-terminal domain (NTD) and C-terminal domain (CTD) are illustrated. Numbers are Gag amino acid residues, HXB2 positions.

The CA protein consists of two interdependently folded domains, an N-terminal domain (NTD) or the ‘‘core’’ domain (Gag residues 133-278) that mediates hexamer formation and a C-terminal domain (CTD) (Gag residues 279-363) or the ‘‘dimerization’’ domain (Briggs *et al.*, 2003; Gamble *et al.*, 1997; Gitti *et al.*, 1996; Li *et al.*, 2000). Studies by NMR spectroscopy (Gitti *et al.*, 1996) and X-ray crystallography (Gamble *et al.*, 1996; Momany *et al.*, 1996; Mortuza *et al.*, 2004) show that the core domain is highly helical and consists of seven α-helices, a short two stranded β-hairpin, and an exposed loop that serves as the binding site for CypA (Figure 7). The β-hairpin (Gag residues 133-145) conformation is thought to be stabilized by a salt bridge between the N-terminal CA amine of P133 and D183 in helix 3 (Fitzon *et al.*, 2000; Gitti *et al.*, 1996; Kelly *et al.*, 2006; Tang *et al.*, 2002; von Schwedler *et al.*, 1998). NMR studies of Gitti *et al.* (1996) suggested that residues P133–N137 in the N-terminal of the mature CA pair with residues Q141–Q145 to form the antiparallel strands of a β-hairpin that packs against helix 6 (Figure 1.5). Furthermore, although differences exist, in both the mature and immature CA, residues H144–I147 interact with the globular portion of the N-terminal CA domain, T180 (helix 3), I243, I247 and M250 in Helix 6 (Tang *et al.*, 2002). Additional interactions between the β-hairpin and the rest of N-terminal domain of the CA include hydrogen bond between the backbone N-H of I134 and the carbonyl of G178. These bonds are further supported by the hydrophobic packing between the side chains of the β-hairpin I134, T180 (helix 3) and, l247 and M250 (helix 6).

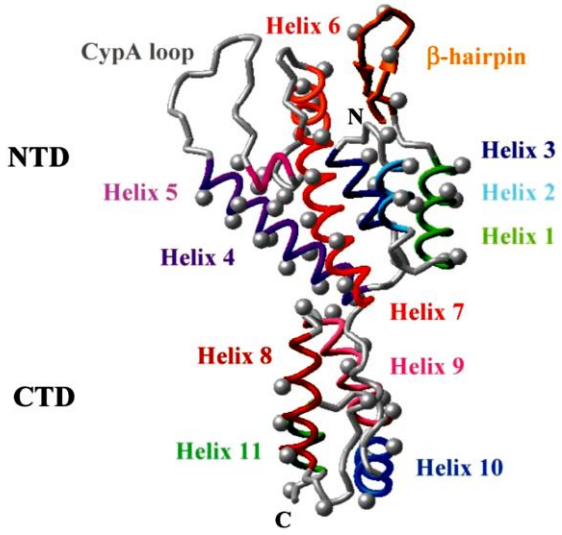


Figure 1.5: Structural model for the HIV-1 CA protein based upon crystal structures of the protein's NTD and CTD (von Schwedler *et al.*, 2003). Residues forming the seven a-helices are S148–K162, S165–S176, T180–V191, H194–H216, R232–A237, T242–T251 and V258–S278 respectively, and are packed together in a flat and triangular shape, while residues P217–P231, which include the CypA-binding site, form a conformationally flexible loop (Tang *et al.*, 2002). Residues forming the CTD α-helices (8 – 11) are P292-L304, S310-V323, N327-L334, and A341-A349. Numbers are Gag amino acid residues, HXB2 positions.

1.2.3.1 The Capsid and cyclophilin A (CypA) binding

The earliest studies on murine leukemia virus (MLV) demonstrated that a host antiretroviral factor, *Fvl,* restricted the incoming (post-entry) viral capsid and prevented infection of mouse cells in a CA-dependent manner or dependent on mouse background (Best *et al.*, 1996; Friend, 1957; Hilditch *et al.*, 2011; Schindler *et al.*, 1981). Replication of HIV in humans, but not in non-human primates, suggested that similar factors could be involved in restricting lentiviral infections in a CA-dependent manner (Besnier *et al.*, 2002; Cowan *et al.*, 2002; Hatziioannou *et al.*, 2005; Hofmann *et al.*, 1999; Munk *et al.*, 2002; Towers *et al.*, 2003). Studies subsequently elucidated the lentiviral restriction factor in non-human primates as TRIM5 (a member of tripartite motif family of proteins) related genes (Nisole *et al.*, 2004; Sayah *et al.*, 2004; Stremlau *et al.*, 2004; Wu *et al.*, 2006). It was shown that CypA decreases

HIV infectivity in the Old World monkey (OWM) cells by interacting with a host restriction factor, TRIM5α (Berthoux *et al.*, 2005; Keckesova *et al.*, 2006; Stremlau *et al.*, 2006a).

CypA is a cytosolic protein that catalyzes the cis–trans isomerization of proline residues (Fischer *et al.*, 1989). Virions incorporate CypA via interactions with a proline-rich stretch in the NTD of the capsid (Braaten *et al.*, 1997). Although the proline rich loop in the viral capsid domain directly binds to the hydrophobic binding pocket of CypA (Braaten *et al.*, 1997; Endrich *et al.*, 1999), residues that are critical for CypA binding in the NTD of the CA are Gag - G221 and P222 (Braaten *et al.*, 1996; Gamble *et al.*, 1997). In humans, CypA is required for maximum viral infectivity and virions deficient in CypA show defects early in postinfection events (Braaten *et al.*, 1996; Luban, 2007; Steinkasserer *et al.*, 1995). The observation that CA point mutants close to the CypA binding site such as CA A224E or G226D appear to lead to restriction of HIV-1 in human cells further supports the idea that CypA is required for viral infectivity (Sokolskaja *et al.*, 2004). Cyclophilin has been implicated in promoting the uncoating stage of the CA (Briones *et al.*, 2010; Saphire *et al.*, 2002). A more recent study has demonstrated that HIV-1 capsid-cyclophilin interactions is required for the nuclear import pathway, integration targeting and in replication maximization (Schaller *et al.*, 2011).

Human TRIM5α_h is substantially less potent than TRIM5α_rh from the old world monkeys although it has been shown to have some *in vitro* anti-HIV activity (Kaumanns *et al.*, 2006; Stremlau *et al.*, 2004; Yap *et al.*, 2004). It has been shown that TRIM5α_h inactivates cyclophilin binding to the incoming retroviral capsid and accelerates viral uncoating, which inhibits reverse transcription and transportation of the viral genome to the nucleus (Sebastian and Luban, 2005; Stremlau *et al.*, 2004; Stremlau *et al.*, 2006a; Ulm *et al.*, 2007). Several studies have also shown that the interaction between host cyclophilin and HIV-1 CA within the host cell modulates the binding of host restriction factors like TRIM5α (Berthoux *et al.*, 2005; Hatziioannou *et al.*, 2005; Sokolskaja *et al.*, 2004). However, preventing the interaction of HIV-1 incoming cores with CypA reduces viral infectivity independently of TRIM5 expression (Keckesova *et al.*, 2006; Stremlau *et al.*, 2006b). Furthermore, other studies have

shown that TRIM5α_h does not recognise HIV-1; but HIV-1 becomes restricted by something else in the absence of CypA activity (Keckesova *et al.*, 2006; Sokolskaja *et al.*, 2004).

CypA also interacts with both p6 and Vpr but the biological significance of this interaction in the HIV life cycle is still not completely understood, although it is thought that CypA modulates host activities to favour viral replication (Solbak *et al.*, 2010)

1.2.4. The Major homology region

The HIV-1 major homology region (MHR) – (CA 153–172) or (Gag 285 – 304) is located near the end of the N-terminus of the capsid or C domain and consists of a stretch of 20 residues (Figure 1.6), that are highly conserved across retroviral genera (Gamble *et al.*, 1997). The most highly conserved residues Q287, G288, E291, and R299 were shown to form a network of hydrogen-bonds that stabilize a conformation required for the Gag–Gag interactions for Gag particle assembly (Campos-Olivas *et al.*, 2000; Gamble *et al.*, 1997). Furthermore, other conserved hydrophobic residues, 293F, 296Y, 297V, 301Y, and 304L lie on one face of helix 1 (Clish *et al.*, 1996), where they contribute to the hydrophobic core of the protein (Gamble *et al.*, 1997). Mutations in Q287 and E291 block HIV replication (Mammano *et al.*, 1994). Mutations in residues K290 and Y296 respectively blocked virus assembly or production of virus-like particle (VLP) *in vitro* (Mammano *et al.*, 1994; von Schwedler *et al.*, 2003), while substitutions at K290, F300 or E307 significantly diminished VLP production (Ganser-Pornillos *et al.*, 2007).

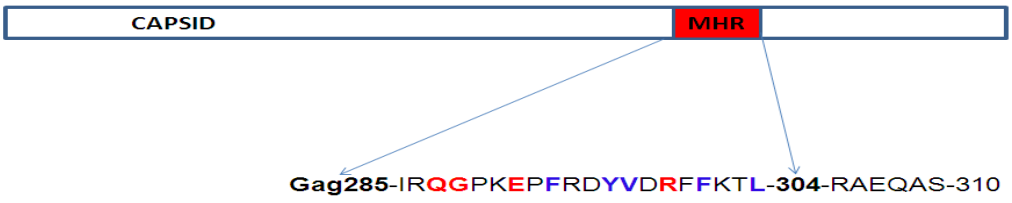


Figure 1.6: Schematic representations of HIV-1 Capsid with MHR, HIV-1 subtype C, 2004 consensus sequence within the CA (20 residues) (Drawn based on Gamble *et al.* 1997). The most conserved MHR residues (Q287, G288, E291, and R299) are shown in red, while the

conserved hydrophobic residues (293F, 296Y, 297V, 301Y, and 304L) are shown in blue. Numbers are Gag - HXB2 amino acid positions.

1.2.5. The Nucleocapsid

The 7kDa HIV Nucleocapsid (NC) is a highly basic protein that is characterized by the presence of two zinc-finger motifs, Cys-X2-Cys-X4-His-X4-Cys (CCHC, X = variable HIV-1 amino acid), which are critical to its functions (Bess *et al.*, 1992; South and Summers, 1993; Summers *et al.*, 1990). NC is found tightly associated with the viral RNA in the mature virion (Meric *et al.*, 1984). Mutations that disrupt the NC zinc fingers affect genomic RNA encapsidation and virus infectivity (Gorelick *et al.*, 1988; Gorelick *et al.*, 1990). These two major mutations in the NC are H23C and H44C (Figure 1.7).

The NC binds, encapsulates and protects viral RNA. However, it is also thought to be involved in RNA dimerization, Gag interactions, membrane binding, reverse transcription and preintergration complex stability (Freed, 1998). Furthermore, the NC employs its nucleic acid chaperone activity to anneal the aminoacyl tRNA Synthetase (tRNALys, 3) primer to the viral genome during reverse transcription (Kleiman *et al.*, 2004; Rein *et al.*, 1998).

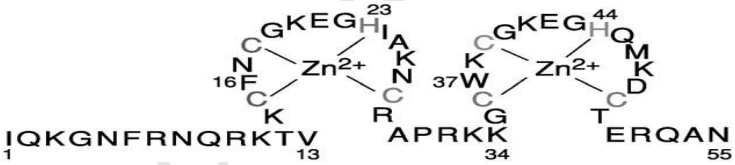


Figure 1.7: Primary sequence of HIV-1 NC - NL4-3 isolate – showing (in grey) the Cysteine and Histidine residues that chelate zinc (Levin *et al.*, 2005)

1.2.6 The viral late domain (HIV-1) – p6

In HIV-1, a highly conserved motif, Pro-[Thr/Ser]-Ala-Pro (PSAP)/(PTAP) and a conserved leucine motif $LYPX_nL$ (n = 1-3 variable residues) located near the N and C terminus of p6, respectively were identified to be critical in releasing virus particles after budding from the plasma membrane (Fisher *et al.*, 2007; Freed, 2002; Strack *et al.*, 2003). The role of the p6

motif in viral budding is to interact with cellular endosomal sorting complexes required for transport (Carlson *et al.*, 2008), a machinery that catalyze the scission of membrane necks (Wollert *et al.*, 2009) in the biogenesis of multivesicular bodies (MVBs) (Hurley, 2008) and a key step in virus release (von Schwedler *et al.*, 2003). The ESCRT machinery consists of four protein complexes, ESCRT-0,-I,-II and III (Hurley, 2008; Saksena *et al.*, 2007) that are involved in the recognition and sorting of membrane proteins to the internal vesicles of MVBs/late endosomes that are subsequently degraded in lysosomes (Bieniasz, 2009; Goff, 2007b; Malim and Emerman, 2008). Budding involves a sequential recruitment of these multiprotein ESCRT to endosomal membranes, which leads to release of endosomal vesicles into multivesicular bodies (MVBs) and incorporation of ubiquitinylated membrane protein cargoes into the inwardly budding vesicles (Bieniasz, 2009; Goff, 2007b; Raiborg and Stenmark, 2009). The process is initiated by HRS, a component protein of ESCRT-0 that links the ubiquitinylated membrane protein cargoes with the ESCRT-I components Tsg101 (tumor susceptibility gene) (Raiborg and Stenmark, 2009). In HIV budding, the p6 PTAP motif mimics the PSAP motif in HRS and thus binds with the Tsg101 component while the second p6 late domain motif LYPXnL binds with ALIX, a component protein in ESCRT- III (Garrus *et al.*, 2001; Pornillos *et al.*, 2003; Strack *et al.*, 2003). A cellular restriction factor BST-2/tetherin has been shown to hamper HIV-1 release by physically tethering fully formed mature particles to the plasma membrane of infected cells (Mangeat *et al.*, 2009; Neil *et al.*, 2008; Perez-Caballero *et al.*, 2009). A study has recently shown that the ESCRT-0 component HRS interacts with HIV accessory protein Vpu to downregulate the cellular host restriction factor BST-2/tetherin (Janvier *et al.*, 2011).

1.2.7 Structural and Biochemical Characteristics of Nef

HIV-1 Nef is a myristoylated accessory protein, about 27 kDa and 206 to 210 amino acid residues in length (Allan *et al.*, 1985; Niederman *et al.*, 1993). The N-terminal contains a hydrophobic myristoylation motif (MGxxxS), which promote its association with cellular membranes. Structural studies have revealed that Nef has three major elements: a flexible N-terminal region (residues 1-80), consisting of acidic region (residues 62- 65) and a type II polyproline helix (residues 69-78); a folded C-terminal core (residues 81-147 and 181-202);

and an exposed, flexible loop (residues 148-180) centred around a dileucine motif (residues 160-165) bounded by negatively charged di-glutamic acid (residues 154-155) and di-aspartic acid (residues 174-175) regions (Arold *et al.*, 1997; Lee *et al.*, 1996).

1.2.8 Major roles of HIV-1 Nef in disease progression

The *nef* gene is unique to primate lentiviruses and absent in the other retroviruses (Piguet and Trono, 1999). The protein Nef was originally thought to be a negative factor that inhibited viral replication (Cheng-Mayer *et al.*, 1989; Luciw *et al.*, 1987). However, later studies demonstrated the importance of Nef for efficient viral replication and pathogenesis *in vivo*. In rhesus macaques, *nef*-deleted SIVmac239 displays attenuated viral replication and pathogenicity (Kestler *et al.*, 1991; Khan *et al.*, 1998). Virus isolates from several long-term non-progressors of HIV-1 infection were found to possess mutations and deletions in *nef* (Deacon *et al.*, 1995; Kestler *et al.*, 1991; Kirchhoff *et al.*, 1995). Nef is incorporated into virions and associates with the viral core (Bukovsky *et al.*, 1997; Forshey and Aiken, 2003; Kotov *et al.*, 1999; Simmons *et al.*, 2001) and enhances infectivity (Qi and Aiken, 2008). Nef is required for maintenance of high viral loads but although macaques infected with *nef*-deleted SIV have attenuated infection, these viruses do finally induce AIDS symptoms (Desrosiers, 1999; Sawai *et al.*, 2000; Sugimoto *et al.*, 2003). Studies have demonstrated *in vitro* that Nef is a multi-functional protein and is involved in three major activities:

a) Downregulation of cell surface molecules such as CD4
b) Downregulation of major histocompatibility complex class I (MHC-1)
c) Modulation of T cell activation state

1.2.8.1 CD4 Downregulation

Nef downregulates CD4 expression by firstly, redirecting CD4 molecules from the trans-Golgi network (TGN) to the endosomal compartment, where they are targeted to the lysosome for degradation; and secondly, by accelerating the internalization of CD4 molecules that have reached the cell surface (Bonifacino and Traub, 2003; Mangasarian *et al.*, 1997; Piguet *et al.*,

1999). The binding of the CD4 cytosolic tail requires a di-leucine motif (residues 413-414) and a surface pocket in Nef (residues 57-58) that contains a tryptophan-leucine (WL) motif (Aiken *et al.*, 1994; Bonifacino and Traub, 2003; Mangasarian and Trono, 1997). Binding between Nef and CD4 appears to be driven, in part, by myristoylation of the Nef protein, which targets and concentrates Nef to membranes, the primary site of its action (Bentham *et al.*, 2003; Bentham *et al.*, 2006; Harris and Neil, 1994), where it is brought in closer proximity to CD4. The presence of basic residues near the N-terminus of Nef supports this subcellular localization, as it may serve to anchor the protein among acidic phospholipids common to cellular membranes (Dennis *et al.*, 2005).

HIV-1 Nef interacts with the adaptor protein complex (AP) of clathrin coated pits (CCP/AP) via a dileucine-based motif (LL165) in the C-terminal region of the viral protein (Bresnahan *et al.*, 1998; Craig *et al.*, 1998; Piguet *et al.*, 1998). Adaptor protein complexes are a family of heterotetramers that recognize cytoplasmic sorting motifs in transmembrane proteins for packaging into clathrin-coated vesicles for intracellular transport (Ohno, 2006; Traub, 2009). Nef mutants defective for AP binding are not able to accelerate CD4 internalization, which demonstrates that adaptor complexes are a major downstream partner of Nef for CD4 endocytosis (Piguet *et al.*, 1999). It also interacts with a subunit of the V-ATPase that could facilitate its association with the endocytic apparatus (Geyer *et al.*, 2002; Lu *et al.*, 1998).

Another downstream cellular partner of Nef in the down regulation of CD4 is Coatomer (or COP I). COP I is a large cytosolic protein complex which forms a coat around vesicles budding from the Golgi apparatus (Cosson *et al.*, 1996; Orci *et al.*, 1986). The COP I coatomer is involved in the ER-Golgi transport, and Nef accomplishes the down regulation of CD4 by connecting the receptor via its acidic dipeptide (Nef, 154-EE-155) to the β-COP subunit of COP I coatomer and prevents the cell surface recycling of CD4 that is instead targeted to a degradation pathway in endosomes (Oldridge and Marsh, 1998; Piguet *et al.*, 1998).

1.2.8.2 MHC class I downregulation

Nef downregulates MHC-1 through a pathway different from that of CD4 (Blagoveshchenskaya *et al.*, 2002; Mangasarian *et al.*, 1999), although the mechanisms involved in MHC-I down-regulation are still controversial. Nef attenuates MHC-I cell surface expression by accelerating endocytosis of class I molecules from the plasma membrane to the trans-Golgi network (TGN). MHC class I is internalized via a clathrin-independent mechanism and is sequestered in the trans-Golgi network (TGN), and eventually degraded (Greenberg *et al.*, 1998; Le Gall *et al.*, 2000). MHC-I downregulation induced by Nef depends on a sorting protein, PACS-1 (phosphofurin acidic cluster sorting protein-1) (Blagoveshchenskaya *et al.*, 2002; Hung *et al.*, 2007; Piguet *et al.*, 2000). In this model, initiation of MHC-1 downregulation occurs when PACS-1 or AP-1 binds to the Nef 62-EEEE-65 acidic cluster, which sorts Nef to the Trans Golgi Network (TGN) (Crump *et al.*, 2001; Piguet *et al.*, 2000). PACS-1 protein is a binding partner for furin, an enzyme that cycles between the TGN and endosomes (Wan *et al.*, 1998). Through its highly conserved proline motif (72PxxP75 [X= any amino acid]), Nef recruits the TGN-localized Src family tyrosine kinase (SFK) which activates a phosphatidyl inositol 3-kinase (Hung *et al.*, 2007) that promotes the assembly of a Nef-SFK-ZAP-70/Syk-PI3Kcomplex (Hung *et al.*, 2007). This complex is translocated by PACS-1 from early endosomes to the plasma membrane where PI3K generates PIP3-containing lipids (Atkins *et al.*, 2008; Peterlin and Trono, 2003; Piguet *et al.*, 2000). PIP3-containing lipids recruits the ARNO guanine nucleotide exchange factor (GEF) to the plasma membrane which activates ARF6 that finally leads to accelerated MHC-I endocytosis through an AP-2/clathrin independent mechanism (Mangasarian *et al.*, 1999; Padron *et al.*, 2003). However, Larsen *et al.* (2004) found no effect on the down-modulation of MHC class I due to mutations that provide more specific perturbations in the GDP-GTP cycling of ARF6 (Larsen *et al.*, 2004). Furthermore, another study found that inhibiting PACS-1 had no effect on the down-regulation of HLA-A2 and the localization of other proteins containing acidic cluster motifs (Lubben *et al.*, 2007).

Other models have suggested that Nef connects the MHC-I cytoplasmic tail through the di-leucine (164-LL-165) motif to subunits of the clathrin adaptor protein complex 1 (AP-1 complex) (Roeth *et al.*, 2004; Williams *et al.*, 2002). However, mutation of this motif does not impair MHC-I down-regulation, which suggests that AP-1 recruitment is a cooperative process, involving determinants present in both Nef and MHC-I cytoplasmic tail (Noviello *et al.*, 2008; Roeth *et al.*, 2004). The structural formation of the Nef–MHC-I–AP1 complex has been recently confirmed by a study of Jia *et al.* (2012) and supports the idea that AP-1 recruitment is a cooperative process, involving determinants present in both Nef and MHC-I cytoplasmic domain. In this model, HIV-Nef down-regulation of MHC-1 requires the first α-helix (aa 17 – 26) containing M-20, an acidic cluster (62-EEEE-65), a proline-rich (PxxP)3 repeat at the junction of the N-terminal loop and the folded core, and residue Asp123 (Atkins *et al.,* 2008; Blagoveshchenskaya *et al.,* 2002; Greenberg *et al.,* 1998; Hung *et al.,* 2007; Mangasarian *et al.,* 1999; Noviello *et al.,* 2008; Roeth *et al.,* 2004). Nef interacts with the MHC-1 cytoplasmic domain through the μ1 subunit of AP-1 that encompasses the cargo-recognition site of μ1 and the proline-rich strand of Nef (Jia *et al.,* 2012).

1.2.8.3 Modulation of T cell activation state

Quiescent T cells do not support optimal reverse transcription, integration, and expression of the HIV-1 genome due to low levels of nucleotides, ATP, and nuclear transcription activators (Coiras *et al.*, 2009; Stevenson, 2003). Several studies suggest that Nef binding to T cell receptor (TCR) induces an up-regulation of FAS ligand, resulting in death of cytotoxic T cells that come in contact with virus-infected cells (Piguet and Trono, 2001; Xu *et al.*, 1997). Moreover, Nef promotes the production of chemokines like the macrophage inflammatory proteins (MIP) 1α and 1β, and other mediators of T cell activation, which induce the chemotaxis and activation of nearby resting T lymphocytes, facilitating rapid infection (Dai and Stevenson, 2010; Fackler and Baur, 2002; Olivetta *et al.*, 2003; Swingler *et al.*, 1999).

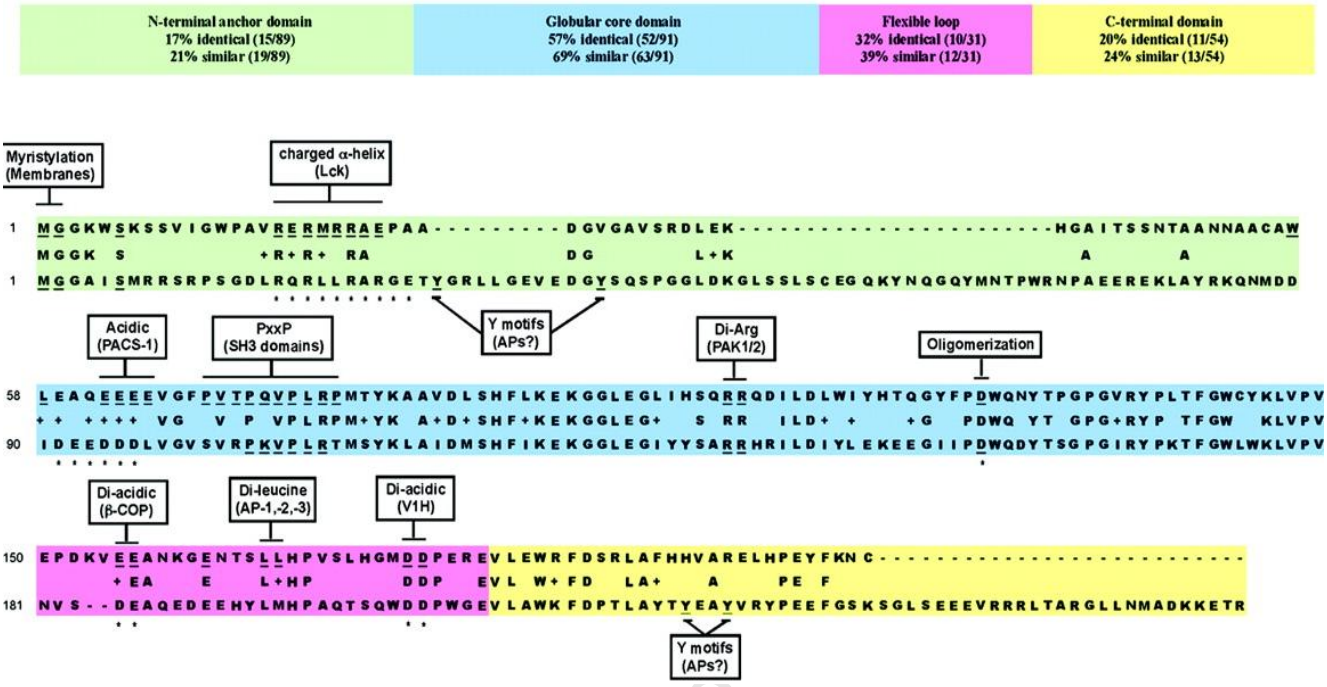


Figure 1.8 HIV-1 and SIV Nef-sequence alignment and protein-protein interaction domains (from Roeth and Collins, 2006). Boxes indicate protein motifs and their putative cellular binding partners (in parentheses). Residues important for protein interactions are underlined. The asterisks indicate residues in SIV_{mac239} Nef that are homologous to important residues in HIV_{NL4-3} Nef that are yet to be tested for functional homology.

1.3 HIV-1 Life Cycle

The HIV virion is enclosed in a roughly spherical outer lipid bilayer membrane 120–200 nm in diameter (Briggs *et al.*, 2003), which is studded with trimeric clusters of the transmembrane Envelope (Env) gp41 and the surface Env glycoprotein gp120. The core consists of two single-stranded genomic RNA in complex with the nucleocapsid (NC) protein, viral protein R (Vpr) and the viral enzymes, reverse transcriptase (RT) and integrase (IN) enclosed in a capsid shell protein (Adamson and Freed, 2007; Ganser-Pornillos *et al.*, 2008).

During infection, HIV-1 surface envelope (Env) glycoprotein gp120 binds to the cellular receptor CD4 and subsequently interacts with a chemokine coreceptor, usually CCR5 or CXCR4 (Berger *et al.*, 1999; Doms, 2000). Binding of a coreceptor triggers conformation

changes in both gp120 and the transmembrane envelope gp41, which causes fusion and the liberation of the viral core into the cytoplasm of the target cell (Doms, 2000; Melikyan, 2008).

During the uncoating of the viral core, CA is lost, while at least some MA, as well as NC, the pol-encoded enzymes and the viral protein R (Vpr) are thought to be retained as part of a high-molecular weight complex (Freed, 1998). The RNA genome is reverse transcribed into a double-stranded DNA copy that is transported as the high-molecular-weight complex, now known as the preintegration complex, into the nucleus where the enzyme IN facilitates the stable integration into the host cell genome (Delelis *et al.*, 2008; Harrich and Hooker, 2002; Suzuki and Craigie, 2007; Vandegraaff and Engelman, 2007). The integrated viral DNA, known as the provirus, serves as the template for the generation of full-length progeny viral RNA and a number of spliced mRNA transcripts that are translated in the cytoplasm to yield several major structural proteins: the Gag polyprotein precursor, the Gag-Pol polyprotein precursor and the Env glycoprotein precursor, gp160 (Freed, 2001). The *env* gene product, the precursor protein, (gp160) is glycosylated within the endoplasmic reticulum. These protein components, together with full-length viral genomic RNA, are each transported to the site of virus particle assembly at the plasma membrane (Adamson and Freed, 2007).

During or shortly after reaching the plasma membrane, the Gag precursor directs the assembly by recruiting two copies of the single-stranded viral genomic RNA and interacting with the Gag-pol precursor into structures (Briggs *et al.*, 2003; Derdowski *et al.*, 2004; Sandefur *et al.*, 2000). This assembly of Gag protein complex causes membrane curvature, which eventually buds off as immature virions (Gould *et al.*, 2003; Ono and Freed, 2001). During budding, the viral Env glycoproteins are incorporated into the nascent particles. Other associates in these structures are viral enzymes [(reverse transcriptase (RT), integrase (IN), and protease (PR)], which are incorporated within the nascent virion as part of the precursor gag-pol polyprotein (Smith *et al.*, 1993), the cellular tRNAlys3 primer and cellular compounds (De Guzman *et al.*, 1998; Greene and Peterlin, 2002; Kleiman and Cen, 2004; Zimmerman *et al.*, 2002).

The genomic viral RNA and Vpr are encapsidated through noncovalent interactions with Pr55 gag subdomains p6 and nucleocapsid (NC), respectively (Aldovini and Young, 1990; Darlix

et al., 1995; Lavallee *et al.*, 1994; Sakaguchi *et al.*, 1993). In the process of, or shortly after budding, the viral protease (PR) is activated to cleave the Gag and Gag-pol polyprotein precursors to the mature Gag and Pol proteins (Briggs *et al.*, 2003). The enzymatic cleavage leads to core condensation and the generation of mature, infectious virions which are now capable of initiating a new round of infection.

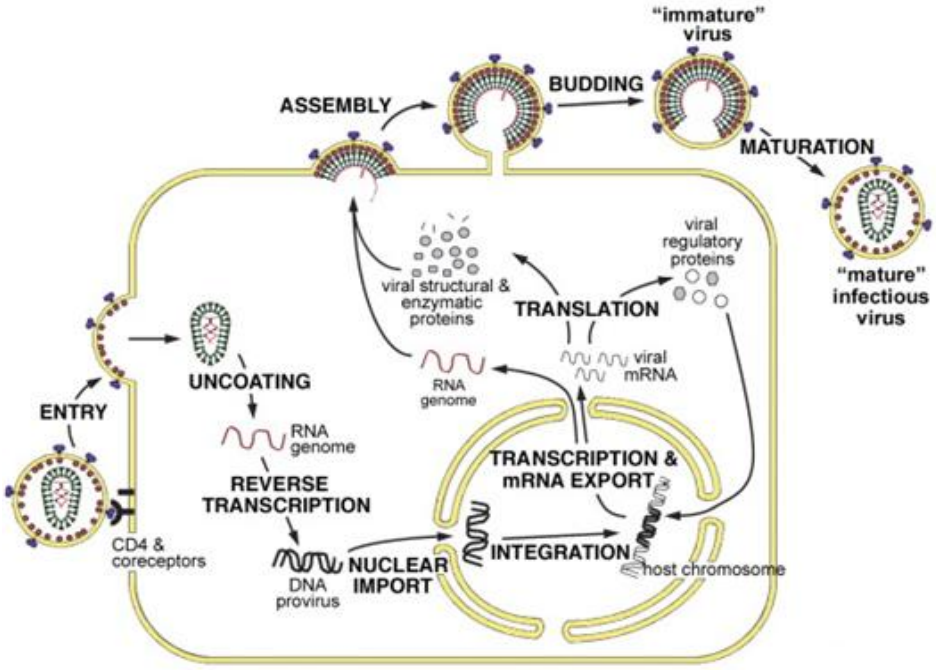


Figure 1.9 Schematic representation of the HIV-1 replication cycle (Ganser-Pornillos *et al.*, 2008).

1.4 Natural History of HIV Infection

Immediately following HIV-1 exposure and transmission, there is a period of about 7 to 21 days, known as the eclipse phase, when viral RNA is not detectable in the plasma (Fiebig *et al.*, 2003; Kleinman *et al.*, 2009; Zhu *et al.*, 2003). When HIV-1 RNA reaches a concentration of 1 to 5 copies per millilitre in plasma, the virus can be detected with quantitative nucleic acid amplification (Palmer *et al.*, 2003) and there is generally a uniform and orderly timing of the appearance of viral and host markers of infection (Fiebig *et al.*, 2003). The chronologic order of appearance in serum of these markers is viral RNA, p24 antigen, and anti-HIV antibody, and this order may distinguish acute from early infection (Fiebig *et al.*, 2003). The

length and definition of acute infection varies in individuals but for the purpose of this thesis will be defined as the 3-month (100 days) period following HIV infection during which time individuals will have undergone complete seroconversion (Kassutto and Rosenberg, 2004; Lyles *et al.*, 2000; McMichael *et al.*, 2010). Events in this phase are thought to play an important role in subsequent disease progression and clinical outcome (Kelley *et al.*, 2007; Piatak *et al.*, 1993). Studies in non-human primates have shown that following transmission, the virus crosses the mucosal barrier within hours (Hu *et al.*, 2000) and expands locally at the foci of infection. Within a few days to a week after infection, the virus is disseminated throughout the body and lymphoid tissue, seeding even the central nervous system (Daar, 1998; Haase, 2010; Miller *et al.*, 2005). In the lymphocyte-rich compartments throughout the body, particularly the spleen and the gut-associated lymphoid tissue (GALT), the virus encounters high densities of primary target CD4+ T cells causing a massive burst of viral replication. These events result in irreversible and severe depletion of memory CD4+ T cells in all tissues, predominantly from the mucosal surfaces (Brenchley *et al.*, 2004; Guadalupe *et al.*, 2003; Mattapallil *et al.*, 2005; Veazey *et al.*, 1998). Additionally, viral latency or the silent integration of HIV-1 DNA into the genomes of resting CD4 T cells is established (Chun *et al.*, 1998). The virus load rises rapidly to peak levels in 21–28 days after infection and coincides with the release of inflammatory cytokines, the major ones being interferon-α and interleukin-15 (Stacey *et al.*, 2009). A crucial period in acute infection occurs between transmission and peak viremia, just before the occurrence of massive destruction of CD4+ T cell and the establishment of viral reservoirs. This has been referred to as the window of opportunity for an HIV-1 vaccine where it could potentially control viral replication and prevent extensive CD4+ T cell depletion and limit generalized immune activation (McMichael *et al.*, 2010).

Following infection, the peak viremia declines to a much lower steady state known as the viral set point (Fiebig *et al.*, 2003; Kaufmann *et al.*, 1998; Lindback *et al.*, 2000; Lyles *et al.*, 2000; Mellors *et al.*, 1996). Viral set-point is one prognostic marker of disease progression with measurements around 12 months shown to be the most reliable predictor (Lyles *et al.*, 2000).

Around peak viremia, some patients may experience symptoms (Lindback *et al.*, 2000). Studies have associated more symptomatic acute infection with faster rates of disease

progression (Kelley *et al.*, 2007; Piatak *et al.*, 1993; Vanhems *et al.*, 2000). In subtype B infection, the peak HIV-1 viral load has ranged from 10^5 - 10^8 RNA copies/ml, lasting for about 2 to 3 weeks and finally dropping to set-point in 4-6 month post infection (Daar *et al.*, 1991; Fiebig *et al.*, 2003; Kaufmann *et al.*, 1998; Little *et al.*, 1999; Lyles *et al.*, 2000). Similar observations on the dynamics of primary infection in non–subtype B epidemics in African studies including subtype A in Kenya, and C in southern Africa, have been reported. Subtype differences in peak viral RNA level in the early stage in HIV-1 infection has been reported in Senegal in which CRF02_AG is associated with lower peak viremia than in HIV-1 subtype A and G infections (Sarr *et al.*, 2005).

High viral loads have been associated with risk of transmission suggesting that people with acute infection are more likely to transmit the virus (Dyer *et al.*, 1998; Granich *et al.*, 2009; Hollingsworth *et al.*, 2008; Tovanabutra *et al.*, 2002; Wawer *et al.*, 1999). This also suggests that reducing viral loads in acute infection, by either vaccination or treatment would reduce the rate of new infections.

HIV-1 primary infection is followed by a chronic phase during which viral load remains relatively stable at patient-specific levels with very slow increase until the onset of AIDS (Douek, 2003; Sabin *et al.*, 2000). However, although the individual may be asymptomatic, the chronic phase is accompanied by persistent viral replication in lymph nodes and rapid turnover of plasma virions, together with a steady decline of the main target-cell population, CD4+ T cells (Mellors *et al.*, 1996; O'Brien *et al.*, 1998).

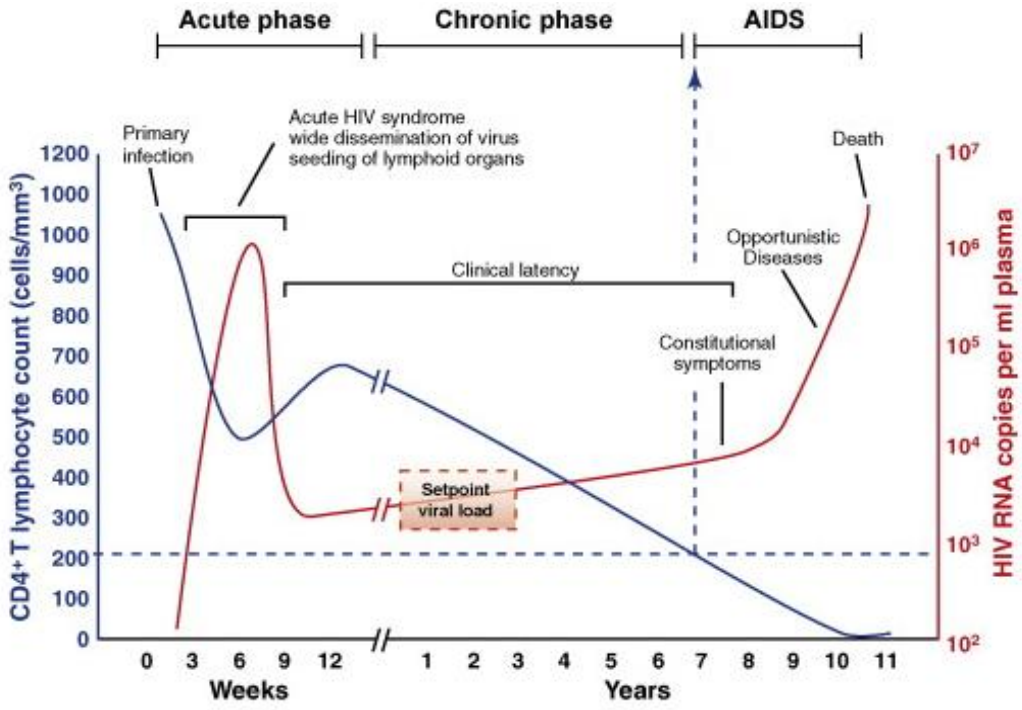


Figure 1.10 A typical natural course of HIV infection (An and Winkler, 2010).

The virus load in chronic infection reflects the equilibrium achieved between high rates of virus production and clearance; perhaps due to CTL and neutralising antibody responses (Ho *et al.*, 1995; Hogan and Hammer, 2001). This period has been described as clinically latent; however, the continuous decline in the number of CD4+ T lymphocytes results in the individual's immune system being unable to control opportunistic pathogens culminating in life-threatening AIDS-defining diseases (Pantaleo *et al.*, 1993; Sierra *et al.*, 2005). CD4+ T-cell count is the strongest predictor of clinical progression. Since 2005 in South Africa, initiation of anti-retroviral therapy was recommended for individuals with CD4 counts < 200 cells/ul or AIDS defining illness, however, the recommendations changed in 2008 and currently treatment may be initiated when individuals reach CD4 counts ≤ 350 cells/μL (www.sahivsoc.org).

1.5 Host factors influencing infection and disease progression

An early population study defined a small percentage of not more than 5% of individuals with untreated HIV-1, who maintain good general conditions and show either delayed or even no

signs of disease progression for several years following infection as long-term nonprogressors' (LTNPs) (Sheppard *et al.*, 1993). LTNPs were, however, further divided in terms of viremia into HIV controllers (HIC) (viral load ≤2000 copies/ml) or elite suppressors/controllers (ES/EC) (viral load <50 copies/ml) (Deeks and Walker, 2007; Lambotte *et al.*, 2005). The rate of HIV-1 disease progression has been shown to depend on a number of host factors. Genome-wide association studies (GWAS) have shown that the major determinant of HIV-1 viral control lies in the major histocompatibility complex (MHC) (Fellay *et al.*, 2007; Pereyra *et al.*, 2010) that encode for the human leukocyte antigen (HLA class I). HLA is also a ligand for the killer cell immunoglobulin-like receptors (KIR), a polymorphic set of molecules that modulate natural killer (NK) cell activity. Polymorphisms in KIR3DS1 (KIR-three domains, short cytoplasmic tail) is associated with HIV disease progression (Carrington *et al.*, 2008). However, among many other host genetic factors that have been shown to affect disease progression are polymorphisms in the HIV-1 co-receptors CCR5 and CCR2 (Dean *et al.*, 1996), the cytokines IL10 (Naicker *et al.*, 2012; Shin *et al.*, 2000) and IFN gamma (An *et al.*, 2003), host HIV-restriction factors like APOBEC3G (apolipoprotein B mRNA-editing enzyme-cytidine deaminase enzymes) (An *et al.*, 2004; Reddy *et al.*, 2010) and TRIM5 (Singh *et al.*, 2011; Stremlau *et al.*, 2004). In the CAPRISA acute infection cohort, which is the focus of this study, the following host genetic factors have been shown to affect susceptibility to infection: APOBEC3G (Reddy *et al.*, 2010), IL-10 polymorphisms (Naicker *et al.*, 2012) and Trim22 (Singh *et al.*, 2011).

1.5.1 The major histocompatibility complex (MHC)

MHC class 1 molecules are membrane-associated proteins that specifically bind and present processed pathogen-derived peptides on the cell surface and, in humans, are termed human leukocyte antigen or HLA (Hewitt, 2003). MHC class II molecules present foreign exogenous peptides to CD4+ T-lymphocytes, while MHC class I molecules present foreign endogenous peptides to CD8+ T lymphocytes (Unanue and Cerottini, 1989). A study of Fellay *et al.* (2007) using genome-wide approach identified the most significant polymorphism associated with differential set point viral load that were located in the HLA-complex P5 (HCP5) gene and in the 5′ region of the HLA-C gene, the single nucleotide polymorphism (SNP) database

numbers, rs2395029 and rs9264942, respectively. The SNP database numbers rs2395029 is known to be in high linkage disequilibrium (LD) with the HLA-B*57 (Fellay *et al.*, 2007; Pereyra *et al.*, 2010), an allele with the strongest impact on HIV-1 disease progression (Altfeld *et al.*, 2003; Migueles *et al.*, 2000). In addition, significant associations of differential HIV-1 disease progression have been detected with polymorphisms in or close to the the zinc ribbon domain–containing 1 (ZNRD1) gene in the HLA locus, which encodes an RNA polymerase I subunit (Ballana *et al.,* 2010; Fellay *et al.,* 2007; Limou *et al.,* 2009).

Furthermore, studies with HIV-1 LTNPs in European and North American cohorts show increased frequencies of certain specific HLA class I (HLA-A1, HLA-A2, HLA-B13, HLA-B14, HLA-B17, HLA-B27, HLA-51) and HLA class II (HLA-DR5, HLA-DR6) alleles (Carrington and O'Brien, 2003; Magierowska *et al.*, 1999; McMichael, 2007). In southern African cohorts, HLAs associated with lower viral loads in chronic infection include B*81:01, B*58:01/57 and B*39:10 (Kiepiela *et al.*, 2004; Leslie *et al.*, 2010). Furthermore, HLA-B alleles appear to have a greater impact on markers of disease progression compared to HLA-A and HLA-Cw alleles (Kiepiela *et al.*, 2004). However, combinations of HLAs have been found to have an additive effect, for example Cw*0401-B*8101, HLA-Cw*1203-B*3910, and HLA-A*7401-B*5703 haplotypes (Leslie *et al.*, 2010).

In contrast, the presence of HLA-B35, HLA-DR3, and HLA-DQ1 was shown to be more frequent in individuals with rapid disease progression (Carrington and O'Brien, 2003; MacDonald *et al.*, 2000; Magierowska *et al.*, 1999), while HLA-B*5802 and B*18 were found to be associated with high viral loads and lower CD4 counts in southern African cohorts (Kiepiela *et al.*, 2004; Leslie *et al.*, 2010). Homozygosity at HLA-A and B has also been shown to be disadvantageous for HIV-1 infected individuals probably due to a narrower array of HIV-1 derived peptides and as a consequence, a less diverse HIV-1 specific CTL repertoire (Carrington *et al.*, 1999; Martin and Carrington, 2005).

Nevertheless, HLA does not always predict disease progression (Emu *et al.*, 2008) and understanding the relationship between the virus and the host, and how HLA differentially

influences disease progression, remains a central question in HIV research and optimal vaccine immunogen design (McMichael and McCutchan, 2010).

1.6. Kinetics and impact of immune responses on disease progression

HIV-1 induces vigorous humoral and cellular immune responses. The first CD8+ T cell (CTL) responses arise around the time when viremia is at its' highest point and these responses peak one to two weeks later, as viremia declines (Borrow *et al.*, 1994; Goonetilleke *et al.*, 2009; Koup *et al.*, 1994; Wilson *et al.*, 2000), while neutralizing antibody responses occur only long after viremia has been controlled (Goonetilleke *et al.*, 2009; Gray *et al.*, 2011; Tomaras *et al.*, 2008). The clinical course of HIV-1 disease is likely to be influenced by the nature of the earliest immune response to the virus and the sooner the virus-infected cells can be eliminated - before the virus spreads widely and integrates to generate long-lasting and non-eradicable reservoirs of latent virus - the faster can the HIV-1 infection be controlled (McMichael and McCutchan, 2010).

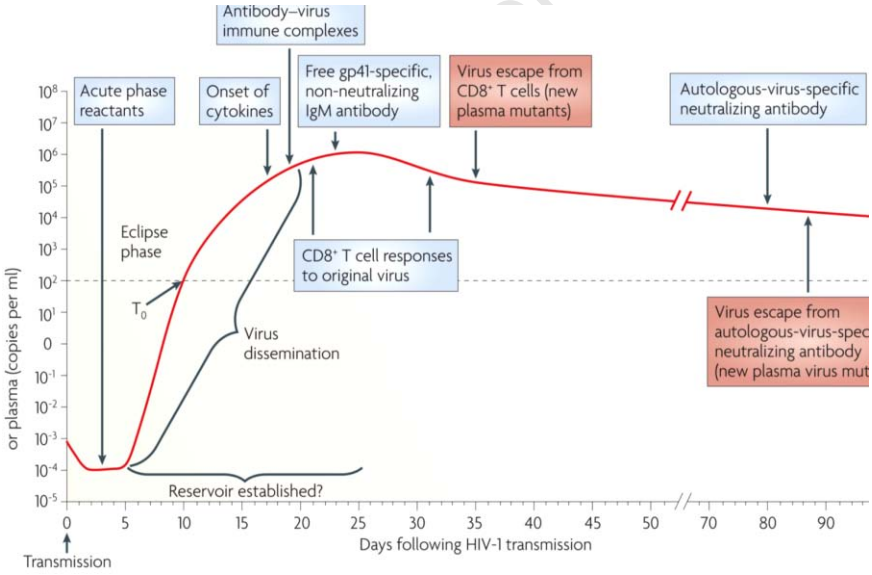


Figure 1.11 An alignment of the earliest innate and adaptive immune responses detected following HIV-1 transmission (from McMicheal *et al.*, 2010)

1.6.1 HIV-1 antibody responses

Numerous studies have found no correlation of early Env binding antibody responses with disease progression (Cecilia *et al.*, 1999; Frost *et al.*, 2005; Loomis-Price *et al.*, 1998; Tomaras *et al.*, 2008). Furthermore neutralizing antibodies are not associated with control of HIV replication in chronic infection and do not mediate HIV-1 suppression in elite controllers (Bailey *et al.*, 2006; Tomaras *et al.*, 2008). Early neutralizing antibodies develop against the autologous "infecting" virus and generally are not able to neutralize the contemporaneous virus due to rapid escape (Gray *et al.*, 2011; Gray *et al.*, 2007; Rong *et al.,* 2007; Wei *et al.,* 2003). Broadly neutralizing antibodies, which develop in only approximately 25% of individuals, occur after a number of years of infection and are able to neutralize a diversity of viruses (Gray *et al.*, 2009b; Richman *et al.*, 2003; Stamatatos *et al.*, 2009; Wei *et al.,* 2003). However, they do not impact on viral replication – possibly because they occur too late in infection to influence the course of disease (Cohen *et al.*, 2011; Gray *et al.*, 2011).

1.6.2 CD4 T-cell responses

The CD4+ T cell response is needed in orchestrating antibody production and primary expansion of CD8+ T cells and their differentiation into cytotoxic effectors (Janssen *et al.*, 2005; Keene and Forman, 1982). HIV-specific CD4+ T cell responses have been shown to expand in acute HIV-1 infection, although there is rapid decline of the responses (Gloster *et al.*, 2004; Oxenius *et al.*, 2001). However, as HIV-1 preferentially target HIV-specific CD4+ T cells, this results in the disruption of central process required for the successful coordination of the antiviral immune response (Dalgleish *et al.*, 1984; Douek *et al.*, 2002; Hel *et al.*, 2006). HIV infected individuals have impaired T helper-cell function which has been linked to the loss of central memory CD4+ T cells in HIV-I infection (Li *et al.*, 2005; Mattapallil *et al.*, 2005; Okoye *et al.*, 2012; Younes *et al.*, 2003).

In addition to their helper function, CD4+ T cells can also directly recognize and kill virally infected cells (Adhikary *et al.*, 2006; Heller *et al.*, 2007; Sacha *et al.*, 2009). Moreover, a

recent study has shown that *in vivo*, HIV-specific granzyme A+ (GrzA+) CD4 T cells at baseline is significantly enriched in individuals who subsequently control viremia (Soghoian *et al.*, 2012), suggesting that cytolytic "killer" CD4 T cells are involved in early control of HIV-1 infection.

1.6.3 CD8+ T-cell responses

Studies of single-genome amplification and sequencing of the first detectable virus, have shown that infection is initiated by a single virus in about 80% of heterosexual transmission of HIV-1 clade B and C infections, 60% in men who have sex with men and 40% in injection-drug users (Abrahams *et al.*, 2009; Keele *et al.*, 2008; Keele *et al.*, 2009; Li *et al.*, 2010; Salazar-Gonzalez *et al.*, 2009). The difference in the severity of the genetic bottleneck is consistent with the presence or absence of a mucosal barrier to transmission (Cohen *et al.*, 2011; Haaland *et al.,* 2009). Following transmission viral diversification is in part due to immune selection pressure in the host (Bernardin *et al.*, 2005; Henn *et al.*, 2012; McMichael *et al.*, 2010; Salazar-Gonzalez *et al.*, 2009). Studies have shown that most amino acid changes in the early virus were a consequence of rapid and sequential selection by mutations by CTL responses that recognize epitopes expressed by the founder virus (Goonetilleke *et al.*, 2009). Of the few mutations that were not associated with CTL selection, some were reversions towards the consensus sequence. Reversions following transmission occur if mutations that may have been advantageous to the virus in the previous host return to consensus amino acids. Other early changes in viral sequence might have been selected by antibody dependent cell-mediated virus inhibition or by NK cells (Goonetilleke *et al.*, 2009).

Emergence of CTL responses coincide with clearance of virus during primary HIV-1 in humans and simian immunodeficiency virus infection in rhesus monkeys (Borrow *et al.*, 1997; Goonetilleke *et al.*, 2009; Henn *et al.*, 2012; Koup *et al.*, 1994; Kuroda *et al.*, 1999). Interestingly, a recent study in five participants demonstrated that potent CTL antiviral responses against heterologous founder viruses occurring during acute infection (Fiebig stages 1/2 to 5) decline concurrently with the resolution of viremia (Freel *et al.*, 2012). Moreover, blocking CTL activity with CD8-specific monoclonal antibodies in the SIV macaque model

prevents the reduction in the early viral load (Jin *et al.*, 1999; Matano *et al.*, 1998; Schmitz *et al.*, 1999). Furthermore, the role of early CD8+ T cell responses in controlling HIV-1 replication is strongly supported by the association between particular HLA class I alleles and improved course of HIV-1 infection (Carrington and O'Brien, 2003; Pereyra *et al.*, 2010).

However, the quality and breadth of CTL responses needed to controls the HIV-1 replication have not been fully elucidated. Some studies have found no correlation between breadth or magnitude of either total HIV-1 or protein-specific CD8+-T-cell responses and plasma viral load (Addo *et al.*, 2003; Cao *et al.*, 2003; Gea-Banacloche *et al.*, 2000; Peretz *et al.,* 2005), while others have found correlations with specific proteins (Jia *et al.*, 2012; Kiepiela *et al.*, 2007; Masemola *et al.*, 2004). CTL responses in the first 3 months post infection commonly target HIV-1 proteins Nef and Gag (Gray *et al.*, 2009a; Lichterfeld *et al.*, 2004) and there have been several studies showing that responses targeting Gag, in particular p24, associate negatively with plasma viral load (Edwards *et al.*, 2002; Jia *et al.*, 2012; Kiepiela *et al.*, 2007; Luo *et al.*, 2012; Masemola *et al.*, 2004; Novitsky *et al.*, 2002; Rolland *et al.*, 2008; Rousseau *et al.*, 2008; Saez-Cirion *et al.*, 2009; Zuniga *et al.*, 2006). In contrast, total HIV-1 Env- or Nef-specific $CD8^+$ T-cell responses have been found to associate positively with plasma viral load (Jia *et al.*, 2012; Masemola *et al.*, 2004; Novitsky *et al.*, 2002). However, some studies found no correlation between viral load and Gag-specific CD8+/CD4+ T-cell responses (Betts *et al.*, 2001; Peretz *et al.,* 2005), while other studies found no correlation between plasma viral load and total HIV-1 Env- or Nef-specific $CD8^+$ /CD4+ T-cell responses (Dalod *et al.*, 1999; Edwards *et al.*, 2002; Peretz *et al.,* 2005).

Nonetheless, the majority of these studies were cross-sectional and were done in either subtype B setting or chronic HIV-1 infections. There are a few studies examining the timing and impact of CTL targeting in acute infection in subtype C infected African populations (Gray *et al.*, 2009a; Masemola *et al.*, 2004; Radebe *et al.*, 2011), which would differ in HLA background to subtype B infected Caucasian populations (Goulder *et al.*, 2000). Although events in acute infection are thought to influence subsequent disease progression, a study employing longitudinal samples of the CAPRISA acute infection cohort, found that breadth/magnitude of acute (≥3months post infection) CTL responses measured using the

IFN-gamma EliSpot assay did not predict set-point viral load at 12 months post-infection (Gray *et al.*, 2009a). This study on the CAPRISA 002 cohort found that Gag and Nef responses were the most immunodominant with Nef responses significantly more dominant than any other protein. However this study did find, similar to other studies, that contemporaneous CD8+ T-cell responses at 12 months post infection, to Env- and Gag proteins were positively and negatively correlated with plasma viral load, respectively, while there were no correlation with Nef- and Pol-specific CD8+ T-cell responses with viremia (Mlotshwa *et al.*, 2010).

The discrepancies observed in studies correlating CTL responses and disease markers may be attributed, at least in part, to the differences in methodologies, disease stages and diverse population groups studied. However, most studies suggest that individuals progressing slowly to infection have strong virus-specific CTL and robust CD4+ T-cell proliferative responses to Gag, with Gag and Nef being most immunodominant in primary infection, irrespective of other factors (Berger *et al.*, 2011; Kiepiela *et al.*, 2007; Kim *et al.*, 2010; Klein *et al.*, 1995; Rolland *et al.*, 2008).

1.7 CTL responses associated with control of viral replication

The association between some HLAs with slow disease progression is likely due to their capacity to present epitopes in Gag p24 (Berger *et al.*, 2011; Borghans *et al.*, 2007; Matthews *et al.*, 2008; Wang *et al.*, 2009). Beneficial HLAs like B*57/58:01 and B*27 present epitopes in the conserved Gag p24, while non-protective HLAs like B*58:02/B*35:02/B*18 present no/few p24 epitopes in acute/early infection (Borghans *et al.*, 2007; Brockman *et al.*, 2010; Dahirel *et al.*, 2011; Matthews *et al.*, 2008; Wang *et al.*, 2009). Furthermore, increased breadth of Gag specific responses has been shown in several distinct population studies to associate negatively with viral load (Edwards *et al.*, 2002; Jia *et al.*, 2012; Kiepiela *et al.*, 2007; Zuniga *et al.*, 2006).

Studies have also shown that CTL recognition of HIV-infected cells early in the viral life cycle may be important for viral containment in HIV-infected individuals (Payne *et al.*, 2010;

Sacha *et al.*, 2007). *In vitro* studies have demonstrated that Gag epitopes can be processed and presented directly from the incoming virion in as little as 2h following infection, unlike other viral proteins that need about 12 h for expression and recognition (Payne *et al.*, 2010; Sacha *et al.*, 2007). It is thought that this early processing of Gag epitopes enables the CTL targeting Gag to recognise and kill infected cells before new virions are released and Nef downregulation of MHC I (Collins *et al.*, 1998; Payne *et al.*, 2010; Sacha *et al.*, 2007).

Furthermore, some studies have shown that T cells with high functional avidity may recognize virally infected cells at lower surface antigen densities than those with low-avidity (Belyakov *et al.*, 2007; Bennett *et al.*, 2007; Betts *et al.*, 2004; Mothe *et al.*, 2012). Unlike non-functionally avid responses, functionally avid CTL responses select for mutations with a fitness cost to the virus and appear to distinguish controllers from progressors (Allen and Altfeld, 2008; Frater *et al.*, 2007; Mothe *et al.*, 2009; O'Connor *et al.*, 2002). A recent study has further shown that irrespective of the allele, HLA-B alleles restricting Gag specific responses exhibited significantly higher functional avidities than responses restricted by HLA-A or HLA-C molecules or targeting epitopes outside Gag (Berger *et al.*, 2011).

However, the hallmarks of a truly protective immune response remain unclear since individuals with HIV infection cannot clear infection and finally progress to AIDS. Numerous factors responsible for immune failure have been suggested including destruction of CD4+ cells, CD8+ cell exhaustion and viral mutational and non-mutational CTL escape (Day *et al.*, 2006; Feeney *et al.*, 2004; Goonetilleke *et al.*, 2009; Goulder *et al.*, 1997; Letvin and Walker, 2003; Wei *et al.*, 2003).

1.8 Mechanisms of CTL mutational escape

HIV-1 uses two major ways to escape killing by CTL and these are mutations in or proximal to an epitope (Klenerman *et al.*, 1994), or by non-mutational mechanisms (Cohen *et al.*, 1999; Le Gall *et al.*, 2000). Non-mutational escape of CTL response includes Nef-mediated downmodulation of CD4+ T-cells and the expression of HLA-A and HLA-B but not of HLA-

C (Collins *et al.*, 1998; Tomiyama *et al.*, 2002; Yang *et al.*, 2002), which are discussed above (Index 1.2.8).

In mutational escape, there are three mechanisms of escape. Firstly, mutations occurring at the anchor residue of the epitope may prevent binding of the peptide to the MHC class I molecule resulting in the epitope not being presented on the surface of the infected cell (Couillin *et al.*, 1994; Goulder and Watkins, 2004; Kaufmann *et al.*, 2007), in which case, there can be no generation of a *de novo* mutant specific CTL response. Secondly, a mutation may impair recognition by the T cell receptor (TCR) and subsequent CTL activation may be antagonised, preventing CTLs that are capable of recognizing wild-type virus (Franco *et al.*, 1995; Price *et al.*, 2004; Reid *et al.*, 1996). Thirdly, a mutation may prevent epitope-presentation due to interference with intracellular antigen processing in which proteosomes degrade viral proteins in forms that are not presentable to MHC-1 or are less recognized by MHC-1 than wild-type (Cardinaud *et al.*, 2011; Cascio *et al.*, 2001; Klenerman *et al.*, 1994; Rock *et al.*, 1994; Saric *et al.*, 2001; York *et al.*, 2003; Zimbwa *et al.*, 2007)

Other forms of interference with intracellular antigen processing may involve preferential binding of oligopeptides to the transporters associated with antigen processing (TAP), and their translocation in an unprocessed form through the endoplasmic reticulum (ER) or restriction of access of some antigenic peptides to the ER (Draenert *et al.*, 2004; Lauvau *et al.*, 1999). Furthermore, the amino terminal precursors may not be trimmed (Draenert *et al.*, 2004) or trimmed into antigenic peptides of the incorrect length that does not bind to the MHC class I molecule (Seifert *et al.*, 2003; Serwold *et al.*, 2001). Often CTL escape mutations occurring in amino acid residues outside but within five residues flanking known HLA class I restricted epitopes interfere with the processing of optimal peptide antigen (Allen *et al.*, 2004; Draenert *et al.*, 2004; Zimbwa *et al.*, 2007).

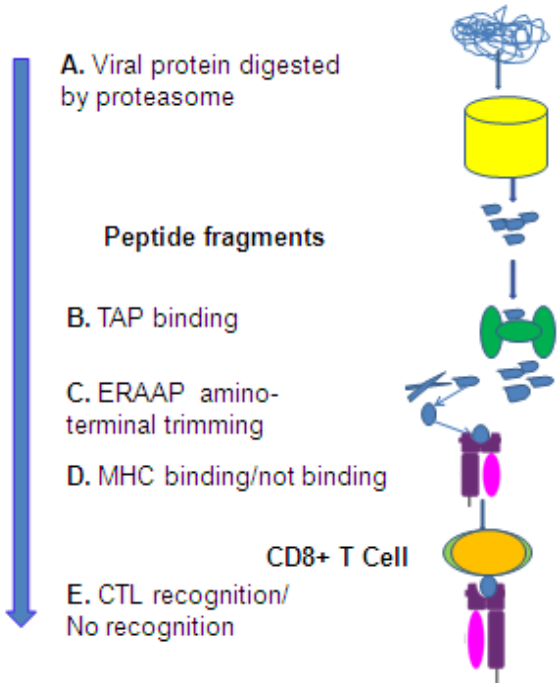


Figure 1.12: Schematic illustrations of antigen processing and presentation and processes that may lead to CTL mutational escape during HIV and SIV infection (Drawn based on Goulder and Watkins, 2004).

1.9 Viral escape from CTL responses

Viral escape offers a survival advantage to the virus as the CTL response generated against the wild-type does not recognise the mutant (Price *et al.*, 1997). However, the extent to which evasion of CTL responses affects the rate of disease progression remains not well understood. Some CTLs that target structurally and/or functionally conserved epitopes may be beneficial to the host if escape mutations impose a replicative fitness cost to the virus (Crawford *et al.*, 2007; Goepfert *et al.*, 2008; Martinez-Picado *et al.*, 2006; Matthews *et al.*, 2008; Schneidewind *et al.*, 2007). Conversely, gradual accumulation of CTL escape mutants may lead to loss of the critical epitopes associated with effective control of viremia that may offer sufficient advantage to viral fitness *in vivo* (Draenert *et al.*, 2004; Goulder and Watkins, 2004; Kawashima *et al.*, 2009). Understanding the nature and impact of CTL selected mutations on disease progression remain important for the design of an effective CTL based vaccine. This study investigated the timing of CTL escape in Gag and Nef and its consequences on HIV-1

disease progression during the first year of infection in subtype C infected individuals from the CAPRISA002 acute infection cohort.

1.10 CTL escape and viral fitness

While viral evasion of CTL response or CTL escape may offer a survival benefit to the virus, there is limitation to the plasticity of HIV-1 proteome (Liu *et al.*, 2007). Such lack of plasticity is seen in reversions of some mutations, which may be detrimental to the virus, when transmitted to individuals lacking the HLA that selected the mutation (Allen and Altfeld, 2008; Allen *et al.*, 2005; Leslie *et al.*, 2004; Navis *et al.*, 2008; Seki *et al.*, 2008).

Most of the escape mutations associated with viral fitness cost are in epitopes restricted by protective HLA class I alleles, B*57/5801, B*27, B*51 and B*8101/B*3910; these epitopes are the TW10 (Gag 240–249, KK10 (Gag 263–272), TI8 (RT 128–135) and TL9 (Gag 180 – 188), respectively (Boutwell *et al.*, 2009; Brockman *et al.*, 2010; Frater *et al.*, 2007; Kawashima *et al.*, 2009; Rolland *et al.*, 2010; Schneidewind *et al.*, 2008; Schneidewind *et al.*, 2007; Wright *et al.*, 2010; Wright *et al.*, 2012). Furthermore, most of these deleterious escape mutations have been identified in the conserved Gag p24 protein (Martinez-Picado *et al.*, 2006; Troyer *et al.*, 2009) and are in structurally constrained epitopes. CTL attenuation of viruses by HLA-B*57/58:01 and B*39:10/81:01 alleles in Gag p24 through selection of mutations in TW10 (Gag 240 – 249) and TL9 (Gag 180 -188) epitopes (Rolland *et al.*, 2010; von Schwedler *et al.*, 2003; Wright *et al.*, 2010) are some of the best examples. Both the TW10 epitope and the TL9 epitope are positioned at functionally important domains involved in triggering viral assembly with common escape mutations at amino acids positions T242 and T180, located at the interface of the N-capping boxes in the sixth α-helix and the third α-helix of the capsid respectively (Martinez-Picado *et al.*, 2006; Tang *et al.*, 2002). Furthermore, studies have found higher numbers of rare or novel CTL escape mutations associated with reduced replication capacities from viruses from elite controllers compare to those from non-controllers (Miura *et al.*, 2009). It remains unknown however, if the rare mutations are selected by the HLA mediating early CTL response in these individuals or from

infection with a founder virus that already contains the escape mutations (Boutwell *et al.*, 2010).

Interestingly, while *in vitro* studies have shown that some CTL mutations selected by HLAs associated with viral control cumulatively incur fitness cost to the virus (Boutwell *et al.*, 2010; Boutwell *et al.*, 2009; Crawford *et al.*, 2009; Rolland *et al.*, 2010), *in vivo* sequential development of these mutations may result in increased viral loads (Crawford *et al.*, 2009). These results emphasize that a complex interplay of factors are involved in determining the HIV-1 overall infection outcome. One way to try and differentiate the role of HLA and viral factors is in looking at transmission of HLA-attenuated viruses to individuals who do not have these so-called beneficial HLAs. Infection in an MHC haplotype *90-120-Ia-negative* macaque with a crippled SIV carrying five *gag* mutations resulted in multiple reversions and progressions to AIDS (Seki *et al.*, 2008). In contrast to this, studies in humans have associated lower viral loads and higher CD4 counts in acute infection in B*57/58:01 HLA negative subjects, who acquired putative attenuated viruses from HLA B*57/58:01 positive individuals (Chopera *et al.*, 2008; Goepfert *et al.*, 2008; Thobakgale *et al.*, 2009). However, the overall impact of transmission of constrained, or even multiple epitope-attenuated viruses to individuals who do not have these HLAs remains controversial, since recent reports failed to detect advantage to recipients of less fit mutations (Novitsky *et al.*, 2010; Wright *et al.*, 2011).

1.11 Compensatory mutations and viral fitness

HLAs that are associated with slow disease progression often select for mutations in structurally and functionally constrained regions (Altfeld and Allen, 2006; Kelleher *et al.*, 2001; Martinez-Picado *et al.*, 2006; Schneidewind *et al.*, 2008). The virus adapts to such deleterious mutations through developing other linked mutations known as compensatory mutations that partially or wholly offset the fitness cost the virus incurs due to the deleterious mutation (Brockman *et al.*, 2010; Brockman *et al.*, 2007; Crawford *et al.*, 2009; Crawford *et al.*, 2011; Martinez-Picado *et al.*, 2006; Schneidewind *et al.*, 2008).

The best example and well-studied CTL escape mutation with substantial fitness cost to the virus has been the mutation at residue 242 (T242N) in the TW10 epitope. Structural modelling has shown that this residue is central to a delicate series of interactions, changes in which might perturb CypA binding (Brockman *et al.*, 2007; Crawford *et al.*, 2009; Martinez-Picado *et al.*, 2006; Pereyra *et al.*, 2010). A number of experiments have shown that viral escape in TW10 through T242N residue is indeed partially or wholly compensated by Gag residues 219, 223 and 228 (mutations H219Q, I223V and M228L respectively) (Brockman *et al.,* 2007; Boutwell *et al.,* 2008).

Studies have also shown that sites of compensatory mutations are more likely than expected by chance to be near the site of the original deleterious mutation (Salazar-Gonzalez *et al.*, 2009). Indeed, studies have shown that mutations like I147L and S165N that partially compensate for A146P and A163G in the B*5703/5801 restricted ISW9 and KF11 epitopes, respectively (Crawford *et al.*, 2007) are quite near the original site of primary escape (Crawford *et al.*, 2009; Draenert *et al.*, 2004; Troyer *et al.*, 2009).

Nevertheless, other compensatory mutations have been detected at sites away from the original deleterious mutations. For instance, the fitness cost of the CTL escape mutation R264K in the B*27 restricted KK10 epitope has been located at S173A (Schneidewind *et al.*, 2007), although a close compensatory substitution at site L268M has also been detected (Kelleher *et al.*, 2001; Schneidewind *et al.*, 2008; Schneidewind *et al.*, 2007; van der Kuyl *et al.*, 2009). Interestingly, however, the mutation L268M has been independently shown to enhance binding to Immunoglobulin-like transcript 4 (ILT4), a MHC class I receptor inhibitor, expressed on monocytes and dendritic cells, causing substantial defects in the antigen-presenting properties of these cells (Lichterfeld *et al.*, 2007). Recent findings suggest that the virus uses this pathway of immune escape in some HLA-restricted escape (HLA-A11-, B8- and B7), while it does not in some others like the B57 (Yang *et al.*, 2010).

Overall, in both SIV and HIV, mutations with a net fitness cost have been shown to revert to wild-type after infecting HLA mismatched hosts due to absence of the CTL pressure that

selected them or compensatory mutations partially restored the viral fitness that would delay reversion to wild-type (Crawford *et al.*, 2007; Leslie *et al.*, 2004; Schneidewind *et al.*, 2007).

1.12 CTL escape and HIV evolution

Following transmission to HLA-mismatched recipients, some escape mutations may persist for years, while others revert to wild-type (Chopera *et al.*, 2008; Friedrich *et al.*, 2004; Leslie *et al.*, 2004; Matthews *et al.*, 2008; Wright *et al.*, 2011). Mutations that have occurred at little fitness cost to the virus or those that have been compensated may accumulate in populations (Crawford *et al.*, 2007; Draenert *et al.*, 2004; Kawashima *et al.*, 2009) resulting in an "HLA footprint" on HIV sequences (Moore *et al.,* 2002; Leslie *et al.,* 2005; Klenerman and McMichael, 2007; Kawashima *et al.,* 2009; Matthews *et al.,* 2009). HLA footprinting was initially reported by Moore *et al.* (2002), although correcting for founder effects resulted in a loss of some of this signal (Bhattacharya *et al.,* 2007). Further studies have suggested that the most common HLA in a given population of HIV infected individuals may influence HIV evolution in a particular direction as specific HLA-HIV codon associations are determined by specific HLA frequency distributions (Avila-Rios *et al.,* 2009). Indeed, a profound effect of HLA-B*51:01, which is a dominant HLA allele in Japanese showed that the protective effects of the TAFTIPSI (reverse transcriptase residues 128–135) epitope was virtually eliminated from the population due to escape through the I135X (Kawashima *et al.*, 2009). This suggests that CTL escape may have a fundamental impact on reducing the immunogenicity of viruses and that gradually, the so-called beneficial HLA alleles may not retain their advantage (Goulder and Watkins, 2008). Furthermore, studies have shown a strong correlation between variant frequency of the rapidly reverting T242N mutation in the B*57:03/58:01 restricted TW10 epitope with HLA prevalence in populations suggesting that this mutation is accumulating over time, which can be explained, either by mutant acquisition exceeding reversion rate or by compensatory mutations that completely prevent reversion (Crawford *et al.*, 2007; Kawashima *et al.*, 2009).

Studies have also shown that transmission of viruses encoding escape mutants in the critical Gag epitopes to individuals expressing the relevant MHC class results in failure to control

viremia (Crawford *et al.*, 2009; Goepfert *et al.*, 2008; Goulder *et al.*, 2001) suggesting that accumulation of these mutations may lead to loss of protection in populations of individuals with beneficial HLAs. The longer-term consequences of the accumulation at the population level of these escape mutations on the immune control of HIV remain unknown however (Goulder and Watkins, 2008).

1.13 Study rationale

The development of a prophylactic HIV vaccine that prevents initial infection remains a global priority. However, the major challenge in developing such a vaccine is defining correlates of sterilising protection against HIV-1 infection (Koup *et al.*, 2011; McElrath and Haynes, 2010). In the development of vaccine candidates that elicit CTL responses, of particular interest is the elucidation of CTL responses associated with control of acute HIV infection. CTL responses have been associated with control of viremia in chronic HIV-1 infection, and these responses contribute to the initial control of peak viremia in acute infection (Borrow *et al.*, 1994; Feeney *et al.*, 2004; Goonetilleke *et al.*, 2009; Goulder *et al.*, 1997; Henn *et al.*, 2012; Kiepiela *et al.*, 2007; Koup *et al.*, 1994).

The study of patterns of immune escape in HIV-1 acute infection provides us with a sensitive tool to identify specific viral regions under active immune selection pressure and contribute to our understanding of the correlates of immune control of HIV-1-infection. While certain CTL responses appear to control HIV replication, escape from these responses does not necessarily result in loss of control. This is because some escape may be detrimental to the virus and result in lower replication capacity - creating a balance between the benefit of escape to the virus and its survival (Crawford *et al.*, 2009; Martinez-Picado *et al.*, 2006; Schneidewind *et al.*, 2007; Woo *et al.*, 2010). Furthermore, mapping immune escape and the fitness cost to the virus provide information relevant to the selection of candidate immunogens for an HIV-1 vaccine.

Understanding events in acute HIV-1 subtype C infection are particularly important since subtype C accounts for over 95% of the HIV infections in southern Africa (Bredell *et al.*,

2007; Novitsky *et al.*, 2009). Several studies have focused on subtype C chronic infection (Kiepiela *et al.*, 2004; Kiepiela *et al.*, 2007) with only a few studies on subtype C acute infection. This study sought to address the question of the timing, location and impact of CTL escape on disease progression in subtype C HIV-1 infection. Furthermore, this study is part of the Centre for AIDS Program of Research in South Africa (CAPRISA) 002 study that focuses on investigating the role of viral and immunologic factors during acute and early subtype C infections. At the time of this study, the cohort had enrolled 62 recently HIV-1 infected individuals (within 3 months of infection) from rural and urban regions of Kwazula Natal.

This study focused on the Gag and Nef proteins as these are immunodominant and targeting of the Gag is associated with better disease outcome (Kiepiela *et al.*, 2007; Miura *et al.*, 2009; Rolland *et al.*, 2008). Furthermore, the majority of deleterious mutations have been found in Gag (Martinez-Picado *et al.*, 2006; Schneidewind *et al.*, 2007; Troyer *et al.*, 2009; Wright *et al.*, 2012).

The overall aim of this study was to determine if networks of attenuating mutations in viruses in acute infection could be responsible for lower viral loads observed in some individuals.

The specific objectives of the project were:

1. To identify mutations associated with CTL escape that emerge during the first 12 months following HIV-1 subtype C infection;
2. To quantify both the frequency of CTL escape-mutations in transmitted viruses and the timing of their reversion over two years following their transmission to HLA mismatched hosts;
3. To determine the impact on disease progression of CTL escape mutations detected in viruses from individuals during acute infection.

Timing and Pathogenic consequences of early CTL escape on HIV-1 Subtype C infection

Abstract

This study investigated the frequency, timing and pathogenic consequences of early cytotoxic T- lymphocyte (CTL) escape on subsequent HIV-1 disease progression. Altogether, 365 and 289 full-length *gag* and *nef* sequences were generated from samples at enrolment (median 1.4 months post infection), 3, 6, 12 and 24 months post infection from 58 subtypes C infected individuals. Putative escape was identified within or flanking known epitopes restricted by the participants HLA. Timing of escape was defined as either acute (≤ 3 months), early (> 3 to ≤ 6 months) or late (> 6 to 12 month).

In participants with single HIV-1 infection, CTL escape was detected in 69% (35/51) and was more frequently observed in the acute than in the early/late infection phases (3.8 versus 1.1 escape/month/100 amino acids; p=0.0231). Participants with viruses escaping CTL responses in Gag and/or Nef in acute and early infection had higher CD4+T-cell counts when compared to those without viral escape in this time period (p=0.0018 and 0.0542, respectively). Although the decline in viral load was similar after 3 months (p=0.998), the initial preservation of CD4+ cells resulted in a higher mean CD4+ T-cell counts in individuals whose viruses escaped in acute infection compared to those without viral escape at 12 months post-infection (p=0.0274). Moreover, a significantly higher number of viral escape mutations (72.4%, 42/58) were selected by HLA-B alleles compared to the number of mutations selected by HLA-A alleles (17.2%, 10/58) (p=0.0134). There was also a trend for higher CD4+ T-cell counts at 12 months in HLA-B restricted escape in Gag and Nef (p= 0.08; median 500 cells/µl [IQR=416 - 645] versus 368 cell/µl [IQR=303 - 624]).

The high frequency of escape during peak viremia provides evidence of CTL pressure on the virus in very early infection. These results demonstrate that escape occurs predominantly in HLA-B restricted epitopes and immune responses selecting escape mutations in acute infection provide benefit within the first year of infection.

2.1 Introduction

Virus-specific cytotoxic T cell (CTL) responses contribute to the control of viremia in acute infection and numerous studies have shown an inverse correlation of CTL responses after seroconversion with control of viremia (Borrow *et al.*, 1994; Freel *et al.*, 2012; Goonetilleke *et al.*, 2009; Henn *et al.*, 2012; Kim *et al.*, 2010; Koup *et al.*, 1994). However, the precise nature of CTL responses associated with successful control of HIV-1 infection is not fully understood. Moreover, most of these studies on the early dynamics of the HIV infection have predominantly been done in subtype B settings, with the few studies done in subtype C suggesting that observations in different subtype populations may not be extrapolated to other subtypes (Gray *et al.,* 2009a; Novitsky *et al.,* 2009; Radebe *et al.*, 2011). Understanding CTL responses and immune escape in subtype C acute infection cohorts is important as firstly, subtype C is the most dominant virus globally; secondly, acute infection is thought to play a disproportional role compared to other stages in infection in influencing disease progression (Brenchley *et al.*, 2004; Kelley *et al.*, 2007; Levesque *et al.*, 2009; Veazey *et al.*, 1998); and lastly, understanding immunological and viral factors that correlate with control of infection during acute infection are critically important for the design of an effective HIV-1 CTL based vaccine.

In subtype C chronic infection, studies have shown that Gag responses were associated with lower viral loads (Kiepiela *et al.,* 2007). However, although events in acute infection are thought to influence subsequent disease progression, a study on the CAPRISA acute infection cohort found none of the HIV-1 protein-specific CTL responses in the first three months of infection, including those targeting Gag, were prognostic of set-point viral load at 12 months post-infection (Gray *et al.,* 2009a). Furthermore, this group of researchers documented that individuals with rapid disease progression were more likely to lose immune responses in the first year of infection compared to controllers, suggesting that CTL escape is detrimental to the individual (Mlotshwa *et al.,* 2010).

Nonetheless, escape from CTLs targeting structurally or functionally conserved epitopes may be beneficial to the host if the escape mutants incur a replicative fitness cost (Martinez-Picado

et al., 2006; Schneidewind *et al.*, 2008; Schneidewind *et al.*, 2007; Smith, 2004). However, these mutations have predominantly been found in Gag epitopes restricted by HLA alleles associated with relative protection and the majority of these have been again in subtype B (Brockman *et al.*, 2010; Crawford *et al.*, 2009; Frater *et al.*, 2007; Martinez-Picado *et al.*, 2006; Schneidewind *et al.*, 2007) (see Chapter 1.10). There are a few studies in subtype C infected cohorts that have investigated the effect of escape in acute infection. These studies focused on B*57/58:01 restricted epitopes in Gag (Chopera *et al.,* 2011; Crawford *et al.,* 2009) and although *in vitro* studies demonstrated that the accumulation of escape mutations reduced viral replicative capacity, there was increase in viral load *in vivo* in association with escape in all three HLA-B*5703–restricted CTL responses (Crawford *et al.,* 2009).

To better understand events in acute infection and subsequent influence on disease progression, this study analyzed viruses in recently infected women from the CAPRISA 002 acute infection cohort. The design of the CAPRISA study was to detect HIV infection at the very earliest stage of HIV infection through monthly screening of HIV negative women by PCR or women with a reactive HIV antibody test within 3 months of a previously negative antibody result (van Loggerenberg *et al.*, 2008). Because viruses rapidly adapt in a new host, this cohort attempted to enrol individuals close to the period of infection such that early viruses would have had limited time to evolve, making this cohort suitable for studying acute infection. This chapter reports on the characterisation of *gag* and *nef* sequences from 58 women recruited during acute HIV infection and followed for at least one year. Seven women were found to be either co-infected (n=4) or superinfected (n=3) with two distinct HIV strains. The timing of CTL escape and its consequences on HIV-1 disease progression was investigated in the 51 women with single infection.

2.2 Material and Methods

2.2.1 Study subjects

The CAPRISA 002 Acute Infection cohort enrolled subjects within 3 months of infection, and followed them until initiation of antiretroviral therapy according to South African National Guidelines (defined at that time as CD4+ < 200 cells/µl or AIDS defining illness) (van Loggerenberg *et al.*, 2008). Women were enrolled from both the HIV negative cohort, and other seroincidence cohorts in Durban, South Africa. HIV-1 infection was identified using rapid antibodies assay and PCR, and was confirmed using an enzyme immunoassay (EIA) test as previously described (van Loggerenberg *et al.*, 2008). CD4+ T cell counts were assessed using a FACSCalibur flow cytometer while viral loads were measured using a COBAS AMPLICORTM HIV-1 Monitor Test v1.5 (Roche Diagnostics, Branchburg, New Jersey, USA). Samples were collected at enrolment, weekly for three weeks, fortnightly until 3 months, monthly until a year and quarterly thereafter. Written informed consent was obtained from all participants. This study received ethical approval from the University of KwaZulu-Natal, University of the Witwatersrand and University of Cape Town (REC NO: 025/2004)

2.2.2 HIV-1 Plasma RNA isolation and sequencing

RNA was isolated from plasma samples using either the Magna-Pure Compact Nucleic Extractor (Roche) or manually using QIAamp viral RNA extraction mini kit (QIAGEN) for samples with plasma viral load of less than 2000 copies per mL. SuperScript lll Reverse transcription kit (Invitrogen) and gene specific primers, the Gag D reverse (5′-AAT TCC TCC TAT CAT TTT TGG-3’; HXB pos 2382-2402) and Nef OR: (5′-AGG CAA GCT TTA TTG AGG-3’; HXB pos 9608-9625) for *gag* and *nef* respectively, were used to transcribe cDNA from RNA. Gag PCR primers for the first round were the Gag D reverse and Gag D forward (5′-TCT CTA GCA GTG GCG CCC G-3’; HXB pos 626–644). Gag A forward (5′-CTC TCG ACG CAG GAC TCG GCT T-3’; HXB pos 683–704) and Gag C reverse (5′-TCT TCT AAT ACT GTA TCA TCT GC-3’; HXB pos 2334–2356) were the second round primers. Nef PCR primers for the first round were the Nef OR and Nef SQ15FC (5′-GAG AGC GGT GGA ACT

TCT-3′; HXB pos 8561-8578), while the second round primers were Nef F (5′-CCT AGA AGA ATA AGA CAG GGC TT-3′; HXB pos 8754-8776) and Nef R (5′-CCT GGA ACG CCC CAG TGG-3′; HXB pos 9443-9461). Gag sequencing primers were Gag A forward, Gag A reverse (5′-ACA TGG GTA TCA CTT CTG GGC T-3′; HXB pos 1282–1303), Gag B forward (5′-CCA TAT CAC CTA GAA CTT TGA AT-3′; HXB pos 1226–1246), Gag B reverse (5′-CTC CCT GAC ATG CTG TCA TCA T-3′; HXB pos 1825–1846), Gag C reverse and Gag C forward (5′-CCT TGT TGG TCC AAA ATG CGA-3′; HXB pos 1748–1768). Nef sequencing primers were Nef Reverse and Nef forward. Sequences were assembled using Chromaspro (Technelysium Pty Ltd) and aligned using the ClustalW in Bioedit (default settings, Thompson *et al.,* 1994).

2.2.3 Timing of CTL escape

The timing of CTL escape was estimated as the time of the first appearance (observation) of a CTL mutation either in or proximal to a known epitope restricted by the participants' HLAs. Timing was then defined as either acute (0 to 3 months), early (>3 to 6 months) or late (>6 to 12 month). Time to escape was estimated as a mid-point between the last sequenced time point and the timing of the first observation of the escape mutation.

2.2.4 Virus-specific PCR

First strand synthesis was transcribed as above described, but using the unique individual-specific reverse (R) primer (sequence shown below). Primers below were then used to amplify *gag* sections specific to the individual's sequence in non-nested 2 rounds. Specific primers used were CAP237GagR: (5'- TCA TTA TGT TTG CAT TGC C -3'; patient's virus sequence pos 1115 - 1122); CAP237GagF: (5'- GGT ACA CAA TCC CCT ATC G -3'; HXB2 pos 420 - 438) CAP281GagR: (5'- CAC ATT GTT TAC CTG GCT CAT -3'; patient's virus sequence pos 1190 - 1110) and CAP281GagF: (5'- AAG TAA TAG AGG AGA AGG C -3'; HXB2 pos 461 - 479).

2.2.5 DNA sequence divergence and phylogenetic analyses

Phylogenetic and molecular evolutionary analyses were conducted using MEGA software version 4 (Tamura *et al.,* 2007). Maximum pairwise distances were estimated using Kimura Two parameter model (Kimura, 1980). Where maximum likelihood phylogenetic trees were constructed, this was implemented in RDP3 (Martin *et al.,* 2005) and edited in MEGA (Tamura *et al.,* 2007). Recombination in the dually infected sequences was assessed using Genetic Algorithms for Recombination Detection (GARD) implemented in Datamonkey (Kosakovsky Pond *et al.,* 2006) and nucleotide transition and transversion plots using the highlighter plot tool (www.hiv.lanl.gov).

2.2.6 IFN-γ ELISpot assay

This data was provided by CM. Gray, NICD/UCT; HIV-1-specific T cell responses across the entire HIV-1 clade C proteome were quantified by interferon gamma (IFN-γ) enzyme-linked immunospot (ELISpot) assay using freshly isolated and prepared PBMCs as previously described (Masemola *et al.*, 2004). Peptide pools used were HIV-1 subtype C consensus sequences and positive responses at the single-peptide level within the peptide pools were confirmed in a second IFN-γ ELISpot assay and monitored longitudinally in participants positive in the initial screening.

2.2.7 HLA typing

HLA Data was provided by the C. Gray, NICD/UCT; HLA class I (four digits) genotyping was performed on DNA extracted from either PBMCs or granulocytes using Pel-Freez DNA isolation Kit (Pel-Freez). Exons 2, 3, and 4 were then sequenced using Atria Allele SEQR kits (Abbott Diagnostics) and Assign SBT 3.5 software (Conexio Genomics). Ambiguities were removed using sequence specific primers.

2.2.8 Statistical analyses

Statistical analysis and graphical presentations were implemented in GraphPad Prism 5.0 (GraphPad Software, Inc.). To control for fluctuations at viral set-points at 12 and 36 months post-infection, we calculated viral load and CD4+ T-cell counts as medians of measurements taken at three time-points closest to these intervals. To determine if DNA distances fell outside of the normal Gaussian distribution, extreme upper outliers were defined as those whose DNA distances fell above an exploratory data analysis value (EDA), which was calculated from the following box plot formula: ([3 x IQR + [upper quartile - lower quartile value]); [John *et al.,* 1983; Iglewicz and Hoaglin, 1993]). Statistical analysis of significance was based on Mann-Whitney two-tailed t test or Kruskal-Wallis (Dunn's test) for multiple comparisons. Variations in timing and number of mutations selected by different HLAs were compared using analysis of variance (ANOVA - two-way t test) model. Rates of viral load and CD4+ count changes over time were estimated using locally weighted regression statistics (Loess) (Cleveland *et al.,* 1988)

2.3 Results

2.3.1 Cohort characteristics

Sixty two women were enrolled into the CAPRISA 002 acute infection study within a median of 6 weeks of infection (IQR = 4-8 weeks) (enrolment visit), of whom one participant, CAP259, opted out in early stages, while samples for three were not available for sequencing at the time of this study. HIV infection in 58 women was characterised in this thesis.

Laboratory staging of acute infection was available for 52 individuals (generated by R. Thebus, IIDMM, UCT). At the time of sampling for sequencing, two individuals were HIV RNA positive and enzyme immunoassay (EIA) antibody negative (Fiebig 1-II), two were EIA antibody positive but Western blot negative (Fiebig III), seven had an indeterminate Western blot result (Fiebig IV), 19 were Western blot positive but had no reactivity to the p31

integrase band (Fiebig V), while 22 had positive western blot results and a p31 band present (Fiebig VI) (Table 2.1) (Fiebig *et al.*, 2003). The median viral load at the enrolment visit was 71 900 copies/ml (IQR=12 500 – 399 000).

CD4+ T-cell counts and viral load measurements were taken weekly for three weeks, monthly until 12 months and quarterly thereafter. Viral load measurements and CD4+ T cell counts at 12 months post-infection were used as prognostic of disease progression (Lyles *et al.*, 2000). To control for fluctuations in viral loads, set-points at 3, 6 and 12 months were determined as medians of three viral load measurements close to these time-points. Median viral loads and CD4+ T-cell counts at 3 and 12 months post were 4.60 (IQR = 4.01 – 5.06) and 4.34 (IQR = 3.46 – 4.82) copies/ml; and 522 (IQR = 393 – 628) and 410 (IQR = 319 – 560) cells/µl, respectively.

Table 2.1: Characterization of single or dual infection and study participants' clinical data

[a]PID	Phase (P) and Visit (V)	[b]Feibig Stage	Days Post-infection	[c]Gag/Nef Infect. Class.	Viral load (copies/ml)		
					3 Months	6 Months	12 Months
CAP008	P1V7	V	23	Single	188000	98400	39300
CAP030*	P2V0	V	35	Single	604000	146000	57400
CAP037	P1V18	IV	14	Co-infec	39700	34700	104000
CAP045*	P2V0	V	35	Single	9790	<400	556
CAP061	P2V0	VI	57	Single	4380	21000	418
CAP063	P1V0	III	34	Single	305000	186000	164500[d]
CAP065	P2V0	VI	42	Single	59 500	104000[d]	122000
CAP069	P1V12	I/II	14	Single	1390000	673000	1230000
CAP084	P2V0	VI	22	Co-infec	28500	2090	6530
CAP085	P2V0	V	23	Single	43400	6450	3030
CAP088	P2V0	VI	36	Single	38700	45300	38700
CAP129	P1V15	IV	14	Single	21700	25000	80100
CAP136	P2V0	V	28	Single	2080	1680	1865
CAP137	P1V13	IV	14	Co-infec	126000	112000	246000
CAP174	P2V0	VI	28	Single	56600	100000	60700
CAP177	P2V0	III	30	Single	57600	89700	42100
CAP188*	P2V0	IV	26	Single	30900	66900	89600
CAP200	P2V0	V/IV	42	Single	172000	279000	108000
CAP206	P2V0	VI	41	Single	140000	138000	252000
CAP210	P2V0	V	36	Single	34200	132000	376000
CAP211*	P2V0	IV	20	Single	62300	8420	1010
CAP217	P2V0	VI	67	Single	162000	177000	26700
CAP220*	P2V0	V	23	Single	<400	<400	<400
CAP221	P1V10	I/II	14	Single	1230	16800	53100
CAP222	P2V0	V	43	Single	33900	1360	448
CAP224	P2V0	V	43	Single	100000	27200	46300
CAP225	P2V0	IV	23	Single	64400	47600	21500
CAP228	P2V0	VI	53	Single	1330	817	1520

CAP229	P2V0	ND	48	Single	14900	14000	22500
CAP237	P2V0	IV	28	Super	52500	7750	11900
CAP239	P2V0	V	48	Single	16100	59700	156000
CAP244	P2V0	VI	58	Single	39900	26100	14500
CAP248	P2V0	V	62	Single	79400	40600	64600
CAP255	P2V0	VI	54	Single	106000	57800	36400
CAP256	P2V0	VI	42	Super	105000	750000	223000
CAP257	P2V0	V	49	Single	173000	14900	10000
CAP258	P2V0	VI	49	Single	268000	228000	122000
CAP260*	P2V0	V	28	Single	6920	3510	3480[d]
CAP261	P2V0	VI	62	Single	445000	65000	65400
CAP262	P2V0	V	48	Single	13600	3260	7160
CAP264	P2V0	V	83	Single	393000	42300	40500
CAP265*	P2V0	VI	86	Single	1825	517	3380
CAP266	P2V0	VI	14	Single	76700	1640	2010
CAP267	P2V0	V	44	Co-infec	12400	19400	19300
CAP268	P2V0	VI	55	Single	4130	15100	7640
CAP269*	P2V0	VI	25	Single	506	790	980
CAP270	P2V0	V	54	Single	527000	365000	738000
CAP271*	P2V0	VI	40	Single	1620	9680	10100
CAP274	P2V0	VI	72	Single	38270	164000	85300
CAP275*	P2V0	VI	60	Single	63400	9700	2830
CAP277*	P2V0	VI	53	Single	6410	2070	483
CAP278*	P2V0	IV	59	Single	870	625	891
CAP279*	P2V0	VI	54	Single	155000	13300	32200
CAP280	P2V0	V	87	ND	77050	37400	2430
CAP281	P2V0	ND	68	Super	2700	1230	400
CAP282*	P2V0	ND	74	Single	10200	28200	1470
CAP285*	P2V0	ND	99	Single	421500	198000	ND
CAP289*	P2V0	ND	56	Single	11600	21100	528000

Viral load value is median of 3 measurements close to the respective time points; Infection date is estimated as a midpoint between the last negative and first positive antibody test or as 14 days for antibody negative but PCR positive participants; [a]participant's identification number; [b]Feibig *et al.*, 2003; [c]Co-infection classified by Woodman (Woodman *et al.*, 2011), and superinfection (super) classified in this study using the *gag/nef* sequences; *These individuals were characterised as single infections (n=16) based on the maximum pairwise DNA distances of either clonal sequences at enrolment or longitudinal Gag and Nef sequences; [d]Value at 9 months post-infection; ND=Not done. VL <400: Below limit of detection

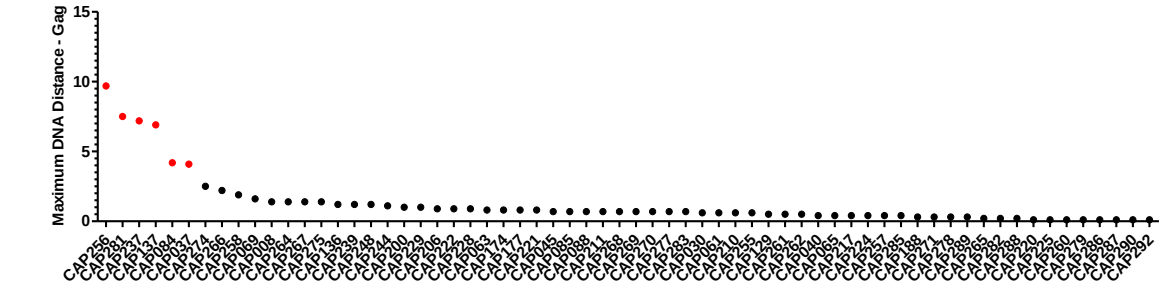
2.3.2 Characterisation of the infecting virus

Altogether, a total of 365 and 289 full-length *gag* and *nef* sequences were generated from 58 participants (median 6 *gag* and 5 *nef* sequences per participant) over the course of the first 12-24 months post-infection. A further 192 full-length *gag* and 204 *nef* sequences were available from 32 participants over the first 6 months of infection (Chopera *et al.*, 2008). All individuals were infected with HIV-1 subtype C (Figure 2.2).

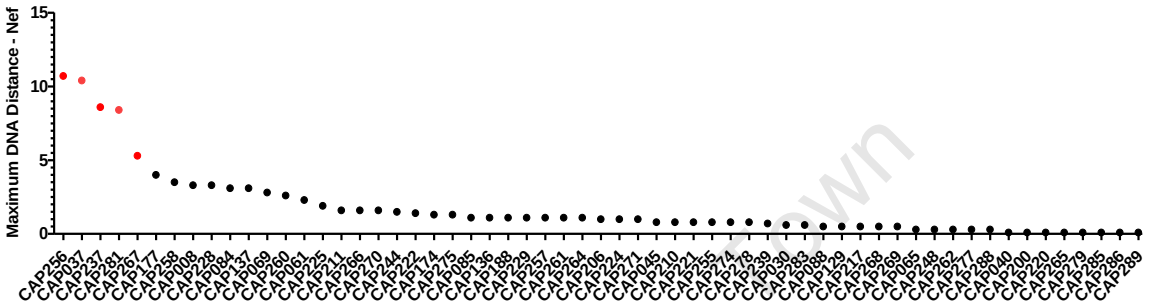
The number of infecting strains following transmission had previously been reported in 39 participants using single genome amplification and sequencing of *env* amplicons (Abrahams *et al.*, 2009); and/or by heteroduplex mobility assay following cloning and sequencing

(Woodman *et al.*, 2011). These studies identified four individuals (CAP037, CAP084, CAP137 and CAP267) as co-infected with two epidemiologically unlinked strains following transmission.

To determine if we could identify dual infection through the analysis of longitudinal sequences, the pairwise distance between the different individuals' sequences for all time-points from enrolment to 12 months post infection was computed (Figure 2.1a and b). We expected individuals infected with a single virus infection to have a normal Gaussian distribution of maximum pairwise DNA distances within the first year post-infection, whereas individuals with dual infection would fall outside of this curve. The group median maximum pairwise DNA distances were 0.7% [IQR 0.4 – 1.2%] and 1.0% [IQR 0.5 – 1.75%] in *gag* and *nef* respectively. Seven individuals fell outside of the normal distribution (≥ extreme outlier values of 3.2% and 5.0% in the *gag* and *nef* respectively) including the four individuals previously identified as dually infected (CAP037, CAP084, CAP137, CAP 267) and a further three (CAP237, CAP256, and CAP281) (Table 2.1). This frequency of superinfection in the first year of infection (5.2%; 3/58) is in agreement with that reported by Kraft *et al.* (2012) in a Zambian subtype C cohort (6.8%; 3/44) although the main purpose of characterizing superinfection in this study was to avoid confusing active CTL escape with superfluous introduction of new residues.



B)



A)

Figure 2.1 Maximum pairwise DNA distance comparison of intrapatient sequences over one year post infection in (A) *gag* and (B) *nef*. Intrapatient distances were calculated using the MEGA package with the Kimura two-parameter model of evolution (Kimura, 1980). Extreme outlier pairwise DNA distances are shown in red colour.

Of the seven individuals with high diversity, four of these individuals separated into distinct clusters in the full-length *gag* and/or *nef* (CAP037, CAP137, CAP256 and CAP281) while a further two separated into distinct clusters in p17/p24 (CAP084 and CAP237) (Figure 2.2a, b, c and d). Although viral diversity was high in CAP267, these sequences clustered together phylogenetically in both the full length *nef* and *gag*/partial *gag* sequences. Nevertheless, CAP267 had previously been identified as infected by more than one HIV-1 viral strains at or very close to the initial infection (co-infected). Three individuals were identified as putative superinfections (SI): CAP256 previously identified by Treurnicht *et al.* (2008), and a further two more individuals, CAP237 and CAP281 were identified in this study. In CAP237, the sequence of the second infecting virus differed from the initial infection by 7.2% and 8.6% in *gag* and *nef* respectively, while in CAP281 the virus differed from the initial infection by 7.5% and 8% in *gag* and *nef* respectively (Table 2.2). Thus looking at diversification over time, this study correctly identified four individuals previously classified as co-infected, and identified a further 3 individuals with putative SI.

Sequences were then analyzed for recombination in order to determine if there was evidence of superinfection prior to first detection. The first evidence of recombination between the two superinfecting strains was detected at 50 weeks for CAP237 with a p17 segment (1-280 nucleotide-lengths) from the sequence at 6 weeks taken up in the sequence at 50 weeks post infection, as estimated using GARD and highlighter plot similarities (Figure 2.3). Similarly, the first evidence of recombination in in CAP256 was detected at 30 weeks post-infection with significant recombination in later sequences. There was no detectable evidence of recombination in CAP281 in either *gag* or *nef* at one year post-infection.

Figure 2.2a:

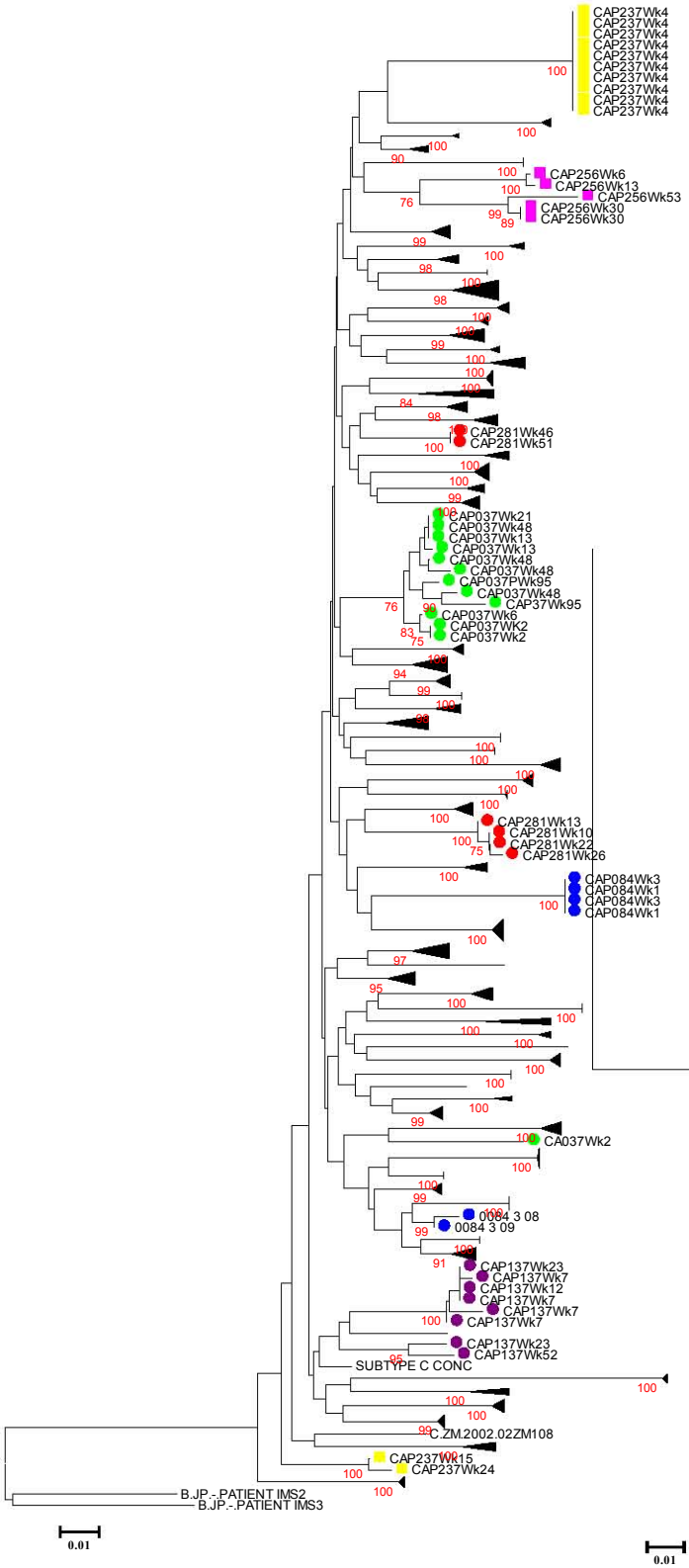


Figure 2.2b

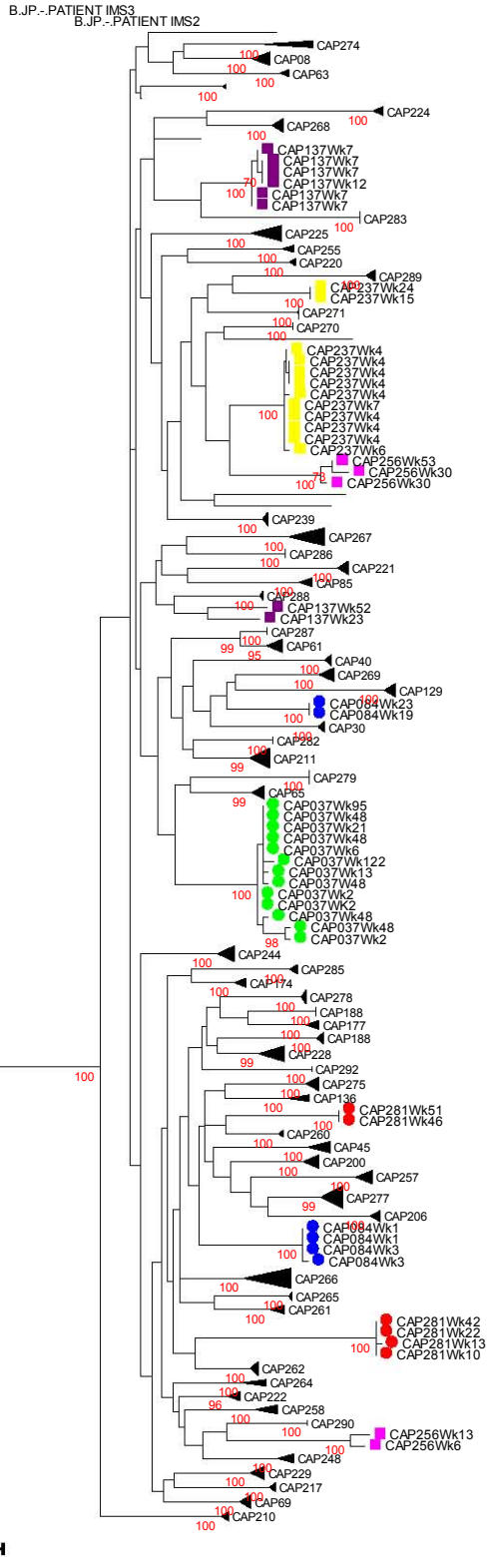


Figure 2.2c:

Figure 2.2d

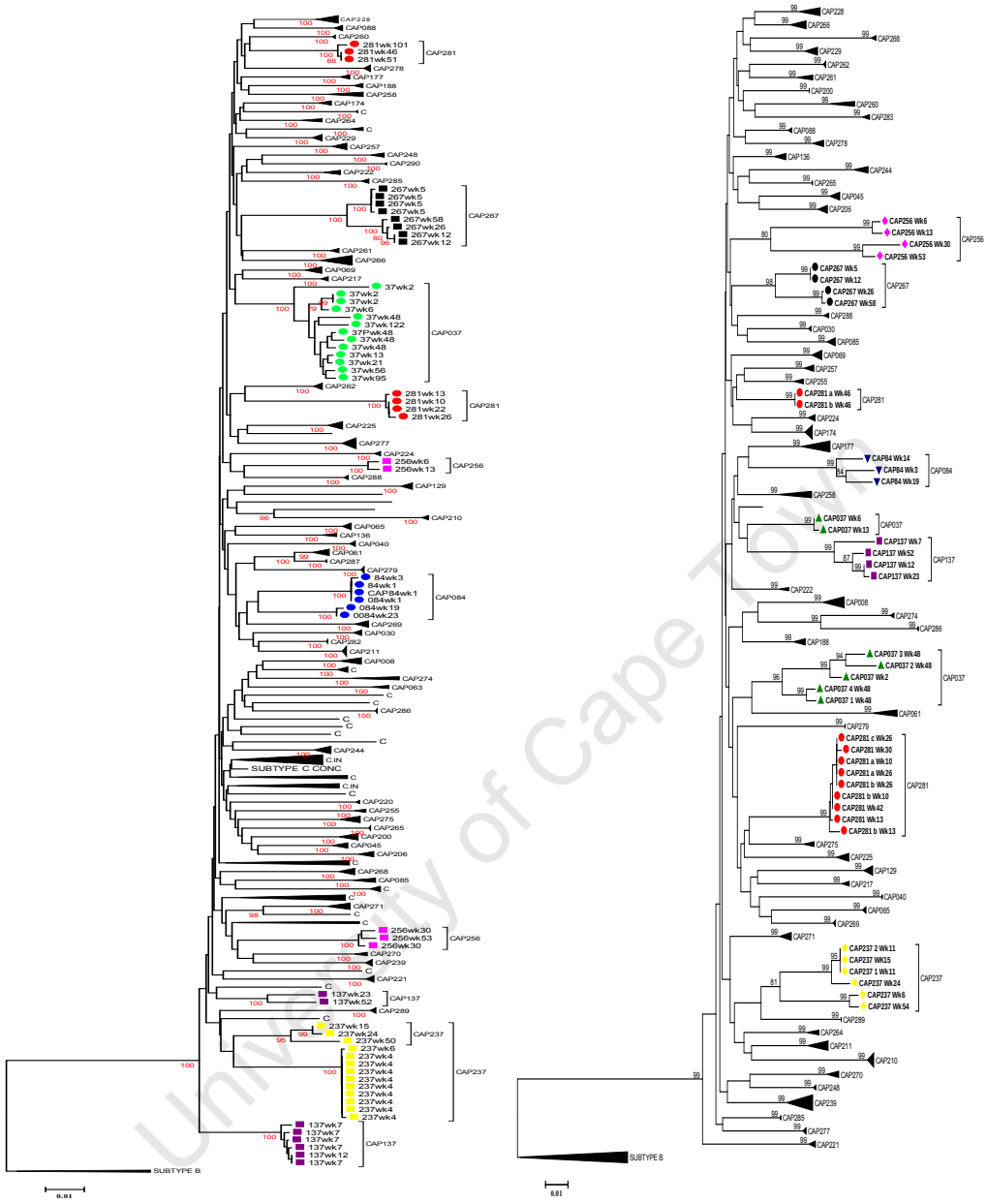
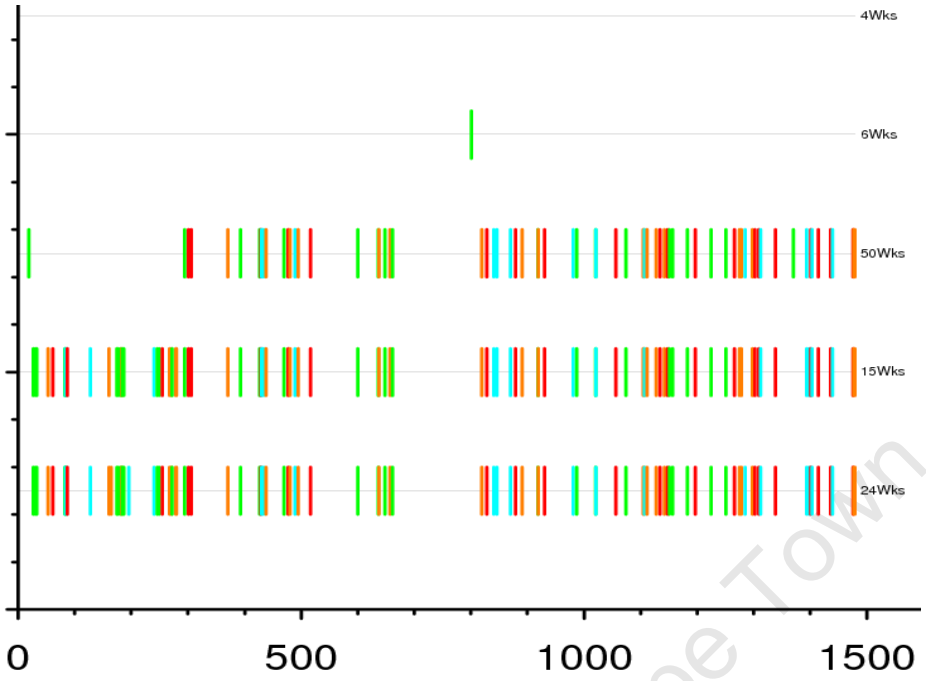
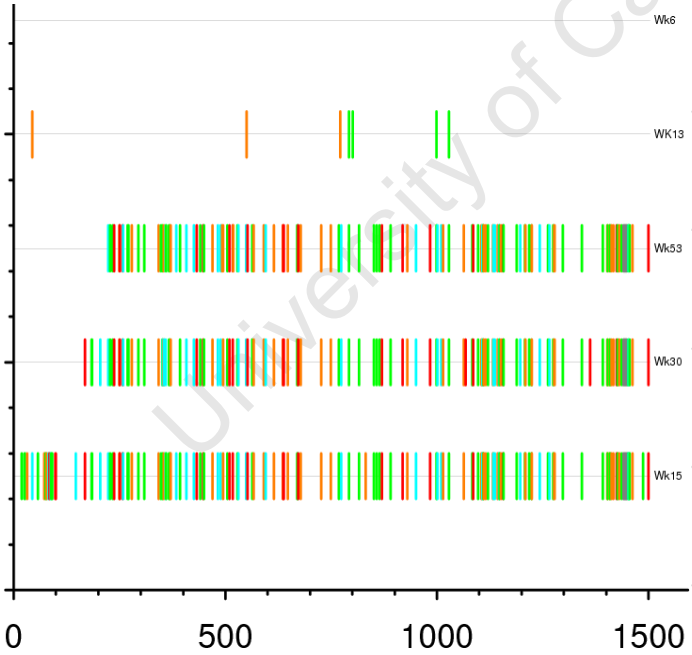


Figure 2.2 Neighbour joining trees showing the phylogenetic clustering of sequences generated from enrolment to a year post infection. Un-labelled sequences are other subtype C sequences from southern Africa. a) Gag p17 and b) p24 nucleotide sequences showing clustering of sequences that are putative dual infections; c) Full lengths longitudinal HIV-1 *gag* nucleotide sequences showing the clustering of sequences that are putative dual infections; d): Full length longitudinal HIV-1 *nef* nucleotide sequences showing clustering of sequences that are putative dual infections; Green: CAP037; Dark Blue: CAP084; Purple: CAP137; Yellow: CAP237; Pink: CAP256; Black: CAP267 and Red: CAP281. Bootstrap values greater than 75% are shown, and scale length of 0.01 is shown as a bar below each tree.

CAP237



CAP256



CAP281

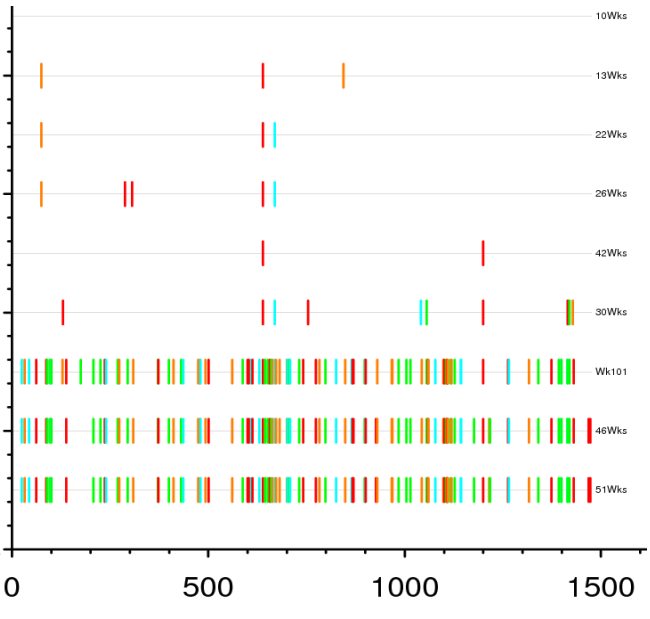


Figure 2.3 Highlighter plot showing longitudinal *gag* sequence similarities in CAP237, CAP256 and CAP281. Later viral sequences for each of the three participants are compared to the individual's earliest available sequence (enrolment); sequences are arranged in a descending order of similarity to the earliest sequence. Nucleotide mismatch are shown as coloured tick (A = green, T = red, G = orange, C = light blue, IUPAC = dark blue, gaps = grey). Numbers on the X-axis scale are positions of the gag nucleotides.

2.3.2.1 Confirmation of superinfection

To confirm superinfection and determine the timing of infection with the second strain, primers specific to unique short sections of the individuals' superinfecting viral sequence, which could not amplify the initial primary infecting virus, were designed. The timing of re-infection was defined as the midpoint between the last PCR negative time-point and the first time-point of detecting the second sequence.

In CAP237, sequencing identified a second incoming strain at week 15, and was not detected at week 4 or 6. Using CAP237 specific primers to the SI virus, there was no amplification detected for samples from 4 and 7 weeks, while the superinfecting virus was detected at weeks 11, suggesting that re-infection occurred between week 7 and 11. An increase of 1.5 copies/mL log viral load occurred between 11 and 13 weeks post infection further supporting this timing (Figure 2.4).

In CAP256, superinfection was first identified at week 15, and was not detected at week 6 or 13. Additionally, typical of superinfection, an increase of 1.7 copies/mL log viral load was observed from week 12 to 15. These timings were later confirmed using CAP256 specific primers by D. Sheward (HIV diversity group of Williamson, IIDMM).

In CAP281, superinfection was first identified at week 46, and was not detected at week 10 or 42. This individual controlled her viral loads to low levels and virus specific primers were designed but were not sensitive enough to detect virus in samples with low viral loads (generally below 1000 copies/ml). However, a significant viral load increase (log 1.1 copies/ml) occurred at 42 weeks post infection, suggesting that superinfection occurred at this time.

Table 2.2 Estimated timing of superinfection in the three participants

PID	Infection Time*[1] (Wks PI)	Superinfection Time*[2] (Wks PI)
CAP237	**2**	**9**
CAP256	**6**	**14**
CAP281	**10**	**42**

Wks PI Weeks Post-Infection

***1** Midpoint between last HIV negative and first HIV positive

***2** Midpoint between first detection of superinfection (SI) and last negative PCR for the SI

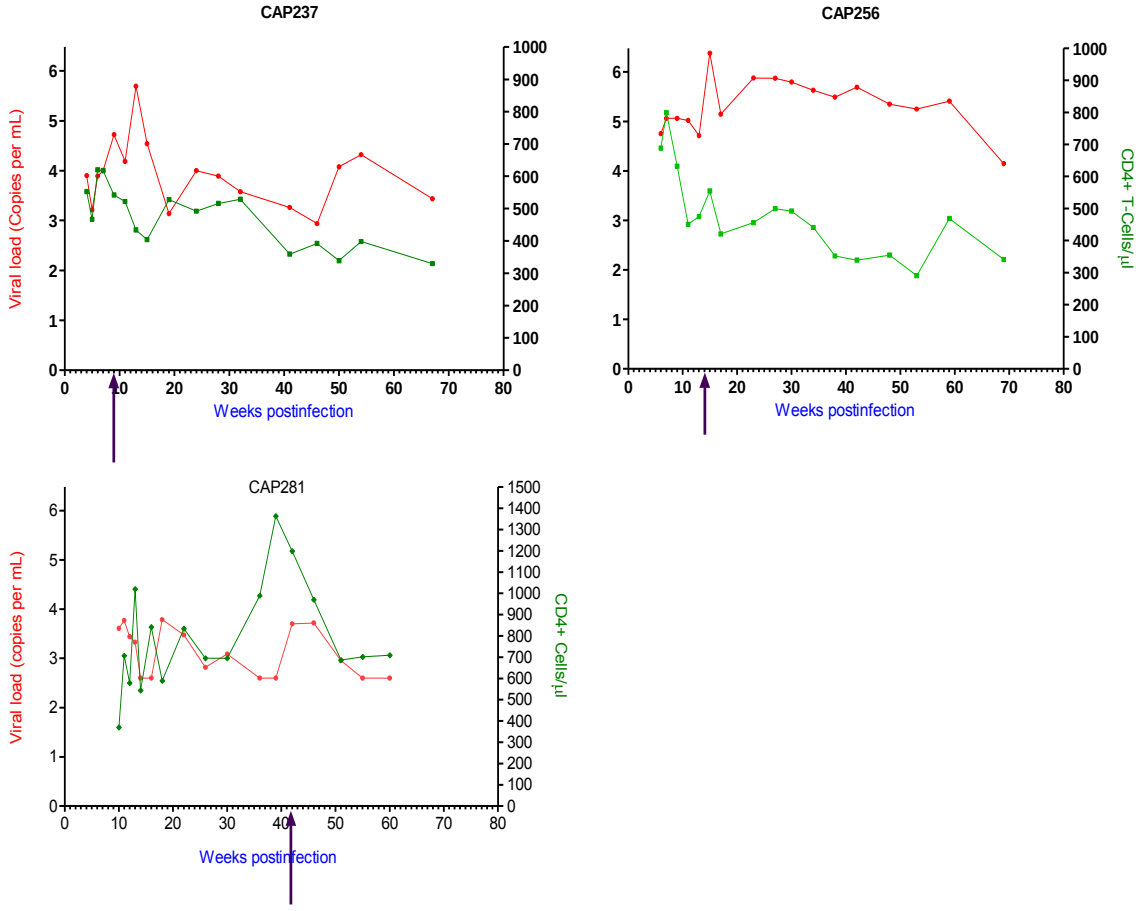


Figure 2.4 Timing of superinfection and impact on disease progression for the superinfected participants, CAP237, CAP256 and CAP281, showing viral load and CD4+ T cell counts. The arrow shows the estimated timing of the new infecting virus or the midpoint between first detection of superinfection (SI) and last negative specific primer PCR for the SI. Characteristic of superinfection, these three individual had a more than one log increase (1.5, 1.7 and 1.1 logs respectively) in viral load at 13, 15 and 42 weeks post infection respectively.

2.3.3 Identification of CTL escape-mutations emerging during the first year of infection

Gag and Nef proteins were selected as they have been shown to elicit dominant immune responses in early infection (Gray *et al.,* 2009a), and as responses to Gag have been shown to be associated with lower viral loads (Berger *et al.,* 2011; Kiepiela *et al.,* 2007). Only individuals with single infection (n=51) were included in this analysis as it was difficult to predict whether changes in epitopes were due to infection with multiple strains or viral evolution in individuals with dual infection or superinfection.

Putative CTL escape mutations were defined as changes that persisted over time in (or within 5 amino acids flanking) known epitopes, defined using the 'A' list of epitopes, restricted by the participants' HLA (Llano *et al.,* 2009; www.hiv.lanl.gov). Significant HLA-associated mutations defined using a large clade C South African cohort occurring in published epitopes were also included in this analysis (Goepfert *et al.,* 2008; Matthews *et al.,* 2008; Rousseau *et al.* 2008). Reversions were defined as changes from low amino acid frequency to high frequency or consensus amino acid in epitopes not restricted by the participants HLA. Gag and Nef IFN-γ ELISpot data was also available for 28 participants (Mlotshwa *et al.,* 2010) (Table 2.3).

2.3.4 Putative escape mutations and CD8+ T-cells recognition of HIV-derived peptides measured by the IFN-γ ELISpot assay

To investigate if mutations associated with immune pressure were accurately predicted, the relationship between putative CTL escape and CD8+ IFN-γ responses measured using IFN-γ ELISpot assay was assessed in 28 participants, whose data was available (Table 2.3; IFN-γ data given by C. Gray: Gray *et al.,* 2009a; Mlotshwa *et al.,* 2010). Individuals' HLAs restricting known epitopes could be mapped out in 94.0% (47/50) of all peptides recognised by CD8+ T cells that were measured by this assay, whereas three peptides were positive in regions of unknown epitopes (Table 2.3). Of these 50 peptides recognised, sequence evolution

was seen in 26 (52%) epitopes, while 24 remained conserved for the first 12 months post-infection.

On an individual level, 64.3% of the 28 individuals (18/28) had escape detected where there were concomitant reductions in IFN-γ ELISpot responses in Gag and/or Nef confirming immune escape. However, there were no IFN-γ ELISpot responses detected in 17.9% (5/28) of these individuals despite the detected CTL escape (HLA-mediated sequence evolution).

Whether the reasons for detecting viral escape with no measureable immune pressure in these five individuals could have been due to differences in the infecting viral sequence and the reagent used to detect responses was investigated. There was one/two mismatch in five of the sequences of the infecting virus and those of the reagent used (n=8) in individuals with escape but without responses, suggesting that use of autologous sequence could have increased response detection.

Taken together, evolution of CTL escape mutations in previously defined Gag and Nef epitopes was detected in 69% of individuals (35/51) at one year post infection (Table 2.4). Escape in Gag was detected in 43% of individuals (22/51), while escape in Nef was detected in 47% individuals (24/51). Escape in both Gag and Nef was observed in 11 (11/51; 22.0%) individuals (Table 2.5).

Table 2.3 Comparison of HLA-mediated sequence evolution in *gag* and *nef* and CTL immune responses, as measured using IFN-γ ELISpot assay (provided by Gray: Gray *et al.*, 2009a; Mlotshwa *et al.*, 2010), commencing in acute infection; Known epitopes are underlined.

PID		[R]Peak	Peptide and Known epitope	HLA	Escape
CAP008	Gag pep4	2695	GKKHYMLKHLVWASREL	B*08:01	+
	Nef pep26/27	2495	IHSKRRQDILDLWVYHTQG	Cw7	+
CAP030	Nef pep20	328	RPMTYKAAFDLSFFL	Unknown	-
	Nef pep21	298	YKAAFDLSFFLKEKG	Unknown	-
	Nef pool2	1990	EEEPEVGFPVRPQVPPLRPM	B*45:01	+
CAP045	Nef pep33/34	2178	PGPGVRYPLTFGWCFKLVP	A*23:01*g	-
	Nef pool2	863	EEEPEVGFPVRPQVPPLRPM	B*45:01	+
CAP085	Gag pep11	420*	TGTEELRSLYNTVATLY	A*30:02	+
	Nef pep29	1638	KRQEILDL-WVYHTQGYFPDWQNY	Cw7	+
	Nef pool2	-	EEEPEVGFPVRPQVPPLRPM	B*45:01	+/No response[M]
CAP088	Gag pep49	465	SHKARVLAEAMSQANSA	B*45:01	+
	Nef pep30	2677.5	TQGYFPDWQNYTPGP	A*29:02	-
	Nef pool2	-	EEEPEVGFPVRPQVPPLRPM	B*45:01	+/No response[M]
CAP129	Nef pep17	138	EVGFPVRPQVPLRPM	B*81:01	+
	Nef pep26	1775	IHSKRRQDILDLWVY	B*18:01	+
	Gag pep40	-	GPKEPFRDYVDRFFKTLR	B*18:01	+/No response
	Gag pep25	-	GATPQDLNTMLNTVGGH	B*81:01	+/No response[M]
CAP174	Nef pep19	923	QVPLRPMTYKAAFDL	A*03:01	+
	Nef pep26	1365	IHSKRRQDILDLWVY	Cw*07:01	+
CAP206	Nef pep32	243	QNYTPGPGRVYPLTF	B*07:02	-
CAP210	Gag pep40	2068	GPKEPFRDYVDRFFKTLR	Cw*03:04	-
	Gag pool2	135	NLQGQMVHQAISPRTLNA	B*15:10	+
CAP217	Gag pep33	2190	SDIAGTTSTLQEQIAWM	B*58:01	+
	Nef pep20	2618	RPMTYKAAFDLSFFL	B*58:01	+
CAP221	Nef pep26	543	IHSKRRQDILDLWVY	Cw*07:01	+
CAP222	Gag pep25	1568	GATPQDLNTMLNTVGGH	B*81:01	-
	Nef pep18	2838	PVRPQVPLRPMTYKA	B*81:01	-
CAP225	Gag pep25	9077.5	GATPQDLNTMLNTVGGH	B*81:01	-
	Nef pep17	1308	EVGFPVRPQVPLRPM	B*81:01	-
	Nef pep32	2468	QNYTPGPGRVYPLTF	B*42:02	-
	Nef pep41	1248	NNCLLHPMSQHGMED	Unknown	+
CAP228	Gag pep23	320	AFSPEVIPMFTALSEGA	A26	-
	Nef pep33	2915	PGPGVRYPLTFGWCF	A*23:01*g	-
CAP229	Gag pep33	1188	SDIAGTTSTLQEQIAWM	B*58:01	+
	Nef pep20	1078	RPMTYKAAFDLSFFL	B*58:01	+
CAP239	Gag pep33	380	SDIAGTTSTLQEQIAWM	B*58:01	+
CAP244	Nef pep33	5735	PGPGVRYPLTFGWCF	A*23:01*g	-
CAP255	Nef pep26	4045	IHSKRRQDILDLWVY	B*18:01	+
	Nef pep33	2688	PGPGVRYPLTFGWCFFKLVP	B*18:01	+
CAP257	Gag pep4	413	RLRPG-GKKHYMLKHLVWASREL	B42	-
	Nef pep33	4928	PGPGVRYPLTFGWCF	A*23:01*g	-
CAP258	Nef pep33	4245	PGPGVRYPLTFGWCF	A*23:01*g	+
CAP260	Nef pep30	2640	TQGYPDWQNYTPGP	A29	-
CAP 261	Nef pep26	300	IHSKRRQDILDLWVY	B*13	+
CAP262	Gag pep25	1335	GATPQDLNTMLNTVGGH	B*81:01	-
	Nef pep18	935	PVRPQVPLRPMTYKA	B*81:01	+
CAP264	Nef pep21	1920	YKAAFDLSFFLKEKG	Cw8	-
CAP266	Gag pep11	1220	TGTEELRSLYNTVATLY	A*29	+
	Nef pep30	6778	TQGYFPDWQNYTPGP	A*29	+
CAP267	Gag pep4	585	RLRPG-GKKHYMLKHLVWASREL	A*03:01	-

	Nef pep33	3065	PGPGVRYPLTFGWCF	A*23:01*g	-
CAP268	Gag pep33	2523	SDIAGTTSTLQEQIAWM	B*58:01	-
	Nef pep26	2898	IHSKRRQDILDLWVY	Cw7	-
	Nef pep18	-	PVRPQVPLRPMTYKA	B*07	+/No response
	Gag pep11	-	TGTEELRSLYNTVATLY	A*02:05	+/No response
	Gag pool 3	-	RLHPVHAGPVAPGQMREP	B*07	+/No response[M]
CAP270	Nef pep20	2065	RPMTYKAAFDLSFFL	B*58:01	+
	Nef pool 1	-	RVIGWPAVRERMRRAR	B*08:01	+/No response[M]

+/- Sequence evolved/did not, respectively; *g Goepfert *et al.*, 2008, JEM; [M]Reagent sequence mismatch the autologous sequence

Table 2.4 Timing, location and HLA restriction of escape mutations and IFN-gamma ELISpot data

PID[a]	Epitope[b]	D. E	Location[H]	[R]P	Timing	HLA[bb]
CAP008	**KRQEILDLWVY**	KY11	Nef (105 - 115)	2495	Acute	Cw*07:01
CAP008	**GGKKQYMLK**	GK9	P17 (Gag 24 - 32)	2695	Acute	B*08:01
CAP030	**EEVGFPVRPQV**	EV11	Nef (64 - 74)	1990	Acute	B*45:01*m
CAP040	**RLRPGGKKHYM**	RY10	p17 (Gag 20 - 29)	ND	Acute	B*42:01
CAP040	**YVDRFFKTL**	YL9	p24 (Gag 296 - 304)	ND	Acute	Cw*03:04
CAP045	**EEVGFPVRPQV**	EV11	Nef (64 - 74)	863	Acute	B*45:01*m
CAP063	**EEVGFPVRPQV**	EV11	Nef (64 - 74)	ND	Acute	B*45:01*m
CAP063	**SLYNTVATL**	SL9	p17(Gag 77 - 85)	ND	Late	A*02:01
CAP085	**EEVGFPVRPQV**	EV11	Nef (64 - 74)	NR	Acute	B*45:01*m
CAP085	RSLYNTVATLY	RY11	p17 (Gag 76 - 86)	420	Late	A*30:02
CAP085	**KRQEILDLWVYHTQGY**	KY11	Nef (64 - 74)	1638	Early	Cw*07:01
CAP088	**EEVGFPVRPQV**	EV11	Nef (64 - 74)	NR	Acute	B*45:01*m
CAP088	**AEAMSQANS**	AS9	p7p/p6 (Gag 364-372)	465	Acute	B*45:01
CAP129	**KRQEILDLWVH**	KY11	Nef (105 - 115)	1775	Early	B*18:01
CAP129	**RPQVPLRPM**	RM9	Nef (71 - 77)	155	Late	B*81:01*m
CAP129	**FRDYVDRFFK**	FK10	p24 (Gag 293 - 302)	NR	Acute	B*18:01
CAP129	**TPQDLNTML**	TL9	P24 (Gag 180 - 188)	NR	Early	B*81:01
CAP174	**KRRDILDLWVY**	KY11	Nef (105 - 115)	1365	Acute	Cw*07:02
CAP174	**QVPLRPMTYK**	QK10	Nef (73 - 82)	923	Late	A*03:01
CAP177	**EEVGFPVRPQV**	EV11	Nef (64 - 74)	ND	Acute	B*45:01*m
CAP188	**RLRPGGRKHY**	RY10	p17 (Gag 20 - 29)	ND	Late	A*74*r
CAP188	**SLYNTVATL**	SL9	p17 (Gag 77 - 85)	ND	Acute	A*02:02
CAP210	**HQAISPRTL**	HL9	p24 (Gag 144 - 152)	135	Late	B*15:10
CAP211	**TSTLQEQIAW**	TW10	p24 (Gag 240 - 249)	ND	Acute	B*57:03
CAP211	**ISPRTLNAW**	IW9	p24 (Gag 147 -155)	ND	Acute	B*57:03
CAP217	**KAAFDLSFF**	KF9	Nef (82 - 90)	3529	Acute	B*58:01
CAP217	**TSTLQEQITW**	TW10	p24 (Gag 240 - 249)	2190	Acute	B*58:01
CAP221	**KRQEILDLWVY**	KY11	Nef (105 - 115)	543	Late	Cw*07:01
CAP224	**AAFDLSFFLK**	AK10	Nef (84 - 92)	ND	Early	A*03:01
CAP228	**KRQEILDLWVY**	KY11	Nef (105 - 115)	NR	Acute	Cw*07:01
CAP229	**KAAFDLSFF**	KF9	Nef (82 - 90)	1808	Early	B*58:01*m
CAP229	**RSLYNTVATLY**	RY11	p17 (Gag 76 - 86)	NR	Early	B58
CAP229	**TSTLQEQIAW**	TW10	p24 (Gag 240 - 249)	1188	Acute	B*58:01
CAP239	**TSTLQEQIAW**	TW10	p24 (Gag 240 - 249)	380	Acute	B*58:01
CAP248	**VKVIEEKAF**	VF9	p24 (Gag 156 - 164)	ND	Late	B*15:03
CAP255	**KRQEILDLWVH**	KY11	Nef (105 - 115)	4045	Early	B*18:01
CAP255	**VRYPLTFGWCY**	YY9	Nef (135–143)	2688	Acute	B*18:01
CAP258	**RYPLTFGW**	RW8	(Nef 134 – 141)	4245	Early	A*23:01
CAP261	**RQEILDLWV**	RV9	Nef (106 - 114)	300	Acute	B*13:02

CAP262	**RPQVPLRPM**	RM9	Nef (71 - 77)	935	Early	B*81:01*m
CAP265	**TPGPGVRYPL**	TL10	Nef (128 - 137)	ND	Acute	B*42:01
CAP266	**YFPDWQNYT**	YT9	Nef (120 - 128)	6778	Acute	A29
CAP266	**LYNTVATL**	LL8	p17 (Gag 78 - 85)	1220	Acute	A29
CAP268	**RPQVPLRPM**	RM9	Nef (71 - 77)	NR	Early	B*07*r
CAP268	**SLYNTVATL**	SL9	p17 (Gag 77 - 85)	NR	Acute	A*02:05
CAP268	**HPVHAGPIA**	HA9	p24 (Gag 216 -224)	NR	Late	B*07
CAP269	**RPQVPLRPM**	RM9	Nef (71 - 77)	ND	Acute	B*07:02
CAP269	**SPRTLNAWV**	SV9	p24 (Gag 148 - 156)	ND	Acute	B*07:02
CAP270	**WPTVRERM**	WM8	Nef (13 -20)	NR	Acute	B*08:01
	KAAFDLSFF	KF9	Nef (82 - 90)	2065	Late	B*58:01
CAP274	**KAAFDLSFF**	KF9	Nef (82 - 90)	ND	Acute	B*58:01
CAP274	**TSTLQEQIAW**	TW10	p24 (Gag 240 - 249)	ND	Acute	B*58:01
CAP275	**RLRPGGKKHY**	RY10	p17 (Gag 20 - 29)	ND	Acute	B*42:01*m
CAP277	**TPQDLNTML**	TL9	p24 (Gag 180 -188)	ND	Late	B*81:01
CAP278	**RPQVPLRPM**	RM9	Nef (71 - 77)	ND	Late	B*39:10
CAP278	**TPQDLNTML**	TL9	p24 (Gag 180 -188)	ND	Acute	B*39:10
CAP282	**TPQDLNTML**	TL9	p24 (Gag 180 -188)	ND	Late	B*42:01
CAP289	**TPQDLNTML**	TL9	p24 (Gag 180 -188)	ND	Early	B*39:10

[a]Participant identification; [b]Escape residue in each epitope underlined; Defined epitope (D. E) [bb]Associated HLA; [H]Defined epitope with residues based on HXB2 sequences in brackets; [R]Peak responses (SFU/10^6 PBMC); [**]Escape sequence is present at enrolment (CAP040 – 3 weeks (Wks); CAP255 – 8 Weeks; CAP265 – 12 Wks; CAP 278 – 9 Wks post-infection); *g Goepfert *et al.,* 2008; *r Rousseau *et al.,* 2008*;* *m Matthews *et al.,* 2008*;* ND: Not done; NR: No response

Table 2.5 Participants with viral escape in Gag and Nef in the first 12 month of infection

Location	Number of individuals with viral escape	%
p17	6	12.0
p24	12	24.0
p17+24	3	5.9
P7p6	1	2.0
Nef	24	47.1
Gag+Nef	11	22.0

2.2.5 Escape occurs more frequently in the acute/early than in the late phase of infection

Timing of CTL escape was followed up in order to estimate the frequency and pattern of escape in the first year of infection. CTL escape mutations were classified according to the infection phase of their occurrence as either acute (≤ 3 months), early (> 3 to ≤ 6 months) or late (> 6 to 12 month).

The frequency of escape in acute infection was higher than in early infection: out of the 58 epitopes identified as having putative CTL escape mutations, 34 (59.0%) were identified in acute infection, a further 11 evolved in early infection, while 13 evolved in the late phase of infection (Table 2.6).

Table 2.6 Timing and number of epitopes developing CTL escape mutations in Gag and Nef in the first 12 months post infection

Location	Gag	Nef	Total
Acute	17	17	34
Early	3	8	11
Late	5	8	13
Total	25	33	58

Within the first 12 months, the rate of escape was significantly higher in acute when compared to the early/late infection-phases (Figure 2.5; 3.8 versus 1.1 escape/month; p=0.0231). Furthermore, when corrected for amino acid length and in individuals escaping per month, *nef* escape was more frequent than *gag* escape although this was not statistically significant (Figure 2.5; 2.7 versus 1.1 escapes/month/100 amino acids; p>0.1824). Taken together, the high frequency of escape in acute compared to early/late infection suggests that CTLs exert considerable pressure on the virus driving early viral diversification during the first year of the infection.

The median time to escape in Gag (estimated as a mid-point between the last sequenced time point and the timing of the first observation of the escape mutation) was 12 weeks (IQR: 8 – 33), while it was 10.5 weeks in Nef (IQR: 8.3 – 17) weeks post infections. In the Gag, however, while escape/escape intermediate mutations were generally observed between 8 and 12 weeks post infection in the majority, in some individuals like the B*5801/B*5703-positive participants in this study, it was observed as early as 2 weeks (14 days) post-infection. This was contrary to escape in the B7 immunodominant TL9 epitope, where escape was generally observed later (about 10 months post-infection). These issues will be examined in details in chapter 3 of this thesis.

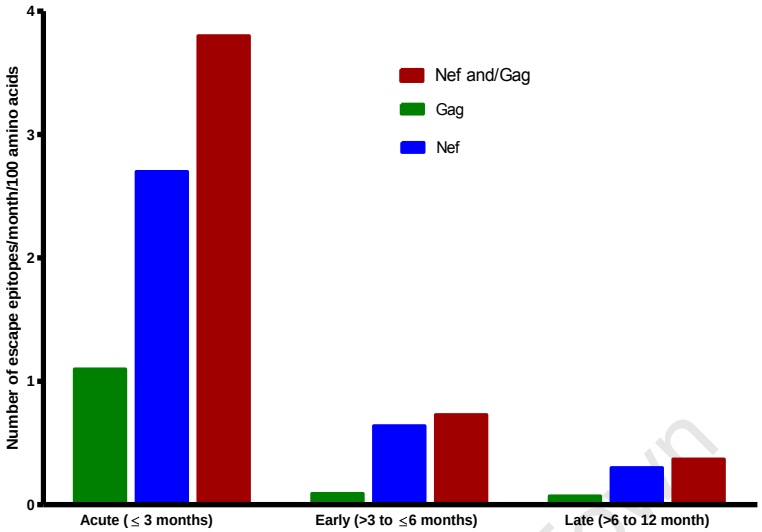


Figure 2.5 Frequency of escape in Gag and Nef proteins in the first year of HIV-1 subtype C infection

2.2.6 Impact of CTL escape on disease progression

2.2.6.1 Acute/early CTL escape associates with higher CD4+ T-cell counts

To determine the impact of CTL escape on disease progression, markers of disease progression including viral loads and CD4+ T-cell counts at 3, 6 and 12 months post-infection were first compared between individuals with viruses developing at least one escape in either Gag or Nef to those individuals where there was no evidence of escape. To explore the effect of the timing of escape on disease progression, we stratified participants according to timing of escape and compared viral loads and CD4+ T cell counts at different time-points (3, 6 and 12 months post-infection).

Although the median viral loads were similar between individuals with and without viral escape (Table 2.7A), participants infected with viruses that evolved acute and early escape had significantly higher CD4+ T-cell counts at 3 and 6 months post-infection (Table 2.7B; p=0.0018 and 0.0542 respectively) suggesting at least in acute and early infection that escape provided contemporaneous benefit. However this benefit was not seen at 12 months post

infection when all individuals with any evidence of viral escape were compared to individuals where the virus never escaped.

Table 2.7 Relationship between escape and contemporaneous viral load and CD4+ T-cell counts

A

Time post infection	No escape Median (IQR)	Escape Median (IQR)	p-value
3 Months	4.60 (3.84 – 5.19) [n=27]	4.54 (4.01 – 4.86 [n=24]	0.685
6 Months	4.37 (4.02 – 5.10) [n=24]	4.40 (3.51 – 4.95) [n=27]	0.485
12 Months[a]	4.29 (3.30 – 4.87) [n=16]	4.58 (3.41 – 4.84) [n=35]	0.556

B

Time post infection	No escape Median (IQR)	Escape Median (IQR)	p-value
3 Months	461 (330 - 548) [n=27]	604 (478 - 735) [n=24]	0.0018
6 Months	415 (330 - 522) [n=24]	522 (389 - 624) [n=27]	0.0542
12 Months[a]	367 (303 - 624) [n=16]	474 (348 - 569) [n=35]	0.7155

A) Medians with Interquartile range (IQR) of Log_{10} viral load consecutively compared at 3, 6 and 12 months post-infection between individuals without and with viral escape; B) Median CD4+ T-cell counts and Interquartile range (IQR) consecutively compared at 3, 6 and 12 months post-infection between individuals without and with viral escape; [a] the analysis at 12 months post-infection constitute individuals whose viruses escape in acute, early and late infection

2.2.6.2 Rates of loss of viral load and CD4+ T-cell counts in individuals with and without viral escape

To further investigate the overall effect of escape on markers of disease progression, linear mixed models, adjusting for repeated measurements for every participant, were fitted to viral loads and CD4+ T-cell counts data. Data in the 0 to 3 (acute) and >3 to 12 (early/late) months

post-infection time periods were separately analyzed in order to assess the effect of acute CTL escape on viral load and CD4+ count and their changes over time. Viral load was log-transformed to ensure normality and adjusted for time post infection, CD4+ counts (in the viral load model) and viral load (in the CD4 counts model).

There was no significant difference in the rate of change or reduction in viral load between participants infected with viruses that escaped CTL responses in either the 0 to 3 or the >3 to 12 months post infection when compared to participants without viral escape in 12 months post-infection in Gag and/or Nef (Table 2.8A; p=0.6029 and p=0.4391, respectively; Figure 2.6A). Furthermore, there was no significant difference in the overall mean viral load between participants infected with viruses that escaped CTL responses in either the 0 to 3 or the >3 to 12 months when compared with participants without viral escape at 12 months post infection (Table 2.8B; p=0.7422 and p=0.8491, respectively).

However, the rate of change of CD4+ T-cell counts in participants infected with viruses that escaped CTL responses differed between the two groups (Table 2.8C; p=0.0439; Figure 2.6B). In the escape group, CD4+ T-cell counts in the first three (0 to 3) months post-infection increased at a rate of approximately 37 cells/µl per month, while they were decreasing at a rate of 16 cells/µl per month for those with no viral escape. Moreover, although this trend was not sustained after acute infection where there was a similar rate of CD4+ decline between individuals with and without viral escape in early and late (>3 to 12 months) infection periods (p=0.9977), the overall mean CD4+ T-cell counts between participants infected with viruses that escaped CTL responses in either the 0 to 3 or the >3 to 12 months were significantly different when compared to that of participants without viral escape at 12 months post infection (Table 2.8D; p=0.0117 and p=0.0274, respectively).
These results highlight the importance of the CTL responses that associate with selection of mutations very early in HIV infection.

Table 2.8 Contemporaneous benefit of acute escape in Gag and/Nef on subsequent disease progression

	Escape	**No escape @ 12 months PI**	
A	**Rate** Log_{10} VL (copies/ml/month	Rate Log_{10} VL (copies/ml/month	p-value
0 -3 months PI	0.16 (SE=0.08)	-0.21 (SE=0.09)	0.6029
>3 to 12 months PI	-0.03 (SE=0.01)	-0.04 (SE=0.01)	0.4391
B	**Mean** Log_{10} VL (copies/ml)	**Mean** Log_{10} VL (copies/ml)	
0 -3 months PI	4.57 (SE=0.17)	4.66 (SE=0.21)	0.7422
>3 to 12 months PI	4.23 (SE=0.15)	4.18 (SE=0.19)	0.8491

	Escape	**No escape @ 12 months PI**	
C	**Rate CD4+ (cells/µl/month)**	**Rate CD4+ (cells/µl/month)**	p-value
0 -3 months PI	36.65 (SE=17.43)	-16.16 (SE=19.78)	**0.0439**
>3 to 12 months PI	-6.55 (SE=2.10)	-6.54 (SE=2.52)	0.9977
D	**Mean (cells/µl)**	**Mean (cells/µl)**	
0 -3 months PI	627 (SE=35.75)	478 (SE=43.84)	**0.0117**
>3 to 12 months PI	583 (SE=34.03)	460 (SE=41.67)	**0.0274**

A) Rates of change in viral load (VL); B) The overall mean VL in each time period estimated using the model; C) Rates of change in CD4+ T cell counts; D) The overall mean CD4+ T cell counts in each time period estimated using the model; Negative (-)/positive (+) signifies decrease/increase of CD4+ T cell counts, respectively; PI: post infection; SE: standard error.

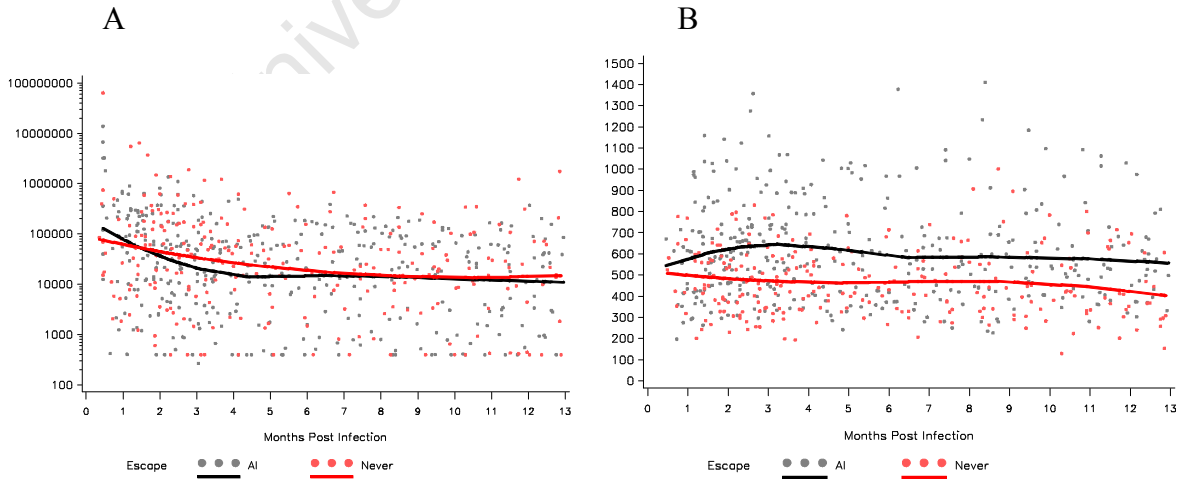


Figure 2.6 Rate of loss of/reduction in viral load (A) and CD4+ T-cell counts (B) compared in individuals with and without viral escape. AI: acute escape; Graphs generated by Lise Werner (CAPRSA, UKZN)

2.2.6.3 Transient correlations of CTL escape in Gag/Gag-Nef with higher CD4+ counts but not with viral load

To determine what CTL responses were impacting on disease progression, viral load and CD4+ T-cell counts in participants with and without viral escape in Gag or Nef epitopes were separately compared.

There were no significant associations in viral load between individuals with and without viral escape in Gag or Nef either at 3, 6 or 12 month post-infection (Figures 2.7A, C, and D). However, when compared to those that did not escape, participants infected with viruses that developed mutations in Gag had a trend towards higher CD4+ T-cell counts, while those with viral escape in both Gag and Nef had significantly higher CD4+ T-cell counts at 3 months post-infection (Figure 2.7 B; p=0.0983 and 0.0248, respectively). Although these results were not sustained at 12 months post-infection, there was still a trend for higher CD4+ T cell counts at 6 months post-infection for participants with viral escape in both Gag and Nef (p=0.068), (Figure 2.7 C and F). These results suggest that escape in both Gag and Nef impacts on disease progression.

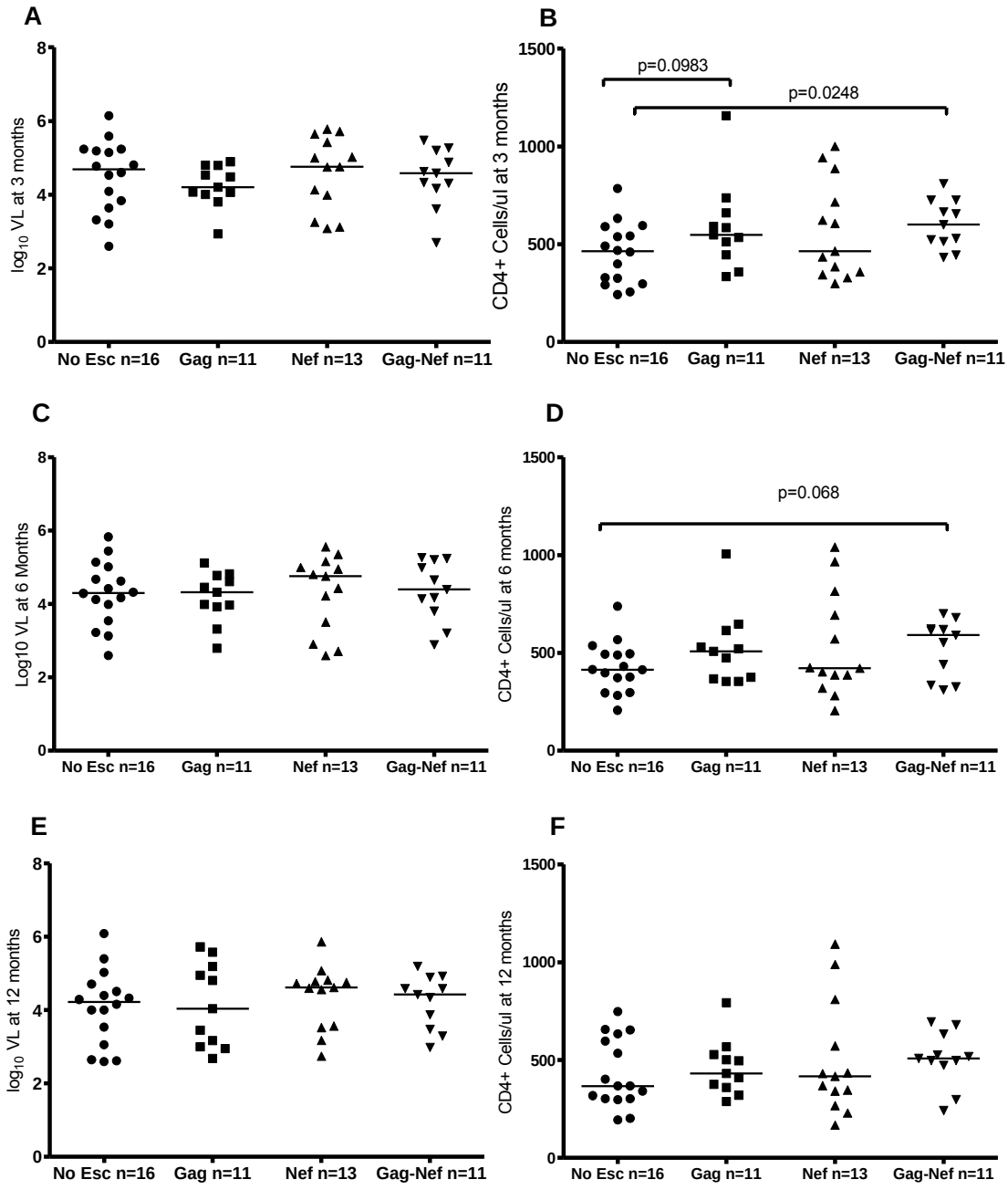


Figure 2.7: Impact of viral escape on markers of disease progression. Figures A, C and E: Log_{10} viral load 3, 6 and 12 months post-infection compared between individuals with and without viral escape in Gag, Nef and both Gag-Nef. Figures B, D and F: CD4+ T-cell counts at 3, 6 and 12 months post-infection compared between individuals with viral and without viral escape in Gag, Nef and both Gag-Nef.

2.2.6.4 HLA-B alleles drive early sequence evolution in HIV-1 infection

HLAs differentially contribute to the total HIV-1-specific CTL response with HLA-B-restricted CTL responses making a stronger contribution and exerting a greater selection pressure compared to HLA-A restricted responses (Kiepiela *et al.,* 2004; Altfeld *et al.,* 2006 Brumme *et al.,* 2008).

Of the 50 Gag and/or Nef reactive peptides identified in the 28 participants with IFN-γ production data available, 52% (26/50) were restricted by HLA-B, while 28% (14/50), 14% (7/50) and 6% (3/50) were restricted by HLA-A, HLA-Cw and unknown-HLAs, respectively (Table 2.3). To determine if differential contribution of HLA-associated polymorphisms could be detected in primary infection in this cohort, HLA mediated sequence evolution in epitopes in the first 12 months of infection were analyzed.

Of the 28 escape mutations identified in Gag, five, 21 and two were in epitopes restricted by HLA-A, HLA-B and HLA-Cw respectively; and of the 30 escape mutations in Nef, 22 were located in epitopes restricted by HLA-B, whereas three and five were in epitopes restricted by HLA-A and HLA-Cw respectively (Table 2.9).

Table 2.9 Differential contribution of HLAs to mutation-selection in Gag and Nef in the first 12 months of HIV-1 subtype C infection

Location	HLA-A	HLA-B	HLA-C
Gag	5	21	2
Nef	5	21	4
Total	10	42	6

The contribution of different HLAs to the number of mutations selected did not significantly differ between *gag* and *nef* genes (p=0.1922). However, there was a significantly higher contribution to the number of mutations selected by HLA-B when compared to other HLAs (HLA-B [72.4%; 42/58] versus both HLA-A [17.2%; 10/58] and HLA-C [10.4%; 6/58];

p=0.0134). Overall, HLA differences accounted for 77.04% of the total variance in the number of mutations selected in the first year of infection.

Next, the differential HLA bias towards selecting mutations in different parts of Gag was determined since response to this protein is associated with better HIV infection control outcome (Kiepiela *et al.*, 2007; Rolland *et al.*, 2008). There was generally no HLA-bias towards selection of mutations in p17; however, 94% (16/17) of the epitopes that evolved sequences in p24 were HLA-B restricted (Table 2.10).

Table 2.10 Contribution of HLAs to mutation-selection in the different Gag proteins in the first 12 months of HIV subtype C infection

Location	HLA-A	HLA-B	HLA-C	Total
p17	5	4	1	**10**
p24	0	16	1	**17**
p2p7p1p6	0	1	0	**1**
Total	**5**	**21**	**2**	**28**

Studies on South African HIV subtype C infection have found that HLA-B*57/58:01 and HLA-B*39:10/81:01 are beneficial, while HLA B*18/08:01 and B*58:02 are deleterious (Kiepiela *et al.,* 2004; Leslie *et al.,* 2010). Of the 16 evolving epitopes in p24 that were HLA-B-mediated, 10 (63%) were driven by these beneficial HLA-B alleles. Participants positive for HLA-B*57/58:01 were more likely than others to evolve sequences in acute infection, while HLA-B*39:10/81:01-positive participants were more likely to evolve sequences in the B7 immunodominant TL9 epitope in late/after 6 months of infection (5/7). Escape in the B*58:01 predominantly occurred in the immunodominant TW10 epitope in p24 in acute/the first 3 months post-infection, while it occurred in both ISW9 and TW10 in the B*57:03. In the seven B*39:10/81:01-positive participants, four evolved late TL9 escape mutations (CAP129, CAP225, CAP277 and CAP289). The role of B*81:01 in controlling viral replication will be presented in detail in Chapter 3.

HLA-B mediated sequence evolution in *nef* epitopes was also common (Table 2.9): of the 21 evolving epitopes in *nef* that were HLA-B-mediated, seven (33.3%) were driven by beneficial

HLA-B alleles, while 10 (47.6%) were mediated by deleterious HLA B*18/08:01. The remaining four (19.1%) were mediated by HLA-B*45:01/42:01/07:02. HLA-mediated evolution in Nef was most likely to occur in the B*45:01 restricted EV11 epitope than in any other epitope, as all the six participants positive for this allele evolved sequence in acute infection.

Overall these results suggest that the association of HLA-B with stronger selection pressure in Gag p24 is driven by beneficial HLA alleles and contribute to our understanding as to why HLA-B alleles are associated with better infection outcome in the Gag. These results also support reports demonstrating a greater selection pressure by HLA-B compared to other HLA class 1 alleles (Kiepiela *et al.*, 2004).

2.2.6.5 HLA-B restricted CTL escape in Nef and Gag associate with modestly higher CD4+T-cell counts at 12 months post-infection

Studies reporting the dominant role of HLA-B were done in chronic infection, which may be enriched with participants controlling the infection (Kiepiela *et al.,* 2004). The impact of HLA-B associated mutations selected in the first year of infection was determined in this thesis. Participants additionally selecting mutations with other HLA types other than HLA-B (n=5) were excluded in this analysis in order to better refine the role of HLA-B although results of this analysis were not affected when participants that selected mutations with both other HLAs and HLA-B were included.

There was no evidence (either in Gag or Nef) to support an association of HLA-B restricted escape with viral load at 12 months post infection (Figure 2.8A). However, there were trends for higher CD4+ cell counts for participants with HLA-B restricted escape in both Gag and Nef, and for participants with HLA-B restricted escape in both Gag and Nef (Figure 2.8B, p=0.076 and 0.08 respectively).

Furthermore, there were no detectable differences in viral load between participants infected with virus that developed escape in either p17 or p24 Gag proteins compared to those that did

not (Figure 2.8C). There was however, a trend towards higher CD4+T-cell counts for individuals infected with virus that developed escape mutations in p24 at one year post-infection (Figure 2.8D p=0.2; medians: 500 cells/µl [IQR = 481 – 646] versus 368 cells/µl [IQR = 301 – 606]) although this was not significant.

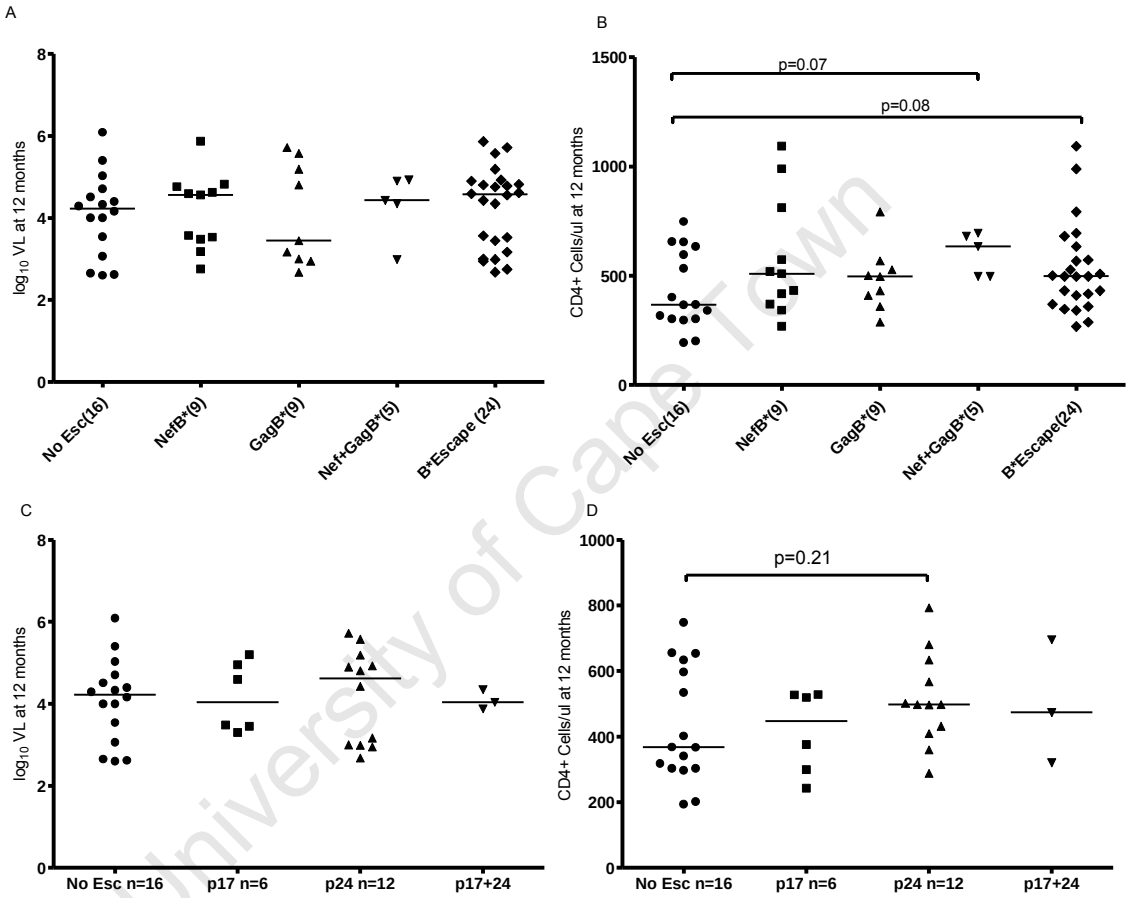


Figure 2.8: Impact of HLA-B restricted viral escape on markers of disease progression in primary infection. Figures A and B: Log_{10} viral load and CD4+ T-cell counts compared between individuals without escape (n=16) and with HLA-B restricted viral escape in Gag, Nef or both. Figures C and D: Log_{10} viral load and CD4+ T-cell counts compared between individuals without and with viral escape in p17, p24 or both - Individuals with both HLA-B and HLA-A or HLA-Cw restricted escape mutations were excluded from this analysis.

2.2.6.6 Association of high CD4+ T-cell counts with acute escape is driven by HLA-B

To better understand how escape effected disease progression in the 51 participants with single infection in this study, two groups were compared: those with high CD4+ T-cell counts (≥ 500 cells/µl at 12 months post infection (n=20)] and those with low CD4+ T-cell counts (< 350 cells/µl at 12 months post infection (n=16). Fifteen individuals with intermediate CD4+ T-cell counts (<500>350 cells/µl) at 12 months post infection were excluded from this analysis. Of the 20 with high CD4+ T-cell counts, 14 (70%) were infected with viruses that escaped CTL responses in Gag and/or Nef, of which 12 (86%) had acute viral escape, while of the 16 individuals with low CD4+ T-cell counts, 8 (50%) were infected with viruses that evolved CTL escape mutations in Gag and/or Nef.

In thirteen of the 14 participants with the high CD4+ T cell counts (13/14) group, viral escape was restricted by HLA-B alleles, which included 11/12 (92%) participants harbouring viruses that evolved acute CTL escape mutations. These included: in Nef, EV11 epitope restricted by B*45:01(CAP030; CAP045; CAP085 & CAP088); RM9 restricted by B*07:02/39:10/42:01 (CAP269, CAP278 & CAP129) and TL10 restricted by B*42:01 (CAP265); and in Gag, RY10 (p17) epitope restricted by B*42:01 (CAP275); TL9 (p24) restricted by B*39:10 & 81:01 (CAP278 & CAP129); SV9 (p24) restricted by B*07:02 (CAP269) and TW10 (p24) restricted by B*58:01(CAP229 &CAP239) (Table 2.4).

Interestingly, the EV11 and the RM9 epitopes that overlap each other and restricted by the B*45:01 and B7 alleles, respectively contain residues that mediate Nef down regulation of MHC and CD4 (Hung *et al.,* 2007), while the RY10 epitope contains the p17 functional basic residues required for nuclear import and export (Dupont *et al.,* 1999).

Discussion

The identification of CTL escape through monitoring sequence evolution associated with the individuals' HLAs provides a sensitive tool to evaluate host immune pressure (Moore *et al.,* 2002; Brumme *et al.,* 2007; Wang *et al.,* 2009). In this study, viral evolution was monitored in 58 individuals from close to infection and at least one year post infection. We confirmed the identification of four individuals who were co-infected with two strains at the earliest time-point, and furthermore identified three individuals who were superinfected within the first year. These 7 individuals were excluded from further analysis due to difficulties in predicting mutations associated with immune escape in individuals with dual infection. This chapter reports on viral escape in Nef and Gag in the 51 participants with single infection.

Here, the epitopes were predicted using the well defined list of optimal epitopes (Frahm, Linde and Brander, www.hiv.lanl.gov; Llano *et al.,* 2009), and HLA-associated residues that were previously identified in published epitopes using a large clade C South African cohort (Rousseau *et al.* 2008; Goepfert *et al.,* 2008; Matthews *et al.,* 2008). These predictions were supported by the fact that individuals' HLA-restricted known epitopes could be mapped out in 94% of all the positive peptides used in IFN-γ ELISpost assay, and also as the majority of participants (64.3%) in this study with predicted escape had concomitant loss/reduction in IFN-γ response support the approach used in this study. The detection of viral escape in the absence of detectable IFN-γ ELISpot responses could be due to a mismatch in peptides reagent and the infecting virus; missed responses due to fluctuation in cellular immune responses (Mlotshwa *et al.*, 2010); or single cytokine IFN-γ ELISpot assay is not detecting all T-cell responses (Boulet *et al.*, 2007).

Despite the increasing breadth and magnitude of responses in the first year of infection (Gray *et al.,* 2009a), escape was more frequent in acute when compared to chronic infection. CTL responses are crucial to the initial control of viremia in acute infection (Borrow *et al.*, 1994; Freel *et al.*, 2012; Goonetilleke *et al.*, 2009; Henn *et al.*, 2012; Kim *et al.*, 2010; Koup *et al.*, 1994), and here we demonstrate that these early focused responses are a major driver of early sequence evolution in the HIV infection. It is possible that the higher frequency of escape in

acute infection was driven by the higher viral loads (increased viral replication), which allow increased rates of viral evolution and CTL escape (Mudd *et al.*, 2012), and this together with functionally avid CTL responses in acute infection (O'Connor *et al.*, 2002) may both be contributing to these rapid evolution of CTL epitopes.

Alleles associated with slow disease progression contribute strongly to the initial CTL response to HIV-infection (Altfeld *et al.,* 2006; Wang *et al.,* 2009) and select mutations in conserved regions, mainly Gag p24 (Borgans *et al.,* 2007; Kiepiela *et al.,* 2007; Streeck *et al.,* 2007). Furthermore, HLA-B is more commonly associated with control and has greater selective pressure than others (Kiepiela *et al.,* 2004; Leslie *et al.,* 2010; Carson *et al.,* 2012). This study has further documented that 72.4% of the number of mutations selected in the first year of infection was attributed to HLA-B alone, with 43% (6/14) of individuals having high CD4+ T-cell counts (≥ 500 cell/µl) harbouring viruses with mutations located in p24 and selected by beneficial HLA-B alleles. The result showing that HLA-B and in particular beneficial HLAs are biased towards selecting mutations in the conserved p24 strongly contribute to our understanding why studies have consistently associated HLA-B with reduced HIV-1 infection severity.

This data augment previous finding of Kiepiela *et al.* (2004), who demonstrated that HLA-B was associated with control of viral loads. Together with later findings in the same cohort, it was suggested that most of the epitopes associated with viral control were those restricted by beneficial HLAs that targeted or selected mutations in the conserved p24 Gag (Kiepiela *et al.*, 2004; Kiepiela *et al.*, 2007). Other studies have shown that beneficial HLAs associated with control select mutations with a fitness cost to the virus (Martinez-Picado *et al.*, 2006; Rolland *et al.*, 2010; Schneidewind *et al.*, 2008; Schneidewind *et al.*, 2007; Wright *et al.*, 2010; Wright *et al.*, 2012). These mutations are in epitopes like ISW9, TW10 and TL9 restricted by beneficial HLAs like the B*57/B*58:01 and HLA-B*81:01/39:10 (Kiepiela *et al.*, 2004; Leslie *et al.*, 2010), which will be investigated in details in Chapter 3 and 4.

Nonetheless, a previous study in this cohort found that loss of CTL responses was more common in rapid progressors (Mlotshwa *et al.*, 2010) and further work is required to

determine the proportions of viral escape explored in this study that result in complete abrogation of immune recognition. In this thesis, there was a CD4+ T cells benefit of escape particularly in early infection. However, analysis of viral load trajectory over time showed that there was no difference in viral loads between individuals with virus escaping in acute infection and those that did not in the first 12 months post-infection in Gag and/Nef. Some studies have suggested that the inability of CTLs to resolve viremia during chronic infection may be due to the loss of the most potent and effective CTL specificities during the first weeks of infection (O'Connor *et al.,* 2002). The initial benefit to the host due to loss of viral fitness may be lost through the development of compensatory mutations that restore the viral replicative fitness (Brockman *et al.,* 2010; Crawford *et al.*, 2009); these issues will be addressed in more detail in chapter 3 and 4.

In conclusion, correlating sequence evolution in CTL epitopes with markers of disease progression is inherently complicated by the large HLA diversity of the population and while viral evolution is intrinsic to understanding the role of different HLAs, this is only one property that may be affecting the host's ability to control viral replication. This study has documented that the major driving force of sequence diversity in Gag and Nef in acute infection is the HLA, with HLA-B driving acute than early/late sequence evolution in these phases of the HIV-1 infection and more than any other HLAs. Furthermore, the study has shown that in Gag, HLA-B preferentially select mutations in p24. Moreover, participants with acute viral escape have significantly higher CD4+ T-cell counts than in participants without viral escape in the first year of the infection suggesting that viral escape provides some benefit to the host in the first year of infection.

Abstract

HLA-B*81:01 and B*39:10 alleles have been associated with viremic control in HIV-1 subtype C infection. Both alleles restrict the TL9 epitope in p24 Gag and cytotoxic T-lymphocyte (CTL) mediated escape mutations in this epitope have been associated with an *in vitro* fitness cost to the virus. In this study, the timing and impact of mutations in the TL9 epitope on disease progression in five B*81:01- and two B*39:10-positive subtype C infected individuals were investigated.

Whereas both B*39:10 participants sampled at two months post-infection had viruses with mutations in the TL9 epitope, in three of the five (3/5) B*81:01 participants, TL9 escape mutations were only detected 10 months after infection, taking an additional 10 to 15 month to reach fixation. In the two remaining B*81:01 individuals, one carried a TL9 escape variant at 2 weeks post-infection, whereas no escape mutations were detected in the virus from the other participant up to 33 months post-infection, despite CTL targeting of the epitope. In all participants, escape mutations in TL9 were linked to co-evolving residues in the region of Gag known to be associated with host tropism. Late escape in TL9, together with co-evolution of putative compensatory mutations, coincided with spontaneous increase in viral loads in two individuals who were otherwise controlling infection.

These results provide *in vivo* evidence of the detrimental impact of B*81:01 mediated viral evolution, in a single Gag p24 epitope, on control of viraemia.

3.1 Introduction

Genome-wide association studies (GWAS) have shown that the major determinant of HIV disease progression is the human leukocyte antigen or HLA class I (Fellay *et al.*, 2007; Pereyra *et al.*, 2010). However, not all individuals carrying beneficial HLAs progress slowly to AIDs defining diseases (Emu *et al.*, 2008). Understanding the relationship between the virus and the host, and how HLA differentially influences disease progression, remains a central question in HIV research and optimal vaccine immunogen design (McMichael and McCutchan, 2010).

CTL responses targeting p24 Gag have been associated with long-term control of HIV-1 replication (Berger *et al.*, 2011; Borghans *et al.*, 2007; Kiepiela *et al.*, 2007). Indeed, HLA alleles associated with control of HIV-1 replication preferentially present p24 Gag epitopes (Borghans *et al.*, 2007) in acute infection (Wang *et al.*, 2009). Apart from HLA B*57/58:01, there is limited information on the impact of CTL immune escape in individuals controlling infection in African cohorts, which may differ from Caucasian cohorts due to differences in HLA allelic frequency.

HLA B-7 supertype is very common in Black Africans compared to Caucasians (Gonzalez-Galarza *et al.*, 2011; Goulder *et al.*, 2000). The HLA-B7 supertype includes HLA-B*07:02, HLA-B*39:10, HLA-B*42:01, HLA-B*42:02 and HLA-B*81:01 (Leslie *et al.*, 2006), and within this supertype, only HLA-B*81:01 and B*39:10 alleles are consistently associated with control of HIV-1 in subtype C chronic infection (Leslie *et al.*, 2010) although all these alleles present the immunodominant TL9 (Gag 180 – 188) epitope in p24 (Leslie *et al.*, 2006). Moreover, the magnitude of B*81:01 and B*39:10 restricted TL9-specific responses (Leslie *et al.*, 2006), and the corresponding polymorphisms in this epitope are significantly higher than in B*42:01 restriction suggesting a stronger selective pressure by B*81:01 and B*39:10 (Geldmacher *et al.*, 2009; Leslie *et al.*, 2006). This difference in recognition and selection pressure may be due to differences in HLA specificity (Leslie *et al.*, 2006) that may be caused by the minor variations in these HLAs (Geldmacher *et al.*, 2009).

Mutations in residues E177, Q182 and T186, located in or proximal to the TL9 epitope, have been associated with immune escape in the B*81:01, B*42:01 and B*39:10 positive individuals (Leslie *et al.*, 2006). *In vitro* studies have shown that the E177D/A and T186S mutations incur a fitness cost to the virus (Rolland *et al.*, 2010; von Schwedler *et al.*, 2003; Wright *et al.*, 2010), while the Q182S mutation reverts after transmission to HLA-mismatched individuals suggesting it is not beneficial to the virus (Matthews *et al.*, 2008).

In this study, a detailed analysis of individuals with alleles belonging to the HLA-B7 supertype recruited during acute/early infection and followed for three years is presented. The timing and frequency of escape in TL9 epitope are reported, and the impact of escape on disease progression in HLA-B*81:01 and HLA-B*39:10 positive HIV-1 subtype C infected participants is elucidated.

3.2 Materials and Methods

3.2.1 Study subjects

Samples used in this study were from participants of the CAPRISA 002 acute infection cohort (van Loggerenberg *et al.*, 2008), described in details in chapter 2.2.1 of this thesis.

3.2.2 HIV-1 Plasma RNA isolation and Sequencing

As described in chapter 2.2.2, cDNA was synthesized from RNA and viral sequencing was done directly on PCR products or was cloned in pGem-T Easy vector according to the manufacturer's instructions (Promega, USA). Where only p24 *gag* sequences were generated, the sequencing primers were only the Gag B forward and Gag B reverse.

3.2.3 IFN-γ ELISPOT assay

This data was provided by C. Gray, UCT and described in detail in chapter 2.2.5.

3.2.4 HLA typing

HLA data was provided by C. Gray, NICD, which was a high resolution (four digits) typing and performed as described in chapter 2.2.6

3.2.5 Statistical analyses

Statistical analysis and graphical presentations were implemented in GraphPad Prism 5.0 (GraphPad Software, Inc.). To control for fluctuations in viral loads and CD4 counts, the 12 month and 36 month set points of these parameters were calculated by taking the median of measurements at 3 time-points closest to 12 and 36 months, respectively. Statistical analysis of significance was based on Mann-Whitney two-tailed t test. To identify networks of interacting sites in the alignment, approaches coupling phylogenetic and Bayesian network

models in Spidermonkey (Poon *et al.*, 2008) were used (available at www.datamonkey.org). Repeated inference with ancestral states sampled from the posterior distribution were used to evaluate robustness (Delport *et al.,* 2010) in identifying networks of interacting sites in an alignment using Bayesian network techniques, which assumes that co-evolving sites will tend to acquire mutations along the same set of branches (Poon *et al.*, 2008).

3.3 Results

Of the 62 women recruited into the CAPRISA 002 acute infection study, five were HLA-B*81:01-positive, and two were B*39:10-positive (Table 3.1). To evaluate *gag* evolution and the timing of escape in the B*81:01 and B*39:10 restricted TL9 epitope, sequencing was performed on samples collected at the earliest time-point and every three months post infection until 3 years or initiation of antiretroviral (ARV) therapy according to national guidelines (Chapter 2.2.1). Samples from additional time-points were sequenced to further refine timing of escape.

Table 3.1 HLA Type 1 for the five B*81:01 and two B*39:10 individuals in this study (data provided by C. Gray, UCT)

PID[a]	Allele pair		
	HLA - A	HLA - B	HLA - C
CAP0129	26:01, 80:01	18:01, **81:01**	02:02, **04:01**
CAP0222	30:01, 33:03	53:01, **81:01**	**04:01**, **04:01**
CAP0225	01:01, 30:01	42:02, **81:01**	17:01, 18:01
CAP0262	01:01, 66:02	42:01, **81:01**	17:01, 18:01
CAP0277	30:09, 43:01	58:02, **81:01**	**04:01, 04:01**
CAP0278	30:01, 43:01	**39:10**, 42:02	**12:03**, 17:01
CAP0289	30:02, 68:01	**39:10**, 58:02	06:02, **12:03**

[a] **Participant identification number**

3.3.1 The influence of HLA-B*81:01 and B*39:10 on early and chronic disease progression

The trajectory of viral loads and CD4+ T-cell counts over a three-year post-infection period was analyzed in order to understand disease progression in HLA-B*81:01/B*39:10 positive participants (Figure 3.1). Two of the B*81:01 (CAP129 and CAP225) were typical progressors with viral load set-points ranging from 80,100 to 21,500 copies/ml. Furthermore, one B*39:10-positive participant (CAP289) maintained high viral loads from the first year of infection (>500,000 copies/ml). There were two controllers with viral loads below 2,000 copies/ml, one B*39:10-positive participant (CAP278) who controlled viral loads from the start of infection, and one B*81:01 positive participant (CAP262) who only controlled her viral load after 18 months of infection. Conversely, two B*81:01-positive participants who controlled viral replication below 2,000 copies/ml for the first 18-24 months of infection had viral loads that increased by one log over a 6 month period and thereafter (CAP222 and CAP277). CD4+ T cell counts were relatively stable except in three participants who had declining CD4+ T cell counts: the one B*39:10-positive participant (CAP289) who had high viral loads; and the two B*81:10-positive participants (CAP222 and CAP277) who had significant increases in viral loads.

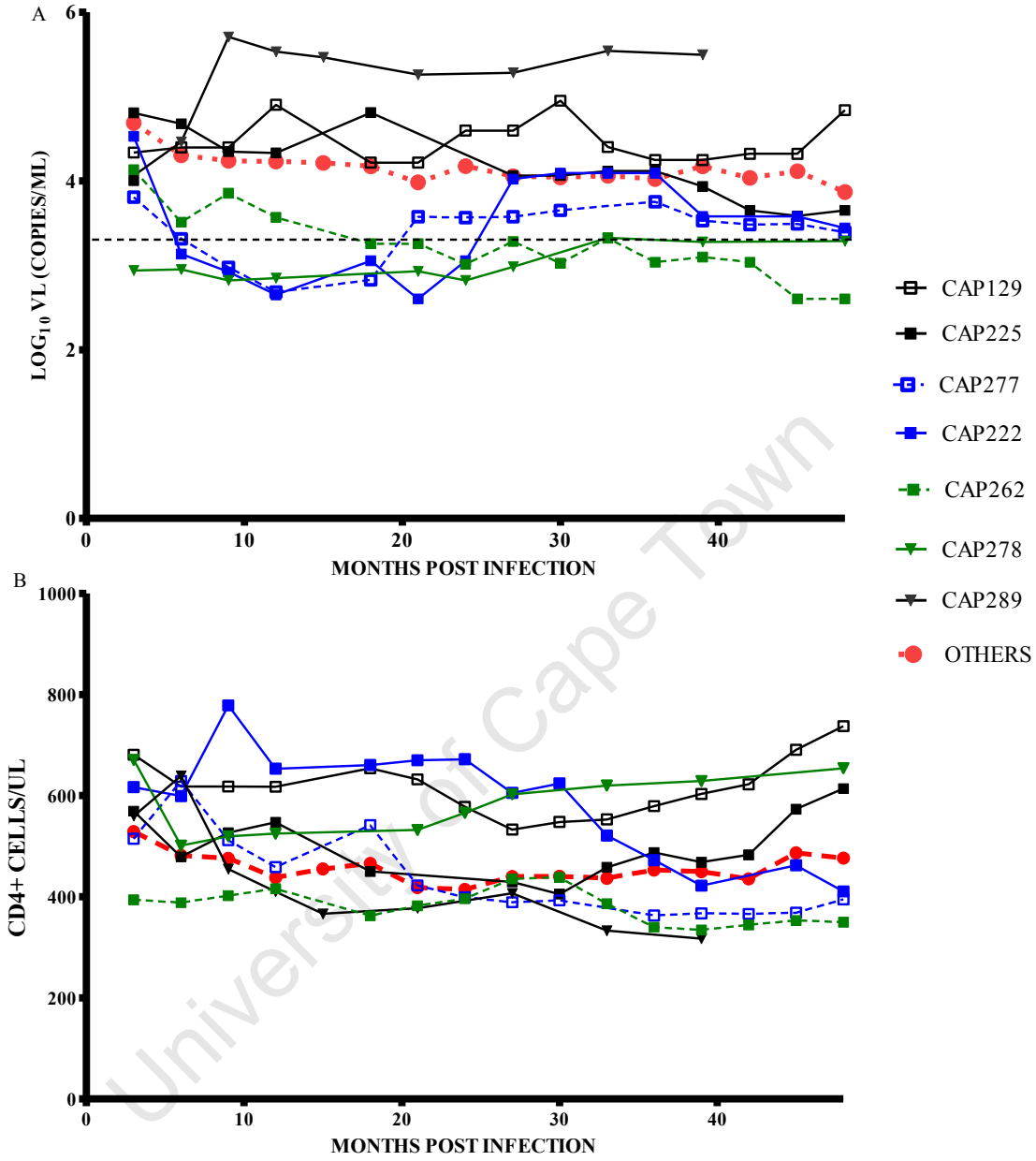


Figure 3.1 Figure 1: (A) Viral load and (B) CD4+ T cell count trajectories for the first 48 months of infection for the five HLA-B*81:01-positive participants (CAP129, CAP222, CAP225, CAP262 and CAP277; represented by squares) and two B*39:10-positive participants (CAP278 and CAP289; represented by triangles). CAP289 had not reached 48 months postinfection by the time of this study. The trajectory of median viral loads and median CD4+ T-cell count of all other CAPRISA participants (OTHERS) excluding the B*39:10/81:01 and B*57/58:01 are shown as dotted red lines. The broken black line is viral load level 2,000 copies/mL, defining controllers. Viral loads and CD4+ T-cell counts were smoothed at each time point (medians of measurements taken at three time-points closest to intervals 3, 6, 9, 12, 15, etc up to 48 months post-infection).

The influence of HLA-B*81:01 and B*39:10 on viral load set-point and CD4+ T-cell counts at one and three years post infection was then investigated. B*57/58:01 participants were

analyzed separately as these "protective" alleles are often associated with better control of HIV-1 replication (Migueles *et al.*, 2000; Miura *et al.*, 2009).

There was a trend for lower viral load in B*81:01-positive participants when compared to those with non-protective alleles (i.e. B*81:01/57/58:01/39:10 negative participants) at set-point (12 months post-infection) although this was not significant (p = 0.2) (Figure 3.2A; median 3.57 $\log_{10}$ copies/ml [IQR = 2.7 – 4.6] versus 4.46 $\log_{10}$ copies/ml [IQR = 3.49 – 4.8]). There were however, significantly higher CD4+ T-cell counts for the B*81:01 positive participants compared to individuals with non-protective alleles at one year post-infection (p=0.049) (Figure 3.2B; median 597 cell/ul [IQR = 467 – 644] versus 369 cell/ul [IQR = 305 – 525]). To determine whether these HLA alleles affected disease progression in chronic infection, both viral loads and CD4+ T-cell counts at 3 years post-infection were analyzed. By three years post-infection, 12 participants had initiated ARV therapy and to not bias the analysis towards controllers, the last available viral load and CD4 count data on these subjects prior to commencing ARV therapy was used. There was no significant difference in either viral load or CD4+ T-cell counts at 36 month post-infection in B*81:01 positive participants compared to those with non-protective alleles suggesting that this potential benefit detected at 12 months was not sustained in chronic infection.

Whether B*81:01 positive participants controlled infection better than B*57/58:01 positive participants was also investigated. Although the CD4+ T-cell counts tended to be higher and viral loads lower in B*81:01 positive, when compared to B*57/58:01-positive participants at 12 months, this difference was not significant, and this trend was not sustained at 36 months (Figure 3.2 C and D).

Of the two B*39:10-positive participants, one was a controller (CAP278), whereas the other was a typical disease progressor (CAP289) (Figure 3.1). The controller had a viral load at 12 months of 2.95 $\log_{10}$ copies/ml compared to 5.72 $\log_{10}$ copies/ml for the progressor, while their CD4+ T-cell counts were 568 and 410 cell/ul respectively (Figure 3.2).

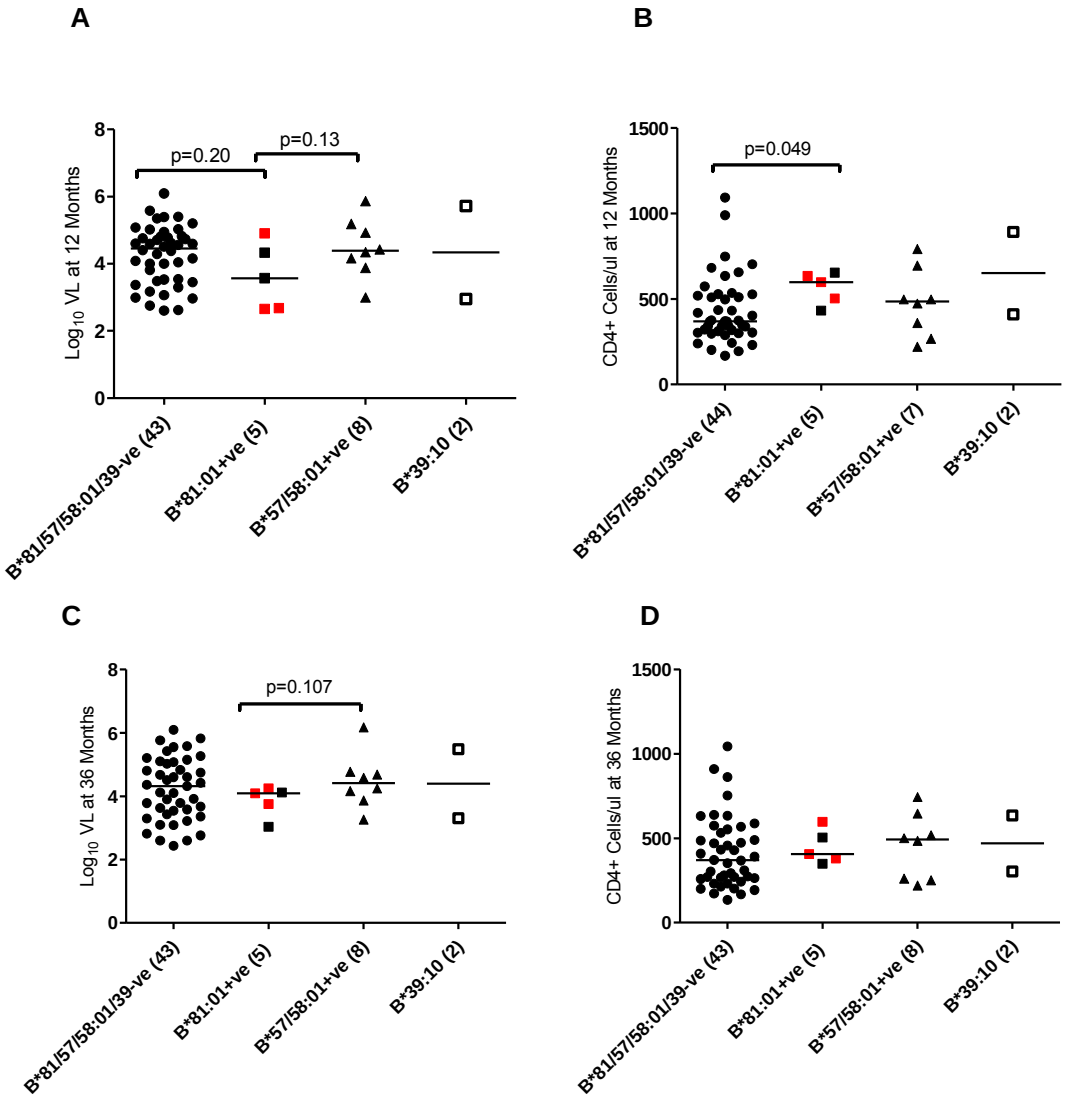


Figure 3.2 Association between viral load (VL) and CD4+ T cell counts prognostic markers and HLA-B*81:01 and B*39:10 alleles at 12 and 36 months post-infection. (A and C) Log_{10} VL compared between individuals without and with HLA-B*81:01, B*39:10 or B*57:03/58:01 alleles at 12 and 36 months post infection; (B and D) CD4+ T cell counts compared between individuals without and with HLA-B*81:01, B*39:10 or B*57/58:01 alleles at 12 and 36 months post infection respectively. In red are B*81:01 participants positive for the Cw*04:01 allele. Both B*39:10-positive participants are also positive for the Cw*12:03 allele. Viral load and CD4+ T cell counts are calculated as medians of measurements taken at three time-points closest to these intervals. To not enrich for slow progressors, we included the last attained median values for viral load and CD4+ T cell counts before initiation of ARV therapy for individuals who had received treatment prior to 12 (n=1) or 36 months (n=12) post infection.

3.3.2 Linkage to HLA-Cw alleles did not impact disease progression

In subtype C infection, a good clinical outcome in individuals positive for B*81:01 and B*39:10 alleles has been linked to an additive impact of Cw*04:01 and Cw*12:03 alleles, respectively (Leslie *et al.*, 2010). Of the five B*81:01-positive participants, three participants were positive for the Cw*04:01 allele of which two were controllers (CAP222 & CAP277) and the third a typical progressor (CAP129). However, of the two B*81:01-positive participants who did not carry the Cw*04:01 allele, one was a typical progressor (CAP225), while the other was a controller (CAP262). Although both of the B*39:10-positive participants were also positive for the Cw*12:03 allele (Table 3.1), one was a progressor (CAP289) and the other a controller (CAP278) (Figure 3.1). These results suggest that the differential disease progression in B*81:01- or B*39:10-positive participants was not simply due to the HLA-Cw allelic combinations.

3.3.3 Kinetics of p24-Gag-TL9 escape in B*39:10 and B*81:01 participants

Substitutions in positions E177, Q182 and T186 Gag residues that are located within or proximal to the TL9 epitope (Gag180 -188) result in loss of immune recognition in B*81:01/B*39:10-positive individuals, with the predominant polymorphisms associated with HLA mediated TL9 restriction *in vivo* being the E177D, Q182S and T186S (Frater *et al.*, 2007; Leslie *et al.*, 2006; Matthews *et al.*, 2008; Rolland *et al.*, 2010).

In B*81:01-positive participants, mutations in TL9 generally emerged later in infection (Table 3.2). The first mutation to emerge was at position 182. In three participants who selected this mutation post-infection, the Q182X (where X=S/T/A/H) was detected between 70 and 83 weeks post-infection in CAP222; 50 and 61 weeks in CAP225; and 48 and 50 weeks in CAP277. In two of the three participants (CAP222 and 277), additional escape mutations were detected within 6 to 12 months following the emergence of the escape mutation at position 182, and reached fixation only 10 to 15 months later (Table 3.2). The T186S mutation emerged either alone as in CAP277, or concurrently with E177D as in CAP222. In the fourth participant, CAP129, sampled prior to a detectable HIV antibody response, the

Q182T mutation was already present. The mutation remained stable for up to 28 weeks; however, by 32 weeks the Q182S mutation emerged. As the Q182X mutation in the TL9 epitope was detected in the earliest sample sequenced from three of the 37 CAPRISA participants who were negative for HLA-B7 supertype alleles, it is possible that this mutation was present in the transmitted virus. This is supported by the fact that there was no evidence of an immune response to this epitope in acute infection (Table 3.3). There was no sequence evolution in the TL9 epitope in the fifth B*81:01 participant (CAP262).

The viruses from the two B*39:10-positive participants had a different pattern of escape. In one B*39:10 participant (CAP278), all three mutations (E177D, Q182S and T186S) were already present in the earliest sequences available at 9 weeks post-infection, with no further evolution within the epitope during 22 months of follow-up. In the second B*39:10 participant (CAP289), the Q182A mutation was present at 8 weeks post-infection and evolved to Q182G at 28 weeks post-infection.

To evaluate the likelihood of the HLA B*39:10-associated mutations in residues E177, Q182 and T186 being present on the transmitted virus in CAP278, we estimated the frequency of TL9 mutations in participants negative for HLA-B7 supertype and Cw*08:02 alleles that recognise this epitope (Llano *et al.*, 2009) in the CAPRISA 002 cohort. We found no sequences from either acute or chronic infection with either the two (Q182S with T186S) or the three (E177D, Q182S and T186S) concurrent mutations. Furthermore, the triple mutation was not present in any of the sequences analyzed from 507 HIV-1 subtype C infected adults in Durban, KwaZulu-Natal, South Africa (Kiepiela *et al.*, 2004; Rousseau *et al.*, 2008).

Together, these results show that the TL9 escape in B*81:01 generally occurs later and can take approximately one year from the first mutations for escape in this epitope to reach fixation. Furthermore, we show that the three concurrent E177D, Q182S and T186S site mutations are rare and might have arisen due to early escape in one of the two B*39:10-positive participants although it cannot be ruled out completely that this was a transmitted variant from a B7 supertype donor.

Table 3.2: HLA-B*81:01, B*39:10 and other B7 allelic restricted sequence variations in and proximal to the TL9 epitope, and putative compensatory mutations. The commonly evolving amino acid residues are in italics in the peptide 176-S*e*GATP*q*DLN*t*ML-188**.**

PID[a]	W. PI	TL9 Mutations							Others			
		E177	T180	Q182	T186	L188	T190	H219	I223	A248	M250	I256
CAP129	2	-	-	T	-	-	-	-	H	T	-	-
B*81:01	23	-	-	T	-	-	-	-	H	T	-	-
	32	-	-	S	-	-	-	-	H	T	-	-
	37	-	-	S	-	-	-	-	H	T	-	-
	42	-	-	S	-	-	-	-	H	T	-	-
	54	-	-	S	-	-	-	-	H	T	-	-
	106	-	-	S	-	-	-	-	H	T	-	-
CAP222	56	-	-	-	-	-	-	-	A	-	-	-
B*81:01	70	-	-	-	-	-	-	-	A	-	-	-
	83	-	-	S (6/9)	-	-	-	-	A	-	-	V
	95	-	-	S (1/6)	-	-	-	-	A	-	-	V
	108	-	-	S (3/11)	-	-	-	-	A	-	-	V (3/11)
	108	[D	A	S	S] 1/11	-	-	-	A	-	-	-
	108	[D	-	S	S] 7/11	-	-	-	V (2/7)	-	-	-
	122	[D	-	S	S]13/13	-	-	-	V 13/13	-	-	-
	133	[D	-	S	S	-	-	-	V] 13/13	-	-	-
	148	D	-	S	S	-	-	-	V	-	-	-
	161	D	-	S	S	-	-	-	V	-	-	-
	174	[D	-	S	S	-	-	-	V] 13/13	-	-	-
	190	D	-	S	S	-	-	-	V	-	-	-
CAP225	50	-	-	-	-	-	-	-	-	-	-	-
B*81:01	61	-	-	S (10/15)	-	-	-	-	-	-	-	V (10/15)
	74	-	-	S	-	-	-	-	-	-	-	V
	85	-	-	S (3/15)	-	-	-	-	-	-	-	V (3/15)
	98	-	-	S	-	-	-	-	-	-	-	V
	117	-	-	S	-	-	-	-	-	-	-	V
	130	-	-	S	-	-	-	-	-	-	-	V
	145	-	-	S	-	-	-	-	-	-	-	V
	158	-	-	S	-	-	-	-	-	-	-	V
	171	-	-	S	-	-	-	-	-	-	-	V
	183	-	-	S	-	-	-	-	-	-	-	V
CAP277	16	-	-	-	-	-	-	-	-	-	I	-
B*81:01	41	-	-	-	-	-	-	-	-	-	I	-
	46	-	-	-	-	-	-	-	-	-	I	-
	50	-	-	H	-	-	-	-	-	-	I	-
	54	-	-	T	-	-	-	-	-	-	I	-
	58	-	-	T	-	-	-	-	-	-	I	-
	71	-	-	A	-	-	-	-	-	-	I	-
	84	-	-	S	-	-	-	-	-	-	I	-
	97	-	-	S (3/13)	-	-	-	-	-	-	I (3/13)	-
	97	-	-	S	S (6/13)	-	-	-	-	-	- (6/13)	-
	97	-	-	-	-	-	-	-	-	-	- (4/13)	-
	110	-	-	S	S	-	-	-	-	T	-	-
	123	-	-	S	S	-	-	-	-	T	-	-
	137	-	-	S	S	-	-	-	-	T	-	-
CAP262	70	-	-	-	-	-	-	-	-	-	-	-
B*81:01	84	-	-	-	-	-	-	-	-	-	-	-
B*42:01	95	-	-	-	-	-	-	-	-	-	-	-
	109	-	-	-	-	-	-	-	-	-	-	-
	122	-	-	-	-	-	-	-	-	-	-	-
	136	-	-	-	-	-	-	-	-	-	-	-
CAP278	9	D	-	S	S	-	-	-	V	-	-	-
B*39:10	54	D	-	S	S	-	-	-	V	-	-	-
	85	D	-	S	S	-	-	Q	V	-	-	-
CAP289	8	-	-	A	-	-	-	-	V	-	-	V
B*39:10	24	-	-	A	-	-	-	-	V	-	-	V
	28	-	-	G	-	-	-	-	V	-	-	V
	41	-	-	G	-	-	-	-	V	-	-	V
CAP061	14	-	-	-	-	-	-	Q	V	-	-	-
B*42:01	33	-	-	-	-	-	-	Q	V	-	-	-
C*08:02	51	-	-	-	-	-	-	Q	V	-	-	-
	76	-	-	-	-	F	-	Q	V	-	I	-
	127	-	-	-	-	-	-	-	V	-	-	-
CAP282	11	-	-	-	-	-	-	-	-	-	-	V
B*42:01	35	-	-	-	-	-	-	-	-	-	-	V
	53	-	-	-	-	-	-	-	-	-	-	V
	60	ND	ND	ND	ND	-	ND	ND	ND	ND	ND	ND

- : No residue substitution or else the incoming residue is indicated; ND: Not done/not available; W.PI: Weeks Post-Infection

Table 3.3: Temporal Pattern of TL9 specific IFN-γ ELIspot responses in B7 positive participants. HLAs known to restrict TL9 are in bold/underlined.

PID[a]	HLA	Weeks (PI)	SFU/10^6 cells	Timing of last wild-type (WT) and first TL9 mutation
CAP129	A*26:01,80:01; B*18:01, **81:01;** Cw*02:02, 04:01	2	Neg	Q182T
		4	Neg	
		6	Neg	
		13	Neg	
		23	Neg	Q182T
		32	ND	Q182S
CAP222	A*30:01,33:03;B**53:01, 81:01;** Cw*04:01, 04:01	6	245	
		8	103	
		9	1568	
		11	368*	
		15	168	
		17	243	
		70	ND	WT
		83	ND	WT/Q182S (mixed pop.)
		108	1233	Q182S/and E177D with T186S (mixed pop)
		122	ND	E177D, Q182S, T186S (fixed)
CAP225	A*01:01,30:01;B**81:01, 42:02;** Cw*17:01,18:01	3	9078	
		4	888	
		5	475	
		7	1288	
		9	1058	
		14	203	
		18	3328	
		21	1513	
		24	1463	
		29	2080	
		33	818	
		37	468	
		50	1710	WT
		61	ND	WT/Q182S (mixed)
		85	Neg	WT/Q182S (mixed)
		145	1221	Q182S (fixed)
CAP262	A*01:01,66:02;B**42:01, 81:01;** Cw*17:01, 18:01	7	365	
		14	Neg	
		16	215	
		18	250	
		26	105	
		136	ND	WT
CAP277	A*30:09,43:01; B*58:02, **81:01;** Cw*04:01, 04:01	37	6147	
		46	ND	WT
		50	ND	Q182H
		97	193	WT / Q182S / T186S (mixed pop)
CAP278	A*30:01,43:01**;B*39:10, 42:02;** Cw*12:03, 17:01	8	117	ND
		9	ND	E177D, Q182S & T186S
		10	Neg	
CAP289	A*30:02,68:01**;B*39:10,** 58:02; Cw*06:02; 12:03	8	ND	Q182A
		9	Neg	
		28	ND	Q182G
CAP037	A*23:01,24:02;**B*07:02, 53:01;** Cw*17:01, 17:01	122	273	WT
CAP040	A*30:01,30:02; B*15:10, **42:01;** Cw*03:04, 17:01	106	293	WT
CAP220	A*30:04,74:01;**B*42:01; 42:01;**Cw*17:01, 17:01	68	373	WT
CAP239	A*01:23,29:02;**B*42:01,** 58:01; Cw*06:02, 17:01	67	258	WT
CAP269	A*02:05,68:02;**B*07:02**, 58:02; Cw*06, 07;	120	Neg	WT
CAP282	A*30:01,66:01;**B*42:01**, 58:02; Cw*: ND	23	129	
		35	ND	WT
		53	ND	Q182T first observed at 12month post-infection
		60	Neg	

PI: Post-Infection; ND: Not done; Neg: Negative; SFU: spot forming units; mixed populations: more than one viral variant identified; fixed: only one viral variant identified

3.3.4 HLA-B*81:01 and B*39:10 exert more selection pressure on TL9 than B*42 and B*07:02

Previous studies have shown that the selection strength of immune responses targeting the TL9 epitope differ according to the HLAs belonging to HLA-B7 supertype with HLA-B*81:01 having a stronger selective capacity than B*4201 (Geldmacher *et al.,* 2009; Leslie *et al.*, 2006). Since three participants in this study were also positive for the HLA-B42 allele, we wished to know the frequency of HLA mediated sequence evolution in the TL9 epitope in participants carrying HLA-B7 supertype (n=17), excluding the B*81:01/39:10 positive participants. Analysis of sequences at 18 months post-infection revealed that only two, of the fourteen B*42:01 participants (including CAP262 who was positive for both B*42:01/B*81:01), developed polymorphisms in the epitope, while no mutations were detected in the four B*07:02-positive participants (Table 3.4). In one of the two B*42:01 participants who selected mutations in the TL9 epitope, an escape mutation L188F emerged between 51 and 76 weeks (CAP061), however, this mutation was not fixed as it was not detected at 102 weeks post-infection (Table 3.2). In the second B*42:01 participant in this study, the Q182T mutation was selected between 48 and 53 weeks post infection (CAP282), while four of the five B*81:01 positive individuals selected for 182S. A recent study employing 2126 chronically HIV-1 subtype C-infected, antiretroviral naïve adults showed that unlike the B*81:01 allele that select against the Q182T mutation, the B*42:01 allele select for Q182T (Carlson *et al.*, 2012). Thus, the difference in escape at residue 182 provides further support that B*81:01 allelic restriction is responsible for the changes observed in the TL9 epitopes.

These results also support the findings, which demonstrated that B*81:01 allele select for mutations in the TL9 epitope over the HLA-B*42:01 allele (Geldmacher *et al.*, 2009; Leslie *et al.*, 2006). Overall, these findings emphasise the differential role of closely related HLA in the timing of epitope escape and sequence evolution.

Table 3.4: Differences in the rate of escape in participants positive for the HLA- B7 supertype and Cw*08:02 alleles that target TL9 epitope.

Timing[a] Months PI	B*39:10 (n=2)	B42 (n=14)[c]	B*07:02 (n=4)	B*81:01 (n=5)[c]	Cw*0802 (n=1)
Enrolment	1 & 1[b]	0	0	0	0
<3 months	0	0	0	0	0
3-6	0	0	0	0	0
6-12	1[b]	1*	0	1 & 1[b]	0
12-18	0	1**	0	2	0

[a] The timing corresponds to the first mutation observed in TL9 epitope in the CAPRISA (AI02) cohort; [b] Participants with mutation at enrollment but with later further evolution of the mutations; [c] Number included one participant (CAP262) who never evolved TL9 mutation in the follow-up period and positive for both HLA-B*42:01 and B*81:01; *: Mutation Q182T; * *: Mutation L188F was transient; PI : Post-Infection

3.3.5 Late sequence evolution in TL9 is not due to lack of CTL responses in acute Infection

A lack of CTL pressure in acute infection has been suggested to partly explain why some epitopes remain invariant through acute infection and only evolve late in infection (Goulder and Watkins, 2004). Of the five B*81:01-positive participants, four (CAP222, CAP225, CAP262 and CAP277) targeted the TL9 epitope with responses detected in acute infection for three participants (CAP222, CAP225 and CAP262) and at 9 months for one participant (CAP277) (earliest sample tested) (Table 3.3). Only one participant (CAP129) who harboured an escape mutant at two weeks post infection showed no response to the TL9 epitope in acute infection (<3 months post-infection). In two of the three B*81:01-positive participants who recognized the TL9 epitope, escape only occurred 10 to 12 month after detectable responses. The lack of TL9 variation in CAP262 was unlikely due to lack of immune targeting since responses to this epitope occurred in acute infection, albeit at a low magnitude. Moreover, there were no TL9 associated mutations in this participant during the 33 months of follow-up (Table 3.2). Only one of the B*39:10-positive participants had detectable responses to TL9 during acute infection (CAP278). The peptide used to detect responses corresponded to the wild-type sequence, and it is possible that responses in individuals harboring TL9 escape viruses in very early infection could be missed due to differences in the viral mutant and the peptide used.

Other than the B*81:01/39:10-positive participants, immune responses at later time points (>12 months postinfection) were generated for six of the 19 participants positive for other HLA-B7 supertype and Cw*08:02 alleles known to target the TL9 epitope. Low-magnitude responses (129 to 273 SFU/106 cells) were detected in five of the six participants tested; however, only one of these participants (CAP282) developed an escape mutant (Q182T).

3.3.6 Sequence evolution in the TL9 epitope is associated with co-varying sites outside of the epitope

Studies have found significant associations between mutations in TL9 epitope and highly variable residues (Gag residues 138, 228, 248, 252, and 256) located in or flanking the three tropism-determining loops of p24 involved in host restriction factor recognition (loops 1 to 3 located at Gag residues 137-147; 214-225; and 248-254 respectively) (Carlson *et al.*, 2008; Crawford *et al.*, 2011; Hatziioannou *et al.*, 2004; Rolland *et al.*, 2010). Whether mutations in these tropism determining loops that were co-selected with TL9 associated mutations could be detected was investigated. Variations in five residues were identified at positions 219, 223, 248, 250 and 256, with the residues 219 and 223 located in the second loop, which contains the cyclophilin A binding sites, and residues 248, 250 and 256 located in or proximal to the third loop (Table 3.2).

Using 463 longitudinal population sequences including over 120 clonal sequences from four B*81:01 participants, co-selection was then investigated using approaches that couple phylogenetic and Bayesian network models in Spidermonkey (Poon *et al.,* 2008; www.data-monkey.org). Significant correlations were found between sites, 186S with 177D, 182S with 256V, and 186S with 250M (posterior probability>0.5) (Table 3.5). In the longitudinal analysis in CAP277, the emergence of T186S mutation was accompanied with evolution of sites, 248 and 250 (Table 3.2). Interestingly, every sequence carrying the TL9 double mutations (Q182S and T186S) was associated with reversion of the M250I mutation, suggesting a selective reversion to accommodate the double TL9 mutations. Similarly, the fixation of the triple mutations E177D, Q182S and T186S at around 122 weeks in CAP222 was preceded by significant toggling in both the epitope, as well as reversion in a putative

compensatory site V223A (Table 3.2) (p=0.0035), suggesting that evolution of mutation V223A was under selective reversion (Delport *et al.*, 2008). Furthermore, in CAP222, as well as in CAP225, evolution of the single Q182S mutation in the TL9 epitope was accompanied with emergence of Gag I256V mutation (Table 3.2). In three other participants, CAP129, CAP278 and CAP289, a combination of mutations Q182T/A248T, E177D/Q182S/T186S/I223V and Q182A/I223V/I256V, respectively, were present in the first sample analyzed.

Sequence evolution at site 190 has been shown to compensate for escape in TL9 epitope through the T186S mutation (Wright *et al.*, 2010; Wright *et al.*, 2012). There was no sequence evolution or transmitted mutations in either any of the B7 positive individuals or any other participant investigated in this study at this residue (T190) in this cohort (Table 3.2).

Table 3.5 Sites co-varying with each other in 3 of the 4 individuals who evolved sequences in the B*81:01 restricted TL9 epitope

PID	HXB2 Position Site 1	HXB2 Position Site 2	Posterior probability
CAP222	186S	177D	0.73
CAP222	182S	256V	0.73
CAP225	182S	256V	0.73
CAP277	186S	250M	0.79

3.3.7 Timing of sequence evolution influenced the differential disease progression among the B*81:01 and B*39:10 positive participants

In vitro studies have demonstrated that substitution in the TL9 epitope residue T186 alone (Wright *et al.*, 2010; Wright *et al.*, 2012) or in both T186 and E177 (Rolland *et al.*, 2010) incurs a fitness cost to the virus. Mutation at residue Q182 reverts to consensus following transmission to HLA mismatched recipients (Matthews *et al.*, 2008) and a recent study of Wright *et al.* (2012) demonstrated in a subtype C Gag backbone that Q182S incurs a fitness cost to the virus *in vitro*.

In order to understand CTL mediated escape and its impact on disease progression *in vivo*, the temporal relationship between sequence evolution in TL9 epitope and, viral loads and CD4+ T-cell counts were investigated in the B*81:01 and B*39:10 participants (Figure 3.3A). No escape was detected in virus from CAP262, who elicited immune responses to TL9 in early infection and gradually controlled viral load to levels below 2000 copies/ml after 24 months post-infection. Evolution of a single site, Q182 in the TL9 epitope in tandem with putative compensatory mutations A248T or I256V, did not affect viral loads or CD4+ T-cell counts in the three typical progressors, CAP129, CAP225 and CAP289 at least for the first 24 months following escape. These participants maintained consistently high viral loads for at least the first three years post-infection. Interestingly, characteristic of complete escape from a dominant response (Crawford *et al.*, 2009; Feeney *et al.*, 2004; Goulder *et al.*, 1997), the spontaneous increase in viral load and reduction in CD4+ T-cell counts in two individuals controlling viraemia, CAP222 and CAP277, coincided with late sequence evolution in three (E177D, Q182S, T186S) and two (Q182S with T186S) sites respectively, together with reversion of putative compensatory mutations I/V223A and M250I, respectively. However, when viral loads in these participants, were compared to those who evolved single site escape mutation, and to the rest of the cohort (Figure 3.1), it was found that CAP277 maintained lower viral loads, while CAP222 had fluctuating viral loads that were similar to the median of the cohort and generally lower than individuals infected with viruses with single site escape mutations. Viral loads were measured quarterly and the cost of escape in these two participants was estimated by comparing the median viral load measurements for 9 months before escape and 9 months later. Multiple residue escape was associated with a one $\log_{10}$ increase in viral load (Figure 3.3B). This higher viral load was sustained for approximately 9 months post escape, after which it subsided, although to a level higher than before escape (Figure 3.3B). Interestingly, the B*39:10-positive participant (CAP278), who had the same mutations in acute infection maintained low viral loads below 2,000 copies/ml over the three years period.

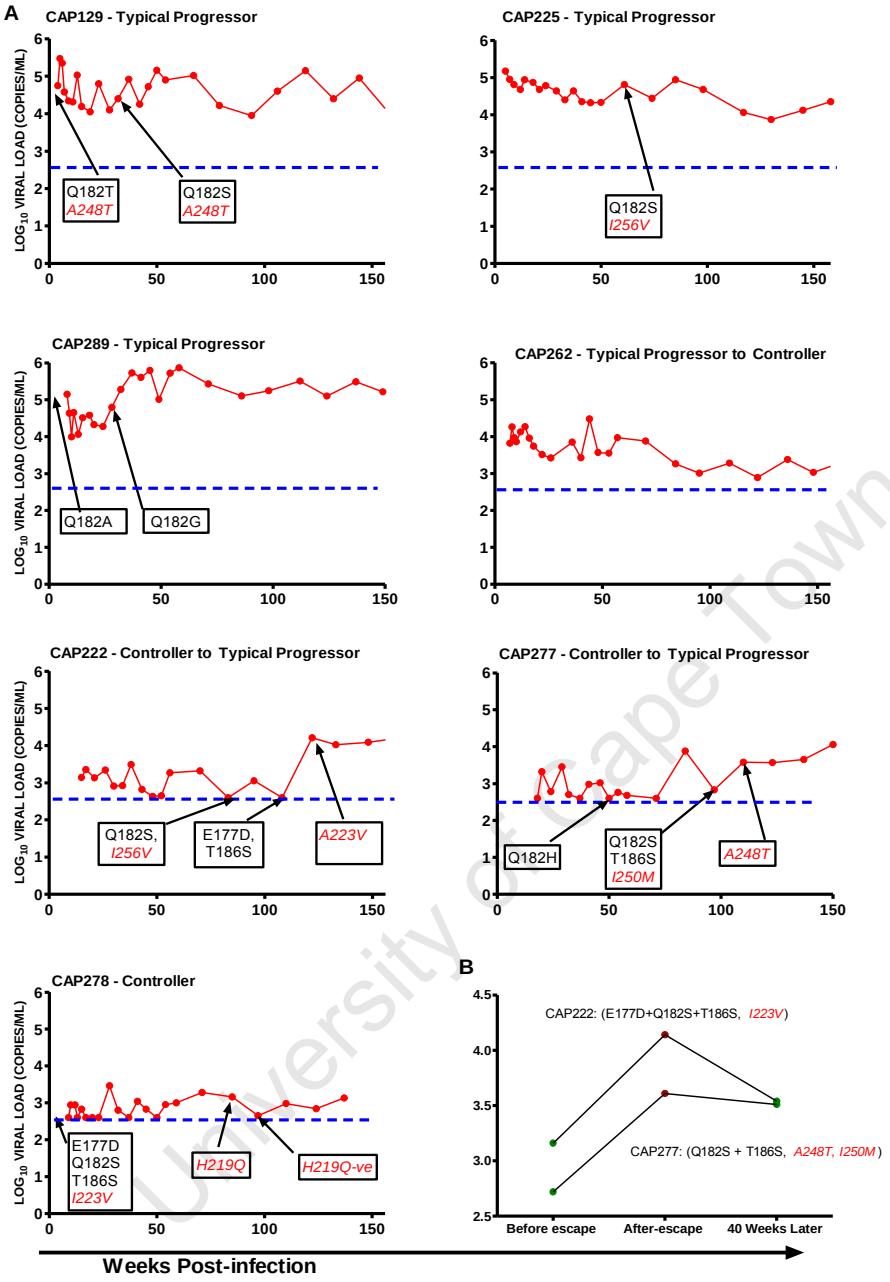


Figure 3.3 (A) Viral load trajectories for the first 156 weeks (36 months) in B*81:01- (CAP129, CAP225, CAP222, CAP262 and CAP277) and B*39:10- (CAP278 and CAP289) positive participants, illustrating the timing of TL9 associated escape. Broken blue line illustrates the limit of detection of the assay (400copies/mL or 2.6 $\log_{10}$). Putative compensatory mutations are shown in red italics. (B) Cost of multiple sequence evolution in or flanking the TL9 epitope in individuals who were controlling the infection as measured by the change in viral load and longer follow up showing reduction in maximum viral load but to levels higher than before escape. Four values of $\log_{10}$ viral loads over 9 months were used to calculate the median viral load before multiple residue-escapes, after and 40 weeks later. Putative compensatory mutations are shown in red italics.

3.3.8 Discussion

The B-7 HLA restriction alleles, including HLA-B*07:02, HLA-B*39:10, HLA-B*42:01, HLA-B*42:02 and HLA-B*81:01, are common in Black Africans and were found at a frequency of 40% (25/62) in the CAPRISA cohort. The TL9 epitope, located in p24 Gag, is an immunodominant epitope restricted by this HLA supertype. These HLAs are less common in Caucasians and the differential CTL targeting of Gag in Caucasians compared to Black South Africans is largely due to differences in frequency of TL9 recognition (Goulder *et al.*, 2000). The significance of evaluating this epitope is further emphasized by the accumulating evidence showing that HLA-B*39:10 and -B*81:01 alleles are associated with better infection outcome in subtype C infection in southern Africa (Kiepiela *et al.*, 2004; Leslie *et al.*, 2010). Aside from well characterised epitopes associated with viral control such as the B*57/58:01 restricted TW10, there is little known about the influence of epitope escape on HIV infection of Africans. Here we report on a longitudinal study of subtype C infected women from Kwazulu-Natal South Africa, from acute to chronic infection. This study provided opportunity to determine the timing of escape in TL9, identify co-evolving sites, and provided insights into the impact of these mutations on disease progression.

Previous studies have shown that individuals carrying any of the closely related HLA‑B7 supertype alleles can mount a CD8+ T cell response that targets the TL9 epitope (Leslie *et al.*, 2006), with the most dominant being HLA-B*81:01, followed by B*39:10, and then B*42:01 (Leslie *et al.*, 2006). While the B42 alleles may have contributed to the selection of mutations in TL9 in each of the three participants positive for B*81:01 or B*39:10 in combination with B42 allele (Table 3.1), there were far fewer polymorphisms in viruses in individuals positive for only B*42:01/42:02 alleles, at least for the first 18 months post-infection confirming previous observations that B*81:01 and B*39:10 exert the greatest selective pressure on the virus (Leslie *et al.*, 2006) (Table 3.4).

The emergence of known B*81:01 escape mutations (E177D, Q182S and T186S) (Leslie *et al.*, 2006) is documented and a common pattern of evolution with the Q182S mutation generally emerging prior to T186S is identified. Unlike the B*57/58:01-restricted TW10

escape, which usually occur rapidly and during acute and early infection (Chopera *et al.*, 2008; Crawford *et al.*, 2009; Martinez-Picado *et al.*, 2006), escape in women with a B*81:01 background was usually late and took nearly a year to reach fixation. In late infection, slow escape has been postulated to be due to a number of factors including low selection pressure, increased diversity of specific responses, and raised viral fitness cost (Ganusov *et al.*, 2011; Loh *et al.*, 2008; Schneidewind *et al.*, 2008). Indeed, i*n vitro* studies have demonstrated that T186S mutation alone, or in combination with E177D, incurs a fitness cost to the virus (Rolland *et al.*, 2010; Wright *et al.*, 2010; Wright *et al.*, 2012). Interestingly, late evolution of T186S either alone, or in combination with E177D, coincided with an increase in viral load and although viral replication was subsequently brought under control, it returned to a level higher than before escape. This reflects the complex balance between the host and the virus, where immune escape results in loss of control and is thus detrimental to the host; however, the impact of the mutation on viral replication capacity provides some benefit.

While escape mutations that reduce viral fitness often select for variants that affect structural and/or reduced function (Schneidewind *et al.*, 2008; Schneidewind *et al.*, 2007), the development of mutations that compensate for the fitness cost in the TL9 epitope are not well described. The analysis of Wright *et al.* (2010) found that individuals carrying both Q182S and T186S had higher viral loads than individuals carrying single T186S. However, although assays for their recent *in vitro* studies for the Q182S mutation alone yielded discordant replication results, *in vitro* experiments in a subtype C Gag backbone found that Q182S incurs a fitness cost to the virus and same level of replication for either T186S alone or both T186S and Q182S (Wright *et al.*, 2012). A recent cross-sectional study on sequences from 662 individuals infected with HIV-1 subtype C found that polymorphism at I256 was significantly associated with the B*81:01 restricted Q182 and T186 sites (Crawford *et al.*, 2011). Here we find concomitant evolution of I256V in each sequence carrying the Q182X mutation providing further support that this mutation compensates for the fitness cost due to sequence evolution at site 182. However, the replication cost for this site mutation with the co-evolving site remains to be determined *in vitro*.

In vitro studies have demonstrated that mutations in residues H219, I223, M228 compensate for the fitness cost due to the T242N escape mutation in B*57/58:01 restricted TW10, while A248T and M250I have been associated with TW10 escape (Brockman *et al.*, 2007; Crawford *et al.*, 2011; Martinez-Picado *et al.*, 2006). This study has found an overlap in sites associated with B*81:01/39:10 restricted TL9. The development of both Q182S and T186S mutations in TL9 was associated with reversion to M250. Similarly, the fixation of the three escape residues at two years post-infection in CAP222 was associated with selective reversion of the transmitted mutation I/V223A in the cyclophilin binding loop, while the H219Q mutation in Gag subsequently developed 17 months post-infection in one of the controllers (CAP278). These common evolutionary pathways associated with TW10/TL9 restriction may be due to functional and structural interactions between the region harbouring these epitopes (Martinez-Picado *et al.*, 2006; Tang *et al.*, 2002). These results support the model proposed by Crawford *et al.* (2011), in which exposed capsid variable-residues generically compensate for fitness cost due to selection in structurally and functionally conserved capsid residues. The findings in this study also support common compensatory pathways of viral evolution in variable domains associated with host tropism.

In conclusion, this study expands our understanding on why some individuals with so-called beneficial alleles do not control viral replication. This study also provides evidence that late escape at multiple sites in/and proximal to TL9 epitope was associated with increased viraemia in subtype C HIV-1 infected individuals. This study expands our knowledge of Gag epitopes that are important to take into account when designing vaccine strategies.

Abstract

This study aimed to determine the impact of HLA mediated viral polymorphisms transmitted to HLA-mismatched individuals on disease progression. Although some studies have shown that infection with viruses carrying beneficial HLA-mediated escape mutations provide survival benefit, this finding remains controversial.

Gag was sequenced from 56 participants with acute HIV infection and HLA associated mutations known to impact on replication capacity were mapped. Specific amino acid sites that were associated with low viral loads and high CD4+ counts at 12 months post-infection were also analyzed. For the newly identified mutation (Q65H), replication of wild-type and mutant viruses in PBMCs was compared.

In HLA-B*57/58:01-negative participants, 51% (25/49) were infected with viruses carrying at least one of the B*57/58:01 attenuating escape mutations in ISW9 (A146P:19/49), KF11 (A163G: 10/49) or TW10 (T242N/S: 11/49) epitope. In HLA-B7 negative participants, only 9.4% (3/32) had viruses carrying mutations in the HLA-B7 immunodominant TL9 epitope. A novel mutation in Gag p17, Q65H was identified, which in PBMCs assays reduced replication by 22% compared to wild-type. This mutation, thought to be restricted by HLA-B7 and A*30, was in 5.5% (2/37) of B7 and A*30 negative participants. Unlike a previous study in this cohort which analysed a subset of 21 individuals, infection with viruses carrying mutations in HLA-B*57/58:01 restricted epitopes alone did not impact on viral load setpoint in HLA-mismatched participants, although CD4 counts at 3 months post infection were higher in participants with greater than any 2 of these mutations (p=0.031). However, HLA-mismatched participants infected with viruses carrying 3 or more of any of the B*57/58:01 or B*39:10/81:01 attenuating escape mutations had significantly lower viral load and higher CD4 counts at 3 months (p=0.017 and p=0.041) and at 12 months (p=0.008 and 0.091) post-infection, respectively. These results suggest that multiple mutations generated when viruses are passaged through individuals with beneficial HLAs are needed to attenuate the virus, supporting vaccination approaches that aim to render the virus less fit.

4.1 Introduction

It has long been known that infection with viruses containing deleterious mutations, such as deletions in *nef*, can result in delayed disease progression (Deacon *et al.,* 1995; Michael *et al.,* 1995). Recently, a study of Alizon *et al.* (2010), who investigated the relationship between viral genotype and viral load set-point in 1100 MSM (men who have sex with men) infected with phylogenetically linked viruses, demonstrated that more than a half of variations in viral load at set point are due to inheritable HIV pathogenic traits (virulence). Furthermore, studies of transmission pairs ranging from 24 to 112 heterosexual couples found that there was a correlation between HIV-1 viral load levels within couples (Hecht *et al.,* 2010; Hollingsworth *et al.,* 2010).These studies support the idea that shared characteristics of the transmitted virus are important determinants of HIV-1 disease progression; however, the identity and contribution of some of these viral factors on differential disease outcome remain poorly defined. Studying HLA associated escape mutations in viruses following their transmission to HLA-mismatched hosts may further provide insights into viral determinants affecting disease progression.

Studies have shown that transmission of viruses carrying attenuating CTL escape mutations to individuals lacking the HLA that selected the attenuating mutations (HLA mismatched individuals) is beneficial to recipients (Chopera *et al.,* 2008; Goepfert *et al.,* 2008; Thobakgale *et al.,* 2009; Prado *et al.,* 2010).These mutations were shown to affect viral loads up to 12 months post infection (Chopera *et al.,* 2008), and this may be a possible mechanism that explains why viral load is an inheritable trait. However, the role of transmitted escape mutations on viral load remains controversial as some studies could not find any advantage of infection with viruses containing CTL-associated attenuated mutations (Novitsky *et al.,* 2010; Wright *et al.,* 2011).In an attempt to resolve the conflict in the field, this chapter reports on an extension of an earlier study by Chopera *et al.* 2008. Here, 56 individuals were analyzed, including the 21 reported by Chopera *et al.,* (2008). Furthermore the impact on disease progression of other transmitted CTL escape mutations, including a novel mutation in p17 identified in this study, were investigated.

4.2 Material and Methods

4.2.1 Study subjects

Participants in this study were those of the CAPRISA 002acute infection cohort, investigating the role of viral and immunological factors in acute and early HIV-1infections (van Loggerenberg *et al.,* 2008), described in detail in chapter 2.2.1 of this thesis.

4.2.2 Viral sequencing

As described in chapter 2, cDNA was synthesized from RNA and viral sequencing was done directly on PCR products or was cloned in pGem-T Easy vector according to the manufacturer's instructions (Promega, USA).

4.2.3 HLA typing

HLA data was provided by C. Gray, NICD, which was high resolution (four digits) typing and performed as described in Chapter 2.

4.2.4 Site-directed mutagenesis for generation of infectious molecular clones

The HIV-1 subtype C infectious molecular clone, pBR246-F10 was used as the backbone. The vector component of the original plasmid was digested with XhoI enzyme to delete the XhoI restriction site and this was followed by blunting with Klenow and blunt-end ligation. Cloning sites BssHIIand XhoI (by silent mutations) were then inserted in the backbone by site directed mutagenesis using the QuickChangeII site-directed mutagenesis kit (Stratagene). The BssHIIsite was inserted at the beginning of the *gag* gene (second and third amino acids) using primers Gag-BssHII F and Gag-BssHII R, while an XhoI site was inserted just at the end, at ~60 nucleotides downstream of the *gag* stop codon using primers Gag-XhoIF and Gag-XhoI R (all primer sequences used in this chapter are shown below).

The p17 65H mutation was introduced into a consensus subtype C full length *gag* by PCRsite directed mutagenesis using Gag-BssHIIF and Q65HR, generating two fragments ~200-300bp and~1200-1800pb, respectively. Ligation of these two fragments was done using the Gag-BssHII Fand Gag CRM primers. This product was then cloned into the pGEM-T Easy vector (Promega, USA) and the mutation was confirmed by sequencing using Gag Af and Ar primers (Primers described in chapter 2).

Both the wide-type (65Q) and mutant (65H) full length *gag* clones were each ligated into the subtype C backbone and propagated in transformed E. coli OneShot Stable 3 (Invitrogen) cells, generating full-length chimeric plasmids pBR246-F10_WT and pBR246-F10_MT clones respectively, stocks of which were purified using a Qiagen Plasmid Midi Kit (Qiagen, Valencia, CA).

Sequences of the primers used in this chapter were: Gag-XhoIF (HXB2 position 2338– 2369; 5'ATG ATA CAG TAC TCG AGG ACA TAA ATT TGC CA); Gag-XhoIR(HXB2 position 2338 – 2369; 5'TGG CAA ATT TAT GTC CTC GAG TAC TGTATCAT), Gag-BssHIIF (HXB2 position 778– 810; 5'-AGA AGG AGA GAG ATG GGCGCGCCA ACT CGAGGA-3');Gag-BssHIIR (HXB2 position 778–810; 5'-TCC TCG AGTTGGCGCGCCCATCTCTCTCCTTCT-3');Q65HR (HXB2 position 974–996; 5'-CTGAAGAGCTGGGTGTAGCTGTT-3'); Q65HF (HXB2 position974–996; 5'-AACAGCTACACCCAGCTCTTCAG-3'); Gag CRM (HXB2 position 2334–2353; 5'- TCC TCG AGT ACT GTA TCA TCT GC-3').

4.2.5 Viral stocks

HEK293T cells were transfected with 7.5 μg of plasmid DNA in Dulbecco modified Eagle medium (Gibco) supplemented with 10% fetal bovine serum (Atlantic Biologicals) using PolyFect(Qiagen, Valencia, CA). Supernatants were harvested after 72 hours and aliquots were stored at -80°C. Virus titration was performed in quadruplicate in TZM-bl cells. Eleven serial four-fold dilutions of viral stocks were assayed. After 48 hours of incubation at 37°C with 5% CO_2in Dulbecco modified Eagle medium (Gibco), cells were assayed for b-

galactosidase-luciferase activity and viral titres were expressed as tissue culture infective dose (TCID), calculated using Reed and Muench,(1938) method (Excel sheet).

4.2.6 Viral replication assays

CD8+ T cells were depleted from PBMCs isolated from 3 healthy HIV-negative donorsaccording to the kit manufacturer's instructions (Dynal, Invitrogen). The PBMCs were then stimulated in RPMI 1640 supplemented with 20% FBS, 5 ug/mL of phytohemagglutinin (Sigma, St. Louis, Mo.,USA), 1%PenS and 20 U/mL of IL-2 (Becton Dickinson Labware) for 72 h at 37°C with 5% CO_2 and were then washed with PBS 1X (without Ca++/Mg++) supplemented with 1% (vol/vol) FBS (Wash media). In order to compare the replication between the two viruses (wild-type and mutant), the stimulated PBMCs were inoculated with viral stocks (MOI = 0.001) in independent experiments (parallel infections) in RPMI 1640 supplemented with 20% FBS, 1%PenS and 20 U of interleukin-2 (Growth media) and after 24 hours, the media (inocula) were removed, cells washed (with wash media) and replaced with fresh growth media (1mL). Sampling was done on days 3, 5, 7, 10, 12 and 14 with replacement of an equal volume of growth medium. Fresh PBMCs were added on day 7. Growth was followed up as increases in the concentrations of p24 as measured using the p24 ELISA kit (Perkin Elmer; GibcoBRL Life technologies) and following the manufacturer's instruction.

4.2.7 Statistical analyses

Amino acid sites encoded by the earliest *gag* sequences associated with low viral loads and high CD4+ counts at 12 months post infection were identify using Wilcoxon rank-sum T-tests and significance of p values was evaluated for multiple comparisons using the Bonferroni-corrected alpha values for each site (Marasuilo & McSweeney, 1977). To calculate relative viral fitness, the mean p24 concentration of wild type virus at each sampling interval was divided by the mean p24 concentration of mutant virus at a coincident time point post-infection [relative viral fitness = (Mean of triplicate p24 concentration of MT/mean of triplicate p24 concentration of WT). Furthermore, since parallel infection assays may be

seriously affected by variations in virus input (Dykes and Demeter, 2007), the exponential growth constant q for each virus in parallel cultures was estimated from the ratio of the concentration of p24 produced at each sampling time ($p24_t$) and the peak concentration ($p24_{max}$) of the virus produced – this method is adopted from Resch *et al.* (2002). Statistical analysis and graphical presentations were implemented in GraphPad Prism 5.0 (GraphPad Software, Inc.). To control for fluctuations of viral set-points at 12 and 36 months post-infection, we calculated viral load and CD4+ T-cell counts as medians of measurements taken at three time-points closest to these intervals. Statistical analysis of significance was based on Mann-Whitney two-tailed *t* test.

4.3 Results

Virus adaptations in the donor, including those mediated by the donor HLAs, will be contained in the sequence of the transmitted virus. This study aimed to determine the impact on disease progression of HLA mediated polymorphisms on the virus transmitted to HLA-mismatched individuals. Previously, Chopera *et al.* (2008) detected a survival advantage in nine out of 21 HLA-B*57/58:01 negative CAPRISA individuals infected with viral strains at enrolment, carrying the T242N mutation in the TW10 and/or the A146X processing mutation in the ISW9 epitope. These individuals had significantly lower viral load and higher CD4 counts at 3 months and 12 month post-infection (p=0.0077 and p=0.0129;and p=0.0275 and p=0.0172; for viral load and CD4 count differences at 3 and 12 months, respectively) (Chopera *et al.*, 2008). In a more detailed analysis, the study was expanded to include 56 individuals and to investigate the role of other HLAs that are associated with replication fitness cost to the virus in Gag. Furthermore, this study screened for novel mutations that impacted on viral loads and CD4 counts.

4.3.1 Identification of mutations that affect viral replication capacity

To assess the frequency of these attenuating mutations, and possibly an additive contribution of mutations on disease progression, HLA associated Gag mutations known to incur replicative fitness cost to the virus were mapped in viruses from 56 participants whose

sequences were available and longitudinally followed for at least 24 months post-infection or initiation of antiretroviral therapy (ART) (Table 4.1).To ensure that we were evaluating mutations that evolved in the previous host(s), transmitted escape mutations were defined as changes in or near epitopes not restricted by the HLAs of the recipient (HLA-mismatched individuals). These changes were mapped onto the earliest available sequence (median of 42 days post infection, IQR = 28 -56).

Mutations were in epitopes restricted by HLA B*57/B*58:01, HLA-B*81:01/39:10 and HLA-B*13 and included mutations in or proximal to ISW9, KF11 and TW10 (restricted by HLA B*57/B*58:01); TL9 (restricted by B*81:01/39:10); and RI9 (restricted by HLA-B*13). Mutations proven with *in vitro* experiments to reduce viral fitness including A146Pproximal to the ISW9 (Gag 146-**A**ISPRTLNAW-155), A163G in the KF11 (Gag 162-KA**A**FSPEVIPMF-172), E177D, T186S proximal and in TL9 (Gag 176-S**E**GATPQDLN**T**ML-188), respectively, T242N in the TW10 (Gag 240-TS**T**LQEQIAW-249) and I437L/M in the RI9 (Gag 429-RQANFLGK**I**-437) epitopes (Table 4.1) (Martinez-Picado *et al.,* 2006; Boutwell *et al.* 2008; Crawford *et al.,* 2009; Prado *et al.,* 2009; Troyer *et al.,* 2009; Rolland *et al.,* 2010; Wright *et al.,* 2010).

Table 4.1 Identification of transmitted escape mutations associated with a replicative fitness cost to the virus and participants' clinical data (See Appendix 3). Shown are mutations in the HLAs -B*57/58:01 restricted epitopes, ISW9, KF11 and TW10; the B*39:10/81:01restricted epitopeTL9; and the B*13 restricted epitope RI9. Participants infected with viruses carrying a novel putative CTL escape mutation in p17 affecting viral replication capacity are also included.

PID	ISW9 (147-155)		KF11 (162-172)		TL9 (180-188)	TW10 (240-249) and associated mutations						RI9 (429-437)	p17*
	A146P	I147L[h]	A163X	S165X		T242X	A248X	M250X	H219Q	I223V	M228		
CAP008[J]			A163G	S165N			A248G						
CAP030[J]					T180I					I223V	M228L		
CAP037	A146S	I147L								I223V		K436RI437L	
CAP040						T242S				I223V	M228L		
CAP045[a, g]	A146P	I147M			L188F			M250I	H219Q				
CAP061[g]	A146P	I147L				T242N			H219Q	I223V	M228I		Q65H
CAP063	A146N	I147L	A163G	S165N			A248G						
CAP065[a, g]	A146S	I147L											
CAP069[J]							A248T				M228I		
CAP085[g]	A146P	I147L				T242N					M228L		Q65H
CAP084[J]													Q65H***
CAP088[g]	A146P		A163G		Q182H, P181S	T242N				I223N			
CAP129	A146P						A248T				M228L		
CAP137						T242S							
CAP174[J]								M250I					
CAP177[a]		I147L					A248T						
CAP188		I147L											
CAP200[a, g]	A146P	I147L				T242N				I223N			
CAP206[J]									H219P	I223P	M228L		
CAP220													Q65H
CAP221	A146P	I147L	A163N			T242N				I223V			
CAP224	A146P	I147L											
CAP225[g]	A146P	I147L				T242N							
CAP228[g]	A146P	I147L	A163S			T242N				I223V	M228I		
CAP244[J]							A248N						
CAP255[g]	A146P	I147L	A163S					M250I					Q65H
CAP257		I147L											Q65H**
CAP258[J]		I147L					A248N						Q65H
CAP260							A248T						
CAP261[a]		I147L	A163G	S165N				M250I					
CAP264[a]							A248T						
CAP265			A163G	S165N				M250I					
CAP266		I147L									M228L		
CAP267	A146P		A163G	S165N		T242N				I223P			
CAP269	A146P												Q65H
CAP271	A146S	I147L											Q65H
CAP275		I147L											
CAP277	A146P	I147L						M250I					Q65H
CAP278		I147L											Q65H**
CAP279	A146P					T242N	A248Q		H219Q		M228I		
CAP281		I147L	A163N					M250I					
CAP136													
CAP210[J]													
CAP222													
CAP237													
CAP248[J]													
CAP256[J]													
CAP262													
CAP282													
CAP211													Q65H
CAP217													
CAP229													
CAP239													
CAP268													
CAP270													
CAP274													

Blank cell (s): the specified mutation(s) is/are absent; Individuals who are positive for HLAs - B*57/58:01are shown in red PID numbers; X: change away from HIV-1 C clade consensus; [a]Participant has HLA-B*1510 or HLA-B*13 associated with ISW9 residues A146X/I147X (Honeyborne *et al.,* 2007); Participants [J]negative (n=11) and [g]positive (n=9) for T242N/X in the TW10 and/or A142P/X in the ISW9 epitope included in the study (n=21) of Chopera *et al.* (2008); [h]Though previously thought to compensate for fitness cost due to escape through A146X, Crawford *et al.* (2009) showed that I147L, together with A146P mutation incur fitness cost to the virus; *Gag p17 mutation at HXB2 position 65; **Participant infected with virus that developed the mutation at HXB2 position 65 in Gag p17 or else was there at transmission –in purple colour: the polymorphism is present in virus in participants positive for HLA-B7 supertype or A*30 alleles that may be linked to formation of the polymorphism. *** Mutation reverted to subtype C consensus residue; Mutations in the cyclophilin binding loop known to compensate for escape in TW10 (T242X) are highlighted in Green.

4.3.2 The frequency of mutations associated with replicative fitness cost in HLA mismatched participants

4.3.2.1 Frequency of HLA B*57/58:01 associated escape mutations in HLA mismatched participants

HLA B*57 and B*58:01 restricted responses select mutations in the ISW9, KF11 and TW10 epitopes in p24 in acute or early HIV-1 infection (Martinez-Picado *et al.,* 2006; Crawford *et al.,* 2009). Of the 56 participants, 49 were negative for the B*57/58:01 alleles. The frequency of the ISW9 processing mutation A146P, and the KF11 epitope mutation A163G was 39% (19/49) and 20% (10/49) respectively, while the frequency of the T242N in the TW10 epitope was 23% (11/49) (Table 4.2). Taken together, a high percentage of individuals were infected with viruses carrying the HLA B*57 and B*58:01 associated escape mutations: 43% (21/49) participants were infected with virus carrying either the T242N and/or the A146X processing mutation in the ISW9 epitope, while 51% (25/49) were infected with at least one of the A146P/T242X/163X HLA B*57/B*58:01associated mutations.

Furthermore, a study of Martinez-Picado *et al.* (2006) describing escape in TW10 epitope suggested that escape through T242N mutation is rare if there is either M250I or S252N mutation downstream of the TW10. In addition, the M250I mutation has been shown to affect viral replication capacity (Chopera *et al.,* unpublished). In agreement with this finding, none of the B*57/58:01-negative participants who were infected with T242N variants had the M250I mutation, suggesting that co-existence of T242N and M250I might indeed be under constraint. The S252X (X=N/H) mutation was present in five participants infected with the T242N variants – thus, suggesting that the two mutations can co-exist. There was co-variation of sites M250 and S252, where six of the seven (6/7) participants with variation at site M250 also had variation at site S252. The frequency of M250I and S252X mutation was 14 % (7/49) and 23% (11/49) respectively.

4.3.2.2 Frequency of HLA-B7 supertype allele-associated mutations in HLA mismatched participants

Of the 32 participants negative for the HLA-B7supertype alleles (HLAs B*81:01, B*42:01, B*39:10, B*07:02) and Cw*08:02 capable of selecting mutations in the TL9 epitope (Leslie *et al.,* 2006; Llano *et al.,* 2009), none were infected with viruses containing the well described E177D/A and T186S mutations associated with reduction in viral fitness (Von Schwedler *et al.,* 2003; Rolland *et al.,* 2010; Wright *et al.,* 2012). Furthermore, only 9.4% (3/32) participants had mutations (T180I, P181S, Q182X and L188F) in the TL9 epitope. However, although no study *in vitro* has been done to prove the fitness cost incurred to the virus due to mutations T180I, P181S and L188F, these reverted *in vivo* between 1.5 and 2.5, 6 and 10.6, and 31 and 46 months post-infection, respectively suggesting a fitness cost of these mutations (Appendix 4).

These results suggest that mutations in the B*39:10/81:01 TL9 epitope known to incur replicative fitness cost to the virus, unlike HLA B*57/58:01 associated mutations, are not common in viruses transmitted to HLA mismatched participants, despite the high frequency of HLAs that are capable of selecting mutations in the epitope in HIV-1, subtype C population (Goulder *et al.,* 2000; Gonzalez-Galarza *et al.,* 2011) and in this cohort (40%; 25/62).

4.3.2.3 Frequency of HLA B*13 associated mutations in HLA mismatched participants

Although HLA- B*13 is not always significantly associated with slow disease progression in HIV-1 subtype C infection, in a study of 578 infected individuals from Durban, Honeyborne *et al.* (2007) found that individuals with CTL responses to the RI9 epitope and who select for CTL escape mutations I437L/M had lower viral loads than others. Furthermore, mutations in this residue were associated with replicative fitness cost *in vitro* (Prado *et al.,* 2009). In this study, it was found that these mutations were rare and were present in only 2% (1/55) of the individuals who did not have the HLA-B*13 allele. The low occurrence of this mutation is probably due to the low frequency of B*13 allele in black African populations as only 3.2% (2/62) participants in this cohort were positive for the allele, which is in agreement with the

study of Honeyborne *et al.* (2007), who found that the phenotypic frequency of this allele in 1212 participants was 3.9%.

Table 4.2: Summary of frequencies of mutations in Gag due to HLAs known to restrict dominant CTL epitopes associated with viral control.

Defined epitope/ HLA Restriction	[b]HLA +ve	[c]HLA -ve	HXB2 Position of Mutation	HLA –ve individuals with the mutation	%
ISW9 B*57/58:01	*7*	*49*	A146P[f]/X	5	10.2
			S147L	9	18.4
			A146P[f]/X and *S147X*	14	29.0
KF11 HLA-B*57/58:01	*7*	*49*	A163G[f]/X	5	10.2
			S165X	0	0
			A163G[f]/X and S165X	5	10.2
TW10 and Associated mutations HLA-B*57/58:01	*7*	*49*	T242N[f]/X	10	20.4
			T242N[f]/X, H219Q[k] and A248Q[d]	1	2.0
			[e]M250I	1	2.0
			M250I and S252N	2	4.1
			M250I and S252G	4	8.2
			S252N	5	10.2
TL9 B*81:01/42:01/39:10/07 :02 or Cw*08	*24*	*32*	E177D[f]	0	0
			T186S[f]	0	0
			T180A/X	1	3.1
			P181S/X	1	3.1
			L188F	1	3.1
RI9 HLA-B*13	*1*	*55*	I437M/L[f]	1	2.0
p17 Q65H[g] HLA-B7Supertype (B*81:01/42/39:10/07:02)/A*30	*29*	*37*	Q65H[g]	2	5.5

[b]Number of HLA positive participants; [c]Number of HLA mismatched participants; ***Number of individuals with mutation; X: change away from HIV-1 clade C consensus; [d]T242N + G248A in subtype B completely abrogates recognition at low peptide concentrations (Leslie *et al.,* 2004); [e]Included as a study of Martinez-Picado *et al.* (2006) associated this mutation with lack of variation in TW10 epitope and also associated with *in vitro* fitness cost (Chopera *et al.,* Unpublished); [f]Proven *in-vitro* to incur fitness cost to the virus; [g]Statistically associated with lower viral loads; exact HLA associations remain controversial (B*81: Wright *et al.,* 2010; A*30: Huang *et al.,* 2011); [k]Compensatory mutation (Brockman *et al.,* 2007).

4.3.3 Screening for novel mutations

Specific amino acid sites that are associated with low viral loads and high CD4+ counts at 12 months post-infection were analyzed in the earliest sequences using Wilcoxon rank-sum tests. This method compared independently the median viral loads and CD4+ counts between groups of viruses with and without the consensus amino acid at each site. Data for participants positive for the B*57/58:01 alleles were initially excluded as these "protective" alleles are often associated with better control of HIV-1 replication (Migueles *et al.*, 2000; Miura *et al.*, 2009). When corrected for multiple comparison a significant association was detected for a mutation at site 65 (n=10) in p17 with lower viral load (p=0.0033) and a trend towards higher CD4+ cell counts at 12 months post-infection (p=0.1758) although this trend for higher CD4 counts was lost when individuals positive for the B*57/58:01 alleles were included in this analysis (section 4.3.4.5). At this site, glutamine (Q) is mutated to histidine (H). The Q65H mutation was not located in a known epitope; however, longitudinal analysis of sequences demonstrated that this site evolved in two participants. In CAP278 (HLA-A*30:01,43:01; B*39:10, 42:02; Cw*12:03, 17:01), a viral load controller, the mutation Q65H was first observed between 54 and 85 weeks, while it was formed between 50-107 weeks post infection in CAP257 (HLA-A*23:01, 29:02; B*42:02, 44:03; Cw*17:01,17:01), an intermediate disease progressor.

A recent study of Huang *et al.* (2011) in a cohort of 1053 black South Africans from Bloemfontein and Durban, associated this Q65H polymorphism with HLA-A*30. However, of the two individuals in the CAPRISA cohort with viruses that evolved this polymorphism, one was negative for HLA-A30 suggesting that this is either not the restricting allele or both alleles are associated with this residue. Interestingly, however, 11 of the 16 HLA-A30 participants in this study were also positive for an HLA-B7 supertype allele, suggesting an over representation of A*30 allele in participants positive for B7 alleles. Furthermore, both individuals were positive for the HLA-Cw*17:01 and HLA-B7 supertype alleles (HLA-B*42:02 and B*39:10) suggesting that these may be the restricting alleles. The hypothesis that this polymorphism is associated with HLA-B7 supertype is further supported by a study of Wright *et al.* (2010), who analysed 406 participants in Durban, South Africa and found an

association between this residue and escape in TL9, a dominant B7 epitope. Nevertheless, further analysis using bioinformatic tools that predict binding of peptides to any known MHC molecule using artificial neural networks (ANNs) (Hoof *et al.*, 2009; available at: www.cbs.dtu.dk/services/NetMHCpan) suggested that both B7 supetype and A30 alleles bind the epitope (60-IMKQL**Q**PAL-68) harbouring the Q65H mutation.

4.3.3.1 Frequency of Q65H mutation in participants negative for HLA-B7 supertype and A*30 alleles

Sequences were further analyzed to determine the frequency of the p17 mutation Q65H in the cohort and 18% (10/56) participants were found to be infected with viruses carrying the mutation. However, of the 37 participants negative for B7 and A30 alleles, only two were infected with virus carrying the polymorphism suggesting that in this cohort, the frequency of the Q65H polymorphism in HLA-mismatched participants was 5.5% (2/37).

4.3.3.2 Polymorphism at p17 Gag Q65H affects viral replication

Studies *in vivo* showed the Q65H mutation to be associated with lower viral loads. Furthermore, this residue is situated in the fourth α-helix of p17, where mutations in this helix are known to affect envelope incorporation, dimerization and, therefore, viral assembly (Dorfman *et al.,* 1994; Morikawa *et al.,* 1998). *In vitro* studies were performed to determine the impact of the p17 Gag Q65H mutation on viral replication. The Q65H mutation was introduced into a consensus subtype C *gag*, and both the wide-type (65Q) and mutant (65H) full length *gag* were each ligated into a subtype C backbone, pBR246-F10 forming chimeric pBR246-F10_WT and pBR246-F10_MT infectious molecular clones, respectively.

Infective viral stocks were successfully generated for both the wide-type and the mutants. Results of this study on the growth kinetics of viruses generated in parallel infections in PBMCs demonstrated that there was increasing concentrations for both mutant and wide-type over the entire 3 to 10 days of follow-up suggesting that the mutation Q65H did not abrogate virus formation (Figure 4.1A and B). However, p24 concentrations of these viruses in PBMCs suggested that the mutant virus replicated slower that the wide-type virus. Moreover, when an exponential growth model (Resch *et al.*, 2002) was fitted to the data from each culture (Figure 4.1C), the net growth rate (q) suggested that the mutant virus was less fit ($q=2.1\pm0.4$ day^{-1}; doubling time: 3.34 days) than the wild-type virus ($q=2.5\pm0.6$ day^{-1}; doubling time: 2.81 days) although the differences were small. This analysis was confined to replication titres between day 3 and 7 as the growth kinetics of this study demonstrated that exponential increase levelled off after day 7. The net growth rate (q) for the mutant was also lower (4.19 ± 0.7 day^{-1}; doubling time: 1.7 days) than that of the wild-type (5.43 ± 1.3 day^{-1}; doubling time: 1.3 days) when the exponential growth curve was fitted between day 3 and day 5. Taken together, it was found that the mutant replicated 78% of the wild-type (22% less than the wild-type) (Figure 4.1D). This polymorphism was thus included in the analysis of mutations associated with lower viral loads.

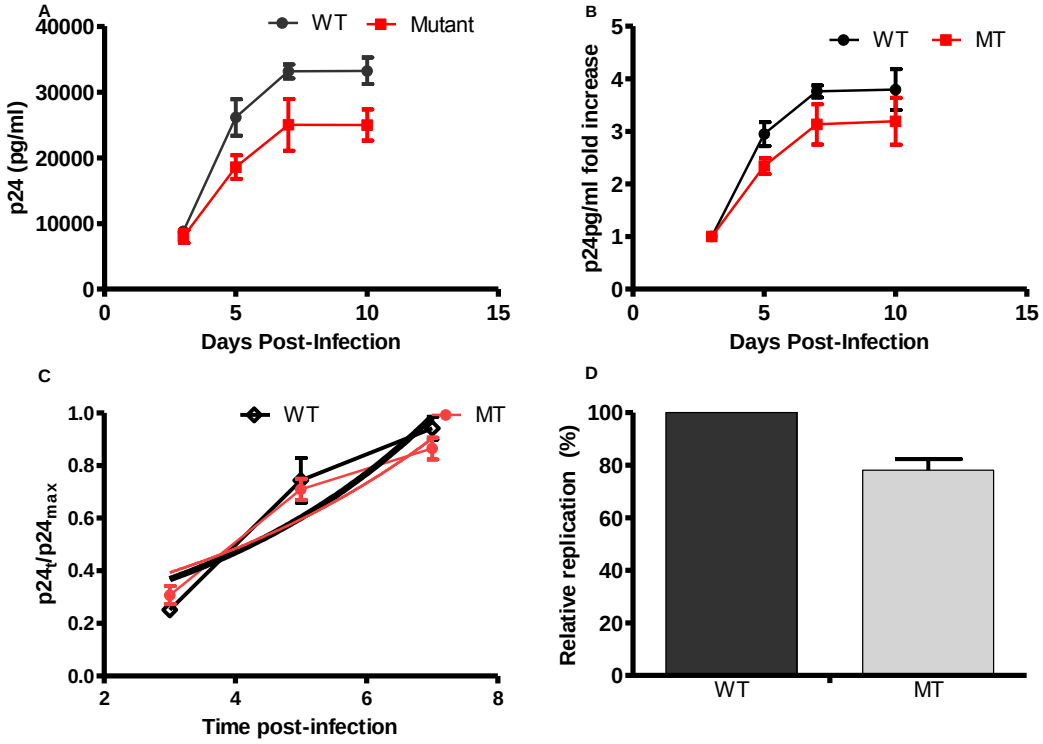


Figure 4.1 Impact of the polymorphism at p17 Gag Q65H on viral replication. Replication kinetics of chimeric wild-type (pBR246-F10_WT) and mutant (pBR246-F10_MT) viruses carrying the Gag p17 65Q and 65H mutations, showing changes in p24 concentrations over the first 10 days (3 repeats) (A) and fold increases (B), respectively; Mean values of triplicates and standard deviation at each sampling point time post-infection are shown; (C): Viral spread is normalized/measured as ratio of concentrations of p24 at time t ($p24_t$) to the maximum (peak) p24 ($p24_{max}$) produced. The exponential growth curves for the wild-type is in black, while that of the mutants is in red colour (D): Relative replication of the mutant, pBR246-F10_MT viruses compared to that of the wide-type pBR246-F10_WT viruses. Relative replication is calculated as percentage of mean ratio of mutant/wild type (MT/WT).

4.3.4 Impact of transmitted CTL escape mutations on disease progression

4.3.4.1 Impact of HLA-B*57/58:01 footprints in HLA mismatched participants on disease progression

Firstly, the analysis of Chopera *et al.* (2008) was repeated where, in this study in 21 HLA-B*57/58:01-negative individuals from the CAPRISA 002 cohort, they had found that nine were infected with viruses containing either the T242N or A146P, or both mutations and these were significantly associated with lower viral load and higher CD4 counts at both 3 and 12 months post infection. In the current study in 56 participants (including the initial 21), 43% of the HLA-B*57/58:01 negative participants (21/49) were infected with viruses carrying either the T242N and/or the A146X processing mutation in the ISW9 epitope at enrolment (Table 4.1). However with this increased sample size, and unlike the Chopera *et al.* (2008), there were no significant differences in either viral load or CD4counts between those infected at enrolment with and without virus carrying the T242N and/or the A146X mutations, either at 3 or 12 months post-infection. At 3 months post infection, the medians for viral load and CD4 counts for individuals infected with and without viruses carrying the mutations were 4.59 log10 copies/ml (IQR= 3.72 – 5.01) versus 4.74 copies/ml [IQR= 4.01 - 5.22] for viral loads, respectively (Figure 4.2A; p=0.28); and 535 cells/ul [IQR = 448-691] versus 467 cells/ul [IQR=358-612] for CD4 counts, respectively (Figure 4.2B; p= 0.24). Similarly, at 12 months post infection, the medians for viral load and CD4 counts for individuals infected with and without viruses carrying the mutations were 4.51 log10 copies/ml (IQR= 3.33 – 4.82) versus 4.12 copies/ml [IQR= 3.34-4.82] for viral load, respectively (Figure 4.2C; p=0.96); and 402 cells/ul [330-566] versus 425 cells/ul [IQR=301-572] for CD4 counts, respectively (Figure 4.2D; p=0.95).

Thus, using expanded numbers, it was not possible to reproduce the findings that individuals infected with virus carrying the T242N and/or the A146X mutations in HLA-B*57/58:01 mismatched participants affected viral loads or CD4 counts.

To further understand this result, it was investigated if the selective pressure of the B*57/58:01 alleles and their additionally associated mutations played a role. Since mutations

in the ISW9 epitopes may be selected by HLA-B*13 and B*15:10 (Honeyborne *et al.,* 2007), viruses were defined as carrying B*57/58:01 signature mutations (Table 4.3) only if individuals were infected with viruses containing either the T242N/S mutation in the TW10 epitope (as this mutation is unique to B*57/58:01); or if either of the two additional HLA-B*57/58:01 restricted mutations in epitopes ISW9 (A146X/I147X), KF11 (A163G) and M250I mutation upstream of the TW10 were present.

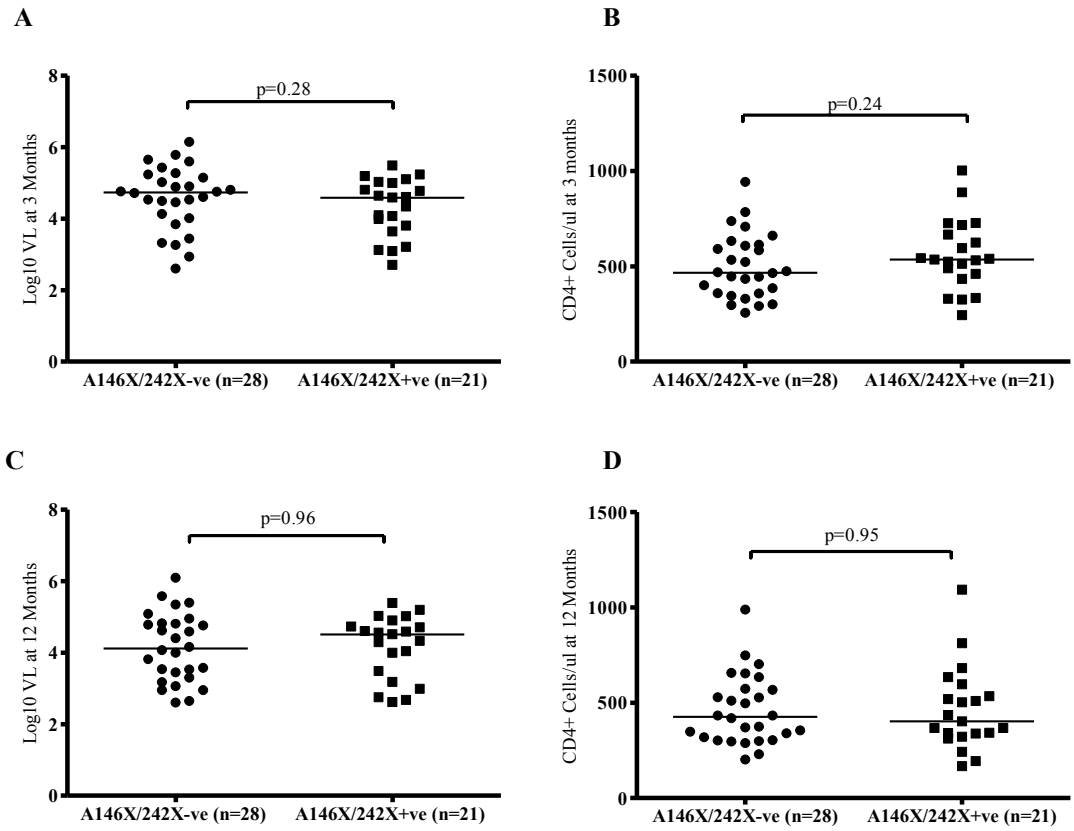


Figure 4.2 Impact of infection with viruses carrying the T242X (X=N/S) and/or A146X (X=P/S) mutations on viral load and CD4 counts. Viral load (A and C) and CD4 counts (B and D) at 3 and 12 months post-infection for the B*57/B58:01 negative individuals (n=49) compared between participants infected with viral strains carrying either both or any of the TW10 escape mutation and the ISW9 processing escape mutation (n = 21) and those that did not (n = 28).

Thirty-eight % (18/49) of the cohort were grouped as infected with viruses carrying the B*57/B*58:01 signature, with 61% of these (11/18) infected with viruses carrying the T242N/S mutation (Table 4.3). To determine if B*57/58:01 negative individuals infected with

viruses carrying B*57/58:01 signature mutations had lower viral loads and higher CD4 counts, these markers of disease progression were compared between participants infected with viruses with (n=18) and without the B*57/58:01 signature mutations (n=31) at enrolment. There were no significant differences in viral load between participants infected with and without virus carrying the B*57/58:01 signature mutations, either at 3, 6 or 12 months post-infection (Figure 4.3A, C and E). However, individuals infected at enrolment with virus carrying three of any of the signature mutations had higher CD4 counts at 3 months post-infection (p=0.031; Figure 4.3B). Nonetheless, this trend was not sustained either at 6 months or at one year post-infection (Figure 4.3D and F) suggesting that the potential benefit derived from B*57/58:01 associated mutations was limited to the early phase of the infection.

Table 4.3 HLA-B*57/58:01 negative participants infected with viruses carrying HLA-B*57/58:01 signature mutations

PID	ISW9 (147-155)		KF11 (162-172)		TW10 (240-249) and associated mutations					
	A146P	**I147L**	**A163X**	**S165X**	**T242X**	**A248X**	**M250X**	**H219Q[a]**	**I223V[b]**	**M228X[d]**
CAP040[y]					T242S				I223V	M228L
CAP045	A146P	I147M					M250I	H219Q[c]		
CAP061	A146P	I147L			T242N			H219Q[c]	I223V	M228I
CAP063	A146N	I147L	A163G	S165N		A248G[c]				
CAP085	A146P	I147L			T242N					M228L
CAP088	A146P		A163G		T242N				I223N	
CAP137[y]					T242S					
CAP200	A146P	I147L			T242N				I223N	
CAP221	A146P	I147L	A163N		T242N				I223V	
CAP225	A146P	I147L			T242N					
CAP228	A146P	I147L	A163S		T242N				I223V	M228I
CAP255	A146P	I147L	A163S				M250I			
CAP261*		I147L	A163G	S165N			M250I			
CAP265*			A163G	S165N			M250I			
CAP267	A146P		A163G	S165N	T242N				I223P	
CAP277	A146P	I147L					M250I			
CAP279	A146P				T242N	A248Q[c]		H219Q[c]		M228I
CAP281*		I147L	A163N				M250I			

[a,b,d] :Known (approved *in vitro*) B57/5801 compensatory mutations ; [y]Participant carrying T242S variant associated with minimal fitness cost to the virus *in vitro* (Boutwell *et al.,* 2008); [c]Mutations that independently increase infectivity (Martinez-Picado *et al.,* 2006; Boutwell *et al.,* 2008). [*]Participant not fitting the Chopera *et al.* (2008) definition i.e., viruses do not have the T242N/A146X mutation.

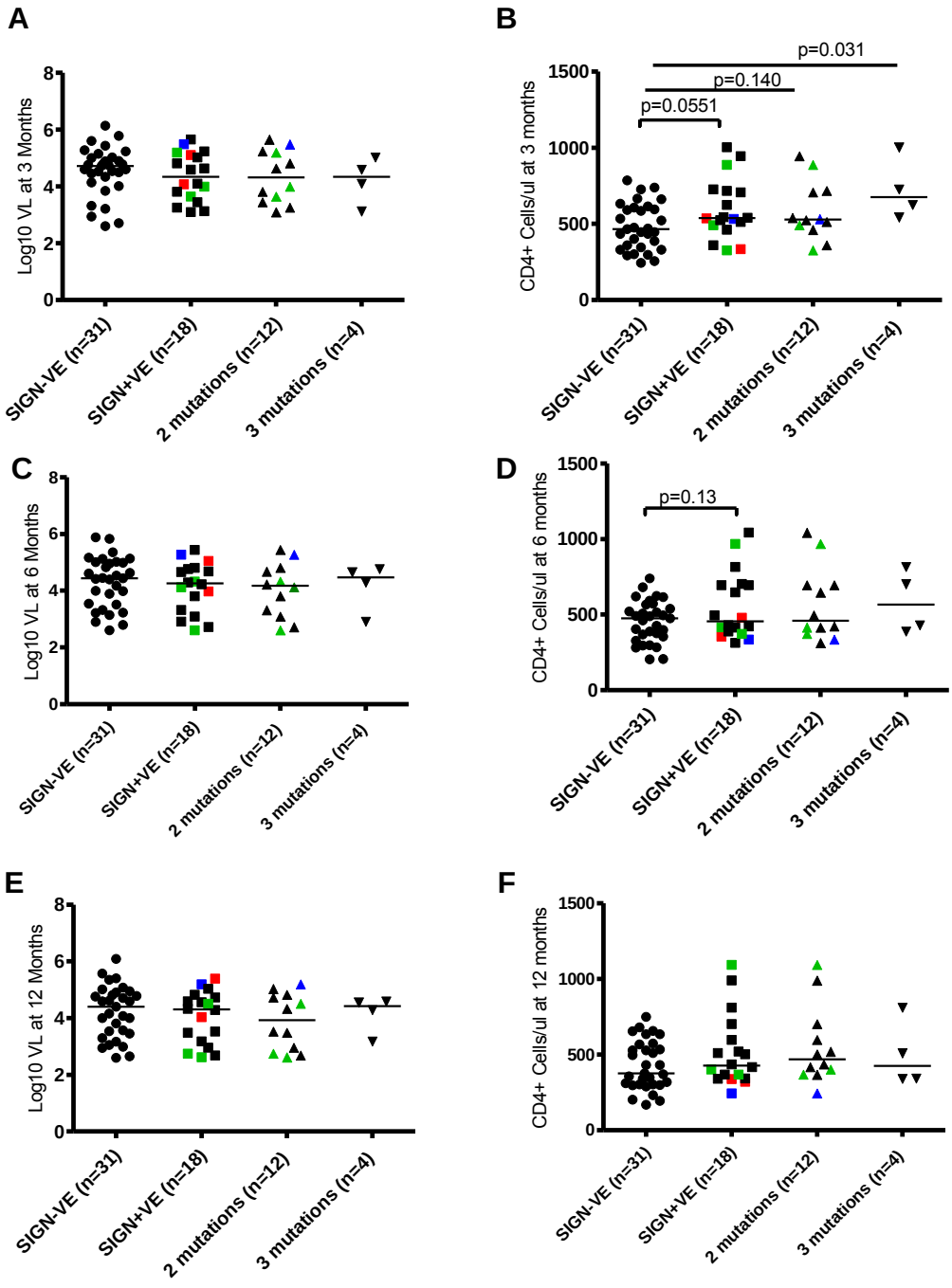


Figure 4.3 Impact of B*57/58:01 footprints in HLA mismatched participants on disease progression- viral load and CD4 counts were grouped according to presence or absence of B*57/58:01 signature mutations (having T242N/S in the TW10 epitope or else, two mutations from any two of the different HLA-B*57/58:01 restricted epitopes ISW9 (A146X/I147X), KF11 (A163G) and M250I upstream of the TW10). Viral load (A, C and E) and CD4 counts (B, D and F) compared at 3, 6 and 12 months post-infection between individuals with and without the B*57/58:01 signature mutations. Colour codes Green, Red and Blue are participants infected with viruses carrying H219Q, T242S and A248T variants, respectively independently associated with wide-type level of fitness or more fitter virus (Brockman *et al.*, 2007; Boutwell *et al.*, 2008; Martinez-Picado *et al.*, 2006).

4.3.4.2 No defined influence of transmitted B*57/58:01 compensatory mutations on the course of disease progression

Gag mutations known to compensate for the replicative fitness cost incurred due to escape through the T242N mutation are H219Q, I223V and M228L (Brockman *et al.,* 2007; Crawford *et al.,* 2007), with mutations I223V and M228L reported to partly compensate for the replicative fitness cost, whereas the H219Q mutation alone is able to restore wild-type fitness to the T242N mutation *in vitro* (Boutwell *et al.,* 2009).

Sequences were analyzed for compensatory mutations in order to answer the question that the extended data might have had more compensatory mutations than there were in the data of Chopera *et al.* (2008) resulting in a loss of signal. The frequency of transmitted compensatory mutations was generally high in these participants. Of the 11 participants infected with viruses carrying the T242X, viruses in 64% (7/11) carried the I223X (X=V/N/P) mutation while viruses in 46% (5/11) had the M228X (X=L/I). Furthermore, 27% (3/11) of these participants were infected with viruses carrying both I223X and M228X mutations together. Viruses in two participants 18% (2/11) had the H219Q mutation. Nevertheless, when the numbers of mutations compensating escape through T242N/S in TW10 epitope present in the dataset of Chopera *et al.* (2008) were compared with those in the expanded dataset, there were no significant variations in either the number or the location of compensatory mutations (Figure 4.4).

Furthermore, data was investigated to determine if grouping participants according to number of compensatory mutations could affect viral load or CD4 counts (Table 4.4). However, there was no evidence suggesting that the number of compensatory mutations in the infecting virus had an influence on either viral load or CD4 counts.

Overall, this study found limited influence of the transmitted B*57/58:01 compensatory mutations on the course of disease progression in these HLA mismatched participants.

Table 4.4 Participants infected with viruses carrying the T242X mutation grouped according to the number of compensatory mutation (s) in the virus

No. of comp. mutations[a]	PID	VL[b]@12 months (Copies/ML)	[f]CD4 counts @12 months(Cells/uL)
0	137[e]	246000	338
	225[c]	21500	597
1	85[c]	3030	519
	88[c]	38700	510
	200[c]	108000	367
	221[e]	53100	435
	267[e]	19300	341
2	40[e]	11000	321
	228[c]	1520	811
	279[e]	32200	368
3	61[c]	418	402

[a]Number of compensatory mutations; [bf]Viral loadand CD4 counts are median of 3 values around 12 months, respectively; [c]Participant previously included in the dataset ofChopera *et al.* (2008)/[e]extended data

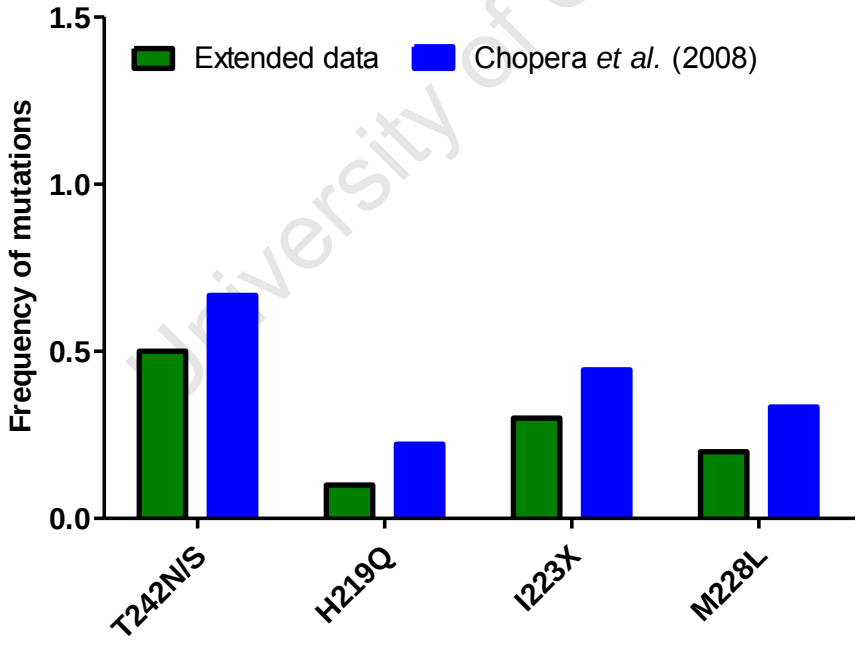


Figure 4.4 Comparison of frequency of mutations compensating for the fitness cost incurred due to escape through T242N/S in TW10 epitope present in viruses in the extended data with that of Chopera *et al.* (2008).

4.3.4.3 Protective HLA (B*57/B*58:01 or B*81:01/39:10) associated mutations revert to consensus in HLA mismatched participants

Next, the timing of reversion was analyzed in mutations conferring escape in protective HLA-positive individuals that were present in mismatched recipients in order to understand the constraint associated with these mutations and to determine if their reversion over time could have had a measurable impact on markers of disease progression. Reversion was defined as change from a low to high frequency amino acid in an epitope not restricted by the HLA of the host i.e. change towards the subtype C consensus (most common) residue.

Reversion of mutations at residues T242N, M250I and T180I/P181S generally occurred earlier in infection and at: 11 [IQR = 7 – 16]; 11 [5 – 20] and 1.2-2.5/6-10.6 months post-infection respectively, while it occurred later at residues A146P or A146P/I147L: 30 [16 – 34] and L188F (31 - 46) months post-infection (Table 4.5).In one of the 3 participant negative for HLA-B7 alleles, reversion occurred at the Q65H novel mutation in the p17 between 3.7 and 5.3 months post infection. There was no reversion in residues, A163G conferring escape in KF11 epitope observed over 36 months of follow-up.

Interestingly, however, there was reversion occurring within 1.5 years post-infection in 82% (9/11) of participants infected with virus carrying the T242N/X mutation, while only 57% (4/7) of those with M250I mutation had shown reversion. In participants infected with viruses carrying the A146P or A146P/I147L mutations, only 6% (1/15) of those investigated had reverted within 1.5 years post-infection, while only 33% (5/15) participants had the mutation reverting by 46 months post infection. On the other hand, there was reversion occurring within 1.5 years post-infection in 66.7% (2/3) of HLA-mismatched participants infected with virus carrying the TL9 mutation (T180I/P181S), while L188F reverted later (31-46months post-infection)(Table 4.5).

Table 4.5: Summary of timing of reversion of B*57/58:01 associated mutations over 36 months period.

Mutation	Median and [IQR] (Months) Post-infection	% Reversion[a]
A147P/I147L	30 [16 – 34]	33 (5/15)
A163G[b]	-	0 (0/10)
T242N	11 [7 – 16]	82 (9/11)
M250I	11 [5 – 20]	57 (4/7)
T180I/X[c]	1.2 – 2.5	100 (1)
P181S/X[d]	6 – 10.6	100 (1)
Q182H[d]	-	0 (0/1)
L188F[e]	31 - 46	100 (1)
p17Q65H[f]	3.7 – 5.3	22[f] (1/3)

[a]Percentage of (No. of individuals with mutations reverting/Total No. of those with the mutation) individuals with viral mutation reverting to consensus within the median time; [b]No reversion observed over the 36 months period; [c-e]Only one participant had the mutation (three participants altogether) and thus too few to estimate the median timing of reversion; however, these mutations reverted *in vivo*; [f]There was reversion in one participant – reverts in other cohorts (Huang *et al.,* 2010)

Of interest in this study, 9/18 participants that were defined as carrying B*57/58:01 signature mutations associated with fitness cost to the virus (Table 4.3) had either 1 or 2 mutations reverting to consensus before one year post-infection (Table 4.6). Furthermore, viruses carrying the T242N mutation had reverted to consensus in 6/11 participants by 12 months post infection, which suggested that the sequential loss of impact of higher CD4 counts due to B*57/58:01 signature mutations, observed at 3 months post-infection probably were at least in part, due to reversion.

Table 4.6 Summary of participants infected with HLA associated viral mutations in the HLA mismatched participants and changes in viral load and CD4 counts accompanying reversions over one-year time post-infection

PID	[b]No.	Timing of reversion of mutation in month post infection & ΔVL[c]/ΔCD4[d]				
		<3	>3 - 6	ΔVL[c]/ΔCD4[d]	>6 – 12	ΔVL[c]/ΔCD4[d]
CAP085	2		T242N	+0.83/-211		
CAP088	5		T180I	+0.07/-24	T242N, P181S	-0.07/-193
CAP137	1		T242S	-0.051/-46		
CAP200	2				T242N	-0.41/-47
CAP221	3				T242N	+0.50/-160
CAP255	4	M250I				
CAP261	2				M250I	0.00/-4
CAP267	3		A146P	+0.19/-112	T242N	0.00/-90
CAP281	2				M250I	-0.133/+4

[b]Number of mismatched HLA associated viral mutations present at enrolment ; **c/d:** Change in $\log_{10}$ viral load (Copies/mL)/Change in CD4 counts (Cells/uL) respectively; **+/-** Increase/decrease, respectively.

4.3.4.4 Modest impact of early reversion of protective HLA (B*57/B*58:01 or B*81:01/39:10) associated mutations on disease progression

Studies have shown that reversion of escape mutations to consensus amino acid residues in viruses transmitted to individuals lacking the selecting HLA is a measure of the fitness cost the mutant virus incurred due to donor HLA mediated escape (Liu *et al.,* 2007; Novitsky *et al.,* 2011; Henn *et al.,* 2012).

To better quantify the overall impact of reversion of some of the protective HLAs (B*57/B*58:01 or B*81:01/39:10) transmitted mutations on viral load and CD4+ counts, we compared changes in $\log_{10}$ viral load and CD4+ counts between individuals infected with viruses carrying these mutations that reverted and those in whom mutations did not revert before the first 12 months post-infection. Since reversion occurring before one year would be expected to raise viral load or lower CD4 counts at 12 months set-point, we calculated change as a difference in viral load/CD4 counts at 12 months and viral load/CD4 counts at 6 months.

There was no evidence to support the idea that reversion of B*57/58:01 signature mutations in viruses infecting these participants resulted in an increase in viral load (Figure 4.5A). Although there was a net negative median change in CD4 counts for individuals infected with viruses carrying the B*57/58:01 signature mutations that reverted before 1 year post-infection (suggesting that reversion influenced the decrease in CD4 counts), this was not significant (Figure 4.5B).

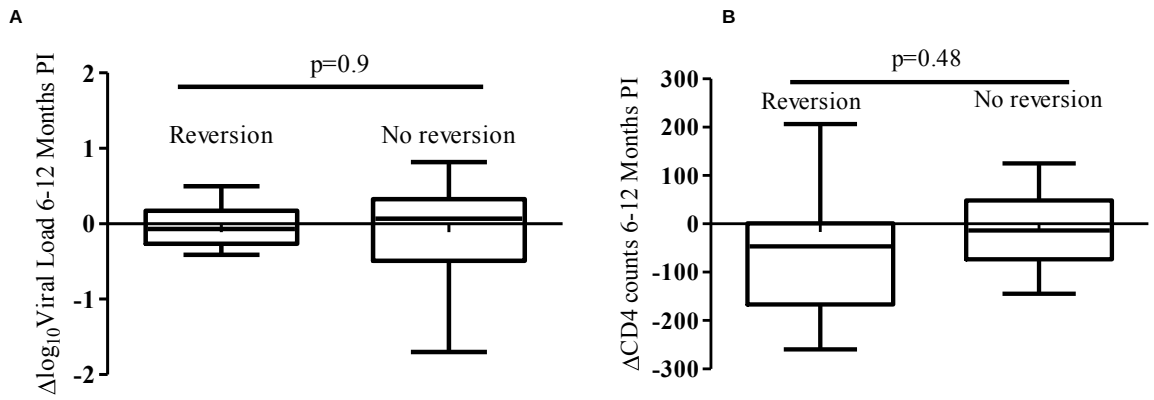


Figure 4.5 Impact of early reversion of protective HLA B*57/58:01 and B*39:10/81:01 mutations on disease progression. Figures A and B: Comparison of change (Δ) in log_{10} viral load (A) and CD4 counts (B) between individuals infected with viruses carrying the protective HLAs signature mutations that reverted (n=9) and those in which mutations did not revert (n=9) before the first 12 months post-infection. (Δ) = Viral load/CD4 counts at 12 months - Viral load/CD4 counts at 6 months, respectively.

4.3.4.5 Polymorphism at p17 Gag Q65H associates with lower viral load

In order to estimate the impact of this mutation in this cohort, participants infected with viruses that evolved sequence at site 65 (n=2) were excluded and viral load and CD4+ counts were compared between participants with (n=10) and without (n=44) the Q65H mutation, irrespective of their HLAs. There was lower viral load at set point for participants infected with viruses carrying the Q65H mutation (Figure 4.6A; p=0.0056), while CD4+ counts did not significantly differ between participants infected with and without viruses carrying the polymorphism (Figure 4.6B; p>0.05).

It was not possible to compare viral load and CD4+ counts *in vivo* in HLA-mismatched individuals between participants negative and those positive for the Q65H mutation due to the small number of participants who were negative for HLA-A*30 and B7 supertype alleles and infected with this mutation (n=2) (Figure 4.6 C and D; open squares). However, when we compared viral load and CD4+ counts at set point between participants negative for supertype -B7 alleles (n=32) infected with viruses carrying the wild-type Q65 (n=28) and with those infected with the polymorphisms H65 (n=4), there was lower viral load at set point for participants infected with viruses carrying the H65 mutation (Figure 4.6C; p=0.0402), while

CD4+ counts did not significantly differ between participants infected with and without viruses carrying the polymorphism in these participants (Figure 4.6D; p>0.05). However, these findings were supported by those of Huang *et al.* (2011), who demonstrated that participants infected with viruses carrying the Q65H mutation have significantly higher CD4+ counts in two chronic cohorts totalling 1053 individuals.

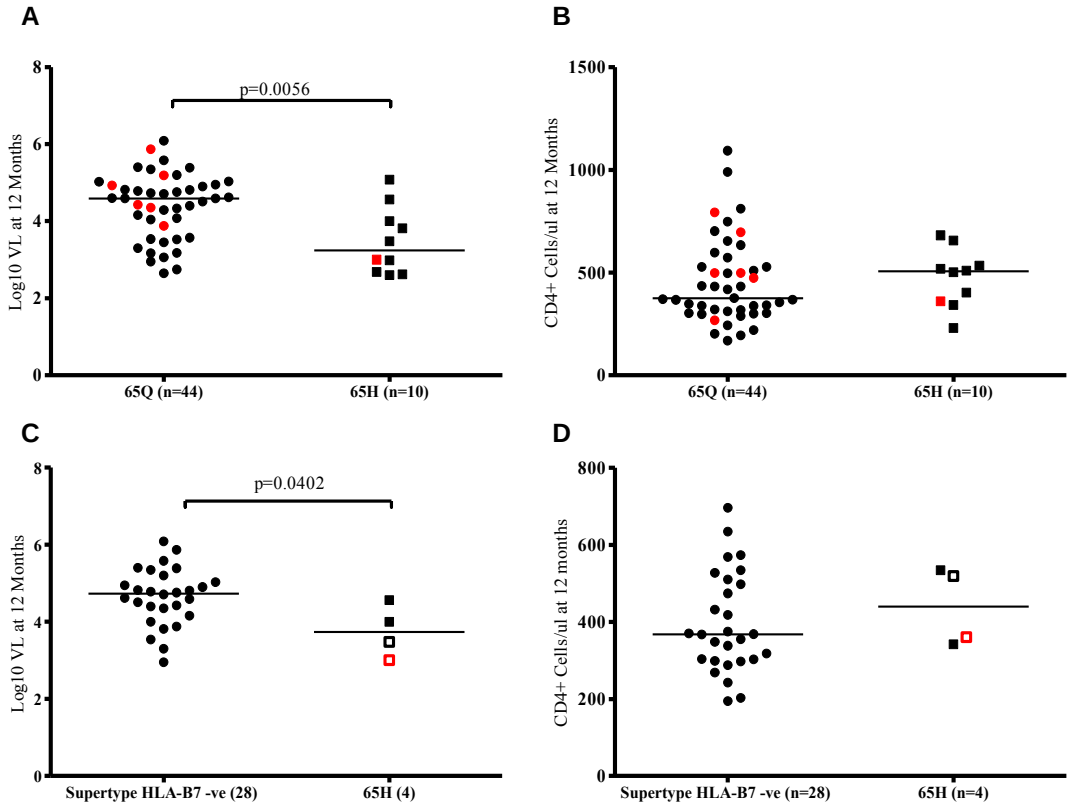


Figure 4.6 Association of the amino acid Histidine at position 65 in p17 Gag with lower viral load. Viral load (A) and CD4 counts (B) compared at 12 months post infection for individuals infected with viral strains with and without the polymorphism – individuals infected with virus that evolved this polymorphism (n=2) were excluded from this analysis; Viral load (C) and CD4+ counts (D) at 12 months post infection compared between participants negative for supertypeHLA-B7 alleles (n=32) infected with viruses carrying the wide-type Q65 (n=28) and with the variant H65 (n=4). Participants in Red colour carry the B*57/58:01 alleles known for their associations with better infection outcome in southern African subtype C clade. Open squares are participants positive for HLA-A*30, also likely to be associated with the mutation. P values are Mann-Whitney t-test results.

4.3.4.6 CTL imprints due to protective HLAs additively attenuate the virus

A study of Boutwell *et al.* (2008) demonstrated that the replication capacity *in vitro* due to the three B*57/58:01 footprint mutations was in the order A146P<A163G<T242N (less to greatest) and multiple mutations had an additively negative effect on viral replication capacity.

A list of participants infected with viruses carrying three or more mutations associated with fitness cost to the virus was then compiled (Table 4.7) and viral loads and CD4 counts for HLA-mismatched participants infected with viruses carrying 3 or more cumulative mutations of HLAs associated with survival benefits were compared in order to determine if these mutations had any impact on disease progression. Thus, all participants with protective HLA-B57/58:01 or HLA-B7supertype alleles (B*81:01, B*42 and B*39:10 and B*07:02) were excluded from this analysis as individuals in this cohort were associated with selection of the Gag, p17 mutation Q65H included in this analysis and also mutations in the B7 immunodominant TL9 epitope (Table 4.7).

HLA-mismatched participants infected with viruses carrying three or more replicative fitness cost-associated mutations had significantly lower viral load at 3 and 12 months post-infection (p=0.0147 and p=0.0241, respectively; Figure 4.6 A and C). There were also significantly higher CD4 counts for participants infected with these viruses at 3 months and 12 months post-infection (p=0.0004 and 0.0438, respectively; Figure 4.6 B and D). Of interest was that none of the individuals included in this analysis (CAP045, CAP085, CAP088, CAP221, CAP255 and CAP267) carried any protective HLAs as all HLAs associated with protection (HLA-B*81:01, B*58:01/57 and B*39:10) in HIV-1, subtype C (Kiepiela *et al.,* 2004; Leslie *et al.,* 2010) were excluded from this analysis. These results suggest that viruses passaged through participants carrying beneficial HLAs may additively attenuate these viruses, with consequential reduction in severity of infection in HLA mismatched participants.

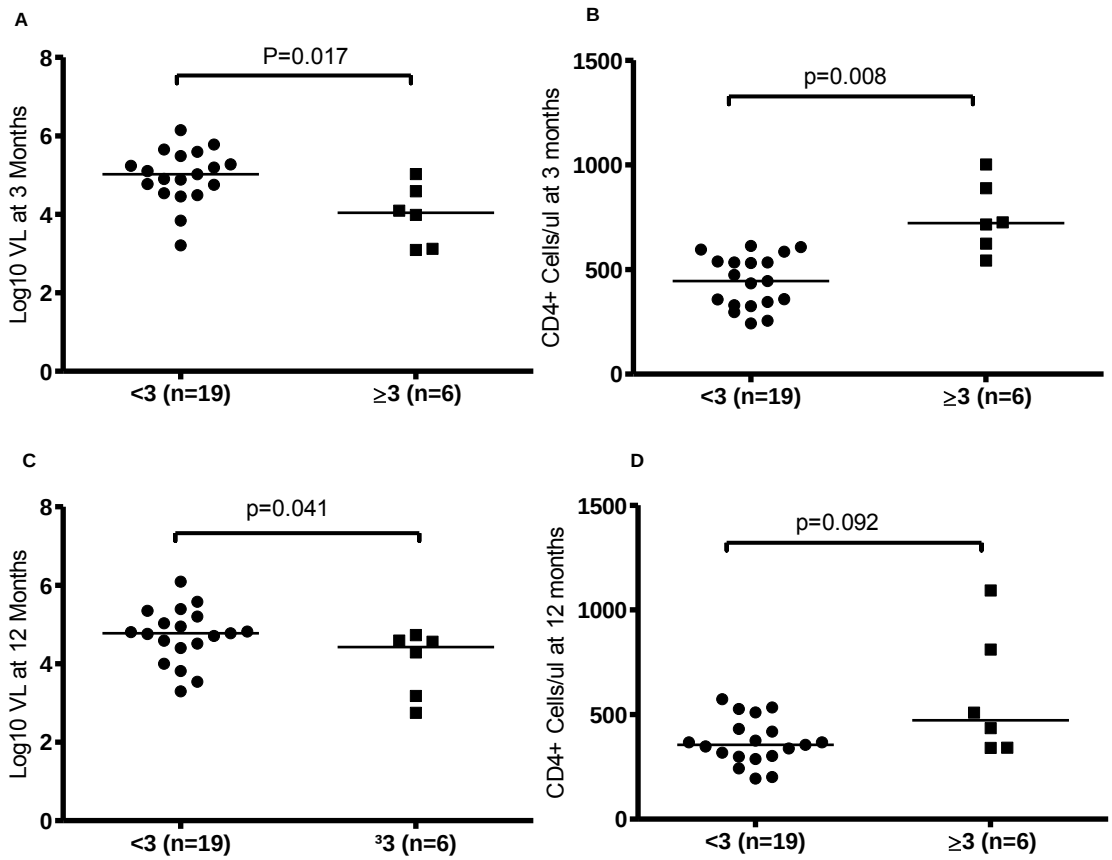


Figure 4.7 Analysis of individuals with non-beneficial HLAs comparing the effect of at least three mutations on viral and CD4 counts. Additive impact of three or more HLA mediated mutations associated with fitness cost to the virus in non-beneficial HLA participants on disease progression- viral load and CD4 counts were grouped according to presence or absence in mismatched recipients at enrolment of three or more associated mutations, ISW9 (A146X), KF11 (A163X), M250I/T242X, TL9 associated mutations (E177D, T180X, P181S/X, D184X, T186S, L188F) and Q65H in p17, due to B*57, B*58:01, B*81:01, B*42:01, B*39:10, B*07:02 and Cw*08:02. Viral load (A and C) and CD4 counts (B and D) compared at 3, 6 and 12 months post-infection.

Next, a critical review of the list of participants infected with viruses carrying cumulative mutations associated with fitness cost to the virus was undertaken in relation to mutations mediated by HLA-B*57/58:01 in order to understand the results of the additive influence of mutations mediated by other beneficial HLAs.

By comparison, of the 9 participants previously described by the study of Chopera *et al.* (2008) – CAP045, CAP061, CAP065, CAP085, CAP088, CAP200, CAP225, CAP228 and CAP255, five carried additional mutations due to other HLAs. The HLA-B7 allele-restricted TL9 epitope anchor (L188F) and rare (P181S) residues respectively, were present at enrolment in CAP045 and CAP088, while in CAP061, CAP085 and CAP255, the p17 mutation associated with lower viral loads and higher CD4 counts, both in this cohort and other cohorts (Huang *et al.,* 2011) were present at enrolment (Table 4.7).

Table 4.7 Participants with viruses carrying three or more mutations due to mismatched HLA at enrolment (transmitted mutations) and are associated with fitness cost

PID	A146P[a]	A163X[a]	TL9[a] Epitope Gag (180-188)	T242X[a]	M250X[a]	p17[a]
CAP045	A146P		L188F		M250I	
CAP061[b]	A146P			T242N		**Q65H**
CAP085[b]	A146P			T242N		**Q65H**
CAP088	A146P	A163G	P181S Q182H	T242N		
CAP221	A146P	A163N		T242N		
CAP228	A146P	A163S		T242N		
CAP255	A146P	A163S			M250I	**Q65H**
CAP267	A146P	A163G		T242N		
CAP277[b]	A146P				M250I	**Q65H**

[a]: Mutation associated with reduced viral replication capacity either *in vitro* or *in vivo*/both; [b]Participants carrying HLAB7supertype/A*30 alleles were excluded from disease progression analyses

4.4 Discussion

Here we have shown that the frequency of HLA-B*57/58:01-associated signatures are far greater than mutations associated with other HLAs that are associated with beneficial outcome in HIV-1 subtype C infection: in HLA-mismatched participants, HLA-B*57/58:01 associated escape mutations occur at a higher frequency than HLA-B7 mediated escape mutations despite the fact that HLA-B7 has a much higher frequency in this cohort (40%; 25/62), compared to HLA-B57/58:01 (13%; 8/62). This result may be due to B*57/58:01 alleles selecting for escape mutations in Gag in early infection when chances of viral transmission are very high due to the high acute phase viral loads (Dyer *et al.,* 1998; Wawer *et al.,* 2005; Cohen *et al.,* 2010). In contrast, the HLA-B7 alleles normally select late mutations, a period when the risk of transmission is relatively lower.

This difference in frequency of HLA-B*57/58:01 associated mutations when compared to those of B7 mutations is unlikely to be due to rapid reversions as most reversions tend to be fixed only after acute (90 – 100 days post-infection) infection (Leslie *et al.,* 2004; Henn *et al.,* 2012; Chopera *et al.,* 2008) and viral sequences in this study were mainly from acute infection [42 days (IQR: 28-56) days post-infection]. Moreover, reversion of the B*57/58:01 associated mutations in and proximal to the TW10 (residues T242N and M250I) [IQR = 5 – 20] in this study occurred no later than reversions in the B7 restricted TL9 at T180I/P181S (around 1.2-2.5/6-10.6 months) post-infection. However, there were too few participants infected with viruses carrying the TL9 or associated mutations to accurately estimate the timing of their reversion (Table 4.5).

This study identified a novel putative escape mutation in Gag p17, Q65H, found in longitudinal viral sequences from two participants with two similar HLA class1 alleles, HLA-B*42:02 and Cw*1701. These findings harmonise with those of Wright *et al.*(2010), who associated this mutation with escape in individuals positive for B*81 alleles and those of Huang *et al*. (2011), who associated the mutation with the A*30 allele. This study has extended previous HLA association of the evolution of the Q65H mutation in p17 Gag and here is documented a novel putative HLA-B7 supertype and HLA-A30 epitope, consequently

providing opportunities to map out and further study the virological and immunological implication of evolution in this epitope.

People infected with viruses carrying the mutation Q65H generally had lower viral loads [first reported in a keystone symposium on HIV Pathogenesis (X6) in Santa Fe, January 12 - 17, 2010, Ntale *et al.,* abstract number 307]. Analyzing over 1000 participants, a study of Huang *et al.* (2011) also found that this mutation was associated with higher CD4 counts. This residue is located in the fourth α-helix of p17, in which mutations (residue 54 to 68) that form part of a finger-like projection from one side of the molecule abolish dimerization (Chazal *et al.,* 1995; Morikawa *et al.,* 1995). Furthermore, MA amino acids between 41 and 68, which include (Q65), were found to be critical for soluble Gag secretion and Env-incorporation (Chazal *et al.,* 1994; Dorfman *et al.,* 1994, Morikawa *et al.,* 1995).

In vitro assays demonstrated for the first time that viruses carrying the mutation (65H) replicate 22% slower than viruses carrying wild-type residue (65Q) in parallel infections in PBMCs, in a subtype C infectious molecular clone. Studies have demonstrated that parallel infections are insensitive to small replication differences and are much more seriously affected by variations in the starting concentrations of the viruses than competition infections (Quiñones-Mateu and Arts, 2001; Dykes and Demeter, 2007). The effects of variations in the starting concentrations were reduced by normalising the concentration of p24 produced at each sampling time ($p24_t$) to the peak concentration ($p24_{max}$) of the virus produced as had been previously done by Resch *et al.* (2002). Interestingly, a recent study of Wright *et al.* (2012) associating this mutation with compensation of viral escape through residue T186S in TL9 found that viruses carrying T186S/Q65H mutation could not replicate in the GXR cell line, and suggested that 65H mutation does not compensate for the fitness cost of T186S.

This study documented that HLA-mismatched participants infected with viruses carrying 3 or more HLA footprints associated with fitness cost in Gag had significantly lower viral load and higher CD4 counts at 3 and 12 months post-infection. Unlike a previous study in this cohort, infection with viruses carrying mutations in HLA-B*57/58:01 restricted epitopes, ISW9 (A146P) and TW10 (T242N/S) alone did not impact on viral load setpoint in HLA-

mismatched participants. However, the majority (5/9) of the participants in the dataset of the study of Chopera *et al.* (2008) carried mutations due to other HLAs that are associated with viral replication *in vivo*. This result suggested that although HLA-B*57/58:01 is the major driver of the process, the advantage previously detected by the study was due to an additive influence, together with other beneficial HLAs, other than HLA-B*57/58:01 alone.

Taken together, these results support the idea that individuals infected with attenuated viruses have better viral control than those without these viral factors; however this benefit it is likely to be transient and is only one of multiple factors contributing to long-term control of viremia.

CHAPTER 5: SUMMARY AND CONCLUSION

Defining correlates of protection against HIV-1 infection is the major challenge in developing a prophylactic HIV vaccine that will prevent infection (Koup *et al.*, 2011; McElrath and Haynes, 2010). However, while it remains generally accepted that an effective vaccine should elicit both broadly cross-neutralizing antibodies and CTL responses (Koup *et al.*, 2011; Richman *et al.*, 2009; Richman *et al.*, 2003), the nature of the CTL response associated with control of viremia remains elusive (Koup *et al.,* 2011).

Studying patterns of immune escape in HIV-1 acute infection provides a sensitive tool to identify specific viral regions under active immune selection pressure and contribute to our understanding of the correlates of immune control of HIV-1-infection. Escape that is detrimental to the virus will result in lower replication capacity - creating a balance between the benefit of escape to the virus and its survival (Crawford *et al.*, 2009; Martinez-Picado *et al.*, 2006; Schneidewind *et al.*, 2007; Woo *et al.*, 2010).

This study sought to determine the frequency, timing and location of CTL escape, and to address the question of how CTL escape impacts on disease progression in subtype C HIV-1 infection. Most studies to date have focused on the characterization of acute infection in subtype B viruses, and a few studies have investigated CTL escape in acute infection in subtype C (Crawford *et al.,* 2009; Gray *et al.,* 2009; Radebe *et al.,* 2011). Studies of different subtypes, together with differences in HLA frequencies between populations have exposed differences in characteristics of these viruses suggesting that observations in subtype B infections may not be directly extrapolated to subtype C infections.

This study found that Gag and/or Nef escape occur in the majority of people (69%) in the first year of infection and is predominantly driven by HLA-B alleles targeting p24. Escape occurs most frequently in the acute infection phase, a period when viremia is at its highest level. High viral load turnover, together with early responses that control replication may both contribute to the rapid evolution of CTL epitopes. This thesis shows an overall benefit of early

escape with individuals with acute escape having increasing CD4+ T cell counts in the first 3 months compared to individuals without viral escape at 12 months. Although this trend was not sustained after 3 months, this early preservation of CD4+ T cell counts appeared to provide a survival benefit within the first year of infection.

Furthermore, whereas there is a lot of information on the role of HLA-B*57/58:01 associated mutations and disease progression (Brockman *et al.*, 2010; Chopera *et al.*, 2011; Crawford *et al.*, 2009; Migueles *et al.*, 2000; Miura *et al.*, 2009), there is limited information on B*81:01, an allele that has been associated with control of replication in southern Africa. Although CTL mediated escape mutations in the B*8101 restricted p24, TL9 epitope had been associated with *in vitro* fitness cost to the virus, the impact on disease progression *in vivo* had never been investigated.

This study documented the kinetics and emergence of known B*81:01 escape mutations (E177D, Q182S and T186S) and identified a common pattern of evolution with the Q182S mutation generally emerging prior to T186S. Escape was linked to co-evolving residues in the region of Gag known to be associated with host tropism and these residues were identified as putative compensatory mutations. Furthermore, unlike the B*57/58:01-restricted TW10 escape, which usually occurs rapidly and during acute and early infection (Chopera *et al.,* 2011; Crawford *et al.,* 2009; Martinez-Picado *et al.,* 2006), escape in individuals positive for the B*81:01 allele was often late and took nearly a year to reach fixation. Moreover, while *in vitro* studies demonstrated that the T186S mutation alone, or in combination with E177D incurs a fitness cost to the virus (Rolland *et al.,* 2010; Wright *et al.,* 2010), late evolution of T186S either alone, or in combination with E177D/and Q182S, coincided with an increase in viral load and although viral replication was subsequently brought under control, it returned to a level higher than before escape. These results provided not only *in vivo* evidence of the detrimental impact of late B*81:01 mediated viral evolution in a single Gag p24 epitope on control of viremia, but also the impact of the response targeting this epitope and consequently highlighting the complexity of immune escape that need to be considered in developing an HIV-1 vaccine.

Furthermore, in an attempt to resolve the conflict in the field over the impact of transmitted CTL escape mutations on disease progression, viral load and CD4 counts were compared in participants infected with viruses carrying polymorphisms associated with beneficial HLA-B*57/58:01 and B*39:10/81:01 selective pressure in HLA mismatched participants.

This study further documented that HLA-mismatched participants infected with viruses carrying 3 or more HLA footprints associated with fitness cost in Gag had significantly lower viral load and higher CD4 counts at 3 and 12 months post-infection. Unlike a previous study in this cohort (Chopera *et al.*, 2008), infection with viruses carrying mutations in HLA-B*57/58:01 restricted epitopes, ISW9 (A146P) and TW10 (T242N/S) alone did not impact on viral load set-point in HLA-mismatched participants. However, majority (5/9) of the participants in the dataset of the study of Chopera *et al.* (2008) carried mutations due to other HLAs that are associated with viral replication *in vivo*. This result suggested that although HLA-B*57/58:01 is the major driver of the process, the advantage previously detected by the study was due to an additive influence of beneficial HLAs, and not HLA-B*57/58:01 alone.

In conclusion, this study has elucidated the kinetics of CTL escape and has shown that escape occurs predominantly in HLA-B restricted epitopes and in acute infection. It finds that escape in acute infection provides a survival benefit in the first year of infection. This study has also provided *in vivo* evidence that some escape is detrimental, with late B*81:01 mediated viral evolution, in a single Gag p24 epitope, significantly influencing control of viremia. Moreover, results of this study suggest that multiple mutations generated when viruses are passaged through individuals with beneficial HLAs are needed to attenuate the virus. Overall, these results support vaccination methodologies that aim to render the virus less fit.

APPENDIX

Abebe,A., Kuiken,C.L., Goudsmit,J., Valk,M., Messele,T., Sahlu,T., Yeneneh,H., Fontanet,A., De,W.F., and Rinke De Wit,T.F. (1997). HIV type 1 subtype C in Addis Ababa, Ethiopia. AIDS Res. Hum. Retroviruses *13*, 1071-1075.

Abrahams,M.R. *et al.* (2009). Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. J. Virol. *83*, 3556-3567.

Adamson,C.S. and Freed,E.O. (2007). Human immunodeficiency virus type 1 assembly, release, and maturation. Adv. Pharmacol. *55*, 347-387.

Addo,M.M. *et al.* (2003). Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. J. Virol. *77*, 2081-2092.

Adhikary,D., Behrends,U., Moosmann,A., Witter,K., Bornkamm,G.W., and Mautner,J. (2006). Control of Epstein-Barr virus infection in vitro by T helper cells specific for virion glycoproteins. J. Exp. Med. *203*, 995-1006.

Aiken,C., Konner,J., Landau,N.R., Lenburg,M.E., and Trono,D. (1994). Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. Cell *76*, 853-864.

Aldovini,A. and Young,R.A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. J Virol *64*, 1920-1926.

Alizon,S. *et al.* (2010). Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. PLoS. Pathog. 6, e1001123.

Allan,J.S., Coligan,J.E., Lee,T.H., McLane,M.F., Kanki,P.J., Groopman,J.E., and Essex,M. (1985). A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients. Science *230*, 810-813.

Allen,T.M. and Altfeld,M. (2008). Crippling HIV one mutation at a time. J. Exp. Med. *205*, 1003-1007.

Allen,T.M. *et al.* (2005). Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. J. Virol. *79*, 13239-13249.

Allen,T.M. *et al.* (2004). Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. Journal of Virology *78*, 7069-7078.

Altfeld,M. *et al.* (2003). Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. AIDS. *17*, 2581-2591.

Altfeld,M. and Allen,T.M. (2006). Hitting HIV where it hurts: an alternative approach to HIV vaccine design. Trends in Immunology *27*, 504-510.

An,P. *et al.* (2004). APOBEC3G genetic variants and their influence on the progression to AIDS. J. Virol. *78*, 11070-11076.

An,P., Vlahov,D., Margolick,J.B., Phair,J., O'Brien,T.R., Lautenberger,J., O'Brien,S.J., and Winkler,C.A. (2003). A tumor necrosis factor-alpha-inducible promoter variant of interferon-gamma accelerates CD4+ T cell depletion in human immunodeficiency virus-1-infected individuals. J. Infect. Dis. *188*, 228-231.

An,P. and Winkler,C.A. (2010). Host genes associated with HIV/AIDS: advances in gene discovery. Trends Genet. *26*, 119-131.

Arold,S., Franken,P., Strub,M.P., Hoh,F., Benichou,S., Benarous,R., and Dumas,C. (1997). The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain

suggests a role for this complex in altered T cell receptor signaling. Structure. *5*, 1361-1372.

Atkins,K.M., Thomas,L., Youker,R.T., Harriff,M.J., Pissani,F., You,H., and Thomas,G. (2008). HIV-1 Nef binds PACS-2 to assemble a multikinase cascade that triggers major histocompatibility complex class I (MHC-I) down-regulation: analysis using short interfering RNA and knock-out mice. J Biol. Chem. *283*, 11772-11784.

Avila-Rios,S. *et al.* (2009). Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. Retrovirology 6.

Bailey,J.R., Lassen,K.G., Yang,H.C., Quinn,T.C., Ray,S.C., Blankson,J.N., and Siliciano,R.F. (2006). Neutralizing antibodies do not mediate suppression of human immunodeficiency virus type 1 in elite suppressors or selection of plasma virus variants in patients on highly active antiretroviral therapy. J. Virol. *80*, 4758-4770.

Ballana,E., Senserrich,J., Pauls, E., Faner,R., Mercader,J.M., Uyttebroeck,F., Palou,E., Mena,M.P., Grau,E., Clotet,B., Ruiz,L., Telenti,A., Ciuffi,A., Esté J.A. (2010). Zinc ribbon domain-containing 1 (ZNRD1) is a host cellular factor influencing HIV-1 replication and disease progression. Clin. Infect. Dis. 50, 1022–1032

Belyakov,I.M., Isakov,D., Zhu,Q., Dzutsev,A., and Berzofsky,J.A. (2007). A novel functional CTL avidity/activity compartmentalization to the site of mucosal immunization contributes to protection of macaques against simian/human immunodeficiency viral depletion of mucosal CD4+ T cells. J. Immunol. *178*, 7211-7221.

Bennett,M.S., Ng,H.L., Dagarag,M., Ali,A., and Yang,O.O. (2007). Epitope-dependent avidity thresholds for cytotoxic T-lymphocyte clearance of virus-infected cells. J. Virol. *81*, 4973-4980.

Bentham,M., Mazaleyrat,S., and Harris,M. (2003). The di-leucine motif in the cytoplasmic tail of CD4 is not required for binding to human immunodeficiency virus type 1 Nef, but is critical for CD4 down-modulation. J Gen. Virol *84*, 2705-2713.

Bentham,M., Mazaleyrat,S., and Harris,M. (2006). Role of myristoylation and N-terminal basic residues in membrane association of the human immunodeficiency virus type 1 Nef protein. J Gen. Virol *87*, 563-571.

Berger,C.T. *et al.* (2011). High-Functional-Avidity Cytotoxic T Lymphocyte Responses to HLA-B-Restricted Gag-Derived Epitopes Associated with Relative HIV Control. Journal of Virology *85*, 9334-9345.

Berger,E.A., Murphy,P.M., and Farber,J.M. (1999). Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. Annual Review of Immunology *17*, 657-700.

Bernardin,F., Kong,D., Peddada,L., Baxter-Lowe,L.A., and Delwart,E. (2005). Human immunodeficiency virus mutations during the first month of infection are preferentially found in known cytotoxic T-lymphocyte epitopes. J. Virol. *79*, 11523-11528.

Berthoux,L., Sebastian,S., Sokolskaja,E., and Luban,J. (2005). Cyclophilin A is required for TRIM5 alpha-mediated resistance to HIV-1 in old world monkey cells. Proceedings of the National Academy of Sciences of the United States of America *102*, 14849-14853.

Besnier,C., Takeuchi,Y., and Towers,G. (2002). Restriction of lentivirus in monkeys. Proc. Natl. Acad. Sci. U. S. A *99*, 11920-11925.

Bess,J.W., Powell,P.J., Issaq,H.J., Schumack,L.J., Grimes,M.K., Henderson,L.E., and Arthur,L.O. (1992). Tightly Bound Zinc in Human-Immunodeficiency-Virus Type-1, Human T-Cell Leukemia-Virus Type-I, and Other Retroviruses. Journal of Virology *66*, 840-847.

Best,S., Le,T.P., Towers,G., and Stoye,J.P. (1996). Positional cloning of the mouse retrovirus restriction gene Fv1. Nature *382*, 826-829.

Betts,M.R., Ambrozak,D.R., Douek,D.C., Bonhoeffer,S., Brenchley,J.M., Casazza,J.P., Koup,R.A., and Picker,L.J. (2001). Analysis of total human immunodeficiency virus

(HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection. J. Virol. *75*, 11983-11991.

Betts,M.R. *et al.* (2004). The functional profile of primary human antiviral CD8+ T cell effector activity is dictated by cognate peptide concentration. J. Immunol. *172*, 6407-6417.

Bhatia,A.K., Campbell,N., Panganiban,A., and Ratner,L. (2007). Characterization of replication defects induced by mutations in the basic domain and C-terminus of HIV-1 matrix. Virology *369*, 47-54.

Bhattacharya,J., Repik,A., and Clapham,P.R. (2006). Gag regulates association of human immunodeficiency virus type 1 envelope with detergent-resistant membranes. J. Virol. *80*, 5292-5300.

Bieniasz,P.D. (2009). The cell biology of HIV-1 virion genesis. Cell Host. Microbe *5*, 550-558.

Blagoveshchenskaya,A.D., Hannah,M.J., Allen,S., and Cutler,D.F. (2002). Selective and signal-dependent recruitment of membrane proteins to secretory granules formed by heterologously expressed von Willebrand factor. Mol. Biol. Cell *13*, 1582-1593.

Bonifacino,J.S. and Traub,L.M. (2003). Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu. Rev. Biochem. *72*, 395-447.

Borghans,J.A., Molgaard,A., de Boer,R.J., and Kesmir,C. (2007). HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. Plos One *2*, e920.

Borrow,P., Lewicki,H., Hahn,B.H., Shaw,G.M., and Oldstone,M.B.A. (1994). Virus-Specific Cd8+ Cytotoxic T-Lymphocyte Activity Associated with Control of Viremia in Primary Human-Immunodeficiency-Virus Type-1 Infection. Journal of Virology *68*, 6103-6110.

Borrow,P. *et al.* (1997). Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nat. Med. *3*, 205-211.

Boulet,S., Ndongala,M.L., Peretz,Y., Boisvert,M.P., Boulassel,M.R., Tremblay,C., Routy,J.P., Sekaly,R.P., and Bernard,N.F. (2007). A dual color ELISPOT method for the simultaneous detection of IL-2 and IFN-gamma HIV-specific immune responses. J. Immunol. Methods. *320*, 18-29.

Boutwell,C.L., Rolland,M.M., Herbeck,J.T., Mullins,J.I., and Allen,T.M. (2010). Viral evolution and escape during acute HIV-1 infection. J. Infect. Dis. *202 Suppl 2*, S309-S314.

Boutwell,C.L., Rowley,C.F., and Essex,M. (2009). Reduced viral replication capacity of human immunodeficiency virus type 1 subtype C caused by cytotoxic-T-lymphocyte escape mutations in HLA-B57 epitopes of capsid protein. J. Virol. *83*, 2460-2468.

Braaten,D., Ansari,H., and Luban,J. (1997). The hydrophobic pocket of cyclophilin is the binding site for the human immunodeficiency virus type 1 Gag polyprotein. Journal of Virology *71*, 2107-2113.

Braaten,D., Franke,E.K., and Luban,J. (1996). Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. Journal of Virology *70*, 3551-3560.

Brandano,L. and Stevenson,M. (2012). A Highly Conserved Residue in the C-Terminal Helix of HIV-1 Matrix Is Required for Envelope Incorporation into Virus Particles. Journal of Virology *86*, 2347-2359.

Bredell,H., Martin,D.P., Van Harmelen,J., Varsani,A., Sheppard,H.W., Donovan,R., Gray,C.M., and Williamson,C. (2007). HIV type 1 subtype C gag and nef diversity in southern Africa. Aids Research and Human Retroviruses *23*, 477-481.

Brenchley,J.M. *et al.* (2004). CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. J. Exp. Med. *200*, 749-759.

Bresnahan,P.A., Yonemoto,W., Ferrell,S., Williams-Herman,D., Geleziunas,R., and Greene,W.C. (1998). A dileucine motif in HIV-1 Nef acts as an internalization signal for CD4 downregulation and binds the AP-1 clathrin adaptor. Curr. Biol. *8*, 1235-1238.

Briggs,J.A.G., Simon,M.N., Gross,I., Krausslich,H.G., Fuller,S.D., Vogt,V.M., and Johnson,M.C. (2004). The stoichiometry of Gag protein in HIV-1. Nature Structural & Molecular Biology *11*, 672-675.

Briggs,J.A.G., Wilk,T., Welker,R., Krausslich,H.G., and Fuller,S.D. (2003). Structural organization of authentic, mature HIV-1 virions and cores. Embo Journal *22*, 1707-1715.

Briones,M.S., Dobard,C.W., and Chow,S.A. (2010). Role of human immunodeficiency virus type 1 integrase in uncoating of the viral core. J Virol *84*, 5181-5190.

Brockman,M.A. *et al.* (2010). Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated for in chronic infection. J. Virol. *84*, 11937-11949.

Brockman,M.A. *et al.* (2007). Escape and compensation from early HLA-B57-Mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. Journal of Virology *81*, 12608-12618.

Brumme,Z, et al. (2008). Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. J Virol. ; 82:9216–9227.

Bryant,M. and Ratner,L. (1990). Myristoylation-Dependent Replication and Assembly of Human Immunodeficiency Virus-1. Proceedings of the National Academy of Sciences of the United States of America *87*, 523-527.

Bukovsky,A.A., Dorfman,T., Weimann,A., and Gottlinger,H.G. (1997). Nef association with human immunodeficiency virus type 1 virions and cleavage by the viral protease. J Virol *71*, 1013-1018.

Bukrinsky,M. (2004). A hard way to the nucleus. Mol. Med. *10*, 1-5.

Bukrinsky,M.I., Haggerty,S., Dempsey,M.P., Sharova,N., Adzhubel,A., Spitz,L., Lewis,P., Goldfarb,D., Emerman,M., and Stevenson,M. (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature *365*, 666-669.

Buonaguro,L., Tornesello,M.L., and Buonaguro,F.M. (2007). Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J. Virol. *81*, 10209-10219.

Burton,D.R., Desrosiers,R.C., Doms,R.W., Koff,W.C., Kwong,P.D., Moore,J.P., Nabel,G.J., Sodroski,J., Wilson,I.A., and Wyatt,R.T. (2004). HIV vaccine design and the neutralizing antibody problem. Nat. Immunol. *5*, 233-236.

Butterfield-Gerson,K.L., Scheifele,L.Z., Ryan,E.P., Hopper,A.K., and Parent,L.J. (2006). Importin-beta family members mediate alpharetrovirus gag nuclear entry via interactions with matrix and nucleocapsid. J. Virol. *80*, 1798-1806.

Campos-Olivas,R., Newman,J.L., and Summers,M.F. (2000). Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. Journal of Molecular Biology *296*, 633-649.

Cao,J., McNevin,J., Holte,S., Fink,L., Corey,L., and McElrath,M.J. (2003). Comprehensive analysis of human immunodeficiency virus type 1 (HIV-1)-specific gamma interferon-secreting CD8+ T cells in primary HIV-1 infection. J. Virol. *77*, 6867-6878.

Cardinaud,S. *et al.* (2011). CTL escape mediated by proteasomal destruction of an HIV-1 cryptic epitope. PLoS. Pathog. *7*, e1002049.

Carlson,J.M. *et al.* (2008a). Phylogenetic Dependency Networks: Inferring Patterns of CTL Escape and Codon Covariation in HIV-1 Gag. Plos Computational Biology *4*.

Carlson,J.M. *et al.* (2012). Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. J. Virol. *86*, 5230-5243.

Carlson,L.A., Briggs,J.A.G., Glass,B., Riches,J.D., Simon,M.N., Johnson,M.C., Muller,B., Grunewald,K., and Krausslich,H.G. (2008b). Three-Dimensional Analysis of Budding Sites and Released Virus Suggests a Revised Model for HIV-1 Morphogenesis. Cell Host & Microbe *4*, 592-599.

Carrington,M., Martin,M.P., and van,B.J. (2008). KIR-HLA intercourse in HIV disease. Trends. Microbiol. *16*, 620-627.

Carrington,M., Nelson,G.W., Martin,M.P., Kissner,T., Vlahov,D., Goedert,J.J., Kaslow,R., Buchbinder,S., Hoots,K., and O'Brien,S.J. (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science. *283*, 1748-1752.

Carrington,M. and O'Brien,S.J. (2003). The influence of HLA genotype on AIDS. Annu. Rev Med. *54*, 535-551.

Cascio,P., Hilton,C., Kisselev,A.F., Rock,K.L., and Goldberg,A.L. (2001). 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. Embo Journal *20*, 2357-2366.

Casella,C.R., Raffini,L.J., and Panganiban,A.T. (1997). Pleiotropic mutations in the HIV-1 matrix protein that affect diverse steps in replication. Virology *228*, 294-306.

Catano,G., Kulkarni,H., He,W., et al. (2008). HIV-1 disease-influencing effects associated with ZNRD1, HCP5 and HLA-C alleles are attributable mainly to either HLA-A10 or HLA-B*57 alleles. PLoS One; 3(11):e3636.

Cecilia,D., Kleeberger,C., Munoz,A., Giorgi,J.V., and Zolla-Pazner,S. (1999). A longitudinal study of neutralizing antibodies and disease progression in HIV-1-infected subjects. Journal of Infectious Diseases *179*, 1365-1374.

John,C., Cleveland,W., Kleiner,B. and Tukey,P. (1983). Graphical Methods for Data Analysis, Wadsworth.

Chang,Y.F., Wang,S.M., Huang,K.J., and Wang,C.T. (2007). Mutations in capsid major homology region affect assembly and membrane affinity of HIV-1 gag. Journal of Molecular Biology *370*, 585-597.

Chazal,N., Carriere,H.B., Gay,B. & Boulanger,P. (1994). Phenotypic characterization of insertion mutants of the human immunodeficiency virus type 1 *gag* precursor expressed in recombinant bactflovirus-infected cells. *Journal of Virology* 68, 111-122.

Chazal,N., Gay,B. Carriere,H.B., Tournier,J., and Boulanger,P. (1995). Human immunodeficiency virus type 1 MA deletion mutants expressed in baculovirus-infected cells: cis and trans effects on Gag precursor assembly pathway. *Journal of Virology* 69, 365-375.

Cheng-Mayer,C., Iannello,P., Shaw,K., Luciw,P.A., and Levy,J.A. (1989). Differential effects of nef on HIV replication: implications for viral pathogenesis in the host. Science *246*, 1629-1632.

Chien,A.I., Liao,W.H., Yang,D.M., and Wang,C.T. (2006). A domain directly C-terminal to the major homology region of human immunodeficiency type 1 capsid protein plays a crucial role in directing both virus assembly and incorporation of Gag-Pol. Virology *348*, 84-95.

Chopera,D.R., Mlotshwa,M., Woodman,Z., Mlisana,K., Rosa,D.D., Martin,D.P., Karim,S.A., Gray,C.M., and Williamson,C. (2011). Virological and Immunological Factors Associated with HIV-1 Differential Disease Progression in HLA-B*58:01-Positive Individuals. Journal of Virology *85*, 7070-7080.

Chopera,D.R. *et al.* (2008a). Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. Virology *4*, e1000033.

Chun,T.W., Engel,D., Berrey,M.M., Shea,T., Corey,L., and Fauci,A.S. (1998). Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. *95*, 8869-8873.

Cleveland,W.S. and Devlin,S.J. (1988). Locally Weighted Regression - An Approach to Regression-Analysis by Local Fitting. Journal of the American Statistical Association 83, 596-610.

Clish,C.B., Peyton,D.H., and Barklis,E. (1996). Spectroscopic study of an HIV-1 capsid protein major homology region peptide analog. FEBS Lett. *378*, 43-47.

Cohen,G.B., Gandhi,R.T., Davis,D.M., Mandelboim,O., Chen,B.K., Strominger,J.L., and Baltimore,D. (1999). The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. Immunity *10*, 661-671.

Cohen,M.S., Shaw,G.M., McMichael,A.J., and Haynes,B.F. (2011). Acute HIV-1 Infection. N. Engl. J. Med. *364*, 1943-1954.

Coiras,M., Lopez-Huertas,M.R., Perez-Olmeda,M., and Alcami,J. (2009). Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. Nat. Rev. Microbiol. *7*, 798-812.

Collins,K.L., Chen,B.K., Kalams,S.A., Walker,B.D., and Baltimore,D. (1998). HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. Nature. *391*, 397-401.

Cosson,P., Demolliere,C., Hennecke,S., Duden,R., and Letourneur,F. (1996). Delta- and zeta-COP, two coatomer subunits homologous to clathrin-associated proteins, are involved in ER retrieval. EMBO J *15*, 1792-1798.

Couillin,I., Culmann-Penciolelli,B., Gomard,E., Choppin,J., Levy,J.P., Guillet,J.G., and Saragosti,S. (1994). Impaired cytotoxic T lymphocyte recognition due to genetic

variations in the main immunogenic region of the human immunodeficiency virus 1 NEF protein. J. Exp. Med. *180*, 1129-1134.

Cowan,S., Hatziioannou,T., Cunningham,T., Muesing,M.A., Gottlinger,H.G., and Bieniasz,P.D. (2002). Cellular inhibitors with Fv1-like activity restrict human and simian immunodeficiency virus tropism. Proc. Natl. Acad. Sci. U. S. A *99*, 11914-11919.

Craig,H.M., Pandori,M.W., and Guatelli,J.C. (1998). Interaction of HIV-1 Nef with the cellular dileucine-based sorting pathway is required for CD4 down-regulation and optimal viral infectivity. Proc. Natl. Acad. Sci. U. S. A *95*, 11229-11234.

Crawford,H. *et al.* (2009). Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. J. Exp. Med. *206*, 909-921.

Crawford,H. *et al.* (2011). The hypervariable HIV-1 capsid protein residues comprise HLA-driven CD8+ T-cell escape mutations and covarying HLA-independent polymorphisms. J. Virol. *85*, 1384-1390.

Crawford,H. *et al.* (2007). Compensatory mutation partially restores fitness and delays reversion of escpe mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. Journal of Virology *81*, 8346-8351.

Crump,C.M., Xiang,Y., Thomas,L., Gu,F., Austin,C., Tooze,S.A., and Thomas,G. (2001). PACS-1 binding to adaptors is required for acidic cluster motif-mediated protein traffic. EMBO J *20*, 2191-2201.

Daar,E.S. (1998). Virology and immunology of acute HIV type 1 infection. AIDS Research and Human Retroviruses Supplementary 3 14, S229-S234.

Daar,E.S., Moudgil,T., Meyer,R.D., and Ho,D.D. (1991). Transient High-Levels of Viremia in Patients with Primary Human-Immunodeficiency-Virus Type-1 Infection. New England Journal of Medicine *324*, 961-964.

Dahirel,V. *et al.* (2011). Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proc. Natl. Acad. Sci. U. S. A. *108*, 11530-11535.

Dai,L. and Stevenson,M. (2010). A novel motif in HIV-1 Nef that regulates MIP-1beta chemokine release in macrophages. J Virol *84*, 8327-8331.

Dalgleish,A.G., Beverley,P.C., Clapham,P.R., Crawford,D.H., Greaves,M.F., and Weiss,R.A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature. *312*, 763-767.

Dalod,M. *et al.* (1999). Weak anti-HIV CD8(+) T-cell effector activity in HIV primary infection. Journal of Clinical Investigation *104*, 1431-1439.

Darlix,J.L., Lapadat-Tapolsky,M., de,R.H., and Roques,B.P. (1995). First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. J Mol. Biol. *254*, 523-537.

Davis,M.R., Jiang,J.Y., Zhou,J., Freed,E.O., and Aiken,C. (2006). A mutation in the human immunodeficiency virus type 1 gag protein destabilizes the interaction of the envelope protein subunits gp120 and gp41. Journal of Virology *80*, 2405-2417.

Day,C.L. *et al.* (2006). PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. Nature. *443*, 350-354.

De Guzman,R.N., Wu,Z.R., Stalling,C.C., Pappalardo,L., Borer,P.N., and Summers,M.F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 Psi-RNA recognition element. Science *279*, 384-388.

Deacon,N.J. *et al.* (1995). Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. Science *270*, 988-991.

Dean,M. *et al.* (1996). 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. *273*, 1856-1862.

Deeks,S.G. and Walker,B.D. (2007). Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. Immunity. *27*, 406-416.

Delelis,O., Carayon,K., Saib,A., Deprez,E., and Mouscadet,J.F. (2008). Integrase and integration: biochemical activities of HIV-1 integrase. Retrovirology *5*.

Delport,W., Poon,A.F., Frost,S.D., and Kosakovsky Pond,S.L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. *26*, 2455-2457.

Delport,W., Scheffler,K., and Seoighe,C. (2008). Frequent toggling between alternative amino acids is driven by selection in HIV-1. PLoS Pathog. *4*, e1000242.

Dennis,C.A., Baron,A., Grossmann,J.G., Mazaleyrat,S., Harris,M., and Jaeger,J. (2005). Co-translational myristoylation alters the quaternary structure of HIV-1 Nef in solution. Proteins *60*, 658-669.

Derdowski,A., Ding,L.M., and Spearman,P. (2004). A novel fluorescence resonance energy transfer assay demonstrates that the human immunodeficiency virus type 1 Pr55(Gag)I domain mediates Gag-Gag interactions. Journal of Virology *78*, 1230-1242.

Desrosiers,R.C. (1999). Strategies used by human immunodeficiency virus that allow persistent viral replication. Nat. Med. *5*, 723-725.

Doms,R.W. (2000). Beyond receptor expression: The influence of receptor conformation, density, and affinity in HIV-1 infection. Virology *276*, 229-237.

Dong,X. *et al.* (2005). AP-3 directs the intracellular trafficking of HIV-1 Gag and plays a key role in particle assembly. Cell *120*, 663-674.

Dorfman,T., Mammano,F., Haseltine,W.A., and Gottlinger,H.G. (1994). Role of the Matrix Protein in the Virion Association of the Human-Immunodeficiency-Virus Type-1 Envelope Glycoprotein. Journal of Virology *68*, 1689-1696.

Douek,D.C. (2003). Disrupting T-cell homeostasis: how HIV-1 infection causes disease. AIDS Rev. *5*, 172-177.

Douek,D.C. *et al.* (2002). HIV preferentially infects HIV-specific CD4+ T cells. Nature. *417*, 95-98.

Draenert,R. *et al.* (2004). Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. Journal of Experimental Medicine *199*, 905-915.

Dupont,S., Sharova,N., DeHoratius,C., Virbasius,C.M.A., Zhu,X.C., Bukrinskaya,A.G., Stevenson,M., and Green,M.R. (1999). A novel nuclear export activity in HIV-1 matrix protein required for viral replication. Nature *402*, 681-685.

Dyer,J.R. *et al.* (1998). High levels of human immunodeficiency virus type 1 in blood and semen of seropositive men in sub-Saharan Africa. Journal of Infectious Diseases *177*, 1742-1746.

Dykes,C., and Demeter,L.M. (2007). Clinical significance of human immunodeficiency virus type 1 replication fitness. Clin. Microbiol. Rev. 20:550–578.

Edwards,B.H., Bansal,A., Sabbaj,S., Bakari,J., Mulligan,M.J., and Goepfert,P.A. (2002). Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. J. Virol. *76*, 2298-2305.

Emu,B., Sinclair,E., Hatano,H., Ferre,A., Shacklett,B., Martin,J.N., McCune,J.M., and Deeks,S.G. (2008). HLA class I-restricted T-cell responses may contribute to the control

of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. J. Virol. *82*, 5398-5407.

Endrich,M.M., Gehrig,P., and Gehring,H. (1999). Maturation-induced conformational changes of HIV-1 capsid protein and identification of two high affinity sites for cyclophilins in the C-terminal domain. Journal of Biological Chemistry *274*, 5326-5332.

Fackler,O.T. and Baur,A.S. (2002). Live and let die: Nef functions beyond HIV replication. Immunity. *16*, 493-497.

Feeney,M.E., Tang,Y., Roosevelt,K.A., Leslie,A.J., McIntosh,K., Karthas,N., Walker,B.D., and Goulder,P.J. (2004). Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child. J. Virol. *78*, 8927-8930.

Fellay,J. *et al.* (2007). A whole-genome association study of major determinants for host control of HIV-1. Science *317*, 944-947.

Fiebig,E.W. *et al.* (2003). Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS *17*, 1871-1879.

Fiorentini,S., Giagulli,C., Caccuri,F., Magiera,A.K., and Caruso,A. (2010). HIV-1 matrix protein p17: a candidate antigen for therapeutic vaccines against AIDS. Pharmacol. Ther. *128*, 433-444.

Fischer,G., Berger,E., and Bang,H. (1989). Kinetic Beta-Deuterium Isotope Effects Suggest A Covalent Mechanism for the Protein Folding Enzyme Peptidylprolyl Cis Trans-Isomerase. Febs Letters *250*, 267-270.

Fisher,R.D., Chung,H.Y., Zhai,Q., Robinson,H., Sundquist,W.I., and Hill,C.P. (2007). Structural and biochemical studies of ALIX/AIP1 and its role in retrovirus budding. Cell *128*, 841-852.

Fitzon,T., Leschonsky,B., Bieler,K., Paulus,C., Schroder,J., Wolf,H., and Wagner,R. (2000). Proline residues in the HIV-1NH(2)-terminal capsid domain: Structure determinants for proper core assembly and subsequent steps of early replication. Virology *268*, 294-307.

Forshey,B.M. and Aiken,C. (2003). Disassembly of human immunodeficiency virus type 1 cores in vitro reveals association of Nef with the subviral ribonucleoprotein complex. J Virol *77*, 4409-4414.

Franco,A., Ferrari,C., Sette,A., and Chisari,F.V. (1995). Viral mutations, TCR antagonism and escape from the immune response. Curr. Opin. Immunol. *7*, 524-531.

Frater,A.J. *et al.* (2007). Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression. J. Virol. *81*, 6742-6751.

Freed, E.O. and Martin,M.A. Retroviridae: the retroviruses and their replication. Peter M.Howley, MD, Diane E.Griffin, MD PhD, Robert A.Lamb, PhD ScD, Malcolm A.Martin, MD, Bernard Roizman, ScD, and Stephen E.Straus, MD. Fields Virology II. 2007. Lippincott William and Wilkins, Philadelphia, USA. Fields virology.

Freed,E.O. (1998). HIV-1 Gag proteins: Diverse functions in the virus life cycle. Virology *251*, 1-15.

Freed,E.O. (2001). HIV-1 replication.Somatic Cell and Molecular Genetics, 26: 1-6.

Freed,E.O. (2002). Viral late domains. J Virol *76*, 4679-4687.

Freed,E.O., Orenstein,J.M., Buckler-White,A.J., and Martin,M.A. (1994). Single amino acid changes in the human immunodeficiency virus type 1 matrix protein block virus particle production. J. Virol. *68*, 5311-5320.

Freel,S.A. *et al.* (2012). Initial HIV-1 Antigen-Specific CD8+ T Cells in Acute HIV-1 Infection Inhibit Transmitted/Founder Virus Replication. J. Virol. *86*, 6835-6846.

Friedrich,T.C. *et al.* (2004). Reversion of CTL escape-variant immunodeficiency viruses in vivo. Nature Medicine *10*, 275-281.

Friend,C. (1957). Cell-free transmission in adult Swiss mice of a disease having the character of a leukemia. J. Exp. Med. *105*, 307-318.

Frost,S.D. *et al.* (2005). Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc. Natl. Acad. Sci. U. S. A *102*, 18514-18519.

Gallo,R., Wongstaal,F., Montagnier,L., Haseltine,W.A., and Yoshida,M. (1988). Hiv/Htlv Gene Nomenclature. Nature *333*, 504.

Gamble,T.R., Vajdos,F.F., Yoo,S.H., Worthylake,D.K., Houseweart,M., Sundquist,W.I., and Hill,C.P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell *87*, 1285-1294.

Gamble,T.R., Yoo,S.H., Vajdos,F.F., vonSchwedler,U.K., Worthylake,D.K., Wang,H., McCutcheon,J.P., Sundquist,W.I., and Hill,C.P. (1997). Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science *278*, 849-853.

Ganser-Pornillos,B.K., Cheng,A., and Yeager,M. (2007). Structure of full-length HIV-1CA: a model for the mature capsid lattice. Cell *131*, 70-79.

Ganser-Pornillos,B.K., Yeager,M., and Sundquist,W.I. (2008). The structural biology of HIV assembly. Current Opinion in Structural Biology *18*, 203-217.

Ganusov,V.V., Goonetilleke,N., Liu,M.K., Ferrari,G., Shaw,G.M., McMichael,A.J., Borrow,P., Korber,B.T., and Perelson,A.S. (2011). Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. J. Virol. *85*, 10518-10528.

Garrus,J.E. *et al.* (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell *107*, 55-65.

Gea-Banacloche,J.C. *et al.* (2000). Maintenance of large numbers of virus-specific CD8+ T cells in HIV-infected progressors and long-term nonprogressors. J. Immunol. *165*, 1082-1092.

Gelderblom,H.R. (1991). Assembly and morphology of HIV: potential effect of structure on viral function. AIDS *5*, 617-637.

Geldmacher,C. *et al.* (2009). Minor viral and host genetic polymorphisms can dramatically impact the biologic outcome of an epitope-specific CD8 T-cell response. Blood. *114*, 1553-1562.

Geyer,M., Yu,H., Mandic,R., Linnemann,T., Zheng,Y.H., Fackler,O.T., and Peterlin,B.M. (2002). Subunit H of the V-ATPase binds to the medium chain of adaptor protein complex 2 and connects Nef to the endocytic machinery. J Biol. Chem. *277*, 28521-28529.

Gitti,R.K., Lee,B.M., Walker,J., Summers,M.F., Yoo,S., and Sundquist,W.I. (1996). Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science *273*, 231-235.

Gloster,S.E., Newton,P., Cornforth,D., Lifson,J.D., Williams,I., Shaw,G.M., and Borrow,P. (2004). Association of strong virus-specific CD4 T cell responses with efficient natural control of primary HIV-1 infection. AIDS. *18*, 749-755.

Goepfert,P.A. *et al.* (2008). Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J. Exp. Med. *205*, 1009-1017.

Goff,S.P. Retroviridae: the retroviruses and their replication. 2007a. Lippincott William and Wilkins, Philadelphia, PA, USA.

Goff,S.P. (2007b). Host factors exploited by retroviruses. Nat. Rev. Microbiol. *5*, 253-263.

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

Goonetilleke,N. *et al.* (2009). The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J. Exp. Med. *206*, 1253-1272.

Gorelick,R.J., Henderson,L.E., Hanser,J.P., and Rein,A. (1988). Point Mutants of Moloney Murine Leukemia-Virus That Fail to Package Viral-Rna - Evidence for Specific Rna Recognition by A Zinc Finger-Like Protein-Sequence. Proceedings of the National Academy of Sciences of the United States of America *85*, 8420-8424.

Gorelick,R.J., Nigida,S.M., Bess,J.W., Arthur,L.O., Henderson,L.E., and Rein,A. (1990). Noninfectious Human-Immunodeficiency-Virus Type-1 Mutants Deficient in Genomic Rna. Journal of Virology *64*, 3207-3211.

Gottlinger,H.G., Sodroski,J.G., and Haseltine,W.A. (1989). Role of Capsid Precursor Processing and Myristoylation in Morphogenesis and Infectivity of Human Immunodeficiency Virus Type-1. Proceedings of the National Academy of Sciences of the United States of America *86*, 5781-5785.

Gould,S.J., Booth,A.M., and Hildreth,J.E.K. (2003). The Trojan exosome hypothesis. Proceedings of the National Academy of Sciences of the United States of America *100*, 10592-10597.

Goulder,P.J. *et al.* (2000). Differential narrow focusing of immunodominant human immunodeficiency virus gag-specific cytotoxic T-lymphocyte responses in infected African and caucasoid adults and children. J. Virol. *74*, 5679-5690.

Goulder,P.J. *et al.* (1997). Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat. Med. *3*, 212-217.

Goulder,P.J. and Watkins,D.I. (2004). HIV and SIV CTL escape: implications for vaccine design. Nat. Rev. Immunol. *4*, 630-640.

Goulder,P.J. and Watkins,D.I. (2008). Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat. Rev. Immunol. *8*, 619-630.

Goulder,P.J. *et al.* (2001). Evolution and transmission of stable CTL escape mutations in HIV infection. Nature *412*, 334-338.

Granich,R.M., Gilks,C.F., Dye,C., De Cock,K.M., and Williams,B.G. (2009). Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet *373*, 48-57.

Gray,C.M. *et al.* (2009a). Human Immunodeficiency Virus-Specific Gamma Interferon Enzyme-Linked Immunospot Assay Responses Targeting Specific Regions of the Proteome during Primary Subtype C Infection Are Poor Predictors of the Course of Viremia and Set Point. Journal of Virology *83*, 470-478.

Gray,E.S. *et al.* (2011). The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. J. Virol. *85*, 4828-4840.

Gray,E.S., Madiga,M.C., Moore,P.L., Mlisana,K., Karim,S.S.A., Binley,J.M., Shaw,G.M., Mascola,J.R., and Morris,L. (2009b). Broad Neutralization of Human Immunodeficiency Virus Type 1 Mediated by Plasma Antibodies against the gp41 Membrane Proximal External Region. Journal of Virology *83*, 11265-11274.

Gray,E.S. *et al.* (2007). Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. J. Virol. *81*, 6187-6196.

Greenberg,M., DeTulleo,L., Rapoport,I., Skowronski,J., and Kirchhausen,T. (1998). A dileucine motif in HIV-1 Nef is essential for sorting into clathrin-coated pits and for downregulation of CD4. Curr. Biol. *8*, 1239-1242.

Greene,W.C. and Peterlin,B.M. (2002). Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nature Medicine *8*, 673-680.

Guadalupe,M., Reay,E., Sankaran,S., Prindiville,T., Flamm,J., McNeil,A., and Dandekar,S. (2003). Severe CD4+ T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy. J. Virol. *77*, 11708-11717.

Haaland,R.E. *et al.* (2009). Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS. Pathog. *5*, e1000274.

Haase,A.T. (2010). Targeting early infection to prevent HIV-1 mucosal transmission. Nature. *464*, 217-223.

Haffar,O.K., Popov,S., Dubrovsky,L., Agostini,I., Tang,H., Pushkarsky,T., Nadler,S.G., and Bukrinsky,M. (2000). Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. J. Mol. Biol. *299*, 359-368.

Harrich,D. and Hooker,B. (2002). Mechanistic aspects of HIV-1 reverse transcription initiation. Reviews in Medical Virology *12*, 31-45.

Harris,M.P. and Neil,J.C. (1994). Myristoylation-dependent binding of HIV-1 Nef to CD4. J Mol. Biol. *241*, 136-142.

Hatziioannou,T., Cowan,S., von Schwedler,U.K., Sundquist,W.I., and Bieniasz,P.D. (2004). Species-specific tropism determinants in the human immunodeficiency virus type 1 capsid. Journal of Virology *78*, 6005-6012.

Hatziioannou,T., Perez-Caballero,D., Cowan,S., and Bieniasz,P.D. (2005). Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells. Journal of Virology *79*, 176-183.

Haynes,B.F. *et al.* (2012). Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial. New England Journal of Medicine *366*, 1275-1286.

Hecht,F.M., Hartogensis,W., Bragg,L., Bacchetti,P., Atchison,R., Grant,R., Barbour,J., and Deeks,S.G. (2010). HIV RNA level in early infection is predicted by viral load in the transmission source. AIDS. *24*, 941-945.

Hel,Z., McGhee,J.R., and Mestecky,J. (2006). HIV infection: first battle decides the war. Trends. Immunol. *27*, 274-281.

Heller,K.N., Upshaw,J., Seyoum,B., Zebroski,H., and Munz,C. (2007). Distinct memory CD4+ T-cell subsets mediate immune recognition of Epstein Barr virus nuclear antigen 1 in healthy virus carriers. Blood. *109*, 1138-1146.

Hemelaar,J., Gouws,E., Ghys,P.D., and Osmanov,S. (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS *25*, 679-689.

Henn,M.R. *et al.* (2012). Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS. Pathog. *8*, e1002529.

Hewitt,E.W. (2003). The MHC class I antigen presentation pathway: strategies for viral immune evasion. Immunology *110*, 163-169.

Hilditch,L., Matadeen,R., Goldstone,D.C., Rosenthal,P.B., Taylor,I.A., and Stoye,J.P. (2011). Ordered assembly of murine leukemia virus capsid protein on lipid nanotubes directs specific binding by the restriction factor, Fv1. Proc. Natl. Acad. Sci. U. S. A *108*, 5771-5776.

Hill,C.P., Worthylake,D., Bancroft,D.P., Christensen,A.M., and Sundquist,W.I. (1996). Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. Proceedings of the National Academy of Sciences of the United States of America *93*, 3099-3104.

Ho,D.D., Neumann,A.U., Perelson,A.S., Chen,W., Leonard,J.M., and Markowitz,M. (1995). Rapid Turnover of Plasma Virions and Cd4 Lymphocytes in Hiv-1 Infection. Nature *373*, 123-126.

Hofmann,W., Schubert,D., LaBonte,J., Munson,L., Gibson,S., Scammell,J., Ferrigno,P., and Sodroski,J. (1999). Species-specific, postentry barriers to primate immunodeficiency virus infection. J. Virol. *73*, 10020-10028.

Hogan,C.M. and Hammer,S.M. (2001). Host determinants in HIV infection and disease part 1: Cellular and humoral immune responses. Annals of Internal Medicine *134*, 761-776.

Hollingsworth,T.D., Anderson,R.M., and Fraser,C. (2008). HIV-1 transmission, by stage of infection. J. Infect. Dis. *198*, 687-693.
(Hecht, 2010 148 /id)
Hollingsworth,T.D. *et al.* (2010). HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. PLoS. Pathog. *6*, e1000876.

Honeyborne,I. *et al.* (2007). Control of human immunodeficiency virus type 1 is associated with HLA-B*13 and targeting of multiple gag-specific CD8(+) T-cell epitopes. Journal of Virology *81*, 3667-3672.

Hoof,I., Peters,B., Sidney,J., Pedersen,L.E., Sette,A., Lund,O., Buus,S., and Nielsen,M. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics *61*, 1-13.

74, 6087-6095.Huang,K.H. et al. (2011). Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations. PLoS. One. 6, e19018.

Hung,C.H. *et al.* (2007). HIV-1 Nef assembles a Src family kinase-ZAP-70/Syk-PI3K cascade to downregulate cell-surface MHC-I. Cell Host. Microbe *1*, 121-133.

Hunter,E. (1994). Macromolecular Interactions in the Assembly of Hiv and Other Retroviruses. Seminars in Virology *5*, 71-83.

Hurley,J.H. (2008). ESCRT complexes and the biogenesis of multivesicular bodies. Curr. Opin. Cell Biol. *20*, 4-11.

Hussein,M., Abebe,A., Pollakis,G., Brouwer,M., Petros,B., Fontanet,A.L., and De Wit,T.F.R. (2000). HIV-1 subtype C in commercial sex workers in Addis Ababa, Ethiopia. Journal of Acquired Immune Deficiency Syndromes *23*, 120-127.

Iglewicz,B. and Hoaglin,D.C. (1993). How to Detect and Handle Outliers, American Society for Quality Control, Milwaukee, WI

Janssen,E.M., Droin,N.M., Lemmens,E.E., Pinkoski,M.J., Bensinger,S.J., Ehst,B.D., Griffith,T.S., Green,D.R., and Schoenberger,S.P. (2005). CD4+ T-cell help controls CD8+ T-cell memory via TRAIL-mediated activation-induced cell death. Nature. *434*, 88-93.

Janssens,W. *et al.* (1994). Genetic and Phylogenetic Analysis of Env Subtype-G and Subtype-H in Central-Africa. Aids Research and Human Retroviruses *10*, 877-879.

Janvier,K., Pelchen-Matthews,A., Renaud,J.B., Caillet,M., Marsh,M., and Berlioz-Torrent,C. (2011). The ESCRT-0 component HRS is required for HIV-1 Vpu-mediated BST-2/tetherin down-regulation. PLoS. Pathog. *7*, e1001265.

Jia,M. *et al.* (2012). Preferential CTL targeting of Gag is associated with relative viral control in long-term surviving HIV-1 infected former plasma donors from China. Cell. Res. *22*, 903-914.

et al. (1999). Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques (vol 189, pg 991, 1999). Journal of Experimental Medicine *189*, 1999.

Jia,X., Singh,R., Homann,S., Yang,H., Guatelli,J., and Xiong,Y. (2012). Structural basis of evasion of cellular adaptive immunity by HIV-1 Nef. Nat Struct. Mol. Biol. 19, 701-706.

Jouvenet,N., Neil,S.J.. Bess,C., Johnson, M.C., Virgen,C.A., Simon,S.M., and Bieniasz P.D. (2006). Plasma membrane is the site of productive HIV-1 particle assembly. PLoS Biol., 4, e435

Kaplan,A.H., Manchester,M., and Swanstrom,R. (1994). The Activity of the Protease of Human-Immunodeficiency-Virus Type-1 Is Initiated at the Membrane of Infected-Cells Before the Release of Viral-Proteins and Is Required for Release to Occur with Maximum Efficiency. Journal of Virology *68*, 6782-6786.

Kaslow,R.A. *et al.* (1990). A1, Cw7, B8, DR3 HLA antigen combination associated with rapid decline of T-helper lymphocytes in HIV-1 infection. A report from the Multicenter AIDS Cohort Study. Lancet. *335*, 927-930.

Kassutto,S. and Rosenberg,E.S. (2004). Primary HIV type 1 infection. Clin. Infect. Dis. *38*, 1447-1453.

Kaufmann,D.E. *et al.* (2007). Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. Nat Immunol. *8*, 1246-1254.

Kaufmann,G.R., Cunningham,P., Kelleher,A.D., Zaunders,J., Carr,A., Vizzard,J., Law,M., and Cooper,D.A. (1998). Patterns of viral dynamics during primary human immunodeficiency virus type 1 infection. Journal of Infectious Diseases *178*, 1812-1815.

Kaumanns,P., Hagmann,I., and Dittmar,M.T. (2006). Human TRIM5alpha mediated restriction of different HIV-1 subtypes and Lv2 sensitive and insensitive HIV-2 variants. Retrovirology. *3*, 79.

Kawashima,Y. *et al.* (2009). Adaptation of HIV-1 to human leukocyte antigen class I. Nature *458*, 641-U108.

Keckesova,Z.F., Ylinen,L.M., Towers,G., and Towers,G.J. (2006). Cyclophilin A renders human immunodeficiency virus type 1 sensitive to Old World monkey but not human TRIM5 alpha antiviral activity. J. Virol. *80 (10)*, 4683-90.

Keele,B.F. *et al.* (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. *105*, 7552-7557.

Keele,B.F. *et al.* (2009). Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. J. Exp. Med. *206*, 1117-1134.

Keene,J.A. and Forman,J. (1982). Helper activity is required for the in vivo generation of cytotoxic T lymphocytes. J. Exp. Med. *155*, 768-782.

Kelleher,A.D. *et al.* (2001). Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J. Exp. Med. *193*, 375-386.

Kelley,C.F., Barbour,J.D., and Hecht,F.M. (2007). The relation between symptoms, viral load, and viral load set point in primary HIV infection. J. Acquir. Immune. Defic. Syndr. *45*, 445-448.

Kelly,B.N., Howard,B.R., Wang,H., Robinson,H., Sundquist,W.I., and Hill,C.P. (2006). Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1-278) and CA(N)(133-278). Biochemistry *45*, 11257-11266.

Kestler,H.W., Ringler,D.J., Mori,K., Panicali,D.L., Sehgal,P.K., Daniel,M.D., and Desrosiers,R.C. (1991). Importance of the nef gene for maintenance of high virus loads and for development of AIDS. Cell *65*, 651-662.

Keulen,W., Nijhuis,M., Schuurman,R., Berkhout,B., and Boucher,C. (1997). Reverse transcriptase fidelity and HIV-1 variation. Science *275*, 229.

Khan,I.H., Sawai,E.T., Antonio,E., Weber,C.J., Mandell,C.P., Montbriand,P., and Luciw,P.A. (1998). Role of the SH3-ligand domain of simian immunodeficiency virus Nef in interaction with Nef-associated kinase and simian AIDS in rhesus macaques. J Virol *72*, 5820-5830.

Kiepiela,P. *et al.* (2004). Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature *432*, 769-774.

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

Kim,G.J., Lee,H.S., Hong,K.J., and Kim,S.S. (2010). Dynamic correlation between CTL response and viral load in primary human immunodeficiency virus-1 infected Koreans. Virol. J. *7*, 239.

Kimura,M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111-120.

Kirchhoff,F., Greenough,T.C., Brettler,D.B., Sullivan,J.L., and Desrosiers,R.C. (1995). Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. N. Engl. J Med. *332*, 228-232.

Kleiman,L. and Cen,S. (2004). The tRNALys packaging complex in HIV-1. International Journal of Biochemistry & Cell Biology *36*, 1776-1786.

Kleiman,L., Halwani,R., and Javanbakht,H. (2004). The selective packaging and annealing of primer tRNA(Lys3) in HIV-1. Current Hiv Research *2*, 163-175.

Klein,M.R., van Baalen,C.A., Holwerda,A.M., Kerkhof Garde,S.R., Bende,R.J., Keet,I.P., Eeftinck-Schattenkerk,J.K., Osterhaus,A.D., Schuitemaker,H., and Miedema,F. (1995). Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. J. Exp. Med. *181*, 1365-1372.

Kleinman,S.H., Lelie,N., and Busch,M.P. (2009). Infectivity of human immunodeficiency virus-1, hepatitis C virus, and hepatitis B virus and risk of transmission by transfusion. Transfusion *49*, 2454-2489.

Klenerman,P. *et al.* (1994). Cytotoxic T-Cell Activity Antagonized by Naturally-Occurring Hiv-1 Gag Variants. Nature *369*, 403-407.

Kosakovsky Pond,S.L., Posada,D., Gravenor,M.B., Woelk,C.H., and Frost,S.D. (2006). GARD: a genetic algorithm for recombination detection. Bioinformatics. *22*, 3096-3098.

Kotov,A., Zhou,J., Flicker,P., and Aiken,C. (1999). Association of Nef with the human immunodeficiency virus type 1 core. J Virol *73*, 8824-8830.

Koup,R.A., Graham,B.S., and Douek,D.C. (2011). The quest for a T cell-based immune correlate of protection against HIV: a story of trials and errors. Nature Reviews Immunology *11*, 65-70.

Koup,R.A., Safrit,J.T., Cao,Y.Z., Andrews,C.A., Mcleod,G., Borkowsky,W., Farthing,C., and Ho,D.D. (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*, 4650-4655.

Kraft,C.S., Basu,D., Hawkins,P.A., Hraber,P.T., Chomba,E., Mulenga, J., Kilembe,W., Khu,N.H., Derdeyn,C.A., Allen,S.A., *et al.* (2012). Timing and source of subtype-C HIV-1 superinfection in the newly infected partner of Zambian couples with disparate viruses. Retrovirology, 9:22.

Kuiken,C., Foley,B., Hahn,B.H., McCutchan,F.E., Mellors,J.W., Mullins,J., Sodroski,J., Wolinsky,S. and Korber,B. Human retroviruses and AIDS 2000: a compilation and analysis of nucleic acids and amino acid sequences. 2000. Los Alamos National Laboratory, Theoretical Biology and Biophysics Group.

Kuroda,M.J., Schmitz,J.E., Charini,W.A., Nickerson,C.E., Lifton,M.A., Lord,C.I., Forman,M.A., and Letvin,N.L. (1999). Emergence of CTL coincides with clearance of virus during primary simian immunodeficiency virus infection in rhesus monkeys. Journal of Immunology *162*, 5127-5133.

Lal,R.B., Chakrabarti,S., and Yang,C. (2005). Impact of genetic diversity of HIV-1 on diagnosis, antiretroviral therapy & vaccine development. Indian J. Med. Res. *121*, 287-314.

Lambotte,O., Boufassa,F., Madec,Y., Nguyen,A., Goujard,C., Meyer,L., Rouzioux,C., Venet,A., and Delfraissy,J.F. (2005). HIV controllers: a homogeneous group of HIV-1-infected patients with spontaneous control of viral replication. Clin. Infect. Dis. *41*, 1053-1056.

Larsen,J.E., Massol,R.H., Nieland,T.J.F., and Kirchhausen,T. (2004). HIVNEF-mediated major histocompatibility complex class I down-modulation is independent of Arf6 activity. Molecular Biology of the Cell *15*, 323-331.

Lauvau,G., Kakimi,K., Niedermann,G., Ostankovitch,M., Yotnda,P., Firat,H., Chisari,F.V., and van Endert,P.M. (1999). Human transporters associated with antigen processing (TAPs) select epitope precursor peptides for processing in the endoplasmic reticulum and presentation to T cells. Journal of Experimental Medicine *190*, 1227-1239.

Lavallee,C., Yao,X.J., Ladha,A., Gottlinger,H., Haseltine,W.A., and Cohen,E.A. (1994). Requirement of the Pr55gag precursor for incorporation of the Vpr product into human immunodeficiency virus type 1 viral particles. J Virol *68*, 1926-1934.

Le Gall,S., Buseyne,F., Trocha,A., Walker,B.D., Heard,J.M., and Schwartz,O. (2000). Distinct tra trafficking pathways mediate Nef-induced and clathrin-dependent major histocompatibility complex class I down-regulation. Journal of Virology *74*, 9256-9266.

Lee,C.H., Saksela,K., Mirza,U.A., Chait,B.T., and Kuriyan,J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. Cell *85*, 931-942.

Lemey,P., Kosakovsky Pond,S.L., Drummond,A.J., Pybus,O.G., Shapiro,B., Barroso,H., Taveira,N., and Rambaut,A. (2007). Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS. Comput. Biol. *3*, e29.

Leslie,A. *et al.* (2010). Additive Contribution of HLA Class I Alleles in the Immune Control of HIV-1 Infection. Journal of Virology *84*, 9879-9888.

Leslie,A. *et al.* (2006). Differential selection pressure exerted on HIV by CTL targeting identical epitopes but restricted by distinct HLA alleles from the same HLA supertype. (vol 177, pg 4699, 2006). Journal of Immunology *177*, 8878.

Leslie,A.J. *et al.* (2004). HIV evolution: CTL escape mutation and reversion after transmission. Nature Medicine *10*, 282-289.

Letvin,N.L. and Walker,B.D. (2003). Immunopathogenesis and immunotherapy in AIDS virus infections. Nat Med. *9*, 861-866.

Levesque,M.C. *et al.* (2009). Polyclonal B cell differentiation and loss of gastrointestinal tract germinal centers in the earliest stages of HIV-1 infection. PLoS. Med. *6*, e1000107.

Levin,J.G., Guo,J.H., Rouzina,I., and Musier-Forsyth,K. (2005). Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: Critical role in reverse transcription and molecular mechanism. Progress in Nucleic Acid Research and Molecular Biology, Vol 80 *80*, 217-286.

Lewis,P., Hensel,M., and Emerman,M. (1992). Human-Immunodeficiency-Virus Infection of Cells Arrested in the Cell-Cycle. Embo Journal *11*, 3053-3058.

Li,H. *et al.* (2010). High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. PLoS. Pathog. *6*, e1000890.

Li,Q.S., Duan,L.J., Estes,J.D., Ma,Z.M., Rourke,T., Wang,Y.C., Reilly,C., Carlis,J., Miller,C.J., and Haase,A.T. (2005). Peak SIV replication in resting memory CD4(+) T cells depletes gut lamina propria CD4(+) T cells. Nature *434*, 1148-1152.

Li,S., Hill,C.P., Sundquist,W.I., and Finch,J.T. (2000). Image reconstructions of helical assemblies of the HIV-1 CA protein. Nature *407*, 409-413.

Lichterfeld,M. *et al.* (2007). A viral CTL escape mutation leading to immunoglobulin-like transcript 4-mediated functional inhibition of myelomonocytic cells. Journal of Experimental Medicine *204*, 2813-2824.

et al. (2004). HIV-1 Nef is preferentially recognized by CD8 T cells in primary HIV-1 infection despite a relatively high degree of genetic diversity. AIDS. *18*, 1383-1392.

Limou,S., Le Clerc,S., Coulonges,C., et al. (2009). Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). J Infect Dis;199(3):419-426.

Lindback,S., Karlsson,A.C., Mittler,J., Blaxhult,A., Carlsson,M., Briheim,G., Sonnerborg,A., and Gaines,H. (2000). Viral dynamics in primary HIV-1 infection. AIDS *14*, 2283-2291.

Lindwasser,O.W. and Resh,M.D. (2004). Human immunodeficiency virus type 1 Gag contains a dileucine-like motif that regulates association with multivesicular bodies. J. Virol. *78*, 6013-6023.

Little,S.J., McLean,A.R., Spina,C.A., Richman,D.D., and Havlir,D.V. (1999). Viral dynamics of acute HIV-1 infection. J. Exp. Med. *190*, 841-850.

Liu,Y. *et al.* (2007). Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape. J. Virol. *81*, 12179-12188.

Llano,A., Frahm N, and Brander C. How to Optimally Define Optimal Cytotoxic T Lymphocyte Epitopes in HIV Infection. HIV Molecular Immunology , 3-5. 2009. Los Alamos National Laboratory, Theoretical Biology and Biophysics Group.

Loh,L., Petravic,J., Batten,C.J., Davenport,M.P., and Kent,S.J. (2008). Vaccination and timing influence SIV immune escape viral dynamics in vivo. Plos Pathogens *4*.

Loomis-Price,L.D. *et al.* (1998). Correlation between humoral responses to human immunodeficiency virus type 1 envelope and disease progression in early-stage infection. Journal of Infectious Diseases *178*, 1306-1316.

Lu,X., Yu,H., Liu,S.H., Brodsky,F.M., and Peterlin,B.M. (1998). Interactions between HIV1 Nef and vacuolar ATPase facilitate the internalization of CD4. Immunity. *8*, 647-656.

Luban,J. (2007). Cyclophilin A, TRIM5, and resistance to human immunodeficiency virus type 1 infection. Journal of Virology *81*, 1054-1061.

Lubben,N.B., Sahlender,D.A., Motley,A.M., Lehner,P.J., Benaroch,P., and Robinson,M.S. (2007). HIV-1 Nef-induced down-regulation of MHC class I requires AP-1 and clathrin but not PACS-1 and is impeded by AP-2. Mol. Biol. Cell *18*, 3351-3365.

Luciw,P.A., Cheng-Mayer,C., and Levy,J.A. (1987). Mutational analysis of the human immunodeficiency virus: the orf-B region down-regulates virus replication. Proc. Natl. Acad. Sci. U. S. A *84*, 1434-1438.

Luo,M. *et al.* (2012). For protection from HIV-1 infection, more might not be better: a systematic analysis of HIV Gag epitopes of two alleles associated with different outcomes of HIV-1 infection. J. Virol. *86*, 1166-1180.

Lyles,R.H., Munoz,A., Yamashita,T.E., Bazmi,H., Detels,R., Rinaldo,C.R., Margolick,J.B., Phair,J.P., and Mellors,J.W. (2000). Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Journal of Infectious Diseases *181*, 872-880.

MacDonald,K.S. *et al.* (2000). Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. J. Infect. Dis. *181*, 1581-1589.

Magierowska,M., Theodorou,I., Debre,P., Sanson,F., Autran,B., Riviere,Y., Charron,D., and Costagliola,D. (1999). Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. Blood. *93*, 936-941.

Malim,M.H. and Emerman,M. (2008). HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host. Microbe *3*, 388-398.

Mammano,F., Ohagen,A., Hoglund,S., and Gottlinger,H.G. (1994). Role of the Major Homology Region of Human-Immunodeficiency-Virus Type-1 in Virion Morphogenesis. Journal of Virology *68*, 4927-4936.

Mangasarian,A., Foti,M., Aiken,C., Chin,D., Carpentier,J.L., and Trono,D. (1997). The HIV-1 Nef protein acts as a connector with sorting pathways in the Golgi and at the plasma membrane. Immunity. *6*, 67-77.

Mangasarian,A., Piguet,V., Wang,J.K., Chen,Y.L., and Trono,D. (1999). Nef-induced CD4 and major histocompatibility complex class I (MHC-I) down-regulation are governed by distinct determinants: N-terminal alpha helix and proline repeat of Nef selectively regulate MHC-I trafficking. J Virol *73*, 1964-1973.

Mangasarian,A. and Trono,D. (1997). The multifaceted role of HIV Nef. Res. Virol *148*, 30-33.

Mangeat,B., Gers-Huber,G., Lehmann,M., Zufferey,M., Luban,J., and Piguet,V. (2009). HIV-1 Vpu Neutralizes the Antiviral Factor Tetherin/BST-2 by Binding It and Directing Its Beta-TrCP2-Dependent Degradation. Plos Pathogens *5*.

Mannioui,A., Nelson,E., Schiffer,C., Felix,N., Le,R.E., Benichou,S., Gluckman,J.C., and Canque,B. (2005). Human immunodeficiency virus type 1 KK26-27 matrix mutants display impaired infectivity, circularization and integration but not nuclear import. Virology *339*, 21-30.

Marasuilo,L.A., and McSweeney,M. (1977). Nonparemetric and Distribution-Free Methods for the Social Sciences, Monterey, CA: Brooks-Cole, p. 312,

Martin,D.P., Williamson,C. and Posada,D. (2005) RDP2: Recombination detection and analysis from sequence alignments. Bioinformatics 21(2): 260–262.

Martin,M.P. and Carrington,M. (2005). Immunogenetics of viral infections. Curr. Opin. Immunol. *17*, 510-516.

Martinez-Picado,J. *et al.* (2006). Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. Journal of Virology *80*, 3617-3623.

Masemola,A. *et al.* (2004a). Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8(+) T cells: Correlation with viral load. Journal of Virology *78*, 3233-3243.

Masemola,A. *et al.* (2004b). Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8(+) T cells: Correlation with viral load. Virology *78*, 3233-3243.

Matano,T., Shibata,R., Siemon,C., Connors,M., Lane,H.C., and Martin,M.A. (1998). Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. Journal of Virology *72*, 164-169.

Mattapallil,J.J., Douek,D.C., Hill,B., Nishimura,Y., Martin,M., and Roederer,M. (2005). Massive infection and loss of memory CD4(+) T cells in multiple tissues during acute SIV infection. Nature *434*, 1093-1097.

Matthews,P.C. *et al.* (2008). Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. Journal of Virology *82*, 8548-8559.

McElrath,M.J. and Haynes,B.F. (2010). Induction of immunity to human immunodeficiency virus type-1 by vaccination. Immunity. *33*, 542-554.

McMichael,A. and McCutchan,F.E. (2010). Host Genetics and Viral Diversity: Report from a Global HIV Vaccine Enterprise Working Group. Nature precedings, 1-4.

McMichael,A.J. (2007). Triple bypass: complicated paths to HIV escape. Journal of Experimental Medicine *204*, 2785-2787.

McMichael,A.F.B.P., Borrow,P.F., Tomaras GD FAU - Goonetilleke,N., Goonetilleke,N.F., and Haynes,B.F. (2010). The immune response during acute HIV-1 infection: clues for vaccine development. Nat Rev Immunol. *10(1)*, 11-23.

Melikyan,G.B. (2008). Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. Retrovirology *5*.

Mellors,J.W., Rinaldo,C.R., Gupta,P., White,R.M., Todd,J.A., and Kingsley,L.A. (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science *272*, 1167-1170.

Meric,C., Darlix,J.L., and Spahr,P.F. (1984). It is Rous sarcoma virus protein P12 and not P19 that binds tightly to Rous sarcoma virus RNA. J. Mol. Biol. *173*, 531-538.

Michael,N.L. (1999). Host genetic influences on HIV-1 pathogenesis. Curr. Opin. Immunol. *11*, 466-474.

Migueles,S.A. *et al.* (2000). HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proceedings of the National Academy of Sciences of the United States of America *97*, 2709-2714.

Miller,C.J. *et al.* (2005). Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. J. Virol. *79*, 9217-9227.

Miura,T. *et al.* (2009). HLA-B57/B*5801 Human Immunodeficiency Virus Type 1 Elite Controllers Select for Rare Gag Variants Associated with Reduced Viral Replication Capacity and Strong Cytotoxic T-Lymphocyte Recognition. Journal of Virology *83*, 5961.

Mlotshwa,M. *et al.* (2010). Fluidity of HIV-1-specific T-cell responses during acute and early subtype C HIV-1 infection and associations with early disease progression. J. Virol. *84*, 12018-12029.

Momany,C. *et al.* (1996). Crystal structure of dimeric HIV-1 capsid protein. Nat. Struct. Biol. *3*, 763-770.

Moore,J.P. (1997). Co-receptors for HIV-1 entry. Current Opinion in Immunology *9*, 551-562.

Moore,J.P., Parren,P.W., and Burton,D.R. (2001). Genetic subtypes, humoral immunity, and human immunodeficiency virus type 1 vaccine development. J. Virol. *75*, 5721-5729.

Morellet,N., Jullian,N., Derocquigny,H., Maigret,B., Darlix,J.L., and Roques,B.P. (1992). Determination of the Structure of the Nucleocapsid Protein Ncp7 from the Human-Immunodeficiency-Virus Type-1 by H-1-Nmr. Embo Journal *11*, 3059-3065.

Morikawa,Y., Zhang,W.H., Hockley,D.J., Nermut,M.V., and Jones,I.M. (1998). Detection of a trimeric human immunodeficiency virus type 1 gag intermediate is dependent on sequences in the matrix protein, p17. Journal of Virology *72*, 7659-7663.

Mortuza,G.B., Haire,L.F., Stevens,A., Smerdon,S.J., Stoye,J.P., and Taylor,I.A. (2004). High-resolution structure of a retroviral capsid hexameric amino-terminal domain. Nature 431, 481-485.

Mothe,B., Ibarrondo,J., Llano,A., and Brander,C. (2009). Virological, immune and host genetics markers in the control of HIV infection. Dis. Markers. *27*, 105-120.

Mothe,B. *et al.* (2012). CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. PLoS. One. *7*, e29717.

Mudd,P.A., Ericsen,A.J., Burwitz,B.J., Wilson,N.A., O'Connor,D.H., Hughes,A.L., and Watkins,D.I. (2012). Escape from CD8(+) T cell responses in Mamu-B*00801(+) macaques differentiates progressors from elite controllers. J. Immunol. *188*, 3364-3370.

Muesing,M.A., Smith,D.H., Cabradilla,C.D., Benton,C.V., Lasky,L.A., and Capon,D.J. (1985). Nucleic-Acid Structure and Expression of the Human Aids Lymphadenopathy Retrovirus. Nature *313*, 450-458.

Munk,C., Brandt,S.M., Lucero,G., and Landau,N.R. (2002). A dominant block to HIV-1 replication at reverse transcription in simian cells. Proc. Natl. Acad. Sci. U. S. A *99*, 13843-13848.

Murakami,T. and Freed,E.O. (2000). The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. Proc. Natl. Acad. Sci. U. S. A *97*, 343-348.

Naicker,D.D. *et al.* (2012). Association of IL-10-Promoter Genetic Variants With the Rate of CD4 T-Cell Loss, IL-10 Plasma Levels, and Breadth of Cytotoxic T-Cell Lymphocyte Response During Chronic HIV-1 Infection. Clinical Infectious Diseases *54*, 294-302.

Navis,M., Matas,D.E., Rachinger,A., Koning,F.A., van,S.P., Kootstra,N.A., and Schuitemaker,H. (2008). Molecular evolution of human immunodeficiency virus type 1 upon transmission between human leukocyte antigen disparate donor-recipient pairs. PLoS. One. *3*, e2422.

Neil,S.J., Zang,T., and Bieniasz,P.D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature *451*, 425-430.

Niederman,T.M., Hastings,W.R., and Ratner,L. (1993). Myristoylation-enhanced binding of the HIV-1 Nef protein to T cell skeletal matrix. Virology *197*, 420-425.

Nisole,S., Lynch,C., Stoye,J.P., and Yap,M.W. (2004). A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. Proc. Natl. Acad. Sci. U. S. A *101*, 13324-13328.

Noviello,C.F.B.S., Benichou S FAU - Guatelli,J., and Guatelli,J.C. (2008). Cooperative binding of the class I major histocompatibility complex cytoplasmic domain and human immunodeficiency virus type 1 Nef to the endosomal AP-1 complex via its mu subunit. J Virol. *82(3)*, 1249-58.

Novitsky,V. *et al.* (2002). Magnitude and frequency of cytotoxic T-lymphocyte responses: identification of immunodominant regions of human immunodeficiency virus type 1 subtype C. J. Virol. *76*, 10155-10168.

Novitsky,V., Wang,R., Baca,J., Margolin,L., McLane,M.F., Moyo,S., van,W.E., Makhema,J., and Essex,M. (2011). Evolutionary gamut of in vivo Gag substitutions during early HIV-1 subtype C infection. Virology. *421*, 119-128.

Novitsky,V., Wang,R., Margolin,L., Baca,J., Moyo,S., Musonda,R., and Essex,M. (2010). Dynamics and timing of in vivo mutations at Gag residue 242 during primary HIV-1 subtype C infection. Virology 403, 37-46.

Novitsky,V. *et al.* (2009). Viral load and CD4+ T-cell dynamics in primary HIV-1 subtype C infection. J. Acquir. Immune. Defic. Syndr. *50*, 65-76.

O'Brien,T.R., Rosenberg,P.S., Yellin,F., and Goedert,J.J. (1998). Longitudinal HIV-1 RNA levels in a cohort of homosexual men. J. Acquir. Immune. Defic. Syndr. Hum. Retrovirol. *18*, 155-161.

O'Connor,D.H. *et al.* (2002). Acute phase cytotoxic T lymphocyte escape is a hallmark of simian immunodeficiency virus infection. Nature Medicine *8*, 493-499.

Ohno, H. (2006). Clathrin-associated adaptor protein complexes. J. Cell Sci. 119, 3719–3721.

Okoye,A.A. *et al.* (2012). Naive T cells are dispensable for memory CD4+ T cell homeostasis in progressive simian immunodeficiency virus infection. J. Exp. Med. *209*, 641-651.

Oldridge,J. and Marsh,M. (1998). Nef--an adaptor adaptor? Trends Cell Biol. 8, 302-305.

Olivetta,E. *et al.* (2003). HIV-1 Nef induces the release of inflammatory factors from human monocyte/macrophages: involvement of Nef endocytotic signals and NF-kappa B activation. J Immunol. *170*, 1716-1727.

Ono,A., Demirov,D., and Freed,E.O. (2000). Relationship between human immunodeficiency virus type 1 Gag multimerization and membrane binding. J. Virol. *74*, 5142-5150.

Ono,A. and Freed,E.O. (1999). Binding of human immunodeficiency virus type 1 Gag to membrane: role of the matrix amino terminus. J. Virol. *73*, 4136-4144.

Ono,A. and Freed,E.O. (2001). Plasma membrane rafts play a critical role in HIV-1 assembly and release. Proc. Natl. Acad. Sci. U. S. A *98*, 13925-13930.

Op de Coul,E.L. *et al.* (2001). Using phylogenetic analysis to trace HIV-1 migration among western European injecting drug users seroconverting from 1984 to 1997. AIDS *15*, 257-266.

Orci,L., Glick,B.S., and Rothman,J.E. (1986). A new type of coated vesicular carrier that appears not to contain clathrin: its possible role in protein transport within the Golgi stack. Cell *46*, 171-184.

Oxenius,A., Fidler,S., Brady,M., Dawson,S.J., Ruth,K., Easterbrook,P.J., Weber,J.N., Phillips,R.E., and Price,D.A. (2001). Variable fate of virus-specific CD4(+) T cells during primary HIV-1 infection. Eur. J. Immunol. *31*, 3782-3788.

Padron,D., Wang,Y.J., Yamamoto,M., Yin,H., and Roth,M.G. (2003). Phosphatidylinositol phosphate 5-kinase Ibeta recruits AP-2 to the plasma membrane and regulates rates of constitutive endocytosis. J Cell Biol. *162*, 693-701.

Paillart,J.C. and Gottlinger,H.G. (1999). Opposing effects of human immunodeficiency virus type 1 matrix mutations support a myristyl switch model of gag membrane targeting. J. Virol. *73*, 2604-2612.

Pal,R., Reitz,M.S., Tschachler,E., Gallo,R.C., Sarngadharan,M.G., and Veronese,F.D. (1990). Myristoylation of Gag Proteins of Hiv-1 Plays An Important Role in Virus Assembly. Aids Research and Human Retroviruses *6*, 721-730.

Palmer,S. *et al.* (2003). New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. J Clin. Microbiol. *41*, 4531-4536.

Pantaleo,G., Graziosi,C., Demarest,J.F., Butini,L., Montroni,M., Fox,C.H., Orenstein,J.M., Kotler,D.P., and Fauci,A.S. (1993). HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. Nature. *362*, 355-358.

Payne,R.P. *et al.* (2010). Efficacious Early Antiviral Activity of HIV Gag- and Pol-Specific HLA-B*2705-Restricted CD8(+) T Cells. Journal of Virology *84*, 10543-10557.

Peeters,M. and Sharp,P.M. (2000). Genetic diversity of HIV-1: the moving target. AIDS *14 Suppl 3*, S129-S140.

Peeters,M., Toure-Kane,C., and Nkengasong,J.N. (2003). Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. AIDS *17*, 2547-2560.

Peretz,Y., Alter,G., Boisvert,M.P., Hatzakis,G., Tsoukas,C.M., and Bernard,N.F. (2005). Human immunodeficiency virus (HIV)-specific gamma interferon secretion directed against all expressed HIV genes: relationship to rate of CD4 decline. J. Virol. 79, 4908-4917.Pereyra,F. *et al.* (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science *330*, 1551-1557.

Perez-Caballero,D., Zang,T., Ebrahimi,A., McNatt,M.W., Gregory,D.A., Johnson,M.C., and Bieniasz,P.D. (2009). Tetherin inhibits HIV-1 release by directly tethering virions to cells. Cell *139*, 499-511.

Peterlin,B.M. and Trono,D. (2003). Hide, shield and strike back: how HIV-infected cells avoid immune eradication. Nat. Rev. Immunol. *3*, 97-107.

Peyerl,F.W., Bazick,H.S., Newberg,M.H., Barouch,D.H., Sodroski,J., and Letvin,N.L. (2004). Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. J. Virol. *78*, 13901-13910.

Piatak,M.Jr., Yang,L.C., Luk,K.C., Lifson,J.D., Saag,M.S., Clark,S.J., Kappes,J.C., Hahn,B.H., and Shaw,G.M. (1993). Viral dynamics in primary HIV-1 infection. Lancet. *341*, 1099.

Piguet,V., Chen,Y.L., Mangasarian,A., Foti,M., Carpentier,J.L., and Trono,D. (1998). Mechanism of Nef-induced CD4 endocytosis: Nef connects CD4 with the mu chain of adaptor complexes. EMBO J *17*, 2472-2481.

Piguet,V., Gu,F., Foti,M., Demaurex,N., Gruenberg,J., Carpentier,J.L., and Trono,D. (1999). Nef-induced CD4 degradation: a diacidic-based motif in Nef functions as a lysosomal targeting signal through the binding of beta-COP in endosomes. Cell *97*, 63-73.

Piguet,V. and Trono,D. (1999). The Nef protein of primate lentiviruses. Rev. Med. Virol *9*, 111-120.

Piguet,V. and Trono,D. (2001). Living in oblivion: HIV immune evasion. Semin. Immunol. *13*, 51-57.

Piguet,V., Wan,L., Borel,C., Mangasarian,A., Demaurex,N., Thomas,G., and Trono,D. (2000). HIV-1 Nef protein binds to the cellular protein PACS-1 to downregulate class I major histocompatibility complexes. Nat. Cell Biol. *2*, 163-167.

Plantier,J.C., Leoz,M., Dickerson,J.E., De Oliveira,F., Cordonnier,F., Lemee,V., Damond,F., Robertson,D.L., and Simon,F. (2009). A new human immunodeficiency virus derived from gorillas. Nature Medicine *15*, 871-872.

Poon,A.F., Lewis,F.I., Frost,S.D., and Kosakovsky Pond,S.L. (2008). Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. Bioinformatics. *24*, 1949-1950.

Pornillos,O., Higginson,D.S., Stray,K.M., Fisher,R.D., Garrus,J.E., Payne,M., He,G.P., Wang,H.E., Morham,S.G., and Sundquist,W.I. (2003). HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. J Cell Biol. *162*, 425-434.

Prado,J.G., Honeyborne,I., Brierley,I., Puertas,M.C., Martinez-Picado,J., and Goulder,P.J.R. (2009). Functional Consequences of Human Immunodeficiency Virus Escape from an HLA-B*13-Restricted CD8(+) T-Cell Epitope in p1 Gag Protein. Journal of Virology *83*, 1018-1025.

Price,D.A., Goulder,P.J., Klenerman,P., Sewell,A.K., Easterbrook,P.J., Troop,M., Bangham,C.R., and Phillips,R.E. (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. Proc. Natl. Acad. Sci. U. S. A. *94*, 1890-1895.

Price,D.A. *et al.* (2004). T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. Immunity. *21*, 793-803.

Qi,M. and Aiken,C. (2008). Nef enhances HIV-1 infectivity via association with the virus assembly complex. Virology *373*, 287-297.

Quinones-Mateu,M., and Arts,E. (2001). HIV-1 fitness: Implications for drug resistance, disease progression, and global epidemic evolution, p. 134-170. *In* C. Kuiken, B. Foley, B. Hahn, P. Marx, F. McCutchan, J. Mellors, S. Wolinsky, and B. Korber (ed.), HIV sequence compendium 2001. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Radebe,M. *et al.* (2011). Limited immunogenicity of HIV CD8+ T-cell epitopes in acute Clade C virus infection. J. Infect. Dis. *204*, 768-776.

Raiborg,C. and Stenmark,H. (2009). The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. Nature *458*, 445-452.

Reddy,K., Winkler,C.A., Werner,L., Mlisana,K., Karim,S.S.A., and Ndung'u,T. (2010). APOBEC3G expression is dysregulated in primary HIV-1 infection and polymorphic variants influence CD(4+) T-cell counts and plasma viral load. AIDS *24*, 195-204.

Reed,L.J. and Muench,H. (1938). A simple method of estimating fifty per cent endpoints. Am. J. Epidemiol. 27:493– 497.

Reid,S.W. *et al.* (1996). Antagonist HIV-1 Gag peptides induce structural changes in HLA B8. J. Exp. Med. *184*, 2279-2286.

Rein,A., Henderson,L.E., and Levin,J.G. (1998). Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. Trends Biochem. Sci. *23*, 297-301.

Rerks-Ngarm,S. *et al.* (2009). Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. New England Journal of Medicine *361*, 2209-2220.

Resch,W., Ziermann,R., Parkin,N., Gamarnik,A., and Swanstrom,R. (2002). Nelfinavir-resistant, amprenavir-hypersusceptible strains of human immunodeficiency virus type 1 carrying an N88S mutation in protease have reduced infectivity, reduced replication capacity, and reduced fitness and process the Gag polyprotein precursor aberrantly. J. Virol. *76*, 8659-8666.

Richman,D.D., Margolis,D.M., Delaney,M., Greene,W.C., Hazuda,D., and Pomerantz,R.J. (2009). The Challenge of Finding a Cure for HIV Infection. Science *323*, 1304-1307.

Richman,D.D., Wrin,T., Little,S.J., and Petropoulos,C.J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc. Natl. Acad. Sci. U. S. A *100*, 4144-4149.

Roberts,J.D., Bebenek,K., and Kunkel,T.A. (1988). The accuracy of reverse transcriptase from HIV-1. Science *242*, 1171-1173.

Rock,K.L., Gramm,C., Rothstein,L., Clark,K., Stein,R., Dick,L., Hwang,D., and Goldberg,A.L. (1994). Inhibitors of the Proteasome Block the Degradation of Most Cell-Proteins and the Generation of Peptides Presented on Mhc Class-I Molecules. Cell *78*, 761-771.

Roeth,J.F. and Collins,K.L. (2006). Human immunodeficiency virus type 1 Nef: adapting to intracellular trafficking pathways. Microbiol. Mol. Biol. Rev. *70*, 548-563.

Roeth,J.F., Williams,M., Kasper,M.R., Filzen,T.M., and Collins,K.L. (2004). HIV-1 Nef disrupts MHC-I trafficking by recruiting AP-1 to the MHC-I cytoplasmic tail. J Cell Biol. *167*, 903-913.

Rolland,M. *et al.* (2010). Amino-Acid Co-Variation in HIV-1 Gag Subtype C: HLA-Mediated Selection Pressure and Compensatory Dynamics. Plos One *5*.

Rolland,M., Heckerman,D., Deng,W., Rousseau,C.M., Coovadia,H., Bishop,K., Goulder,P.J., Walker,B.D., Brander,C., and Mullins,J.I. (2008a). Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads. PLoS One. *3*, e1424.

Rong,R. *et al.* (2009). Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. PLoS. Pathog. 5, e1000594.

Rousseau,C.M. *et al.* (2008). HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. J. Virol. *82*, 6434-6446.

Sabin,C.A., Devereux,H., Phillips,A.N., Hill,A., Janossy,G., Lee,C.A., and Loveday,C. (2000). Course of viral load throughout HIV-1 infection. J. Acquir. Immune. Defic. Syndr. *23*, 172-177.

Sacha,J.B. *et al.* (2007). Gag-specific CD8+ T lymphocytes recognize infected cells before AIDS-virus integration and viral protein expression. J. Immunol. *178*, 2746-2754.

Sacha,J.B. *et al.* (2009). Gag- and Nef-specific CD4+ T cells recognize and inhibit SIV replication in infected macrophages early after infection. Proc. Natl. Acad. Sci. U. S. A. *106*, 9791-9796.

Saez-Cirion,A. *et al.* (2009). Heterogeneity in HIV suppression by CD8 T cells from HIV controllers: association with Gag-specific CD8 T cell responses. J. Immunol. *182*, 7828-7837.

Sakaguchi,K., Zambrano,N., Baldwin,E.T., Shapiro,B.A., Erickson,J.W., Omichinski,J.G., Clore,G.M., Gronenborn,A.M., and Appella,E. (1993). Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. Proc. Natl. Acad. Sci. U. S. A *90*, 5219-5223.

Saksena,S., Sun,J., Chu,T., and Emr,S.D. (2007). ESCRTing proteins in the endocytic pathway. Trends Biochem. Sci. *32*, 561-573.

Salazar-Gonzalez,J.F. *et al.* (2009). Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J. Exp. Med. *206*, 1273-1289.

Sandefur,S., Smith,R.M., Varthakavi,V., and Spearman,P. (2000). Mapping and characterization of the N-terminal I domain of human immunodeficiency virus type 1 Pr55(Gag). Journal of Virology *74*, 7238-7249.

Saphire,A.C., Bobardt,M.D., and Gallay,P.A. (2002). Cyclophilin a plays distinct roles in human immunodeficiency virus type 1 entry and postentry events, as revealed by spinoculation. J Virol *76*, 4671-4677.

Saric,T., Beninga,J., Graef,C.I., Akopian,T.N., Rock,K.L., and Goldberg,A.L. (2001). Major histocompatibility complex class I-presented antigenic peptides are degraded in cytosolic extracts primarily by thimet oligopeptidase. Journal of Biological Chemistry *276*, 36474-36481.

Sarr,A.D., Eisen,G., Gueye-Ndiaye,A., Mullins,C., Traore,I., Dia,M.C., Sankale,J.L., Faye,D., Mboup,S., and Kanki,P. (2005). Viral dynamics of primary HIV-1 infection in Senegal, West Africa. J. Infect. Dis. *191*, 1460-1467.

Sawai,E.T., Hamza,M.S., Ye,M., Shaw,K.E., and Luciw,P.A. (2000). Pathogenic conversion of live attenuated simian immunodeficiency virus vaccines is associated with expression of truncated Nef. J Virol *74*, 2038-2045.

Sayah,D.M., Sokolskaja,E., Berthoux,L., and Luban,J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature *430*, 569-573.

Schaller,T. *et al.* (2011). HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. PLoS. Pathog. *7*, e1002439.

Schindler,J., Gautsch,J.W., Lerner,R.A., and Hopkins,N. (1981). Biochemical analysis of the p30's of N-, B-, and B leads to NB-tropic murine leukemia viruses of BALB/c origin. J. Virol. *39*, 703-712.

Schmitz,J.E. *et al.* (1999). Control of viremia in simian immunodeficiency virus infection by CD8(+) lymphocytes. Science *283*, 857-860.

Schneidewind,A. *et al.* (2008). Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. Journal of Virology *82*, 5594-5605.

Schneidewind,A. *et al.* (2007). Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. J. Virol. *81*, 12382-12393.

Sebastian,S. and Luban,J. (2005). TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. Retrovirology. *2*, 40.

Seifert,U. *et al.* (2003). An essential role for tripeptidyl peptidase in the generation of an MHC class I epitope. Nature Immunology *4*, 375-379.

Seki,S., Kawada,M., Takeda,A., Igarashi,H., Sata,T., and Matano,T. (2008). Transmission of simian immunodeficiency virus carrying multiple cytotoxic T-lymphocyte escape mutations with diminished replicative ability can result in AIDS progression in rhesus macaques. Journal of Virology *82*, 5093-5098.

Serwold,T., Gaw,S., and Shastri,N. (2001). ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. Nature Immunology *2*, 644-651.

Sheppard,H.W., Lang,W., Ascher,M.S., Vittinghoff,E., and Winkelstein,W. (1993). The characterization of non-progressors: long-term HIV-1 infection with stable CD4+ T-cell levels. AIDS. *7*, 1159-1166.

Shin,H.D. *et al.* (2000). Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of IL10. Proc. Natl. Acad. Sci. U. S. A. *97*, 14467-14472.

Sierra,S., Kupfer,B., and Kaiser,R. (2005). Basics of the virology of HIV-1 and its replication. J. Clin. Virol. *34*, 233-244.

Simmons,A., Aluvihare,V., and McMichael,A. (2001). Nef triggers a transcriptional program in T cells imitating single-signal T cell activation and inducing HIV virulence mediators. Immunity. *14*, 763-777.

Simon,F., Mauclere,P., Roques,P., Loussert-Ajaka,I., Muller-Trutwin,M.C., Saragosti,S., Georges-Courbot,M.C., Barre-Sinoussi,F., and Brun-Vezinet,F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat. Med. *4*, 1032-1037.

Singh,R., Gaiha,G., Werner,L., McKim,K., Mlisana,K., Luban,J., Walker,B.D., Karim,S.S., Brass,A.L., and Ndung'u,T. (2011). Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. J. Virol. *85*, 208-216.

Smith,A.J., Srinivasakumar,N., Hammarskjold,M.L., and Rekosh,D. (1993). Requirements for incorporation of Pr160gag-pol from human immunodeficiency virus type 1 into virus-like particles. J. Virol. *67*, 2266-2275.

Smith,S.M. (2004). HIV CTL escape: at what cost? Retrovirology. *1*, 8.

Soghoian,D.Z. *et al.* (2012). HIV-specific cytolytic CD4 T cell responses during acute HIV infection predict disease outcome. Sci. Transl. Med. *4*, 123ra25.

Sokolskaja,E., Sayah,D.M., and Luban,J. (2004). Target cell cyclophilin a modulates human immunodeficiency virus type 1 infectivity. Journal of Virology *78*, 12800-12808.

Solbak,S.M. *et al.* (2010). The intriguing cyclophilin A-HIV-1 Vpr interaction: prolyl cis/trans isomerisation catalysis and specific binding. BMC. Struct. Biol. *10*, 31.

South,T.L. and Summers,M.F. (1993). Zinc-Dependent and Sequence-Dependent Binding to Nucleic-Acids by the N-Terminal Zinc Finger of the Hiv-1 Nucleocapsid Protein - Nmr Structure of the Complex with the Psi-Site Analog, Dacgcc. Protein Science *2*, 3-19.

Spearman,P., Wang,J.J., Vander,H.N., and Ratner,L. (1994). Identification of human immunodeficiency virus type 1 Gag protein domains essential to membrane binding and particle assembly. J. Virol. *68*, 3232-3242.

Stacey,A.R. *et al.* (2009). Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. J. Virol. *83*, 3719-3733.

Stamatatos,L., Morris,L., Burton,D.R., and Mascola,J.R. (2009). Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? Nat. Med. *15*, 866-870.

Steinkasserer,A. *et al.* (1995). Mode of Action of Sdz Nim-811, A Nonimmunosuppressive Cyclosporine-A Analog with Activity Against Human-Immunodeficiency-Virus Type-1 (Hiv-1) - Interference with Early and Late Events in Hiv-1 Replication. Journal of Virology *69*, 814-824.

Stevenson,M. (2003). HIV-1 pathogenesis. Nat. Med. *9*, 853-860.

Strack,B., Calistri,A., Craig,S., Popova,E., and Gottlinger,H.G. (2003). AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. Cell *114*, 689-699.

Streeck,H., Frahm,N., and Walker,B.D. (2009). The role of IFN-gamma Elispot assay in HIV vaccine research. Virology *4*, 461-469.

Stremlau,M., Owens,C.M., Perron,M.J., Kiessling,M., Autissier,P., and Sodroski,J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature *427*, 848-853.

Stremlau,M., Perron,M., Lee,M., Li,Y., Song,B., Javanbakht,H., Diaz-Griffero,F., Anderson,D.J., Sundquist,W.I., and Sodroski,J. (2006a). Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proc. Natl. Acad. Sci. U. S. A *103*, 5514-5519.

Stremlau,M., Song,B.W., Javanbakht,H., Perron,M., and Sodroski,J. (2006b). Cyclophilin A: An auxiliary but not necessary cofactor for TRIM5 alpha restriction of HIV-1. Virology *351*, 112-120.

Sugimoto,C., Tadakuma,K., Otani,I., Moritoyo,T., Akari,H., Ono,F., Yoshikawa,Y., Sata,T., Izumo,S., and Mori,K. (2003). nef gene is required for robust productive infection by simian immunodeficiency virus of T-cell-rich paracortex in lymph nodes. J Virol *77*, 4169-4180.

Summers,M.F., South,T.L., Kim,B., and Hare,D.R. (1990). High-Resolution Structure of An Hiv Zinc Fingerlike Domain Via A New Nmr-Based Distance Geometry Approach. Biochemistry *29*, 329-340.

Suzuki,Y. and Craigie,R. (2007). The road to chromatin - nuclear entry of retroviruses. Nature Reviews Microbiology *5*, 187-196.

Swingler,S. *et al.* (1999). HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages. Nat. Med. *5*, 997-103.

Tamura,K., Dudley,J., Nei,M., and Kumar,S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

Tang,C., Ndassa,Y., and Summers,M.F. (2002). Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. Nat. Struct. Biol. *9*, 537-543.

Tang,S., Murakami,T., Agresta,B.E., Campbell,S., Freed,E.O., and Levin,J.G. (2001). Human immunodeficiency virus type 1 N-terminal capsid mutants that exhibit aberrant core morphology and are blocked in initiation of reverse transcription in infected cells. J. Virol. *75*, 9357-9366.

Taylor,B.S., Sobieszczyk,M.E., McCutchan,F.E., and Hammer,S.M. (2008). The challenge of HIV-1 subtype diversity. N. Engl. J. Med. *358*, 1590-1602.

Temin,H.M. (1993). Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc. Natl. Acad. Sci. U. S. A *90*, 6900-6903.

Thobakgale,C.F. *et al.* (2009). Impact of HLA in mother and child on disease progression of pediatric human immunodeficiency virus type 1 infection. J. Virol. *83*, 10234-10244.

Tomaras,G.D. *et al.* (2008). Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. J. Virol. *82*, 12449-12463.

Tomiyama,H., Akari,H., Adachi,A., and Takiguchi,M. (2002). Different effects of Nef-mediated HLA class I down-regulation on human immunodeficiency virus type 1-specific CD8(+) T-cell cytolytic activity and cytokine production. J. Virol. *76*, 7535-7543.

Tovanabutra,S., Robison,V., Wongtrakul,J., Sennum,S., Suriyanon,V., Kingkeow,D., Kawichai,S., Tanan,P., Duerr,A., and Nelson,K.E. (2002). Male viral load and

heterosexual transmission of HIV-1 subtype E in northern Thailand. J. Acquir. Immune. Defic. Syndr. *29*, 275-283.

Towers,G.J., Hatziioannou,T., Cowan,S., Goff,S.P., Luban,J., and Bieniasz,P.D. (2003). Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. Nat. Med. *9*, 1138-1143.

Traub, L.M. (2009). Tickets to ride: selecting cargo for clathrin-regulated internalization. Nat. Rev. Mol. Cell Biol. 10, 583–596..

Treurnicht,F.K. *et al.* (2010). Adaptive changes in HIV-1 subtype C proteins during early infection are driven by changes in HLA-associated immune pressure. Virology 396, 213-225.

Troyer,R.M. *et al.* (2009). Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. PLoS. Pathog. *5*, e1000365.

Ulm,J.W., Perron,M., Sodroski,J., and Mulligan,C. (2007). Complex determinants within the Moloney murine leukemia virus capsid modulate susceptibility of the virus to Fv1 and Ref1-mediated restriction. Virology *363*, 245-255.

UNAIDS World AIDS Day Report 2011. (2011). UNAIDS Report on the global HIV/AIDS epidemic.

Unanue,E.R. and Cerottini,J.C. (1989). Antigen presentation. FASEB J. *3*, 2496-2502.

Vallari,A. *et al.* (2011). Confirmation of Putative HIV-1 Group P in Cameroon. Journal of Virology *85*, 1403-1407.

van der Kuyl,A.C., Zorgdrager,F., Jurriaans,S., Back,N.K., Prins,J.M., Brinkman,K., van,E.A., Bakker,M., and Cornelissen,M. (2009). Incidence of human immunodeficiency virus type 1 dual infections in Amsterdam, The Netherlands, during 2003-2007. Clin. Infect. Dis. *48*, 973-978.

van Loggerenberg,F. *et al.* (2008). Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. Plos One *3*, e1954.

van Regenmortel. (2011). Limitations to the structure-based design of HIV-1 vaccine immunogens. Journal of molecular recognition [24], 741-753.

van Regenmortel,M. *et al.* (2000). Seventh Report of the International committee on taxonomy of viruses.

Vandegraaff,N. and Engelman,A. (2007). Molecular mechanisms of HIV integration and therapeutic intervention. Expert. Rev. Mol. Med. *9*, 1-19.

Vanhems,P., Hirschel,B., Phillips,A.N., Cooper,D.A., Vizzard,J., Brassard,J., and Perrin,L. (2000). Incubation time of acute human immunodeficiency virus (HIV) infection and duration of acute HIV infection are independent prognostic factors of progression to AIDS. J. Infect. Dis. *182*, 334-337.

Veazey,R.S., DeMaria,M., Chalifoux,L.V., Shvetz,D.E., Pauley,D.R., Knight,H.L., Rosenzweig,M., Johnson,R.P., Desrosiers,R.C., and Lackner,A.A. (1998). Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science. *280*, 427-431.

von Schwedler,U., Kornbluth,R.S., and Trono,D. (1994). The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. Proc. Natl. Acad. Sci. U. S. A *91*, 6992-6996.

von Schwedler,U.K., Stemmler,T.L., Klishko,V.Y., Li,S., Albertine,K.H., Davis,D.R., and Sundquist,W.I. (1998). Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. Embo Journal *17*, 1555-1568.

von Schwedler,U.K., Stray,K.M., Garrus,J.E., and Sundquist,W.I. (2003). Functional surfaces of the human immunodeficiency virus type 1 capsid protein. Journal of Virology *77*, 5439-5450.

Wan,L., Molloy,S.S., Thomas,L., Liu,G., Xiang,Y., Rybak,S.L., and Thomas,G. (1998). PACS-1 defines a novel gene family of cytosolic sorting proteins required for trans-Golgi network localization. Cell *94*, 205-216.

Wang,Y.E. *et al.* (2009). Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. J. Virol. *83*, 1845-1855.

Wawer,M.J. *et al.* (1999). Control of sexually transmitted diseases for AIDS prevention in Uganda: a randomised community trial. Lancet *353*, 525-535.

Wei,X. *et al.* (2003). Antibody neutralization and escape by HIV-1. Nature. *422*, 307-312.

Williams,M., Roeth,J.F., Kasper,M.R., Fleis,R.I., Przybycin,C.G., and Collins,K.L. (2002). Direct binding of human immunodeficiency virus type 1 Nef to the major histocompatibility complex class I (MHC-I) cytoplasmic tail disrupts MHC-I trafficking. J Virol *76*, 12173-12184.

Wills,J.W. and Craven,R.C. (1991). Form, function, and use of retroviral gag proteins. AIDS *5*, 639-654.

Wilson,J.D. *et al.* (2000). Direct visualization of HIV-1-specific cytotoxic T lymphocytes during primary infection. AIDS. *14*, 225-233.

Wollert,T., Yang,D., Ren,X., Lee,H.H., Im,Y.J., and Hurley,J.H. (2009). The ESCRT machinery at a glance. J Cell Sci. *122*, 2163-2166.

Woo,J., Robertson,D.L., and Lovell,S.C. (2010). Constraints on HIV-1 Diversity from Protein Structure. Journal of Virology *84*, 12995-13003.

Woodman,Z., Mlisana,K., Treurnicht,F., Abrahams,M.R., Thebus,R., Karim,S.A., and Williamson,C. (2011). Short communication decreased incidence of dual infections in

South african subtype C-infected women compared to a cohort ten years earlier. AIDS. Res. Hum. Retroviruses. *27*, 1167-1172.

Woodman,Z. and Williamson,C. (2009). HIV molecular epidemiology: transmission and adaptation to human populations. Curr. Opin. HIV. AIDS *4*, 247-252.

Wright,J.K. *et al.* (2010). Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. J. Virol. *84*, 10820-10831.

Wright,J.K., Naidoo,V.L., Brumme,Z.L., Prince,J.L., Claiborne,D.T., Goulder,P.J., Brockman,M.A., Hunter,E., and Ndung'u,T. (2012). Impact of HLA-B*81-associated mutations in HIV-1 Gag on viral replication capacity. J. Virol. *86*, 3193-3199.

Wright,J.K. *et al.* (2011). Influence of Gag-Protease-Mediated Replication Capacity on Disease Progression in Individuals Recently Infected with HIV-1 Subtype C. Journal of Virology *85*, 3996-4006.

Wu,X., Anderson,J.L., Campbell,E.M., Joseph,A.M., and Hope,T.J. (2006). Proteasome inhibitors uncouple rhesus TRIM5alpha restriction of HIV-1 reverse transcription and infection. Proc. Natl. Acad. Sci. U. S. A *103*, 7465-7470.

Xu,X.N. *et al.* (1997). Evasion of cytotoxic T lymphocyte (CTL) responses by nef-dependent induction of Fas ligand (CD95L) expression on simian immunodeficiency virus-infected cells. J Exp. Med. *186*, 7-16.

Yang,O.O., Nguyen,P.T., Kalams,S.A., Dorfman,T., Gottlinger,H.G., Stewart,S., Chen,I.S., Threlkeld,S., and Walker,B.D. (2002). Nef-mediated resistance of human immunodeficiency virus type 1 to antiviral cytotoxic T lymphocytes. J. Virol. *76*, 1626-1631.

Yang,Y., Huang,J.H., Toth,I., Lichterfeld,M., and Yu,X.G. (2010). Mutational Escape in HIV-1 CTL Epitopes Leads to Increased Binding to Inhibitory Myelomonocytic MHC Class I Receptors. Plos One *5*.

Yap,M.W., Nisole,S., Lynch,C., and Stoye,J.P. (2004). Trim5alpha protein restricts both HIV-1 and murine leukemia virus. Proc. Natl. Acad. Sci. U. S. A *101*, 10786-10791.

Yoo,S., Myszka,D.G., Yeh,C., McMurray,M., Hill,C.P., and Sundquist,W.I. (1997). Molecular recognition in the HIV-1 capsid/cyclophilin A complex. J. Mol. Biol. *269*, 780-795.

York,I.A., Mo,A.X.Y., Lemerise,K., Zeng,W.Y., Shen,Y.L., Abraham,C.R., Saric,T., Goldberg,A.L., and Rock,K.L. (2003). The cytosolic endopeptidase, thimet oligopeptidase, destroys antigenic peptides and limits the extent of MHC class I antigen presentation. Immunity *18*, 429-440.

Younes,S.A., Yassine-Diab,B., Dumont,A.R., Boulassel,M.R., Grossman,Z., Routy,J.P., and Sekaly,R.P. (2003). HIV-1 viremia prevents the establishment of interleukin 2-producing HIV-specific memory CD4+ T cells endowed with proliferative capacity. J. Exp. Med. *198*, 1909-1922.

Zhou,W., Parent,L.J., Wills,J.W., and Resh,M.D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. J. Virol. *68*, 2556-2569.

Zhu,T., Corey,L., Hwangbo,Y., Lee,J.M., Learn,G.H., Mullins,J.I., and McElrath,M.J. (2003). Persistence of extraordinarily low levels of genetically homogeneous human immunodeficiency virus type 1 in exposed seronegative individuals. J Virol *77*, 6108-6116.

Zimbwa,P. *et al.* (2007). Precise identification of a human immunodeficiency virus type 1 antigen processing mutant. Journal of Virology *81*, 2031-2038.

Zimmerman,C., Klein,K.C., Kiser,P.K., Singh,A.R., Firestein,B.L., Riba,S.C., and Lingappa,J.R. (2002). Identification of a host protein essential for assembly of immature HIV-1 capsids. Nature *415*, 88-92.

Zuniga,R. *et al.* (2006). Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. J. Virol. *80*, 3122-3125.

Appendix 1 Table of evidence of CTL escape in Gag and Nef; in individuals (n=18) with yellow highlights, escape was confirmed with IFN-γ ELISpot assay

PID	Wks PI	EPITOPE	SFU/10^6 PBMC	FREQUENCY	HLA	Location
CAP008	2	RPGGKKQYMLKHL	918	M→K; 80.2-5.1%	B*0801	p17 Gag
	3	.....	2220			
	6	.....	2610	Escape confirmed		
	10	.....K....	2695			
	13	.....R....	1345			
	25	.....R....				
	56	.....R....				
	2	YSKKRQEILDLWVYHTQ	0	Y-H; (71.33-11.6)		
	3	.....	70	K-Q; (83.45-8.36)	Cw*0701	Nef
	6	.....	280			
	10	.....	555	Escape confirmed		
	13	H.Q.....	2495	K-R; (82.94-8.19%)		
	25	H.Q.....	2118			
	56	...R.....				
CAP030	5	EGEDVGFPVKPQVP	-	(E - 68.6%)		
	11	..DE.....R....	155	(D - 6.83%)	B*4501	Nef
	29	..DE.....	90			
	54	..DE.....		Escape confirmed		
CAP040	4	KIRLRPGGKKHYMLK	NOT DONE	M→R; 80.2-11.4%		
	12	.....R..			B*4201	p17 Gag
	22	.....R..				
	48	.....R..				
	93	.....R..				
	4	RDYVDRFFKTLRAE		T-C; (87.65-1.41%)		p24 Gag
	12	.....C....			Cw*0304	
	22	.....C....				
	48	.....C....				
	93	.....C....				
CAP045	2	EEEEVGFPVRPQVPVRP		Escape confirmed	B*4501	Nef
	5	.....	863	E→G; (68.6-20.8)%		
	7	..G.....	153			
	12	..G.....	113			
	43	..G.....	0			
	133	..G.....	0			
CAP063	2	EEEVGFPVRPQVPLRPM			B*4501	Nef
	4	.....				
	5	..G.....				
	11	..G.....		E→G; (86.5-4.6%)		
	29	..D.....				
	37	..D.....				
	2	LRSLYNTVATLYCVHK		L-I; (98.06-0.97)		p17 Gag
	4	.....			A*0201	
	5	.....				
	11	.....				
	29	.....				
	37	.....I.....				
CAP085	5	EEEKEVGFPVRPQVPL	NO RESPONSE			
	13	...ED.....		E→D; (86.5-7.17%)		
	29	...ED.....			B*4501	Nef
	39	...EG.....				
	55	...GD.....				
	5	KRQEILDLWVYHTQGY	350	Y-F (44.57-54.4%)		
	13	.....	1322.5			
	29	.....F	352.5	Escape confirmed	Cw*0701	Nef
	39	.....F				
	55	.....F				

	5	ELRSLYNTVATLY	0	L-I;(93.22-5.57%)		
	13	.....	180		A*3002	p17 Gag
	29	.....	110			
	39	.....		Escape confirmed		
	55	.I.....				
CAP088	5	EEEKEVGFPVRPQVPVR	NO RESPONSE		B*4501	Nef
	13	...K.....		E→K; (86.5-0.34%)		
	54	...K.....				
	5	VLAEAMSQANSANIM			B*4501	p7/p6 Gag
	6	.....	465	A-V; (19.37-7.02%)		
	13	.....V...	138			
	26	.....V...	0			
	46	.....T...	0	Escape confirmed		
	54	.....T...	0			
CAP129	2	YSKKRQEILDLWVHHTQG	1775		B*1801	Nef
	13	.....Y....	120	E→D; (59.56.39.93%)		
	26	.....D.....Y....	0			
	54	.....D.....Y....		Escape confirmed		
	2	PVRPQVPVRPMTYK	155	V→I; (98.63-0.17%)	B*8101	Nef
	13	.....	137.5			
	26	.....	0	Escape confirmed		
	54	.....I.....	0			
	2	SEGATPTDLNTML	0	Q/T-S (91.53/6-3.63%)	B*8101	p24 Gag
	4	.....	0			
	11	.....	NO RESPONSE			
	23	.....	0			
	32	.....S.....				
	54	.....S.....				
	2	EPFRDYVDRFFKVLR	NO RESPONSE	K-R; (97.1-2.91%)	B*1801	p24 Gag
	4	.....				
	11	.....				
	23	.....				
	42	.....R...				
	54	.....R...				
	106	.....R...				
CAP174	2	RPQVPLRPMTYKAA	405		A*0301	Nef
	11	.....	135	V→A; (98.63-0%)		
	24	.....	172.5	Escape confirmed		
	40	...A.....	67.5			
	53	...A.....				
	2	KKRRDILDLWVYN			Cw*0702	Nef
	4	.E.....	717.5	K-E (82.94 - 0.17)		
	11	.....H.	295			
	13	.....H.	257.5	N: 0.51%		
	17	.N.....H.	197.5	Escape confirmed		
	40	.N.....				
	53	.N.....				
CAP177	2	EEEEVGFPVRPQVPL	NOT DONE	E→D;(86.5-7.17%)	B*4501	Nef
	4	.....				
	13	..G.....				
	25	..D.....				
	54	..D.....				
CAP188	4	SLYNTVATLY	NOT DONE		A*0202	p17 Gag
	13	..X.A.....		Y-H; (55.55-2.6)%		
	48	..H....V..				
	93	..X....V..				
						140

	4	RLRPGGRKHY		R-K; (88.6-8.96%)	A*7401	p17 Gag
	13	....				
	48	K.........				
	93	K.........				
CAP210	5	HQAISPRTL	0	A-S (71.19-6.54%)	B*1510	p24 Gag
	12	....	70	Escape confirmed		
	22	....	135			
	50	..S......				
CAP211	3	AISPRTLNAW	NOT DONE	A-L (71.19-27.12%)	B*57	p24 Gag
	10	.L........		P (19.37)		
	23	P.........				
	27	P.........				
	66	P.........				
	3	TSTLQEQITW	NOT DONE		B*57	p24 Gag
	10	....H.....				
	23	..N.......		T-N (84.75-10.41)		
	27	..N.......				
	66	..N.......				
CAP217	9	TYKAAFDLSFFLKE	1238	A-E; (47.6-0.17%)	B*5801	Nef
	12	...E......	1030			
	31	...E......		Escape confirmed		
	60	...E......				
	9	TTSTLQEQIAWM	217	T-N (84.75-10.41)	B*5801	p24 Gag
	12	....	2190			
	15	.....P.....	2008	Escape confirmed		
	17	...N.......	1380			
	19	...N.......	485			
	23	...N.......	75			
CAP221	5	YSKKRQEILDLWVYN	70	L-M;(99.15-0.55%)	Cw7	Nef
	12	....	217.5			
	24	.........M....	475	Escape confirmed		
	46	...R..D........				
CAP224	3	KAAFDLSFFLKEKG	NOT DONE	A-G;(47.6-46.8)	A*0301	Nef
	12	....				
	25	.G............				
	46	.G............				
CAP228	8	IIWSKRRQDILDLWVYNT	NO RESPONSE			
	11	L.............I...		V-I; (97.78-0.85%)	Cw7	Nef
	47	L.............I...				
	147	L.............I...				
CAP229	7	TGTEELRSLYNTVTLY	0	R-K;(51.82-45.76%)	B58	p17 Gag
	8	....	0			
	21	....	0			
	25	......K........	NO RESPONSE			
	80	......K........				
	7	TTSTLQEQIAWM	1188			
	8	...N.....T..	105	T-N (84.75-10.41)	B*5801	p24 Gag
	9	...N.....T..	340			
	21	...N.....T..		Escape confirmed		
	25	...N.....T..				
CAP239	2	TTSTLQEQVAWM				
	2	....I....T..		T-N (84.75-10.41)	B*5801	p24 Gag
	5	...N.....T..	380	Escape confirmed		
	6	...N.....T..	80			
	22	...N.......	0			
						141

CAP248	9	NAWVKVIEEKAFSP	NOT DONE	K-N (99.76-0%)	B*1503	p24 Gag
	12	.....				
	22	.....				
	52	.....N....				
CAP255	8	KKRQEILDLWVHNT	1250	Escape confirmed	B*1801	Nef
	13	...R.....Y..	1070	Q-R (95.9-0%)		
	36	....D.....Y..		E→D; (59.56.39.93%)		
	47	E..K.....Y..				
	8	GIRYPLTFGWCYKLV	100	V-I; (83.96-8.01%)	B*1801	Nef
	13	.....	0			
	36	.....		Escape confirmed		
	47	.....	990			
CAP258	7	RYPLTFGWCFK	2273	F-Y; (97.27-0.85%)	A*2301	Nef
	24	.....Y.....	627.5			
	40	.....Y.....		Escape confirmed		
CAP261	9	SKQRQEILDLWVYN	275		B*1302	
	12	.RK.....	115	K-R; (83.45-5.63%)		Nef
	26	.RK.....	148	Escape confirmed		
CAP262	7	VGFPVRPQVPLRPMT	925	R-K; (87.88-8.53%)	B*8101	
	12	.....	760	Escape confirmed		Nef
	16	.....K.....	290			
	26	.....K.....				
CAP265		NYTPGPGVRYPLTFG	NOT DONE	Y-L; (80.89-14.16)	B*4201	Nef
		.....L.....				
		.....L.....				
CAP266	2	RSLYNTVATLYCVH				
	12	......I.....	1155	V-I (93.7-5.8%)		
	32	......T.....	1063	T (0%)	A*2901	p17 Gag
	49	......T.....				
	86	......T.....		Escape confirmed		
	124	.....I....				
	150	.....I....				
	7	YNTQGYFPDWQNYT	5636			
	12	.....F.....	5043	Y-F; (54.44-44.37%)	A*2901	Nef
	32	.....F.....		Escape confirmed		
	49	.....F.....				
	124	.....F.....				
CAP268	8	LHPVHAGPIAP	0	I-V (47.7-42.6)	B*07	p24 Gag
	24	.....	73			
	53	.....V..	NO RESPONSE			
	53	.....N..				
	86	.....V..				
	137	.....V..				
	8	RSLFNTMATLYCVH	NO RESPONSE	M (0%)	A*0205	p17 Gag
	24	......I.....		I (5.8%)		
	53	......I.....				
	53	......I.....				
	86	......I.....				
	137	......I.....				
	8	PVRPQVPLRPMT	NO RESPONSE		B*0702	Nef
	15	.....		R-K; (87.88-8.53%)		
	24	..K.....				
CAP269	4	PVRPQVPLRPMT	NOT DONE			
	11	..K.....		R-K; (87.88-8.53%)	B*0702	Nef
	20	..K.....				
	131	..K.....				

	4	HQPISPRTLNAWV-154			B*0702	p24 Gag
	11	....T.....		S-T; (98.8-0.97%)		
	24	..A.....				
	45	.....A.....				
	93	..A.....				
	131	.....A.....				
CAP270	8	WPNVRERMRRAE	NOT DONE	N-D (5.8-7.5)	B*0801	Nef
	29	..D...K...T.		R-K (93.17-3.4%)		
	58	..E.....T.				
	8	KAAFDLSFFLK	2065	S-G (76.28-23.21%)	B*5801	Nef
	29	.....	335	Escape confirmed		
	58	......G....				
CAP274	10	KGAVDLGFFLK	NOT DONE	G (46.76%)		Nef
	22	.G.V..G....		V (10.76)	B*5801	
	60	.G.V..G....		G (23.21)		
	10	TTSNLQEQVQWM		T-N (84.75-10.41)		
	22	...N....VQ..			B*5801	p24 Gag
	60	...N....VQ..				
CAP275	8	KIMLRPGGK	NOT DONE		A*7401	p17 Gag
	11	..G.....		M-G/S (0/0-1.25%)		
	13	..S.....				
	15	..S.....				
	27	..S.....				
	46	..S.....				
CAP277	8	TPQDLNTML	NOT DONE		B*8101	p24 Gag
	13	.....		Q-S (91.53-3.63%)		
	16	.....				
	20	.....				
	24	.....				
	48	..S.....				
	84	..S.....				
	97	..S...S..				
CAP278	9	SDGTPSDLNSML	NOT DONE		B*3910	p24 Gag
	12	.....		E-D (98.3-1.45%)		
	23	.....		Q-S (91.53-3.63%)		
	54	.....		T-S (96-2.66%)		
	85	.....				
	89	.....				
	9	FPVRPQVPVRPMTYKA	NOT DONE	R-K; (87.88-8.53%)	B*3910	Nef
	12	.....				
	23	.....L.....				
	54	...T....L.....				
CAP282	11	SEGATPQDLNTML	NOT DONE	Q-S (91.53-3.63%)	B*4201	p24 Gag
	14	.....				
	26	.....				
	53	......T.....				
CAP289	8	SEGATPADLNTML	NOT DONE		B*3910	p24 Gag
	18	.....		Q/A-G; (91.5/0.97-0.97%)		
	24	.....				
	28	......G.....				
	41	......G.....				

APPENDIX 2 HLA Data for the CAPRISA participants in this study

SAMPLE	A	A	B	B	Cw	Cw
CAP008	2301	2301	0801	1510	0701	1601
CAP030	0201	3402	4403	4501	0401	1601
CAP037	2301	2402	0702	5301	1701	1701
CAP040	3001	3002	1510	4201	0304	1701
CAP045	2301	2902	1510	4501	0602	1601
CAP061	6602	6802	1401	4201	0802	1701
CAP063	0201	2301	4501	4501	0401	1601
CAP065	2301	6802	1510	5802	0804	0602
CAP069	0301	2301	1503	5802	0210	0602
CAP084	2902	7401	1503	4403	0210	0701
CAP085	3002	3002	0801	4501	0701	1601
CAP088	2902	6601	4501	5802	0602	0602
CAP129	2601	8001	1801	8101	0202	0401
CAP136	3004	7401	4201	4201	1701	1701
CAP137	2902	6801	4101	5802	0602	1701
CAP174	0301	7401	4901	5802	0602	0701
CAP177	0301	3002	1510	4501	0401	1601
CAP188	0202	7401	1503	1516	0210	1402
CAP200	0205	4301	1510	4101	0401	1701
CAP206	0301	3201	0702	4403	0210	0702
CAP210	6802	6802	1510	1510	0304	0304
CAP211	0205	3002	5703	5703	1801	1701
CAP217	0202	2901	1503	5801	0210	0602
CAP220	3004	7401	4201	4201	1701	1701
CAP221	6802	6802	4403	4403	0701	0701
CAP222	3001	3303	5301	8101	0401	0401
CAP224	0301	6602	4201	5802	0602	1701
CAP225	0101	3001	4202	8101	1701	1801
CAP228	2301	2638	4403	5101	0303	0701
CAP229	0123	0123	5801	5801	0602	0602
CAP237	6602	6802	4201	4201	1701	1701
CAP239	0123	2902	4201	5801	0602	1701
CAP244	2301	3004	4403	5802	0401	0602
CAP248	0205	2902	1401	1503	0210	0804
CAP255	0301	8001	0801	1801	0202	0702
CAP256	2901	6601	1503	5802	0401	0602
CAP257	2301	2902	4202	4403	1701	1701
CAP258	2301	2902	4101	4201	1701	1701
CAP259	0202	6601	1401	3910	08	12
CAP260	2902	4301	1503	4403		18
CAP261	2911	4301	1302	1503	0602	0602
CAP262	0101	6602	4201	8101	1701	1801
CAP264	3601	6802	1510	5301	0401	0804
CAP265	0201	3001	4201	4507	1601	1701
CAP266	2901	6601	1503	5802	0401	0602
CAP267	0301	2301	1401	5802	0210	0602
CAP268	0205	2601	0705	5801	0701	0702
CAP269	0205	6802	0702	5802	06	07
CAP270	0301	3002	0801	5801	0701	0701
CAP271	0205	2301	1401	4403	0303	0804
CAP274	0201	3001	4201	5801	0302	1701
CAP275	2902	7401	1503	4201	0210	1701

CAP276	0201	2301	1510	5801	06	08
CAP277	3009	4301	5802	8101	0401	0401
CAP278	3001	4301	3910	4202	1203	1701
CAP279	2902	6801	5802	5802	0602	0602
CAP280	2902	7401	1503	1510	0210	0804
CAP281	02	30	4201	4403		
CAP282	3001	6601	4201	5802		
CAP285	0301	2402	0702	5802	0602	0702
CAP289	3002	6801	3910	5802	0602	1203
CAP294	29	02	1302	5802	0602	0602

APPENDIX 3 Identification of transmitted escape mutations associated with a replicative fitness cost to the virus and participants' clinical data. Shown are mutations in the HLAs - B*57/58:01 restricted epitopes, ISW9, KF11 and TW10; the B*39:10/81:01 restricted epitope TL9; and the B*13 restricted epitope RI9. Participants infected with viruses carrying a novel putative CTL escape mutation in p17 affecting viral replication capacity are also included.

PID	ISW9 (147-155)		KF11 (162-172)		TL9 (180-188)	TW10 (240-249) and associated mutations						RI9 (429-437)	p17*	[N]Viral Load at 3 months PI	[N]CD4 count at 3 months PI	[N]Viral Load at 12 months PI	[N]CD4 count at 12 months PI
	A146P	I147L[h]	A163X	S165X		T242X	A248X	M250X	H219Q	I223V	M228						
CAP008[J]			A163G	S165N			A248G							4.59	299	5.27	445
CAP030[J]					T180I					I223V	M228L			4.76	573	5.78	607
CAP037	A146S	I147L								I223V		K436RI437L		5.02	311	4.60	434
CAP040						T242S				I223V	M228L			4.04	321	4.07	334
CAP045[a, g]	A146P	I147M			L188F			M250I	H219Q					2.75	1093	3.99	888
CAP061[g]	A146P	I147L				T242N			H219Q	I223V	M228I		Q65H	2.62	402	3.64	490
CAP063	A146N	I147L	A163G	S165N			A248G							5.20	243	5.48	531
CAP065[a,g]	A146S	I147L												4.71	194	4.78	243
CAP069[J]							A248T				M228I			6.09	202	6.14	297
CAP085[g]	A146P	I147L				T242N					M228L		Q65H	3.82	510	4.46	613
CAP084[J]													Q65H***	3.48	519	4.64	524
CAP088[g]	A146P		A163G		Q182H, P181S	T242N				I223N				4.59	509.5	4.59	726
CAP129	A146P						A248T				M228L			4.90	634	4.34	727
CAP137						T242S								5.10	535	5.39	338
CAP174[J]								M250I						4.75	345	4.78	348
CAP177[a]		I147L					A248T							4.76	465	4.62	370
CAP188		I147L												4.49	585	4.95	375
CAP200[a, g]	A146P	I147L				T242N				I223N				5.24	539	5.03	367
CAP206[J]									H219P	I223P	M228L			5.15	292	5.40	297
CAP220													Q65H	2.60	400	2.60	656
CAP221	A146P	I147L	A163N			T242N				I223V				3.09	716	4.73	435
CAP224	A146P	I147L												5	330	4.60	168
CAP225[g]	A146P	I147L				T242N								4.33	597	4.81	461
CAP228[g]	A146P	I147L	A163S			T242N				I223V	M228I			3.18	811	3.12	1002
CAP244[J]							A248N							4.16	303	4.60	469
CAP255[g]	A146P	I147L	A163S					M250I					Q65H	4.56	342	5.03	624
CAP257		I147L											Q65H**	4.00	634	5.24	591
CAP258[J]		I147L					A248N						Q65H	5.08	231	5.43	300
CAP260							A248T							3.54	303	3.84	330
CAP261[a]		I147L	A163G	S165N				M250I						4.82	418	5.65	359
CAP264[a]							A248T							5.59	256	4.40	318
CAP265			A163G	S165N				M250I						3.26	944	3.53	989
CAP266		I147L									M228L			4.89	434	3.30	527
CAP267	A146P		A163G	S165N		T242N				I223P				4.09	543	4.29	341
CAP269	A146P												Q65H	2.70	666	2.99	681

CAP271	A146S	I147L											**Q65H**	3.21	595	4.00	534
CAP275		I147L												4.80	737	3.45	528
CAP277	A146P	I147L						M250I					**Q65H**	3.81	513	2.68	502
CAP278		I147L											**Q65H****	2.94	661	2.95	568
CAP279	A146P					T242N	A248Q		H219Q		M228I			5.19	326	4.51	368
CAP281		I147L	A163N					M250I						3.44	707	2.98	702
CAP136														3.32	785	3.06	748
CAP210[J]														4.53	534	5.58	288
CAP222														4.53	632	2.65	654
CAP237														4.72	522	4.08	339
CAP248[J]														4.90	358	4.81	432
CAP256[J]														5.02	475	5.35	355
CAP262														4.13	386	3.57	432
CAP282														4.01	446	3.17	497
CAP211													**Q65H**	4.90	592	3.00	360
CAP217														5.21	515	4.43	498
CAP229														4.17	601	4.35	696
CAP239														4.21	1158	5.19	793
CAP268														3.62	656	3.88	474
CAP270														5.72	436	5.87	268
CAP274														4.32	811	4.93	498

Blank cell (s): the specified mutation(s) is/are absent; X: change away from HIV-1 C clade consensus; Individuals who are positive for HLAs - B*57/58:01 are shown in red PID numbers; [a]Participant has HLA-B*1510 or HLA-B*13 associated with ISW9 residues A146X/I147X (Honeyborne *et al.*, 2007); Participants [J]negative (n=11) and [g]positive (n=9) for T242N/X in the TW10 and/or A142P/X in the ISW9 epitope included in the study (n=21) of Chopera *et al.* (2008); [h]Though previously thought to compensate for fitness cost due to escape through A146X, Crawford *et al.* (2009) showed that I147L, together with A146P mutation incur fitness cost to the virus;[N]Log_{10}Viral load (copies/mL) and CD4+ T-cell counts (cells/µL) are medians of measurements taken at three time-points closest to these intervals.*Gag p17 mutation at HXB2 position 65; **Participant infected with virus that developed the mutation at HXB2 position 65 in Gag p17 or else was there at transmission –in purple colour: the polymorphism is present in virus in participants positive for HLA-B7 supertype or A*30 alleles that may be linked to formation of the polymorphism. *** Mutation reverted to subtype C consensus residue; Mutations in the cyclophilin binding loop known to compensate for escape in TW10 (T242X) are highlighted in Green.

Appendix 4 Timing of reversion of protective HLA associated mutations in mismatched participants

PID	Mutation	Sample sequenced latest time-point	Timing (Wks) of reversion
CAP030	T180I		5 - 11
CAP040	242X	93	N
CAP045	A146X	199	133-199
	I147X		133-199
	F188L		133-199
	M250I	199	N
CAP061	A146X	122	76-122
	I147X		76-122
	242X		58-76
CAP063	A146X	37	N
	I147X		N
	F163X		N
	F165X		N
CAP085	A146X	107	N
	I147X		N
	242X		13-29wks
CAP088	A146X	108	N
	A163X		N
	T242X		46-54
	P181S		26-46
	Q182H		N
CAP137	242X	52	12-23
CAP174	M250I	53	N
CAP200	A146X	106	N
	I147X		N
	242X		25-50
CAP221	A146X	46	N
	I147X		N
	F163X		N
	242X		20-46
CAP225	A146X	183	130-145
	I147X		N
	242X		61-74
CAP228	A146X	203	N
	I147X		N
	F163X		N
	242X		96-147
CAP255	A146X	51	N
	I147X		N
	A163X		N
	M250I		10-13wKS
CAP261	I147X	48	N
	A163X		N
	S165X		N
	M250I		26-49
CAP265	A163X	75	N
	S165X		N
	M250I		N
CAP267	A146X	58	12-26WKS
	A163X		N
	S165X		N
	M242X		26-48
CAP277	A146X	137	58-71
	I147X		123-137
	M250X		84-97
CAP279	A146X	50	N
	T242X		N
CAP281	I147X		N
	A163X		N
	M250X		42-46

N=Not yet reverted by the latest time-point sequenced.