# The Class I Human Leukocyte Antigen in Immune Control and Intra-host Evolution of HIV-1

**Jacqueline Roxanne Brener**

Wolfson College

Michaelmas Term 2015


*A thesis submitted to the University of Oxford in partial fulfillment of the requirements for the degree of Doctor of Philosophy*

**For the Levine Stekels**

*without whom there would be no thesis*

"Science appears calm and triumphant when it is completed; but science in the process of being done is only contradiction and torment, hope and disappointment."

Pierre Paul Emile Roux (1853-1933)

Pasteur Institute, Paris

# Abstract


**The Class I Human Leukocyte Antigen in Immune Control and Intra-host Evolution of HIV-1**

Jacqueline Brener, Wolfson College

Thesis submitted for the degree of Doctor of Philosophy, Michaelmas 2015

The CD8+ T (CTL) cell response is an important component of the immune response to HIV-1 infection. Presentation of peptides by the human leukocyte antigen (HLA) class I is key in directing that response. HLA-B*27:05 and HLA-B*57:01 have been consistently associated with immune control of HIV. To gain novel insights into the determinants of an effective CTL response mediated by these alleles, this work investigated cases of HIV disease control and progression in HLA-B*27:05 and B*57:01-expressing subjects. These subjects were either infected with clades of HIV infrequently found in combination with these alleles (C clade in HLA-B*27:05, and CRF01_AE clade in HLA-B*27:05/B*57:01) and/or represented the rare scenario of dual expression of both alleles. Transmission pair studies, viral replicative capacity assays and full-length ultra-deep sequencing were used to characterise CTL-mediated intra-host viral evolution, viral fitness and their effect on disease outcome.

In the context of HLA-B*27:05 expression, this work demonstrated that the patterns of CTL immune escape and the associated viral fitness are considerably different in C clade compared to B clade HIV infection. This work also showed that rapid intra-host viral evolution and selection of escape may lead to disease progression despite expression of both favourable HLA class I alleles. The critical role of the Gag-specific CTL response in maintaining disease control, despite applying an unbiased genome-wide approach to detecting viral evolution, was demonstrated.

This work has described the interactions between expression of HLA-B*27:05 and/or HLA-B*57:01, intra-host viral evolution and HIV disease progression, contributing to our understanding of CTL-mediated immune control. This is an important approach for elucidating correlates of protective immunity against HIV infection that may inform the design of intervention strategies.

## Acknowledgements

I would like to thank my supervisor Professor Philip Goulder for offering me a position in his lab and for ideas, input and supervision that have shaped this work. I am indebted to my co-supervisor Dr Philippa Matthews for her mentorship, support and invaluable insights. Thank you for all the hours of patient proofreading of draft after draft after draft, constant encouragement, and calm, unfailing discretion. I could not have hoped to work with a more inspirational woman in science.

I would like to thank all the individuals who provided technical assistance and advice in completing this work, especially Ellen Leitman, Rebecca Payne, Claudia Juarez, Joannah Fergusson, Ayako Kurioko, and Hannah Roberts. I would also like to thank the clinicians and study participants without whom this work would not be possible. I am particularly grateful to the study participants willing to share very personal and sensitive information so that we could piece together their transmission histories. I would like to thank all the collaborators who have contributed such valuable work to these pages, with particular mention of Dr Astrid Gall for her patience and mentorship over email and during my visits to the Wellcome Trust Sanger Institute.

I would like to thank the funders that have made my time at Oxford possible; the wonderful Skye Foundation, the Commonwealth Scholarship Commission, and the Oppenheimer Memorial Trust.

To Di, Dov, Lorraine, Ada and Rosie: you are my refuge. Thank you for invaluable conversations over posh chocolate cake, rescue missions, coconut water rooibos tea and for teaching me what 'family' really means. My DPhil experience would have ended at 6 months, …and at 3 years …and at 3.95 years …and several other times in between without you.

Thanks to Moms and Pops for being a constant source of support and encouragement always just a Skype call away, and to Sam for the same, with a healthy dose of absolute rationality thrown in when it mattered most.

Thank you to my wonderful housemate and friend Jen for being the queen of word(s), both the dictionary and Microsoft types, for countless meals during thesis write up and for putting up with me and all the black dogs and the thesis monsters that have filled our house in recent months. Your support and advice have been invaluable

Thank you to all the individuals at the Peter Medewar Building who made coming to work each day a little easier. There are too many of you to mention. I take life-long friendships away from my time here. I am particularly grateful for overwhelming support from Jo, Ayako, Alba and Ellen. Thank you also to the wonderful friends who I've met in other capacities during my time in Oxford, especially Iona, with whom I have shared so many DPhil trials.

Thank you to all my wonderful South African friends and family who have cheered me on from further away, and offered invaluable pick-me-ups when I came home for visits, especially Amy and Nadia, for your constant encouragement.

Thank you all for traveling the long and bumpy road with me.

## Contributions to the Thesis

The text and data presented here are my own work unless otherwise stated. The contributions of colleagues and collaborators are detailed below.

All Chapters

- My supervisors Professor Philip Goulder and Dr Philippa Matthews have contributed ideas and supervision to all work presented here
- Ultra-deep sequencing was performed in collaboration with Dr Astrid Gall and Prof Paul Kellam (Wellcome Trust Sanger Institute, UK). Dr Astrid Gall performed quality control on raw reads, *de novo* assembly of consensus sequences, and minor variant calls.
- Dr Rebecca Batorsky performed minor variant analysis including determination of epitope haplotypes.

Chapter 3

- The following individuals contributed population level Sanger sequencing data; Claudia Juarez, Allison Hempensall, Supriya Singh, Nora Lavandier.
- Dr Jacob Hurst (University of Oxford, UK) wrote the code for logisitic curve modeling of viral replicative capacity data.
- Dr Jonathan Carlson (Microsoft Research, US) performed the analysis of co-variation.
- Dr Astrid Gall provided training and support for performing Bayesian phylogenetic analysis using BEAST v1.8.0.

Chapter 4

- Nora Lavandier contributed clonal and population level Sanger sequencing data.

Chapter 5

- The majority of the work presented in this chapter has been published in (Brener et al., 2015) with the following co-authors: Astrid Gall, Rebecca Batorsky, Lynn Riddell, Soren Buus, Ellen Leitman, Paul Kellam, Todd Allen, Philip Goulder, Philippa Matthews.
- Dr Rebecca Batorsky (Ragon Institute, US) performed the heat map analyses. Troubleshooting and interpretation of these analyses was performed collaboratively.
- Dr Soren Buus produced the peptide-MHC tetramers.

# List of Publications and Sequence Accession Numbers

**Brener** J, Gall A, Batorsky R, Riddell L, Buus S, Leitman E, Kellam P, Allen T, Goulder P, Matthews PC. 2015. Disease progression despite protective HLA expression in an HIV-infected transmission pair. Retrovirology 12:55.

> The Illumina MiSeq sequencing data obtained for the transmission pair in this paper and in Chapter 5 are available from the EMBL/GenBank/ DDBJ Sequence Read Archive under accession numbers: ERS250039, ERS250040, ERS250041, ERS250042, ERS394610 and ERS394611. Consensus sequences have been deposited in GenBank under the accession numbers: KP873161-KP873166.

Hunt M, Gall A, Ong SH, **Brener** J, Ferns B, Goulder P, Nastouli E, Keane JA, Kellam P, Otto TD. 2015. IVA: accurate de novo assembly of RNA virus genomes. Bioinformatics 31:2374–2376.

Juarez-Molina CI, Payne R, Soto-Nava M, Avila-Rios S, Valenzuela-Ponce H, Adland E, Leitman E, **Brener** J, Muenchhoff M, Branch S, Landis C, Reyes-Teran G, Goulder P. 2014. Impact of HLA selection pressure on HIV fitness at a population level in Mexico and Barbados. J Virol 88:10392–10398.

Kloverpris HN, McGregor R, McLaren JE, Ladell K, Stryhn A, Koofhethile C, **Brener** J, Chen F, Riddell L, Graziano L, Klenerman P, Leslie A, Buus S, Price DA, Goulder P. 2014. Programmed death-1 expression on HIV-1-specific CD8+ T cells is shaped by

epitope specificity, T-cell receptor clonotype usage and antigen load. AIDS 28:2007–2021.

*Manuscripts in preparation:*

**Brener** *et al.* Patterns of immune escape and associated fitness cost are clade specific in HLA-B*27:05-mediated immune control of HIV

**Brener** *et al*. Characterising failed immune control of HIV in adults with HLA-B*27:05 and HLA-B*57:01.

# Contents

# List of Figures

# List of Table

# Abbreviations

| | |
|---|---|
| Ab | Antibody |
| ADCC | Antibody Dependant Cell-mediated Cytotoxicity |
| AIDS | Acquired Immuno-Deficiency Syndrome |
| APC | Allophycocyanin |
| APC | Antigen Presenting Cell |
| APOBEC3G | apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G |
| ART | Antiretroviral Therapy |
| BEAST | Bayesian Evolutionary Analysis Sampling Trees |
| bnAb | Broadly Neutralising Ab |
| CCR | C-C motif Chemokine Receptor |
| CD | Cluster of Differentiation |
| CP | Chronic Progressor |
| CRF | Circulating Recombinant Form |
| CTL | Cytotoxic T Lymphocyte (CD8+ T lymphocyte) |
| CTLA-4 | Cytotoxic T Lymphocyte-associated protein 4 |
| CXCR | C-X-C motif Chemokine Receptor |
| CypA | Cyclophilin A |
| DC | Dendritic Cell |
| DC-SIGN | Dendritic Cell ICAM3-grabbing non-integrin |
| DMSO | Dimethyl Sulfoxide |
| DNA | Deoxyribonucleic Acid |
| dNTP | Deoxynucleotide Triphosphate |
| EC | Elite Controller |

| | |
|---|---|
| EDTA | Ethylene-Diamine-Tetra-acetic Acid |
| EliSpot | Enzyme-Linked Immunosorbent Spot Assay |
| Env | Envelope glycoproteins |
| ER | Endoplasmic Reticulum |
| ERAAP | Endoplasmic Reticulum Aminopeptidase associated with Antigen Processing |
| FACS | Flourescence Activated Cell Sorting |
| FCS | Foetal Calf Serum |
| FITC | Fluorescein Isothiocyanate |
| FSc | Forward Scatter |
| Gag | Group-specific Antigen (capsid protein) |
| GALT | Gut Associated Lymphoid Tissue |
| GFP | Green Fluorescent Protein |
| GWAS | Genome-Wide Association Study |
| HAART | Highly Active Antiretroviral Therapy |
| HCV | Hepatitis C Virus |
| HESN | HIV Exposed Sero-Negative |
| HIV | Human Immunodeficiency Virus |
| HLA | Human Leukocyte Antigen |
| HTLV | Human T-Lymphotropic Virus |
| ICAM | Intracellular Adhesion Molecule |
| IFN | Interferon |
| IL | Interleukin |
| Int | Integrase |

| | |
|---|---|
| KIR | Killer Immunoglobulin-like Receptor |
| LB | Luria-Bertani Medium |
| LILR | Leukocyte Immunoglobulin-Like Receptors |
| LTNP | Long-Term Non-Progressor |
| LTR | Long Terminal Repeat |
| MFI | MedianFlourescence Intensity |
| MHC | Major Histocompatibility Complex |
| ML | Maximum Likelihood |
| MOI | Multiplicity of Infection |
| MRCA | Most Recent Common Ancestor |
| MSM | Men who have Sex with Men |
| nAb | Neutralising Antibody |
| Nef | Negative Regulatory Factor Protein |
| NHP | Non-Human Primate |
| NK | Natural Killer cell |
| NS5B | HCV Non-Structural protein 5b (RNA polymerase) |
| OLP | Over-Lapping Peptide |
| PBMC | Peripheral Blood Mononuclear Cell |
| PBS | Phosphate Buffered Saline |
| PCR | Polymerase Chain Reaction |
| PD-1 | Programmed cell Death receptor 1 |
| PE | Phycoerythrin |
| PEG | Polyethylene Glycol |
| PHA | Phytohaemagglutinin |
| PIC | Pre-integration Complex |

| | |
|---|---|
| Pol | protease, reverse transcriptase and integrase polyprotein |
| Rev | Regulator of Expression of Virion Proteins |
| RNA | Ribonucleic Acid |
| RP | Rapid Progressor |
| RPMI | Roswell Park Memoral Institute (media) |
| RT | Reverse Transcriptase |
| SAMHD1 | SAM domain and HD domain-containing protein 1 |
| SDM | Site Directed Mutagenesis |
| SFU | Spot Forming Unit |
| SIV | Simian Immunodeficiency Virus |
| SSc | Side Scatter |
| TAP | Transporter Associated with Antigen Processing |
| Tat | Trans-activator of HIV-1 gene expression |
| TCID | Tissue Culture Infective Dose |
| TCR | T Cell Receptor |
| TF | Transmission Founder |
| TIM-3 | T cell Immunoglobulin domain and Mucin domain-containing protein 3 |
| TRIM5α | Tripartite Motif containing 5α |
| VC | Viraemic Controller |
| Vif | Viral Infectivity Factor |
| VL | Viral Load |
| Vpr | Viral Protein R |
| Vpu | Viral Protein U |
| WT | Wild Type |

# CHAPTER 1: Introduction

## 1.1 The HIV/AIDS pandemic

The first reports of a condition characterised by rare malignancies and opportunistic infections that later became known as Acquired Immune Deficiency (AIDS) virus occurred in 1981 among men who have sex with men (MSM) in the Unites States. Two years later the causative agent, a retrovirus initially classified as Human T-lymphotropic Virus (HTLV) III was isolated (Barré-Sinoussi et al., 1983; Gallo et al., 1983). The virus later became known as Human Immunodeficiency Virus (HIV).

Today, HIV/AIDS is a global pandemic. In 2014, it was estimated that 36.9 million people were infected with HIV, with 2 million becoming newly infected that year and 1.2 million dying of AIDS-related illness (UNAIDS, 2014). Disease burden is concentrated in sub-Saharan Africa with 25.8 million people living with HIV in this region alone in 2014.

## 1.2 HIV diversity and evolution

### 1.2.1 Origin of HIV-1, HIV-2 and HIV-1 subtypes

HIV-1 and HIV-2 originate from two separate simian immunodeficiency virus (SIV) zoonotic transmission events[1]. HIV-1 (group M and N, described below) is closely related to $SIV_{CPZ}$ in chimpanzees (*Pan troglodytes troglodytes)* (Sharp & Hahn,

[1] SIV tends to be non-pathogenic in its natural host (sooty mangabeys and African green monkeys). It is however, pathogenic in the rhesus macaque species. SIV infection in rhesus macaques is thus a common non-human primate (NHP) model of HIV infection.
[2] HIV-1 group M is referred to simply as HIV throughout this thesis
[3] The terms LTNP, EC, VC and RP may be defined differently by some research groups. Alternative definitions are reviewed in (Migueles & Connors, 2010).
[4] R5-tropic and X4-tropic virus may also be referred to as M-tropic and T-tropic

2010). Cross-over of $SIV_{CPZ}$ into human populations occurred in Central Africa, probably as a result of bushmeat hunting, and can be traced back as early as the 1920s (Faria et al., 2014), well before the earliest reported HIV-1 sequence from 1959 (Nahmias et al., 1986; Zhu et al., 1998). HIV-2, a less pathogenic virus that accounts for a minority of HIV infections largely confined to West Africa, has been traced back to zoonosis from $SIV_{SM}$ in sooty mangabeys.

HIV-1 strains can be subdivided into four distinct groups, namely M (major), O (outlier), N (non-M non-O) and the most recently discovered P groups (Figure 1.1) (Hemelaar, Gouws, Ghys, & Osmanov, 2006; Plantier et al., 2009)[2]. Each group originated from an independent zoonotic transmission event; group M and N from $SIV_{CPZ}$, while group O and P are most closely related to $SIV_{gor}$ from Western lowland gorillas (*Gorilla gorilla gorilla*). The vast majority of global HIV infections are caused by HIV-1 sub-group M (Hemelaar, 2012).

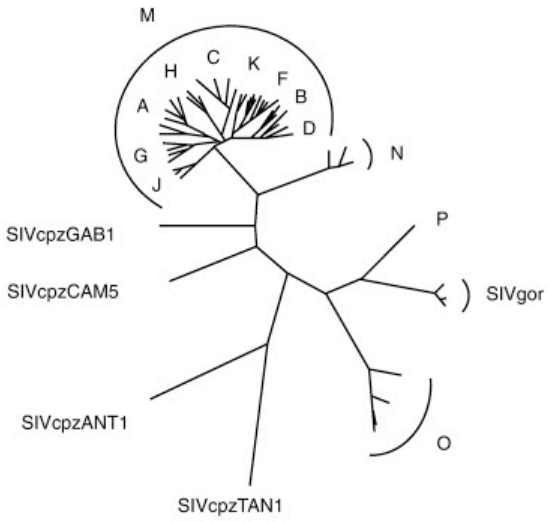



**Figure 1.1 Maximum likelihood phylogenetic tree of HIV and SIV polymerase sequences.**
This illustrates the genetic relatedness of HIV-1 sub-groups and SIV variants. Reproduced from (Hemelaar, 2012) with permission from Elsevier.

Within pandemic HIV-1 group M, nine clades (subtypes) are recognised; A-D, F-H, J and K. Recombination between strains gives rise to circulating recombinant forms

[2] HIV-1 group M is referred to simply as HIV throughout this thesis

(CRFs). Seventy-two CRFs are currently listed on the Los Alamos database (www.hiv.lanl.gov). The clades show a distinct geographical distribution pattern (Figure 1.2). This is an important consideration for designing studies and clinical interventions for different parts of the world. The clades of particular relevance to this thesis are clade B (most prevalent in Europe and the Americas), clade C (most prevalent in Sub-Saharan Africa) and CRF01_AE (prevalent in Thailand).

At the epicentre of the HIV pandemic in Central Africa, all the M group subtypes were found early in the epidemic (Faria et al., 2014). The chance dispersal and establishment of particular strains to other geographical regions (the founder effect) is likely to be a major determinant of the current global distribution of subtypes.s

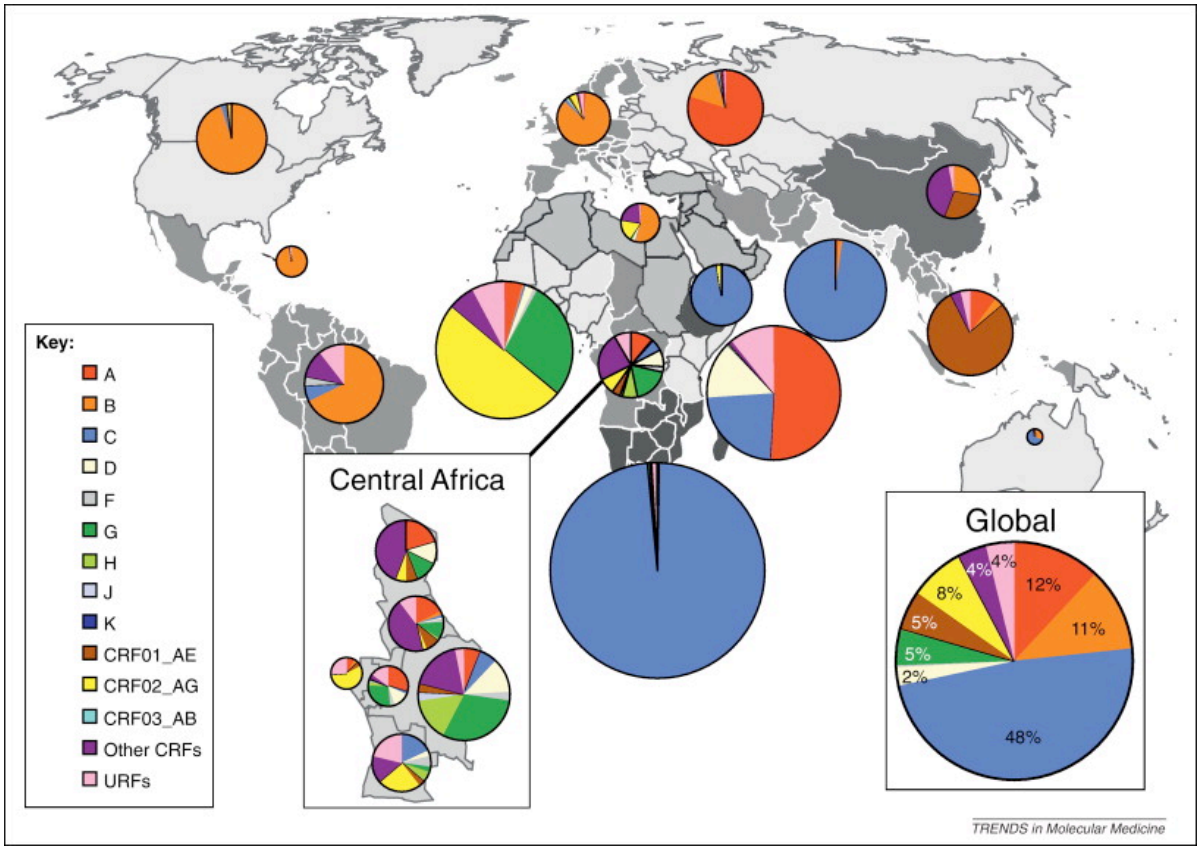



**Figure 1.2 Global distribution of HIV-1 group M clades**. Reproduced from (Hemelaar, 2012) with permission from Elsevier.

### 1.2.2 HIV diversity and evolutionary rate

HIV diversity exists on multiple levels. The existence of different clades and circulating recombinant forms represents diversity at a global level. Variation between these subtypes at the amino acid level across the proteome ranges from 17-35%, but can be as high as 43% (Yang, 2009). The next level of diversity exists within clades. Within-clade variation at the amino acid level ranges from 8-17% (up to 30%) (Hemelaar, 2012). Finally, HIV exists as a diverse population of variants within each infected individual. Viral sequences within a single individual can differ by up to 10% (Korber et al., 2001), with intra-host evolutionary rates estimated at $10^{-3}$-$10^{-4}$ nucleotide substitutions/site/year (Salemi, 2013). Within a single genome, diversity varies by gene, with the *gag* and *pol* genes being the most conserved and *env* and *nef* the most variable (Figure 1.3)

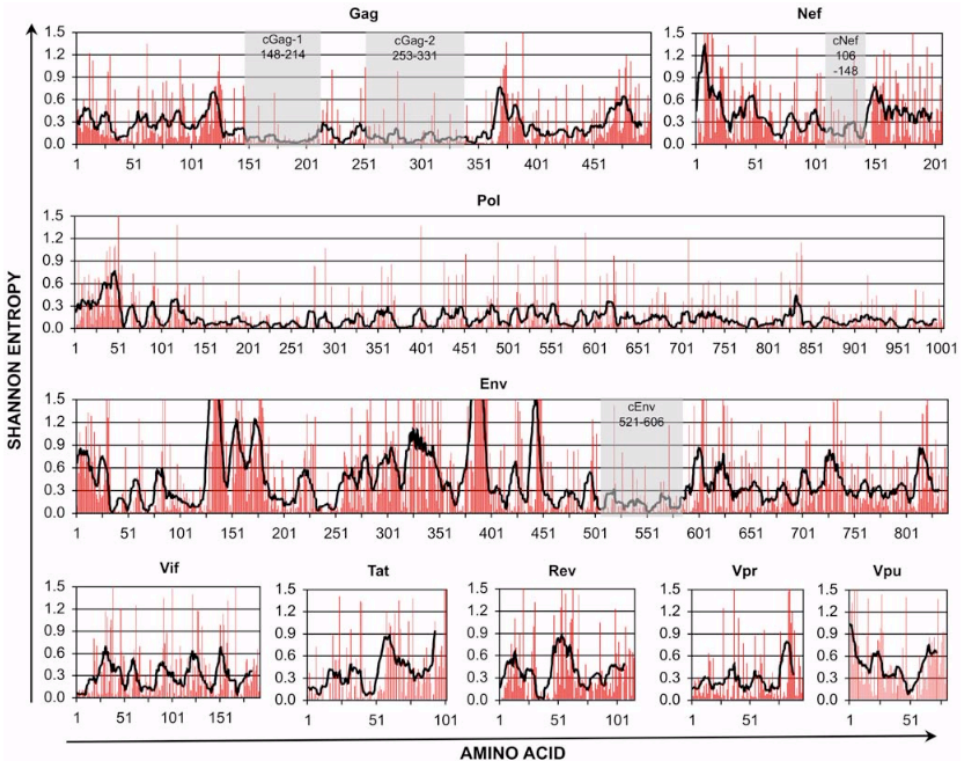



**Figure 1.3 Variability across the HIV genome.** Shannon entropy (a quantitative measure of uncertainty (Shannon, 1948) that provides an estimate of variability at each residue when applied to sequence data) at each amino acid position among B clade sequences is shown. Reproduced from (Yang, 2009) with permission under CC BY licence.

The HIV reverse transcriptase (RT) enzyme lacks proofreading capability. The resulting error-prone replication and high mutation rate is a major contributor to high rates of viral diversity. Mutation rate is estimated at $3.4\times10^{-5}$ mutations/nucleotide/replication cycle, or one nucleotide substitution in every 2-3 newly synthesised virions (Salemi, 2013). Rapid viral generation times (1.5-2.6 days) and high rates of recombination further contribute to the rapid evolution of HIV.

## 1.3 The HIV-1 virion and genome

### 1.3.1 Classification of HIV-1

HIV belongs to the *Retroviridae,* a family of viruses that replicate through reverse transcription of a single-stranded RNA genome. It belongs to the *Lentivirus* genus, a grouping characterised by delayed onset of disease symptoms.

### 1.3.2 Structure of the HIV-1 virion

The mature HIV-1 particle is 100-120nm in diameter (Figure 1.4). It comprises a cone-shaped core composed of p24 capsid protein that encloses two copies of the positive-sense single-stranded RNA genome and associated nucleocapsid proteins. The viral enzymes integrase (Int) and RT are also contained within the capsid. Vif and Nef regulatory proteins are closely associated with the core, while the viral accessory gene product is likely found outside of the core. The core is enclosed by a layer comprising p17 matrix proteins. During the process of viral budding from infected host cells, the virion becomes encased in a host derived lipid bilayer. The envelope glycoproteins form the outer-most surface of the virion, and comprise

external gp120 proteins and transmembrane gp41 proteins that associate to form trimers (Ley, 2007).

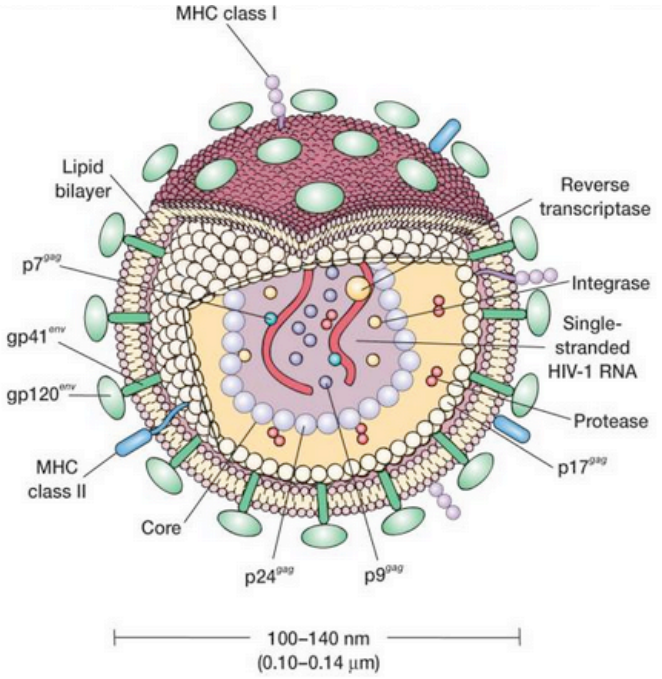



**Figure 1.4 Structure of the HIV-1 virion. s** Reproduced from http://classroom.sdmesa.edu/eschmid/lecture9-microbio.htm.

### 1.3.3 Organisation and gene function of the HIV-1 genome

The HIV-1 genome is 9719 nucleotides in length (Figure 1.5). Translation in all three reading frames is required to produce the full HIV proteome.

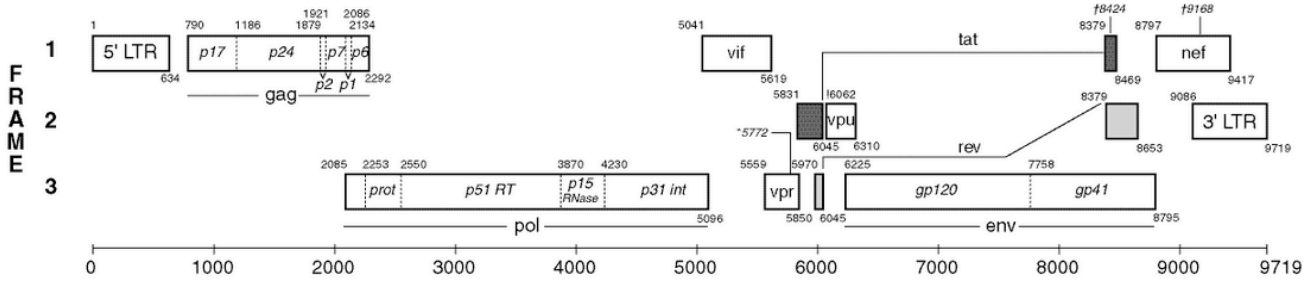



**Figure 1.5 Schematic representation of the HIV-1 genome.** Reproduced from the Los Alamos database (http://www.hiv.lanl.gov).

*Non-coding regions*

The long terminal repeat (LTR) is a DNA sequence flanking both ends of integrated proviruses. It has regulatory functions especially important for transcription initiation and polyadenylation.

*Structural Proteins*

The structural proteins are essential components of the retroviral particle. The *gag* (group-specific antigen) gene encodes the p55 (assemblin) polyprotein precursor that is cleaved by viral protease to produce the p17 matrix, p24 capsid, p7 nucleocapsid and p6 proteins. The *env* (envelope) gene encodes the gp160 glycoprotein that forms the gp120-gp41 complex that assembles to form trimers on the surface of the virion.

*Viral Enzymes*

The *pol* gene encodes a Gag-Pol polyprotein produced by a frameshift near the 3'end of *gag.* This polyprotein is cleaved by viral protease to produce the viral protease, RT, Int and RNAse H enzymes.

*Regulatory Proteins*

The regulatory proteins modulate viral gene expression and are essential for viral replication. The *tat* (trans-activator of gene expression) gene products (Tat-1 and Tat-2 forms) localise at the nucleus and activate viral RNA transcription and elongation from the LTR. The *rev* (regulator of expression of virion proteins) gene product is a phosphoprotein that regulates viral RNA splicing. It stabilises viral

mRNA for nuclear export into the cytoplasm by binding the Rev response element (RRE) in the *env* gene.

*Accessory Proteins*

The accessory proteins are non-essential for virus replication *in vitro* but their conservation suggests that they perform important functions *in vivo.* Vif, Vpr and Vpu proteins recruit ubiquitin ligases to induce ubiquitination and proteasomal degradation of host factors that may inhibit pathogenicity. Vif (viral infectivity factor) is a cytoplasmic protein that promotes free-virus infectivity and inhibits the host restriction factor APOBEC-3G by preventing its incorporation into the virion and targeting it for degradation. The *vpr* (viral protein R) gene product localises to the nucleus where it performs a number of functions including promoting nuclear import of pre-integration complexes, cell growth arrest and apoptosis, activation of cellular genes, and regulation of reverse transcription. The *vpu* (viral protein U) gene product promotes ubiquitin-dependent degradation of the CD4 receptor in the ER, enhances virion release from the membrane of infected cells and plays a role in Env protein maturation. It also counteracts the host restriction factor tetherin. *Nef* (negative regulatory factor) gene products are produced in large numbers early in infection and are important for HIV-1 pathogenicity and infectivity. Nef functions in evasion of the cytotoxic T lymphocyte (CTL) response through down-regulation of HLA-A and –B on the cell surface. It also down-regulates CD4, CCR5 and CXCR4 expression, limiting superinfection. Nef interacts with components of host cell signal transduction and activation pathways.

## 1.4 The HIV-1 replication cycle

*Attachment, fusion and entry*

During infection of a CD4+ target cell, the Env gp120 glycoprotein of the HIV-1 virion engages the CD4 receptor and a CCR5 or CXCR4 co-receptor (Figure 1.6). This triggers fusion of the viral and cellular membranes, allowing entry of the core virus particle into the cell.

*Uncoating and reverse transcription*

The viral capsid uncoats to release viral RNA and protein into the cytoplasm. Reverse transcription produces DNA that is incorporated into the pre-integration complex (PIC).

*Nuclear import and integration*

Nuclear import and integration of the provirus by the Int enzyme and host factors follows reverse transcription of the viral genome.

*Transcription and translation*

Host RNA polymerase II induces proviral transcription of RNA (spliced and unspliced mRNA and full-length genomes) that is exported from the nucleus in a host-protein dependent manner. mRNA is translated by the host machinery to produce the viral proteins that are translocated to the cell surface.

*Assembly, release and maturation of viral particles*

Genome-length RNA molecules and viral proteins are assembled to produce immature viral particles. Budding of the viral particle from the cell surface encases the virion in a host-derived lipid bilayer, and the virion is released. Protease-mediated maturation (cleavage of polyproteins to form mature Gag proteins) produces an infectious virus that is able to infect other CD4+ target cells.

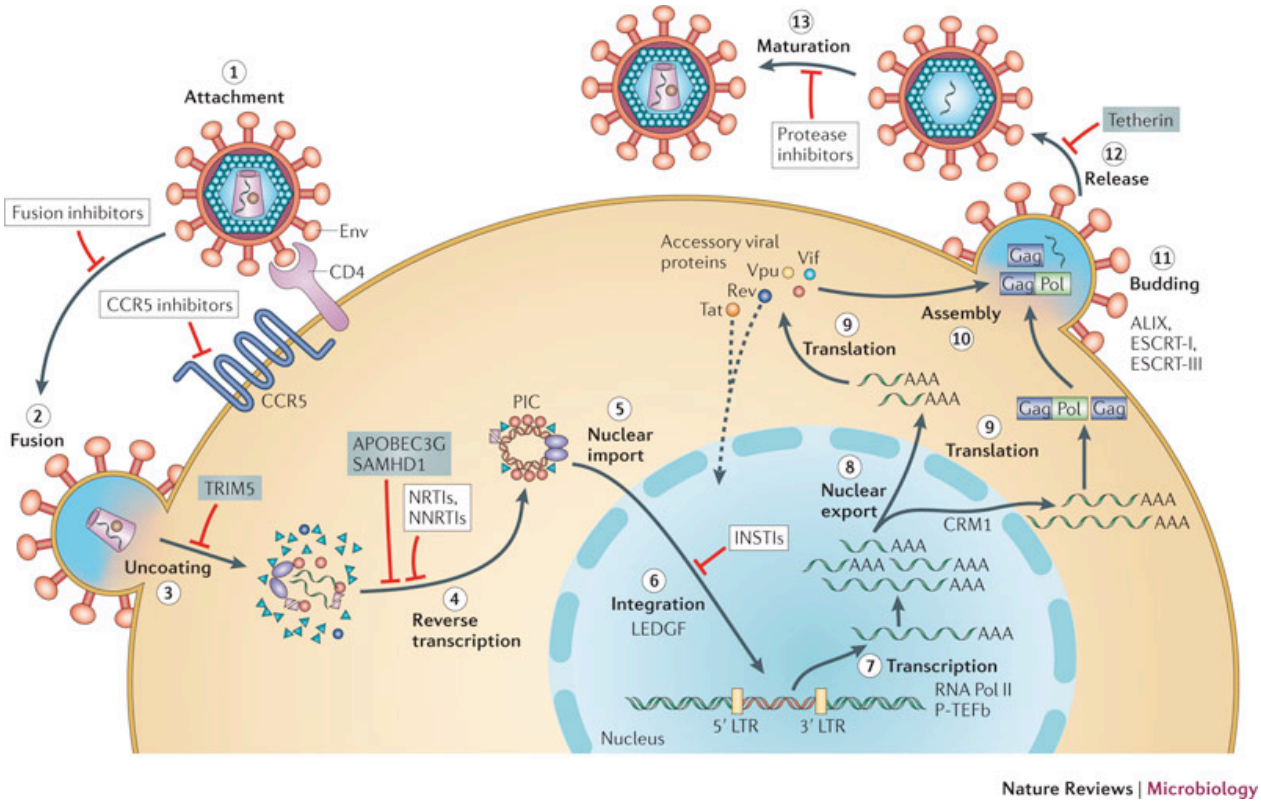



**Figure 1.6 The replication cycle of HIV-1.** Reproduced from (Engelman & Cherepanov, 2012) with permission from Nature Publishing Group.

### 1.4.1 Host factor targets of the HIV-1 replication cycle

Cellular restriction factors also target specific steps in the HIV replication cycle (Figure 1.6). APOBEC3G is an ancient anti-viral defence mechanism that involves deamination of cytidine to uridine in foreign mRNA molecules (Simon, Bloch, & Landau, 2015). Binding of TRIM5α to the viral capsid protein after entry disrupts

uncoating of the virion. SAMHD1, an enzyme with phosphohydrolase activity, inhibits reverse transcription by depleting dNTP stores required for the process. Tetherin inhibits budding of new virus particles by tethering them to the cell plasma membrane. HIV has evolved to express antagonists of these host factors. Vif and Vpu inhibit APOBEC3G, and tetherin respectively (Simon et al., 2015).

## 1.5 The clinical course of HIV-1 infection

The typical course of HIV infection procedes though acute infection, chronic infection and the onset of AIDS (Figure 1.7). During the acute phase of HIV infection, individuals may experience influenza-like symptoms, lymphadenopathy, rash and fever. The eclipse phase of acute infection is a period of approximately 10 days before viral RNA is detectable in peripheral blood. This is followed by the period of peak viraemia. Amplification of the virus occurs through infection of activated CD4+ T cell targets in the draining lymph node and systemic spread to other lymphoid tissues. Spread to the gut-associated lymphoid tissue (GALT) in particular leads to an exponential increase in plasma viraemia. A concurrent decrease in the CD4+ T cell count occurs during this phase. Following peak viraemia and peak CD8+ T cell expansion shorly after the acute phase, the viral setpoint is established at 4-6 months post-infection, where the viral load (VL) declines to a more stable level. The viral setpoint is strongly predictive of the rate of disease progression (Mellors et al., 1996). Fiebig staging is a system for further classifying the earliest stages of infection based on detection of HIV RNA, p24 protein and HIV-specific Abs (Fiebig et al., 2003) (Figure 1.8).

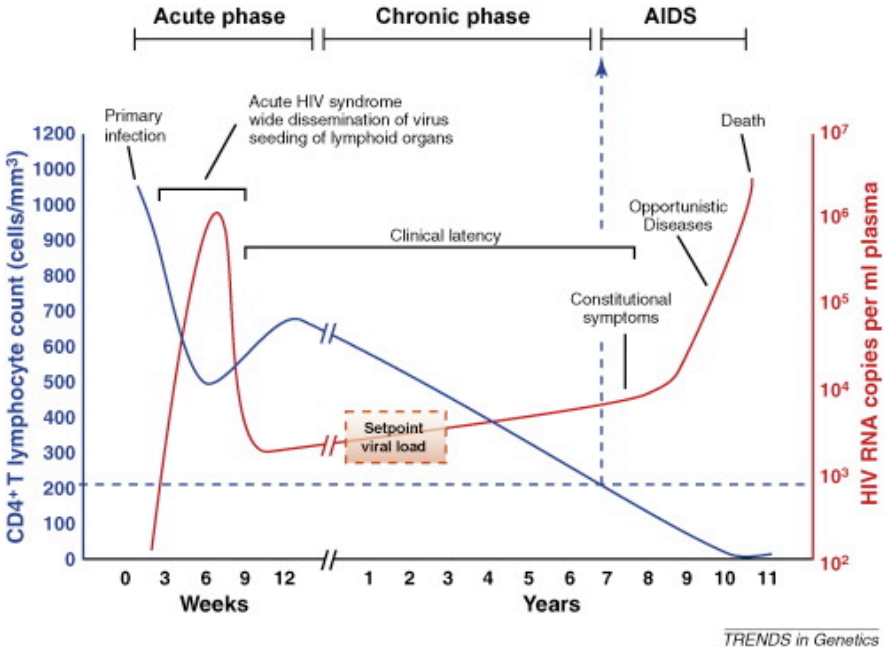



**Figure 1.7 The typical course of HIV infection.** Reproduced from (An & Winkler, 2010) with permission from Elsevier Ltd.

The chronic phase of infection is characterised by an asymptomatic period of clinical latency during which the CD4+ T cell count recovers partially from the decline during the acute phase (An & Winkler, 2010). The length of clinical latency is highly variable among individuals, but lasts 10 years on average. During chronic infection the CD4+ T cell count declines at an average rate of 50 cells/μl/year while VL gradually rises (Wolbers et al., 2010).

The final stage of untreated infection is defined by the development of AIDS; a CD4+ T cell count of less than 200 cells/μl, or AIDS-defining illness (defined by the Centre for Disease Control). Opportunistic infections and malignancies ultimately lead to death. AIDS-defining illnesses include, among others, wasting syndrome, candidiasis, CMV disease, Karposi's sarcoma, lymphoma, mycobacterial infection, and pneumocystis pneumonia. Anti-retroviral therapy (ART) has played a major role in delaying the onset of these conditions and drastically extended life

expectancy in infected individuals who have access to it. Non-AIDS comorbidities and age-related diseases related to prolonged immune activation, including liver, cardiovascular and kidney disease, and co-infections have since become important clinical features of the syndrome (Rockstroh, Guaraldi, & Deray, 2010).

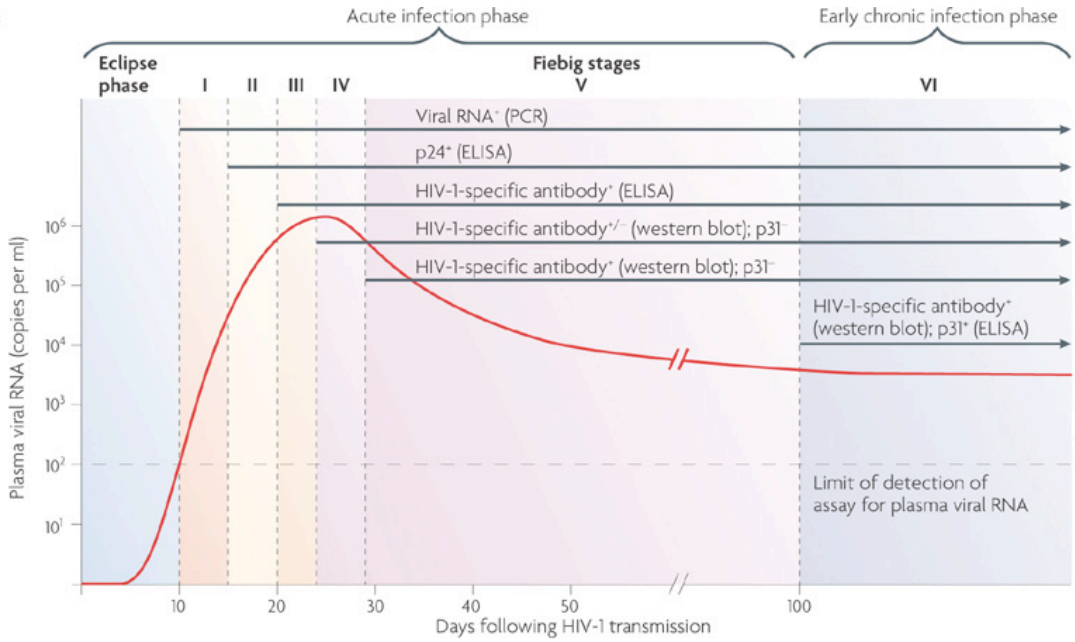



**Figure 1.8 Fiebig staging of acute HIV infection.** Reproduced from (McMichael, Borrow, Tomaras, Goonetilleke, & Haynes, 2010) with permission from Nature Publishing Group.

## 1.6 Rates of HIV disease progression

The rate at which an individual progresses from acute HIV infection to AIDS varies considerably, with periods of 1 to 35 years reported (Goulder & Walker, 2012). A rare group of HIV infected individuals termed long-term non-progressors (LTNPs) maintain CD4+ T cell counts above 500 cell/mm$^3$ and remain asymptomatic for extended period without ART (8-10 years or more following seroconversion) (Casado et al., 2010). Elite controllers (ECs) are defined as individuals who maintain HIV VLs of less than 50 copies/ml for at least one year, while

viraemic/virological controllers (VCs) maintain HIV VLs of less than 2000 copies/ml for at least one year, also in the absence of ART. Progressors are defined as individuals who advance to symptomatic disease or to ART initiation within 10 years of seroconversion, while this occurs within three years in rapid progressors (RPs) (Casado et al., 2010).[3]

Studying immune control of HIV in LTNPs, ECs and VCs has been an important approach for elucidating correlates of protective immunity against HIV infection. These represent complex phenotypes resulting from the interplay of multiple host and viral factors (McLaren & Carrington, 2015; Pereyra et al., 2008). The two major host genetic determinants of susceptibility to HIV infection or disease progression are HLA type (discussed in section 1.12) and CCR5 co-receptor expression phenotype. Individuals who are homozygous for the CCR5 deletion (CCR5Δ32) represent the only genotype consistently associated with complete protection from R5-tropic HIV infection. Resistance to HIV infection in rare groups of HIV exposed uninfected individuals (HESN) without this genotype is likely due to a more complex multi-factorial set of circumstances (Lederman et al., 2010).

## 1.7 HIV transmission

### 1.7.1 Routes of transmission and target cells

HIV is transmitted through contact of infected body fluid, sexually, from mother to child or through blood-to-blood contact. High donor VL is associated with higher rates of transmission (Quinn et al., 2000). Viral dissemination within an individual

[3] The terms LTNP, EC, VC and RP may be defined differently by some research groups. Alternative definitions are reviewed in (Migueles & Connors, 2010).

occurs not only from cell-free virions, but also by direct cell-to-cell transmission through virological synapse formation. This is an important mode by which dendritic cells (DCs) that internalise virions may spread infection to CD4+ T cells. CD4+ targets of HIV infection include T cells, monocytes/macrophages, microglial cells, DCs, and astrocytes. Infecting virions are referred to as R5- or X4-tropic[4] depending on their co-receptor usage (CCR5 and CXCR4 respectively), which in turn determines the target cell type that can be infected. DCs, monocytes/macrophages and activated CD4+ T cells are among the CCR5-expressing targets of R5-tropic HIV, so most blood-derived viruses are R5 tropic. Although both R5 and X4-tropic strains can infect primary CD4+ T cells including memory and naïve CD4+ subsets (Clapham & McKnight, 2001), expression of CXCR4 on naïve CD4+ T cells makes this subset the primary target of X4-tropic HIV. The presence of R4-tropic virus is typically associated with the later stages of HIV/AIDS.

Although there is some evidence that LTNP phenotypes may be attributed to infection with an attenuated viral strain (Claiborne et al., 2015; Mariani et al., 1996; Michael et al., 1995; Miura et al., 2010; Yue et al., 2015), this provides only a partial explanation for the phenotype. Replication competent virus can be recovered from LTNPs and ECs, and has been shown to result in progressive disease when transmitted (Bailey et al., 2008; Blankson et al., 2007; Buckheit et al., 2012; Lamine et al., 2007). These studies highlight the importance of the host immune response in establishing LTNP and EC phenotypes.

[4] R5-tropic and X4-tropic virus may also be referred to as M-tropic and T-tropic respectively

### 1.7.2 The transmission bottleneck

Approximately 80% of heterosexual HIV transmissions are thought to be established from a single transmitted founder (TF) virus that is selected from a diverse population in the infecting donor (Joseph, Swanstrom, Kashuba, & Cohen, 2015; Keele et al., 2008). A strong "bottleneck" effect thus limits transmission to a single viral variant. A number of factors contribute to this transmission bottleneck. The most transmissible phenotypes are selected first in the donor at the blood/transmission fluid barrier. The most stringent bottleneck, however, occurs in the genital tract of the recipient. Co-receptor usage is a major determinant of transmissibility, with TF viruses typically being R5 T cell-tropic (Joseph et al., 2014; Ping et al., 2013). TF viruses have also been shown to target cells more likely to be trafficked across the mucosa (Cicala, Arthos, & Fauci, 2011) and that to express higher Env concentrations for binding CD4 target molecules (Parrish et al., 2013).

A strong body of evidence supports the favourable transmission of viruses with high viral replicative capacity (VRC) during heterosexual HIV transmission. Env proteins that are closest to ancestral sequence (Herbeck et al., 2006; Sagar et al., 2009), and have reduced glycosylation (Gnanakaran et al., 2011) have been shown to be preferentially transmitted. A recent study of 137 linked transmission pairs demonstrated a selection bias in the non-Env genes (Gag, Pol and Nef) of the TF for residues that are closest to the cohort consensus, and thus confer increased viral fitness (Carlson et al., 2014). This leads to a transmission advantage in quasispecies dominated by high fitness variants (Carlson et al., 2014). However, another recent study of mother/child pairs found that VRC is slightly higher in

mothers than in children, suggesting that lower fitness virus may be transmitted in this setting (Adland et al., 2015). In addition, a more stringent bottleneck is thought to occur during heterosexual transmission than in MSM and vertical transmission (Salemi, 2013). The factors defining the transmission bottleneck may thus be influenced by the route of transmission.

## 1.8 The latent HIV reservoir

Within days of initial infection with HIV, a pool of latently infected cells, where expression of proviral DNA is silenced, is established (Whitney et al., 2014). Latently infected cells are phenotypically identical to uninfected cells, and are thus hidden from immune targeting. There is evidence that both infection of activated CD4+ T cells that subsequently return to a resting state, and direct infection of resting CD4+ T cells contribute to establishment of the reservoir (Chavez, Calvanese, & Verdin, 2015). The largest cellular reservoir resides in resting memory CD4+ T cells (Chomont et al., 2009). Latently infected cells have also been found in the central nervous system (cerebrospinal fluid cells, astrocytes and microglia) (Sturdevant et al., 2015), GALT, and genitourinary and respiratory tracts, and adipose tissues (Couturier et al., 2015) with ongoing development of technologies for identifying sites of infection likely to extend this list (Santangelo et al., 2015). Recently haematopoietic stem cells have been described as an additional reservoir (McNamara, Ganesh, & Collins, 2012). The latent reservoir represents a stable and persistent source of inducible, replication competent virus that is a major barrier to viral eradication.

## 1.9 The immune response to HIV

### 1.9.1 Establishment of infection at mucosal surfaces

The exact mechanism by which HIV breaches the mucosal barrier is an area of ongoing research. Transcytosis across columnar epithelial cells, where the virion is transported from the luminal to basal side of the cell in a coated vesicle and released into the extracellular space, is one proposed mechanism (Bhardwaj, Hladik, & Moir, 2012). A second mechanism involves infection of dendritic cells (DCs) of the squamous epithelium (Langerhans cells), through binding of CD4 and CCR5 as well as C-type lectin receptors such as DC-SIGN (dendritic cell specific ICAM3-grabbing non-integrin) and langerin (Bhardwaj et al., 2012). The dendrites of DCs may extend through epithelial tight junctions towards the luminal surface, and may thus be the first leukocyte encountered by an HIV virion. Binding of gp120 to DC-SIGN induces phagocytosis of the virion/DC-SIGN complex. In this way, HIV can be transferred to CD4+ T cells and stromal DCs without directly infecting the first DCs contacted (Geijtenbeek et al., 2015). CCR5+ macrophages at the mucosal surface are another likely target of initial infection (Bhardwaj et al., 2012). Transmission is enhanced by physical damage to the mucosal epithelium allowing easy passage of the virion across the mucosa. Local amplification of a single focus of CD4+ T cells establishes infection and DCs bearing HIV migrate to the lymph nodes allowing systemic spread (McMichael et al., 2010).

### 1.9.2 Innate immunity

The earliest stages of the immune response to HIV infection comprise a rise in acute-phase proteins and an intense “cytokine storm” characterised by a rapid

increase in IL-15, type I IFNs, CXCL10, IL-18, TNF, IFN-Ɣ, IL-22, and IL-10. Although some of this response may promote viral clearance, it is also associated with promoting viral spread and immunopathology (McMichael et al., 2010).

Both DCs and macrophages are activated by pattern recognition receptor (PRR) binding, either DC-SIGN recognition of carbohydrate moieties on gp120 or TLR7 recognition of single-stranded viral RNA. Activated DCs migrate to the lymph nodes and present HIV-specific peptides to T cells, priming the adaptive immune response. Activated macrophages also present viral peptides and produce restriction factors and chemotactic cytokines.

Natural killer (NK) cells are important players in immunity to HIV that straddle the innate and adaptive response. NK cells have been shown to suppress replication *in vitro* in a dose- and cell contact-dependent manner (Alter et al., 2007). They play a role in lysis of virally infected cells through direct recognition of viral proteins or virus-induced stress ligands. Infected cells with low expression of MHC (downregulated by Nef) are also targeted for killing by NK cells through loss of NK inhibitory receptor signaling from MHC binding. NK cells also secrete pro-inflammatory cytokines, and mediate antibody-dependent cellular cytotoxicity (ADCC), and modulate the adaptive immune response (Altfeld & Gale, 2015). Evidence that immune pressure drives selection of viral escape from the NK cell killer immunoglobulin-like receptors (KIRs) highlights the importance of the NK cell response to HIV (Alter et al., 2011).

### 1.9.3 Humoral immunity

HIV-specific antibodies (Abs) can be classed as non-neutralising, neutralising (nAbs), that mediate loss of viral infectivity, and broadly neutralising (bnAbs), that neutralise a broad range of HIV strains (Burton & Mascola, 2015). nAbs target the native Env trimer, including the gp120 variable loops, CD4 binding site and membrane-proximal external region, while non-neutralising Abs typically bind other regions, or Env epitopes expressed on isolated proteins but not on the native trimer (Burton & Mascola, 2015).

The earliest detectable antibodies following HIV infection tend to be non-neutralising and do not select for viral escape (McMichael et al., 2010). These function through ADCC, Ab-mediated complement and opsonisation for phagocytosis of virions. nAbs that neutralise autologous virus begin to appear 3-36 months post-infection, and develop in a large proportion of infected individuals (McMichael et al., 2010; Richman, Wrin, Little, & Petropoulos, 2003; Wei et al., 2003). bnAbs are very rarely produced and only appear after 20-30 months of infection. This is due, in part, to the requirement for extensive hypermutation that is slowed by an impaired CD4+ T cell response (Derdeyn, Moore, & Morris, 2014). Recently described bnAbs include the CD4 binding site Abs VRC01 (Zhou et al., 2010) and 3BNC11 (Scheid et al., 2011), and the Env V1V2 binding Ab CAP256 Abs (Doria-Rose et al., 2014). The 3BNC117 and VRC01 Abs have been used in the first phase I trials of passive immunisation with a bnAb, and the former showing a transient reduction in viraemia (Caskey et al., 2015; Ledgerwood et al., 2015). Amongst others the Env membrane proximal external region is another target of bnAbs.

HIV has evolved a number of strategies to evade the Ab response. nAb binding sites are often inaccessible and protected by the surface glycan shield. The delay in mounting an Ab response to the rapidly evolving virus means that the virus has often selected escape mutations by the time Abs to a particular variant are produced, rendering these Abs ineffective. The presentation of self-epitopes derived from budding through the host cell, and exposure of decoy epitopes that produce less effective Abs also contribute to viral evasion of the Ab response.

### 1.9.4 CD8+ T cell-mediated immunity

A considerable body of evidence supports the role of the CTL response in HIV immune control (Goulder & Walker, 2012; Migueles & Connors, 2015). The decline to viral setpoint following peak viraemia is associated temporally with the emergence of the HIV-specific CTL responses (Borrow, Lewicki, & Hahn, 1994; Koup, Safrit, & Cao, 1994), and the rate and magnitude of CTL activation in response to peak viraemia may influence this setpoint (Ndhlovu et al., 2015). In SIV-infected NHPs, administration of anti-CD8 Abs inhibits this decline, and significantly increases VL in chronic infection (Chowdhury et al., 2015; Matano et al., 1998; Schmitz et al., 1999).

The CTL responses that dominate during acute and chronic infection may be substantially different (Goulder, Altfeld, et al., 2001). The earliest T cell responses are often specific for Gag, Nef and, to a lesser extent, Env, and are most important for the decline in VL to setpoint in acute infection. Following peak viraemia, rapid selection of CTL escape mutations begins to occur, the early CTL response wanes

rapidly and the TF viral sequence is lost as the virus rapidly evolves (Goonetilleke et al., 2009). A second wave of CTLs against conserved proteins (Gag and Pol) maintains the VL following establishment of the setpoint (Goonetilleke et al., 2009; Wang et al., 2009). Earlier targeting of conserved epitopes (particularly in Gag) in acute infection is likely to result in a lower viral setpoint (Streeck et al., 2009) and slower disease progression during the chronic phase of infection (Altfeld et al., 2006; Streeck et al., 2014).

A timeline of the earliest immune responses to HIV is given in Figure 1.9.

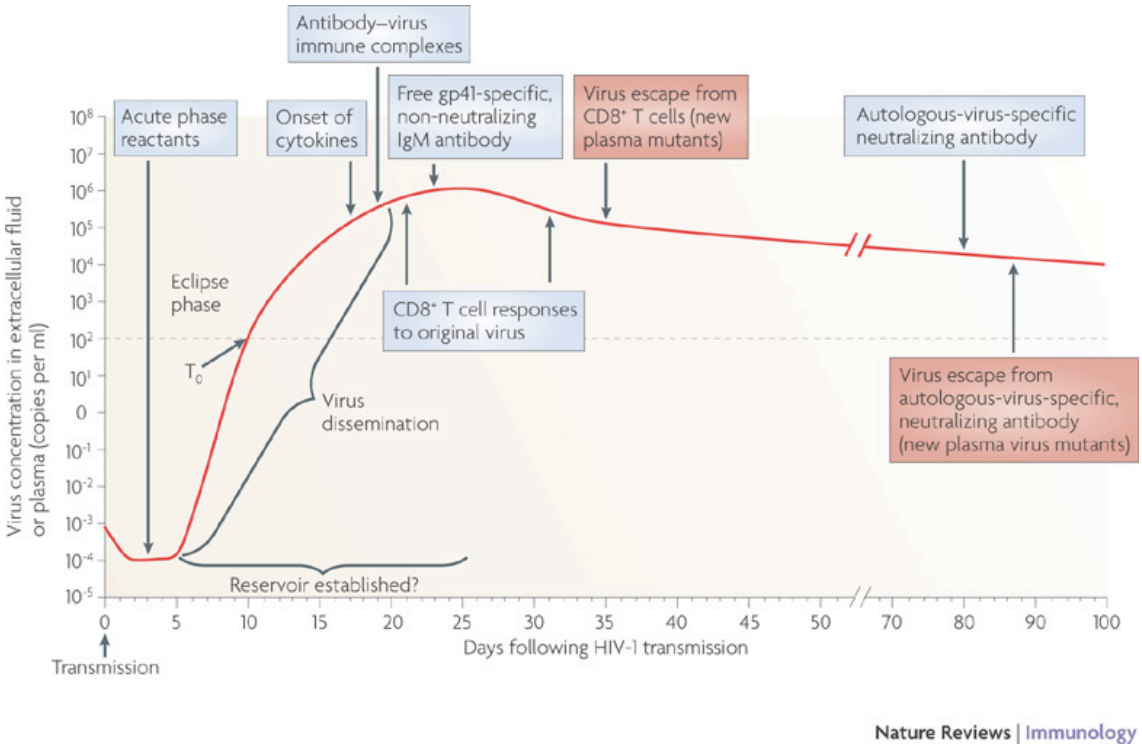



**Figure 1.9 Timeline of the early immune responses to HIV.** Acute phase proteins are detectable when virus replication is largely still limited to mucosal tissues. A cytokine storm accompanies detection of virus in the plasma. The earliest non-neutralising Ab and CTL responses are mounted just prior to peak viraemia. Escape from the earliest CTL response can occur within as little as 10 days after initial T cell expansion. The earliest nAbs can be detected at about day 80 post-infection as VL is declining to setpoint. nAb escape begin within a week of this. Reproduced from (McMichael et al., 2010) with permission from Nature Publishing Group.

## 1.10 Immunopathology of HIV/AIDS

### 1.10.1 CD4+ T cell depletion

HIV disease progression is associated with substantial immune dysfunction, the earliest and strongest of which is substantial depletion of CD4+ T cells. This is particularly apparent in the GALT, where as much as 80% of CD4+ T cells can be depleted within the first 3 weeks of infection without later recovery (Brenchley et al., 2004; Mattapallil et al., 2005). The mechanism of depletion of CD4+ T cells is incompletely understood. Although the primary method of viral release from infected cells involves budding rather than lysis, direct killing of infected cells occurs by virus-induced apoptosis. Infected CD4+ cells are also targets of CTL- and NK-mediated killing. CD4+ T cell death is not, however, limited to infected cells. Activation induced cell death (AICD), a normal consequence of T cell activation, may contribute to depletion of activated bystander CD4+ T cells. [5]

### 1.10.2 Immune activation and T cell exhaustion

Extensive immune activation of innate cells, B cells and T cells that begins in acute infection and persists in chronic infection, and is not limited to HIV-specific cells, represents a second major cause of immunopathology. This is characterised by increased T cell turn over and expression of activation markers (CD69, CD38, Ki57, CD25, HLA-DR), an increase in pro-inflammatory cytokines and chemokines (particularly IFN-α produced by DCs) and macrophage dysfunction. Release of

[5] An interesting hypothesis proposes that binding of gp120 to CD4 on uninfected cells disrupts immunological synapse formation by disrupting signalling through the CD4/TCR/MHC complex. This may contribute to premature T cell death. (Suchard, M, personal communication)

apoptotic microparticles and increased expression of apoptotic ligands (Fas ligand and TRAIL) causes extensive bystander apoptosis of B and T cells..

Increased gut permeability due to GALT depletion, leading to microbial translocation and increased levels of circulating lipopolysaccharide (LPS) has been linked to generalised immune activation through chronic PRR stimulation (Brenchley et al., 2006). The loss of Tregs (regulatory T cells), which are largely CD4+, may also contribute. Recurrent antigen stimulation in the presence of a dysregulated immune system is an alternative hypothesis for inducing immune activation (Boasso & Shearer, 2008), although the persistence of residual immune activation in ART treated patients with suppressed viraemia suggests that it is not driven solely by high viral and antigen load (Hunt et al., 2003).

The level of immune activation is a strong predictor of the rate of HIV disease progression (Deeks et al., 2004). Control of HIV has been associated with lower levels of microbial translocation and immune activation (Choudhary et al., 2007). Non-pathogenic SIV infection of natural hosts (sooty mangabeys) is characterised by low levels of both immune activation and bystander apoptosis, which likely protects the host from progressive disease (Bosinger et al., 2009; Silvestri et al., 2003). Persistent antigen exposure and immune activation lead to T cell exhaustion, characterised by loss of T cell proliferative capacity and effector functions, and expression of negative regulatory molecules (TIM3, PD1, CTLA4), and is associated with disease progression (Day et al., 2006).

### 1.10.3 Immunodeficiency

CD4+ T cells are at the centre of directing both the innate and adaptive immune response, but are extensively depleted during HIV infection. In the setting of loss of CD4+ T helper function required for B cell proliferation, Ab production and Ab class-switch, circulating memory B cells are severely depleted. HIV infection is also associated with lysis of follicular B cells and B cell apoptosis, and early loss (up to 50% within the first 80 days) of the germinal centres, particularly in the gut (McMichael et al., 2010). CD8+ T cell dysfunction related to exhaustion, lack of CD4+ T cell help and DC dysfunction occurs. Thus, in HIV infection, major players of the immune response are compromised, resulting in immunodeficiency and susceptibility to opportunistic infections and malignancy, especially in the late stages of untreated disease (Okoye & Picker, 2013). Paradoxically, the immune system is extensively activated at the same time as infected individuals experience immunodeficiency. HIV disease progression occurs in the setting of vigorous anti-HIV immune responses that control neither HIV nor opportunistic infections.

## 1.11 HIV vaccine approaches

### 1.11.1 Prophylactic vaccine strategies

Designing an effective vaccine for HIV-1 has proved a significant challenge that is yet to be resolved. There is an enormous scientific challenge involved in designing a vaccine for such a diverse and continually evolving target. Integration into the host genome and establishment of latency, immune evasion and the lack of a clear immune correlate of protection in natural infection, present significant barriers to the HIV vaccine endeavour. Furthermore, there is a lack of robust animal models

for testing interventions that consistently predict outcome in humans, and very large and expensive studies are required for measuring the incidence of new infections as a study endpoint (Esparza, 2013).

The main approaches to HIV vaccine development have attempted to induce nAbs using Env subunit vaccines or CTL responses using DNA and viral vectored vaccines. A small number of vaccines have entered efficacy trials (clinical trial stage II/III) (Table 1.1). The AIDSVAX gp120 vaccine produced non-neutralising Abs to Env that failed to protect against subsequent infection (Pitisuttithum et al., 2006). The Step and Phambili trials used attenuated adenovirus serotype 5 (Ad5) vectors expressing Gag, Pol and Nef to stimulate CTL responses, but also demonstrated no protective effect (Buchbinder et al., 2008; Gray et al., 2011). Approaches using Ad5 vectors expressing Gag, Pol, Nef and Env following initial priming with a DNA vaccine (HTVN505) have also been unsuccessful (Hammer et al., 2013).

The RV144 trial (ALVAC/AIDSVAC) used a canary pox viral vector expressing gp120, Gag and Pol in a prime regimen, followed by boosting with the AIDSVAX gp120 vaccine. This showed a moderate protective effect with 30% fewer recipients becoming infected following vaccination (Rerks-Ngarm et al., 2009). Production of IgG antibodies to the V1V2 region of gp120 that mediate ADCC were found to be an important correlate of this protection (Haynes et al., 2012; Kim, Excler, & Michael, 2015), while Env-specific IgA was associated with an increased risk of acquisition possibly related to HLA class II expression (Prentice et al., 2015). This highlighted the importance of non-neutralising antibodies for vaccine

approaches. Trials in South Africa (HVTN097) are underway to determine whether a similar strategy targeted using C clade antigens might be effective.

**Table 1.1 Clinical HIV vaccine efficacy trials.** Reproduced with permission from (Barouch & Michael, 2014).

| Study | Vaccines | Phase | Risk Group | HIV Incidence per 100 Person-Years | Location | Result |
|---|---|---|---|---|---|---|
| Vax003 | AIDSVAX B/E gp120 in alum | III | injecting drug users | 3.40% | Thailand | no vaccine efficacy |
| Vax004 | AIDSVAX B/B gp120 in alum | III | high-risk women and MSM | 2.60% | United States, Europe | no vaccine efficacy |
| HVTN 502 Step | MRKAd5 HIV-1 gag/pol/nef B | IIb | high-risk women and MSM | 3.00% | United States | halted at interim analysis for futility; early transient increased infection in vaccinees |
| HVTN 503 Phambili | MRKAd5 HIV-1 gag/pol/nef B | IIb | high-risk heterosexual men and women | 3.70% | South Africa | no vaccine efficacy; late increased HIV infection in unblinded male vaccinees |
| RV144 | ALVAC-HIV vCP1521, AIDSVAX B/E rgp120 in alum | III | community risk heterosexual men and women | 0.28% | Thailand | 31.2% efficacy at 42 months as primary endpoint; 60% efficacy at 12 months |
| HVTN 505 | DNA, rAd5 (A, B, C) | IIb | circumcised MSM without pre-existing Ad5 antibodies | 1.80% | United States | halted at interim analysis for futility |

MSM, men who have sex with men; Ad5, adenovirus serotype 5.

The most promising advancement in the field of CTL vaccine development recently has come from work in rhesus macaques vaccinated with rhesus CMV-based vectors. The use of a chronically infecting viral vector is intended to result in persistent stimulation of the CTL response, allowing an earlier response to HIV challenge than one that has to be recalled from memory cell populations. Rapidly replicating mucosal infections require a rapid response to afford protection, and this is particularly important in HIV where there is rapid integration of the genome and establishment of the latent reservoir. These vectors express Gag, Rev, Tat, Nef, and Env. (Hansen et al., 2009, 2011). In 50% of vaccinated animals infected with challenge virus, VL is reduced to undetectable levels (Hansen et al., 2013). Interestingly, at least half the response was shown to be mediated by unconventional MHC-II-restricted CTL (Hansen et al., 2013).

## 1.12 Human leukocyte antigen in HIV infection

### 1.12.1 HLA class I biology and peptide presentation

The HLA class I molecule comprises two polypeptide chains; the α heavy chain, comprising the $\alpha_1$, $\alpha_2$ and $\alpha_3$ domains (encoded on chromosome 6) and the invariant light chain $\beta_2$ microglobulin (Figure 1.10). The peptide binding groove, formed by $\alpha_1$, and $\alpha_2$ domains, accommodates peptides 8-11 amino acids in length for presentation to CD8+ T cells. Polymorphic residues in the peptide binding groove form the binding pockets A-F to which anchor residues in the presented peptide bind. Peptide binding pockets B and F represent anchor positions to which the amino acid residue at position 2 and at the C terminus of the peptide bind respectively. Specific binding of peptide anchor residues to the HLA-binding pockets determines which peptides can be presented by a particular HLA molecule, such that there is a specific peptide motif that binds each HLA allele. The peptide binding motifs for the most important HLA types in this thesis are shown in (Table 1.2).

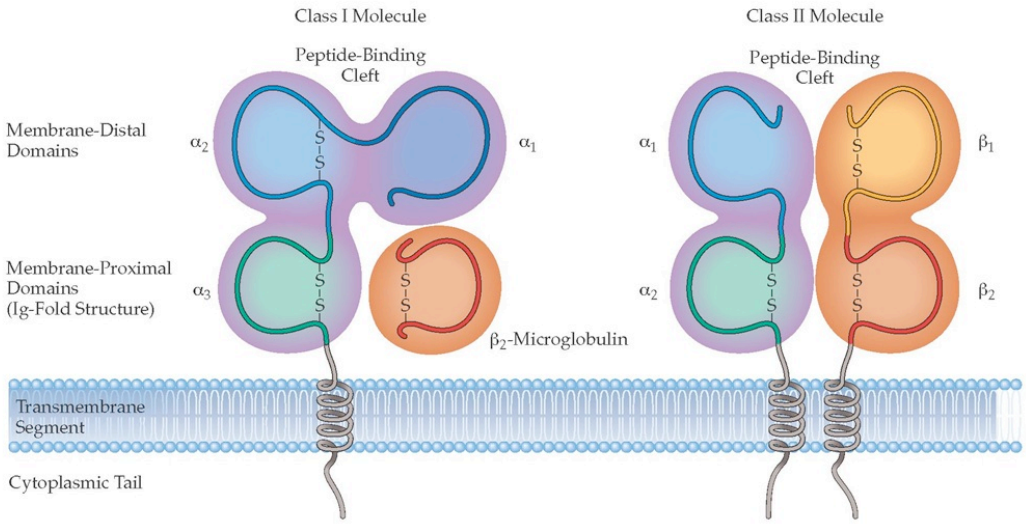



**Figure 1.10 The structure of the HLA class I and class II molecule.** Reproduced from http://what-when-how.com/acp-medicine/adaptive-immunity-histocompatibility-antigens-and-immune-response-genes-part-1/

The classical HLA class I genes are encoded at three loci on chromosome 6 for expression of the HLA-A, B and C genes respectively. These genes are highly polymorphic[6] in the human population, with HLA-B representing the most diverse region. Most individuals are heterozygous at the HLA-A, -B, and -C loci and thus express six different class I molecules, with one set inherited from each parent.

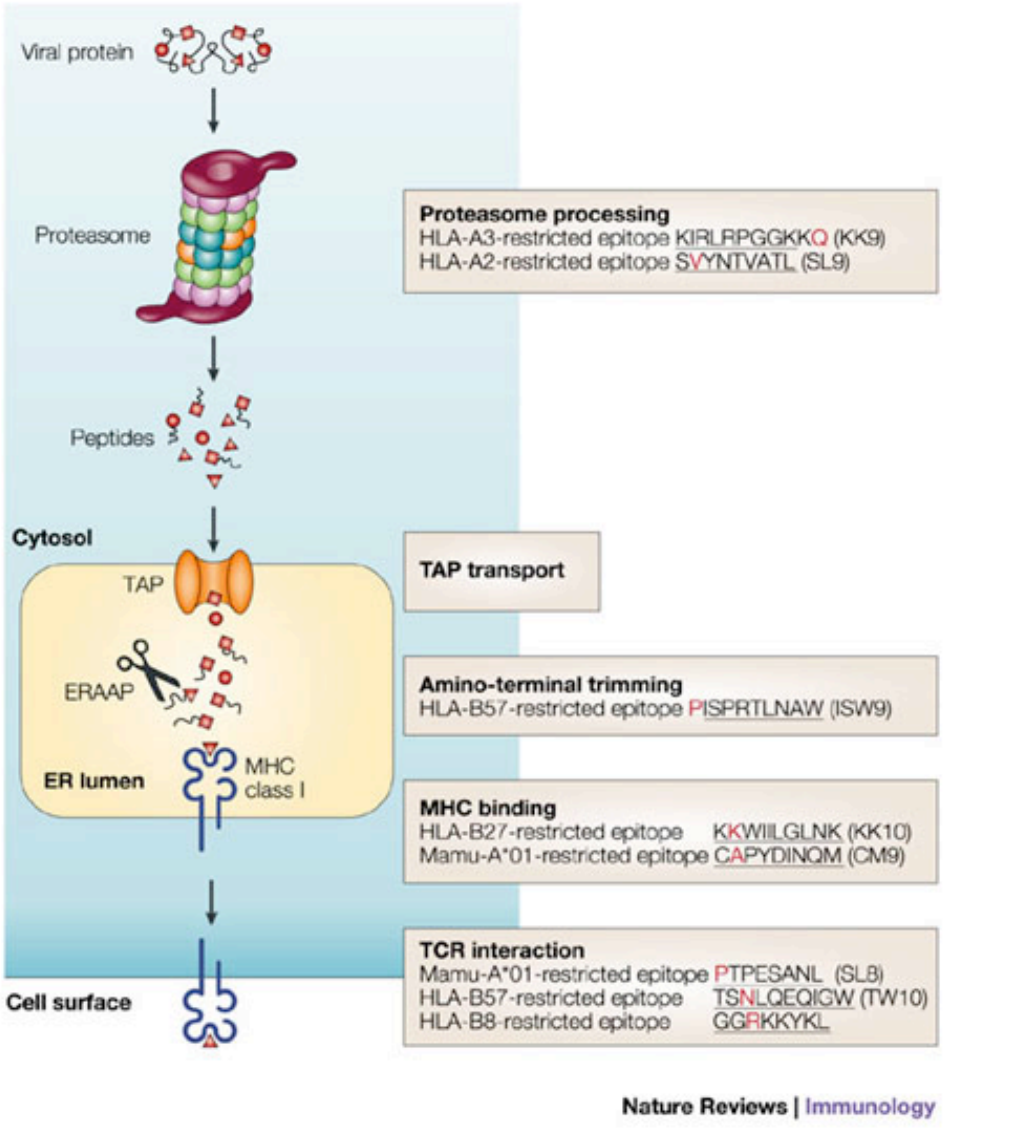



**Figure 1.11 Processing of HIV peptides for presentation on HLA class I and steps at which escape mutations affect the process.** Example epitopes with escape mutations highlighted in red are shown. Reproduced from (Goulder & Watkins, 2004) with permission from Nature Publishing Group.

[6] Currently, there are 3192, 3977 and 2740 HLA-A, B and C allele sequences respectively listed in the IMGT/HLA database (Robinson et al., 2015).

**Table 1.2 HLA-class I peptide binding motifs.** Binding of peptides to HLA-B*27:05 involves hydrogen bonding between the conserved arginine at position 2 of the peptide and HLA B pocket and binding of a hydrophobic or positively charged residue at the C terminus of the peptide to the F pocket (Madden, Gorga, Strominger, & Wiley, 1991; Stewart-Jones et al., 2005). In HLA-B*57:01-peptide interactions there is a strong preference for binding of a C terminal tryptophan to the F pocket of the HLA. Table adapted from (Goulder & Walker, 2012).

| **HLA Binding pocket:** | **A** | **B** | **D** | **C** | **F** | |
|---|---|---|---|---|---|---|
| **Peptide position:** | **1** | **2** | **3** | **5-7** | **C** | |
| **HLA-B*27:05** | R<br>K | R | Y<br>F<br>W<br>I | I<br>L<br>P<br>V | R<br>K<br>H<br>Y<br>F<br>M<br>L<br>I | (hydrophobic residues at C terminus) |
| **HLA-B*57:01** | K | A<br>T<br>S | K<br>R<br>L<br>F<br>Y | | F<br>W | (strong preference for tryptophan (W) at C terminus) |

Proteasomal degradation of viral proteins in infected cells occurs as part of normal processing and presentation of cellular proteins (Figure 1.11). The peptides produced are transported by TAP1 and TAP2 (transporters associated with antigen processing) proteins into the endoplasmic reticulum (ER). In the ER, ERAAP (ER aminopeptidase associated with antigen processing) trims the peptides to the ideal length for association with HLA. The HLA-peptide complex is assembled in the ER. Newly synthesised partially folded HLA heavy chains are bound to the chaperone protein calnexin, which dissociates upon light chain binding. The HLA molecule associates with Erp57, calreticulin and tapasin (TAP associated protein) chaperone proteins. Tapasin acts as a bridge between the HLA and TAP transporter for transferring peptides to the HLA binding groove. The fully folded HLA-peptide complex disassociates from the chaperones and is transported out of the ER and presented on the cell surface.

The HLA alleles can be classified into supertypes (such as the B7 supertype that includes HLA-B*07:02, B*39:10, B*42:01, B*42:02, B*81:01) that have similar binding motifs, and can sometimes bind the same peptides (Lund et al., 2004). The B alleles can be further classified into the Bw4 and Bw6 groups based on specificity of peptide binding at the C terminus to the F pocket. Most (but not all) class I HLA alleles that are protective in HIV infection are Bw4 (Martin et al., 2002). There are four Bw4 motifs formed by amino acid 77-83 of the class I HLA heavy chain. These motifs are involved in recognition of the HLA by the NK KIRs.

### 1.12.2 HLA associations with HIV disease outcome

HLA class I variation has been consistently linked to differential disease outcome in HIV infection (Bartha et al., 2013; Fellay et al., 2007; Kaslow et al., 1996; Kiepiela et al., 2004; Leslie et al., 2010; McLaren et al., 2015; Pereyra et al., 2010). In several genome wide association studies (GWAS), the HLA class I alleles have been shown to be the most significant genetic determinant of HIV disease outcome (Dalmasso et al., 2008; Fellay et al., 2007, 2009; Limou et al., 2009). This association is particularly strong for the HLA-B alleles (Kiepiela et al., 2004). In a GWAS study of the individual amino acids associated with control of HIV, variation at amino acids in the HLA-B binding groove that are likely to affect peptide binding were associated with both disease control and disease progression (Pereyra et al., 2010).

### 1.12.3 KIRs in HLA associations with HIV disease outcome

One mechanism proposed for the strong association of HLA with HIV disease outcome relates to the interactions of HLA with the NK cell KIRs. The most important HLA-KIR interactions for determining HIV disease outcome are those involving KIR3DL1 and KIR3DS1. KIR3DL1 is an inhibitory receptor that binds the Bw4 motif on HLA class I. A lack of signaling through KIR3DL1 on infected cells with down-regulated HLA expression targets these cells for NK-mediated killing. An advantage in HIV outcome is associated with expression of both KIR3DL1 or the activating receptor KIR3DS1 and an HLA Bw4 motif with isoleucine at position 80 of the heavy chain (Martin et al., 2007).

### 1.12.4 HLA-mediated CTL responses in HIV disease outcome

There is a substantial body of evidence that demonstrates HLA-specific effects on the CTL response in determining HIV disease outcome. Quantification of the breadth and magnitude of the CTL response alone does not predict immune control of viraemia (Addo et al., 2003). A growing number of studies have attempted to define additional correlates of a protective CTL response. CTL proliferative capacity (Day et al., 2007; Migueles et al., 2002), polyfunctionality (Almeida et al., 2007; Betts et al., 2006), functional avidity (antigen sensitivity) (Almeida et al., 2009; Bennett, Ng, Dagarag, Ali, & Yang, 2007), cytotoxic capacity (Hersperger et al., 2010; Migueles et al., 2002; Migueles et al., 2008, 2015) and immune regulation (Elahi et al., 2011; Kaufmann et al., 2007) are among the proposed correlates. Since peptide presentation on HLA class I is critical for CTL activation, HLA class I is likely to be an important determinant of these features. Proliferative capacity has a strong association with disease outcome. The highest levels of CTL proliferation have been associated with the most protective HLA alleles (Horton et al., 2006; Migueles et al., 2002). It is however, important to consider whether a particular feature of the CTL response in LTNPs is a cause or a consequence of low VL (Migueles & Connors, 2015). Proliferation and polyfunctionality in particular may be determined by antigen load (Streeck, Brumme, et al., 2008). The degree to which features of the CTL response are restored in the setting of low viraemia on ART may shed light on this (Migueles et al., 2009).

A consistent and somewhat predictable pattern of immunodominance during HIV infection in individuals with a particular HLA type has been described (Draenert et

al., 2006). Favourable HLA alleles are associated with immunodominance hierarchies that may mediate advantageous outcome (Goulder et al., 1996; Schneidewind et al., 2008). A shift in this immunodominance hierarchy may have consequences for disease outcome (Goulder, Sewell, et al., 1997), and redirecting the hierarchy to target the most vulnerable regions of the HIV genome may be an important aim to consider in designing CTL vaccines (Hancock et al., 2015).

TCR usage may be another important determinant of HLA-mediated disease outcome, although there are conflicting data on the nature of the most effective TCR repertoire. Studies have linked both a preference for and absence of 'public'[7] clonotypes (Chen et al., 2012; Gillespie et al., 2006; Iglesias et al., 2011; Mendoza et al., 2012; Price et al., 2009; Yu et al., 2007), diversity in the repertoire (Kosmrlj et al., 2010; Yu et al., 2007) and cross-reactivity to variants (Chen et al., 2012) to favourable HIV disease outcome. Recent work has linked TCR usage and the immunodominance hierarchy, representing an additional mechanism for how the TCR repertoire determines disease outcome (Kloverpris et al., 2015). Regardless of the mechanism, evidence of parallel evolution of the virus and TCR repertoire highlight the importance of TCR usage in immune control (Costa et al., 2015).

### 1.12.5 Epitope specificity in HLA associations with HIV disease outcome

There is evidence to suggest that, apart from the effect of the HLA-peptide complex in determining the nature of the CTL response, the origin and specificity of the presented epitope itself plays an important role. Responses to specific HIV proteins, especially Gag, have been linked to lower VLs in chronic HIV infection

[7] Shared TCR clonotypes present in many individuals.

(Dinges et al., 2010; Edwards et al., 2002; Geldmacher et al., 2007; Kiepiela et al., 2007; Masemola et al., 2004; Mothe et al., 2011; Sáez-Cirión et al., 2009; Streeck et al., 2007). In part, this may be due to the fact that Gag is abundantly expressed within the infecting virus, and does not require *de novo* synthesis following infection (Payne et al., 2010). Early HLA-mediated presentation of Gag epitopes may be important in recruiting immune responses early after a cell has become infected, and means CTL responses can take effect before Nef-mediated downregulation of HLA-A and –B (Borghans, Mølgaard, de Boer, & Keşmir, 2007; Brockman et al., 2010; Sacha et al., 2007).

A high degree of sequence conservation, particularly in the *p24* gene, is another explanation for the benefit of Gag-specific responses. This conservation is preserved by a considerable cost to viral fitness associated with selection of mutations in this gene (Crawford et al., 2009). CTL directed at Gag are thus particularly difficult for the virus to evade through selection of escape mutations (Troyer et al., 2009). Many HLA alleles that mediate a particularly favourable immune response are thus associated with CTL responses to highly conserved Gag epitopes (Boutwell, Rowley, & Essex, 2009; Brockman et al., 2007; Crawford et al., 2009; Kloverpris et al., 2012; Martinez-picado et al., 2006; Matthews et al., 2008; Miura, Brockman, Schneidewind, et al., 2009; Schneidewind et al., 2007).

Targeting of Gag alone does not, however, ensure disease control, since non-controllers also target Gag with comparable breadth and magnitude (Migueles et al., 2015; Mothe et al., 2012) and epitope specificity does not explain immune control in LTNPs who lack protective alleles.

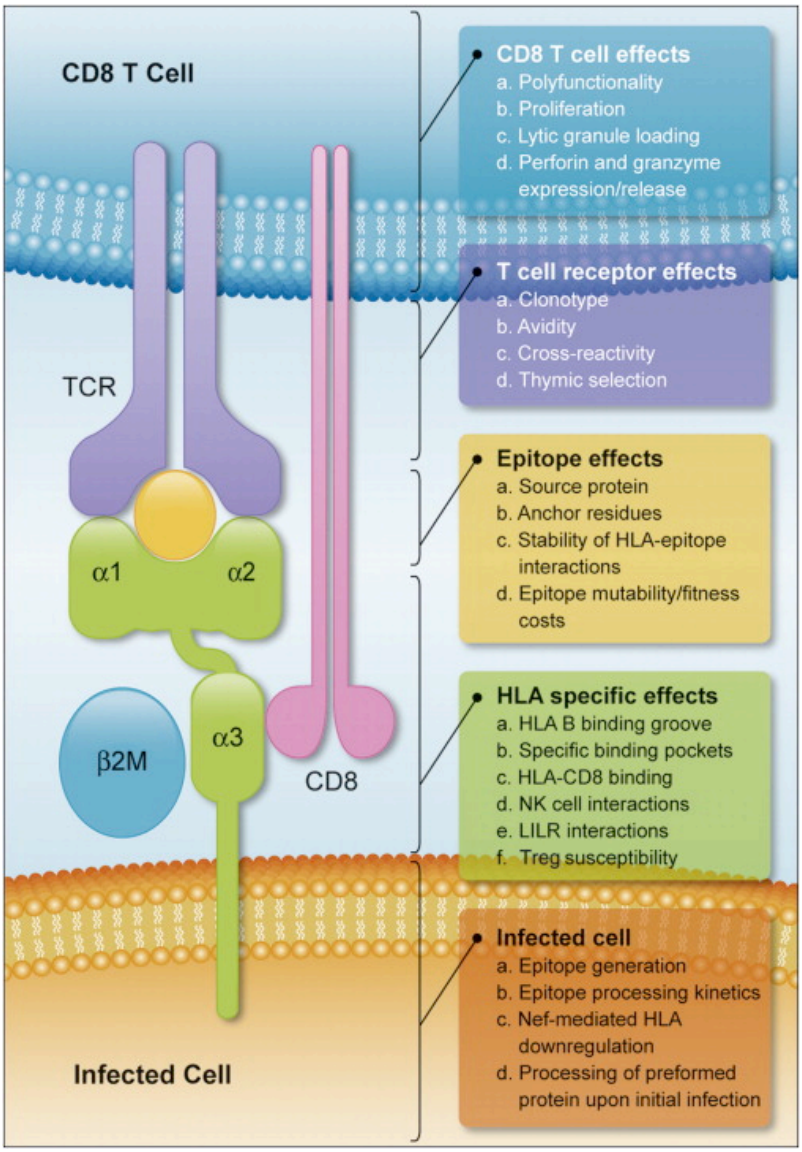



**Figure 1.12 Mechanisms of HLA-mediated control of HIV.** Reproduced with permission from (Goulder & Walker, 2012).

## 1.13 The effect of CTL and HLA on HIV evolution and adaptation

### 1.13.1 CTL immune escape

Evasion of the CTL-mediated immune response through selection of escape mutations in HIV epitopes was first described in 1990's (Borrow et al., 1997; Phillips et al., 1991). Escape mutations may affect proteasomal processing, peptide trimming by ERAAP, MHC binding, or TCR recognition (Figure 1.11). These escape mutations often incur a fitness cost to the virus (Crawford et al., 2009; Martinez-

picado et al., 2006; Miura, Brockman, Schneidewind, et al., 2009; Peyerl et al., 2004; Schneidewind et al., 2007; Troyer et al., 2009) and acquisition involves a balance between the selection pressure driving CTL escape and the pressure to maintain the viral sequence with optimal fitness. Highly conserved epitopes in which escape incurs a considerable fitness cost are particularly effective targets of CTL, but are also a source of selective pressure driving mutation in targeted epitopes. Where selection of escape is too costly to the virus to be selected alone, a compensatory mutation that restores viral fitness and stabilises the virus may accompany escape. CTL-driven escape and compensatory mutation thus drive the adaptation of HIV to the host immune response at a rate dependent on the strength of immune pressure and viral population size (Figure 1.13). The rate at which specific escape mutations are selected (and revert upon transmission) varies across genes and across HLA types (Brumme et al., 2008) reflecting the complexity of the selective pressures at play.

Predictable patterns of escape mutations or 'footprints' are associated with particular HLA class I alleles (Leslie et al., 2004; Matthews et al., 2009). In a GWAS study, paired viral data was included to identify associations between host genetic polymorphisms and HIV-1 sequence (Bartha et al., 2013). All significant SNP associations with HIV amino acid variants mapped to the HLA class I region, highlighting the strong influence of HLA genes on viral sequence (Bartha et al., 2013). The amino acid sites associated with adaptation to HLA have been identified across the full B clade HIV genome (Carlson, Listgarten, et al., 2012).

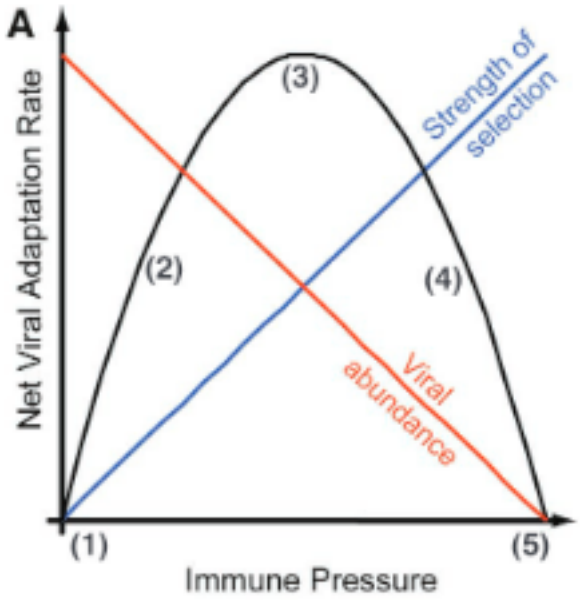



**Figure 1.13 Relationship between the immune response, viral adaptation and viral load.** Low immune pressure is associated with low rates of viral adaptation (2). Intermediate immune pressure promotes the fastest rate of viral adaptation (3). High immune pressure limits viral turn over and opportunity for mutation, and is thus associated with lower rates of viral adaptation (4). Reproduced from (Grenfell et al., 2004).

There are two possible consequences of CTL escape with regard to HIV disease outcome. Accumulation of mutations that are detrimental to viral fitness or replicative capacity may contribute to control of viraemia, particularly in an HLA-mismatched transmission partner whose CTL response targets different epitopes (Boutwell et al., 2009; Chopera et al., 2008; Goepfert et al., 2008; Martinez-picado et al., 2006; Matthews et al., 2009). Alternatively, escape from CTL mutations that control viraemia may lead to loss of immune control and disease progression (Ammaranond et al., 2011; Crawford et al., 2009; Goulder, Brander, et al., 2001; Goulder, Phillips, et al., 1997).

### 1.13.2 Transmission of CTL escape mutations to HLA-mismatched donors

Reversion of escape mutations upon transmission to an HLA mismatched recipient (thus relieving specific HLA-mediated selection pressure) is one possible consequence of transmission of such mutations. This provides strong evidence that these mutations incur a considerable fitness cost to the virus (Crawford et al., 2009; Friedrich et al., 2004; Leslie et al., 2004; Matthews et al., 2008) and promotes the maintenance of consensus HIV sequences at the population level (Eriksson et al., 2015). While transmission of weakened viruses with escape may benefit the recipient, in partners who share HLA alleles, transmission of pre-adapted virus may result in poor outcome (Goepfert et al., 2008). This is seen most frequently in mother-to-child transmission (Thobakgale et al., 2009). In addition to reversion driven by pressure to maximise viral fitness, recent evidence suggests that reversion to consensus may actively be driven by the recipient's CTL response to the mutated sequence (Eriksson et al., 2015).

Transmission of a costly escape mutation in combination with a compensatory mutation that restores the fitness cost may mitigate the need for reversion in the HLA-mismatched recipient, allowing stable transmission (Brockman et al., 2007; Crawford et al., 2007; Schneidewind et al., 2008, 2009). Persistence of mutations without compensation suggests a limited impact on viral fitness.

### 1.13.3 Adaptation of HIV to HLA at the population level

The stable transmission of mutations may lead to adaptation of the circulating virus in a population as a whole. A body of evidence suggests that population-specific differences in HIV sequences can be attributed to the accumulation of

polymorphisms driven by prevalent HLA class I alleles in that population (Chikata et al., 2014; Cotton et al., 2014; Juarez-Molina et al., 2014; Karakas, Brumme, & Poon, 2015; Kawashima et al., 2009; Kinloch et al., 2015; Schellens et al., 2011). Clade specific differences, for example, may reflect the adaptation of HIV to HLA at the population level. In the case of protective HLA types, this adaptation may mitigate the protective effects of these alleles over time, but may also contribute to lowering the viral replicative capacity of the virus at the population level (Nomura et al., 2013; Payne et al., 2014).

One example is the CRF01_AE clade subtype, which closely resembles an HLA-B*57/B*58:01 escape mutant. The Gag CRF01_AE consensus bears the polymorphisms which match HLA-B*57/B*58:01-selected footprints including A146P, I147L, A163G and S165N (Buranapraditkun et al., 2011; Matthews et al., 2009). Although a founder effect is one possible explanation for this pattern of polymorphisms in the CRF01_AE clade sequence, the high frequency of B*58:01 (e.g. 17.6% in a recently published Thai cohort of 557 individuals (Mori et al., 2014)) seems a likely driving force. It is unsurprising, therefore, that HLA-B*58:01 is frequently not associated with control in the Thai population (Techakriengkrai, Tansiri, & Hansasuta, 2012) despite its significant link with disease control in other cohorts (Carlson, Listgarten, et al., 2012; Kiepiela et al., 2004; Lazaryan et al., 2010).

## 1.14 Immune control of HIV mediated by HLA-B*27:05 and HLA-B*57:01

HLA-B*57 and HLA-B*27 are the genes most widely and consistently associated with immune control of HIV (Fellay et al., 2009; Pereyra et al., 2010). These genes are highly enriched among elite and viraemic controllers (Emu et al., 2008; Migueles et al., 2000; Pereyra et al., 2008). Understanding the mechanism of immune control associated with these alleles may provide valuable insights into the nature of the most effective CTL responses to HIV.

HLA-B*27:05-meditated immune control of HIV has been shown to depend largely on an immunodominant CTL response to the Gag-KK10 (KRWIILGLNK, Gag 263-272) epitope (Goulder, Phillips, et al., 1997; Nixon et al., 1988) with some contribution from a CTL response to Pol-KY9 (KRKGGIGGY, Pol 901-909) (Payne et al., 2010). Escape from the Gag-KK10 CTL response tends to occur late in infection with selection of an R264K mutation, incurring a considerable fitness cost to the virus and requiring co-selection of a compensatory mutation. (Goulder, Phillips, et al., 1997; Schneidewind et al., 2007). These pathways of escape and compensation are explored in more detail in Chapter 3. An additional polymorphism associated with KK10 escape, L268M, often precedes R264K selection. This residue is likely to interact with the TCR (Stewart-Jones et al., 2005) and selection of this polymorphism is associated with changes to the Gag-KK10-specific CTL population, allowing escape from the most effective KK10-specific TCR clonotypes

(Iglesias et al., 2011). This may facilitate selection of R264K escape, although the mechanism has not been elucidated.

Immune control mediated by HLA-B*57 and HLA-B*58 (a closely related allele) is associated with Gag-specific CTL responses to Gag-TW10 (TSTLQEQIAW, Gag-240-249), Gag-ISW9 (ISPRTLNAW, Gag-147-155) and Gag-KF11 (KAFSPEVIPMF , Gag-162-172) epitopes (Kloverpris et al., 2012). Gag-ISW9 and Gag-TW10 responses are usually immunodominant during acute infection, while KF11-specific CTL responses tend to be dominant in chronic infection (Brumme et al., 2008; Crawford et al., 2007; Goulder, Altfeld, et al., 2001). The order in which escape mutations arise in these epitopes is generally reflected by this pattern of immunodominance, where escape in TW10 and ISW9 predictably precede KF11-specific CTL immunodominance and escape (Brumme et al., 2008). The selection of escape mutations in these three epitopes sequentially reduce viral replicative capacity and are associated with a concomitant increase in viral load (Crawford et al., 2009) which may be restored by accumulation of compensatory mutations (Gijsbers et al., 2013).

Escape in Gag-TW10 typically occurs with a T242N mutation, which has a demonstrated fitness cost, but does not completely cripple the virus (Martinez-picado et al., 2006). T242N is selected early during infection, reverts relatively rapidly in the absence of immune pressure (Leslie et al., 2004) and may be compensated by G248A and A248G in B and C Clade infection, respectively (Martinez-picado et al., 2006). Upstream mutations in the cyclophilin A (CypA) binding loop (H219Q, I223V and M228I) may also partially restore the replication

defect of T242N selection (Brockman et al., 2007; Crawford et al., 2011) and having pre-existing compensatory mutations lowers the fitness cost of T242N selection (Liu et al., 2014).

Viral escape from the Gag-ISW9 response is typically mediated by selection of the I147L mutation early during infection (Crawford et al., 2009), which tends to revert rapidly upon transmission. The A146P mutation, selected in the region flanking the epitope, is a processing mutation that prevents N-terminal trimming of peptides by ERAAP to produce the optimal ISW9 peptide (Draenert et al., 2004).

Escape in the KF11 epitope typically occurs at A163G, which has a demonstrated viral fitness cost, and reverts upon transmission to HLA mismatched recipients (Crawford et al., 2007). S165N is subsequently selected to compensate for structural changes imposed on the viral capsid, and partially restores viral fitness.

Differences in the timing of the protective effect of HLA-B*27:05 and HLA-B-57:01 during the course of infection points to different mechanisms of control in each case. HLA-B*57 is associated with delayed decline in CD4+ T cell count early in infection, while HLA-B*27 is associated with delayed progression to AIDS defining illness during later stages (Gao et al., 2005). Although the pathways of HLA-B*27:05 and HLA-B*57:01-mediated immune control have been well studied, much remains to be learned. Expression of a favourable HLA allele does not in itself guarantee immune control of HIV infection, and individuals who control infection do eventually progress. The precise factors that will determine whether

an individual with a given HLA allele will be able to control viral replication remains an incompletely answered question.

## 1.15 Study aims

It is clear that the mechanisms of HLA mediated control of HIV are incompletely understood, and likely to be multifactorial (Carlson, Brumme, et al., 2012; Payne et al., 2014). No single determinant associated with favourable HLA allele expression has proved to be absolute in ensuring disease control. This thesis aims to further our understanding of HLA-mediated immune control of HIV by examining the interplay of protective HLA expression, rate of disease progression, CTL driven intra-host viral evolution and viral replicative capacity in unusual settings. These include rare combinations of HLA alleles and clades of infection and cases of disease progression despite protective HLA expression. Specific aims were:

1. To define the mechanisms of HLA-B*27:05 and HLA-B*57:01-mediated immune control of HIV by studying clade specific difference in disease outcome in C clade and CRF01_AE clade infection (Chapters 3 and 5).

2. To define the determinants of HLA-B*27:05 and HLA-B*57:01-mediated immune control by studying dual expression of these alleles, particularly in individuals who fail to control infection (Chapter 4).

3. To define the pathways of intra-host viral evolution driven by specific HLA class I restricted CTL and their impact on disease outcome in transmission pairs where a baseline TF virus can be approximated (Chapters 3, 4 and 5).

4. To define the role of VRC in determining disease outcome in the setting of favourable HLA expression (Chapters 3 and 4).

5. To develop strategies for meaningful analysis of high resolution sequencing of HIV, including minor variant population analysis, to add to our understanding of the dynamics of CTL driven evolution within the intra-host quasispecies (Chapters 3, 4 and 5).

This work draws on multiple cohorts from across the globe to study unique groups of HLA-B*27:05/B57:01 dual expressing subjects, and HLA-B*27:05 expressing subjects infected with C clade HIV (a rare combination). Several well-characterised examples of HIV transmission pairs are presented and full-length ultra-deep sequencing with novel approaches to data analysis are used to provide detailed insights into the dynamics of HLA-mediated control of HIV and intra-host viral evolution.

# CHAPTER 2: Methods

Detailed methods are provided here. Methods used in each study are listed briefly in each chapter along with additional study-specific details. Supplementary methods including cell culture media constituents, details of cell lines, viral plasmids, and PCR primers are listed in Appendix I.

## 2.1 Patient cohorts

Chronically HIV-1 infected, ART naïve adults from cohorts listed in (Table 2.1) were recruited from antenatal or outpatient clinics at each participating site. A single paediatric, ART naïve subject from the Chronic Pair (CP) Study (Durban Cohort) was included. Ethical approval was granted by institutional review boards associated with each site. All subjects provided written informed consent for participation.

## 2.2 Quantification of HIV-1 viral load and CD4+ T cell count

HIV viral load testing was performed using the Roche Amplicor version 1.5 assay (Roche, Switzerland). CD4+ T cell counts were determined by flow cytometry. These tests were conducted as part of routine monitoring of patients at each hospital managing the clinical care of the subjects.

**Table 2.1 Cohorts and sites of recruitment of study subjects.** The number of subjects from each cohort that were included in these studies and their clade of HIV infection are given. All subjects are chronically infected ART naïve adults unless otherwise indicated.
[α] Includes one paediatric subject

| Cohort | n | Clade of Infection | Ethical Approval | References |
|---|---|---|---|---|
| 1) Durban, South Africa (Gateway Study, Chronic Pair Study, King Edward Hospital, Cato Manor Study, Paediatric Early HAART and Structured Treatment Interruption (PEHS) study, Sinikithemba cohort) | 16 [α] | C | Univeristy of KwaZulu-Natal Biomendical Research Ethics Committee, Oxford Research Ethic Committee, Institutional Review Board of the Massachusetts General Hospital | (Kiepiela et al., 2007) |
| 2) Kimberley Hospital, South Africa | 1 | C | Univeristy of Free State Ethics Committee | (Adland et al., 2015) |
| 3) Bloemfontein Hospital, South Africa | 5 | C | Univeristy of Free State Ethics Committee | (Huang et al., 2009) |
| 4) Thames Valley Cohort, UK (Royal Berkshire Hospital, Reading; Northampton General Hospital, Northampton; Churchill Hospital, Oxford; High Wycombe General Hospital, High Wycombe; Wexham Park Hospital Slough; Luton and Dunstable Hospital, Luton) | 10 | B, C, CRF01_AE | Oxford Research Ethic Committee | (Payne et al., 2010) |
| 5) Zambia Emory HIV Research Project, Lusaka, Zambia | 2 | C | University of Zambia Research Ethics Committee, Emory University Institutional Review Board | (Prince et al., 2012) |
| 6) Mbabane Study, Gaborone, Botswana | 1 | C | Office of Human Research Administration (Harvard School of Public Health) and the Health Research Development Committee (Botswana Ministry of Health | (Matthews et al., 2011) |
| 7) New Delhi HIV clinic, India | 10 | C | National Centre for Disease Control (New Dehli) Ethics Committee | |
| 8) Warsaw HIV clinic, Poland | 2 | B, C | | |
| 9) Massachusetts General Hospital, Boston, US | 6 | B | Institutional Review Board of the Massachusetts General Hospital | (Chen et al., 2012) |
| 10) Hospital Germans Trial I Pujol, Barcelona, Spain | 1 | B | Research Ethics Committee, and Institutional Review Board of Hospital Germans Trias I Pujol | |

## 2.3 Whole blood sample processing

### 2.3.1 Separation of PBMC and plasma from whole blood

Whole blood collected in EDTA was centrifuged for 10 minutes at 900×g, and aliquots of 1-1.5ml of plasma were removed and stored at -80°C. For PBMC separation, the remaining whole blood was diluted with phosphate-buffered saline (PBS) and layered onto a Lymphoprep sucrose gradient (Axis-Shield, Norway) at a blood:PBS:Lymphoprep ratio of 1:1:1. The gradient was centrifuged at 900×g for 20 minutes using slow acceleration and deceleration settings. The PBMC layer was removed, washed in RMPI (Sigma, UK) 3 times, and resuspended in R10 (Appendix I) following the last wash.

### 2.3.2 Cryopreservation and thawing of PBMC

Cells were centrifuged at 900×g for 5 minutes, and the pellet resuspended in freezing medium (Appendix I) added drop-wise. Cells were cooled at 1°C/minute in "Mr Frosty" containers at -80°C overnight before long-term storage in liquid nitrogen (-180°C). To thaw cryopreserved cells, vials were incubated at 37°C until the pellet could be dislodged. Warmed R10 medium was added drop-wise to cell pellets. For functional assays, cells were incubated at 37°C and 5% $CO_2$ overnight before use.

### 2.3.3 DNA extraction from whole blood

DNA extraction for sequencing of autologous virus (population level and clonal sequencing) and HLA-typing was performed using Puregene Reagents (Qiagen, UK). 3ml of whole blood was incubated in 9ml of red cell lysis solution for 10 minutes, centrifuged for 10 minutes at 900×g, and the pellet resuspended in 3ml of white cell lysis solution. After a period of at least 48 hours, 1ml of protein precipitation solution was added and the mixture vortexed for 20 seconds, and centrifuged for 10 minutes at 900×g. The supernatant was removed into 3ml of isopropanol, tubes inverted 50 times, and the mixture centrifuged for 3 minutes at 900×g. The DNA pellet was washed in 3ml of 70% ethanol, allowed to dry, resuspended in 250μl of sterile water and stored at -20°C.

### 2.3.4 RNA extraction from plasma

RNA extraction for sequencing of autologous virus (population level and clonal sequencing) was performed on 500-1,000μl of plasma using a Qiamp Viral RNA Mini Kit (Qiagen, UK). Aliquots of plasma were centrifuged for 1-2 hour at 25,000×g and the volume reduced to 140μl before proceeding according to the manufacturer's instructions. For ultra-deep sequencing, samples with a viral load <3,000 copies/ml were concentrated by processing three aliquots of plasma on the same Qiamp column.

## 2.4 HLA class I typing

Four-digit high resolution Sequence Based Typing of HLA-A. –B, and –C was performed from genomic DNA in the CLIA/ASHI accredited laboratory of William

Hildebrand, PhD, D (ABHI) at the University of Oklahoma Health Sciences Center. Locus specific PCR amplification and heterozygous DNA sequencing of exon 2 and 3 of the class I PCR amplicon, which encode the $\alpha_1$ and $\alpha_2$ extracellular domains, was performed. Relevant ambiguities (Cano et al., 2007) were resolved by homozygous sequencing.

## 2.5 Molecular methods

### 2.5.1 PCR amplification for Sanger sequencing of HIV genes

Nested touchdown PCR was used for the amplification of ~1,000bp fragments of the HIV genome for sequencing. For proviral DNA templates, round 1 PCRs were performed in 50µl (37µl water, 5µl reaction buffer, 1µl 3' primer (at 10nM), 1µl 5' primer (at 10nM), 3µl $MgCl_2$, 0.5µl dNTPs (at 10mM) (Bioline, UK), BioTaq DNA polymerase (Bioline, UK) and 2µl DNA template). The second round PCR was performed in 100µl using double the reagent volumes above. Generalised thermocycling conditions are shown in Table 2.2, where temperatures X and Y represent the optimised temperatures for a given set of primers (usually 2-5°C below the melting temperature of each primer). PCR primer sequences and annealing temperatures are given in Appendix I.

When sequencing from RNA templates, reverse transcription and the first round of amplification were performed simultaneously using Superscript III One-Step RT-PCR Kit with Platinum Taq High Fidelity (Invitrogen, UK) according to the manufacturer's instructions using 10µl of template RNA. Cycling conditions are given in Table 2.3.

Where higher fidelity amplification or blunt-ended PCR products for cloning were required, the second round PCR was carried out using Platinum Taq DNA Polymerase High Fidelity enzyme (Invitrogen, UK). The reaction was carried out in a 100µl volume (75.6µl water, 10µl reaction buffer, 4µl $MgSO_4$, 2µl 3' primer (at 10nM), 2µl 5' primer (at 10nM), 2µl 10mM dNTP Mix (Invitrogen, UK), 0.4µl Platinum Taq High Fidelity and 4µl of round 1 PCR product). PCR cycling conditions shown in Table 2.2 were used with the annealing temperatures adjusted to 68°C.

**Table 2.2 Generalised thermocycling conditions for PCR amplification of HIV genes for sequencing.** Temperatures "X" and "Y" represent the optimised annealing temperature for the forward and reverse primer respectively.

| Step | PCR Stage | Temperature | Time |
|---|---|---|---|
| **1** | Denaturation | 94°C | 2 minutes |
| **2** | Denaturation | 94°C | 15 seconds |
| **3** | Annealing | X°C | 30 seconds |
| **4** | Elongation | 72°C | 1 minute |
| **5** | Cycle Repeat (19 times to step 2) | | |
| **6** | Denaturation | 94°C | 15 seconds |
| **7** | Annealing | Y°C | 30 seconds |
| **8** | Elongation | 72°C | 1 minute |
| **9** | Cycle Repeat (19 times to step 6) | | |
| **10** | Elongation | 72°C | 7 minutes |
| **11** | Hold | 4°C | Indefinitely |

**Table 2.3 Thermocycling conditions for simultaneous reverse transcription and PCR amplification of RNA templates**

| Step | RT-PCR Stage | Temperature | Time |
|---|---|---|---|
| **1** | cDNA synthesis | 55°C | 30 minutes |
| **2** | Denaturation | 94°C | 2 minutes |
| **3** | Denaturation | 94°C | 15 seconds |
| **4** | Annealing | 55°C | 30 seconds |
| **5** | Elongation | 68°C | 2 minutes |
| **6** | Cycle Repeat (29 times to step 3) | | |
| **7** | Elongation | 68°C | 5 minutes |
| **8** | Hold | 4°C | Indefinitely |

### 2.5.2 PCR amplification for Sanger sequencing of HCV genes

A hemi-nested PCR was used for amplification of a ~1,200bp fragment of the HCV NS5B region. The first round PCR was performed from plasma RNA using Superscript III One-Step RT-PCR Kit with Platinum Taq High Fidelity (Invitrogen, UK) according to the manufacturer's instructions adjusting the annealing temperature to 53°C for the cDNA amplification cycles. Cycling conditions for the second round PCR are given in Table 2.4 below. PCR primer sequences are given in Appendix I.

### 2.5.3 Gel electrophoresis of PCR product.

PCR products were visualised (and separated for gel extraction purification) by loading 5µl of PCR product and 2µl of loading dye (Bioline, UK) on a 1% agarose (Sigma, UK) gel in TBE buffer with 5µl of ethidium bromide per 250ml of gel. 280V

were applied for 40 minutes for separation of 1,000bp PCR products. Hyperladder I (Bioline, UK) was used for estimation of PCR product size.

**Table 2.4 Thermocycling conditions for PCR amplification of HCV NS5B.**

| Step | PCR Stage | Temperature | Time |
|---|---|---|---|
| **1** | Denaturation | 94°C | 2 minutes |
| **2** | Denaturation | 94°C | 15 seconds |
| **3** | Annealing | 54°C | 30 seconds |
| **4** | Elongation | 72°C | 2 minutes |
| **5** | Cycle Repeat (39 times to step 2) | | |
| **6** | Elongation | 72°C | 7 minutes |
| **7** | Hold | 4°C | Indefinitely |

### 2.5.4 Purification of PCR product

For population level sequencing, PCR products were purified using polyethylene glycol (PEG) precipitation. 60µl of 20% PEG with 2.5M NaCl were added to 100µl of PCR product. This was incubated for 30 minutes at room temperature, and pelleted by centrifugation at 2,700×g for 1.5 hours at 4°C. Pellets were washed twice in 150µl chilled 70% ethanol, dried at 94°C for 1 minute, and resuspended in 16µl-20µl sterile water.

For other applications, PCR products were purified using QIAquick PCR Purification Kit (Qiagen, UK) according the manufacturer's instructions. Where multiple PCR product bands were seen on the agarose gel, or where a large volume

of PCR product required concentration, the full PCR product was seperated by gel electrophoresis and purified using QIAquick Gel Extraction Kit (Qiagen, UK).

### 2.5.5 Sanger sequencing

Sanger sequencing involves 'sequencing by synthesis' using polymerase for DNA replication of the sequence template. Di-dNTPs (ddNTPs) that lack the 3'OH group required by polymerase for addition of subsequent nucleotides are added to the reaction. This results in termination of the reaction, producing nucleotide fragments of varying length. Each of the four terminator ddNTPs are labelled with a different flourescent dye. Detection of the ddNTP that terminates the reaction at each position in the template strand allows resolution of the template nucleotide sequence.

For each ~1,000bp PCR product, six sequencing primers (three forward and three reverse) were designed to produce overlapping reads across the fragment. The sequences of these primers are given in Appendix I. Purified PCR products for sequencing were prepared using the BigDye Terminator v3.1 reaction Mix (Applied Biosystems, UK). Reactions were carried out in a 6µl volume (1.5µl reaction buffer, 4µl of water, 0.5µl BigDye, 2µl sequencing primer (at 3.3nM) and 2µl PCR product). The chain termination reaction was carried out in a thermocycler (30 cycles of 96°C for 10 seconds, 50°C 5 seconds and 60 °C for 4 minutes). Chain termination reaction products were precipitated using a sodium acetate precipitation (30-minute room temperature incubation with 10µl water, 7µl NaOAC and 50µl of 100% ethanol). Precipitates were pelleted by centrifugation

for 1.5 hours at 4°C at 2,700×g . Pellets were washed in chilled 70% ethanol, and dried in a thermocycler at 94°C for 1 minute. Template sequencing was carried out on an ABI 3730xl DNA Analyzer (Applied Biosystems, UK) by the Department of Zoology Sequencing Facility, University of Oxford. Editing of sequence reads and assembly of a single contig from the 6 overlapping reads (one from each sequencing primer for a given template) was performed using Sequencher v5.0.1 software (Gene Codes Corp, US).

### 2.5.6 Clonal sequencing

For higher resolution sequencing, blunt-end PCR products were cloned into TOPO vectors using Zero Blunt TOPO PCR Cloning Kit (Invitrogen, UK) according to the manufacturer's instructions using 4µl of purified PCR product. The full ligation reaction volume was added to One Shot TOP10 chemically competent *E. coli* cells (Invitrogen, UK) for transformation. Transformed cells were spread on LB agar plates containing 50µg/ml kanamycin and incubated overnight at 37°C. A kanamycin resistance gene in the TOPO vector allows for the selective growth of transformed cells on the kanamycin-containing agar plates. Successful ligation of the PCR product into the TOPO vector disrupts a lethal gene in the vector allowing selective growth of recombinant plasmids.

Colonies on the LB agar plates were selected for culture in LB broth for mini prep extraction (see section 2.5.7). Sanger sequencing was performed as described in section 2.5.5 above using the plasmid DNA as the sequencing template. M13F and M13R primers included in the Zero Blunt PCR cloning kit (that bind the TOPO

vector on either side of the cloned insert) were added to the set of six sequencing primers.

### 2.5.7 Mini, midi and maxi prep extraction of plasmid DNA

Plasmid DNA was amplified and extracted using the Qiagen plasmid mini, midi or maxi prep kits (Qiagen, UK) as appropriate. 1.5ml, 100ml and 500ml of overnight bacterial culture in LB broth were used respectively, proceeding according to the manufacturer's instructions. Where a large batch of mini-prep extractions was required for simultaneous sequencing of a large number of clones from multiple samples, mini-prep extractions were performed using Montage 96-well plasmid preparation kits (Millipore, US) according to the manufacturer's instructions.

## 2.6 Cellular methods and flow cytometry

### 2.6.1 IFN-γ ELISpot assays

CD8+ T cell responses were measured using IFN-γ ELISpot assays. To screen for responses to the full HIV proteome, a panel of 210 18mer overlapping peptides (OLPs) spanning the full proteome was screened. Peptides were grouped into 72 pools of 11 or 12 peptides (final concentration 33μg/ml) arranged in a matrix, with each peptide represented in two pools. A response to both pools containing a particular OLP was suggestive of a response to that OLP. These responses were confirmed by screening of individual OLPs in a subsequent IFN-γ ELISpot assay. Peptides representing optimally defined CD8+ T cell epitopes restricted by relevant HLA alleles were also screened.

IFN-γ ELISpot assays were performed with fresh or thawed and rested PBMC in 96 well Polyvinylidene Fluoride membrane plates (Millipore, USA). Plates were coated with anti-human IFN-γ monoclonal antibody (Mabtech, Sweden) at 1:2,000 in PBS (100µl/well) for at least 16 hours. Plates were washed in 100µl/well blocking buffer (Appendix I), and 50µl R10 added per well. 10µl of OLP at 33µg/ml or optimal peptide at 20µg/ml was added to each well. 10µl phytohaemagglutinin (PHA) at 250µg/ml was added to positive control wells. PBMC were counted using a haemocytometer and 50,000-100,000 cells added in 100µl/well. Plates were incubated at 37°C and 5% $C0_2$ for 14-16 hours, washed six times in PBS, and incubated with 100µl/well of a 1:2,000 dilution of biotinylated IFN-γ antibody (Mabtech, Sweden) for 90 minutes in the dark. Plates were then washed 6 times in PBS and incubated with 100µl/well of a 1:2,000 dilution of streptavidin-alkaline phosphatase-conjugated antibody (Mabtech, Sweden) for 45 minutes in the dark. The plate was developed by addition of substrate colour solution (Bio-Rad Laboratories, UK) and read using an ELISpot plate reader (ELISpot v4.0, Autoimmune Diagnotiska, Germany). The SFU/million PBMC were calculated as follows, with a positive response counted as a value ≥50 SFU:

$$[\text{spot count} - (\text{mean} + 3(\text{SD}) \text{ of 4 negative control wells})] \times \frac{1\text{M PBMC}}{\text{PMBC per well}}$$

### 2.6.2 Cell surface marker staining for flow cytometry

The antibody panel used for cell surface marker staining for flow cytometry is given below in Table 2.5. PBMC for staining were pelleted by centrifugation (900×g for 5 minutes). Antibodies were prepared in a 50µl volume for each sample and cells resuspended in this mixture. Cells were incubated for 30 minutes at room temperature in the dark. 150µl of PBS was added before pelleting the cells as

before, and fixing the cells in 2% paraformaldehyde for 20-30 minutes. Samples were analysed using an LSRII flow cytometer (BD, UK) or MACSQuant Analyzer 10 flow cytometer (Miltenyi Biotec, Germany). Data were analysed using FlowJo version 10.0.7.

**Table 2.5 Antibody panel for cell surface marker staining.**

| Antibody | Working dilution |
|---|---|
| near-IR Live/Dead marker (Invitrogen, UK) | 1:50 |
| CD3 Pacific Orange (Invitrogen, UK) | 1:50 |
| CD4 AlexaFlour700 (BD Biosciences, UK) | 1:100 |
| CD8 V450 (BD Biosciences, UK) | 1:200 |

### 2.6.3 Peptide-MHC tetramer staining

Peptide-MHC tetramers were generated as previously described (Leisner et al., 2008) by the Buus laboratory (University of Copenhagen, Denmark). 0.5-1 million PBMC were used for each tetramer stain. PE-conjugated HLA-B*27:05-KK10 (W/T) and HLA-B*57:01-KF11 peptide-MHC tetramers and APC-conjugated HLA-B*27:05-KK10 variant peptide-MHC tetramers were used. PBMC were pelleted in a 96-well plate (900×g , 3 minutes) and pellets washed in 200µl PBS. Tetramers were used at a 30nM concentration and cells stained in a 25µl volume for single tetramer stains, or 50µl for double tetramers stains for 20 minutes in the dark at

room temperature. 160µl PBS was added and cells pelleted before proceeding with surface staining (section 2.6.2).

## 2.7 Viral culture and viral replicative capacity assays using p6.5/p83-2 + p83-10$_{GFP}$

### 2.7.1 Construction of recombinant p6.5 and p83-2 plasmids

The p6.5 plasmid was provided by Dr Julia Prado. In brief, this plasmid was constructed by cloning the p24 gene, amplified from a C clade infected subject, into the p83-2 plasmid between the SapI and ApaI restriction sites (Martinez-picado et al., 2006; Prado et al., 2010). To produce a p6.5 plasmid with the wildtype C clade p24 gene (p6.5_C_p24_WT), six amino acid differences between the patient-derived p24 sequence and the C clade consensus sequence were changed to C clade consensus residues by site-directed mutagenesis (SDM).

*Site-directed mutagenesis (SDM)*

SDM was performed on the p83-2 or p6.5_C_p24_WT plasmid as a backbone for production of recombinant HIV variants. For introduction of a single mutation, Stratagene QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, US) was used according to the manufacturer's instructions. For simultaneous introduction of multiple mutations, Stratagene QuikChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies, US) was used according to the manufacturer's instructions. Primers designed for SDM are given in Appendix I.

For QuikChange II SDM, a pair of complementary primers (one for each complementary plasmid DNA strand) that encode a desired mutation is designed. During a thermal cycling reaction, the plasmid is denatured, mutagenic primers containing the desired mutation anneal to the plasmid, and the primers are extended by *Pfu* DNA polymerase. This produces a nicked single-stranded mutant plasmid. The complementary strand is synthesised during a subsequent cycling step using complementary primers in the reaction mix. In a subsequent reaction, parental DNA is digested by DpnI endonuclease, an enzyme that digests only methylated DNA. This treatment is selective since only DNA replicated in dam+ *E.coli* will be methylated. The mutant plasmid is transformed into competent cells for nick repair and mutant plasmid propagation.

For QuikChange Multi SDM, up to five sites can be mutated simultaneously. A single primer is designed for each mutation. During thermocycling, the double-stranded plasmid is denatured, multiple mutagenic primers anneal to the template, the primers are extended by *Pfu* DNA polymerase, and nicks are repaired. Parental DNA is digested by DpnI endonuclease digest, and the single-stranded mutant plasmid is used for transformation of competent *E. coli.* The complementary plasmid DNA strand is synthesised *in vivo* in these cells.

An ampicillin resistance gene in the plasmids allowed selection of transformants by growth on LB agar plates containing 50μg/ml ampicillin. Selected colonies were cultured overnight in LB broth for mini-prep plasmid extraction (section 2.5.7). Plasmids were sequenced to confirm successful mutagenesis (section 2.5.5).

*Ligation of recombinant plasmids*

To reduce the chance that PCR errors were introduced into plasmid DNA outside of the p24 insert during SDM, the mutant p24 region was ligated into fresh p83-2 or p6.5 plasmid. Recombinant and fresh plasmids were digested using ApaI and SapI restriction enzyme and electrophoretically separated on a 1% agarose gel. The p24 insert and p6.5/p83-2 vector were purified by gel extraction using Qiaquick gel extraction kit (Qiagen, UK). The p24 insert was ligated into fresh vector using Quick Ligase (New England Biolabs, UK) according to the manufacturer's instructions, combining 50ng of vector with a 3 fold molar excess (approximately 18ng) of p24 insert and proceeding with transformation of One Shot TOP10 chemically competent *E. coli* cells (Invitrogen, UK). Stocks of each recombinant plasmid were prepared from larger overnight cultures in LB broth by midi or maxi prep plasmid extraction (section 2.5.7).

### 2.7.2 Transfection of MT4 cells for production of recombinant viral stocks

The p6.5, p83-2 and p83-$10_{GFP}$ plasmids and mutant variants were linearised by *EcoRI* (New England Biolabs, UK) restriction digest at 37°C for 2 hours. The digested product was electrophoretically separated and purified by gel extraction (section 2.5.4).

5µg of each plasmid (p6.5+p83-$10_{GFP}$ or p83-2+p83-$10_{GFP}$) was used for transfection of MT4 cells (further details provided in Appendix I). For each transfection 5 million MT4 cells in 500µl of electroporation medium were electroporated in 0.4cm cuvettes in a BioRad GenePulser II at 250V, 950µF capacitance and infinite resistance. Electroporated cells were transferred to a flask

with an additional 1 million MT4 cells and 10ml R10 medium and incubated at 37°C and 5% $C0_2$.

Once transfected, the two plasmids recombine to form a single template for complete HIV virion production. These virions express GFP and infect neighbouring target cells. Infected cells are thus detectable by GFP expression.

Virus stocks were expanded over a period of approximately 25 days and 3-4ml of media removed and replaced every 2-3 days. Viral growth was measured every 3 days using percentage GFP expression (measured by flow cytometry) as a marker of viral spread. 400-500µl of resuspended cell culture was centrifuged at 900×g for 5 minutes, the cell pellets washed in PBS, and fixed in 2% formaldehyde for flow cytometry.

Viral supernatants were harvested when 60-80% of live cells expressed GFP. RNA extraction was performed on a sample of supernatant for *gag* sequencing to confirm viral stock sequence (section 2.3.4 and 2.5.5).

### 2.7.3 Titration of viral stocks by end-point dilution assay

For the titration of recombinant virus expressing GFP (i.e. produced from p6.5 + p83-$10_{GFP}$ or p83-2 + p83-$10_{GFP}$ and mutant derivatives of these plasmids), viral titration was determined using an endpoint dilution assay. A 96-well plate was set up as in Figure 2.1, with 200µl PBS in the outside wells, and 100µl R10 in all other wells. 25µl of virus stock was added to each well in the first column, and a 1:5 serial dilution performed by removing 25µl into each adjacent column. As a

negative control, no virus stock was added to the final column. 30,000 MT4 cells were added to each well in 100µl. The plate was incubated for 5 days at 37°C, 5% $CO_2$. GFP expression was detected using a MACSQuant Analyzer 10 flow cytometer (Miltenyi Biotec, Germany).

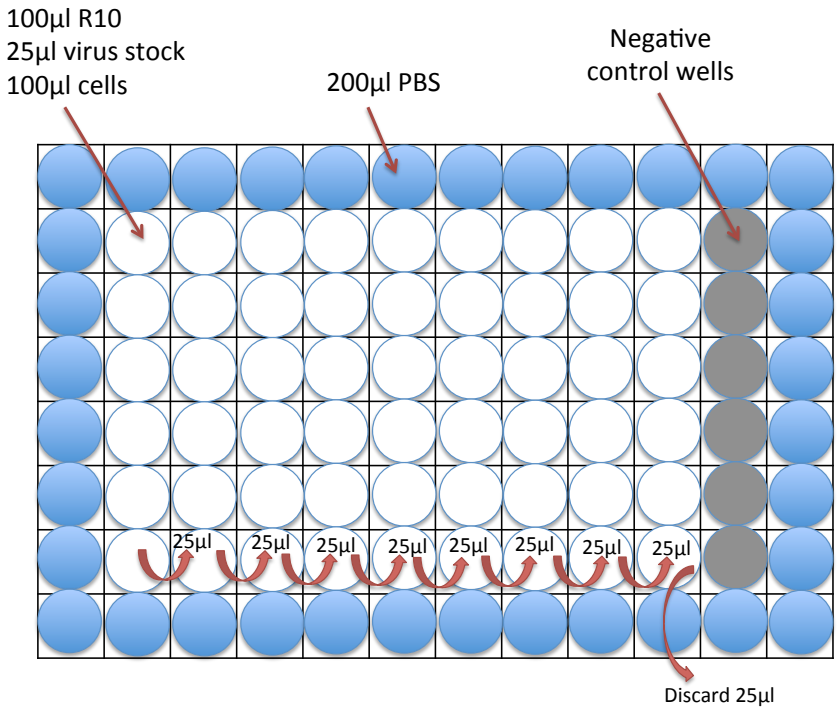



**Figure 2.1 llustration of endpoint dilution assay for virus titration**

The $TCID_{50}$ (tissue culture infective dose at which 50% of cells are infected) was calculated using the Reed-Muench calculation (Reed & Muench, 1938) using the Microsoft Excel calculator developed by B Lindenbach (Lindenbach, 2009). The dilution factor that gives higher than 50% infectivity and dilution factor that gives lower than 50% infectivity are used to determine the $TCID_{50}$, which lies between these values (the interpolated value/proportionate distance of the 50% endpoint (I)):

$$I = \left\{ \frac{\text{(\% of wells infected at dilution above 50\% - 50\%)}}{\text{(\% of wells infected at dilution above 50\% - \% of wells infected at dilution below)}} \right\}$$

Where h=dilution factor:

$$\text{50\% endpoint titer} = 10^{\text{log total dilution above 50\% - (I x log h)}}$$

$TCID_{50}$ (units/volume of virus dilution tested)=1/50% endpoint titre

This value can then be converted to log units/ml.

The MOI (multiplicity of infection) (ratio of infectious particles to target cells) was then calculated:

MOI=volume of virus stock × $TCID_{50}$ (log U/ml) / number of cells to be infected

### 2.7.4 Viral replicative capacity assays

MT4 cells were cultured and used in active growth phase for VRC assays. Each sample was run in triplicate using 1 million MT4 cells and infecting virus at an MOI of 0.001 in 1ml of R10 medium. MT4 cells were added in 100µl and incubated for 4 hours at 37°C, vortexing gently every 30 minutes to keep cells in suspension. Cells were washed 3 times in PBS to remove cell-free virus and resuspended in 2ml R10.

Viral spread was measured at the same time every day for 10 days by removing 200µl resuspended culture (~100,000 cells) (replaced with 200µl R10) for GFP detection by flow cytometry. Cells were stained with near-IR Live/Dead marker (Invitrogen, UK) to allow gating on viable cells by incubating the cell pellet in a 1:50 dilution of Live/Dead marker for 10 minutes in the dark at room temperature.

The percentage of live GFP+ cells was determined by flow cytometry using the MACSQuant Analyzer 10 flow cytometer (Miltenyi Biotec, Germany). Data were analysed using FlowJo version 10.0.7.

## 2.8 Viral culture and viral replicative capacity assays using ΔGag-Pro NL4-3

### 2.8.1 Construction of Gag-Pro chimera NL4-3 plasmids

The method of Brockman *et al,* where the *gag* gene and *protease* genes are amplified by PCR and cloned into a *Gag-Pro* deleted HIV backbone (ΔGag-Pro NL4-3, further details provided in Appendix I) was used to produce chimeric virus for *in vitro* VRC assays (Brockman, Tanzi, Walker, & Allen, 2006; Miura, Brockman, Brumme, et al., 2009). This method is designed to test the VRC related to the *gag* sequence. *Protease* is included in the construct to allow cleavage of *gag* gene products by autologous protease (Miura, Brockman, Brumme, et al., 2009).

Amplification of the *Gag-Pro* insert from autologous virus was conducted by nested PCR from plasma RNA as described in section 2.5.1 and Table 2.3 (round 1 PCR conditions with annealing temperatures X=58°C and Y=55°C). Round 1 PCR was performed using Superscript III One-Step RT-PCR kit with Platinum Taq High Fidelity enzyme (Invitrogen, UK). Round 2 PCR was performed using High Fidelity Platinum Taq (Invitrogen, UK). The cycling conditions for the second PCR are given below (Table 2.6). Primers for *Gag-Pro* amplification are given in Appendix I.

*Gag-Pro* PCR product was purified using QIAquick PCR Purification Kit (Qiagen, UK) according to the manufacturer's instructions. *Gag-pro* amplicons were sequenced to ensure correct amplification of patient-derived sequence (section 2.5.5) using the *gag-pro* PCR primers and F1b sequencing primers (listed in Appendix I).

**Table 2.6 Thermocycling conditions for second round PCR amplification of *Gag-Pro.***

| Step | RT-PCR Stage | Temperature | Time |
|---|---|---|---|
| **1** | Denaturation | 94°C | 2 minutes |
| **2** | Denaturation | 94°C | 30 seconds |
| **3** | Annealing | 60°C | 30 seconds |
| **4** | Elongation | 68°C | 2 minutes |
| **5** | Cycle Repeat (19 times to step 2) | | |
| **6** | Denaturation | 94°C | 30 seconds |
| **7** | Annealing | 57°C | 30 seconds |
| **8** | Elongation | 68°C | 2 minutes |
| **9** | Cycle Repeat (19 times to step 6) | | |
| **10** | Elongation | 68°C | 7 minutes |
| **11** | Hold | 4°C | Indefinitely |

### 2.8.2 Transfection of CEM-GXR cells for production of *gag-pro* chimeric virus stocks

The ΔGag-Pro NL4-3 was linearised by BstEII restriction digest. For each transfection, 10µg of plasmid DNA was digested with 20 units (2µl) BstEII restriction enzyme (New England Biolabs, UK) and 2µl BSA supplemented buffer 3.1 (New England Biolabs, UK) for 2 hours at 60°C.

10µg of linearised plasmid and 2.5µg of purified *gag-pro* PCR product was used for transfection of the CEM-GXR reporter T cell line (further details provided in Appendix I). For each transfection, 2 million CEM-GXR cells in 800µl of electroporation media were electroporated in 0.4cm cuvettes in a BioRad GenePulser II at 300V, 500µF capacitance and infinite resistance. Cells were rested for 45 minutes before transferring to a culture flask with 5ml R10 medium. Flasks were incubated at 37°C and 5% $C0_2$.

Virus stocks were expanded over a period of approximately one month. 5ml of fresh media was added on day 3, and 2-3ml of media removed and replaced every 2-3 days. Viral growth was measured every 2 days using percentage GFP expression (measured by flow cytometry) as a marker of viral spread. 200µl of resuspended cell culture was centrifuged at 900×g for 5 minutes, the cell pellet washed in PBS, and fixed in 2% formaldehyde for flow cytometry.

Viral supernatants were harvested when at least 30% of live cells expressed GFP. RNA extraction was performed on a sample of supernatant for *gag-pro* sequencing to confirm viral stock sequence (section 2.3.4 and 2.5.5)

### 2.8.3 Titration of viral stocks

Viral titers were determined by infecting 1 million CEM-GXR cells in 100µl with 400µl of virus stock and incubating at 37°C and 5% $C0_2$. 1ml R10 medium was added after 12 hours and the percentage of GFP+ cells determined by flow cytometry after 48 hours. 200µl resuspended cell culture was removed, pelleted,

fixed in 2% paraformaldehyde and analysed using a MACSQuant Analyzer 10 flow cytometer (Miltenyi Biotec, Germany).

An MOI of 0.025% infected cells at day 2 was used for viral replicative capacity assays. The volume of viral stock was calculated using:

VOLUME (in µl)=(MOI required) × (GFP titer % / volume of stock used in titration).

### 2.8.4 Viral replicative capacity assays

CEM-GXR cells (Brockman et al., 2006) were cultured and used in active growth phase for VRC assays. Each sample was run in triplicate using 1 million CEM-GXR cells. The volume of viral stock required for MOI of 0.025% infected cells at day 2 was diluted to a total volume of 500µl in R10 medium. Viral spread was measured by GFP expression by flow cytometry as described in section 2.7.4.

## 2.9 Logistic curve modeling of viral replicative capacity data[8]

Viral growth curves of VRC assay data were modeled on the logistic curve function:

$$y=m(x,\theta) + \varepsilon$$

$$=\theta_1 / 1 + exp(-\theta_2 + \theta_3 x)$$

where $\theta_1$ represents the asymptotic y point, $\theta_2$ the midpoint, and $\theta_3$ the scaling factor (rate of change) of the curve. An illustration of how each of these parameters affects the curve is given in (Figure 2.2). Curves were fitted using the nonlinear regression nls function in R version 3.2.1, plotting the best fit logistic

[8] The code for running this analysis was written by Dr Jacob Hurst.

growth curve. To assess whether each parameter was statistically different when comparing curves, models were built where one of the parameters was fixed while the others were allowed to vary. A model where all parameters were allowed to vary was then fitted and a statistical comparisons of parameters was made using ANOVA.

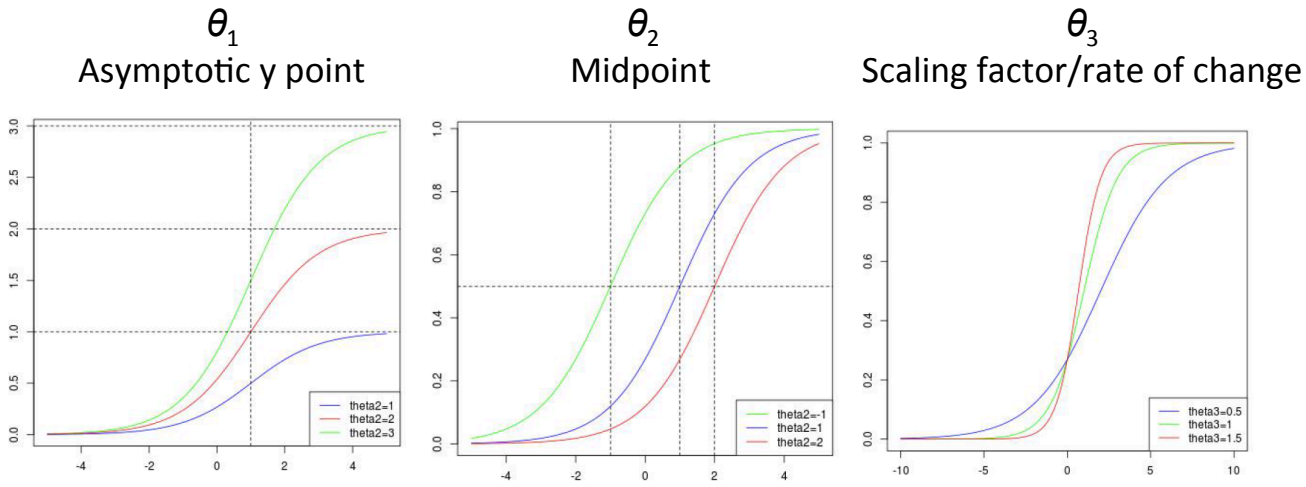



**Figure 2.2 The effect of varying each parameter in the logistic curve function used to analyse viral replicative capacity data.** The curves produced by setting a given parameter to three different values, while keeping the other parameters constant, are shown for $\theta_1$ (parameter defining the asymptotic y point), $\theta_2$ (parameter defining the midpoint of the curve) and $\theta_3$ (the scaling factor, or parameter defining the rate of change).

## 2.10 Ultra-deep sequencing with Illumina MiSeq technology[9]

Further explanation of this method is provided in Appendix II.

### 2.10.1 RNA extraction and PCR amplification of full-length HIV genomes

RNA extractions were performed using the Qiamp Viral RNA Mini Kit (Qiagen, UK) concentrating samples with a viral load <3,000 copies/ml by processing 3 aliquots of plasma on the same Qiamp column (section 2.3.4). PCR amplification of the full HIV genome (complete amino acid coding region and partial LTR) was performed in 4 fragments (1.9kb, 3.6kb, 3kb, 3.5kb) using Superscript III One-Step RT PCR Kit with Platinum Taq High Fidelity enzyme (Invitrogen, UK) in a 25µl reaction with the enzyme in 2x excess, as previously described (Gall et al., 2012; Gall, Morris, & Kellam, 2014) (Table 2.7). The four amplicons were pooled in approximately equimolar amounts for single-stranded library preparation.

[9] The methods in this section were performed in collaboration with Dr Astrid Gall and Prof Paul Kellam at the Wellcome Trust Sanger Institute and Dr Rebecca Batorsky and Dr Todd Allen at the Ragon Institute. Ultra-deep sequencing was performed by the Wellcome Trust Sanger Institute. Dr Astrid Gall performed quality control on raw reads, *de novo* assembly of consensus sequences, and minor variant calls. Dr Rebecca Batorsky performed minor variant analysis including epitope haplotype analysis and heat map analysis.

**Table 2.7 PCR cycling conditions for the full-length HIV PCR amplification for Illumina sequencing**

| Step | RT-PCR Stage | Temperature | Time |
|---|---|---|---|
| **1** | cDNA synthesis | 50°C | 30 minutes |
| **2** | Denaturation | 94°C | 2 minutes |
| **3** | Denaturation | 94°C | 15 seconds |
| **4** | Annealing | 58°C | 30 seconds |
| **5** | Elongation | 68°C | 4.5 minutes |
| **6** | Cycle Repeat (39 times to step 3) | | |
| **7** | Elongation | 68°C | 10 minutes |
| **8** | Hold | 4°C | Indefinitely |

### 2.10.2 Ultra-deep sequencing of amplicons and *de novo* assembly of consensus sequences

Amplicons were pooled for Illumina library preparation, including a unique bar code for each sample (for multiplex sequencing), and sequenced using MiSeq 250bp paired-end technology (Bentley et al., 2008). Quality control (removing reads of <50 bp and trimming low-quality bases from the 3'-end of the reads until the median quality of the read was 30) for removing sequence related errors that obscure true biological variants was carried out using QUASR (Quality Assessment of Short Reads) (http://sourceforge.net/projects/quasr/) (Watson et al., 2013). A *de novo* assembly was constructed using SPAdes version 2.4.0 (Bankevich et al., 2012). Resulting contiguous sequences were aligned with a clade-specific

reference strain and a consensus sequence was generated using Abacas version 1.3.1 and MUMmer version 3.2 (Kurtz et al., 2004).

### 2.10.3 Minor variant analysis

Raw reads were assembled using *Vicuna* (Yang et al., 2012) and *V-FAT* (Charlebois, Yang, Newman, Henn, & Zody, 2012) to form a single genome, which represents the majority base at each nucleotide position (the consensus assembly). The reads were then aligned to the consensus assembly using *Mosaik* (Lee et al., 2014). *V-Phaser*2 (Yang, Charlebois, Macalalad, Henn, & Zody, 2013) was used in order to call variants. This program uses both quality scores as well as covariation between variants (observation of two variants on the same read) to separate real variants from sequencing artifacts. We applied a modified strand bias cut-off to the variant calls. We required the odds-ratio of the appearance of a mutation between the two directions to be larger than 3.

### 2.10.4 Determination of epitope haplotypes

Haplotypes in the epitope regions were determined using *Vprofiler* (Henn et al., 2012) by selecting reads that span the epitope region and which contain only accepted variants. This analysis is limited to positions that are within the sequence read length of 250bp.

### 2.10.5 Heat map analysis

Heat maps of intra-host diversity were created using *Vprofiler* (Henn et al., 2012) as well as custom programs written in Perl and R. This method provides colour

plots that represent the extent of variability across the HIV proteome, either comparing sequences between two individuals or representing diversity within one individual at a given time point providing a snapshot of within-host diversity.

## 2.11 Phylogenetic analysis

### 2.11.1 Sequence subtyping and recombination detection

Sequence subtypes were determined using REGA HIV-1 Subtyping Tool version 3.0 (Pineda-Peña et al., 2013). Recombination analysis was performed using RDP 4.46 (Recombination Detection Programme) (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) and SimPlot 3.5.1 (Lole et al., 1999).

### 2.11.2 Maximum likelihood phylogenetic analysis

Maximum likelihood phylogenetic trees were constructed using Mega 6.06 software under the General Time Reversible model of nucleotide substitution as determined by jModelTest version 0.1.1 (Posada, 2008) with 1,000 bootstrap replicates. Phylogenetic trees were viewed and edited using FigTree v1.4.0 software.

### 2.11.3 Bayesian phylogenetic analysis[10]

A Bayesian time-measured phylogeny was constructed using Bayesian Evolutionary Analysis Sampling Trees (BEAST) v1.8.0 (Drummond, Suchard, Xie, &

[10] Training and support in conducting this analysis was provided by Dr Astrid Gall.

Rambaut, 2012). In this analysis, the time of sampling for each tip in the phylogenetic tree is known, and the time elapsed between samples can thus be used to estimate an evolutionary rate and extrapolate back to estimate the timing of branching events deeper in the tree. Because HIV tends to undergo a population bottleneck at the time of transmission, the time to the most recent common ancestor (MRCA) of a recipient approximates the time of transmission within a 95% height posterior density (HPD range) (Poon et al., 2011). The General Time Reversible (GTR) model of nucleotide substitution with gamma-distributed rate heterogeneity was selected using jModelTest version 0.1.1 (Posada, 2008). A relaxed clock model with gamma-distributed clock rate (Drummond, Ho, Phillips, & Rambaut, 2006) and GMRF Bayesian Skyride coalescent model (Minin, Bloomquist, & Suchard, 2008) were used. The Markov chain Monte Carlo (MCMC) search was set to 50,000,000 iterations with tree sampling every 2,500 iterations (producing 20,000 trees) with a 20% burn-in discarded. To ensure chain convergence, four individual chains were run. Phylogenetic trees were viewed and edited using FigTree v1.4.0 software.

## 2.12 Statistical analysis

### 2.12.1 Statistical tests

Statistical analysis was performed using Graphpad Prism v6.0 software. Fisher's exact test was used to test for associations between 2 categorical variables (such as presence or absence of a particular HLA allele, and presence or absence of a particular mutation) using 2×2 contingency tables. The Mann-Whitney U test, a

nonparametric test that compares the distribution of two unmatched groups, was used to compare VL and CD4+ T cell count data of two groups.

### 2.12.2 Analysis of covariation[11]

To identify HLA-associated and covarying polymorphisms, maximum likelihood phylogenetic trees (one per HIV-1 gene) were constructed using Phyml and a model of conditional adaptation inferred for each observed HIV-1 amino acid at each codon, as previously described (Carlson, Listgarten, et al., 2012; Carlson et al., 2008). The amino acid was assumed to evolve independently along the tree until reaching the tips, representing the present host. Selection via host HLA-mediated pressure and HIV-1 amino acid covariation was directly modeled using a weighted logistic regression, in which the individual's HLA repertoire and covarying HIV-1 amino acids were used as predictors, and the bias determined by the inferred possible transmitted sequences. To identify which factors (HLA and/or HIV-1 covariation) contribute to selection pressure, a forward selection procedure was employed where the most significant association was added to the model in an iterative fashion, with p-values computed using the likelihood ratio test. Statistical significance is reported using q-values, the p-value analogue of the false discovery rate (FDR). Q-values denote the expected proportion of false positives among results deemed significant at a given p-value threshold; for example, at $q \leq 0.05$, 5% of identified associations are expected to be false positives.

[11] This analysis was performed by Dr Jonathan Carlson

# CHAPTER 3: Patterns of immune escape and associated fitness cost are clade-specific in HLA-B*27:05-mediated immune control of HIV-1

## 3.1 Introduction

Human leukocyte antigen (HLA) class I is the genetic determinant that has been most strongly associated with extremes of disease outcome (Bartha et al., 2013; Fellay et al., 2007; Kaslow et al., 1996; Kiepiela et al., 2004; Leslie et al., 2010). HLA-B*27:05 is consistently associated with immune control (Fellay et al., 2009; Pereyra et al., 2010) and this allele is enriched accordingly among elite and viraemic controllers (Pereyra et al., 2008).

HLA-B*27:05-mediated control of HIV has been shown to hinge substantially on the immunodominant CTL response to the Gag-KK10 (KRWIILGLNK, Gag 263-272) epitope (Goulder, Phillips, et al., 1997). Selection of the Gag-R264K mutation results in effective escape from this CTL response, but imposes a considerable fitness cost on the virus (Schneidewind et al., 2007). It is therefore unsurprising that associated compensatory changes are co-selected. R264K is most commonly accompanied by an upstream change, S173A, that restores replicative capacity (Schneidewind et al., 2007). R264K rarely reverts when transmitted to an HLA-mismatched host with an S173A compensatory substitution (Goulder, Brander, et al., 2001; Schneidewind et al., 2009), providing further evidence that this compensation provides near complete restoration of viral fitness. A glycine is less

frequently selected as an alternative escape variant (R264G) and is associated with incomplete compensation by a Gag-E260D mutation (Schneidewind et al., 2008). The R264G substitution is less costly to viral fitness than R264K, but incomplete compensation may result in lower selection frequency of this variant (Schneidewind et al., 2008).

The phenotypic frequency of the HLA-B*27:05 allele is 8-10% in Caucasian populations, but is much lower (<1%) in sub-Saharan African populations (Gonzalez-Galarza, Christmas, Middleton, & Jones, 2011). Due to the geographical distribution of B clade infection in the largely Caucasian populations of Europe and the Americas, immune control and viral escape seen in subjects with HLA-B*27:05 have largely been studied in the context of B clade virus. The well-recognised suite of HLA-associated footprints described above is thus characteristic of selection pressure in B clade infection.

HIV disease outcome can be clade-specific in the context of expression of HLA alleles associated with both favourable and poor outcome (Brener et al., 2015; Kløverpris et al., 2013; Matthews, Koyanagi, et al., 2012). For HLA-B*27:05, sequence differences between B and C clade may affect the availability of key CTL epitopes. In particular, the Gag-173 consensus residue is serine (S) in B clade but threonine (T) in C clade, and the Gag-260 consensus is glutamic acid (E) in B clade but aspartic acid (D) in C clade (Table 3.1). The selection of HLA-B*27:05 escape and compensatory mutations may thus differ considerably between C clade and B clade infection.

**Table 3.1 The amino acid at Gag-173 and Gag-260 in B and C clade HIV sequences.**

| | Gag Amino Acid Position | |
|---|---|---|
| | 173 | 260 |
| B clade wildtype | S | E |
| C clade wildtype | T | D |
| Compensation for R264X escape in B clade | A | D |

A group of HLA-B*27:05-positive adults infected with C clade HIV were studied. Individuals with this rare combination were identified from several large cohorts (cohorts 1-8 listed in (Table 2.1). Full-length HIV ultra-deep sequencing, was used to identify longitudinal changes in viral sequence and study their effect on the rate of HIV disease progression. Viral replicative capacity (VRC) assays were used to investigate the effect of escape and co-selected mutations on viral fitness. The data offer unique insights into the clade-specific patterns of HLA-B*27:05-mediated escape and have wider implications for how HLA mediates differential disease outcome in HIV infection.

## 3.2 Methods

The following methods were used in this study as described in Chapter 2:

- Quantification of HIV viral load and CD4+ T cell count (section 2.2)
- HLA Class I typing (section 2.4)
- DNA extraction from whole blood (for viral sequencing and HLA Class I typing) (section 2.3.3)
- RNA extraction from plasma (for viral sequencing) (section 2.3.4)
- IFN-γ ELISpot assays (section 2.6.1)
- Sanger sequencing (section 2.5.5)
- Ultra-deep sequencing with Illumina MiSeq technology (incuding *de novo* assembly of consensus sequences) (section 2.10.1 and 2.10.2)
- Analysis of ultra-deep sequencing data (minor variant analysis and determination of epitope haplotypes) (section 2.10.3 and 2.10.4)
- Phylogenetic analysis (sequence subtyping, maximum likelihood phylogenetic analysis, and Bayesian phylogenetic analysis) (section 2.11)
- Viral culture and VRC assays using p6.5/p83-2 + p83-$10_{GFP}$ (including generation of recombinant plasmids by SDM). (section 2.7)
- Logisitic curve modeling of viral replicative capacity data (section 2.9)
- Statistical analysis of covariation (section 2.12.2)

### 3.2.1 Additional details

*Study subjects*

As the expression of HLA-B*27:05 and infection with C clade HIV is a rare combination, 26 patients were identified from a large number of diverse cohorts (cohorts 1-8 listed in (Table 2.1). A pool of 15 HLA-B*27:05-negative subjects whose autologous virus carried HLA-B*27:05-associated footprints were identified from the same cohorts. The HLA-B*27:05-negative transmission partner of an HLA-B*27:05-positive subject from the Thames Valley Cohort, and the HLA-B*27:05-negative transmission partner and child of an HLA-B*27:05-positive subject from the Chronic Pair study were also studied.

*Plasmids produced by site directed mutagenesis*

The p83-2 and p6.5_C_p24_WT plasmids, which express wildtype B and C clade p24 respectively, were used as the backbone for production of a panel of variant plasmids using SDM. These plasmids, which were used to produce viral variants for VRC assays, are listed in Table 3.2.

**Table 3.2 Plasmids produced by SDM for expression of HIV variants for VRC assays.**

| | C clade p24 Backbone | B clade p24 Backbone |
|---|---|---|
| **Wildtype** | p6.5_C_p24_WT | p83-2 |
| **R264K Variants** | p6.5_C_p24 + R264K | p83-2 + R264K |
| | p6.5_C_p24 + T173S | |
| | p6.5_C_p24 + T173A | |
| | p6.5_C_p24 + R264K + T173A | p83-2 + R264K + S173A |
| | p6.5_C_p24 + R264K + T173S | p83-2 + R264K + S173T |
| | p6.5_C_p24 + S165N | |
| | p6.5_C_p24 + R264K + S165N | |
| | p6.5_C_p24 + R264K+ S165N + V168I | |
| | p6.5_C_p24 + R264K+ S165N + V218M | |
| | p6.5_C_p24 + V218M | |
| | p6.5_C_p24 R264K + V218M | |
| **R264G Variants** | p6.5_C_p24 + R264G | p83-2 + R264G |
| | p6.5_C_p24 + R264G + D260E | p83-2 + R264G + E260D |
| | p6.5_C_p24 + R264G + D260N | |

## 3.3 Results

### 3.3.1 HIV disease progression in a C clade infected HLA-B*27:05-positive adult

A C clade infected adult transmission pair (subjects R036 and R056) was studied over 86 months from April 2007 (referred to as 'time 0' for both partners) (Figure 3.1). The HLA-B*27:05-negative donor R036 (a South African female) infected the HLA-B*27:05-positive recipient R056 (a UK Caucasian male) between November 2007 (negative HIV test) and September 2008 (diagnosis). Maximum likelihood phylogenetic analysis and Rega HIV-1 subtyping of the full-length HIV genomes confirmed that the donor and recipient sequences were closely related (Figure 3.1C) and that the pair were infected with C clade HIV. A time-measured phylogeny was constructed using BEAST and used to estimate the date of infection as October 2007 (Figure 3.1D)[12]. These results and the patients' clinical histories suggest that the recipient was infected soon after he tested HIV negative in November 2007. The recipient was thus followed up from within a few months of transmission.

[12] Bayesian phylogenetic analysis with BEAST (Bayesian Evolutionary Analysis Sampling Trees) allows construction of a phylogeny with dated tips to be constructed based on the time of sampling (Drummond et al., 2012). The time elapsed between samples is used to estimate the rate of viral evolution. This can be used to extrapolate the timing of branching events deeper in the tree. As HIV undergoes a transmission bottleneck, the most recent common ancestor (MRCA) for multiple samples from a single subject provides a reasonable estimate of the time of HIV transmission in that subject (Poon et al., 2011). With this method, the recipient was estimated to have been infected in October 2007, with 95% HPD (height posterior density, similar to a confidence interval) between mid-2006 and mid-2008.

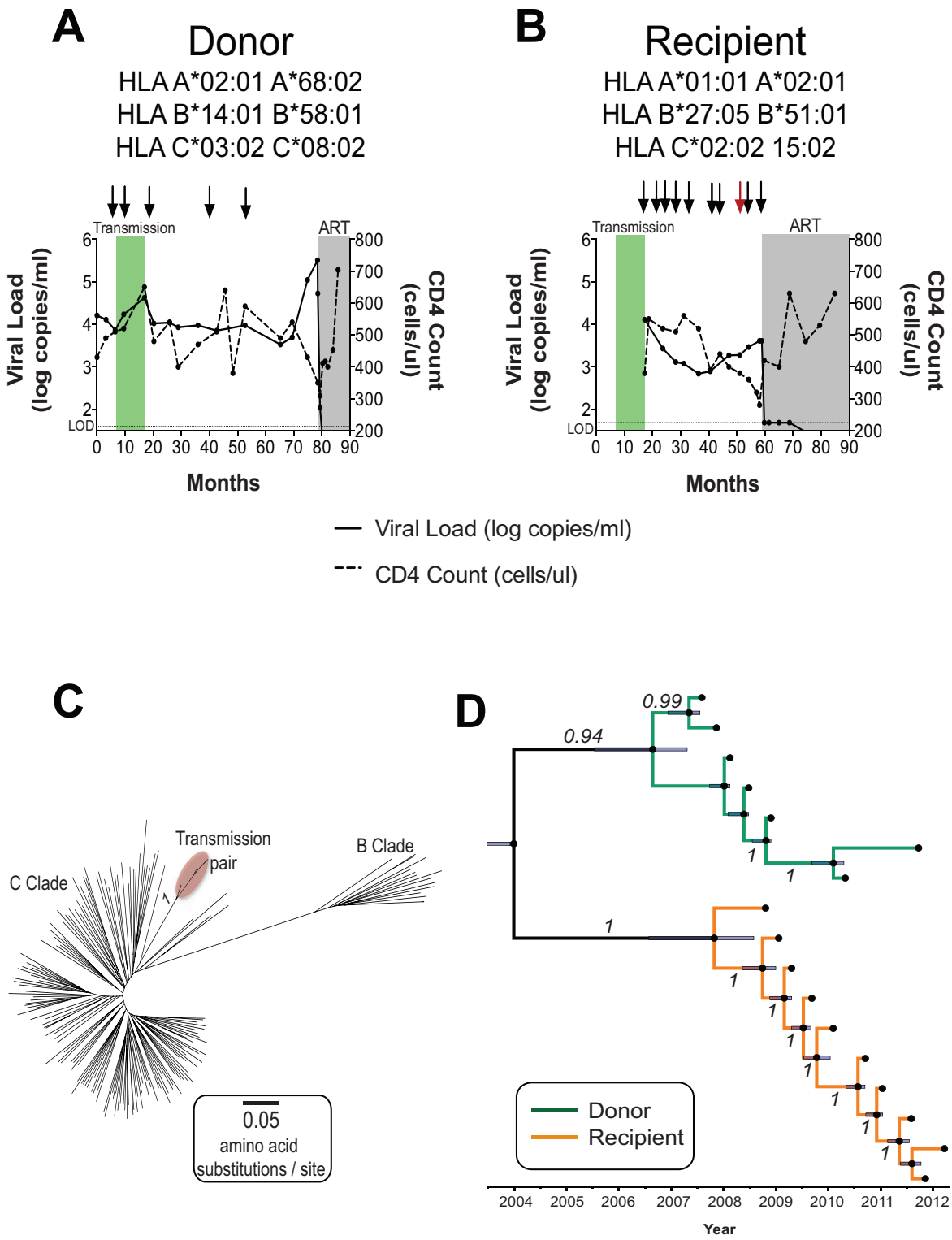



**Figure 3.1 Characteristic of HIV infection in a C clade infected adult transmission pair.** A) Female donor R036 and B) male recipient R056, showing plasma HIV RNA viral load (solid line) and CD4+ T cell counts (dashed line). Time '0' for both subjects represents the time of diagnosis of the recipient. The horizontal dotted line represents the limit of detection (LOD) of the viral load assay (40 copies/ml). Arrows indicate sampling for ultra-deep sequencing. Grey shading represents the periods during which the subjects received antiretroviral therapy (ART). The green shaded region represents the window during

which the recipeint was infected (time between negative and positive HIV tests). The red arrow indicates the deep sequencing timepoint at which the R264K escape mutation was first detected in the Gag-KK10 epitope. C) Maximum likelihood phylogenetic tree of 8,739bp sequence alignment (full-length HIV genome), including the donor and recipient pair (red circle, n=7 and 10 sequences respectively), and 150 C clade and 15 B clade reference sequences from South Africa and the US, respectively collected between 2001 and 2011 from the Los Alamos database (http://www.hiv.lanl.gov). Donor and recipient sequences cluster with C clade reference sequences, confirming that the transmission pair was C clade infected. The bootstrap value based on 1,000 bootstrap replicates for the transmission pair cluster is shown in italics. D) Bayesian time-measured phylogenetic analysis of full-length HIV sequences from multiple timepoints from the donor (green) and recipient (orange) date the MRCA of the recipient, and thus the timing of infection, at approximately October 2007 within the 95% height posterior density (HPD) range (blue bars). Posterior probability values are shown in italics.

### 3.3.2 Distinct pattern of Gag-KK10 escape in the HLA-B*27:05-positive recipient

Longitudinal ultra-deep sequencing of the HLA-B*27:05-positive recipient's autologous virus was used to track sequence changes over 42 months as he progressed to a CD4+ T cell count of <350cells/mm$^3$ and ART initiation (Table 3.3). This study first focused on Gag, which contains the immunodominant HLA-B*27:05-restricted epitope KK10 (KRWIILGLNK, Gag 263-272) (Figure 3.2). A pattern of escape was observed in the Gag-KK10 epitope sequence that was distinct from that typically seen in B clade infection. R264K selection was first observed at 44 months in 30% of the sequences before reaching fixation at 51 months. This was some time after disease progression was already evident from a declining CD4+ T cell count and rising viral load (Figure 3.1B). Importantly, the Gag-173A compensatory mutation was not selected to offset the fitness cost of this mutation, which is typical of B clade infection. Instead, Gag-S165N and Gag-V218M were co-selected. These mutations may be novel compensatory mutations in the context of C clade infection.

The L268M mutation, which typically precedes R264K selection in B clade infection, was lost in 95% of the sequences by the time the R264K mutation reached fixation. Mutations other than L268M and R264K, such as I267V and N271H, also occurred within minor variant populations.

**Table 3.3 Deep sequencing of the HLA-B*27:05-restricted Gag-KK10 epitope and co-varying amino acid positions in a C clade infected transmission pair.** Longitudinal Gag-KK10 haplotype and R264K co-selected polymorphisms are shown. Haplotype percentages have been adjusted to show the proportion of variants out of the total number after exclusion of variants with <1.5% frequency. Depth of coverage ranges from 174-192,000 reads. The C clade consensus sequence is shown in bold. R264K escape variants are shown in red.

| | C Clade Consensus and Autologous Variants | Frequency (%) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Donor (Month) | Recipient (Month) | | | | | | | | | |
| | Month | 19 | 19 | 21 | 24 | 28 | 33 | 41 | 44 | 51 | 54 | 59 |
| HLA-B*27:05 Gag-KK10 haplotype variants | $^{263}$**KRWIILGLNK**$^{272}$ | 100 | 72 | 98 | 39 | 43 | 54 | 10 | 32 | - | - | - |
| | -K--------- | - | 6 | - | - | - | - | - | 30 | 95 | 100 | 100 |
| | -K---M---- | - | - | - | - | - | - | - | - | 5 | - | - |
| | -----M---- | - | 13 | 2 | 61 | 48 | 25 | 5 | 3 | - | - | - |
| | ----VM---- | - | 6 | - | - | - | - | 67 | 15 | - | - | - |
| | -----M--H- | - | - | - | - | - | 3 | 14 | 8 | - | - | - |
| | -----M--T- | - | - | - | - | 10 | 19 | 2 | - | - | - | - |
| | ----V----- | - | 3 | - | - | - | - | - | - | - | - | - |
| | --------H- | - | - | - | - | - | - | - | 5 | - | - | - |
| | --------T- | - | - | - | - | - | - | 2 | - | - | - | - |
| | -----I--T- | - | - | - | - | - | - | - | 8 | - | - | - |
| Variants co-selected with Gag-R264K | **T**$^{173}$ → **X** | - | - | - | - | - | - | - | - | - | - | - |
| | **S**$^{165}$ → **N** | - | 6 | - | - | - | - | - | 31 | 100 | 100 | 77 |
| | **V**$^{218}$ → **M** | - | 5 | - | - | - | - | - | - | 81 | 88 | 96 |

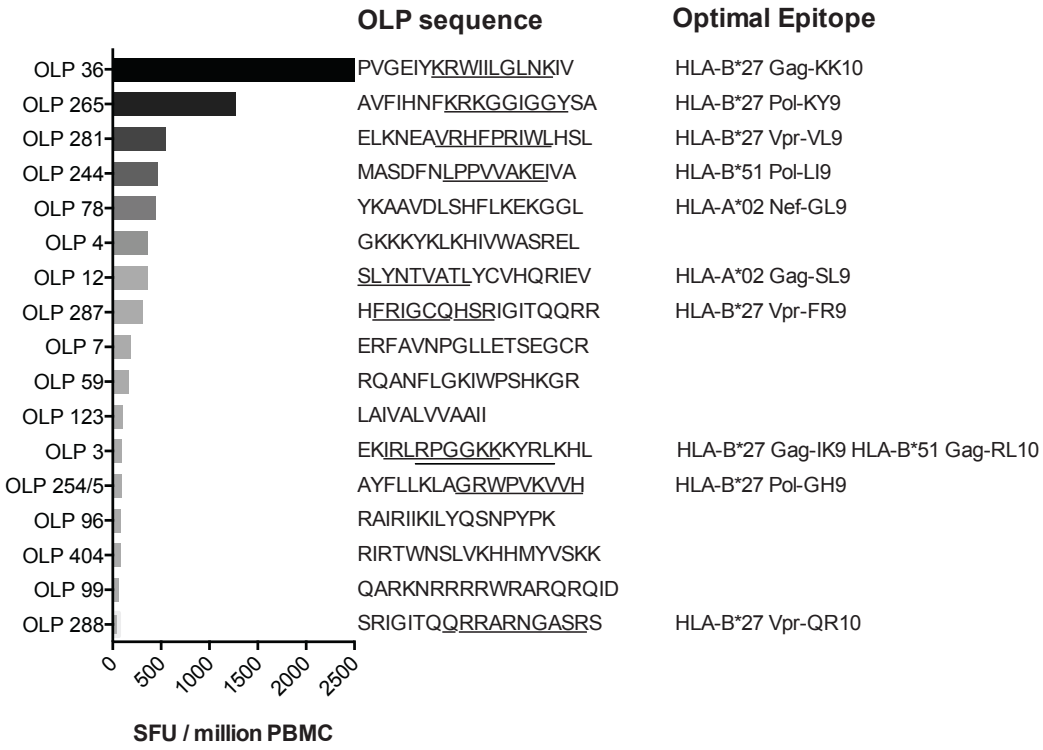



**Figure 3.2 CTL response hierarchy for an HLA-B*27:05-positive C clade HIV infected adult recipient R056 at 44 months.** IFN-γ ELISpot responses show an immunodominant response to the HLA-B*27:05-restricted Gag-KK10 epitope. The sequence of each OLP is shown, with the optimal epitope within the peptide underlined where known

### 3.3.3 Clade-specific pathways of Gag-KK10 escape and compensation

The pathways of KK10 escape and compensation in C clade infection were further investigated to determine whether these differed consistently from those seen in B clade infection. Viral Gag sequences from 25 additional C clade infected subjects expressing HLA-B*27:05 were studied (Table 3.4). Eighteen subjects carried variants within the Gag-KK10 epitope. Fifteen HLA-B*27:05-negative subjects whose autologous virus sequences contained a R264X mutation (assumed to have acquired the virus from an HLA-B*27:05-positive donor) were also identified. Co-selection of S165N in 9/23 subjects with R264K and absence of S165N in all 11 HLA-B*27:05-positive subjects without R264X supports the hypothesis that S165N may act as a C clade-specific compensation (p=0.02, Fisher's exact test).

Differences were consistently observed between the pathways used to compensate for R264X escape in B and C clade infection (Table 3.5). In a statistical analysis of co-variation with R264X, 12/19 sites were differentially selected in C compared to B clade infection (Table 3.5). In B clade infection, there is an almost universal requirement for S173A to be co-selected with R264K (Goulder, Phillips, et al., 1997; Schneidewind et al., 2007). In contrast, in C clade infection the putative compensatory mutation S165N was seen in only 9/23 cases of R264K selection (Table 3.4). These data suggest that, in contrast to the almost complete compensation by Gag-S173A in B clade infection, S165N may not fully compensate for reduced viral fitness resulting from R264K selection in C clade infection. An additional putative compensation, V218M, was identified in the HLA-B*27:05-positive recipient of the transmission pair described above. However, it was not identified consistently enough in other individuals to reach statistical significance.

E260D partially compensates for R264G selection in B clade infection (Schneidewind et al., 2008) and is present in the C clade consensus sequence. Its effect on R264G escape and compensation was next investigated. R264G selection was observed in three cases, with co-selection of D260N, which may act as a clade-specific compensation, in one case.

**Table 3.4 HIV Gag sequence haplotypes of proviral DNA from C clade infected subjects.** The Gag-KK10 epitope is shown with the associated compensatory position in B clade infection (Gag-173), and candidate compensatory positions in C clade infection (Gag-S165). *=one sample sequenced from RNA.

| | S[165] | T[173] | [263]KRWIILGLNK[272] | |
|---|---|---|---|---|
| CONSENSUS_C | S | T | KRWIILGLNK | |
| CONSENSUS_B | - | S | ---------- | |
| | | | | No. of Patients |
| HLA-B*27:05+ patients with KK10 R264X Escape (n=15) | - | - | -K---M---- | n=3* |
| | - | S | -K---I-X-- | n=1 |
| | N | - | -K---M---- | n=1 |
| | - | - | -K---M---- | n=1 |
| | N | - | -K---I--H- | n=1 |
| | - | - | -K---M---- | n=1 |
| | N | S | -K---M--H- | n=1 |
| | N | - | -K-------- | n=1* |
| | - | - | -K---M---- | n=1 |
| | - | I | -K---M---- | n=1 |
| | - | - | -G-------- | n=2 |
| | - | S | -G---M---- | n=1 |
| HLA-B*27:05+ patients without R264X KK10 Escape (n=11) | - | - | ---------- | n=3 |
| | - | - | ---------- | n=2 |
| | - | - | -----M---- | n=1 |
| | - | - | --------H- | n=1 |
| | - | - | ---V------ | n=1 |
| | - | - | -------F-- | n=1 |
| | - | - | ---------- | n=1 |
| | - | - | ---------- | n=1 |
| B27 Negative Patients with R264X escape (n=15) | - | - | -K-------- | n=4 |
| | N | - | -K-------- | n=1 |
| | N | - | -K---M---- | n=2 |
| | - | - | -K-------- | n=1 |
| | - | - | -K---I---- | n=1 |
| | N | I | -K--V----- | n=1 |
| | N | - | -K---M---- | n=1 |
| | - | - | -G-------- | n=1 |
| | - | - | -G-------- | n=1 |
| | - | - | -G---M---- | n=1 |
| | - | - | -S-------- | n=1 |
| B27 Negative Patients with other KK10 variants (n=76) | - | - | R--------- | n=6 |
| | - | - | ----VM---- | n=1 |
| | - | - | ----V----- | n=16 |
| | N | - | ----V----- | n=11 |
| | K | - | ----V----- | n=2 |
| | N | - | ----V----- | n=2 |
| | - | - | ----V----- | n=3 |
| | - | V | ----V----- | n=1 |
| | - | - | ----N----- | n=1 |
| | - | - | -----M---- | n=22 |
| | - | S | -----M---- | n=3 |
| | N | - | -----M---- | n=1 |
| | R | - | -----M---- | n=1 |
| | - | - | -----I---- | n=2 |
| | - | - | -----R---- | n=2 |
| | - | - | --------K- | n=2 |

**Table 3.5 Co-variation analysis of Gag-R264X in C and B clade HIV infection.** Co-variation of R264X with amino acid positions within Gag-KK10 and at Gag-173 and Gag-165 are shown with p and q values for B and C clade, and a comparison of B and C clade to determine which sites are differentially selected in each clade.

| | Amino Acid Position | C clade Consensus Residue | B clade Consensus Residue | Associated/ Covarying Residue | B Clade P value | C Clade P Value | Comparison of B and C clade | | Differential Selection In B vs C clade |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | P value | Q value | |
| R264 (W/T) | 163 | A | A | A | ns | ≤0.05 | ≤0.0001 | ≤0.0001 | Y |
| | 256 | I | I | V | ≤0.01 | ns | ≤0.0001 | ≤0.0001 | Y |
| | 256 | I | I | I | ≤0.01 | ns | ≤0.0001 | ≤0.0001 | Y |
| | 268 | L | L | I | ≤0.0001 | ≤0.0001 | ≤0.001 | ≤0.001 | Y |
| | 215 | L | V | M | ≤0.001 | ns | ns | ns | N |
| R264K escape | 173 | T | S | A | ≤0.0001 | ns | ≤0.0001 | ≤0.0001 | Y |
| | 173 | T | S | S | ≤0.0001 | ns | ≤0.01 | ≤0.01 | Y |
| | 182 | Q | Q | H | ns | ≤0.0001 | ≤0.01 | ≤0.01 | Y |
| | 200 | M | M | I | ns | ≤0.01 | ≤0.05 | ≤0.05 | Y |
| | 165 | S | S | N | ≤0.0001 | ≤0.001 | ≤0.05 | ≤0.05 | Y |
| | 165 | S | S | S | ≤0.0001 | ≤0.001 | ≤0.05 | ≤0.05 | Y |
| | 173 | T | S | T | ≤0.05 | ≤0.01 | ns | ≤0.05 | N |
| | 268 | L | L | L | ≤0.0001 | ≤0.0001 | ns | ≤0.05 | N |
| | 268 | L | L | M | ≤0.0001 | ≤0.0001 | ns | ns | N |
| | 268 | L | L | I | ≤0.0001 | ≤0.0001 | ns | ns | N |
| | 9 | R | S | R | ns | ≤0.05 | ns | ns | N |
| | 26 | K | K | K | ≤0.01 | ns | ns | ns | N |
| R264G escape | 136 | Q | Q | R | ≤0.0001 | ns | ≤0.001 | ≤0.001 | Y |
| | 136 | Q | Q | Q | ≤0.0001 | ns | ≤0.001 | ≤0.001 | Y |

### 3.3.4 Clade-specific Gag-KK10 epitope variant selection

To determine the frequency of KK10 variant selection in B and C clade sequences, 3,438 B and 1,729 C clade sequences from the Los Alamos HIV database (http://www.hiv.lanl.gov/) were compared. Gag-KK10 variants were significantly more common in B clade sequences (599/3428, 17.5%) than in C clade (155/1729, 9%) ($p<0.0001$, Fishers exact test). Among the sequences that carried variants, a significantly greater proportion of R264X and L268X variants were present in B clade sequences (Figure 3.3 and Table 3.6). These data are consistent with the higher prevalence of HLA-B*27:05 in B clade infected populations, resulting in a greater impact of HLA-B*27:05 on B clade viral sequences (Kawashima et al., 2009). I267V is likely to be found more frequently in C clade sequences because it is also associated with HLA-B*57:03, an allele that is much more prevalent in C clade infected populations. However, the differences between B and C clade remained statistically significant when sequences with I267V were excluded from the analysis (data not shown).

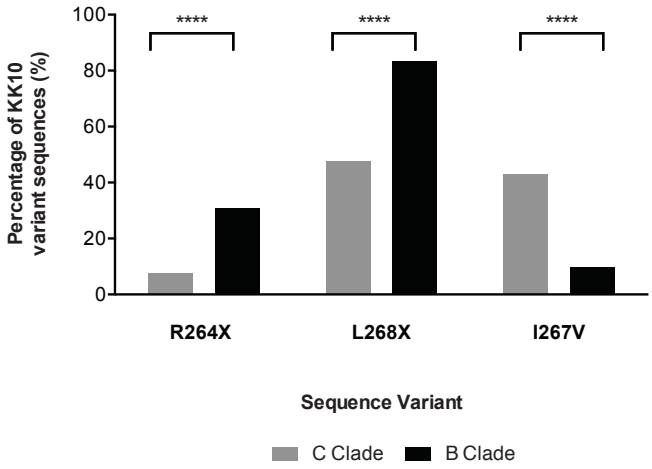



**Figure 3.3 Frequency of R264X, L268X, and I267V variants in the Gag-KK10 epitope in C clade and B clade sequences.** 155 C clade and 499 B clade Gag-KK10 variant sequences from the Los Alamos database (http://www.hiv.lanl.gov/) were compared. The frequency of each variant is given as a percentage of the total number of variant sequences. Statistical significance using Fishers Exact test on the absolute numbers of sequences is indicated. **** indicates $p<0.0001$.

**Table 3.6 Frequency of variants in the Gag-KK10 epitope in C clade and B clade sequences.** 155 C clade and 499 B clade Gag-KK10 variant sequences from the Los Alamos database (http://www.hiv.lanl.gov/) were compared. The frequency of each variant is given as a percentage of the total number of variant sequences.

| | | Percentage of Variant Sequences | |
|---|---|---|---|
| Variant | | C clade n=155 | B clade n=599 |
| R264K | -K-------- | 1.9 | 3.5 |
| | -K---M---- | 3.9 | 16.2 |
| | -K---I---- | - | 3.7 |
| | -K---M---R | 0.6 | - |
| | -K--VM---- | - | 0.8 |
| | -K--V----- | - | 0.3 |
| | -K--VI---- | - | 0.7 |
| | -K---M-F-- | - | 0.2 |
| R264G | -G-------- | - | 2.7 |
| | -G---M---- | 0.6 | - |
| R264T | -T-------- | 0.6 | 0.5 |
| | -T---M---- | - | 0.3 |
| | -T--VM---- | - | 0.2 |
| R264S | -S-------- | - | 0.3 |
| R264Q | -Q-------- | - | 0.2 |
| | -Q---M---- | - | 1.2 |
| R264N | -N-------- | - | 0.2 |
| L268M | -----M---- | 36.8 | 53.3 |
| | -----M--H- | - | 0.3 |
| | -----M--I- | - | 0.2 |
| | ----VM---- | 3.9 | 1.8 |
| I267V | ----V----- | 38.7 | 5.5 |
| | ----VI---- | 0.6 | 0.3 |
| Other Rare | -----I---- | 0.6 | 4.2 |
| | ----T----- | 0.6 | - |
| | ----TM---- | 0.6 | - |
| | ----FR---- | - | 0.3 |
| | ---L-M---- | - | 0.2 |
| | ---V------ | - | 0.3 |
| | ---T------ | - | 0.3 |
| | E--------- | 1.3 | - |
| | R--------- | 3.9 | 0.2 |
| | N-R------- | - | 0.3 |
| | R--V------ | 0.6 | - |
| | -P----RK-- | 0.6 | - |
| | --R------- | 0.6 | - |
| | ---M------ | 0.6 | - |
| | ---V------ | 0.6 | - |
| | -----V--T- | - | 0.2 |
| | ------R--- | - | 0.2 |
| | -------FD- | - | 0.2 |
| | ------E--- | - | 0.2 |
| | -----I--T- | - | 0.2 |
| | --------H- | - | 0.3 |
| | --------T- | 0.6 | 0.5 |
| | --------D- | 1.3 | - |
| | ---------I | - | 0.2 |

### 3.3.5 Effect of Gag-KK10 escape and compensation on viral replicative capacity in B and C clade infection

The effect of the escape and compensatory pathways observed in B and C clade Gag sequences *in vivo,* was investigated using *in vitro* VRC assays (Figure 3.4, Table 3.7). In the context of B clade Gag, R264K significantly reduced *in vitro* viral fitness compared to wildtype ($p<0.0001$ for $\theta_3$ and overall fit) (Figure 3.4A) which is consistent with previous reports (Schneidewind et al., 2007). The addition of the S173A compensatory change restored VRC to wildtype levels. The R264G variant also significantly reduced VRC compared to wildtype ($p<0.05$ for $\theta_3$, $p<0.0001$ for overall fit) but less so than R264K. The previously reported compensatory mutation E260D did not, however, significantly restore VRC (Schneidewind et al., 2008).

Similarly, in the context of a C clade p24 backbone, R264K significantly reduced VRC ($p<0.05$ for $\theta_3$, $p<0.0001$ for overall fit) compared to wildtype (Figure 3.4B). However, consistent with the lack of covariation at Gag-173 in association with R264K in *in vivo* viral sequences, addition of T173A only partially restored VRC (not statistically significant for $\theta_3$). As in B clade infection, in the C clade p24 virus R264G reduced VRC, although this did not reach statistical significance for $\theta_3$ (Figure 3.4C). The addition of the D260N mutation to the R264G variant did not restore VRC to this variant (data not shown).

Next, the impact of the C clade putative compensatory mutations on the VRC of C clade viruses with R264K escape was investigated (Figure 3.4B). The addition of

both the S165N and V218M mutations to the R264K mutant partially restored VRC. These compensatory mutations did not, however, fully restore VRC to wildtype levels as seen with addition of S173A in the B clade R264K escape variant. It is evident that compensation for R264K escape in C clade is less effective in restoring VRC than in B clade infection.

Unexpectedly, there was no significant difference in the VRC of the B and C clade wildtype p24 recombinant virus strains ($p=0.58$ for $\theta_3$, $p=0.19$ for overall fit) (Figure 3.4D). As the B- and C-clade p24 recombinant strains had comparable VRC values, the impact of introducing the Gag-R264K mutation into each backbone could be compared. Gag-R264K reduced the VRC substantially more in C-clade p24 than in B-clade.

These data indicate that the same escape mutation may have different effects on VRC in B and C clade variants. The compensatory mutations selected to restore VRC are clade-specific, and do not necessarily restore VRC to the same extent in each variant.

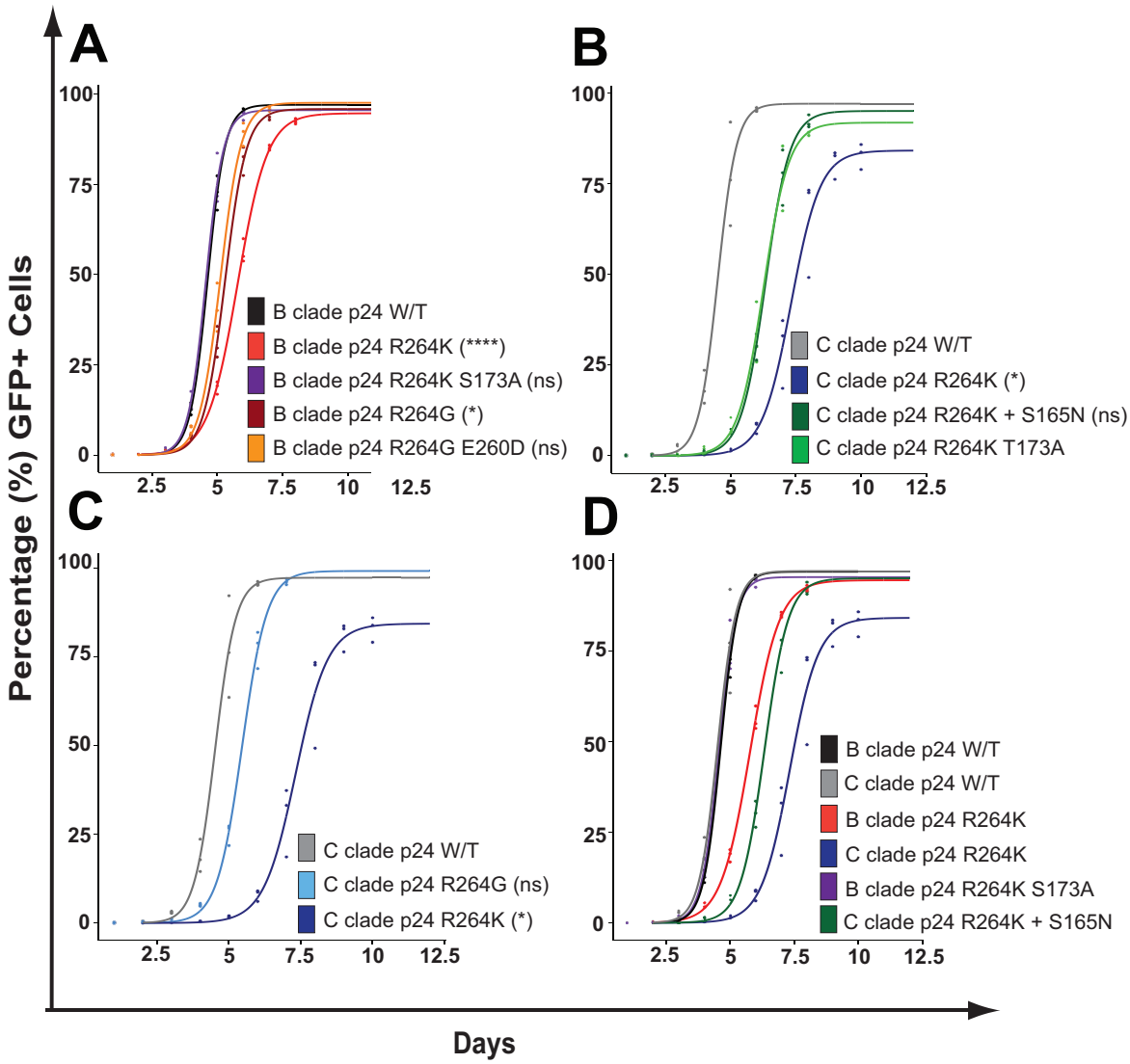



**Figure 3.4 Viral replicative capacity (VRC) assays for HIV B and C clade p24 sequences containing combinations of HLA-B*27:05-associated polymorphisms.** A) The effect of R264X escape and compensation on VRC in a B clade backbone. B) Effect of R264X mutation and the clade specific candidate compensatory mutation S165N in a C clade p24 backbone. C) The effect of R264G compared to R264K in a C clade p24 backbone. D) Comparison of R264K, with and without compensation, in a B and C clade p24 backbone. Statistical significance (of $\theta_3$ parameter) compared to the wildtype variant is shown in the key (ns not significant, *p≤0.05, **p≤0.01. ***p≤0.001, ****p≤0.0001).

**Table 3.7 Statistical comparison (ANOVA) of viral replicative capacity (VRC) assay data for HIV B and C C clade p24 sequences containing combinations of HLA-B*27:05-associated polymorphisms.** P values are shown for the overall fit of parameters $\theta_1$, $\theta_2$ and $\theta_3$ (asymptotic y value, midpoint, and scaling factor/rate of change, respectively) and for $\theta_3$. Letters in column 1 correspond to the panels in Figure 3.4.

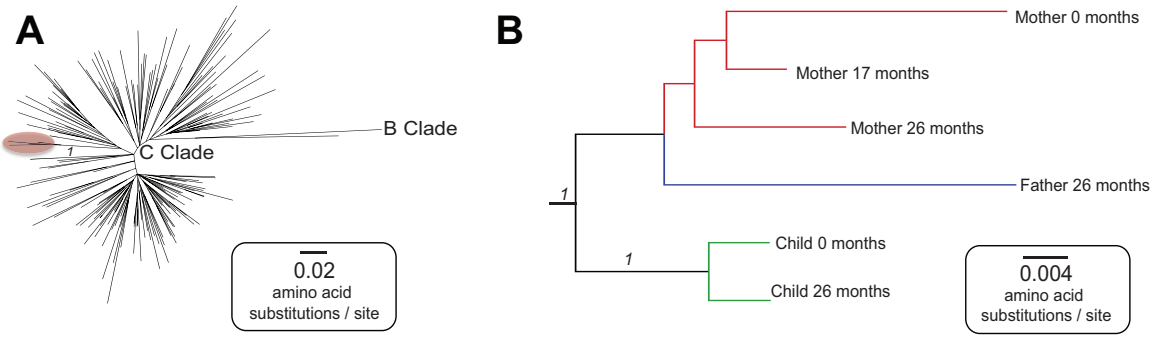
| | Statistical Comparison | | | θ3 (P value) | Overall fit (P value) |
|---|---|---|---|---|---|
| A | B clade p24 W/T | vs | B clade p24 R264K | <0.0001 | <0.0001 |
| | B clade p24 W/T | vs | B clade p24 R264K S173A | 0.97 (ns) | 0.38 (ns) |
| | B clade p24 W/T | vs | B clade p24 R264G | 0.01 | <0.0001 |
| | B clade p24 W/T | vs | B clade p24 R264G E260D | 0.07 (ns) | <0.0001 |
| B | C clade p24 W/T | vs | C clade p24 R264K | 0.02 | <0.0001 |
| | C clade p24 W/T | vs | C clade p24 R264K + S165N | 0.1 (ns) | <0.0001 |
| | C clade p24 W/T | vs | C clade p24 R264K T173A | 0.1 (ns) | <0.0001 |
| C | C clade p24 W/T | vs | C clade p24 R264G | 0.17 (ns) | <0.0001 |
| D | B clade p24 W/T | vs | C clade p24 W/T | 0.58 (ns) | 0.19 (ns) |
| | B clade p24 R264K | vs | C clade p24 R264K | 0.87 (ns) | <0.0001 |
| | B clade p24 R264K S173A | vs | C clade p24 R264K S165N | 0.002 | <0.0001 |

### 3.3.6 Stable transmission of R264K escape with S165N compensation in C clade infection

In B clade HIV infection, R264K in combination with S173A is stably transmitted to HLA-mismatched recipients and rarely reverts (Goulder, Brander, et al., 2001; Schneidewind et al., 2009). In the context of C clade infection, the S165N mutation may restore viral fitness sufficiently to allow stable transmission of the mutations to a mismatched recipient. To investigate this hypothesis, a C clade infected family trio consisting of an HLA-B*27:05-positive mother, HLA-B*27:05-negative father, and HLA-B*27:05-negative child, was studied. Phylogenetic analysis confirmed that the sequences from these individuals were epidemiologically linked (Figure 3.5A and B). HIV VL and CD4+ T cell counts for the trio are consistent with advanced chronic infection in the adults and a typical childhood disease trajectory in the child (Figure 3.5C).

The HLA-B*27:05-associated footprints within the Gag-KK10 epitope present in all three individuals suggested that transmission occurred from the mother to both the father and child (Figure 3.5). R264K was stably transmitted in combination with S165N with no reversion in either the father or child over 26 months. Selection of S165N, may thus have reduced the fitness cost of R264K selection such that no reversion occurred upon transmission. This result supports the restoration of viral fitness by S165N after R264K selection in C clade infection.

Other sequence polymorphisms in the autologous viral sequence may have contributed to the stable transmission of Gag-R264K escape from the mother to the father and child. In particular, the Gag-H219Q mutation in the CypA binding loop, which has been associated with enhanced VRC (Crawford et al., 2011), was present in all three subjects (Figure 3.5C).




C

| Subject | Month | S165 – | T173 S | H219 H | 263KRWIILGLNK272 ---------- | Viral load (log copies/ml) | CD4+ T cell count (cells/µl) |
|---|---|---|---|---|---|---|---|
| CONSENSUS_C | | N | – | Q | -K---I--H- | | |
| CONSENSUS_B | | N | – | Q | -K---I--H- | | |
| **Mother** HLA-A*30:02/30:04 B*27:05/58:02 C*02:02/06:02 | **0** | N | – | Q | -K---I--H- | 5.27 | 246 |
| | **17** | N | – | Q | -K---I--H- | 5.09 | 264 |
| | **26** | N | – | Q | -K---I--H- | 5.16 | 286 |
| **Child** HLA-A*03:01/30:02 B*08:01/58:02 C*06:02/07:01 | **0** | N | – | Q | -K---I--H- | 5.48 | 2205 |
| | **26** | N | – | Q | -K---I--H- | 5.80 | 885 |
| **Father** HLA-A*03:01/30:02 B*08:01/58:02 C*06:02/07:01 | **26** | N | – | Q | -K---I--H- | 5.71 | 129 |

**Figure 3.5: Gag-264K and Gag-S165N are stably transmitted from HLA-B*27:05-positive donor to HLA-mismatched recipients in a C clade HIV infected family trio.** A) Maximum likelihood phylogenetic analysis of Gag sequences (870bp) of an HIV infected mother, father and child transmission trio (red) form a cluster with 180 C clade reference sequences from Durban, confirming the clade of infection as C clade. B) The close phylogenetic relationship of the sequences from the mother, father and child confirm that they are a likely transmission trio. Bootstrap values based on 1000 bootstrap replicates >0.75 are shown in italics. Time since diagnosis is expressed in months. C) Alignment of proviral DNA sequences for the Gag-KK10 epitope.

### 3.3.7 Disease outcome associated with HLA-B*27:05 in C clade infection

To investigate whether HLA-B*27:05 is associated with favourable disease outcome in C clade infection, HIV VL and CD4+ T cell counts of these individuals were compared to C clade infected subjects not expressing protective HLA alleles (B*27, B*57, B*58:01 and B*81:01 ) (Figure 3.6 A and B). No significant difference was observed for either CD4+ T cell count or VL between the HLA-B*27:05-positive group and 'no protective HLA' groups (p=0.08 and p=0.12 respectively, Mann Whitney U test). To assess the effect of KK10 escape on disease outcome, the HIV VL and CD4+ T cell counts of HLA-B*27:05-positive subjects with and without

the R264X mutation were compared (Figure 3.6 C and D). No significant difference was found for either comparison (p=0.5 and p=0.2 respectively, Mann Whitney U test), although sample numbers were small. In contrast, HIV VL would be significantly higher and CD4+T cell count significantly lower in HLA-B*27:05+ groups in the context of B clade infection, since selection of escape in KK10 is strongly associated with HIV disease progression (Goulder, Phillips, et al., 1997).

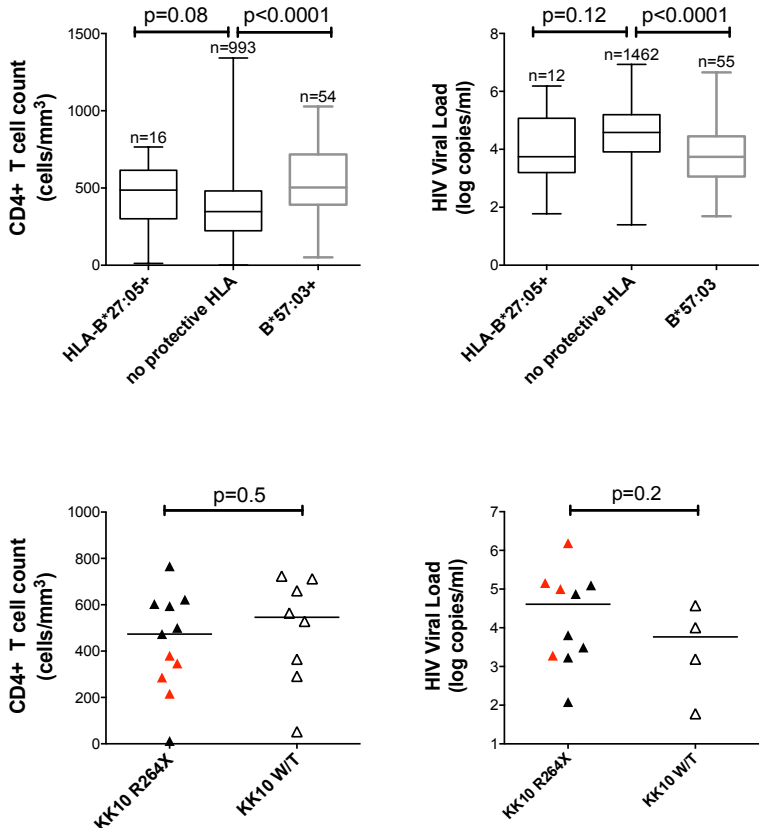



**Figure 3.6 Comparison of CD4+ T cell count and HIV viral load in C clade infected HLA-B*27:05-positive and negative subjects and in HLA-B*27:05-postive subjects with and without R264X mutation.** Comparison of A) CD4+ T cell count and B) HIV viral load in a "no protective HLA" control group (excluding individuals who expressed the protective HLA alleles B*27, B*57, B*58:01 and B*81:01) and the HLA-B*27:05-positive C group (excluding individuals who express an additional protective HLA B allele) both infected with C clade HIV. Subjects from a single cohort (Durban) expressing the protective allele HLA-B*57:03 are shown as an additional control group. C) CD4+ T cell count and D) HIV viral load in C clade infected HLA-B*27:05-positive subjects with and without R264X escape in the KK10 epitope. Box and whisker plots show the interquartile range and minimum and maximum values. The C clade HLA-B*27:05-positive group is compared with C clade infected individuals without protective alleles from the Durban, Bloemfontein, Kimberley, Botswana and TVC cohorts. Subjects with co-selection of the putative compensatory mutation S165N are shown in red.

## 3.4 Discussion

This study described the unusual scenario of HLA-B*27:05-expression in C clade infected subjects. One case was followed longitudinally from soon after transmission, with HIV disease progression occurring more rapidly than expected (Figure 3.1). Ultra-deep sequencing and VRC assays demonstrated how the patterns of escape and compensation and their effect on viral fitness differ substantially in B and C clade infection (Table 3.3 and Figure 3.4).

These findings were extended to a unique cohort of HLA-B*27:05-positive C clade infected subjects. The cohort showed a characteristic pattern of HLA-B*27:05-mediated selection pressure on C clade virus and confirmed that Gag-S165N may be an important candidate compensatory mutation (Table 3.4). VRC assays confirmed that selection of R264K was costly to the fitness of C clade virus, and that this is only partially compensated by S156N (Figure 3.4). This compensation appeared to be sufficient to allow stable transmission of R264K, together with S165N, to HLA-mismatched recipients without reversion (Figure 3.5).

This study has highlighted the differences between the escape mutations, fitness costs and disease outcomes associated with expression of HLA-B*27:05 in B and C clade infection. Although HLA-B*27:05 is strongly associated with immune control in B clade infection (Fellay et al., 2009; Pereyra et al., 2008, 2010), no protective effect could be demonstrated in C clade infected subjects. Similar to previous reports of the considerable fitness cost of R264K escape on the virus in B clade infection (Schneidewind et al., 2007), selection of this mutation significantly affected VRC in the context of C clade p24. However, the pathways of

compensation for this fitness cost differ substantially in C clade infection. Whereas almost complete compensation by Gag-S173A is seen in B clade infection, alternative mutations including S165N partially, but do not fully compensate for R264K in C clade infection,

These results are consistent with previous reports of B clade infection that suggest that R264K is selected more frequently than R264G, not because it is less costly to viral fitness, but because the virus can select a compensatory mutation that restores VRC more effectively (Schneidewind et al., 2008). This finding was extended to C clade infection, which showed that R264G is less costly, but also less well compensated than R264K, providing an explanation for why it is not selected more frequently.

Similar clade-specific effects have been observed for other HLA alleles. HLA-B*35 (especially B*35:01) is associated with rapid disease progression in B clade, but not in C clade infection. This is attributable to the targeting of the Gag-NY10 (NPPIPVGDIY, Gag 253-262) epitope in C clade infection that cannot be targeted in B clade infection due to the D260E polymorphism (Matthews, Koyanagi, et al., 2012). Likewise, HLA-B*07:02 is associated with disease progression in B clade, but not in C clade infection (Kløverpris et al., 2013). HLA-B*51:01 is protective in B clade infection, but linked with disease progression in C clade (Carlson, Listgarten, et al., 2012). This study adds HLA-B*27:05 to the list of alleles that are dependent on clade to mediate their effect.

Previous studies have shown that C clade HIV is considerably less fit than B clade HIV (Abraha et al., 2009; Ariën et al., 2005). C and B clade HIV have not, however, undergone convergent evolution over time, with the consensus sequences maintaining only 87% amino acid sequence similarity in the relatively conserved Gag protein. This may be in part due to the geographical distribution of these clades among different population groups, since population-specific selection pressure has been shown to shape HIV sequences at the population level (Kawashima et al., 2009; Kinloch et al., 2015, Matthews et al. 2009). Unexpectedly, in this study there was no significant difference in the VRC between the B and C clade wildtype p24 recombinant virus strains (Figure 3.4D). Replacing only the C clade p24 gene in the B clade backbone seems not to reduce VRC of the recombinant strain. The artificial system, therefore, allowed the direct comparison of the effect of a given mutation on viral fitness in B and C clade p24. In this system, Gag-R264K reduced the VRC substantially more in C-clade p24 than in B-clade p24. It is possible that differences between the wildtype B and C clade p24 sequences (6% amino acid difference) other than those examined in detail in this study may have contributed to this effect (Figure 3.7).

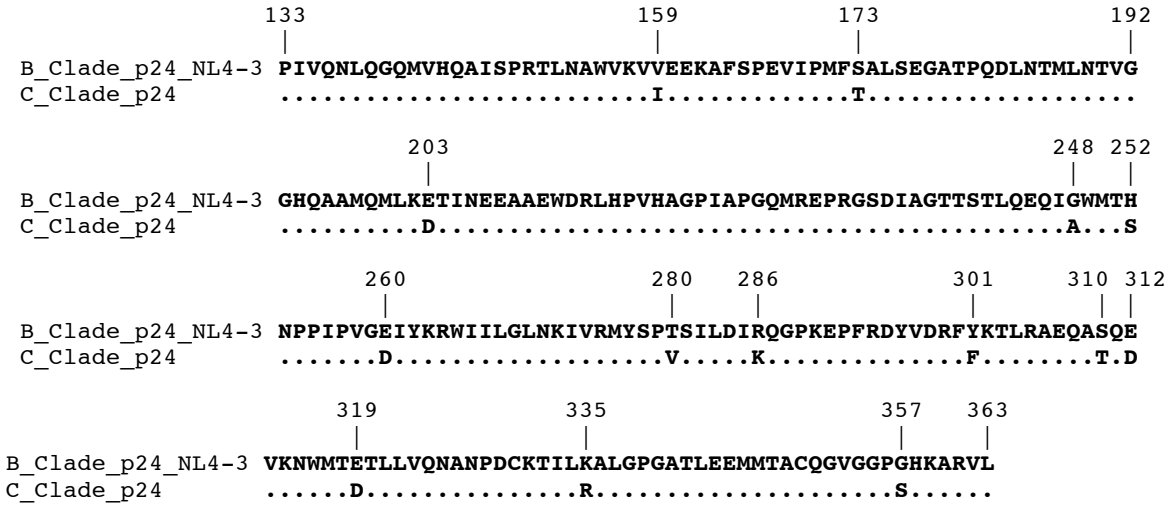


**Figure 3.7 Alignment of wildtype B clade (NL4-3) and C clade p24 sequences of the p83-2 and p6.5_C_p24 backbones used for production of viral variants for viral replicative capacity assays.** The position of 14 amino acid mismatches between the two sequences are shown.

A number of methods for culturing viral variants for VRC assays have been developed in recent years. These have included introduction of single point mutations into viral backbones, cloning of autologous viral genes between restriction sites in a viral backbone (Prado et al., 2010), and cloning of autologous viral genes into gene-deleted viral backbones (Miura, Brockman, Brumme, et al., 2009). Most recently, methods for producing full length infectious molecular clones from autologous virus, that most closely mimic *in vivo* VRC, have been developed (Claiborne et al., 2015; Deymier et al., 2015; Yue et al., 2015). In this study, introduction of single point mutations into an adapted viral backbone was used. This is the most controlled method of examining the effect of single point mutations on VRC. The extent to which these assays, where single mutations are introduced into an artificial viral backbone represent *in vivo* VRC is, however, an important caveat for consideration. Methods using autologous viral sequence that more closely represents the VRC of individual study subjects could be considered for future work.

A number of statistic approaches could be used to assess the VRC curves generated in this study (Figure 3.4) including independent comparison of each of the parameters built into the logistic curve model and comparison of the overall fit of all the parameters. In addition to the overall fit, the $\theta_3$ parameter, which defines the rate of change across the curves, and thus estimates viral growth rate, was analysed independently. This parameter represents the most biologically

important measure of VRC, and is most consistent with previous reporting of VRC data (Juarez-Molina et al., 2014; Miura et al., 2010; Miura, Brockman, Brumme, et al., 2009; Prado et al., 2010).

Expression of HLA-B*27:05 with infection of C clade HIV is an unusual combination. This study was therefore limited by small sample numbers, making it difficult to draw definitive conclusions about the overall association of HLA-B*27:05 with disease progression in C clade infection. The individuals studied were also pooled from many different cohorts to collect a sufficient number of study subjects, which adds complexity to the interpretation of the results. Furthermore, data from matched HLA-B*27:05-negative control subjects was not available from all cohorts.

Although the sample numbers available for the *in vivo* studies presented here are a limiting factor, expression of HLA-B*27:05 appears not to provide an advantage to C clade infected subjects. Counter intuitively, escape from Gag-KK10 in C clade p24 is costly to viral fitness and incompletely compensated in *in vitro* VRC assays. Larger studies of HLA-B*27:05-positive C clade infected subjects will be required to reconcile these observations. This work has, however, described a unique pathway of HLA-B*27:05-mediated CTL escape and compensation in the context of C clade infection. This shows that the transmitted founder sequence has a significant effect on the mechanism by which favourable HLA alleles mediate immune control of HIV infection.

# CHAPTER 4: Characterising failed immune control of HIV in adults with HLA-B*27:05 and HLA-B*57:01

## 4.1 Introduction

HIV elite and viraemic controllers represent a diverse and genetically heterogeneous group (Pereyra et al., 2008). Although HLA-B*27:05 and HLA-B*57:01 have consistently been associated with favourable disease outcome (Fellay et al., 2007, 2009; Goulder & Walker, 2012; Kiepiela et al., 2004; Pereyra et al., 2008), there is considerable variability in the rate of disease progression between individuals expressing these alleles (Chen et al., 2012; Norstrom et al., 2012) and most infected individuals do eventually progress. Defining the factors that lead to HIV disease progression within this group may provide important information for understanding the mechanisms by which protective HLA alleles mediate HIV disease control.

Each of the HLA class I alleles that affect disease outcome may have an additive or synergistic effect (Leslie et al., 2010; Matthews et al., 2011; Matthews, Listgarten, et al., 2012). The expression of multiple favourable HLA alleles might thus be expected to provide a considerable immunological advantage for the control of HIV infection. This study investigated a rare group of individuals expressing both the highly favourable alleles HLA-B*27:05 and HLA-B*57:01, with particular interest in defining the factors associated with failure to control HIV infection within this group. This chapter presents comparative case studies of a progressor (R097) and two disease controllers (RI088 and RI422) followed longitudinally over a period of 3.5, 12 and 20 years respectively. An additional 10 HLA-B*27:05/B*57:01-positive

subjects, four of whom progressed during the period of follow up, were studied to elucidate changes in viral sequence, rate of intra-host evolution, CTL response and viral fitness.

## 4.2 Methods

The following methods were used in this study as described in Chapter 2:

- Quantification of HIV-1 viral load and CD4+ T cell count (section 2.2)
- HLA Class I typing (section 2.4)
- DNA extraction from whole blood (for viral sequencing and HLA-typing) (section 2.3.3)
- RNA extraction from plasma (for viral sequencing) (section 2.3.4)
- IFN-γ ELISpot assays (section 2.6.1)
- Sanger sequencing (including population level and clonal sequencing) (section 2.5.5 and 2.5.6)
- Ultra-deep sequencing with Illumina MiSeq technology (incuding *de novo* assembly of consensus sequences) (section 2.10.1 and 2.10.2)
- Analysis of ultra-deep sequencing data (minor variant analysis and determination of epitope haplotypes) (section 2.10.3 and 2.10.4)
- Maximum likelihood phylogenetic analysis and recombination detection (section 2.11.1 and 2.11.2)
- Viral culture and viral replicative capacity assays using ΔGag-Pro NL4-3 (section 2.8)
- Logistic curve modeling of viral replicative capacity data (section 2.9)

### 4.2.1 Additional details

*Study Subjects*

A total of 13 HLA-B*27:05/B*57:01+ B clade HIV chronically infected subjects were recruited from the Massachusetts General Hospital (n=7) Thames Valley and RI Cohorts (n=4), Hospital Germans Trias I Pujol, Barcelona (n=1) and Warsaw Cohorts (n=1) (Table 2.1). Ethical approval was granted by the Institutional Review Board of the Massachusetts General Hospital, Oxford Research Ethics Committee, and Institutional Review Board of Hospital Germans Trias I Pujol (Barcelona).

*HLA types*

The full HLA haplotypes of the study subjects are given in Table 4.1

**Table 4.1 Full HLA haplotypes of HIV infected study subjects.**
ND=not done

| | | HLA Type | | |
|---|---|---|---|---|
| | | HLA A | HLA B | HLA C |
| **Controllers (n=8)** | B001 | 02:01 02:01 | 27:05 57:01 | 01:02 06:02 |
| | B003 | 01:01 26:01 | 27:05 57:01 | 02:02 06:02 |
| | RI088 | 01:01 03:01 | 27:05 57:01 | 02:02 06:02 |
| | RI422 | 24:02 31:01 | 27:05 57:01 | 02:02 06:02 |
| | B002 | 01:01 02:01 | 27:05 57:01 | 01:02 06:02 |
| | B004 | 01:01 11:01 | 27:05 57:01 | 02:02 06:02 |
| | B005 | 01:01 24:01 | 27:05 57:01 | 01:02 06:02 |
| | R039 | 01:01 31:01 | 27:05 57:01 | 02:02 06:02 |
| **Progressors (n=5)** | B006 | 02:01 02:01 | 27:05 57:01 | 01:02 06:02 |
| | B007 | 03:01 26:01 | 27:05 57:01 | 02:02 06:02 |
| | CJB001 | 01:01 02:01 | 27:05 57:01 | 01:02 06:02 |
| | R097 | 11:01 24:01 | 27:05 57:01 | 02:02 06:02 |
| | Warsaw79 | ND | 27:05 57:01 | ND |

## 4.3 Results

### 4.3.1 Progression in association with selection of multiple escape mutations in an HLA-B*27:05/B*57:01-positive subject

HLA-B*27:05/B*57:01-positive subject R097 was diagnosed with HIV infection in March 2010 (referred to as "time 0") after infection at an unknown timepoint up to 4 years earlier. This subject was infected by a known donor, subject R096 (HLA-A*02:01/A*11:01 B*13:02/B*51:01 C*03:03 C*06:02). Phylogenetic analysis of viral sequences confirmed that these subjects were epidemiologically linked (Figure 4.1B). The presence of HLA-B*51:01 associated footprints in the early viral sequence from R097 confirmed the direction of transmission from R096 to R097 (Figure 4.1C).

Despite expression of the two most protective HLA class I molecules in Caucasians infected with B clade HIV (Goulder & Walker, 2012; Kaslow et al., 1996; Pereyra et al., 2010), progression in R097 was rapid, with absolute CD4+ T cell counts declining to 470 cells/mm$^3$ and VL rising from 65 copies/ml at diagnosis to >10$^5$ copies/ml in a space of 2 years (Figure 4.1A). At approximately the same time as peak viraemia (February 2012), the recipient was diagnosed with Hepatitis C virus (HCV) genotype 1a co-infection by PCR, but remained HCV antibody negative for a further 2.5 years. A 24-week course of IFN-α/Ribovirin therapy was initiated 5 months after HCV diagnosis. This was followed by a rapid decline in the CD4+ T cell count that is typically associated with this treatment regimen (Chung et al., 2004; Soriano et al., 1994). The CD4+ T cell count did not recover after HCV

treatment was completed, and anti-retroviral (ART) treatment was initiated at 3.4 years.

Phylogenetic analysis of full-length ultra-deep sequences collected from subject R097 at the time of diagnosis and less than 2 years later showed rapid intra-host viral evolution, with sequences forming two distinct phylogenetic clusters (Figure 4.1B). Across the full-length genome, sequences collected less than 2 years apart were only 87.5% identical at the nucleotide level, and 90.2% identical in the *gag* gene. This translates to a 13% difference at the amino acid level in Gag, which is usually a highly conserved protein. This intra-host sequence variability is far greater than the previously reported values of 5%, and at most 10%, in the absence of superinfection (Korber et al., 2001; Salemi, 2013; Shriner, Liu, Nickle, & Mullins, 2006).

To assess the degree to which CTL escape in HLA-B*27:05 and HLA-B*57:01 epitopes contributed to this rapid viral sequence change, the HLA-B*27:05 and B*57:01-restricted epitopes in the donor and recipient were compared (Table 4.2). Sequence polymorphisms were selected in 11/21 epitopes in subject R097. Of note, escape mutations were selected in 3/4 p24 Gag CTL epitopes that are known to be particularly important for HLA-B*27:05 and B*57:01-mediated immune control (Crawford et al., 2007, 2009; Draenert et al., 2004; Goulder, Brander, et al., 2001; Goulder, Phillips, et al., 1997; Leslie et al., 2004; Martinez-picado et al., 2006; Schneidewind et al., 2007); R264K in Gag-KK10 ($^{263}$KRWIILGLNK$^{272}$), I47L and A146P in Gag-ISW9 ($^{147}$ISPRTLNAW$^{155}$) and T242N in Gag-TW10

([240]TSTLQEQIGW[249]). The wildtype sequence was maintained only at Gag-KF11 ([162]KAFSPEVIPMF[172]), which is associated with late escape (Brumme et al., 2008).

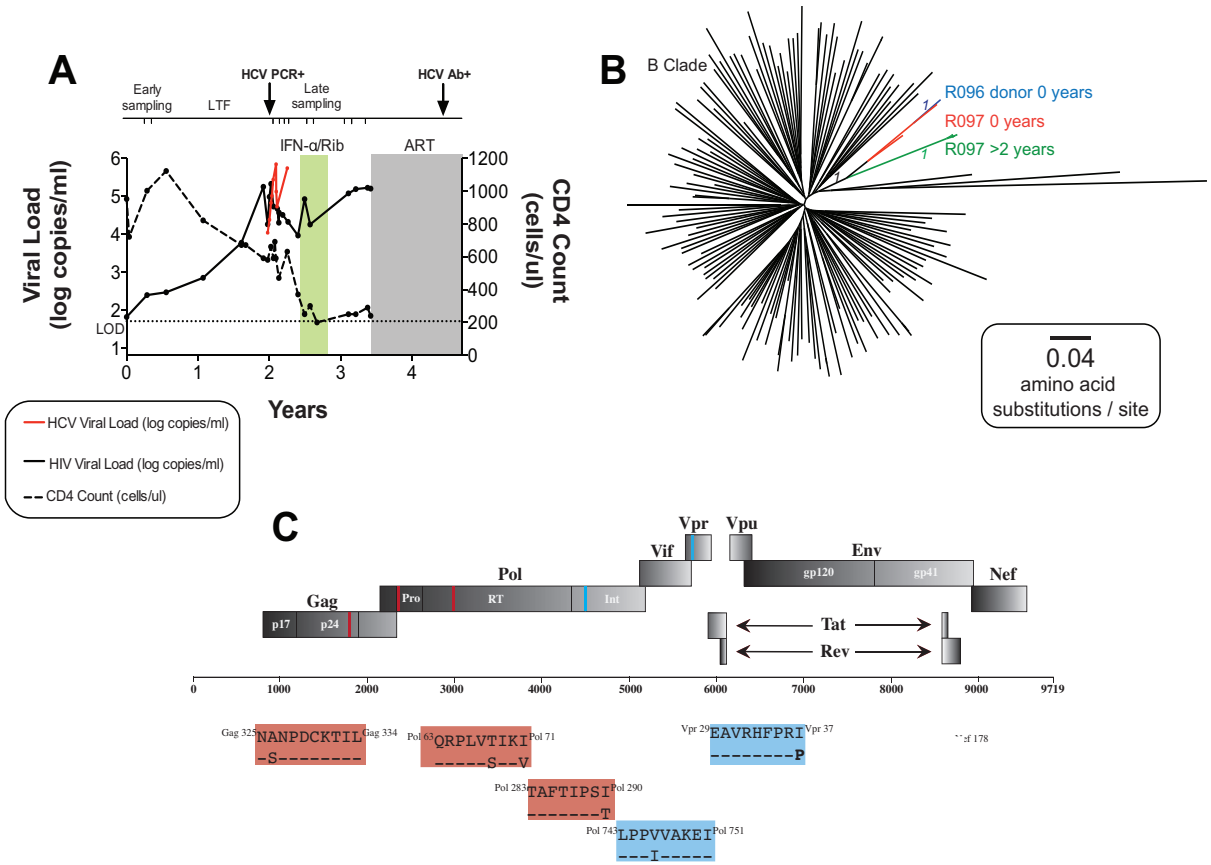



**Figure 4.1 Clinical course of infection and epidemiological linkage in a transmission pair with HLA-B*27:05/B*57:01-positive recipient.** A) Plasma HIV and HCV viral load and CD4+ T cell count for HLA-B*27:05/B*57:01-positive recipient R097. Time '0' represents the time of diagnosis of R097. The horizontal dotted line represents the limit of detection (LOD) of the viral load assay (40 copies.ml). A timeline indicates the time of sampling. The HCV viral load is shown in red. The time of first HCV PCR positive and antibody (Ab) positive tests are indicated with arrows. Grey shading indicates the period during which the subject received antiretroviral (ART) therapy. Green shading indicates the period during which the subject received interferon-α/Ribovirin therapy (IFN-α/Rib). B) Maximum likelihood phylogenetic tree of 7,588bp alignment of RNA (ultra-deep consensus) sequences across the full-length HIV genome. The donor R096 sequence is shown in blue. The early R097 sequences are shown in red (n=2). The late (post-2 year) R097 sequences are shown in green (n=11). 145 B clade reference sequences from the US and UK collected between 2003 and 2011 from the Los Alamos database (http://www.hiv.lanl.gov/) are shown in black. Bootstrap values for the R097 sequence clusters based on 1,000 bootstrap replicates are shown in italics. C) HLA-B*51:01 footprints in the recipient R097 sequence. The R097 sequence from time 0 is aligned to the B clade consensus sequence of five HLA-B*51:01-restricted epitopes with sequence polymorphism. Epitopes where polymorphisms that have previously been described HLA-B*51:01-associated footprints (Carlson, Brumme, et al., 2012) are shown in red.

**Table 4.2 Donor R096 and recipient R097 viral sequences of HLA-B*27:05/B*57:01-restricted epitopes.** Eleven epitopes showing evidence of CTL driven evolution in R097 are highlighted in grey. Sequences containing a described footprint (Carlson, Brumme, et al., 2012) are shown in red.

| HLA-B*57-restricted epitope | |
|---|---|
| **Gag-ISW9** | $^{146}$**A ISPRTLNAW**$^{155}$ |
| R096 0 years | - --------- |
| R097 0 years | - --------- |
| R097 2 years | P LT------- |
| **Gag-KF11** | $^{162}$**KAFSPEVIPMF**$^{172}$ |
| R096 0 years | ---------- |
| R097 0 years | ---------- |
| R097 2 years | ---------- |
| **Gag-TW10** | $^{240}$**TSTLQEQIGW**$^{249}$ |
| R096 0 years | ---------- |
| R097 0 years | --------E- |
| R097 2 years | --N------- |
| **Pol-VY10** | $^{272}$**VPLDEDFRKY**$^{280}$ |
| R096 0 years | ---------- |
| R097 0 years | ---------- |
| R097 2 years | ---------- |
| **Pol-ISW9** | $^{399}$**IVLPEKDSW**$^{407}$ |
| R096 0 years | --------- |
| R097 0 years | -Q--D---- |
| R097 2 years | -E------- |
| **Pol-IW9** | $^{530}$**IATESIVIW**$^{538}$ |
| R096 0 years | --------- |
| R097 0 years | --------- |
| R097 2 years | --L------ |
| **Pol-SW10** | $^{838}$**STTVKAACWW**$^{847}$ |
| R096 0 years | ---------- |
| R097 0 years | -NA------- |
| R097 2 years | -N-------- |
| **Pol-KF9** | $^{888}$**KTAVQMAVF**$^{896}$ |
| R096 0 years | --------- |
| R097 0 years | --------- |
| R097 2 years | --------- |
| **Vif-IF9** | $^{31}$**ISKKAKGWF**$^{39}$ |
| R096 0 years | ------R-V |
| R097 0 years | ------D-- |
| R097 2 years | T-R---D-R |
| **Vif-HL9** | $^{73}$**HTGERDWHL**$^{81}$ |
| R096 0 years | -----E--- |
| R097 0 years | Q-------- |
| R097 2 years | Q-------- |
| **Vpr-AW9/AL10** | $^{30}$**AVRHFPRIW**$^{38}$ **L**$^{39}$ |
| R096 0 years | -------P- - |
| R097 0 years | -------P- - |
| R097 2 years | --------- - |

| | |
|---|---|
| **Rev-KY10** | $^{14}$**KTVRLIKFLY**$^{23}$ |
| R096 0 years | ---------H |
| R097 0 years | -----V-R-- |
| R097 2 years | -----V-R-- |
| **Nef-KF9** | $^{82}$**KAAVDLSHF**$^{90}$ |
| R096 0 years | -G-L----- |
| R097 0 years | -G-F---F- |
| R097 2 years | -G-L----- |
| **Nef-HW9** | $^{116}$**HTQGYFPDW**$^{124}$ |
| R096 0 years | --------- |
| R097 0 years | not resolved |
| R097 2 years | N-------- |

| HLA-B*27-restricted epitope | |
|---|---|
| **Gag-IK9** | $^{19}$**IRLRPGGKK**$^{27}$ |
| R096 0 years | --------- |
| R097 0 years | ----R---- |
| R097 2 years | --------- |
| **Gag-KK10** | $^{263}$**KRWIILGLNK**$^{272}$ |
| R096 0 years | -----I---- |
| R097 0 years | -----I---- |
| R097 2 years | -K-------- |
| **Gag-QK10** | $^{379}$**QRGNFRNQRK**$^{388}$ |
| R096 0 years | -K-------- |
| R097 0 years | -K-------- |
| R097 2 years | -K-------- |
| **Pol-KY9** | $^{901}$**KRKGGIGGY**$^{909}$ |
| R096 0 years | --------- |
| R097 0 years | --------- |
| R097 2 years | --------- |
| **Vpr-VL9** | $^{31}$**VRHFPRIWL**$^{39}$ |
| R096 0 years | ------P-- |
| R097 0 years | ------P-- |
| R097 2 years | --------- |
| **Env-GY10** | $^{786}$**GRRGWEALKY**$^{795}$ |
| R096 0 years | ---------- |
| R097 0 years | ---------- |
| R097 2 years | ------V-R- |
| **Nef-RV10** | $^{105}$**RRQDILDLWV**$^{114}$ |
| R096 0 years | Q--------- |
| R097 0 years | K--E------ |
| R097 2 years | ---------- |

### 4.3.2 Rapid viral adaptation masquerades as an HIV superinfection in an HLA-B*27:05/B*57:01-positive progressor

HIV disease progression, rapid emergence of escape mutations in HLA-B*27:05 and B*57:01-restricted epitopes and an extraordinary burst of viral sequence change in just 2 years in subject R097 prompted the question of whether this was caused by an HIV superinfection.

The fluctuations in the HCV VL of subject R097 at the time of HCV diagnosis (Figure 4.2) and an HCV RNA negative test 10 months earlier were consistent with an acute HCV infection occurring around the time of peak HIV viraemia at 2 years post HIV diagnosis. Although HCV transmission primarily occurs through blood-to-blood contact, particularly in needle-sharing intra-venous drug users, a growing epidemic of sexually transmitted HCV co-infection in HIV infected MSMs has been described in recent years (Kouyos et al., 2014; Myers et al., 2009; Urbanus et al., 2009; van de Laar et al., 2007). An acute infection with HCV at the time of peak HIV viraemia in R097 indicated that this subject was exposed to ongoing risk of acquiring sexually transmitted infections, supporting the hypothesis that he may have acquired an HIV superinfection at this time.

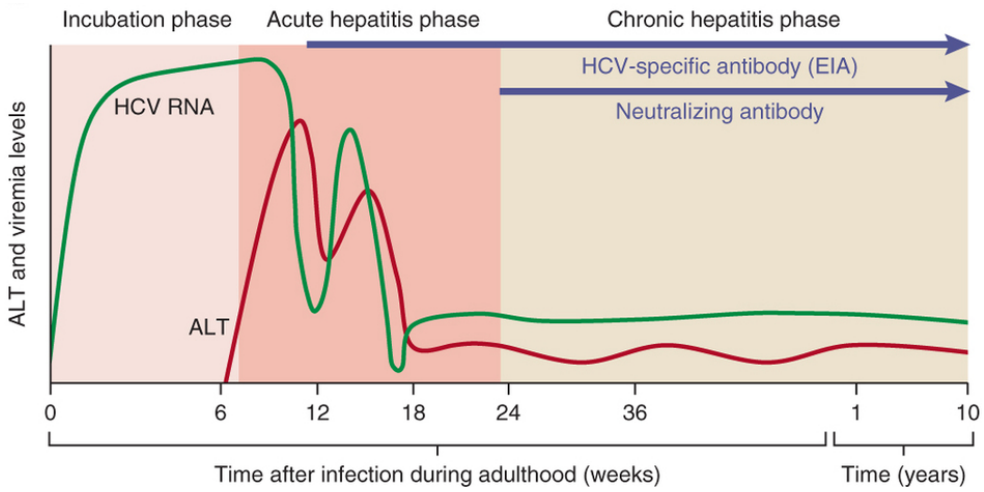



**Figure 4.2 Typical course of HCV infection.** During a typical course of HCV infection, the HCV VL fluctuates significantly during the first weeks of infection before stabilising during the chronic phase. (ALT = alanine aminotransferase, a liver enzyme used as a marker of liver damage). Reproduced from (Rehermann, 2013) with permission from Nature Publishing Group.

The selection of HLA-B*27:05 and -B*57:01-associated escape mutations in the autologous HIV sequence 2 years post-diagnosis suggest that the putative donor of the HIV superinfection may also be HLA-B*27:05 and/or HLA-B*57:01-positive, and thus transmitted an HIV variant carrying HLA associated footprints. This prompted the hypothesis that co-incident sexually transmitted HCV and HIV infection may have been transmitted from the same donor. To investigate this, the HCV genome was sequenced to look for evidence of HLA-B*27- or B*57-associated footprints. These HLA alleles are similarly associated with CTL-mediated selection pressure driving selection of HLA-associated footprints in the HCV genome (Dazert et al., 2009; Neumann-Haefelin et al., 2006, 2011).

The HCV NS5B protein (non-structural protein 5B, RNA polymerase) contains two important HLA-B*57- and three HLA-B*27-restricted epitopes (Table 4.3). The immunodominant NS5B-AF9 and subdominant NS5B-GY9 in particular, have been associated with selection of HLA-B*27-associated footprints (Neumann-Haefelin et

al., 2006, 2011). The NS5B gene of the autologous HCV virus from subject R097 was sequenced at 2 years post HIV diagnosis. HLA-B*27-associated polymorphisms were found in both the NS5B-AF9 (A2841V and M2846L) and NS5B-GY9 (I2940M) epitopes (Table 4.3). The presence of these footprints supports the hypothesis that HCV and an HIV superinfection may have been transmitted from the same HLA-B*27:05-positive donor.

**Table 4.3 Sequencing of HLA-B*27:05 and –B*57:01-restricted epitopes of the HCV NS5B gene from subject R097 at 2 years post HIV diagnosis.** The R097 sequence is aligned to the HCV genotype 1a consensus sequence. Epitope variants containing described HLA-B*27-associated footprints are shown in red.

| Epitope and HLA restriction | Epitope Sequence |
|---|---|
| HLA-B*57 NS5B-KF9 | $^{2629}$**KSKKTPMGF**$^{2637}$ |
| R097 2 years | --------- |
| HLA-B*27 NS5B-AW9 | $^{2820}$**ARHTPVNSW**$^{2828}$ |
| R097 2 years | --------- |
| HLA-B*27 NS5B-AF9 | $^{2841}$**ARMILMTHF**$^{2849}$ |
| R097 2 years | V----L--- |
| HLA-B*57 NS5B-LR10 | $^{2912}$**LGVPPLRAWR**$^{2921}$ |
| R097 2 years | --N------- |
| HLA-B*27 NS5B-GY9 | $^{2936}$**GRAAICGKY**$^{2944}$ |
| R097 2 year | ----M---- |

To further investigate the hypothesis that R097 acquired an HIV superinfection, evidence of multiple divergent HIV sequences in the intra-host population in ultra-deep sequencing data was investigated. Importantly, the escape mutations in Gag-IW9, Gag-TW10, Gag-KK10 and Pol-SW10 that had reached fixation at 2 years post-diagnosis had already been selected in minor variant populations at time 0 (Table 4.4), indicating that they may not have been acquired from a superinfection. Recombinant detection software was unable to find evidence of recombination

that would suggest superinfection. These data strongly suggest that, rather than being superinfected with a second strain carrying these escape mutations, rapid viral adaption at multiple epitopes occurred in this individual, and escape variants began to accumulate very early on in infection.

**Table 4.4 Ultra-deep RNA sequencing of HLA-B*57:01 and HLA-B*27:05-restricted epitopes and associated compensatory positions from R097 and his transmission partner R096.** Depth of coverage ranges from 46-85,000 reads. Epitopes at which predictable footprints (Carlson, Brumme, et al., 2012) are selected in the donor are shown. The early selection of escape and compensatory mutations in the minor variant populations is shown in red.

| HLA Class I Restriction and Epitope | B Clade Consensus and Autologous Variants | Haplotype Frequency (%) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Donor R096 (HLA-B*13:02/B*51:01) | Recipient R097 (HLA-B*27:05/57:01) | | | | | | | | | | |
| | | 0 Years | 0 Years | 2 Years | 2.1 Years | 2.2 Years | 2.25 Years | 2.3 Years | 2.4 Years | 2.6 Years | 2.7 Years | 3.1 Years | 3.2 Years | 3.4 Years |
| HLA-B*57 Gag-IW9 | $^{147}$ISPRTLNAW$^{155}$ | 98 | 86 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | |
| | L-------- | 2 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 7 | 13 | 9 | 14 | |
| | LT------- | 0 | 0 | 99 | 99 | 100 | 50 | 3 | 0 | 7 | 8 | 35 | 44 | |
| | -T------- | 0 | 0 | 0 | 0 | 0 | 48 | 96 | 99 | 86 | 78 | 55 | 41 | |
| | $^{146}$A | 99 | 80 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | P | 0 | 19 | 99 | 99 | 100 | 99 | 99 | 99 | 99 | 99 | 100 | 99 | 76 |
| | S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24 |
| HLA-B*57 Gag-KF11 | $^{162}$KAFSPEVIPMF$^{172}$ | 100 | 100 | 99 | 98 | 100 | 100 | 100 | 100 | 100 | 99 | 100 | 100 | 100 |
| HLA-B*57 Gag-TW10 | $^{240}$TSTLQEQIGW$^{249}$ | 96 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --------A- | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --------E- | 0 | 87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --N------- | 0 | 4 | 100 | 100 | 98 | 99 | 98 | 100 | 99 | 99 | 99 | 100 | 100 |
| | --N-----N- | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | $^{219}$H $^{223}$I $^{228}$M | 100 | 90 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | H V M | 0 | 10 | 93 | 94 | 95 | 97 | 97 | 98 | 99 | 99 | 99 | 99 | 99 |
| | H A M | 0 | 0 | 6 | 6 | 4 | 3 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| HLA-B*27 Gag-KK10 | $^{263}$KRWIILGLNK$^{272}$ | 17 | 5 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -----I---- | 83 | 92 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -K-------- | 0 | 3 | 100 | 100 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | $^{173}$S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | T | 98 | 87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | A | 0 | 10 | 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 | 100 | 99 |
| HLA-B*57 Pol-SW10 | STTVKAACWW | 98 | 0 | 0 | 0 | 0 | 0 | 3 | | 0 | 0 | 0 | 0 | 0 |
| | -N-------- | 0 | 50 | 74 | 86 | 57 | 69 | 63 | | 53 | 60 | 76 | 81 | 100 |
| | --A------- | 0 | 50 | 0 | 0 | 4 | 0 | 2 | | 0 | 0 | 0 | 0 | 0 |
| | -A-------- | 0 | 0 | 24 | 13 | 38 | 30 | 32 | | 45 | 39 | 23 | 18 | 0 |
| | -AM------- | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 1 | 0 | 0 | 0 | 0 |

### 4.3.3 Control of HIV infection in two HLA-B*27:05/B*57:01-postive subjects is associated with viral sequence conservation

To investigate the rate of viral sequence change over time in cases of slow disease progression (in contrast to subject R097), two HLA-B*27:05/B57:01-positive slow progressing subjects, RI088 and RI422 were studied. Both subjects maintained high CD4+ T cell counts and low VL for over a decade of follow up (Figure 4.3A and B). In contrast to the viral sequence divergence seen in R097 in just two years, sequences from these subjects were remarkably conserved over time (Figure 4.3C). Both subjects maintained nucleotide sequence identity of ≥98.8% across all sequences.

As RI088 had very low VL values, ultra-deep viral sequencing from RNA could not be performed. Clonal sequencing of *gag* from proviral DNA for RI088 from six timepoints spanning 10 years was used to look for evidence of CTL selection pressure on the virus in minor variant populations, despite tight conservation of the consensus sequence. Similar to the rapid progressor R097, there was evidence of selection pressure driven by p24 Gag-specific CTL (Table 4.5). However, only the T242N escape mutation in Gag-TW10 and A146P mutation in Gag-ISW9 reached fixation. Although transient variants arose in the KF11 and KK10 epitopes, these did not reach fixation. In contrast to the rapid fixation of the escape mutations seen in R097, fixation of mutations in the TW10 and ISW9 epitopes were only seen in RI088 after more than 10 years.

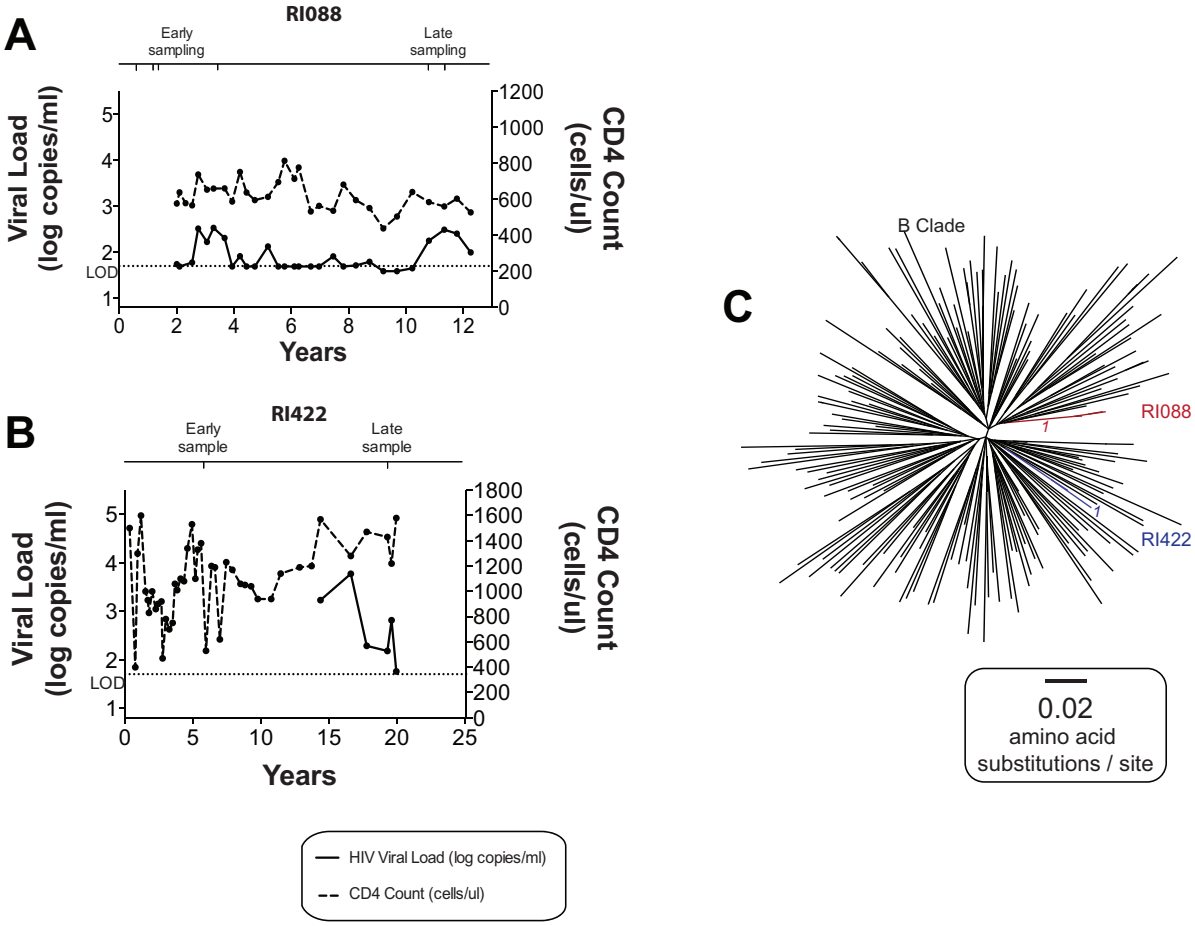



**Figure 4.3 Clinical course of infection and phylogenetic analysis of viral sequences from HLA-B*27:05/57:01-positive controllers RI088 and RI422.** A) Plasma HIV viral load and CD4+ T cell count for HLA-B*27:05/B*57:01-positive recipient RI088 and B) RI422. Time '0' represents the time of diagnosis in each case. The horizontal dotted line represents the limit of detection (LOD) of the viral load assay (40 copies/ml). A timeline indicates the time of sampling for each subject. C) Maximum likelihood phylogenetic tree of 1089bp alignment of the Gag gene sequenced from proviral DNA. Subjects RI088 (n=6) and EC RI422 (n=2) are shown in red and blue, spanning 10 and 13 years respectively. 180 B clade reference sequences from the US and UK collected between 2003 and 2011 from the Los Alamos database (http://www.hiv.lanl.gov/) are shown in black. Bootstrap values for the RI088 and RI422 sequence clusters based on 1,000 bootstrap replicates are shown in italics.

**Table 4.5 Clonal sequencing of HIV proviral DNA at HLA-B*57:01 and HLA-B*27:05-restricted Gag epitopes from controller RI088.** The number of clones sequenced at each timepoint and number of independent PCR reactions from which PCR product was cloned are given. The percentage of clones with each sequence haplotype is shown. The most prevalent (consensus) epitope sequence is shown in black. Lower frequency variant sequences are shown in grey.

| | | | HLA-B*57:01-Restricted Epitopes | | | | | | HLA-B*27:05-Restricted Epitopes | |
|---|---|---|---|---|---|---|---|---|---|---|
| Years Since Diagnosis | Number of clones sequenced | Number of independent PCRs | $^{146}$A ISPRTLNAW$^{155}$ | % | $^{162}$KAFSPEVIPMF$^{172}$ | % | $^{240}$TSTLQEQIGW$^{249}$ | % | $^{263}$KRWIILGLNK$^{272}$ | % |
| 0.6 | 17 | 1 | - --------- | 100 | ---------- | 100 | ---------- | 94 | ---------- | 100 |
| | | | | | | | I--------- | 6 | | |
| 1.2 | 17 | 2 | - --------- | 88 | ---------- | 82 | ---------- | 82 | ---------- | 100 |
| | | | P L-------- | 12 | ------T--- | 6 | --N------- | 12 | | |
| | | | | | -------L-- | 6 | ----R----- | 6 | | |
| | | | | | ------M--- | 6 | | | | |
| 1.4 | 20 | 2 | P L-------- | 55 | ---------- | 95 | --N------- | 55 | ---------- | 90 |
| | | | - --------- | 35 | --------T- | 5 | ---------- | 45 | ---V------ | 5 |
| | | | - T---A---- | 5 | | | | | ----V----- | 5 |
| | | | - ------S-- | 5 | | | | | | |
| 3.4 | 6 | 1 | - --------- | 83 | ---------- | 83 | ---------- | 100 | ---------- | 100 |
| | | | - --T------ | 17 | ------M--- | 17 | | | | |
| 10.8 | 26 | 2 | P --A------ | 92 | ---------- | 96 | --N------- | 100 | ---------- | 100 |
| | | | P --A---S-- | 8 | --------I- | 4% | | | | |
| 11.3 | 17 | 2 | P --A------ | 100 | ---------- | 88 | --N------- | 94 | ---------- | 100 |
| | | | | | --------T- | 6% | --NP------ | 6 | | |
| | | | | | --------V- | 6% | | | | |

### 4.3.4 Selection of unusual variants in controller RI088

Previous work has shown that unusual mutations may be selected in HLA-B*57/B*58:01+ controllers, but not in progressors (Miura, Brockman, Schneidewind, et al., 2009). In keeping with this observation, in subject RI088 the well-described I147L mutation in Gag-ISW9 was lost in viral sequences from later timepoints, while the much more unusual mutation P149A reached fixation. This mutation is present in only 3.3% of B clade sequences (n=5,204) from the Los Alamos database (http://www.hiv.lanl.gov/) and may represent an additional variant to those previously described.

### 4.3.5 The effect of viral sequence polymorphisms on viral replicative capacity in HLA-B*27:05/B*57:01-positive controllers and progressors

The HLA-B*27:05/B*57:01-positive controller RI088 and progressor R097 thus both accumulated polymorphisms in Gag. The effect of these multiple polymorphisms on VRC was investigated (Figure 4.4 A and B).

Compensatory mutations may partially or completely restore VRC of CTL escape mutations. To cause progressive disease in subject R097, the VRC of the autologous virus must have been sufficiently compensated against the multiple costly escape mutations. Indeed, co-selection of S173A that compensates for the fitness cost of R264K in Gag-KK10, and I223V (in the CypA binding loop), which compensates for T242N in Gag-TW10 at the time of escape was observed (Table 4.4). The VRC of

virus from R097 at the time of diagnosis was substantially reduced compared to NL4-3 and to the virus from the donor R096 (Figure 4.4). A further drop in VRC in the virus from 2 years later was observed with some apparent recovery of VRC at later timepoints. Although selection of compensatory mutations likely contributed to the partial restoration of VRC, the VRC of R097 viral sequences remained consistently low compared to W/T.

In contrast, the VRC of the autologous virus from RI088 was a high replicative variant, with no significant difference compared to W/T (Figure 4.4B). Thus, RI088 maintained control of infection with a high VRC virus, while R097 progressed despite low VRC of the autologous virus. Therefore, differences in the rate of disease progression in these cases cannot be attributed to the VRC of the autologous virus.

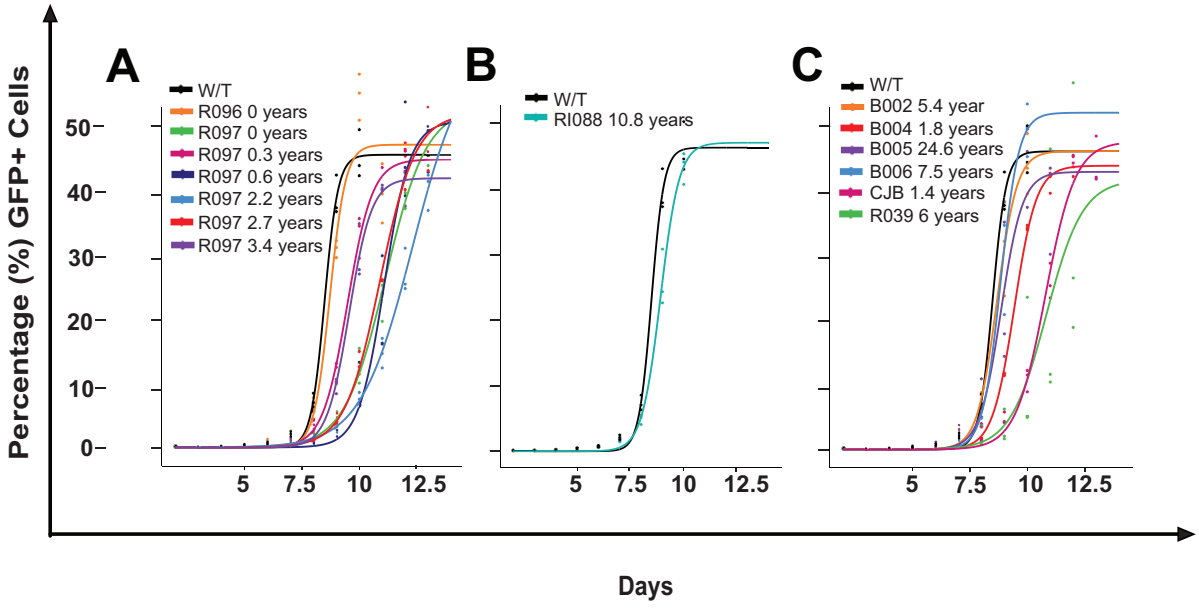



**Figure 4.4 Viral replicative capacity (VRC) of recombinant viruses produced from autologous *gag-pro* from HLA-B*27:05/B*57:01-positive subjects.** A) VRC of *gag-pro* chimeric virus derived from A) longitudinal sampling of progressor R097 and his transmission partner R096, B) controller RI088, and C) other HLA-B*27:05/B*57:01-positive subjects, all compared to a wildtype (WT) NL4-3 control. VRC is given by the percentage of infected (GFP-expressing) target cells.

### 4.3.6 Hierarchy of the CTL response differs in HLA-B*27:05/B*57:01-positive controllers and progressors

To compare the hierarchy of the HIV-specific CTL responses in controller RI088 and progressor R097, PBMC from these subjects were tested against a panel of peptides representing optimal HLA-B*27 and HLA-B*57-restricted epitopes in IFN-γ ELISpot assays (Figure 4.5). The CTL response in RI088 was dominated by responses to Gag, with an immunodominant KK10 response while Env responses were least dominant. This hierarchy in the CTL response in a case of control is consistent with reports that a Gag-driven immune response is the most effective response against HIV (Kiepiela et al., 2007). In contrast, the immunodominant CTL responses were directed against Vpr, Pol and Nef in R097. The escape mutations selected in multiple Gag epitopes in this subject (Table 4.2) resulting in a shift in the immunodominance hierarchy away from a Gag-directed response, may be associated with loss of immune control in this case.

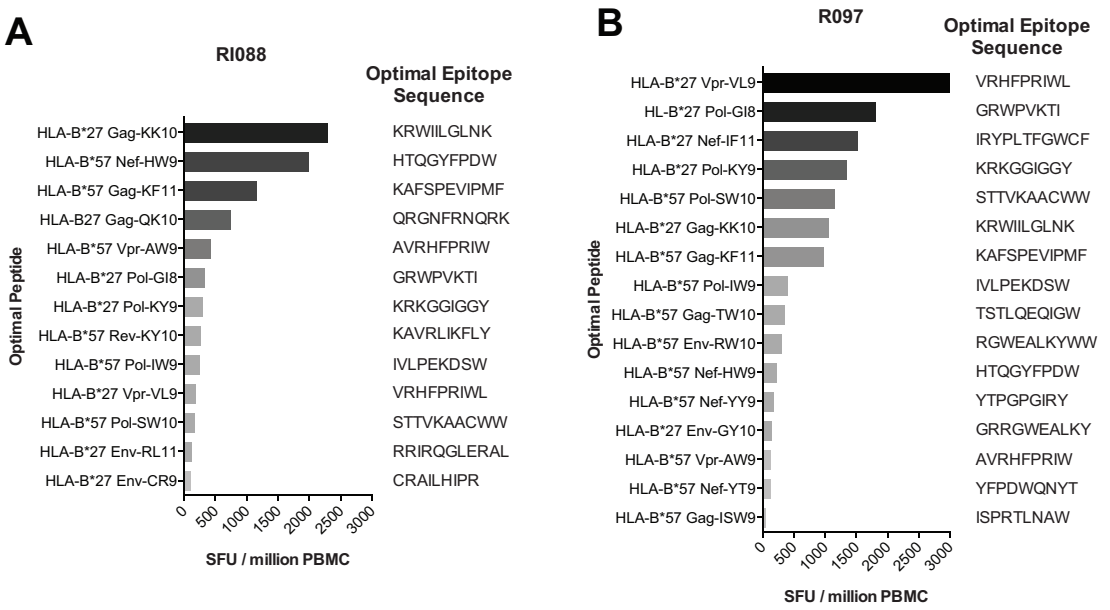



**Figure 4.5 The CTL response hierarchy in an elite controller and chronic progressor.** CTL responses were measured using IFN-γ ELISPot assays testing a panel of HLA-B*27:05 and HLA-B*57:01-restricted optimal epitope peptides. The sequence of the optimal peptide is given alongside the bar graph. A) CTL response hierarchy of elite controller RI088 at 11.3 years of follow up. B) The CTL response hierarchy of progressor R097 at 2.7 years of follow up.

### 4.3.7 Characteristics of failed immune control in additional HLA-B*27:05/B*57:01-positive subjects

Ten additional subjects expressing HLA-B*27:05/B*57:01, including five disease progressors, were studied to further investigate the causes of failure of these alleles to control HIV infection (Table 4.6). First, CTL mediated selection pressure was characterised in these subjects. The I147L and T242N escape mutations in Gag-ISW9 and Gag-TW10, respectively, were both present in 10/13 subjects, including some who maintained control of viraemia. It thus seems possible to maintain immune control in the presence of these escape mutations, even in the presence of co-selected compensatory mutations. Selection of R264K in Gag-KK10 was seen in 3/13 subjects, all of whom failed to control HIV viraemia. Selection of

Gag-KK10 escape may thus be an important marker of disease progression in HLA-B*27:05/B*57:01-positive subjects.

The selection of G248D in Gag-TW10, often in the absence of T242N, has been associated with HLA-B*57/B*58:01-positive controllers, but not with progressors (Miura, Brockman, Schneidewind, et al., 2009). Consistent with this report, G248D was observed in the absence of T242N in three subjects, and other G248 variants in four additional subjects. 5/7 of these subjects maintained viraemic control.

Next, the association of VRC of the autologous virus with disease outcome was investigated (Figure 4.4C). Although VRC may have contributed to disease outcome in some cases, for example, transmission of a low VRC variant in controller R039, this was not true of all the subjects studied. For example, subject CJB 021111 had one of the weakest virus variants but still lost immune control. Although VRC of the autologous variant may contribute to disease outcome in some cases, it was not a defining feature of the controller and progressor phenotypes.

**Table 4.6 Consensus sequences of HLA-B*57:01 and HLA-B*27:05-restricted Gag epitopes in ART naïve adults expressing these alleles.** Subjects are listed in ascending order of HIV viral load. Time is shown in years since diagnosis or enrollment (timing of infection unknown). ND=not done, Green=clonal sequencing from plasma RNA, black=ultra-deep sequencing from plasma RNA, dark blue=clonal sequencing from proviral DNA, light blue=population level sequencing from proviral DNA.

| | | Years | CD4+ T cell count (cells/µl) | HIV viral load (copies/ml) | HLA-B*57:01-Restricted Epitopes | | | | HLA-B*27:05-Restricted Epitopes | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $^{146}$A ISPRTLNAW$^{155}$ | $^{162}$KAFSPEVIPMF$^{172}$ | $^{240}$TSTLQEQIGW$^{249}$ | $^{219}$HAGPIAPGQM$^{228}$ | $^{173}$S | $^{263}$KRWIILGLNK$^{272}$ |
| Controllers | B003 | 22.8 | 425 | <50 | - L-------- | ---------- | --N-----A- | ---------- | T | -----M---- |
| | B005 | 24.6 | 1096 | 65 | P L-------- | ---------- | --N-----A- | ---------- | - | ---------- |
| | B001 | 26.6 | 963 | 103 | - --------- | ---------- | --------D- | ---------- | - | ---------- |
| | RI422 | 5.8 | 530 | ND | - --------- | ---------- | --------D- | ---------- | - | ---------- |
| | | 19.3 | 1430 | 152 | - --------- | ---------- | --------D- | ---------- | - | ---------- |
| | RI088 | 10.8 | 584 | 178 | P --A------ | ---------- | --N------- | ---------- | - | ---------- |
| | B004 | 1.8 | 955 | 348 | - L-------- | ---------- | --N------- | A---V----- | T | ---------- |
| | B002 | 5.4 | 996 | 879 | P L-------- | ---------- | --N-----Q- | Q---V----I | A | ---------- |
| | R039 | 1.8 | 790 | 504 | - L-------- | ---------- | --N------- | Q---V----- | - | -----M---- |
| | | 2.4 | 590 | 972 | - L-------- | ---------- | --N------- | ---------- | - | -----M---- |
| Progressors | CJB | 1.4 | 718 | 980 | - L-------- | ---------- | --N----V-- | Q---V----- | - | -----M---- |
| | B006 | 1.7 | 870 | <50 | - L-------- | ---------- | --N-----A- | ---------- | T | ---------- |
| | | 4.9 | 791 | 3020 | P --------- | ---------- | --N-----N- | ---------- | A | -K---I---- |
| | | 7.5 | 777 | 4180 | S L-------- | -----I---- | --N-----N- | ---------- | A | -K---I---- |
| | B007 | 22.7 | 1031 | 6360 | - L-------- | ---------- | --------D- | ---------- | - | ---------- |
| | Warsaw79 | 2.7 | 807 | 13869 | P L-------- | ---N------ | --N------- | ---------- | - | -K-------- |
| | R097 | 3.2 | 250 | 150443 | P LT------- | ---------- | --N------- | ----V----- | A | -K-------- |

## 4.4 Discussion

A number of studies have attempted to define the qualitative and quantitative differences in CTL response that lead to divergent HIV disease outcome. Determinants include targeting of the most conserved viral epitopes, in particular Gag p24 (Borghans et al., 2007; Dahirel et al., 2011; Kiepiela et al., 2007), viral escape from the CTL response (Goulder, Phillips, et al., 1997) and CTL driven selection of mutations that impair viral fitness (Miura, Brockman, Schneidewind, et al., 2009). In a study that controlled for factors related to viral sequence variation and CTL escape, divergent outcomes in progressor and controller groups expressing the same protective HLA alleles was attributed to selection of a more potent TCR repertoire that was more cross-reactive to epitope variants in the controller group (Chen et al., 2012). Here, the role of intra-host viral evolution, CTL escape and immunodominance, and viral fitness in divergent disease outcome in a similar setting of shared HLA allele expression was investigated. In addition to dual expression of both HLA-B*27:05 and B*57:01, the HLA-A and C alleles expressed by this cohort were fairly conserved (Table 4.1).

The case studies of subjects RI088, RI422 and R097 here have shown that disease progression in the setting of HLA-B*27:05/B*57:01 expression may be associated with rapid intra-host viral evolution and selection of CTL escape, while control is associated with longitudinal sequence conservation and minimal selection of CTL escape. These data are consistent with a previous study that showed that the rate of viral evolution, specifically within Gag p24, distinguished between HLA-

B*57:01-postive subjects who were low- and high-risk progressors (Norstrom et al., 2012).

Although evolution is tightly constrained in controllers by low viral turn-over, some evidence of low level evolution in minor variant populations with selection of unusual mutations was found, consistent with previous reports (Mens et al., 2010; Miura, Brockman, Schneidewind, et al., 2009). As sequencing for subject RI088 was done from proviral DNA, these variants may have been archival sequences not represented in the circulating viral population, but may none-the-less reflect a low level of viral evolution over several years.

HIV superinfection has been associated with accelerated disease progression and loss of immune control (Clerc, Colombo, Yerly, Telenti, & Cavassini, 2010; Jost et al., 2002; Streeck, Li, et al., 2008). This was considered as a factor contributing to poor outcome in the case study of R097. However, very early selection of escape mutations was detected in minor variant populations. Although this does not exclude the possibility of superinfection with a virus containing an abundance of mutations within HLA-B*27:05/B*57:01 epitopes prior to time 0, it is more likely that multiple escape mutations were rapidly selected in R097, with consequent reduction in CTL recognition and loss of immune control. A lack of existing data on how early low frequency escape mutations can be detected in minor variant populations in late-escaping epitopes is, however, a caveat of this interpretation of the data. Although the degree of sequence diversity seen in subject R097 over the period of follow up is greater than expected in a single individual in the absence of superinfection, phylogenetic analysis shows that all R097 sequences have the same

MRCA (Figure 4.1C), supporting a case of unusually rapid intra-host evolution, rather than a case of superinfection.

The development of immunological memory that protects against subsequent infection after the first exposure to a pathogen is an important premise of vaccination strategies. Understanding how often and under what circumstances HIV superinfection occurs thus has important implications for understanding the lack of protection against multiple infections (Altfeld et al., 2002; Blish et al., 2012). The growing number of global CRFs attests to the occurrence of inter-clade dual infection which can be observed through detection of recombination between strains. This suggests that HIV infection may not provide cross-clade protection. In contrast, the case study of R097 presented here demonstrates the complexity involved in accurately detecting an intra-clade superinfection. This is important for making inferences regarding the extent to which infection does or does not afford intra-clade protection against subsequent infection. Here, analysis of genetic distance and phylogenetic clustering alone have been shown to be insufficient. A number of studies of intra-clade HIV superinfection have, however, relied heavily on these methods (Altfeld et al., 2002; Casado et al., 2007; Gottlieb et al., 2007; Gottlieb, 2004). Detection of recombination between strains in addition to sequence diversity provides a stronger case for superinfection (Streeck, Li, et al., 2008). Simultaneous detection of multiple distinct intra-host populations also provides strong evidence. Ultra-deep sequencing has thus emerged as an important method for detecting superinfection (Redd, Quinn, & Tobian, 2013; Ronen et al., 2013; Wagner et al., 2014).

It is well recognised that HIV and HCV co-infection is associated with poorer outcome of both infections (Sulkowski, 2013; Webster, Klenerman, & Dusheiko, 2015). IFN-α/Ribovirin therapy may further contribute to poor HIV disease outcome related to the CD4+ T cell decline typically associated with this treatment (Chung et al., 2004; Soriano et al., 1994). Although HCV co-infection in subject R097 may have contributed to the rate of HIV disease progression from 2 years post-diagnosis, a decline in CD4+ T cell count and rising HIV VL was evident before this time. Early selection of CTL escape likely set this subject on the path to disease progression prior to HCV co-infection.

A number of factors may have contributed to differences in viral control in HLA-B*27:05/B*57:01-positive subjects. Although the distinction between the controllers and progessors in terms of the levels of CTL escape is not as clear in all subjects as in the case studies of subjects RI088, RI422 and R097, the mutations selected in these subjects likely played an important role in shaping the CTL response and autologous viral fitness that lead to divergent outcomes. Other factors beyond the scope of this study may also have contributed, including levels of T cell activation and exhaustion (Deeks et al., 2004), CTL proliferative capacity (Day et al., 2007; Migueles et al., 2002) and polyfunctionality (Almeida et al., 2007; Betts et al., 2006) and NK cell responses (Altfeld & Gale, 2015). The overall CTL response breadth and immunodominance hierarchy (as studied in subjects RI088 and R097) may also have contributed to the disease outcome in these groups. Importantly, there is some evidence to suggest that elite and viraemic control can be attributed to the VRC of the transmitted founder virus (Claiborne et al., 2015; Mariani et al., 1996; Michael et al., 1995; Miura et al., 2010; Yue et al., 2015).

Although the subjects in this study were chronically infected, precluding determination of transmitted founder VRC, this study presented a rare opportunity to address the importance of VRC in disease outcome in the setting of shared HLA-alleles, thus partially removing HLA as a contributing factor to immune control. Although VRC of the autologous virus may contribute to disease outcome in some cases, it was not consistently associated with outcome in all of the subjects. This supports the hypothesis that host factors are equally important for achieving disease control of HIV.

Dual expression of HLA-B*27:05 and B*57:01 is a rare combination that occurs in <1% of the Caucasian population, based on an estimated prevalence of HLA-B*27:05 and –B*57:01 expression of 10% and 7%, respectively. Sample size was thus a limiting factor in this study. Obtaining viral sequence data from subjects with extremely low VL can be problematic, and could only be achieved using proviral DNA in some cases. The use of B clade consensus peptides in ELISpot assays, rather than peptides matched to the autologous viral sequence, may have resulted in an underestimation of the CTL response to autologous epitope variants. Despite these limitations, this study has highlighted the interaction between host and viral factors associated with disease control and progression in subjects expressing favourable HLA alleles, including rate of intra-host evolution, patterns of CTL escape, and VRC.

# CHAPTER 5: Disease progression despite protective HLA expression in HIV-infected transmission pairs

The majority of the work presented in this chapter, has been published:
Brener *et al. Retrovirology* 2015 12(55). Disease progression despite protective HLA expression in an HIV-infected transmission pair.

## 5.1 Introduction

Studies of HIV infected transmission pairs may provide a number of valuable insights for understanding HIV pathogenesis, since these studies allow for direct comparison of donor and recipient sequence. In cases where a very early sequence is available from the recipient, this sequence may provide a close approximation of of the TF virus. In cases where an early recipient sequence is not available, the donor sequence (obtained as close to the time of transmission as possible) may be used to approximate the TF sequence. The latter approach is, however, limited by the assumption that the most prevalent sequence in the donor is most likely to be transmitted. Comparison of donor and recipient sequences are particularly informative for studying intra-host evolution and the effect of HLA-specific selection pressure on a particular baseline (TF) sequence. These studies also allow approximation of the viral fitness of the TF virus, which may be an important determinant of disease outcome (Claiborne et al., 2015; Mariani et al., 1996; Michael et al., 1995; Miura et al., 2010; Yue et al., 2015).

One mechanism believed to be important in contributing to HLA associations with characteristic disease outcomes is the targeting of specific CTL epitopes (Borghans et al., 2007; Boutwell et al., 2009; Brockman et al., 2010; Crawford et al., 2009;

Kiepiela et al., 2007; Kloverpris et al., 2012; Matthews et al., 2008; Miura, Brockman, Schneidewind, et al., 2009; Sacha et al., 2007). The subtype (clade) of HIV infection may therefore impact on disease control, by affecting the availability of specific T cell epitopes (Kløverpris et al., 2013; Matthews et al., 2009; McKinnon et al., 2009). The profile of CD8+ T cell responses to CRF01_AE clade infection may be substantially different from that of other clades (Buranapraditkun et al., 2011; Kantakamalakul et al., 2006; Mori et al., 2011; Sriwanthana et al., 2012). The Gag CRF01_AE clade consensus sequence bears the polymorphisms which match HLA-B*57/B*58:01-selected footprints including A146P, I147L, A163G and S165N (Buranapraditkun et al., 2011; Matthews et al., 2009).

This case study of an HIV infected transmission pair pursues two hypotheses; 1) failure to mediate immune control of HIV in HLA-B*27:05/B*57:01 positive adults was related to the presence of existing escape mutations in the CRF01_AE clade sequence of the TF virus, and 2) the CD8+ T cell epitopes important for immune control are those in which escape is selected in association with, or prior to, disease progression.[13]

[13] A second transmission pair with an unusual combinations of viral clade and HLA type; subjects R036 and R056 described in chapter 3, were also studied. Ongoing work analysing this pair is described in Appendix III.

## 5.2 Methods

The following methods were used in this study as described in Chapter 2:

- Quantification of HIV viral load and CD4+ T cell count (section 2.2)
- HLA Class I typing (section 2.4)
- DNA extraction from whole blood (for HLA Class I typing) (section 2.3.3)
- RNA extraction from plasma (for viral sequencing) (section 2.3.4)
- IFN-γ ELISpot assays (section 2.6.1)
- Ultra-deep sequencing with Illumina MiSeq technology (incuding *de novo* assembly of consensus sequences) (section 2.10.1 and 2.10.2)
- Analysis of ultra-deep sequencing data (minor variant analysis, determination of epitope haplotypes, and heat map analysis) (section 2.10.3 - 2.10.5)
- Phylogenetic analysis (sequence subtyping and maximum likelihood phylogenetic analysis) (section 2.11.1 and 2.11.2)
- Peptide-MHC tetramer staining (section 2.6.3)

### 5.2.1 Additional details

*Study subjects*

This adult Caucasian transmission pair was recruited from the Thames Valley Cohort (Table 2.1). The male donor, infected prior to 2007 subsequently infected his female partner. Both subjects gave written informed consent for their participation. Ethical approval was given by the Oxford Research Ethics Committee.

*Identification of known and predicted epitopes*

Known epitopes were identified from the Los Alamos Immunology Database CTL Epitopes A-List (Llano, Williams, Olvera, Silvia-Arrieta, & Brander, 2013). Predicted epitopes were identified from described HLA footprints (Llano et al., 2013) and peptide motifs (Table 1.2) and the HLArestrictor tool (Erup Larsen et al., 2011).

## 5.3 Results

### 5.3.1 Progression in a UK transmission pair with CRF01_AE virus infection

A Caucasian transmission pair from the UK was studied. The male donor N094 (HLA-A*02:01/32:01 B*13:02/14:01 C*06:02/08:02) is believed to have acquired HIV infection in Thailand, and to have infected the female recipient N107 (HLA-A*02:01/02:01 B*27:05/57:01 C*01:02/06:02). Both partners were diagnosed more than 2 years later when the recipient was HIV-tested during pregnancy (referred to as 'time 0').

Maximum likelihood analysis and the Rega HIV-1 Subtyping Tool were used to confirm that the transmission pair was infected with CRF01_AE clade virus (Figure 5.1A), the recombinant virus prevalent in Thailand (Buonaguro, Tornesello, & Buonaguro, 2007; Buranapraditkun et al., 2011). The close phylogenetic relationship of the donor and recipient viruses suggested that these subjects were likely to be transmission partners (Figure 5.1B). An HLA-B*14:01-associated escape mutation, Gag-K302R (within the Gag-DA9 epitope (Carlson, Brumme, et al., 2012)) was present in the HLA-B*14:01-positive donor's autologous virus. This was transmitted to the HLA-B*14:01-negative recipient where it reverted to wildtype (Table 5.1). This provides evidence to support the direction of transmission suggested by the clinical history. In contrast, the HLA-B*27:05 and HLA-B*57:01-driven mutants observed in the recipient were not present in the donor.

The HLA-B*27:05/57:01-positive recipient progressed to an absolute CD4 count of <350cells/mm$^3$, meeting the criteria for initiation of ART within 4.5 years of diagnosis (Figure 5.2B). The donor also progressed to ART initiation over a similar time period from diagnosis (Figure 5.1A) despite expressing three HLA molecules that have also been associated with some degree of protection against disease progression, HLA-A*32:01, HLA-B*13:02 and HLA-B*14:01 (Fellay et al., 2009; Honeyborne et al., 2007; Pereyra et al., 2010).

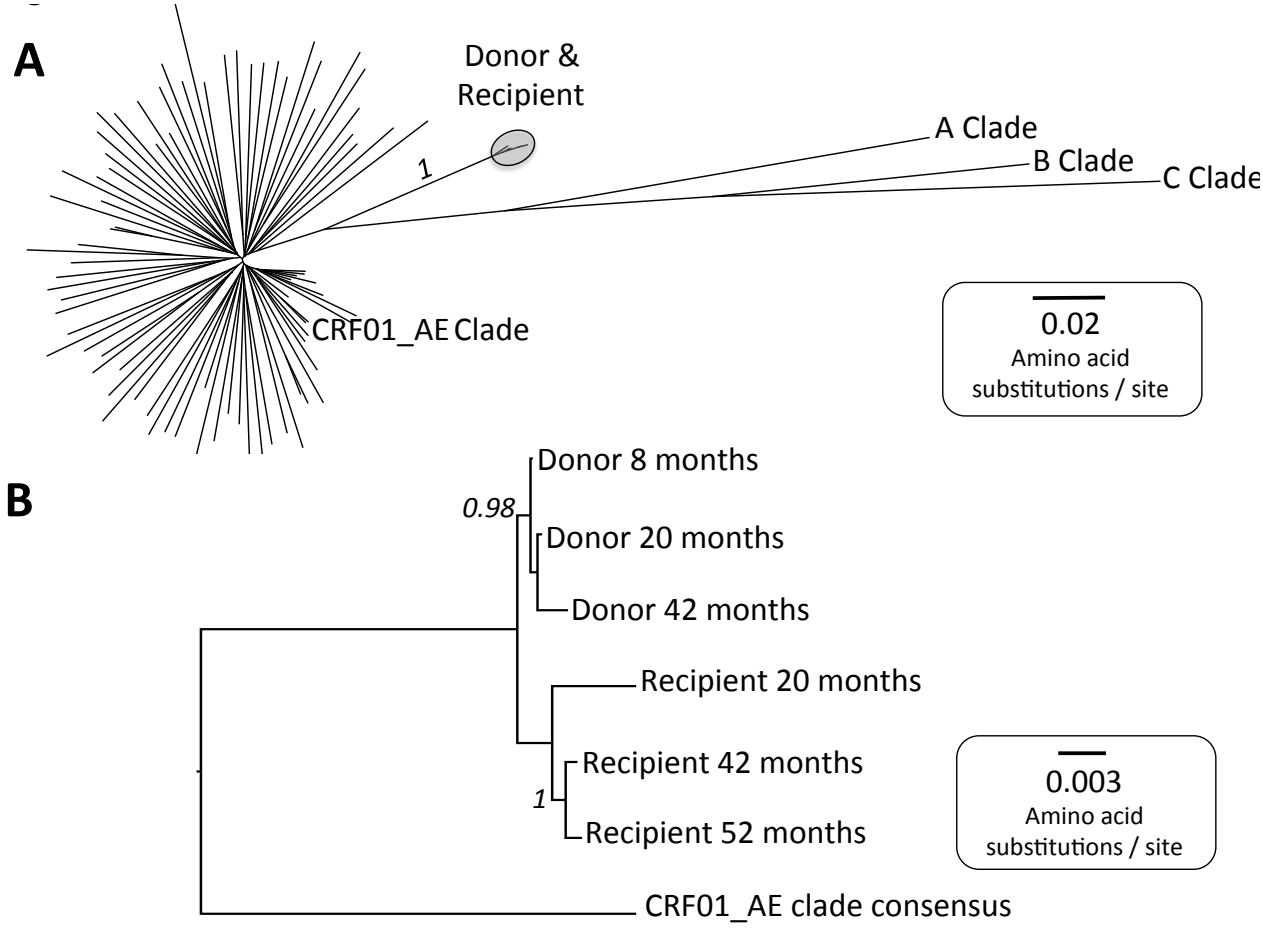



**Figure 5.1 Phylogenetic trees demonstrating subtype and genetic proximity of sequences from HIV transmission pair N094/N107** A) Maximum likelihood phylogenetic tree of 92 6,803bp nucleotide sequences, including consensus sequences for clades A, B, C and CRF01_AE, donor sequences from 8, 20 and 42 months post-diagnosis and recipient sequences from 20, 42 and 52 months post-diagnosis and 82 CRF01_AE Clade sequences from Thailand (Los Alamos HIV database, http://www.hiv.lanl.gov/). Donor and recipient sequences (circled) cluster with CRF01_AE clade sequences from Thailand, confirming that the Thai epidemic was the likely source of infection. The bootstrap value based on 1,000 bootstrap replicates for the donor-recipient cluster is shown in italics. B) The close relationship between the donor and recipient sequences supports transmission between these two subjects. Bootstrap values >0.75 based on 1,000 bootstrap replicates are shown in italics. Scale bars the show number of substitutions per site.

**Table 5.1 Sequence changes within Gag epitopes restricted by HLA-B*14, B*27 and B*57 identified by ultra-deep sequencing of HIV transmission pair N094/N107 aligned to CRF01_AE clade consensus (bold) and B clade consensus sequence (grey).** The frequency of minor variant haplotypes at the HLA-B*14 Gag-DA9, HLA-B*27 Gag-KK10, HLA-B*57 Gag-ISW9, HLA-B*57 Gag-KF11 and HLA-B*57 Gag-TW10 epitopes above a cut-off of 1% are shown. The depth of coverage ranges from 560 to 240,000 reads. Position 146 (flanking HLA-B*57 Gag-ISW9), associated with selection of an HLA-B*57/58:01-selected A146P processing mutation in B clade infection, is also shown. Minor variant frequencies are rounded off to the nearest 1%.

| HLA Class I Restriction and Epitope | AE and B Clade Consensus and Autologous Variants | Haplotype Frequency (%) | | | | | |
|---|---|---|---|---|---|---|---|
| | | Donor N094 (HLA-B*13:02/14:01) | | | Recipient N107 (HLA-B*27:05/57:01) | | |
| | | 8 months | 20 months | 42 months | 20 months | 42 months | 52 months |
| HLA-B*57 Gag-IW9 | $^{146}$**P LSPRTLNAW**$^{155}$ | 2 | 21 | 0 | 100 | 0 | 2 |
| | A I--------- | | | | | | |
| | S --------- | 86 | 77 | 100 | 0 | 100 | 98 |
| | - I--------- | 1 | 0 | 0 | 0 | 0 | 0 |
| | - V--------- | 10 | 2 | 0 | 0 | 0 | 0 |
| HLA-B*57 Gag-KF11 | $^{162}$**KGFNPEVIPMF**$^{172}$ | 100 | 100 | 100 | 100 | 85 | 45 |
| | -A-S-------- | | | | | | |
| | --------V--- | 0 | 0 | 0 | 0 | 15 | 55 |
| HLA-B*57 Gag-TW10 | $^{240}$**TSTLQEQIGW**$^{249}$ | 99 | 100 | 99 | 0 | 0 | 0 |
| | ---------- | | | | | | |
| | --N-------- | 0 | 0 | 0 | 100 | 100 | 100 |
| HLA-B*27 Gag-KK10 | $^{263}$**KRWIILGLNK**$^{272}$ | 100 | 100 | 100 | 0 | 4 | 0 |
| | ---------- | | | | | | |
| | -----M----- | 0 | 0 | 0 | 0 | 0 | 15 |
| | -----I----- | 0 | 0 | 0 | 0 | 0 | 28 |
| | ---------H- | 0 | 0 | 0 | 100 | 27 | 56 |
| | -----M--H- | 0 | 0 | 0 | 0 | 61 | 0 |
| | -----V--H- | 0 | 0 | 0 | 0 | 8 | 0 |
| HLA-B*14 Gag-DA9 | $^{298}$**DRFYKTLRA**$^{306}$ | 36 | 12 | 4 | 0 | 100 | 100 |
| | --------- | | | | | | |
| | ----R---- | 64 | 87 | 95 | 100 | 0 | 0 |

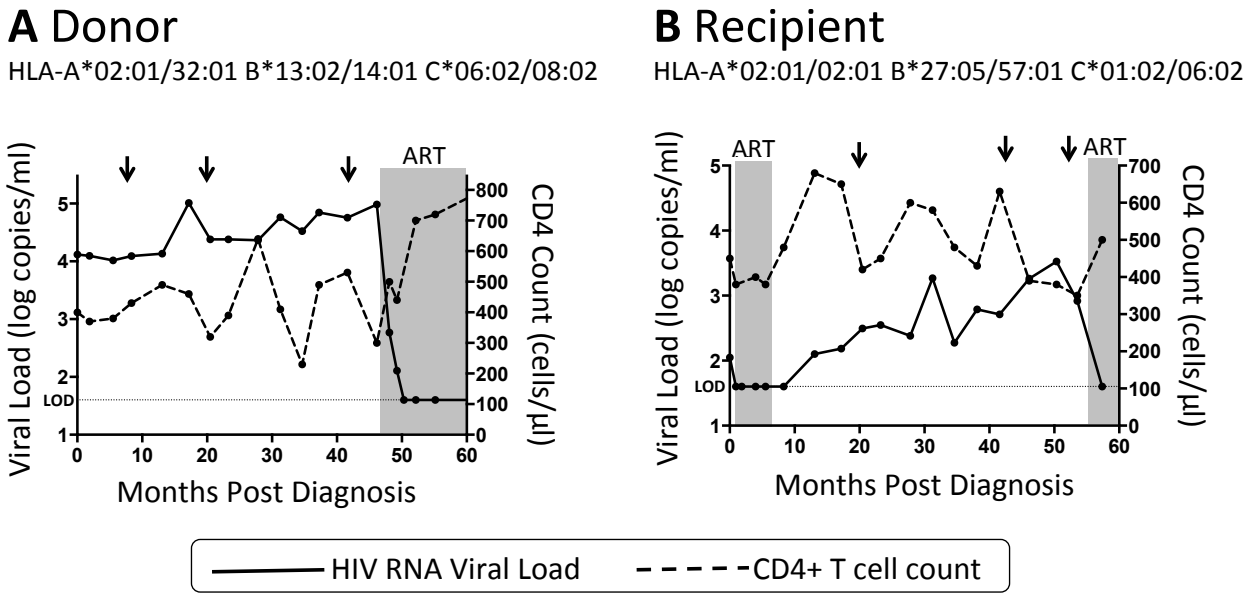



**Figure 5.2 Course of infection in HIV transmission pair N094/N107.** A) Longitudinal plasma HIV RNA viral load and CD4+ T cell counts for the donor over 60 months of follow up. Grey shading represents the period during which the subject received antiretroviral therapy (ART) following decline of the CD4+ T cell count to <350cells/mm3 (Williams et al., 2014). The horizontal dotted line represents the limit of detection (LOD) of the viral load assay (40copies/ml). Arrows indicate the time of sampling for ultra-deep sequencing. B) Longitudinal plasma HIV RNA viral load and CD4+ T cell counts for the recipient showing disease progression over 54 months of follow up. Grey shading represents the periods during which the subject received antiretroviral therapy (ART), initially as peripartum prophylaxis and subsequently following decline of the CD4+ T cell count to 350cells/mm3 at 55 months post-diagnosis (Williams et al., 2014). The horizontal dotted line represents the limit of detection (LOD) of the viral load assay (40copies/ml). Arrows indicate the time of sampling for ultra-deep sequencing.

### 5.3.2 HLA-B*27 and –B*57 Gag escape mutations in the recipient N107

Initially, the well-studied Gag epitopes restricted by HLA-B*27:05 and HLA-B*57:01, believed to play a central role in immune containment in subjects expressing these alleles, were studied (Goulder & Walker, 2012; Goulder et al., 1996; Goulder, Brander, et al., 2001; Goulder, Phillips, et al., 1997). The presence in the CRF01_AE clade consensus of the very residues that are selected in B or C clade infected subjects as escape mutants in two of the four HLA-27:05/B*57:01-restricted p24 Gag-specific epitopes Gag-ISW9 and Gag-KF11, prompted the question of whether well-defined HLA-B*27:05/57:01-restricted epitopes are accessible in AE clade infection. Only six out of 20 HLA-B*27/B*57-restricted

epitopes (HLA-B*57 Gag-TW10, Pol-IW9, Pol-KF9; HLA-B*27 Gag-IK9, Gag-KK10, Pol-KY9) previously shown to drive the selection of escape mutants (Carlson, Brumme, et al., 2012), share the same consensus sequence in B and AE clades (Table 5.2).

In the HLA-B*27:05/57:01-positive recipient, the earliest sample available for sequencing was from 20 months after diagnosis, by which time progression was already evident (Figure 5.2B). The T242N mutation within the B*57:01-restricted epitope Gag-TW10 had already reached fixation by this time, being present in 100% of the intra-host population detected by ultra-deep sequencing (Table 5.1). This occurred in the absence of selection of the compensatory mutations G248A, H219Q or M228I, which are associated with restoring viral replicative capacity in the presence of the T242N mutation (Crawford et al., 2011; Leslie et al., 2004). The other two HLA-B*57:01-restricted Gag epitopes, Gag-ISW9 and Gag-KF11, already carry polymorphisms A146P/I147L and A163G/S165N in the consensus sequence of CRF01_AE Clade HIV. These polymorphisms are well-characterised escape mutants within the B-clade versions of these epitopes (Table 5.2) (Leslie et al., 2004; Leslie et al., 2005).

### 5.3.3 Transmission of a minor variant from donor N094 to recipient N107

In the HLA-B*13:02 positive donor, selection of a Gag-P146S substitution was observed in a considerable proportion of the intra-host population, reaching fixation by 42 months (Table 5.1). However, proline was present at this position at the earliest time-point in the recipient, suggesting possible transmission of a minor variant bearing the CRF01_AE clade consensus residue. Subsequently, the Gag-

P146S variant was also selected in the recipient. HLA-B*13:02 in the donor and HLA-B*57:01 in the recipient are thus likely to have independently selected P146S in the autologous virus. The possibility that P146S was transmitted to the recipient, reverted shortly therefore, and was later re-selected, cannot, however, be excluded.

**Table 5.2 Alignment of HLA-B*27:05 and HLA-B*57:01-restricted epitopes in an HIV-1 transmission pair, showing sequences derived from donor and recipient, compared to consensus sequences for B clade and CRF01_AE clade.** Known escape mutations in B clade infection are indicated in bold.

| | | |
|---|---|---|
| **HLA-B*57-Restricted Epitopes** | | |
| **Gag-ISW9** | B Clade Consensus: | $^{146}$A ISPRTLNAW$^{155}$ |
| | CRF01_AE Clade Consesnsus: | P L-------- |
| | Donor Consensus (8 months): | S L-------- |
| | Recipient Consensus (20 months): | P L-------- |
| | Recipient Consensus (52 months): | S L-------- |
| **Gag-KF11** | B Clade Consensus: | $^{162}$KAFSPEVIPMF$^{172}$ |
| | CRF01_AE Clade Consesnsus: | -G-N------- |
| | Donor Consensus (8 months): | -G-N------- |
| | Recipient Consensus (20 months): | -G-N------- |
| | Recipient Consensus (52 months): | -G-N---V--- |
| **Gag-TW10** | B Clade Consensus: | $^{240}$TSTLQEQIGW$^{249}$ |
| | CRF01_AE Clade Consesnsus: | ---------- |
| | Donor Consensus (8 months): | ---------- |
| | Recipient Consensus (20 months): | --N------- |
| | Recipient Consensus (52 months): | --N------- |
| **Pol-VY10** | B Clade Consensus: | $^{272}$VPLDEDFRKY$^{280}$ |
| | CRF01_AE Clade Consesnsus: | -----S---- |
| | Donor Consensus (8 months): | -----N---- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | -----N---- |
| **Pol-ISW9** | B Clade Consensus: | $^{399}$IVLPEKDSW$^{407}$ |
| | CRF01_AE Clade Consesnsus: | -E------- |
| | Donor Consensus (8 months): | -K------- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | -K------- |
| **Pol-IW9** | B Clade Consensus: | $^{530}$IATESIVIW$^{538}$ |
| | CRF01_AE Clade Consesnsus: | --------- |
| | Donor Consensus (8 months): | --------- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | --------- |
| **Pol-SW10** | B Clade Consensus: | $^{838}$STTVKAACWW$^{847}$ |
| | CRF01_AE Clade Consesnsus: | -AA------- |
| | Donor Consensus (8 months): | -TA------- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | -NA------- |
| **Pol-KF9** | B Clade Consensus: | $^{888}$KTAVQMAVF$^{896}$ |
| | CRF01_AE Clade Consesnsus: | --------- |
| | Donor Consensus (8 months): | --------- |
| | Recipient Consensus (20 months): | --------- |
| | Recipient Consensus (52 months): | --------- |
| **Vif-IF9** | B Clade Consensus: | $^{31}$ISKKAKGWF$^{38}$ |
| | CRF01_AE Clade Consesnsus: | ------K-- |
| | Donor Consensus (8 months): | R-----E-- |
| | Recipient Consensus (20 months): | R-----E-- |
| | Recipient Consensus (52 months): | R-----E-- |
| **Vif-HL9** | B Clade Consensus: | $^{73}$HTGERDWHL$^{81}$ |
| | CRF01_AE Clade Consesnsus: | Q---K--Q- |
| | Donor Consensus (8 months): | Q---K---- |
| | Recipient Consensus (20 months): | Q---K---- |
| | Recipient Consensus (52 months): | Q---K---- |
| **Vpr-AW9/AL10** | B Clade Consensus: | $^{30}$AVRHFPRIW$^{38}$ L$^{39}$ |
| | CRF01_AE Clade Consesnsus: | -------P- - |
| | Donor Consensus (8 months): | -------L- - |
| | Recipient Consensus (20 months): | -------L- - |
| | Recipient Consensus (52 months): | -------I- - |
| **Rev KY10** | B Clade Consensus: | $^{14}$KTVRLIKFLY$^{23}$ |
| | CRF01_AE Clade Consesnsus: | RA--I--I-- |
| | Donor Consensus (8 months): | AA--T--I-- |
| | Recipient Consensus (20 months): | AA--T--I-- |
| | Recipient Consensus (52 months): | VA--T--I-- |
| **Nef-KF9** | B Clade Consensus: | $^{82}$KAAVDLSHF$^{90}$ |
| | CRF01_AE Clade Consesnsus: | -G-F---F- |
| | Donor Consensus (8 months): | -G-F---F- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | -G-F---F- |
| **Nef-HW9** | B Clade Consensus: | $^{116}$HTQGYFPDW$^{124}$ |
| | CRF01_AE Clade Consesnsus: | N---F---- |
| | Donor Consensus (8 months): | N-------- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | N-------- |
| **HLA-B*27-Restricted Epitopes** | | |
| **Gag-IK9** | B Clade Consensus: | $^{19}$IRLRPGGKK$^{27}$ |
| | CRF01_AE Clade Consesnsus: | --------- |
| | Donor Consensus (8 months): | -------R- |
| | Recipient Consensus (20 months): | -------R- |
| | Recipient Consensus (52 months): | -------R- |
| **Gag-KK10** | B Clade Consensus: | $^{263}$KRWIILGLNK$^{272}$ |
| | CRF01_AE Clade Consesnsus: | ---------- |
| | Donor Consensus (8 months): | ---------- |
| | Recipient Consensus (20 months): | --------N- |
| | Recipient Consensus (52 months): | --------N- |
| **Pol-KY9** | B Clade Consensus: | $^{901}$KRKGGIGGY$^{909}$ |
| | CRF01_AE Clade Consesnsus: | --------- |
| | Donor Consensus (8 months): | --------- |
| | Recipient Consensus (20 months): | --------- |
| | Recipient Consensus (52 months): | --------- |
| **Vpr-VL9** | B Clade Consensus: | $^{31}$VRHFPRIWL$^{39}$ |
| | CRF01_AE Clade Consesnsus: | ------P-- |
| | Donor Consensus (8 months): | ------L-- |
| | Recipient Consensus (20 months): | ------L-- |
| | Recipient Consensus (52 months): | |
| **Env-GY10** | B Clade Consensus: | $^{786}$GRRGWEVLKY$^{795}$ |
| | CRF01_AE Clade Consesnsus: | L-----G--- |
| | Donor Consensus (8 months): | L-----G--- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | L-----G--- |
| **Nef-RI10** | B Clade Consensus: | $^{105}$KRQDILDLWV$^{114}$ |
| | CRF01_AE Clade Consesnsus: | ---E------ |
| | Donor Consensus (8 months): | ---E-I---- |
| | Recipient Consensus (20 months): | not done |
| | Recipient Consensus (52 months): | ---E-I---- |

To investigate which HIV-specific CD8+ T-cell responses were detectable at the earliest timepoint available in the recipient (20 months post-diagnosis), IFN-γ ELISpot assays were conducted. Responses to 6 of these 18mers were identified (Figure 5.3A). The dominant Gag responses were to the HLA-B*27-restricted epitope KK10 and to the B*57:01-restricted epitope TW10. There was also a subdominant Vpr response and a high magnitude response to the HLA-B*27:05-restricted epitope in Integrase, Pol-KY9 ($^{901}$KRKGGIGGY$^{909}$), a response that is typically co-dominant with HLA-B*27:05 Gag-KK10 (Payne et al., 2010). Whereas the HLA-B*27:05-KK10 and HLA-B*57:01-TW10 epitopes had escaped by the first timepoint (20 months in the recipient), this was not the case for the other epitopes. HLA-B*27:05 Pol-KY9 did not drive selection of escape even at 52 months post-diagnosis. These data support the hypothesis that the dominant Gag epitopes, including HLA-B*27:05-KK10 and the HLA-B*57:01-TW10, are critical for maintaining immune control.

Although an IFN-γ ELISpot response to the Gag-TW10 epitope was observed, no response to the autologous T242N variant was observed, and no CD8+ T cell responses were detectable in this subject to either the B clade or AE clade version of these Gag-ISW9 and Gag-KF11 epitopes (Figure 5.3). At 20 months after diagnosis, the HLA-B*27-restricted epitope Gag-KK10, believed to play an important role in HLA-B*27-mediated immune control of HIV (Goulder & Walker, 2012; Goulder et al., 1996; Goulder, Brander, et al., 2001; Goulder, Phillips, et al., 1997; Nixon et al., 1988), also contained the escape mutation N271H in 100% of the recipient sequences (Table 5.1), despite persistence of a substantial *ex vivo* T cell response to the wildtype and N271H variant epitopes (Figure 5.3B and C). Thus, disease progression was seen in the HLA-B*27:05/57:01-positive recipient

in association with mutations in all four HLA-B27:05/*57:01-restricted p24 Gag epitopes.

### 5.3.4 The majority of sequence changes selected in the recipient are escape polymorphisms in known epitopes

To investigate whether other sequence changes outside of the well-studied region of p24 Gag might also have contributed to progression in the HLA-B*27:05/57:01-positive recipient, ultra-deep sequence data of the full-length HIV genome was examined. Heat maps were generated to visualise the proportion of amino acid variants at each position compared to a given baseline. All sites of complete amino acid mismatch between the donor and recipient that reflect inter-host evolution were identified using the donor consensus sequence at 8 months as the baseline for comparison to the recipient at 52 months (Figure 5.4A). Sites of amino acid diversity in the recipient at 52 months, demonstrating intra-host evolution, were also identified using the recipient consensus sequence at the same timepoint as the baseline for comparison (Figure 5.4B) were also identified. The heat map analyses that were generated highlight the location of the residues that changed most rapidly in the recipient and which arose within CD8+ T cell epitopes (Figure 5.4, Figure 5.5).

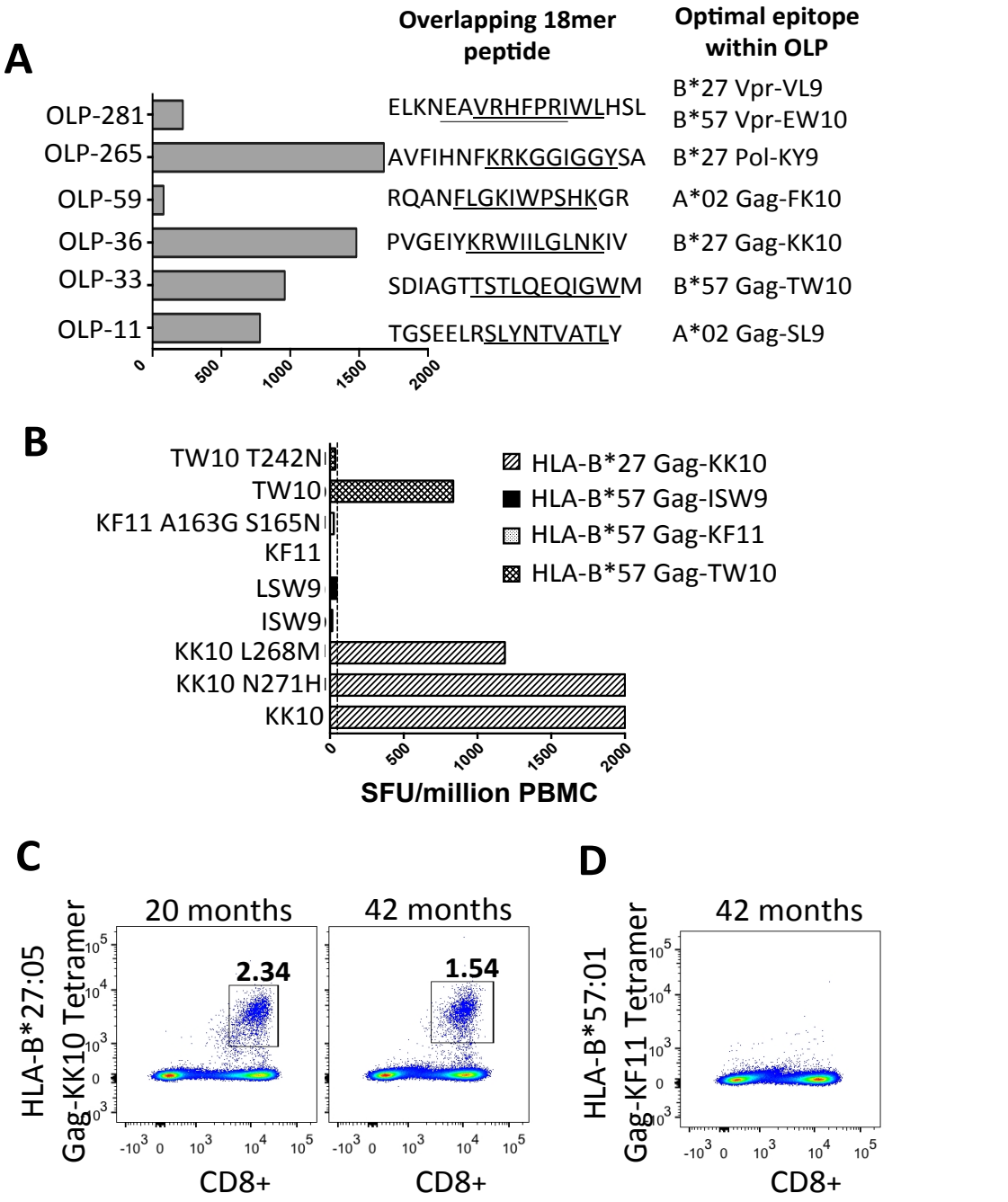



**Figure 5.3 Quantification of CD8+ T cell responses to HLA-B*57 and HLA-B*27-restricted Gag epitopes in the recipient from an HIV transmission pair.** A) IFN-$\gamma$ ELISpot responses at 20 months post-diagnosis in response to 410 18mer peptides spanning the B clade proteome. Responses to six 18mer peptides were detected. B) IFN-$\gamma$ ELISpot responses at 20 months post-diagnosis, showing maintenance of large responses to HLA-B*27 Gag-KK10 (KRWIIGLNK) and the N271H variant of this epitope and moderate responses to the HLA-B*57 Gag-TW10 epitope (TSTLQEQIGW), regardless of sequence changes within the autologous virus, with no responses to KF11 (KAFSPEVIPMF) or ISW9 (ISPRTLNAW) and their variants above the cut-off. Spot forming units (SFU) per million PBMC above a cut-off of 50 SFU/million are reported. Responses >2000 SFU/million PBMC (OLP 36, KK10 and KK10-N271H) could not be quantified precisely. C) HLA-B*27:05-KK10 tetramer stains at 20 and 42 months post-diagnosis show maintenance of a large Gag-KK10-specific CD8+ T cell population. D) HLA-B*57:01-Gag-KF11 tetramer stain at 42 months post-diagnosis shows no detectable Gag-KF11-specific CD8+ T cell population.

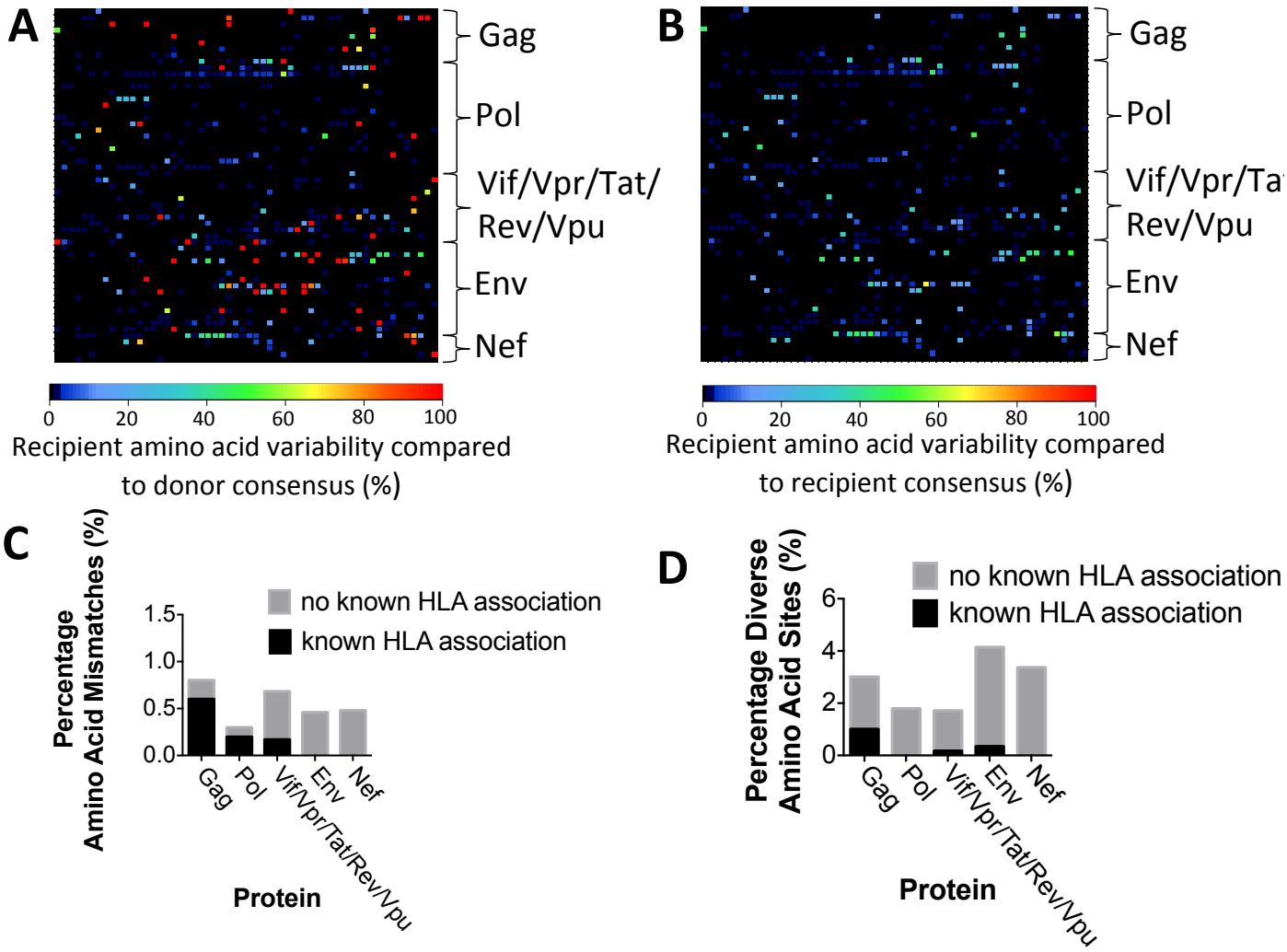



**Figure 5.4 Comparison of inter- and intra-host diversity of HIV quasispecies in a transmission pair.** A) Heat map representation of inter-host amino acid diversity across the full-length HIV genome, comparing donor and recipient. For the baseline, the donor sequence at 8 months post-diagnosis (which in this case represents the closest approximation of the founder virus) was used; variation in the recipient at 52 months post-diagnosis is compared to this baseline. Each square represents a single codon, coloured to reflect the percentage of sequences in the recipient that differ from the consensus residue in the donor. B) Heat map representation of intra-host amino acid diversity in the recipient at 52 months post-diagnosis. For the baseline, the recipient consensus sequence at 52 months post-diagnosis was used, to which the intra-host recipient population at the same timepoint is compared. Each square represents a single codon coloured to reflect the percentage of minor variants in the recipient that differ from the consensus ('baseline') residue. C) Percentage of true amino acid mismatches (excluding positions where the recipient's sequence was represented as a minor variant in the donor) between donor and recipient sequences by gene. The proportion of mismatches at sites where there is a known or predicted association with the recipient's HLA alleles is indicated. D) Percentage of diverse amino acid sites (variability >10%) in the recipient intra-host population by gene. The proportion of diverse sites where there is a known or predicted association with the recipient's HLA alleles is indicated.

Using the donor consensus sequence at 8 months post-diagnosis as the closest approximation of the transmitted founder virus, 16 sites were identified across the full-length genome where the residue in the donor (including minor variant residues) had been entirely replaced in the recipient by 52 months post-diagnosis. Four of these residues were within Env, so were most likely to be susceptible to changes driven by neutralising antibody responses. Eight of the remaining 12 residues were in or flanking known epitopes, and seven of these were restricted by HLA-B*27:05 or HLA-B*57:01 (Figure 5.5). None of these sites were within epitopes restricted by HLA alleles expressed by the donor. These sequence changes were thus attributable to active selection in the recipient, rather than reversion of transmitted mutants selected in the donor.

These data from a single transmission pair cannot definitively limit the most effective CD8+ T cell responses to this group of seven epitopes. However, the data are consistent with the hypothesis that the most effective responses are among this group. Of note, these do not include many of the well-studied HLA-B*27:05/57:01-restricted epitopes that, like the Gag epitopes ISW9 and KF11, are mutated in AE clade compared to B clade virus (Table 5.2), and in this transmission pair did not differ between donor and recipient at the timepoints compared.

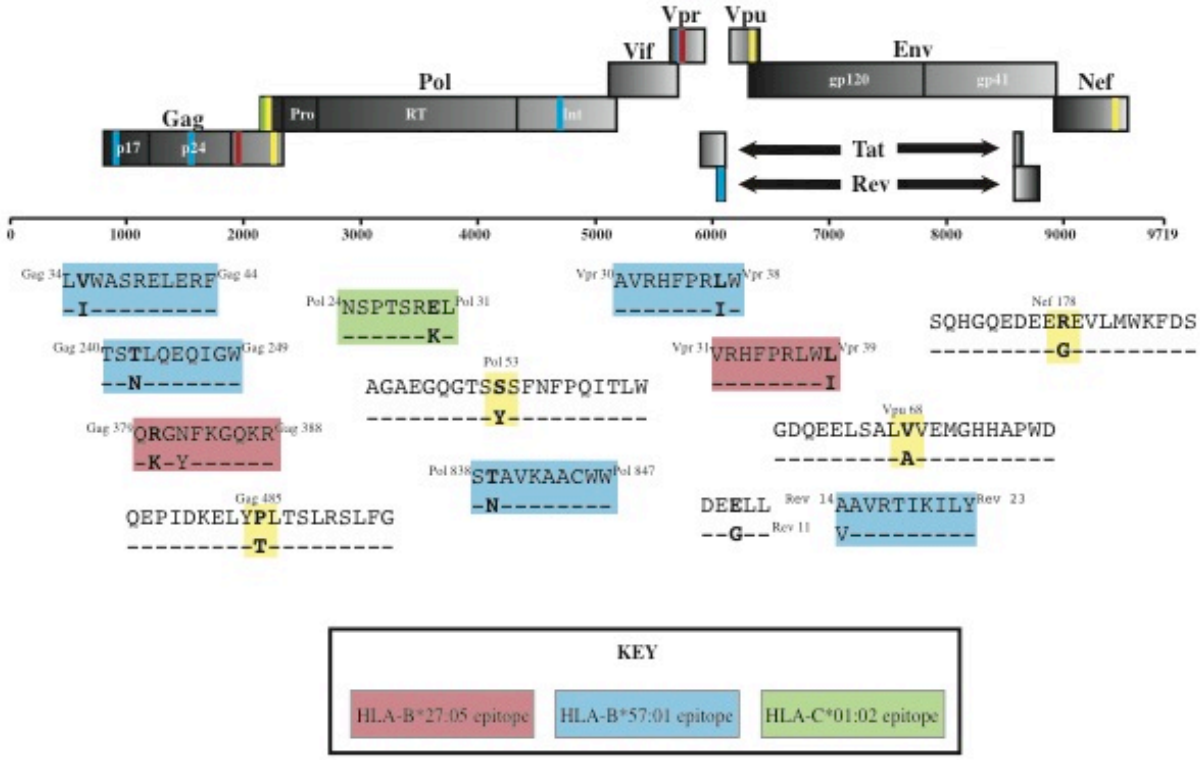



**Figure 5.5 Schematic representation of sites of complete amino acid mismatch between the donor and recipient full-length HIV sequences.** The recipient consensus sequence at 52 months is aligned to the donor consensus sequence at 8 months (the latter being the closest representation of the transmitted HIV sequence). The sites shown in this figure are complete mismatches between donor and recipient identified in the heat map analysis (red squares, **Figure 5.4**A) The mismatched residue is indicated in bold. HLA-B*27:05, B*57:01 and C*01:02-restricted epitopes are shown in red, blue and green respectively. Mismatched residues that do not fall within a relevant HLA-Class I epitope are shown in yellow.

To identify additional sites across the full-length genome in the recipient that were subject to turnover without having reached fixation yet, sites at which amino acid diversity of at least 10% was present in the intra-host population at 52 months post diagnosis were identified (Figure 5.4B, Figure 5.6). Diversity was found at only 2.6% of all amino acid residues, of which the majority (1.1%) were in Env, a highly variable region of the genome where mutations are driven largely by neutralising antibody responses. Of the remaining sites of diversity, 16% were within or flanking recognised HLA-B*27 or HLA-B*57 epitopes. In both the donor/recipient comparison (Figure 5.4A) and intra-host diversity plot (Figure 5.4B) the evolving sites in Gag were frequently within known or predicted CTL

epitopes restricted by the recipient's HLA alleles. In contrast those outside of Gag, especially in Env or Nef, were rarely within known or predicted epitopes (Figure 5.4C, D).

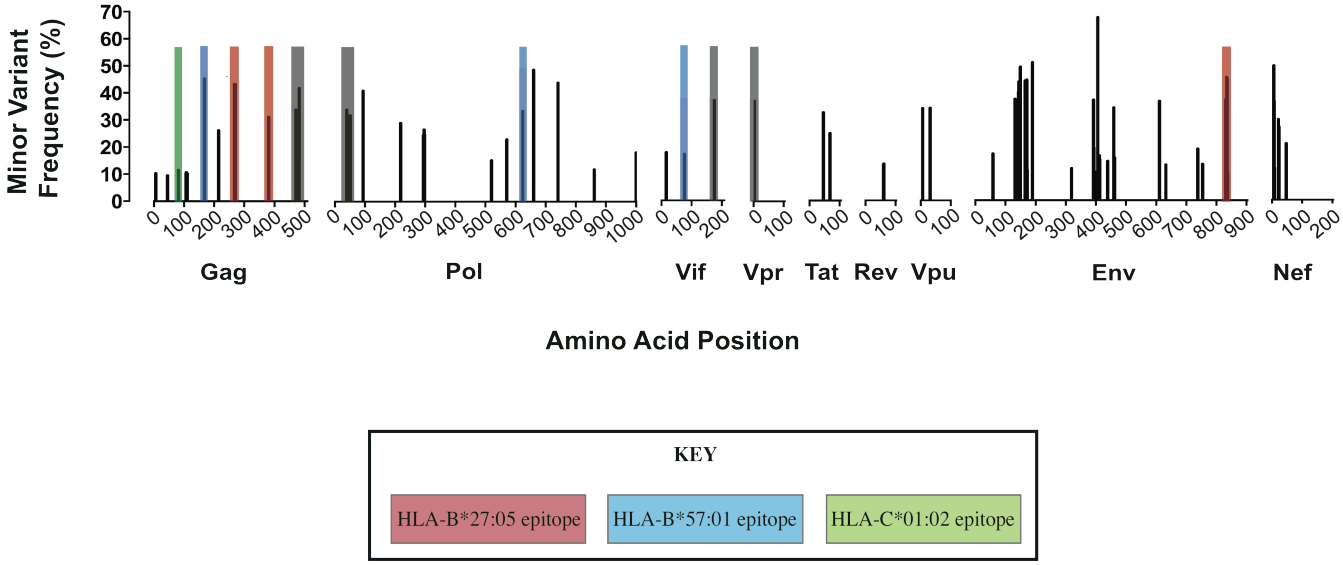



**Figure 5.6 Sites of amino acid diversity in the recipient from an HIV transmission pair at 52 months post-diagnosis.** Diverse sites were defined as amino acid positions with diversity of ≥10% in the intra-host population. Sites that fall within or flanking known or predicted HLA-B*27:05, B*57:01 and C*01:02-restricted epitopes are shown in red, blue and green respectively. Duplicated sites, present due to overlap in the Gag/Pol and Vif/Vpr reading frames, are shown in grey.

### 5.3.5 Sequence changes in the donor reflect escape polymorphisms selected in known epitopes

Finally, the sequences in the donor, who progressed despite possessing the protective HLA-A*32:01, HLA-B*13:02 and HLA-B*14:01 alleles (Fellay et al., 2009; Honeyborne et al., 2007) were examined. Compared to the full-length CRF01_AE clade consensus sequence, there were six epitopes at which HLA-associated mutations were present in the donor, two of which were in p24 Gag. These are within epitopes restricted by HLA-B*13:02 (Gag 135-143, 'VV9") and B*14:01 (Gag 298-306, 'DA9') respectively (Table 5.3). Thus, as in the recipient, progression to HIV disease in the donor was associated with mutations in critical p24 Gag epitopes.

**Table 5.3 Sites of HLA-associated footprints in the donor from an HIV transmission pair at 8 months post-diagnosis aligned to B and CRF01_AE clade consensus sequence.** Epitopes restricted by the favourable alleles expressed by the donor, A*3201, B*1302 and B*1401, are shown.

| | | |
|---|---|---|
| **HLA-A*32:01-Restricted Epitopes** | | |
| **Pol-PW10** | B Clade Consensus: | $^{547}$PIQKETWETW$^{556}$ |
| | Described B Clade Footprint: | ---R----I- |
| | CRF01_AE Clade Consesnsus: | ---R------ |
| | Donor Consensus (8 months): | ---R------ |
| **Pol-KF9** | B Clade Consensus: | $^{930}$KQITKIQNF$^{938}$ |
| | Described B Clade Footprint: | ---I-M--- |
| | CRF01_AE Clade Consesnsus: | --------- |
| | Donor Consensus (8 months): | --------- |
| **Vif-KY11** | B Clade Consensus: | $^{34}$KAKGWFYRHHY$^{44}$ |
| | Described B Clade Footprint: | -------K--- |
| | CRF01_AE Clade Consesnsus: | ---K------- |
| | Donor Consensus (8 months): | ---E------- |
| **Tat-KI11** | B Clade Consensus: | $^{29}$KCCFHCQVCFI$^{39}$ |
| | Described B Clade Footprint: | Q---------- |
| | CRF01_AE Clade Consesnsus: | ---Y------L |
| | Donor Consensus (8 months): | ---W---L--L |
| **Env-RW9** | B Clade Consensus: | $^{419}$RIKQIINMW$^{427}$ |
| | Described B Clade Footprint: | -------**L**- |
| | CRF01_AE Clade Consesnsus: | K-------- |
| | Donor Consensus (8 months): | K-R--VR-- |
| **Env-SY10** | B Clade Consensus: | $^{703}$SIVNRNRQGY$^{712}$ |
| | Described B Clade Footprint: | -LA------- |
| | CRF01_AE Clade Consesnsus: | -----V---- |
| | Donor Consensus (8 months): | --A--V---- |
| **HLA-B*14:01-Restricted Epitopes** | | |
| **Gag-DA9** | B Clade Consensus: | $^{298}$DRFYKTLRA$^{306}$ |
| | Described B Clade Footprint: | ----R---- |
| | CRF01_AE Clade Consesnsus: | --------- |
| | Donor Consensus (8 months): | ----R---- |
| **Vif-DK8** | B Clade Consensus: | $^{172}$DRWNKPQK$^{179}$ |
| | Described B Clade Footprint: | ----E-K- |
| | CRF01_AE Clade Consesnsus: | -------- |
| | Donor Consensus (8 months): | -K------ |
| **Env-EL9** | B Clade Consensus: | $^{584}$ERYLKDQQL$^{592}$ |
| | Described B Clade Footprint: | -XF-Q---F |
| | CRF01_AE Clade Consesnsus: | -------KF |
| | Donor Consensus (8 months): | -------KF |

| | | |
|---|---|---|
| **HLA-B*13:02-Restricted Epitopes** | | |
| **Gag-VV9** | B Clade Consensus: | $^{135}$VQNLQGQMV$^{14}$ HQAI$^{147}$ |
| | Described B Clade Footprint: | --------- --SL |
| | CRF01_AE Clade Consesnsus: | ---A----- ---L |
| | Donor Consensus (8 months): | ---A----- --SL |
| **Gag-RI9** | B Clade Consensus: | $^{429}$RQANFLGKI$^{437}$ |
| | Described B Clade Footprint: | -------RL |
| | CRF01_AE Clade Consesnsus: | --------- |
| | Donor Consensus (8 months): | --------- |
| **Pol-RI10** | B Clade Consensus: | $^{113}$RQYDQILIEI$^{122}$ |
| | Described B Clade Footprint: | ------X--- |
| | CRF01_AE Clade Consesnsus: | ---------- |
| | Donor Consensus (8 months): | ------C--- |
| **Pol-KI10-** | B Clade Consensus: | $^{521}$KQLTEAVQKI$^{530}$ |
| | Described B Clade Footprint: | ---V------ |
| | CRF01_AE Clade Consesnsus: | ---------- |
| | Donor Consensus (8 months): | ---------- |
| **Nef-RV9** | B Clade Consensus: | $^{106}$RQDILDLWV$^{114}$ |
| | Described B Clade Footprint: | -R------- |
| | CRF01_AE Clade Consesnsus: | --E------ |
| | Donor Consensus (8 months): | -RE-I---- |
| **HLA-Associated Footprints Outside Epitopes** | | |
| **HLA-A*32:01 Vpu** | B Clade Consensus: | $^{80}$D |
| | Described B Clade Footprint: | V |
| | CRF01_AE Clade Consesnsus: | - |
| | Donor Consensus (8 months): | - |
| **HLA-A*32:01 Env** | B Clade Consensus: | $^{121}$KLTPLCVTLNCTDLM$^{135}$ |
| | Described B Clade Footprint: | ----------R |
| | CRF01_AE Clade Consesnsus: | ----------NAN |
| | Donor Consensus (8 months): | ---------H-KKVD |
| **HLA-B*14:01 Vpr** | B Clade Consensus: | $^{3}$Q |
| | Described B Clade Footprint: | R |
| | CRF01_AE Clade Consesnsus: | - |
| | Donor Consensus (8 months): | - |
| **B*14 Nef** | B Clade Consensus: | $^{53}$A |
| | Described B Clade Footprint: | P |
| | CRF01_AE Clade Consesnsus: | - |
| | Donor Consensus (8 months): | - |
| **B*14 Nef** | B Clade Consensus: | $^{101}$I |
| | Described B Clade Footprint: | V |
| | CRF01_AE Clade Consesnsus: | - |
| | Donor Consensus (8 months): | - |

## 5.4 Discussion

This study capitalises on longitudinal data from a well-characterised transmission pair, for whom the depth (ultra-deep approach) and breadth (full-length HIV genomes) of sequence resolution was maximised. This allowed the evolution of the escape mutations to be precisely quantified, including minor variants, in the context of what would usually be regarded as a highly favourable combination of HLA alleles, HLA-B*27:05 and HLA-B*57:01. Since both these alleles occur at a very low frequency within the Thai population (approximately 0.2% and 1.4% respectively (Mori et al., 2014)) finding this haplotype in the context of CRF01_AE clade infection was an 'accident of nature' that provided a unique opportunity to study the mechanisms of immune control.

There are conflicting data regarding the extent to which HLA-B*57 may be protective in Thai cohorts. Although a recent study in a particular Thai cohort where the median CD4+ T cell count was only 86 cells/mm$^3$ reported that HLA-B*57:01 was protective (Mori et al., 2014), a parallel study of 116 transmission pairs found no benefit of HLA-B*57 (Gesprasert et al., 2010). The latter result fits with the conclusions drawn here about the HLA-B*57-positive recipient N107. This is consistent with the abrogation of HLA-B*57-restricted Gag epitopes due to pre-existing polymorphisms in CRF01_AE clade virus that represent HLA-B*57 escape mutations. This result highlights the extent to which clade of infection may be an important determinant of immunological and clinical outcomes. Even in an individual expressing a favourable combination of HLA alleles that are usually

strongly linked to immune control, rapid progression may result in the context of infection with a viral sequence bearing pre-existing escape mutations.

Although HIV is recognised as a highly polymorphic virus, this study has demonstrated that viral evolution is frequently constrained to specific amino acid residues, with the success of the CD8+ T cell response dependent on these sites. Significant variability (>10%) was evident within the intra-host population at only 2.6% of amino acid residues in the recipient. Ultra-deep sequencing demonstrated a high degree of conservation within key HLA-B*27:05 and HLA-B*57:01-restricted epitopes (Table 1) except at pre-defined sites of escape mutation, which most often corresponded to anchor residues. Selection pressure was thus specifically directed at these particular sites, which is consistent with previous reports showing that selective escape from CD8+ T cell responses follows constrained evolutionary pathways (Allen et al., 2005).

Consistent with previous reports (Gesprasert et al., 2010; Martinez-picado et al., 2006), the recipient showed robust selection and fixation of the Gag HLA-B*57-selected T242N mutation in the Gag-TW10 epitope, which was maintained in the host viral population. Within the Gag-KK10 epitope, strong selection pressure drove N271H selection almost to fixation. Subsequent reversion to the wildtype residue in a substantial proportion of the variants does, however, indicate more complexity in the adaption of the autologous virus at this site. This change demonstrates substantial variability in the circulating intra-host population that is not detected in consensus sequences, and may reflect the delicate balance between selection of an escape mutation that benefits the virus, versus the cost to viral fitness that is imposed by the escape polymorphism.

Explaining variation at certain sites is made more complicated by multiple influences on viral polymorphism. For example, Gag P146S is a common variant in CRF01_AE Clade infection (occurring in approximately 9.5% of sequences). This site is also subject to selection pressure from both HLA-B*13:02 and HLA-B*57:01-mediated T cell responses (Carlson, Brumme, et al., 2012; Crawford et al., 2009; Matthews et al., 2009)[14]. Variation at this position in this study could therefore be attributed to selection pressure from either the donor or recipient CD8+ T cell response, or to a founder virus bearing a serine variant rather than the more common proline. An alternative explanation for sequence variation occurring over time in a transmission pair is that more than one transmission event has taken place; the introduction of a new founder virus could then alter the dominant quasispecies. In this instance, re-infection appears unlikely on the basis of phylogeny demonstrating clear clustering of donor and recipient sequences respectively, but cannot be excluded completely due to the limited number of samples analysed over the time period of follow up.

It is striking that even by applying an unbiased approach to seeking sequence variability across the whole genome, the majority of polymorphisms identified in the recipient were within or flanking known epitopes, with HLA-B*27 and HLA-B*57-restricted epitopes being dominant, and Gag accounting for the greatest number of these. The observations made here, using the approach of this genome-wide search for polymorphisms, therefore corroborate previous data in studies

[14] Mutations at Gag-146 and 147 have been associated with escape from the HLA-B*13:02-restricted Gag-VV9 (VQNLQGQMV) response (Honeyborne et al., 2007) and selection of both a serine (S) and leucine (L) at position 146 and 147 respectively are associated with HLA-B*13:02 (J. M. Carlson, Brumme, et al., 2012).

that have used known CD8+ T cell epitopes or IFN-γ ELISpot assays as their starting point to identify sites of immune selection (Leslie et al., 2005; Martinez-picado et al., 2006; Payne et al., 2010).

The unique nature of the circumstances described in this report mean that the findings are difficult to replicate, and can be presented as a case study only. An additional limitation for this transmission pair was lack of information about the precise timing of infection, and absence of samples from timepoints closer to the time of transmission. Furthermore, a lack of data on the epitopes restricted in the context of this rare combination of HLA alleles and clade of infection has limited analysis of epitopes to those that have been described in the context of B clade infection. Although CRF01_AE clade-specific peptides were tested in IFN-γ ELISpot assays where possible, screening of CTL responses using consensus B clade OLPs may have limited identification of clade-specific and variant-specific responses. It is noteworthy, for example, that the B*27:05-KK10 variant selected in the clade AE-infected recipient was N271H that has been rarely observed in B clade infection. In this case, a strong N271H-specific CTL response was observed, which may appear counter-intuitive if N271H is selected as an escape mutant. However, it has been well described with respect to escape mutants that affect T cell receptor recognition, such as the more commonly observed L268M within KK10 (Almeida et al., 2007; Iglesias et al., 2011; Ladell et al., 2013), that a high frequency response can be observed to a TCR-variant when it is recognised by a subset of CTL clones. Despite these caveats, this transmission pair provided a unique insight, gained by full-length ultra-deep sequencing data, supporting the association between the selection of polymorphisms to allow escape from HLA-B*27 and HLA-B*57-restricted epitopes, and loss of immunological control.

The unique opportunity to study CRF01_AE Clade HIV infection longitudinally in the context of a transmission pair with protective HLA alleles, using ultra-deep sequencing and an unbiased approach to full-length sequence analysis, has shown the extent to which the polymorphisms associated with disease progression are constrained to very specific amino acid sites, frequently within Gag-restricted epitopes. The extent to which selection of escape mutations is robust and predictable is surprising given the overall plasticity of the HIV genome. This observation is encouraging for the development of T cell vaccines for which meeting the challenges presented by viral escape is a major consideration.

# CHAPTER 6: Discussion

## 6.1 Contributions of this work to the field

HLA-B*27:05 and –B*57:01 have been well studied and long recognised as important genetic determinants of disease control in HIV infection (Fellay et al., 2007, 2009; Goulder & Walker, 2012; Kiepiela et al., 2004; Pereyra et al., 2008, 2010). However, the factors that will determine whether an individual expressing these alleles will be able to mount a CTL response that will effectively control viraemia remain incompletely understood. Given that studies of elite and viraemic control of HIV remain an important approach to elucidating correlates of protection that may inform the design of intervention strategies, studying the mechanism(s) of HLA-B*27:05 and HLA-B*57:01-mediated immune control may prove to be an important approach. To elucidate novel aspects of HLA-B*27:05 and B*57:01-mediated immune control, this work has focused on the expression of these alleles in the context of clades of infection in which they are infrequently found, and in the rare setting of dual HLA-B*27:05/B*57:01 expression. The aim of this work has been to examine the interactions between expression of these HLA class I alleles, intra-host viral evolution and HIV disease progression applying new technologies to advance our understanding of HIV specific CTL immunity.

One explanation for the association of HLA with variable disease outcome in HIV comes from *in silico* experiments that relate more extensive binding of self-peptide and negative selection in the thymus with the deletion of a larger proportion of the T cell repertoire that might recognise HIV, leading to poorer disease outcome (Kosmrlj et al., 2010). In this case, disease outcome would be unlikely to be

influenced by clade-specific differences in the viral sequence. Demonstrating that clade specific differences do impact on HLA-mediated immune control therefore implicates the specific sequence of the epitope and flanking regions in immune control. From the work presented in chapters 3 and 5, and in related studies (Carlson, Listgarten, et al., 2012; Kløverpris et al., 2013; Matthews, Koyanagi, et al., 2012) it is clear that the clade of infection does have an important role to play in determining disease outcome in the context of expression of protective HLA alleles. This points to the sequence of the transmitted founder in the mechanism of HLA-mediated control of HIV. With reference to the CRF01_AE clade infected transmission pair discussed in chapter 5, this also provides a possible explanation for disease progression in the context of dual HLA-B*27:05/B*57:01-expression. Chapter 4 further elucidates the factors contributing to disease progression in this context, describing associations with VRC, intra-host viral evolution and CTL escape.

In detailed transmission pair studies, this work has demonstrated how specific CTL pressure mediates intra-host viral evolution, immune evasion and HIV disease progression. These studies, where a known transmission partner provides an approximation of the transmitted founder, allowing extensive analysis of HLA-mediated selection pressure on a particular viral sequence, are a key contribution of this work.

Developing robust techniques to manage and interpret the large volumes of data that are generated by ultra-deep sequencing is a vital component of viral genomics. The methods described here for analysing full-length ultra-deep HIV sequences,

with specific application in studies of CTL immunology and the impact of HLA on intra-host viral evolution are a key contribution of this work. This has allowed high-resolution sequence analysis that is not biased to specific epitopes or regions where diversity is already known or expected to occur. To our knowledge, the heat map analyses developed in collaboration with Dr Rebecca Batorsky and Dr Todd Allen have only been presented in one other publication other than our own (Henn et al., 2012). Analysis of the minor variant populations produced with these methods have allowed detection of transient fluctuation in the intra-host populations before fixation of mutations occurs, demonstrating the complexity of the intra-host evolutionary process. This has also allowed detailed tracking of selection and reversion of mutations in transmission partners. Minor variant analysis was essential in distinguishing the case of rapid intra-host viral evolution (subject R097) from a case of HIV superinfection. Appendix III presents further applications of these methods in detection of minor variant transmission events and longitudinal analysis of intra-host evolution. These studies have thus clearly demonstrated the importance of considering minor variant populations in studies of CTL-mediated selection pressure and intra-host viral evolution. The strides made here in analysing this type of data may help to inform a similar approach to the study of other viruses such as Hepatitis B virus (HBV).

The work presented here is relevant to other literature that is currently emerging. in the field. In similar recent work, Yue *et al* compared two HLA-matched transmission pairs, demonstrating the importance of the VRC of the transmitted founder in determining disease outcome (Yue et al., 2015). This study is consistent with the work presented here, highlighting the importance of the TF virus in

shaping the CTL response and determining HIV disease outcome. Recent work has also demonstrated that intra-host viral diversity during acute infection is associated with disease outcome (Gounder et al., 2015) providing further evidence of the importance of considering minor variant populations in sequence analysis. The findings of this study are also consistent with recent work demonstrating that transient selection of minor variant polymorphisms occurs before mutations reach fixation, guided by a fine balance of viral fitness cost and the most effective route to CTL escape (Henn et al., 2012; Sunshine et al., 2015).

## 6.2 Study limitations

As described in each results chapter, these studies are limited to small sample numbers or individual case studies due to the unusual set of circumstances that they describe. Expression of HLA-B*27:05 and infection with C clade HIV, expression of both HLA-B*27:05 and B*57:01, and expression of HLA-B*27:05/B*57:01 and infection with CRF01_AE clade HIV are all rare combinations. Furthermore, in the subjects expressing these alleles who maintained low VLs, obtaining both full-length and deep sequence data were a limitation. Studying HIV infection in these unusual settings has, however, allowed for detailed analysis yielding unique insights into HLA-mediated immune control.

In the transmission pair studies, the timing of infection of the recipient is an important consideration. In particular, the studies of R096/R097 (B clade infected pair) and N094/N107 (CRF01_AE clade infected pair) have been limited by

recruitment of the subjects during chronic infection, precluding analysis of early sequences that more closely match the TF virus. Inferring the rate of viral evolution over a specified time frame is made more complicated in subjects where the date of transmission cannot be determined with certainty.

The extent to which the *in vitro* VRC assays used in these studies, in which single mutations or autologous genes are introduced into an artificial viral backbone, are representative of the VRC of the full-length autologous virus is an important consideration. Since these assays have included Gag variants only, caution should be exercised in drawing conclusions regarding the overall VRC of full-length autologous viral variants.

## 6.3 Future directions

*Chapter 3*

The question of whether HLA-B*27:05 is protective in C clade HIV infection remains an unanswered question of the work presented in chapter 3. In future work, the cohorts presented here could be extended to populations such as Ethiopia, where C clade infection is prevalent, and HLA-B*27:05 is expressed by 4-5% of the population (Tsegaye et al., 2007; Tully & Wood, 2010). In Senegal, HLA-B*27:05 is expressed by 3-6% of the population (Belachew, Sandu, Schaller, & Guta, 2009) and an HIV infected MSM population with 40% C clade infection has been described (Jung et al., 2012).

*Chapter 4*

The analysis of evolutionary rate in HIV disease controllers and progressors presented in this chapter could be extended in future work to include estimates of viral evolutionary rate using BEAST to build time-measured phylogenies (as presented for R036/R056 in chapter 3). Furthermore, the rate of synonymous substitution, a marker of the rate of viral turnover, has been shown to predict the rate of HIV disease progression in both HLA mismatched (Lemey et al., 2007) and HLA-matched (Norström et al., 2014) subjects. An analysis of the synonymous substitution rate in this cohort is a potential area for future research.

*Chapter 5*

The methods used for analysis of full-length ultra-deep sequencing data for transmission pair N094/N107 can be applied further to the analysis of transmission pair R036/R056 (Chapter 3). Details of the approach for ongoing work are described in Appendix III. In addition, ongoing work in the Goulder laboratory, lead by Ellen Leitman, will involve the application of deep-sequencing methods to mother/child transmission pairs and paediatric slow progressor cohorts.

Manuscripts are in preparation for publication of the work presented in Chapters 3 and 4, with submission planned for November/December 2015.

## 6.4 Importance of the HIV specific CTL response in intervention strategies

Understanding what constitutes an effective CTL response has important applications for prophylactic vaccine, therapeutic vaccine and cure intervention strategies. Despite the considerable advancements in developing effective ART, these are critical endeavours for controlling the epidemic. Although there has been considerable effort toward development of a CTL vaccine for HIV (Buchbinder et al., 2008; Gray et al., 2011; Hammer et al., 2013; Pitisuttithum et al., 2006), a candidate that affords consistent protection has not been produced. One opinion is that this has been the case not because the CTL response in general is not protective, but because the specific response generated by candidate vaccines trialed to date have been ineffective (Migueles & Connors, 2015). There is, therefore, room for elucidating correlates of protection to inform the design of CTL vaccine candidates that might elicit effective CTL responses. Prophylactic vaccination is most likely to be achieved using a combined CTL and nAb approach with the CTL response being particularly important as a second line defence for eliminating cells infected by nAb-resistant virus or virus that has managed to cross the mucosal barrier (Migueles & Connors, 2015). Therapeutic CTL vaccines that do not offer sterilising immunity, but do reduce viraemia are an alternative aim.

Immunogens for the next generation of T cell vaccines are being designed with the aim of mounting more effective CTL than those produced naturally or using natural HIV sequences. The design of mosaic and conserved antigens are two approaches currently in development that aim to tackle the challenge of the global diversity among HIV strains. To increase the breadth of the elicited immune

response, mosaic antigens include sequences from a diversity of HIV strains. This approach has been shown to markedly increase the breadth of the CTL response in NHP models (Barouch et al., 2010). Conserved antigens aim to focus the immune response on the most conserved regions across HIV strains. One such approach has recently shown promising results in Phase I clinical trials (Borthwick et al., 2014; Letourneau et al., 2007). Very recent work has also begun to investigate the benefits of combining mosaic and conserved antigen approaches (Abdul-Jawad et al., 2015)(Ondondo et al., 2016). The work presented in chapter 3 and 5 of this thesis is relevant to both these approaches, which require an intricate knowledge of the effect of clade-specific differences in viral sequence on the CTL response. Other recent work has relied on functional T cell data to design an immunogen expressing the most effective T cell epitopes (Hancock et al., 2015; Mothe et al., 2015). The work presented throughout this thesis, which aims to define the most effective CTL responses in individuals expressing favourable HLA alleles, may inform similar future approaches.

In the field of HIV cure research, one approach to clear the viral reservoir has been to attempt to reactivate latent HIV using histone deacetylase inhibitors (Archin et al., 2012; Nancie M Archin et al., 2009). It is likely that this approach will require a second arm aimed at enhancing immune control following reactivation, since reactivation alone does not kill infected cells (Shan et al., 2012). The autologous CTL response from chronically infected patients may not be sufficient to clear infected cells, but antigen specific prestimulation or *ex vivo* expansion of CTL before reactivation may dramatically enhance killing of these cells (Shan et al., 2012; Sung et al., 2015). Therapeutic vaccination may prove to play an important

role in this area. Recent work has shown that the latent provirus may harbour a large percentage of CTL escape variants, and redirecting the CTL response to unmutated or subdominant epitopes may be an important approach (Deng et al., 2015). Furthermore, CTL may play a role in targeting latently infected cells that express Gag (but not infectious virus) (Graf et al., 2013). Developing an understanding of the most effective CTL responses to HIV may therefore support the HIV cure agenda.

## 6.5 Concluding remarks

The factors that determine the success or failure of the CTL response to HIV remain incompletely understood. Pinpointing the mechanisms of natural immune control remains an important strategy for defining the most effective immune responses. This is a critical area of ongoing research that is important for informing the design of intervention strategies for controlling the HIV epidemic. Defining the CTL response in individuals who express favourable HLA alleles and understanding how, why and in what context HLA alleles mediate immune control, may prove an important means of elucidating the determinants of effective CTL control of HIV. This body of work offers a contribution to that endeavor.

# APPENDIX I: Supplementary methods

## 1.1 Culture media and buffer constituents

### 1.1.1 Preparation of heat-inactivated FCS

FCS was heat-inactivated at 56°C for 30 minutes and filtered through a 0.45µm filter prior to use.

### 1.1.2 R10 cell culture medium

- 500ml RPMI
- 50ml heat inactivated FCS
- 5ml Penicillin-streptomycin
- 5ml L-Glutamine

### 1.1.3 Cell freezing medium

- 10% DMSO in heat-inactivated FCS

### 1.1.4 Electroporation medium

- 500ml RPMI
- 5ml L-Glutamine

### 1.1.5 EliSpot blocking buffer

- 5% FCS in PBS

## 1.2 PCR primers and annealing temperatures

**Table 1.1 PCR primers and annealing temperatures for amplification for population level and clonal sequencing of HIV-1 genes.** For nested PCRs, the first round PCR "outer primer" set is designated "OP". The second round PCR "inner primer" set is designated "IP".

| Fragment | Primer | | Sequence | | Annealing Temp (X/Y) |
|---|---|---|---|---|---|
| 1b (C Clade Gag) | 5'OP | 5' | CTCTAGCAGTGGCGCCCGAA | 3' | 60/57 |
| | 3'OP | 5' | TCCTTTCCACATTTCCAACAGCC | 3' | |
| | 5'IP | 5' | ACTCGGCTTGCTGAAGTGC | 3' | 57/54 |
| | 3'IP | 5' | CAATTTCTGGCTATGTGCCC | 3' | |
| 4 (C Clade Pol) | 5'OP | 5' | CATGGTGGACAGACTATTGGC | 3' | 61/58 |
| | 3'OP | 5' | CCTGCCATCTGTTTTCCATA | 3' | |
| | 5'IP | 5' | CCTCCCCTAGTAAAATTATGG | 3' | 60/57 |
| | 3'IP | 5' | GCTGTCTCTGTAATAAACCCG | 3' | |
| 8 (C Clade Nef) | 5'OP | 5' | GATCCATTCGATTAGTGAACGGA | 3' | 61/58 |
| | 3'OP | 5' | CTTATATGCAGCATCTGAGG | 3' | |
| | 5'IP | 5' | TTCAGCTACCACCGATTGAGA | 3' | 61/58 |
| | 3'IP | 5' | TGAGGGTTGGCCACTCC | 3' | |
| B Clade pol (F4) | 5'OP | 5' | CTGAAAACAGGAAAATATGCAAG | 3' | 58/56 |
| | 3'OP | 5' | TACTTGCCACACAATCATCACC | 3' | |
| | 5'IP | 5' | GAGGGGTGCCCACACTAATGA | 3' | 60/58 |
| | 3'!P | 5' | CTTTTCTTCTTGGCACTACTTTTA | 3' | |
| B Clade Nef N1 | 5'OP | 5' | TGCTCTGGAAAACTCATTTGCACC | 3' | 64/61 |
| | 3'OP | 5' | CAGGCTCAGATCTGGTCTAACCA | 3' | |
| | 5'IP | 5' | AACATGACCTGGATGGAGTGGGAAA | 3' | 64/61 |
| | 3'IP | 5' | TACAGGCAAAAAGCAGCTGCTTATATG | 3' | |
| B Clade Nef | 5'OP | 5' | TTCCAGTCAGACCTCAGGTAC | 3' | 60/57 |
| | 3'OP | 5' | TCTAGTTACCAGAGTCACACAA | 3' | |
| | 5'IP | 5' | CTTTTTAAAAGAAAAGGGGGGAC | 3' | 60/57 |
| | 3'IP | 5' | TCAGATCTGGTCTAACCAGAG | 3' | |

**Table 1.2 PCR primers for amplification of HCV NS5B for population level sequencing.** For nested PCRs, the first round PCR "outer primer" set is designated "OP". The second round PCR "inner primer" set is designated "IP".

| Primer | | Sequence | |
|---|---|---|---|
| Gag-Pro OP | 5' | CCACATCAACTCCGTGTGG | 3' |
| 3'OP | 5' | TTCATCGGTTGGGGAGGAGG | 3' |
| 3'IP | 5' | GGA GTG AGT TTG AGC TTG GT | 3' |

**Table 1.3 PCR primers for amplification of HIV *Gag-Pro* for construction of chimeric virus for VRC assays.** For nested PCRs, the first round PCR "outer primer" set is designated "OP". The second round PCR "inner primer" set is designated "IP".

| Primer | | Sequence | |
|---|---|---|---|
| GagPro OP Fwd | 5' | AAATCTCTAGCAGTGGCGCCCGAACAG | 3' |
| GagPro OP Rev | 5' | TTTAACCCTGCTGGGTGTGGTATYCCT | 3' |
| GagPro IP Fwd | 5' | GACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGC CAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGG | 3' |
| GagPro IP Rev | 5' | GGCCCAATTTTTGAAATTTTTCCTTCCTTTTCCATTTCTGTACAAATTTCTACTAATGCT TTTATTTTTTCTTCTGTCAATGGCCATTGTTTAACTTTTG | 3' |

## 1.3 Sequencing primers

**Table 1.4 Primers designed for sequencing of PCR amplified HIV genes.**

| Fragment | Primer | | Sequence | |
|---|---|---|---|---|
| 1b (Gag) | Frag1_5IP | 5’ | ACTCGGCTTGCTGAAGTGC | 3’ |
| | Frag1_2 | 5’ | CTGCAGAATGGGATAGTTACAT | 3’ |
| | Frag1_3 | 5’ | GACACCAAGGAAGCCTTAG | 3’ |
| | Frag1_4 | 5’ | CTCCCACTGGAATAGGTG | 3’ |
| | Frag1_5 | 5’ | GGAACAAATAGCATGGATGAC | 3’ |
| | Frag1_3IP | 5’ | CAATTTCTGGCTATGTGCCC | 3’ |
| 4 (Pol) | Frag4_5IP | 5’ | CCTCCCCTAGTAAAATTATGG | 3’ |
| | Frag4_2 | 5’ | GATTCACTCTTATCTGGTTGTG | 3’ |
| | Frag4_3 | 5’ | CAGACTCACAGTATGCATTAGG | 3’ |
| | Frag4_4 | 5’ | ACTGGCCATCTTCCTGCTA | 3’ |
| | Frag4_5 | 5’ | AGTGGCTACATAGAAGCAGAGGT | 3’ |
| | Frag4_3IP | 5’ | GCTGTCTCTGTAATAAACCCG | 3’ |
| 8 (Nef) | Frag8_5IP | 5’ | TTCAGCTACCACCGATTGAGA | 3’ |
| | Frag8_2 | 5’ | ACCTGAGGTCTGACTGGAAAG | 3’ |
| | Frag8_3 | 5’ | AGATAAACATGGGGCACTTAC | 3’ |
| | Frag8_4 | 5’ | AGCAGTCTTTGTAATACTCCG | 3’ |
| | Frag8_5 | 5’ | ACCTCAGGTACCTTTAAGAC | 3’ |
| | Frag8_3IP | 5’ | TGAGGGCTGGCCACTCC | 3’ |

## 1.4 Site-directed mutagenesis primers

SDM primers require HPLC purification. Primers should be 25-45 bases in length, have a melting temperature (Tm) of at least 78°C, encode the desired mutation in the middle of the primer, have a GC content of ~40% and terminate in a G or C base.

**Table 1.5 SDM primers for mutation of p6.5 HIV backbone to represent the C clade Gag-p24 consensus sequence.** The mutated amino acid residue in the primer sequence translation is shown in bold.

| Primer Name | Nucleotide Mutation | Primer Sequence | | | Primer length | Tm (°C) | Primer Sequence Translation |
|---|---|---|---|---|---|---|---|
| Revert T242S | T406A | 5' | GTGACATAGCAGGAACTACTAGTACCCTTCAGGAAC | 3' | 36 | 79 | DIAGTTS**T**LQE |
| | | 5' | GTTCCTGAAGGGTACTAGTAGTTCCTGCTATGTCAC | 3' | | | |
| Revert S252G I256V | G448A C450T | 5' | CAAATAGCATGGATGACAAGTAACCCACCTATTCCAGTGGGAGA | 3' | 45 | 79 | IAWMT**S**NPP**I**PVGD |
| | | 5' | TCTCCCACTGGAATAGGTGGGTTACTTGTCATCCATGCTATTTG | 3' | | | |
| Revert R286K | G539A | 5' | CTGTCAGCATTTTGGACATAAAACAAGGGCCAAAGG | 3' | 36 | 78 | VSILDI**K**QGPK |
| | | 5' | CCTTTGGCCCTTGTTTTATGTCCAAAATGCTGACAG | 3' | | | |
| Revert D319E | A639C | 5' | ACACAAGATGTAAAAAATTGGATGACAGACACCTTGTTGGTCCAA | 3' | 45 | 78 | TQDVKNWMT**D**TLL VQ |
| | | 5' | TTGGACCAACAAGGTGTCTGTCATCCAATTTTTTACATCTTGTGT | 3' | | | |
| Revert S357G | G751A | 5' | GGGAGTGGGAGGACCTAGCCACAAAGCAA | 3' | 29 | 79 | GVGGP**S**HKA |
| | | 5' | TTGCTTTGTGGCTAGGTCCTCCCACTCCC | 3' | | | |

**Table 1.6 SDM primers for mutation of C clade Gag p24 wildtype HIV backbone for production of recombinant variants.** The mutated amino acid residue in the primer sequence translation is shown in bold.

| Primer Name | Nucleotide Mutation | Primer Sequence | | | Primer length | Tm (°C) | Primer Sequence Translation |
|---|---|---|---|---|---|---|---|
| R264K | G473A | 5' | CCTATTCCAGTGGGAGACATCTATAAAAAATGGATAATTCTGGG | 3' | 44 | 80 | PIPVGDIYK**K**WIIL |
| | | 5' | CCCAGAATTATCCATTTTTTATAGATGTCTCCCACTGGAATAGG | 3' | | | |
| S165N | G176A | 5' | TAAAAGTAATAGAGGAGAAGGCTTTTAACCCGGAGGTAATAC | 3' | 42 | 79 | KVIEEKAF**N**PEVI |
| | | 5' | GTATTACCTCCGGGTTAAAAGCCTTCTCCTCTATTACTTTTA | 3' | | | |
| V168I (after S165N) | G184A | 5' | GAAGGCTTTTAACCCGGAGATAATACCCATGTTTACAGC | 3' | 39 | 80 | KAF**N**PE**I**IPMFT |
| | | 5' | GCTGTAAACATGGGTATTATCTCCGGGTTAAAAGCCTTC | 3' | | | |
| R264G | A472G | 5' | CTATTCCAGTGGGAGACATCTATAAAGGATGGATAATTCTGG | 3' | 42 | 79 | IPVGDIYK**G**WIIL |
| | | 5' | CCAGAATTATCCATCCTTTATAGATGTCTCCCACTGGAATAG | 3' | | | |
| D260N (after R264G) | G460A | 5' | ACCCACCTATTCCAGTGGGAAACATCTATAAAGGATGG | 3' | 38 | 79 | PPIPVG**N**IYK**G**W |
| | | 5' | CCATCCTTTATAGATGTTTCCCACTGGAATAGGTGGGT | 3' | | | |
| D260E (after R264G) | C462A | 5' | CCCACCTATTCCAGTGGGAGAAATCTATAAAGGATGGATAATT | 3' | 43 | 80 | PPIPVG**E**IYK**G**WII |
| | | 5' | AATTATCCATCCTTTATAGATTTCTCCCACTGGAATAGGTGGG | 3' | | | |
| V218M | G334A A336G | 5' | ATGGGACAGATTACATCCAATGCATGCAGGGCCTATTGCAC | 3' | 41 | 79 | WDRLHP**M**HAGPIA |
| | | 5' | GTGCAATAGGCCCTGCATGCATTGGATGTAATCTGTCCCAT | 3' | | | |
| T173A | A199G | 5' | CCCGGAGGTAATACCCATGTTTGCAGCATTATCAGAAG | 3' | 38 | 79 | PEVIPMF**A**ALSE |
| | | 5' | CTTCTGATAATGCTGCAAACATGGGTATTACCTCCGGG | 3' | | | |
| T173S | A199T | 5' | CCCGGAGGTAATACCCATGTTTTCAGCATTATCAGAAG | 3' | 38 | 79 | PEVIPMF**S**ALSE |
| | | 5' | CTTCTGATAATGCTGAAAACATGGGTATTACCTCCGGG | 3' | | | |
| S173T | A199A | 5' | CCCGGAGGTAATACCCATGTTTACAGCATTATCAGAAG | 3' | 38 | 79 | PEVIPMF**T**ALSE |
| | | 5' | CTTCTGATAATGCTGTAAACATGGGTATTACCTCCGGG | 3' | | | |

**Table 1.7 SDM primers for mutation of B clade Gag p24 wildtype HIV backbone (NL4-3) for production of recombinant variants.** The mutated amino acid residue in the primer sequence translation is shown in bold.

| Primer Name | Nucleotide Mutation | Primer Sequence | | | Primer length | Tm (°C) | Primer Sequence Translation |
|---|---|---|---|---|---|---|---|
| R264K | | 5' | CCTATCCCAGTAGGAGAAATCTATAAAAAATGGATAATCCTGGG | 3' | 44 | 80 | PIPVGEIYK**K**WIIL |
| | | 5' | CCCAGGATTATCCATTTTTTATAGATTTCTCCTACTGGGATAGG | 3' | | | |
| R264G | | 5' | CCTATCCCAGTAGGAGAAATCTATAAAGGATGGATAATCCTGG | 3' | 43 | 80 | PIPVGEIYK**G**WIIL |
| | | 5' | CCAGGATTATCCATCCTTTATAGATTTCTCCTACTGGGATAGG | 3' | | | |
| R264G + E260D | | 5' | CACCTATCCCAGTAGGAGACATCTATAAAGGATGGATAATCCTGGGATT | 3' | 49 | 80 | PIPVG**D**IYK**G**WIILG |
| | | 5' | AATCCCAGGATTATCCATCCTTTATAGATGTCTCCTACTGGGATAGGTG | 3' | | | |
| S173T | | 5' | GCCCAGAAGTAATACCCATGTTTACAGCATTATCAGAAG | 3' | 39 | 79 | PEVIPMF**T**ALSE |
| | | 5' | CTTCTGATAATGCTGTAAACATGGGTATTACTTCTGGGC | 3' | | | |
| S173A | | 5' | GCCCAGAAGTAATACCCATGTTTGCAGCATTATCAGAAG | 3' | 39 | 79 | PEVIPMF**A**ALSE |
| | | 5' | CTTCTGATAATGCTGCAAACATGGGTATTACTTCTGGGC | 3' | | | |

## 1.5 Cell lines

### 1.5.1 MT4 cell line (NIH Reagents)

The MT4 cell line is a human T cell line isolated from an adult with T cell leukemia. Cells are CD4+ and highly sensitive to HIV infection.

### 1.5.2 CEM-GXR cell line

The CEM GXR cell line is a T cell reporter line, which originally expressed CD4 and CXCR4 (Gervaix et al., 1997) and was later modified to express CCR5 as well (Brockman et al., 2006). It is therefore both permissive to R5 and X4-tropic HIV infection. GFP is expressed upon infection with HIV in a Tat-dependent manner.

## 1.6 Viral plasmids

### 1.6.1 NL4-3

NL4-3 (NIH reagents) is a complete genome replication competent laboratory strain of B clade HIV-1, encoded in an ampicillin resistant plasmid (Adachi et al., 1986).

### 1.6.2 p83-2 and p83-$10_{GFP}$

The NL4-3 sequence is available as two vectors expressing approximately half the genome respectively (NIH reagents). The p83-2 plasmid encodes the 5' end of the genome (5' LTR, Gag, Pol, Vif and Vpr). The p83-$10_{GFP}$ plasmid encodes the 3' end of the genome (Tat, Rev, Vpu, Env, and 3' LTR) and a GFP reporter. The plasmids

are linearised by EcoRI restriction digest and cotransfected. Recombination allows for expression of the full length virus and GFP reported for detecting infected cells.

### 1.6.3 ΔGag-Pro NL4-3

ΔGag-Pro NL4-3 is a variant of the NL4-3 plasmid with the *Gag* and *Protease* genes deleted (Brockman et al., 2006; Miura, Brockman, Brumme, et al., 2009). This plasmid serves as a backbone for production of chimeric virus with Gag-Pro region derived from autologous patient virus. The plasmid was engineered to express BstEII restriction enzyme sites on either side of the Gag-Pro region. Restriction digest with BstEII produces a *Gag-Pro* deleted variant with self-ligated BstEII restriction site. The plasmid is linearised for insertion of patient-derived *Gag-Pro* by BstEII restriction digest.

### 1.6.4 p6.5

The plasmid p6.5 is a variant of p83-2 in which the p24 gene has been replaced by a C clade version of this gene derived from a C clade infected subject.

# APPENDIX II: Illumina sequencing technology

## 2.1 Single-stranded DNA library preparation and sequencing-by-synthesis using Illumina technology

The single-stranded (ss) DNA library is prepared by shearing 500ng of double-stranded (ds) DNA template (PCR product) using ultrasound to produce 100-500bp fragments. A 2100 Bioanalyzer (Agilent Technologies) is used to assess fragment length and 200-300bp fragments size selected. Complementary nucleic acid 'adaptors' are attached to each end of the fragments (Figure 2.1 A). The dsDNA template is denatured and immobilised to a solid support (flowcell) densely coated with complementary adaptors. Adaptors at both ends of the ssDNA template bind to the solid support, forming a 'bridge'. The adaptors act as primers for bridge amplification PCR, producing dsDNA molecules. These are denatured to produce ssDNA templates attached at one end for further amplification. Up to 40 million dense clusters of dsDNA with approximately 1000 copies of ssDNA in each cluster can be produced. During bridge amplification PCR terminator nucleotides (with each of the four nucleotides labeled with a different fluorescent dye) are incorporated to allow sequencing by synthesis. Image capture of the colour and position of fluorescent dyes on the flow cell after each nucleotide addition allow elucidation of the DNA template sequence.

## 2.2 Multiplexing

Fragments from multiple samples can be pooled and sequenced in the same run in multiplex Illumina sequencing. One of 96 unique nucleotide barcodes are added

between to the template DNA and adaptor for each sample (Figure 2.1A). Reads are sorted based on this barcode during data analysis. Multiplexing massively increases the number of samples that can be run simultaneously.

## 2.3 Paired-end reads

Illumina sequencing technology allows generation of reads from adaptors at both ends of the DNA fragment (paired-end reads off the same template strand) (Figure 2.1 B and C). This yields much more information than single end reads of the same length. It allows determination of longer sequences from overlapping reads (Figure 2.1 B) and provides an indication of the physical distance between reads (Figure 2.1 C).

## 2.4 Data analysis for *de novo* assembly of consensus sequences

A work-flow for data analysis provided by collaborators at the Wellcome Trust Sanger Institute is shown in (Figure 2.3).

The steps involved in data analysis are;

- Quality control of raw reads. This removes reads <50bp and trims low quality bases from the 3' end until the median Phred quality score is 30 (representing a 1 in 1000 probability of an incorrect base call).
- *De novo* assembly is used to produce an initial consensus sequence
- Quality controlled reads are mapped against the initial consensus sequence and a final consensus sequence assembled.

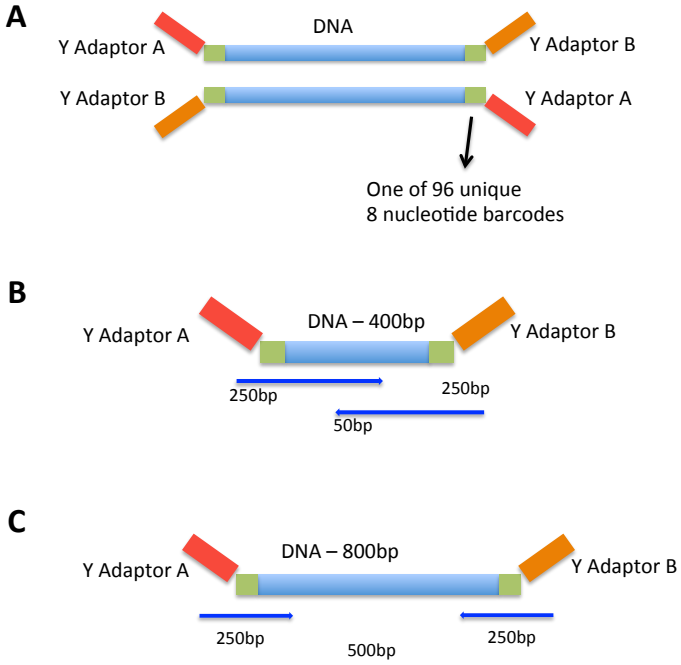



**Figure 2.1 Illustration of DNA templates for multiplexed and paired-end sequencing with Illumina technology.** A) Nucleotide adaptors are attached to each end of the fragmented DNA sequencing templates. These attach the DNA to the flow cell (solid support) and act as the primers for subsequent amplification. Nucleotide barcodes are added for multiplex sequencing of multiple samples. B) Using paired-end reads from both adaptors allows longer sequences to be assembled by overlapping the reads from each end. C) For longer reads, paired-end technology allows estimation of the distance between the reads, allowing easier mapping of reads. Sequence data from the gap between the reads is obtained from other fragments and mapped in a similar way allowing assembly of the full sequence.

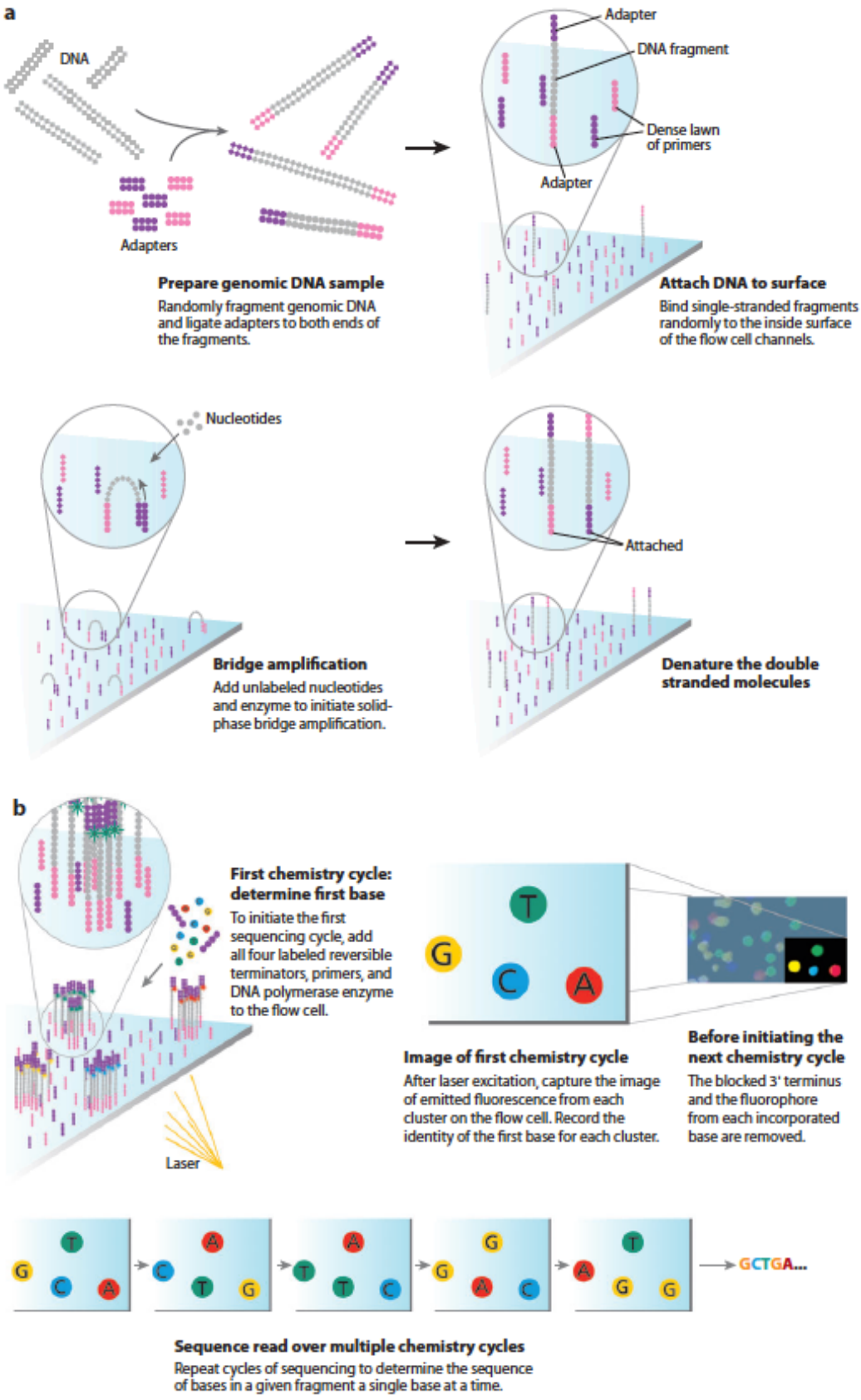
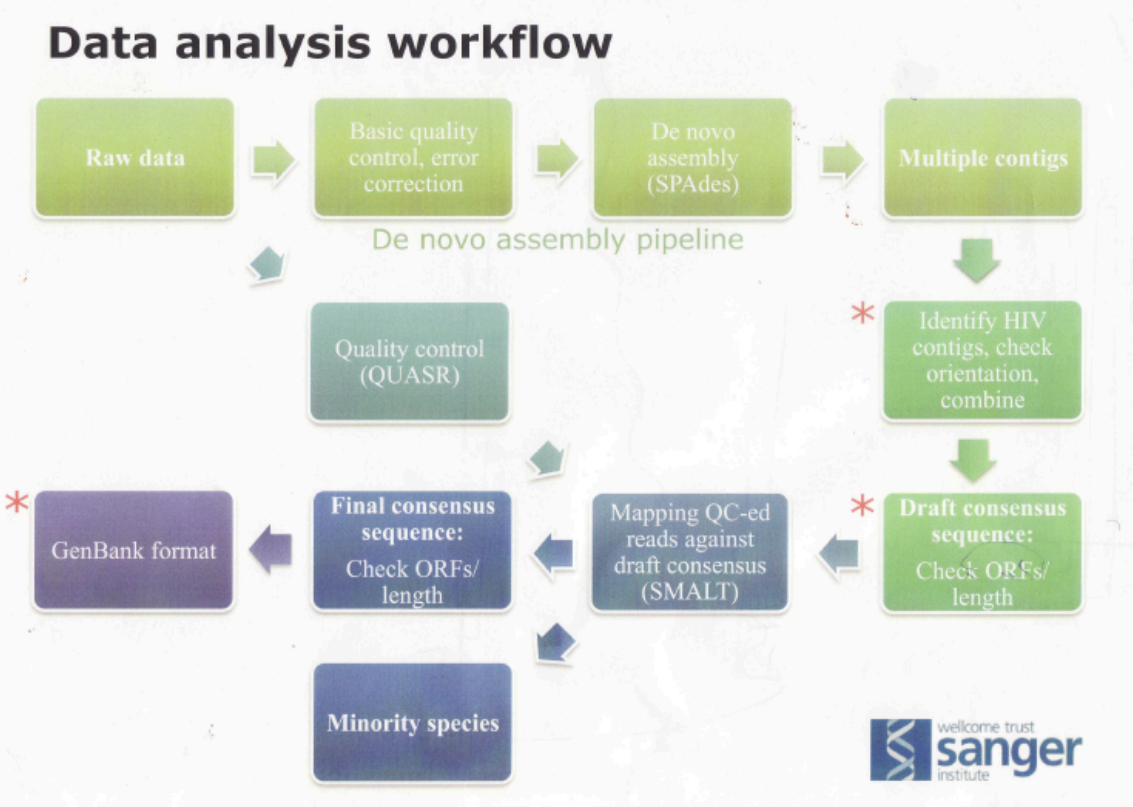



**Figure 2.2 Schematic representation of Illumina sequencing technology.** Reproduced from (Mardis, 2008) with permission from Annual Reviews.

**Figure 2.3 Workflow of data analysis for *de novo* assembly of consensus sequences.**



# APPENDIX III: Directions for ongoing analysis of transmission pair R036/R056

Chapter 3 describes the case of the C clade HIV transmission pair R036/R056, in which the HLA-B*27:05-positive recipient progressed more rapidly than expected (Figure 3.1). Here, additional details obtained from ultra-deep viral sequencing of this pair and directions for ongoing analysis of this data, are presented.

## 3.1 Reversion of transmitted Gag-T242N in recipient R056

In transmission pair studies, there is an opportunity to examine reversion of transmitted mutations in the HLA mismatched recipient (Song et al., 2014). The rate of reversion upon transmission may be used to infer the fitness cost of transmitted mutations. Ultra-deep sequences were examined across the epitopes restricted by the donor's HLA B alleles in both donor and recipient sequences to look for evidence of reversion of mutations associated with these alleles upon transmission (Table 3.1). Consistent with clinical history, HLA-B*58-associated footprints in the early recipient sequence in Gag-TW10 (T242N) and Nef-KF9 (A83G) provide evidence to support the direction of transmission from R036 and R056 (Carlson, Brumme, et al., 2012). No HLA-B*14-associated variants were transmitted to the recipient (data not shown). In the majority of HLA-B*58 epitopes, the transmitted variant was stably transmitted and persisted in the recipient throughout the period of follow up (Table 3.1). A single exception was

observed in Gag-TW10. The Gag-T242N escape variant was transmitted to the recipient but reverted to the consensus residue within 2 years of transmission. This is consistent with the literature on the fitness cost and rate of reversion of this mutation (Leslie et al., 2004; Martinez-picado et al., 2006).

## 3.2 Intra-host diversity in donor R036 and recipient R056

Heat map analysis as described in chapter 5 may be used to represent the intra-host population at a single timepoint to provide a snapshot of diversity in a particular patient. An example of this analysis is shown for the donor R036 at 53 months, representing the minor variants compared to the consensus sequence (Figure 3.1A). Diversity is evident at sites distributed across the full proteome, with the greatest diversity on the Env protein, which is well recognised as the most highly variable protein in HIV.

To examine which sites across the full-length HIV genome were under strong selective pressure in the recipient, we compared the consensus sequence at the earliest timepoint in the recipient (19 months) to the intra-host population at 59 months (Figure 1.1B), thus representing viral sequence diversity selected over 40 months. The sites at which there is complete replacement of the early consensus amino acid residue (red squares in the heatmap) show the sites at which the early intra-host population has been entirely replaced, or sites at which a minor variant present at the early timepoint reaches fixation by the later timepoint. These sites are thus likely to be under strong selective pressure in the recipient. Initial analysis has shown that 5% and 7% of amino acid residues in Gag and Pol are

under strong selective pressure. Ongoing work will examine what percentage of sites under selective pressure across the full genome might be attributed to HLA-specific immune selection.

**Table 3.1 Ultra-deep sequencing of HLA-B*58-restricted epitopes in a C clade infected transmission pair (R036/R056).** Autologous variant haplotypes are aligned to the C clade consensus sequence.

| HLA-B*58-Restricted Epitopes | | Frequency (%) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Donor (Month) | Recipient (Month) | | | | | | | | | |
| | C Clade Consensus and Autologous Variants | 19 | 19 | 21 | 24 | 28 | 33 | 41 | 44 | 51 | 54 | 59 |
| Gag-TW10 | $^{240}$TSTLQEQIAW$^{249}$ | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --N-----N- | 93 | 73 | 99 | | 92 | 78 | 0 | 1 | 0 | 0 | 0 |
| | --N-----G- | 0 | 11 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --N------- | 4 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --N-A----- | 2 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --------N- | 0 | 8 | 0 | | 7 | 22 | 99 | 98 | 99 | 98 | 97 |
| | --------G- | 0 | 6 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -----G---- | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Gag-QW9 | $^{308}$QATQDVKNW$^{316}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | ----E---- | 54 | 83 | 99 | 99 | 99 | 100 | 99 | 98 | 98 | 99 | 100 |
| | --S-E---- | 45 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pol-IW9 | $^{474}$IAMESIVIW$^{482}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | --Q------ | 96 | 83 | 99 | 98 | 98 | 99 | 99 | 97 | 97 | 99 | 100 |
| | --T-G---- | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pol-SW9 | $^{782}$SAAVKAACW$^{790}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -N------- | 98 | 100 | 97 | 99 | 99 | 100 | 98 | 99 | 99 | 99 | 100 |
| | --T------- | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Pol-KF9 | $^{832}$KTAVQMAVF$^{840}$ | 100 | 100 | 99 | 98 | 99 | 100 | 100 | 99 | 99 | 99 | 100 |
| Rev-KY10 | $^{14}$QAVRIIKILY$^{23}$ | 81 | 57 | 100 | 95 | 85 | 0 | 0 | 0 | 0 | 0 | 2 |
| | -------L-- | 0 | 43 | 0 | 0 | 0 | 76 | 97 | 97 | 99 | 99 | 95 |
| | ---K------ | 0 | 0 | 0 | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -T-------- | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | ---K--R--- | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| | ----T----- | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -T-K------ | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -------V-- | 0 | 0 | 0 | 0 | 0 | 22 | 2 | 2 | 0 | 0 | 0 |
| | -------L-H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| | ---A-T-R--- | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | ---A-T----- | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Env-KW11 | $^{59}$KAYEKEVHNVW$^{69}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 |
| | -G--------- | 97 | 100 | 99 | 98 | 99 | 99 | 99 | | 94 | 50 | 29 |
| | ------A---- | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 5 | 41 | 70 |
| | -A-DR-A---- | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 2 |
| | ------M---- | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 8 | 0 |
| | -G--R------ | 2 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 |
| Nef-KF9 | $^{82}$KAAFDLSFF$^{90}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | -G------- | 99 | 100 | 99 | 99 | 99 | 100 | 99 | 98 | 99 | 99 | 96 |
| Nef-HW9 | $^{116}$HTQGYFPDW$^{124}$ | 40 | 100 | 100 | 99 | 99 | 100 | 100 | 100 | 100 | 100 | 96 |
| | N---F---- | 59 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

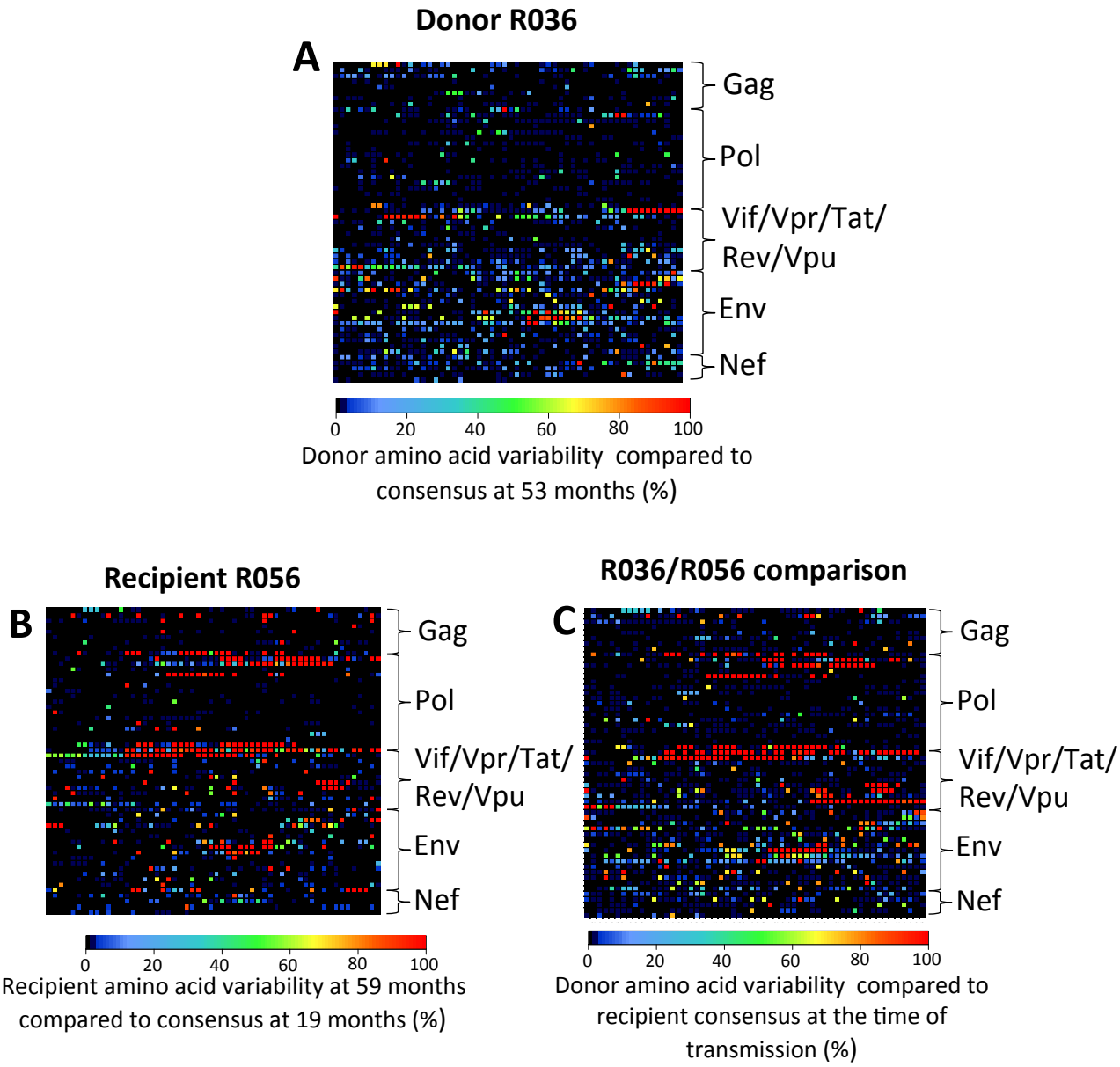



**Figure 3.1 Heat map representations of inter- and intra-host diversity of HIV quasispecies in transmission pair R036/R056.** A) Intra-host amino acid diversity in the donor at 53 months. Variation in the intra-host population is compared to the consensus sequence at the same timepoint to represent a snapshot of diversity at this time. B) Longitudinal intra host amino acid diversity in the recipient over 40 months. The intra-host population at 59 months is compared to the consensus sequence at 19 months to highlight sites selective pressure over 40 months. C) Donor amino acid variability at 19 months compared to the recipient consensus sequence (closest approximation of the transmitted founder) at the same timepoint, highlight sites of possibly minor variant transmission.

## 3.3 Transmission of minor variants to recipient R056

The intra-host population of HIV sequences in a donor undergoes a severe bottleneck at transmission, often limiting the transmitted founder to a single variant (Joseph et al., 2015; Keele et al., 2008). The CTL response may play in important role in this bottleneck in determining which variants are present in the infecting swarm in the donor, and which CTL escape mutations, which affect the VRC of the transmitted virus, might be present (Joseph et al., 2015). A number of studies have provided evidence to support selection of the most fit viral variants and/or the variants closest to consensus at transmission (Carlson et al., 2014; Herbeck et al., 2006; Sagar et al., 2009; Strickland et al., 2012). Information on the selective pressure favouring transmission of particular viral sequence may be useful for directing intervention strategies toward the most frequently transmitted variants. Ultra-deep sequencing of transmission pairs provides an opportunity to examine this. Transmission of minor variant residues more frequently than expected by chance suggests that these variants may be specifically selected for transmission. The effect of these residues on VRC can then be examined to determine whether the selective advantage at transmission is associated with maximising VRC of the transmitted variant.

Follow up of transmission pair R036/R056 from a time close to the time of infection of the recipient provided a rare opportunity to examine how frequently transmission of donor minor variants occurs. In this analysis, the earliest viral sequence from the recipient represents the closest approximation of the transmitted founder virus. Comparison of this sequence to the intra-host population of the donor at a closely matched timepoint allows for detection of

amino acid mismatches that may reflect minor variant transmission events (Figure 1.1C). Initial analysis of Gag has shown that minor variant transmission may be a very rare event, occurring at only 3/505 amino acids. Ongoing analysis will examine minor variant transmission frequency across the full-length genome and examine how these specific residues may affect VRC of the transmitted variant.

# REFERENCES


Abdul-Jawad, S., Ondondo, B., van Hateren, A., Gardner, A., Elliott, T., Korber, B., & Hanke, T. (2015). Increased valency of conserved-mosaic vaccines enhances the breadth and depth of epitope recognition. *Molecular Therapy*, 1–10. doi:10.1038/mt.2015.210

Abraha, A., Nankya, I. L., Gibson, R., Demers, K., Tebit, D. M., Johnston, E., … Arts, E. J. (2009). CCR5- and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than other dominant group M subtypes: implications for the epidemic. *Journal of Virology*, *83*(11), 5592–5605. doi:10.1128/JVI.02051-08

Adachi, A., Gendelman, H., Koenig, S., Folks, T., Willey, R., Rabson, A., & Martin, M. (1986). Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *Journal of Virology*, *59*(2), 284–291.

Addo, M. M., Yu, X. G., Rathod, A., Cohen, D., Eldridge, R. L., Strick, D., … Johnston, M. N. (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. *Journal of Virology*, *77*(3), 2081–2092. doi:10.1128/JVI.77.3.2081

Adland, E., Paioni, P., Thobakgale, C., Laker, L., Mori, L., Muenchhoff, M., … Goulder, P. J. R. (2015). Discordant impact of HLA on viral replicative capacity and disease progression in pediatric and adult HIV infection. *PLOS Pathogens*, *11*(6), e1004954. doi:10.1371/journal.ppat.1004954

Allen, T. M., Altfeld, M., Geer, S. C., Kalife, E. T., Moore, C., O'sullivan, K. M., … Walker, B. D. (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. *Journal of Virology*, *79*(21), 13239–13249. doi:10.1128/JVI.79.21.13239

Almeida, J. R., Price, D., Papagno, L., Arkoub, Z. A., Sauce, D., Bornstein, E., … Appay, V. (2007). Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of Experimental Medicine*, *204*(10), 2473–2485. doi:10.1084/jem.20070784

Almeida, J. R., Sauce, D., Price, D., Papagno, L., So, Y. S., Moris, A., … Appay, V. (2009). Antigen sensitivity is a major determinant of CD8+ T-cell polyfunctionality and HIV-suppressive activity. *Blood*, *113*(25), 6351–6360. doi:10.1182/blood-2009-02-206557

Alter, G., Heckerman, D., Schneidewind, A., Fadda, L., Kadie, C. M., Carlson, J. M., … Altfeld, M. (2011). HIV-1 adaptation to NK-cell-mediated immune pressure. *Nature*, *476*(7358), 96–100. doi:10.1038/nature10237

Alter, G., Martin, M. P., Teigen, N., Carr, W. H., Suscovich, T. J., Schneidewind, A., … Altfeld, M. (2007). Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *The Journal of Experimental Medicine*, *204*(12), 3027–3036. doi:10.1084/jem.20070695

Altfeld, M., Allen, T. M., Yu, X. G., Johnston, M. N., Agrawal, D., Korber, B. T., … Walker, B. D. (2002). HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. *Nature*, *420*(6914), 434–439. doi:10.1038/nature01662

Altfeld, M., & Gale Jr, M. (2015). Innate immunity against HIV-1 infection. *Nature Immunology*, *16*(6), 554–562. doi:10.1038/ni.3157

Altfeld, M., Kalife, E. T., Qi, Y., Streeck, H., Lichterfeld, M., Johnston, M. N., … Walker, B. D. (2006). HLA alleles associated with delayed progression to AIDS contribute strongly to the initial CD8+ T cell response against HIV-1. *PLoS Medicine*, *3*(10), 1851–1864. doi:10.1371/journal.pmed.0030403

Ammaranond, P., van Bockel, D. J., Petoumenos, K., McMurchie, M., Finlayson, R., Middleton, M. G., … Kelleher, A. D. (2011). HIV immune escape at an immunodominant epitope in HLA-B*27-positive individuals predicts viral load outcome. *Journal of Immunology*, *186*(1), 479–488. doi:10.4049/jimmunol.0903227

An, P., & Winkler, C. (2010). Host genes associated with HIV/AIDS: advances in gene discovery. *Trends in Genetics*, *26*(3), 119–131. doi:10.1016/j.tig.2010.01.002

Archin, N. M., Espeseth, A., Parker, D., Cheema, M., Hazuda, D., & Margolis, D. M. (2009). Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid. *AIDS Research and Human Retroviruses*, *25*(2), 207–212. doi:10.1089/aid.2008.0191

Archin, N. M., Liberty, A. L., Kashuba, A. D., Choudhary, S. K., Kuruc, J. D., Crooks, A. M., … Margolis, D. M. (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*, *487*(7408), 482–485. doi:10.1038/nature11286

Ariën, K. K., Abraha, A., Quiñones-Mateu, M. E., Kestens, L., Vanham, G., & Arts, E. J. (2005). The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *Journal of Virology*, *79*(14), 8979–8990. doi:10.1128/JVI.79.14.8979-8990.2005

Bailey, J. R., O'Connell, K., Yang, H.-C., Han, Y., Xu, J., Jilek, B., … Blankson, J. N. (2008). Transmission of human immunodeficiency virus type 1 from a patient who developed AIDS to an elite suppressor. *Journal of Virology*, *82*(15), 7395–7410. doi:10.1128/JVI.00800-08

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. doi:10.1089/cmb.2012.0021

Barouch, D. H., & Michael, N. L. (2014). Accelerating HIV-1 vaccine efficacy trials. *Cell*, *159*(5), 969–972. doi:10.1016/j.cell.2014.10.046

Barouch, D. H., O'Brien, K. L., Simmons, N. L., King, S. L., Abbink, P., Maxfield, L. F., … Korber, B. (2010). Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. *Nat Med*, *16*(3), 319–323.

Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., … Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for Acquired Immune Deficiency Syndrome (AIDS). *Science*, *220*(4599), 868–871. doi:10.2307/1690359

Bartha, I., Carlson, J. M., Brumme, C. J., McLaren, P. J., Brumme, Z. L., John, M., … Fellay, J. (2013). A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife*, *2013*(2), e01123. doi:10.7554/eLife.01123

Belachew, D. A., Sandu, N., Schaller, B., & Guta, Z. (2009). Ankylosing spondylitis in sub-Saharan Africa. *Postgraduate Medical Journal*, *85*(1005), 353–357. doi:10.1136/pgmj.2007.064717

Bennett, M. S., Ng, H. L., Dagarag, M., Ali, A., & Yang, O. O. (2007). Epitope-dependent avidity thresholds for cytotoxic T-lymphocyte clearance of virus-infected cells. *Journal of Virology*, *81*(10), 4973–4980. doi:10.1128/JVI.02362-06

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53–59. doi:10.1038/nature07517

Betts, M. R., Nason, M. C., West, S. M., De Rosa, S. C., Migueles, S. A., Abraham, J., … Koup, R. A. (2006). HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood*, *107*(12), 4781–4789. doi:10.1182/blood-2005-12-4818

Bhardwaj, N., Hladik, F., & Moir, S. (2012). The immune response to HIV. *Nature Poster*. Retrieved from http://www.nature.com/nri/posters/hiv/index.html

Blankson, J. N., Bailey, J. R., Thayil, S., Yang, H.-C., Lassen, K., Lai, J., … Siliciano, R. F. (2007). Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. *Journal of Virology*, *81*(5), 2508–2518. doi:10.1128/JVI.02165-06

Blish, C. A., Dogan, O. C., Jaoko, W., Mcclelland, R. S., Mandaliya, K., Odem-davis, K. S., … Overbaugh, J. (2012). Cellular immune responses and susceptibility to HIV-1 superinfection: a case-control study. *Aids*, *26*, 643–651.

Boasso, A., & Shearer, G. M. (2008). Chronic innate immune activation as a cause of HIV-1 immunopathogenesis. *Clinical Immunology*, *126*(3), 235–242. doi:10.1016/j.clim.2007.08.015

Borghans, J. A. M., Mølgaard, A., de Boer, R. J., & Keşmir, C. (2007). HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. *PLoS ONE*, *2*(9), e920. doi:10.1371/journal.pone.0000920

Borrow, P., Lewicki, H., & Hahn, B. (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*(9), 6103–6110.

Borrow, P., Lewicki, H., Wei, X., Horwitz, M. S., Peffer, N., Meyers, H., … Shaw, G. M.

(1997). Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature Medicine*, *3*(2), 205–211. doi:10.1038/nm0297-205

Borthwick, N., Ahmed, T., Ondondo, B., Hayes, P., Rose, A., Ebrahimsa, U., … Hanke, T. (2014). Vaccine-elicited human T cells recognizing conserved protein regions inhibit HIV-1. *Molecular Therapy : The Journal of the American Society of Gene Therapy*, *22*(2), 464–475. doi:10.1038/mt.2013.248

Bosinger, S. E., Li, Q., Gordon, S. N., Klatt, N. R., Duan, L., Xu, L., … Kelvin, D. J. (2009). Global genomic analysis reveals rapid control of a robust innate response in SIV-infected sooty mangabeys. *The Journal of Clinical Investigation*, *119*(12), 3556–3572. doi:10.1172/JCI40115

Boutwell, C. L., Rowley, C. F., & 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. *Journal of Virology*, *83*(6), 2460–2468. doi:10.1128/JVI.01970-08

Brenchley, J. M., Price, D. A., Schacker, T. W., Asher, T. E., Silvestri, G., Rao, S., … Douek, D. C. (2006). Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nature Medicine*, *12*(12), 1365–1371. doi:10.1038/nm1511

Brenchley, J. M., Schacker, T. W., Ruff, L. E., Price, D. A., Taylor, J. H., Beilman, G. J., … Douek, D. C. (2004). CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *The Journal of Experimental Medicine*, *200*(6), 749–759. doi:10.1084/jem.20040874

Brener, J., Gall, A., Batorsky, R., Riddell, L., Buus, S., Leitman, E., … Matthews, P. C. (2015). Disease progression despite protective HLA expression in an HIV-infected transmission pair. *Retrovirology*, *12*(55). doi:10.1186/s12977-015-0179-z

Brockman, M. A., Brumme, Z. L., Brumme, C. J., Miura, T., Sela, J., Rosato, P. C., … Allen, T. M. (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. *Journal of Virology*, *84*(22), 11937–11949. doi:10.1128/JVI.01086-10

Brockman, M. A., Schneidewind, A., Lahaie, M., Schmidt, A., Miura, T., Desouza, I., … Allen, T. M. (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*(22), 12608–12618. doi:10.1128/JVI.01369-07

Brockman, M., Tanzi, G. O., Walker, B. D., & Allen, T. M. (2006). Use of a novel GFP reporter cell line to examine replication capacity of CXCR4- and CCR5-tropic HIV-1 by flow cytometry. *Journal of Virological Methods*, *131*(2), 134–142. doi:10.1016/j.jviromet.2005.08.003

Brumme, Z. L., Brumme, C. J., Carlson, J., Streeck, H., John, M., Eichbaum, Q., … Walker, B. D. (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. *Journal of Virology*, *82*(18), 9216–9227. doi:10.1128/JVI.01041-08

Buchbinder, S. P., Mehrotra, D. V, Duerr, A., Fitzgerald, D. W., Mogg, R., Li, D., … Robertson, M. N. (2008). Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *The Lancet*, *372*(9653), 1881–1893. doi:10.1016/S0140-6736(08)61591-3

Buckheit, R. W. 3rd, Allen, T. G., Alme, A., Salgado, M., O'Connell, K. A., Huculak, S., … Blankson, J. N. (2012). Host factors dictate control of viral replication in two HIV-1 controller/chronic progressor transmission pairs. *Nature Communications*, *3*, 716. doi:10.1038/ncomms1697

Buonaguro, L., Tornesello, M. L., & Buonaguro, F. M. (2007). Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *Journal of Virology*, *81*(19), 10209–10219. doi:10.1128/JVI.00872-07

Buranapraditkun, S., Hempel, U., Pitakpolrat, P., Allgaier, R. L., Thantivorasit, P., Lorenzen, S.-I. I., … Ruxrungtham, K. (2011). A novel immunodominant CD8+ T cell response restricted by a common HLA-C allele targets a conserved region of Gag HIV-1 clade CRF01_AE infected Thais. *PloS One*, *6*(8), e23603. doi:10.1371/journal.pone.0023603

Burton, D. R., & Mascola, J. R. (2015). Antibody responses to envelope glycoproteins in HIV-1 infection. *Nature Immunology*, *16*(6), 571–576. doi:10.1038/ni.3158

Cano, P., Klitz, W., Mack, S. J., Maiers, M., Marsh, S. G. E., Noreen, H., … Fernández-Viña, M. (2007). Common and well-documented HLA alleles: Report of the ad-hoc committee of the American Society for Histocompatiblity and Immunogenetics. *Human Immunology*, *68*(5), 392–417. doi:10.1016/j.humimm.2007.01.014

Carlson, J. M., Brumme, C. J., Martin, E., Listgarten, J., Brockman, M., Le, A. Q., … Brumme, Z. L. (2012). Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. *Journal of Virology*, *86*(24), 13202–16. doi:10.1128/JVI.01998-12

Carlson, J. M., Brumme, Z. L., Rousseau, C. M., Brumme, C. J., Matthews, P., Kadie, C., … Heckerman, D. (2008). Phylogenetic dependency networks: Inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS Computational Biology*, *4*(11), e1000225. doi:10.1371/journal.pcbi.1000225

Carlson, J. M., Listgarten, J., Pfeifer, N., Tan, V., Kadie, C., Walker, B. D., … Heckerman, D. (2012). Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. *Journal of Virology*, *86*(9), 5230–5243. doi:10.1128/JVI.06728-11

Carlson, J. M., Schaefer, M., Monaco, D. C., Batorsky, R., Claiborne, D. T., Prince, J., … Hunter, E. (2014). Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science*, *345*(6193), 1254031. doi:10.1126/science.1254031

Casado, C., Colombo, S., Rauch, A., Martinez, R., Gunthard, H. F., Garcia, S., … Lopez-

Galindez, C. (2010). Host and viral genetic correlates of clinical definitions of HIV-1 disease progression. *PloS One*, *5*(6), e11079. doi:10.1371/journal.pone.0011079

Casado, C., Pernas, M., Alvaro, T., Sandonis, V., García, S., Rodríguez, C., … Lopez-Galindez, C. (2007). Coinfection and superinfection in patients with long-term, nonprogressive HIV-1 disease. *The Journal of Infectious Diseases*, *196*(6), 895–899. doi:10.1086/520885

Caskey, M., Klein, F., Lorenzi, J. C. C., Seaman, M. S., West, A. P., Buckley, N., … Nussenzweig, M. C. (2015). Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature*, *522*(7557), 487–491. doi:10.1038/nature14411

Charlebois, P., Yang, X., Newman, R., Henn, M., & Zody, M. (2012). V-FAT: a post-assembly pipeline for the finishing and annotation of viral genomes. Retrieved from www.broadinstitute.org/scientific-community/science/projects/viral-genomics/v-fat

Chavez, L., Calvanese, V., & Verdin, E. (2015). HIV latency Is established directly and early in both resting and activated primary CD4 T cells. *PLOS Pathogens*, *11*(6), e1004955. doi:10.1371/journal.ppat.1004955

Chen, H., Ndhlovu, Z. M., Liu, D., Porter, L. C., Fang, J. W., Darko, S., … Walker, B. D. (2012). TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nature Immunology*, *13*(7), 691–700. doi:10.1038/ni.2342

Chikata, T., Carlson, J. M., Tamura, Y., Borghan, M. A., Naruto, T., Hashimoto, M., … Takiguchi, M. (2014). Host-specific adaptation of HIV-1 subtype B in the Japanese population. *Journal of Virology*, *88*(9), 4764–4765. doi:10.1128/JVI.00147-14

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F. A., Yassine-Diab, B., … Sekaly, R.-P. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature Medicine*, *15*(8), 893–900. doi:10.1038/nm.1972

Chopera, D. R., Woodman, Z., Mlisana, K., Mlotshwa, M., Martin, D. P., Seoighe, C., … Williamson, C. (2008). Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. *PLoS Pathogens*, *4*(3), e1000033. doi:10.1371/journal.ppat.1000033

Choudhary, S. K., Vrisekoop, N., Jansen, C., Otto, S., Schuitemaker, H., Miedema, F., & Camerini, D. (2007). Low immune activation despite high levels of pathogenic human immunodeficiency virus type 1 results in long-term asymptomatic disease. *Journal of Virology*, *81*(16), 8838–8842. doi:10.1128/JVI.02663-06

Chowdhury, A., Hayes, T. L., Bosinger, S. E., Lawson, B. O., Vanderford, T., Schmitz, J. E., … Silvestri, G. (2015). Differential impact of in vivo CD8+ T lymphocyte depletion in controller versus progressor simian immunodeficiency virus-infected macaques. *Journal of Virology*, *89*(17), 8677–8686. doi:10.1128/JVI.00869-15

Chung, R. T., Andersen, J., Volberding, P., Robbins, G. K., Liu, T., Sherman, K. E., …

van der Horst, C. (2004). Peginterferon Alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. *The New England Journal of Medicine*, *351*(5), 451–459. doi:10.1056/NEJMoa032653

Cicala, C., Arthos, J., & Fauci, A. S. (2011). HIV-1 envelope, integrins and co-receptor use in mucosal transmission of HIV. *Journal of Translational Medicine*, *9*(Suppl 1), S2. doi:10.1186/1479-5876-9-S1-S2

Claiborne, D. T., Prince, J. L., Scully, E., Macharia, G., Micci, L., Lawson, B., … Hunter, E. (2015). Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4+ T cells, and disease progression. *Proceedings of the National Academy of Sciences*, *112*(12), E1480–9. doi:10.1073/pnas.1421607112

Clapham, P. R., & McKnight, Á. (2001). HIV-1 receptors and cell tropism. *British Medical Bulletin*, *58*, 43–59. doi:10.1093/bmb/58.1.43

Clerc, O., Colombo, S., Yerly, S., Telenti, A., & Cavassini, M. (2010). HIV-1 elite controllers: Beware of super-infections. *Journal of Clinical Virology*, *47*(4), 376–378. doi:10.1016/j.jcv.2010.01.013

Costa, A. I., Koning, D., Ladell, K., McLaren, J. E., Grady, B. P. X., Schellens, I. M. M., … van Baarle, D. (2015). Complex T-cell receptor repertoire dynamics underlie the CD8+ T-cell response to HIV-1. *Journal of Virology*, *89*(1), 110–119. doi:10.1128/JVI.01765-14

Cotton, L. a., Kuang, X. T., Le, A. Q., Carlson, J. M., Chan, B., Chopera, D. R., … Brumme, Z. L. (2014). Genotypic and functional impact of HIV-1 adaptation to its host population during the North American epidemic. *PLoS Genetics*, *10*(4), e1004295. doi:10.1371/journal.pgen.1004295

Couturier, J., Suliburk, J. W., Brown, J. M., Luke, D. J., Agarwal, N., Yu, X., … Lewis, D. E. (2015). Human adipose tissue as a reservoir for memory CD4+ T cells and HIV. *AIDS*, *29*(6), 667–674. doi:10.1097/QAD.0000000000000599

Crawford, H., Lumm, W., Leslie, A., Schaefer, M., Boeras, D., Prado, J. G., … Hunter, E. (2009). Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. *The Journal of Experimental Medicine*, *206*(4), 909–921. doi:10.1084/jem.20081984

Crawford, H., Matthews, P. C., Schaefer, M., Carlson, J. M., Leslie, A., Kilembe, W., … Goulder, P. J. R. (2011). The hypervariable HIV-1 capsid protein residues comprise HLA-driven CD8+ T-cell escape mutations and covarying HLA-independent polymorphisms. *Journal of Virology*, *85*(3), 1384–1390. doi:10.1128/JVI.01879-10

Crawford, H., Prado, J. G., Leslie, A., Hué, S., Honeyborne, I., Reddy, S., … Goulder, P. J. R. (2007). Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *Journal of Virology*, *81*(15), 8346–8351. doi:10.1128/JVI.00465-07

Dahirel, V., Shekhar, K., Pereyra, F., Miura, T., Artyomov, M., Talsania, S., … Chakraborty, A. K. (2011). Coordinate linkage of HIV evolution reveals regions

of immunological vulnerability. *Proceedings of the National Academy of Sciences*, *108*(28), 11530–11535. doi:10.1073/pnas.1105315108

Dalmasso, C., Carpentier, W., Meyer, L., Rouzioux, C., Goujard, C., Chaix, M.-L., … Theodorou, I. (2008). Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. *PloS One*, *3*(12), e3907. doi:10.1371/journal.pone.0003907

Day, C. L., Kaufmann, D. E., Kiepiela, P., Brown, J. A., Moodley, E. S., Reddy, S., … Walker, B. D. (2006). PD-1 expression on HIV-specific T-cells is associated with T-cell exhaustion and disease progression. *Nature*, *443*(7109), 350–354. doi:10.1038/nature05115

Day, C. L., Kiepiela, P., Leslie, A. J., van der Stok, M., Nair, K., Ismail, N., … Klenerman, P. (2007). Proliferative capacity of epitope-specific CD8 T-cell responses is inversely related to viral load in chronic human immunodeficiency virus type 1 infection. *Journal of Virology*, *81*(1), 434–438. doi:10.1128/JVI.01754-06

Dazert, E., Neumann-Haefelin, C., Bressanelli, S., Fitzmaurice, K., Kort, J., Timm, J., … Thimme, R. (2009). Loss of viral fitness and cross-recognition by CD8+ T cells limit HCV escape from a protective HLA-B27-restricted human immune response. *Journal of Clinical Investigation*, *119*(2), 376–386. doi:10.1172/JCI36587

Deeks, S. G., Kitchen, C. M. R., Liu, L., Guo, H., Gascon, R., Narvaez, A. B., … Hecht, F. M. (2004). Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. *Blood*, *104*(4), 942–947. doi:10.1182/blood-2003-09-3333

Deng, K., Pertea, M., Rongvaux, A., Wang, L., Durand, C. M., Ghiaur, G., … Siliciano, R. F. (2015). Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature*, *517*(7534), 381–385. doi:10.1038/nature14053

Derdeyn, C. A., Moore, P. L., & Morris, L. (2014). Development of broadly neutralizing antibodies from autologous neutralizing antibody responses. *Current Opinion in HIV and AIDS*, *9*(3), 210–216. doi:10.1097/COH.0000000000000057

Deymier, M. J., Ende, Z., Fenton-May, A. E., Dilernia, D. A., Kilembe, W., Allen, S. A., … Hunter, E. (2015). Heterosexual Transmission of Subtype C HIV-1 Selects Consensus-Like Variants without Increased Replicative Capacity or Interferon-α Resistance. *PLOS Pathogens*, *11*(9), e1005154. doi:10.1371/journal.ppat.1005154

Dinges, W. L., Richardt, J., Friedrich, D., Jalbert, E., Liu, Y., Stevens, C. E., … Horton, H. (2010). Virus-specific CD8+ T-cell responses better define HIV disease progression than HLA genotype. *Journal of Virology*, *84*(9), 4461–4468. doi:10.1128/JVI.02438-09

Doria-Rose, N. A., Schramm, C. A., Gorman, J., Moore, P. L., Bhiman, J. N., DeKosky, B. J., … Mascola, J. R. (2014). Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature*, *509*(7498), 55–62.

doi:10.1038/nature13036

Draenert, R., Allen, T. M., Liu, Y., Wrin, T., Chappey, C., Verrill, C. L., … Martinez-Picado, J. (2006). Constraints on HIV-1 evolution and immunodominance revealed in monozygotic adult twins infected with the same virus. *The Journal of Experimental Medicine*, *203*(3), 529–539. doi:10.1084/jem.20052116

Draenert, R., Le Gall, S., Pfafferott, K. J., Leslie, A. J., Chetty, P., Brander, C., … Goulder, P. J. R. (2004). Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *The Journal of Experimental Medicine*, *199*(7), 905–915. doi:10.1084/jem.20031982

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*(5), 699–710. doi:10.1371/journal.pbio.0040088

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. doi:10.1093/molbev/mss075

Edwards, B. H., Bansal, A., Sabbaj, S., Bakari, J., Mulligan, M. J., & 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. *Journal of Virology*, *76*(5), 2298–2305.

Elahi, S., Dinges, W. L., Lejarcegui, N., Laing, K. J., Collier, A. C., Koelle, D. M., … Horton, H. (2011). Protective HIV-specific CD8+ T cells evade Treg cell suppression. *Nature Medicine*, *17*(8), 989–995. doi:10.1038/nm.2422

Emu, B., Sinclair, E., Hatano, H., Ferre, A., Shacklett, B., Martin, J. N., … 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. *Journal of Virology*, *82*(11), 5398–5407. doi:10.1128/JVI.02176-07

Engelman, A., & Cherepanov, P. (2012). The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology*, *10*(4), 279–290.

Eriksson, E. M., Liegler, T., Keh, C. E., Karlsson, A. C., Holditch, S. J., Pilcher, C. D., … Hecht, F. M. (2015). Newly exerted T cell pressures on mutated epitopes following transmission help maintain consensus HIV-1 sequences. *Plos One*, *10*(4), e0120787. doi:10.1371/journal.pone.0120787

Erup Larsen, M., Kloverpris, H., Stryhn, A., Koofhethile, C. K., Sims, S., Ndung'U, T., … Nielsen, M. (2011). HLArestrictor-a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides. *Immunogenetics*, *63*(1), 43–55. doi:10.1007/s00251-010-0493-5

Esparza, J. (2013). A brief history of the global effort to develop a preventive HIV vaccine. *Vaccine*, *31*(35), 3502–3518. doi:10.1016/j.vaccine.2013.05.018

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., … Lemey, P. (2014). The early spread and epidemic ignition of HIV-1 in human populations. *Science*, *346*(6205), 56–61.

Fellay, J., Ge, D., Shianna, K. V., Colombo, S., Ledergerber, B., Cirulli, E. T., … Goldstein, D. B. (2009). Common genetic variation and the control of HIV-1 in humans. *PLoS Genetics*, *5*(12), e1000791. doi:10.1371/journal.pgen.1000791

Fellay, J., Shianna, K. V, Ge, D., Colombo, S., Ledergerber, B., Weale, M., … Goldstein, D. B. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science*, *317*(5840), 944–947. doi:10.1126/science.1143767

Fiebig, E. W., Wright, D. J., Rawal, B. D., Garrett, P. E., Schumacher, R. T., Peddada, L., … Busch, M. P. (2003). Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS*, *17*(13), 1871–1879. doi:10.1097/01.aids.0000076308.76477.b8

Friedrich, T. C., Dodds, E. J., Yant, L. J., Vojnov, L., Rudersdorf, R., Cullen, C., … Watkins, D. I. (2004). Reversion of CTL escape-variant immunodeficiency viruses in vivo. *Nature Medicine*, *10*(3), 275–281. doi:10.1038/nm998

Gall, A., Ferns, B., Morris, C., Watson, S., Cotten, M., Robinson, M., … Kellam, P. (2012). Universal amplification, next-generation sequencing and assembly of HIV-1 genomes. *Journal of Clinical Microbiology*, *50*(12), 3838–3844. doi:10.1128/JCM.01516-12

Gall, A., Morris, C., & Kellam, P. (2014). Complete genome sequence of the WHO international standard for HIV-1 RNA determined by deep sequencing. *Genome Announcements*, *2*(1), 10–11. doi:10.1128/genomeA.01254-13.Copyright

Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V. S., … Popovic, M. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*, *220*(4599), 865–867. doi:10.2307/1690358

Gao, X., Bashirova, A., Iversen, A. K. N., Phair, J., Goedert, J. J., Buchbinder, S., … Carrington, M. (2005). AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nature Medicine*, *11*(12), 1290–1292. doi:10.1038/nm1333

Geijtenbeek, T. B. H., Kwon, D. S., Torensma, R., van Vliet, S. J., van Duijnhoven, G. C. F., Middel, J., … van Kooyk, Y. (2015). DC-SIGN, a dendritic cell–specific HIV-1-binding protein that enhances trans-Infection of T cells. *Cell*, *100*(5), 587–597. doi:10.1016/S0092-8674(00)80694-7

Geldmacher, C., Currier, J. R., Herrmann, E., Haule, A., Kuta, E., McCutchan, F., … Hoelscher, M. (2007). CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients. *Journal of Virology*, *81*(5), 2440–2448. doi:10.1128/JVI.01847-06

Gervaix, A., West, D., Leoni, L. M., Richman, D. D., Wong-Staal, F., & Corbeil, J. (1997). A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. *Proceedings of the National Academy of Sciences*, *94*(9), 4653–4658. doi:10.1073/pnas.94.9.4653

Gesprasert, G., Wichukchinda, N., Mori, M., Shiino, T., Auwanit, W., Sriwanthana, B., … Ariyoshi, K. (2010). HLA-associated immune pressure on gag protein in CRF01_AE-Infected individuals and its association with plasma viral load. *PLoS One*, *5*(6), e11179. doi:10.1371/journal.pone.0011179

Gijsbers, E. F., Feenstra, K. A., van Nuenen, A. C., Navis, M., Heringa, J., Schuitemaker, H., & Kootstra, N. (2013). HIV-1 replication fitness of HLA-B*57/58:01 CTL escape variants is restored by the accumulation of compensatory mutations in Gag. *PLoS One*, *8*(12), e81235. doi:10.1371/journal.pone.0081235

Gillespie, G. M. A., Stewart-Jones, G., Rengasamy, J., Beattie, T., Bwayo, J. J., Plummer, F., … Rowland-Jones, S. L. (2006). Strong TCR conservation and altered T cell cross-reactivity characterize a B*57-restricted immune response in HIV-1 infection. *Journal of Immunology*, *177*(6), 3893–3902. doi:10.4049/jimmunol.177.6.3893

Gnanakaran, S., Bhattacharya, T., Daniels, M., Keele, B. F., Hraber, P. T., Lapedes, A. S., … Korber, B. (2011). Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. *PLoS Pathogens*, *7*(9). doi:10.1371/journal.ppat.1002209

Goepfert, P. A., Lumm, W., Farmer, P., Matthews, P., Prendergast, A., Carlson, J. M., … Hunter, E. (2008). Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *The Journal of Experimental Medicine*, *205*(5), 1009–1017. doi:10.1084/jem.20072457

Gonzalez-Galarza, F. F., Christmas, S., Middleton, D., & Jones, A. R. (2011). Allele frequency net: A database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Research*, *39*(Suppl 1), D913–9. doi:10.1093/nar/gkq1128

Goonetilleke, N., Liu, M. K. P., Salazar-Gonzalez, J. F., Ferrari, G., Giorgi, E., Ganusov, V. V, … McMichael, A. J. (2009). The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *The Journal of Experimental Medicine*, *206*(6), 1253–1272. doi:10.1084/jem.20090365

Gottlieb, G. (2004). Dual HIV-1 infection associated with rapid disease progression. *The Lancet*, *363*(9409), 619–622. doi:10.1016/S0140-6736(04)15596-7

Gottlieb, G. S., Nickle, D. C., Jensen, M. A., Wong, K. G., Kaslow, R. A., Shepherd, J. C., … Mullins, J. I. (2007). HIV type 1 superinfection with a dual-tropic virus and rapid progression to AIDS: a case report. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *45*(4), 501–509. doi:10.1086/520024

Goulder, P. J., Altfeld, M. A., Rosenberg, E. S., Nguyen, T., Tang, Y., Eldridge, R. L., … Brander, C. (2001). Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection. *The Journal of Experimental Medicine*, *193*(2), 181–194. doi:10.1084/jem.193.2.181

Goulder, P. J., Brander, C., Tang, Y., Tremblay, C., Colbert, R. A., Addo, M. M., … Walker, B. D. (2001). Evolution and transmission of stable CTL escape

mutations in HIV infection. *Nature*, *412*(6844), 334–338. doi:10.1038/35085576

Goulder, P. J., Bunce, M., Krausa, P., McIntyre, K., Crowley, S., Morgan, B., … McMichael, A. J. (1996). Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS Research and Human Retroviruses*, *12*(18), 1691–1698. doi:10.1089/aid.1996.12.1691

Goulder, P. J., Phillips, R. E., Colbert, R. A., McAdam, S., Ogg, G., Nowak, M. A., … Rowland-Jones, S. (1997). Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature Medicine*, *3*(2), 212–217. doi:10.1038/nm0297-212

Goulder, P. J. R., & Walker, B. D. (2012). HIV and HLA Class I: An evolving relationship. *Immunity*, *37*(3), 426–440. doi:10.1016/j.immuni.2012.09.005

Goulder, P. J. R., & Watkins, D. I. (2004). HIV and SIV CTL escape: implications for vaccine design. *Nature Reviews. Immunology*, *4*(8), 630–640. doi:10.1038/nri1417

Goulder, P. J., Sewell, A. K., Lalloo, D. G., Price, D. A., Whelan, J. A., Evans, J., … McMichael, A. J. (1997). Patterns of immunodominance in HIV-1-specific cytotoxic T lymphocyte responses in two human histocompatibility leukocyte antigens (HLA)-identical siblings with HLA-A*0201 are influenced by epitope mutation. *The Journal of Experimental Medicine*, *185*(8), 1423–1433. doi:10.1084/jem.185.8.1423

Gounder, K., Padayachi, N., Mann, J. K., Radebe, M., Mokgoro, M., van der Stok, M., … Ndung'u, T. (2015). High frequency of transmitted HIV-1 Gag HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection. *Plos One*, *10*(3), e0119886. doi:10.1371/journal.pone.0119886

Graf, E. H., Pace, M. J., Peterson, B. A., Lynch, L. J., Chukwulebe, S. B., Mexas, A. M., … O'Doherty, U. (2013). Gag-positive reservoir cells are susceptible to HIV-specific cytotoxic T lymphocyte mediated clearance. *PLoS One*, *8*(8), 1–13. doi:10.1371/journal.pone.0071879

Gray, G. E., Allen, M., Moodie, Z., Churchyard, G., Bekker, L.-G., Nchabeleng, M., … Kublin, J. G. (2011). Safety and efficacy of the HVTN 503/Phambili study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study. *The Lancet. Infectious Diseases*, *11*(7), 507–515. doi:10.1016/S1473-3099(11)70098-6

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, *303*(5656), 327–332. doi:10.1126/science.1090727

Hammer, S. M., Sobieszczyk, M. E., Janes, H., Karuna, S. T., Mulligan, M. J., Grove, D., … Gilbert, P. B. (2013). Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *New England Journal of Medicine*, *369*(22), 2083–2092. doi:10.1056/NEJMoa1310566

Hancock, G., Yang, H., Yorke, E., Wainwright, E., Bourne, V., Frisbee, A., … Dorrell, L.

(2015). Identification of effective subdominant anti-HIV-1 CD8+ T cells within entire post-infection and post-vaccination immune responses. *PLOS Pathogens*, *11*(2), e1004658. doi:10.1371/journal.ppat.1004658

Hansen, S. G., Ford, J. C., Lewis, M. S., Ventura, A. B., Hughes, C. M., Coyne-Johnson, L., … Picker, L. J. (2011). Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. *Nature*, *473*(7348), 523–527. doi:10.1038/nature10003

Hansen, S. G., Piatak, M., Ventura, A. B., Hughes, C. M., Gilbride, R. M., Ford, J. C., … Picker, L. J. (2013). Immune clearance of highly pathogenic SIV infection. *Nature*, *502*(7469), 100–104. doi:10.1038/nature12519

Hansen, S. G., Sacha, J. B., Hughes, C. M., Ford, J. C., Burwitz, B. J., Scholz, I., … Picker, L. J. (2013). Cytomegalovirus vectors violate CD8+ T cell epitope recognition paradigms. *Science*, *340*(6135). doi:10.1126/science.1237874

Hansen, S. G., Vieville, C., Whizin, N., Coyne-Johnson, L., Siess, D. C., Drummond, D. D., … Picker, L. J. (2009). Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. *Nature Medicine*, *15*(3), 293–299. doi:10.1038/nm.1935

Haynes, B. F. B., Gilbert, P. P. B., McElrath, M. J., Zolla-Pazner, S., Tomaras, G. D., Alam, S. M., … Kim, J. H. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *England Journal of*, *366*(14), 1275–1286. doi:10.1056/NEJMoa1113425

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine*, *18*(3), 1–11. doi:10.1016/j.molmed.2011.12.001

Hemelaar, J., Gouws, E., Ghys, P. D., & Osmanov, S. (2006). Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS*, *20*(16), W13–23. doi:10.1097/01.aids.0000247564.73009.bc

Henn, M. R., Boutwell, C. L., Charlebois, P., Lennon, N. J., Power, K., Macalalad, A. R., … Allen, T. M. (2012). Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathogens*, *8*(3), e1002529. doi:10.1371/journal.ppat.1002529

Herbeck, J. T., Nickle, D. C., Learn, G. H., Gottlieb, G. S., Curlin, M. E., Heath, L., & Mullins, J. I. (2006). Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. *Journal of Virology*, *80*(4), 1637–1644. doi:10.1128/JVI.80.4.1637

Hersperger, A. R., Pereyra, F., Nason, M., Demers, K., Sheth, P., Shin, L. Y., … Betts, M. R. (2010). Perforin expression directly ex vivo by HIV-specific CD8 T-cells is a correlate of HIV elite control. *PLoS Pathogens*, *6*(5), e1000917. doi:10.1371/journal.ppat.1000917

Honeyborne, I., Prendergast, A., Pereyra, F., Leslie, A., Crawford, H., Payne, R., … Goulder, P. J. R. (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*(7), 3667–3672. doi:10.1128/JVI.02689-06

Horton, H., Frank, I., Baydo, R., Jalbert, E., Penn, J., Wilson, S., … McElrath, M. J.

(2006). Preservation of T cell proliferation restricted by protective HLA alleles is critical for immune control of HIV-1 infection. *Journal of Immunology*, *177*(10), 7406–7415.

Huang, K.-H. G., Goedhals, D., Fryer, H., van Vuuren, C., Katzourakis, A., De Oliveira, T., … Frater, J. (2009). Prevalence of HIV type-1 drug-associated mutations in pre-therapy patients in the Free State, South Africa. *Antiviral Therapy*, *14*(7), 975–984. doi:10.3851/IMP1416

Hunt, P. W., Martin, J. N., Sinclair, E., Bredt, B., Hagos, E., Lampiris, H., & Deeks, S. G. (2003). T cell activation is associated with lower CD4+ T cell gains in human immunodeficiency virus-infected patients with sustained viral suppression during antiretroviral therapy. *Journal of Infectious Diseases*, *187*(10), 1534–1543. doi:10.1086/374786

Iglesias, M. C., Almeida, J. R., Fastenackels, S., Van Bockel, D. J., Hashimoto, M., Venturi, V., … Appay, V. (2011). Escape from highly effective public CD8+ T-cell clonotypes by HIV. *Blood*, *118*(8), 2138–2149. doi:10.1182/blood-2011-01-328781

Joseph, S. B., Arrildt, K. T., Swanstrom, A. E., Schnell, G., Lee, B., Hoxie, J., & Swanstrom, R. (2014). Quantification of entry phenotypes of macrophage-tropic HIV-1 across a wide range of CD4 densities. *Journal of Virology*, *88*(4), 1858–69. doi:10.1128/JVI.02477-13

Joseph, S. B., Swanstrom, R., Kashuba, A. D. M., & Cohen, M. S. (2015). Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nature Reviews Microbiology*, *13*(7), 414–425. doi:10.1038/nrmicro3471

Jost, S., Bernard, M., Kaiser, L., Yerly, S., Hirschel, B., Samri, A., … T. (2002). A patient with HIV-1 super-infection. *New England Journal of Medicine*, *347*(10), 731–736.

Juarez-Molina, C. I., Payne, R., Soto-Nava, M., Avila-Rios, S., Valenzuela-Ponce, H., Adland, E., … Goulder, P. (2014). Impact of HLA selection pressure on HIV fitness at a population level in Mexico and Barbados. *Journal of Virology*, *88*(18), 10392–10398. doi:10.1128/JVI.01162-14

Jung, M., Leye, N., Vidal, N., Fargette, D., Diop, H., Toure Kane, C., … Peeters, M. (2012). The origin and evolutionary history of HIV-1 subtype C in Senegal. *PloS One*, *7*(3), e33579. doi:10.1371/journal.pone.0033579

Kantakamalakul, W., de Souza, M., Bejrachandra, S., Ampol, S., Cox, J., & Sutthent, R. (2006). Identification of a novel HIV type 1 CRF01_AE cytotoxic T lymphocyte (CTL) epitope restricted by an HLA-Cw0602 allele and a novel HLA-A0206/peptide restriction. *AIDS Research and Human Retroviruses*, *22*(12), 1271–1282. doi:10.1089/aid.2006.22.1271

Karakas, A., Brumme, Z. L., & Poon, A. F. Y. (2015). Global database-driven assessment of HIV-1 adaptation to the immune repertoires of human populations. *Journal of Virology*, *89*(20), 10693–10695. doi:10.1128/JVI.01355-15

Kaslow, R. A., Carrington, M., Apple, R., Park, L., Muñoz, A., Saah, A. J., … Mann, D. L. (1996). Influence of combinations of human major histocompatibility complex

genes on the course of HIV-1 infection. *Nature Medicine*, *2*(4), 405–411. doi:10.1038/nm0496-405

Kaufmann, D. E., Kavanagh, D. G., Pereyra, F., Zaunders, J. J., Mackey, E. W., Miura, T., … Walker, B. D. (2007). Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. *Nature Immunology*, *8*(11), 1246–1254. doi:10.1038/ni1515

Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., Addo, M., … Goulder, P. (2009). Adaptation of HIV-1 to human leukocyte antigen class I. *Nature*, *458*(7238), 641–645. doi:10.1038/nature07746

Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., … Shaw, G. M. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences*, *105*(21), 7552–7557. doi:10.1073/pnas.0802203105

Kiepiela, P., Leslie, A. J., Honeyborne, I., Ramduth, D., Thobakgale, C., Chetty, S., … Goulder, P. J. R. (2004). Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature*, *432*(7018), 769–775. doi:10.1038/nature03113

Kiepiela, P., Ngumbela, K., Thobakgale, C., Ramduth, D., Honeyborne, I., Moodley, E., … Goulder, P. (2007). CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature Medicine*, *13*(1), 46–53. doi:10.1038/nm1520

Kim, J. H., Excler, J.-L., & Michael, N. L. (2015). Lessons from the RV144 Thai phase III HIV-1 vaccine trial and the search for correlates of protection. *Annual Review of Medicine*, *66*, 423–437. doi:10.1146/annurev-med-052912-123749

Kinloch, N. N., MacMillan, D. R., Le, A. Q., Cotton, L. A., Bangsberg, D. R., Buchbinder, S., … Brumme, Z. L. (2015). Population-level Immune-mediated adaptation in HIV-1 polymerase during the North American epidemic. *Journal of Virology*. doi:10.1128/JVI.02353-15

Kløverpris, H. N., Adland, E., Koyanagi, M., Stryhn, A., Harndahl, M., Matthews, P. C., … Goulder, P. (2013). HIV subtype influences HLA-B*07:02-associated HIV disease outcome. *AIDS Research and Human Retroviruses*, *674*(5), 1–28. doi:10.1089/AID.2013.0197

Kloverpris, H. N., McGregor, R., McLaren, J. E., Ladell, K., Harndahl, M., Stryhn, A., … Goulder, P. J. (2015). CD8+ TCR bias and immunodominance in HIV-1 infection. *The Journal of Immunology*, *194*, 5329–5345. doi:10.4049/jimmunol.1400854

Kloverpris, H. N., Stryhn, A., Harndahl, M., van der Stok, M., Payne, R. P., Matthews, P. C., … Goulder, P. (2012). HLA-B*57 micropolymorphism shapes HLA allele-specific epitope ommunogenicity, selection pressure, and HIV immune control. *Journal of Virology*, *86*(2), 919–929. doi:10.1128/JVI.06150-11

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., & Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *British Medical Bulletin*, *58*, 19–42. doi:10.1093/bmb/58.1.19

Kosmrlj, A., Read, E. L., Qi, Y., Allen, T. M., Altfeld, M., Deeks, S. G., … Chakraborty, A. K. (2010). Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature*, *465*(7296), 350–354. doi:10.1038/nature08997

Koup, R., Safrit, J., & Cao, Y. (1994). Temporal association of the cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of Virology*, *68*(7), 4650–4655.

Kouyos, R. D., Rauch, A., Boni, J., Yerly, S., Shah, C., Aubert, V., … Gunthard, H. F. (2014). Clustering of HCV coinfections on HIV phylogeny indicates domestic and sexual transmission of HCV. *International Journal of Epidemiology*, *43*(3), 887–896. doi:10.1093/ije/dyt276

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), R12. doi:10.1186/gb-2004-5-2-r12

Ladell, K., Hashimoto, M., Iglesias, M. C., Wilmann, P. G., McLaren, J. E., Gras, S., … Appay, V. (2013). A molecular basis for the control of preimmune escape variants by HIV-specific CD8+ T cells. *Immunity*, *38*(3), 425–436. doi:10.1016/j.immuni.2012.11.021

Lamine, A., Caumont-Sarcos, A., Chaix, M.-L., Saez-Cirion, A., Rouzioux, C., Delfraissy, J.-F., … Lambotte, O. (2007). Replication-competent HIV strains infect HIV controllers despite undetectable viremia (ANRS EP36 study). *AIDS*, *21*(8), 1043–1045. doi:10.1097/QAD.0b013e3280d5a7ac

Lazaryan, A., Song, W., Lobashevsky, E., Tang, J., Shrestha, S., Zhang, K., … Kaslow, R. A. (2010). Human leukocyte antigen class I supertypes and HIV-1 control in African Americans. *Journal of Virology*, *84*(5), 2610–2617. doi:10.1128/JVI.01962-09

Lederman, M. M., Alter, G., Daskalakis, D. C., Rodriguez, B., Sieg, S. F., Hardy, G., … Goedert, J. J. (2010). Determinants of protection among HIV-exposed seronegative persons: an overview. *The Journal of Infectious Diseases*, *202*(Suppl 3), S333–S338. doi:10.1086/655967

Ledgerwood, J. E., Coates, E. E., Yamshchikov, G., Saunders, J. G., Holman, L., Enama, M. E., … Graham, B. S. (2015). Safety, pharmacokinetics, and neutralization of the broadly neutralizing HIV-1 human monoclonal antibody VRC01 in healthy adults. *Clinical and Experimental Immunology*, *182*(3), 289–301. doi:10.1111/cei.12692

Lee, W. P., Stromberg, M. P., Ward, A., Stewart, C., Garrison, E. P., & Marth, G. T. (2014). MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS ONE*, *9*(3), e90581. doi:10.1371/journal.pone.0090581

Leisner, C., Loeth, N., Lamberth, K., Justesen, S., Sylvester-Hvid, C., Schmidt, E. G., … Stryhn, A. (2008). One-pot, mix-and-read peptide-MHC tetramers. *PLoS ONE*, *3*(2), e1678. doi:10.1371/journal.pone.0001678

Lemey, P., Kosakovsky Pond, S. L., Drummond, A. J., Pybus, O. G., Shapiro, B.,

Barroso, H., … Rambaut, A. (2007). Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Computational Biology*, *3*(2), 282–292. doi:10.1371/journal.pcbi.0030029

Leslie, A. J., Pfafferott, K. J., Chetty, P., Draenert, R., Addo, M. M., Feeney, M., … Goulder, P. J. R. (2004). HIV evolution: CTL escape mutation and reversion after transmission. *Nature Medicine*, *10*(3), 282–289. doi:10.1038/nm992

Leslie, A., Kavanagh, D., Honeyborne, I., Pfafferott, K., Edwards, C., Pillay, T., … Goulder, P. (2005). Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *The Journal of Experimental Medicine*, *201*(6), 891–902. doi:10.1084/jem.20041455

Leslie, A., Matthews, P. C., Listgarten, J., Carlson, J. M., Kadie, C., Ndung'u, T., … Goulder, P. J. R. (2010). Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. *Journal of Virology*, *84*(19), 9879–9888. doi:10.1128/JVI.00320-10

Letourneau, S., Im, E.-J., Mashishi, T., Brereton, C., Bridgeman, A., Yang, H., … Hanke, T. (2007). Design and pre-clinical evaluation of a universal HIV-1 vaccine. *PloS One*, *2*(10), e984. doi:10.1371/journal.pone.0000984

Ley, J. A. (2007). *HIV and the Pathogenesis of AIDS* (third edit.). ASM Press.

Limou, S., Le Clerc, S., Coulonges, C., Carpentier, W., Dina, C., Delaneau, O., … Zagury, J.-F. (2009). Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *The Journal of Infectious Diseases*, *199*(3), 419–426. doi:10.1086/596067

Lindenbach, B. D. (2009). Measuring HCV infectivity produced in cell culture and in vivo. *Methods in Molecular Biology*, *510*, 329–336. doi:10.1007/978-1-59745-394-3

Liu, D., Zuo, T., Hora, B., Song, H., Kong, W., Yu, X., … Gao, F. (2014). Preexisting compensatory amino acids compromise fitness costs of a HIV-1 T cell escape mutation. *Retrovirology*, *11*(1), 101. doi:10.1186/s12977-014-0101-0

Llano, A., Williams, A., Olvera, A., Silvia-Arrieta, S., & Brander, C. (2013). Best-Characterized HIV-1 CTL Epitopes: The 2013 Update. In *HIV Molecular Immunology. Los Alamos National Laboratory, Theoretical Biology and Biophysics. Los Alamos, New Mexico* (pp. 3–25).

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., … Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*, *73*(1), 152–160.

Lund, O., Nielsen, M., Kesmir, C., Petersen, A. G., Lundegaard, C., Worning, P., … Brunak, S. (2004). Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*, *55*(12), 797–810. doi:10.1007/s00251-004-0647-4

Madden, D. R., Gorga, J. C., Strominger, J. L., & Wiley, D. C. (1991). The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation.

*Nature*, *353*(6342), 321–325. doi:10.1038/353321a0

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, *9*, 387–402. doi:10.1146/annurev.genom.9.081307.164359

Mariani, R., Kirchhoff, F., Greenough, T. C., Sullivan, J. L., Desrosiers, R. C., & Skowronski, J. (1996). High frequency of defective nef alleles in a long-term survivor with nonprogressive human immunodeficiency virus type 1 infection. *Journal of Virology*, *70*(11), 7752–7764.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(5). doi:10.1093/ve/vev003

Martin, M. P., Gao, X., Lee, J.-H., Nelson, G. W., Detels, R., Goedert, J. J., … Carrington, M. (2002). Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature Genetics*, *31*(4), 429–434. doi:10.1038/ng934

Martin, M. P., Qi, Y., Gao, X., Yamada, E., Martin, J. N., Pereyra, F., … Carrington, M. (2007). Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature Genetics*, *39*(6), 733–740. doi:10.1038/ng2035

Martinez-picado, J., Prado, J. G., Fry, E. E., Pfafferott, K., Leslie, A., Chetty, S., … Kiepiela, P. (2006). Fitness cost of escape mutations in p24 gag in association with control of human immunodeficiency virus type 1. *Journal of Virology*, *80*(7), 3617–3623. doi:10.1128/JVI.80.7.3617

Masemola, A., Mashishi, T., Khoury, G., Mohube, P., Mokgotho, P., Vardas, E., … Gray, C. M. (2004). Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by CD8+ T cells: correlation with viral load. *Journal of Virology*, *78*(7), 3233–3243.

Matano, T., Shibata, R., Siemon, C., Connors, M., Lane, H. C., & 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*(1), 164–169.

Mattapallil, J. J., Douek, D. C., Hill, B., Nishimura, Y., Martin, M., & Roederer, M. (2005). Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature*, *434*(7037), 1093–1097. doi:10.1038/nature03501

Matthews, P. C., Adland, E., Listgarten, J., Leslie, A., Mkhwanazi, N., Carlson, J. M., … Goulder, P. J. R. (2011). HLA-A*7401-mediated control of HIV viremia is independent of its linkage disequilibrium with HLA-B*5703. *Journal of Immunology*, *186*(10), 5675–5686. doi:10.4049/jimmunol.1003711

Matthews, P. C., Koyanagi, M., Kloverpris, H. N., Harndahl, M., Stryhn, A., Akahoshi, T., … Goulder, P. J. R. (2012). Differential clade-specific HLA-B*3501 association with HIV-1 disease outcome is linked to immunogenicity of a single gag epitope. *Journal of Virology*, *86*(23), 12643–12654. doi:10.1128/JVI.01381-12

Matthews, P. C., Leslie, A. J., Katzourakis, A., Crawford, H., Payne, R., Prendergast, A.,

… Goulder, P. J. R. (2009). HLA footprints on human immunodeficiency virus type 1 are associated with interclade polymorphisms and intraclade phylogenetic clustering. *Journal of Virology*, *83*(9), 4605–4615. doi:10.1128/JVI.02017-08

Matthews, P. C., Listgarten, J., Carlson, J. M., Payne, R., Huang, K.-H. G., Frater, J., … Goulder, P. J. R. (2012). Co-operative additive effects between HLA alleles in control of HIV-1. *PLoS One*, *7*(10), e47799. doi:10.1371/journal.pone.0047799

Matthews, P. C., Prendergast, A., Leslie, A., Crawford, H., Payne, R., Rousseau, C., … Goulder, P. J. R. (2008). Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. *Journal of Virology*, *82*(17), 8548–8559. doi:10.1128/JVI.00580-08

McKinnon, L. R., Capina, R., Peters, H., Mendoza, M., Kimani, J., Wachihi, C., … Plummer, F. (2009). Clade-specific evolution mediated by HLA-B*57/5801 in human immunodeficiency virus type 1 clade A1 p24. *Journal of Virology*, *83*(23), 12636–12642. doi:10.1128/JVI.01236-09

McLaren, P. J., & Carrington, M. (2015). The impact of host genetic variation on infection with HIV-1. *Nature Immunology*, *16*(6), 577–583. doi:10.1038/ni.3147

McLaren, P. J., Coulonges, C., Bartha, I., Lenz, T. L., Deutsch, A. J., Bashirova, A., … Fellay, J. (2015). Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. *Proceedings of the National Academy of Sciences*, *112*(47), 14658–63. doi:10.1073/pnas.1514867112

McMichael, A. J., Borrow, P., Tomaras, G. D., Goonetilleke, N., & Haynes, B. F. (2010). The immune response during acute HIV-1 infection: clues for vaccine development. *Nature Reviews. Immunology*, *10*(1), 11–23. doi:10.1038/nri2674

McNamara, L. A., Ganesh, J. A., & Collins, K. L. (2012). Latent HIV-1 infection occurs in multiple subsets of hematopoietic progenitor cells and is reversed by NF-kappaB activation. *Journal of Virology*, *86*(17), 9337–9350. doi:10.1128/JVI.00895-12

Mellors, J. W., Rinaldo, C. R., Gupta, P., White, R. M., Todd, J., & Kingsley, L. (1996). Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*, *272*(5265), 1167–1170. doi:10.1126/science.272.5265.1167

Mendoza, D., Royce, C., Ruff, L. E., Ambrozak, D. R., Quigley, M. F., Dang, T., … Connors, M. (2012). HLA B*5701-positive long-term nonprogressors/elite controllers are not distinguished from progressors by the clonal composition of HIV-specific CD8+ T cells. *Journal of Virology*, *86*(7), 4014–8. doi:10.1128/JVI.06982-11

Mens, H., Kearney, M., Wiegand, A., Shao, W., Schønning, K., Gerstoft, J., … Coffin, J. M. (2010). HIV-1 continues to replicate and evolve in patients with natural control of HIV infection. *Journal of Virology*, *84*(24), 12971–12981. doi:10.1128/JVI.00387-10

Michael, N. L., Chang, G., d'Arcy, L. A., Ehrenberg, P. K., Mariani, R., Busch, M. P., …

Schwartz, D. H. (1995). Defective accessory genes in a human immunodeficiency virus type 1-infected long-term survivor lacking recoverable virus. *Journal of Virology*, *69*(7), 4228–4236.

Migueles, S. A., Laborico, A. C., Shupert, W. L., Sabbaghian, M. S., Rabin, R., Hallahan, C. W., … Connors, M. (2002). HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. *Nature Immunology*, *3*(11), 1061–1068. doi:10.1038/ni845

Migueles, S. A., Weeks, K. A., Nou, E., Berkley, A. M., Rood, J. E., Osborne, C. M., … Connors, M. (2009). Defective human immunodeficiency virus-specific CD8+ T-cell polyfunctionality, proliferation, and cytotoxicity are not restored by antiretroviral therapy. *Journal of Virology*, *83*(22), 11876–11889. doi:10.1128/JVI.01153-09

Migueles, S., & Connors, M. (2010). Long-term nonprogressive disease among untreated HIV-infected individuals: clinical implications of understanding immune control of HIV. *Journal of the American Medical Association*, *304*(2), 194–201. doi:10.1001/jama.2010.925

Migueles, S., & Connors, M. (2015). Success and failure of the cellular immune response against HIV-1. *Nature Immunology*, *16*(6), 563–570. doi:10.1038/ni.3161

Migueles, S., Mendoza, D., Zimmerman, M. G., Martins, K. M., Toulmin, S. A., Kelly, E. P., … Connors, M. (2015). CD8+ T-cell cytotoxic capacity associated with human immunodeficiency virus-1 control can be mediated through various epitopes and human leukocyte antigen types. *EBioMedicine*, *2*(1), 46–58. doi:10.1016/j.ebiom.2014.12.009

Migueles, S., Osborne, C. M., Royce, C., Compton, A., Joshi, R. P., Weeks, K. A., … Connors, M. (2008). Lytic granule loading of CD8+ T cells Is required for HIV-infected cell elimination associated with immune control. *Immunity*, *29*(6), 1009–1021. doi:10.1016/j.immuni.2008.10.010

Migueles, S., Sabbaghian, M. S., Shupert, W. L., Bettinotti, M. P., Marincola, F. M., Martino, L., … Connors, M. (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*, *97*(6), 2709–2714. doi:10.1073/pnas.050567397

Minin, V. N., Bloomquist, E. W., & Suchard, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, *25*(7), 1459–1471. doi:10.1093/molbev/msn090

Miura, T., Brockman, M. A., Brumme, Z. L., Brumme, C. J., Pereyra, F., Trocha, A., … Walker, B. D. (2009). HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. *Journal of Virology*, *83*(1), 140–149. doi:10.1128/JVI.01471-08

Miura, T., Brockman, M. A., Schneidewind, A., Lobritz, M., Pereyra, F., Rathod, A., … Walker, B. D. (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*(6), 2743–2755. doi:10.1128/JVI.02265-08

Miura, T., Brumme, Z. L., Brockman, M. A., Rosato, P., Sela, J., Brumme, C. J., … Walker, B. D. (2010). Impaired replication capacity of acute/early viruses in persons who become HIV controllers. *Journal of Virology*, *84*(15), 7581–7591. doi:10.1128/JVI.00286-10

Mori, M., Sriwanthana, B., Wichukchinda, N., Boonthimat, C., Tsuchiya, N., Miura, T., … Sawanpanyalert, P. (2011). Unique CRF01_AE Gag CTL epitopes associated with lower HIV-viral load and delayed disease progression in a cohort of HIV-infected Thais. *PLoS One*, *6*(8), e22680. doi:10.1371/journal.pone.0022680

Mori, M., Wichukchinda, N., Miyahara, R., Rojanawiwat, A., Pathipvanich, P., Maekawa, T., … Sawanpanyalert, P. (2014). HLA-B*35: 05 is a protective allele with a unique structure among HIV-1 CRF01_AE-infected Thais, in whom the B*57 frequency is low. *AIDS*, *28*(7), 959–967. doi:10.1097/QAD.0000000000000206

Mothe, B., Hu, X., Llano, A., Rosati, M., Olvera, A., Kulkarni, V., … Brander, C. (2015). A human immune data-informed vaccine concept elicits strong and broad T-cell specificities associated with HIV-1 control in mice and macaques. *Journal of Translational Medicine*, *13*(1), 60. doi:10.1186/s12967-015-0392-5

Mothe, B., Llano, A., Ibarrondo, J., Daniels, M., Miranda, C., Zamarreño, J., … Brander, C. (2011). Definition of the viral targets of protective HIV-1-specific T cell responses. *Journal of Translational Medicine*, *9*(1), 208. doi:10.1186/1479-5876-9-208

Mothe, B., Llano, A., Ibarrondo, J., Zamarreño, J., Schiaulini, M., Miranda, C., … Brander, C. (2012). CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. *PLoS One*, *7*(1). doi:10.1371/journal.pone.0029717

Myers, T., Allman, D., Xu, K., Remis, R. S., Aguinaldo, J., Burchell, A., … Swantee, C. (2009). The prevalence and correlates of hepatitis C virus (HCV) infection and HCV-HIV co-infection in a community sample of gay and bisexual men. *International Journal of Infectious Diseases*, *13*(6), 730–739. doi:10.1016/j.ijid.2008.11.015

Nahmias, A. J., Weiss, J., Yao, X., Lee, F., Kodsi, R., Schanfield, M., … Motulsky, A. (1986). Evidence for human infection with an HTLV III/LAV-like virus in Central Africa, 1959. *The Lancet*. ENGLAND.

Ndhlovu, Z. M., Kamya, P., Mewalal, N., Kloverpris, H. N., Nkosi, T., Pretorius, K., … Walker, B. D. (2015). Magnitude and kinetics of CD8(+) T cell activation during hyperacute HIV infection impact viral set point. *Immunity*, *43*(3), 591–604. doi:10.1016/j.immuni.2015.08.012

Neumann-Haefelin, C., McKiernan, S., Ward, S., Viazov, S., Spangenberg, H. C., Killinger, T., … Thimme, R. (2006). Dominant influence of an HLA-B27 restricted CD8+ T cell response in mediating HCV clearance and evolution. *Hepatology*, *43*(3), 563–572. doi:10.1002/hep.21049

Neumann-Haefelin, C., Oniangue-Ndza, C., Kuntzen, T., Schmidt, J., Nitschke, K., Sidney, J., … Allen, T. M. (2011). Human leukocyte antigen B27 selects for rare escape mutations that significantly impair hepatitis C virus replication and require compensatory mutations. *Hepatology*, *54*(4), 1157–1166. doi:10.1002/hep.24541

Nixon, D. F., Townsend, A. R., Elvin, J. G., Rizza, C. R., Gallwey, J., & McMichael, A. J. (1988). HIV-1 gag-specific cytotoxic T lymphocytes defined with recombinant vaccinia virus and synthetic peptides. *Nature*, *336*(6198), 484–487. doi:10.1038/336484a0

Nomura, S., Hosoya, N., Brumme, Z. L., Brockman, M., Kikuchi, T., Koga, M., … Miura, T. (2013). Significant reductions in gag-protease-mediated HIV-1 replication capacity during the course of the epidemic in Japan. *Journal of Virology*, *87*(3), 1465–1476. doi:10.1128/JVI.02122-12

Norstrom, M. M., Buggert, M., Tauriainen, J., Hartogensis, W., Prosperi, M. C., Wallet, M. A., … Karlsson, A. C. (2012). Combination of immune and viral factors distinguishes low-risk versus high-risk HIV-1 disease progression in HLA-B*5701 subjects. *Journal of Virology*, *86*(18), 9802–9816. doi:10.1128/JVI.01165-12

Norström, M. M., Veras, N. M., Huang, W., Proper, M. C. F., Cook, J., Hartogensis, W., … Salemi, M. (2014). Baseline CD4+ T cell count correlates with HIV-1 synonymous rate in HLA-B*5701 subjects with different risk of disease progression. *PLoS Computational Biology*, *10*(9), e1003830. doi:10.1371/journal.pcbi.1003830

Okoye, A. A., & Picker, L. J. (2013). CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunological Reviews*, *254*(1), 54–64. doi:10.1111/imr.12066

Ondondo, B., Murakoshi, H., Clutton, G., Abdul-Jawad, S., Wee, E. G.-T., Gatanaga, H., … Hanke, T. (2016). Novel conserved-region T-cell mosaic vaccine with high global HIV-1 coverage is recognized by protective responses in untreated infection. *Molecular Therapy : The Journal of the American Society of Gene Therapy*. doi:10.1038/mt.2016.3

Parrish, N. F., Gao, F., Li, H., Giorgi, E. E., Barbian, H. J., Parrish, E. H., … Hahn, B. H. (2013). Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences*, *110*(17), 6626–33. doi:10.1073/pnas.1304288110

Payne, R., Muenchhoff, M., Mann, J., Roberts, H. E., Matthews, P., Adland, E., … Goulder, P. J. R. (2014). Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proceedings of the National Academy of Sciences*, *111*(50), E5393–E5400. doi:10.1073/pnas.1413339111

Payne, R. P., Branch, S., Kløverpris, H., Matthews, P. C., Koofhethile, C. K., Strong, T., … Goulder, P. J. R. (2014). Differential escape patterns within the dominant HLA-B*57:03-restricted HIV Gag epitope reflect distinct clade-specific functional constraints. *Journal of Virology*, *88*(9), 4668–78. doi:10.1128/JVI.03303-13

Payne, R. P., Kløverpris, H., Sacha, J. B., Brumme, Z., Brumme, C., Buus, S., … Goulder, P. J. R. (2010). Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. *Journal of Virology*, *84*(20), 10543–10557. doi:10.1128/JVI.00793-10

Pereyra, F., Addo, M. M., Kaufmann, D. E., Liu, Y., Miura, T., Rathod, A., … Walker, B. D. (2008). Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *The Journal of Infectious Diseases*, *197*(4), 563–571. doi:10.1086/526786

Pereyra, F., Jia, X., McLaren, P. J., Telenti, A., de Bakker, P. I. W., Walker, B. D., … Zhao, M. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*, *330*(6010), 1551–1557. doi:10.1126/science.1195271.The

Peyerl, F. W., Bazick, H. S., Newberg, M. H., Barouch, D. H., Sodroski, J., & Letvin, N. L. (2004). Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. *Journal of Virology*, *78*(24), 13901–13910. doi:10.1128/JVI.78.24.13901

Phillips, R. E., Rowland-Jones, S., Nixon, D. F., Gotch, F. M., Edwards, J. P., Ogunlesi, A. O., … Rizza, C. R. (1991). Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature*, *354*(6353), 453–459. doi:10.1038/354453a0

Pineda-Peña, A. C., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., … Vandamme, A. M. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infection, Genetics and Evolution*, *19*(126), 337–348. doi:10.1016/j.meegid.2013.04.032

Ping, L.-H., Joseph, S. B., Anderson, J. A., Abrahams, M.-R., Salazar-Gonzalez, J. F., Kincer, L. P., … Swanstrom, R. (2013). Comparison of viral Env proteins from acute and chronic infections with subtype C human immunodeficiency virus type 1 identifies differences in glycosylation and CCR5 utilization and suggests a new strategy for immunogen design. *Journal of Virology*, *87*(13), 7218–33. doi:10.1128/JVI.03577-12

Pitisuttithum, P., Gilbert, P., Gurwith, M., Heyward, W., Martin, M., van Griensven, F., … Choopanya, K. (2006). Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *The Journal of Infectious Diseases*, *194*(12), 1661–1671. doi:10.1086/508748

Plantier, J.-C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., … Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine*, *15*(8), 871–872.

Poon, A. F. Y., McGovern, R. A., Mo, T., Knapp, D. J. H. F., Brenner, B., Routy, J.-P., … Harrigan, P. R. (2011). Dates of HIV infection can be estimated for seroprevalent patients by coalescent analysis of serial next-generation sequencing data. *AIDS*, *25*(16), 2019–2026. doi:10.1097/QAD.0b013e32834b643c

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, *25*(7), 1253–1256. doi:10.1093/molbev/msn083

Prado, J. G., Prendergast, A., Thobakgale, C., Molina, C., Tudor-Williams, G., Ndung'u, T., … Goulder, P. (2010). Replicative capacity of human immunodeficiency virus type 1 transmitted from mother to child is associated with pediatric disease progression rate. *Journal of Virology*, *84*(1), 492–502. doi:10.1128/JVI.01743-09

Prentice, H. A., Tomaras, G. D., Geraghty, D. E., Apps, R., Fong, Y., Ehrenberg, P. K., … Thomas, R. (2015). HLA class II genes modulate vaccine-induced antibody responses to affect HIV-1 acquisition. *Science Translational Medicine*, *7*(296), 296ra112. doi:10.1126/scitranslmed.aab4005

Price, D. A., Asher, T. E., Wilson, N. A., Nason, M. C., Brenchley, J. M., Metzler, I. S., … Douek, D. C. (2009). Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. *The Journal of Experimental Medicine*, *206*(4), 923–936. doi:10.1084/jem.20081127

Prince, J. L., Claiborne, D. T., Carlson, J. M., Schaefer, M., Yu, T., Lahki, S., … Hunter, E. (2012). Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. *PLoS Pathogens*, *8*(11), e1003041. doi:10.1371/journal.ppat.1003041

Quinn, T. C., Wawer, M. J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., … Rakai PRoject Study Group. (2000). Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine*, *342*(13), 921–929.

Redd, A. D., Quinn, T. C., & Tobian, A. A. (2013). Frequency and implications of HIV superinfection. *The Lancet Infectious Diseases*, *13*(7), 622–628. doi:10.1016/S1473-3099(13)70066-5

Reed, L. J., & Muench, H. (1938). A simple method of estimating fifty percent endpionts. *American Journal of Epidemiology*, *27*(3), 493–497.

Rehermann, B. (2013). Pathogenesis of chronic viral hepatitis: differential roles of T cells and NK cells. *Nature Medicine*, *19*(7), 859–68. doi:10.1038/nm.3251

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., … Kim, J. H. (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *New England Journal of Medicine*, *361*(23), 2209–2220. doi:10.1056/NEJMoa0908492

Richman, D. D., Wrin, T., Little, S. J., & Petropoulos, C. J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proceedings of the National Academy of Sciences*, *100*(7), 4144–4149. doi:10.1073/pnas.0630530100

Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P., & Marsh, S. G. E. (2015). The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research*, *43*(Database issue), D423–31. doi:10.1093/nar/gku1161

Rockstroh, J., Guaraldi, G., & Deray, G. (2010). HIV and the body: a review of multidisciplinary management. *HIV Medicine*, *11*(Suppl 2), 1–8.

doi:10.1111/j.1468-1293.2010.00859.x

Ronen, K., McCoy, C. O., Matsen, F. A., Boyd, D. F., Emery, S., Odem-Davis, K., … Overbaugh, J. (2013). HIV-1 Superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. *PLoS Pathogens*, *9*(8), e1003593. doi:10.1371/journal.ppat.1003593

Sacha, J. B., Chung, C., Rakasz, E. G., Spencer, S. P., Jonas, A. K., Bean, A. T., … Watkins, D. I. (2007). Gag-specific CD8+ T lymphocytes recognize infected cells before AIDS-virus integration and viral protein expression. *Journal of Immunology*, *178*(5), 2746–2754. doi:JVI.00926-07 [pii]\n10.1128/JVI.00926-07

Sáez-Cirión, A., Sinet, M., Shin, S. Y., Urrutia, A., Versmisse, P., Lacabaratz, C., … Pancino, G. (2009). Heterogeneity in HIV suppression by CD8 T cells from HIV controllers: association with Gag-specific CD8 T cell responses. *Journal of Immunology*, *182*(12), 7828–7837. doi:10.4049/jimmunol.0803928

Sagar, M., Laeyendecker, O., Lee, S., Gamiel, J., Wawer, M. J., Gray, R. H., … Quinn, T. C. (2009). Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *The Journal of Infectious Diseases*, *199*(4), 580–589. doi:10.1086/596557

Salemi, M. (2013). The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: A phylogenetic perspective. *Infectious Disease Reports*, *5*(Suppl 1), 8–14. doi:10.4081/idr.2013.s1.e3

Santangelo, P. J., Rogers, K. A., Zurla, C., Blanchard, E. L., Gumber, S., Strait, K., … Villinger, F. (2015). Whole-body immunoPET reveals active SIV dynamics in viremic and antiretroviral therapy-treated macaques. *Nature Methods*, *12*(5), 427–432. doi:10.1038/nmeth.3320

Scheid, J. F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T. Y. K., … Nussenzweig, M. C. (2011). Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*, *333*(6049), 1633–1637. doi:10.1126/science.1207227

Schellens, I. M. M., Navis, M., van Deutekom, H. W. M., Boeser-Nunnink, B., Berkhout, B., Kootstra, N., … Borghans, J. A. M. (2011). Loss of HIV-1-derived cytotoxic T lymphocyte epitopes restricted by protective HLA-B alleles during the HIV-1 epidemic. *AIDS*, *25*(14), 1691–1700. doi:10.1097/QAD.0b013e32834981b3

Schmitz, J. E., Kuroda, M. J., Santra, S., Sasseville, V. G., Simon, M. A., Lifton, M. A., … Reimann, K. A. (1999). Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science*, *283*(5403), 857–860.

Schneidewind, A., Brockman, M. A., Sidney, J., Wang, Y. E., Chen, H., Suscovich, T. J., … Allen, T. M. (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*(11), 5594–5605. doi:10.1128/JVI.02356-07

Schneidewind, A., Brockman, M. A., Yang, R., Adam, R. I., Li, B., Le Gall, S., … Allen, T. M. (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. *Journal of Virology*, *81*(22), 12382–12393. doi:10.1128/JVI.01543-07

Schneidewind, A., Brumme, Z. L., Brumme, C. J., Power, K. A., Reyor, L. L., O'Sullivan, K., … Allen, T. M. (2009). Transmission and long-term stability of compensated CD8 escape mutations. *Journal of Virology*, *83*(8), 3993–3997. doi:10.1128/JVI.01108-08

Shan, L., Deng, K., Shroff, N. S., Durand, C. M., Rabi, S. A., Yang, H. C., … Siliciano, R. F. (2012). Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. *Immunity*, *36*(3), 491–501. doi:10.1016/j.immuni.2012.01.014

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technology Journal*, *27*(379-423), 379:423, 623–656. doi:10.1145/584091.584093

Sharp, P. M., & Hahn, B. H. (2010). The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society*, *365*(1552), 2487–2494. doi:10.1098/rstb.2010.0031

Shriner, D., Liu, Y., Nickle, D. C., & Mullins, J. I. (2006). Evolution of intrahost HIV-1 genetic diversity during chronic infection. *Evolution*, *60*(6), 1165–1176. doi:10.1554/05-473.1

Silvestri, G., Sodora, D. L., Koup, R. A., Paiardini, M., O'Neil, S. P., McClure, H. M., … Feinberg, M. B. (2003). Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. *Immunity*, *18*(3), 441–452.

Simon, V., Bloch, N., & Landau, N. R. (2015). Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. *Nature Immunology*, *16*(6), 546–553. doi:10.1038/ni.3156

Song, H., Hora, B., Bhattacharya, T., Goonetilleke, N., Liu, M. K. P., Wiehe, K., … Gao, F. (2014). Reversion and T cell escape mutations compensate the fitness loss of a CD8+ T cell escape mutant in their cognate transmitted/founder virus. *PLoS One*, *9*(7), e102734. doi:10.1371/journal.pone.0102734

Soriano, V., Bravo, R., Samaniego, J. G., Gonzalez, J., Odriozola, P. M., Arroyo, E., … Carballo, E. (1994). CD4+ T-lymphocytopenia in HIV-infected patients receiving interferon therapy for chronic hepatitis C. *AIDS*, *8*(11), 1621–1622.

Sriwanthana, B., Mori, M., Tanaka, M., Nishimura, S., Miura, T., Pathipvanich, P., … Ariyoshi, K. (2012). The effect of HLA polymorphisms on the recognition of Gag epitopes in HIV-1 CRF01_AE infection. *PLoS One*, *7*(7), e41696. doi:10.1371/journal.pone.0041696

Stewart-Jones, G. B. E., di Gleria, K., Kollnberger, S., McMichael, A. J., Jones, E. Y., & Bowness, P. (2005). Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705. *European Journal of Immunology*, *35*(2), 341–351. doi:10.1002/eji.200425724

Streeck, H., Brumme, Z. L., Anastario, M., Cohen, K. W., Jolin, J. S., Meier, A., … Altfeld,

M. (2008). Antigen load and viral sequence diversification determine the functional profile of HIV-1-specific CD8+ T cells. *PLoS Medicine*, *5*(5), e100. doi:10.1371/journal.pmed.0050100

Streeck, H., Jolin, J. S., Qi, Y., Yassine-Diab, B., Johnson, R. C., Kwon, D. S., … Altfeld, M. (2009). Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells. *Journal of Virology*, *83*(15), 7641–7648. doi:10.1128/JVI.00182-09

Streeck, H., Li, B., Poon, A. F. Y., Schneidewind, A., Gladden, A. D., Power, K. A., … Allen, T. M. (2008). Immune-driven recombination and loss of control after HIV superinfection. *The Journal of Experimental Medicine*, *205*(8), 1789–1796. doi:10.1084/jem.20080281

Streeck, H., Lichterfeld, M., Alter, G., Meier, A., Teigen, N., Yassine-Diab, B., … Altfeld, M. (2007). Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles. *Journal of Virology*, *81*(14), 7725–7731. doi:10.1128/JVI.00708-07

Streeck, H., Lu, R., Beckwith, N., Milazzo, M., Liu, M., Routy, J.-P., … Altfeld, M. (2014). Emergence of individual HIV-specific CD8 T cell responses during primary HIV-1 infection can determine long-term disease outcome. *Journal of Virology*, *88*(21), 12793–12801. doi:10.1128/JVI.02016-14

Strickland, S. L., Gray, R. R., Lamers, S. L., Burdo, T. H., Huenink, E., Nolan, D. J., … Salemi, M. (2012). Efficient transmission and persistence of lowfrequency SIVmac251 variants in CD8-depleted rhesus macaques with different neuropathology. *Journal of General Virology*, *93*(5), 925–938. doi:10.1099/vir.0.039586-0

Sturdevant, C. B., Joseph, S. B., Schnell, G., Price, R. W., Swanstrom, R., & Spudich, S. (2015). Compartmentalized replication of R5 T cell-tropic HIV-1 in the central nervous system early in the course of infection. *PLoS Pathogens*, *11*(3), e1004720. doi:10.1371/journal.ppat.1004720

Sulkowski, M. S. (2013). Current management of hepatitis C virus infection in patients with HIV Co-infection. *Journal of Infectious Diseases*, *207*(Suppl. 1), S26–32. doi:10.1093/infdis/jis764

Sung, J. A., Lam, S., Garrido, C., Archin, N., Rooney, C. M., Bollard, C. M., & Margolis, D. M. (2015). Expanded cytotoxic T-cell lymphocytes target the latent HIV reservoir. *The Journal of Infectious Diseases*, *212*(2), 258–263. doi:10.1093/infdis/jiv022

Sunshine, J. E., Larsen, B. B., Maust, B., Casey, E., Deng, W., Chen, L., … Frahm, N. (2015). Fitness-balanced escape determines resolution of dynamic founder virus escape processes in HIV-1 infection. *Journal of Virology*, *89*(20), 10303–10318. doi:10.1128/JVI.01876-15

Techakriengkrai, N., Tansiri, Y., & Hansasuta, P. (2012). Poor HIV control in HLA-B*27 and -B*57/58 non-controllers is associated with limited number of polyfunctional Gag p24-specific-CD8+ T cells. *AIDS*, *27*(1), 17–27.

doi:10.1097/QAD.0b013e32835ac0e1

Thobakgale, C. F., Prendergast, A., Crawford, H., Mkhwanazi, N., Ramduth, D., Reddy, S., … Goulder, P. J. R. (2009). Impact of HLA in mother and child on disease progression of pediatric human immunodeficiency virus type 1 infection. *Journal of Virology*, *83*(19), 10234–10244. doi:10.1128/JVI.00921-09

Troyer, R. M., McNevin, J., Liu, Y., Zhang, S. C., Krizan, R. W., Abraha, A., … Arts, E. J. (2009). Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. *PLoS Pathogens*, *5*(4), e1000365. doi:10.1371/journal.ppat.1000365

Tsegaye, A., Ran, L., Wolday, D., Petros, B., Dorigo, W., Piriou, E., … van Baarle, D. (2007). HIV-1 Subtype C gag-specific T-cell responses in relation to human leukocyte antigens in a diverse population of HIV-infected Ethiopians. *Journal of Acquired Immune Deficiency Syndromes*, *45*(4), 389–400. doi:10.1097/QAI.0b013e318059beaa

Tully, D. C., & Wood, C. (2010). Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. *AIDS*, *24*(10), 1577–1582. doi:10.1097/QAD.0b013e32833999e1

UNAIDS. (2014). Fact Sheet: 2014 Global Statistics. Retrieved August 15, 2015, from http://www.unaids.org/en/resources/documents/2015/20150714_factsheet

Urbanus, A. T., van de Laar, T. J., Stolte, I. G., Schinkel, J., Heijman, T., Coutinho, R. A., & Prins, M. (2009). Hepatitis C virus infections among HIV-infected men who have sex with men: an expanding epidemic. *AIDS*, *23*(12), F1–7. doi:10.1097/QAD.0b013e32832e5631

van de Laar, T. J. W., van der Bij, A. K., Prins, M., Bruisten, S. M., Brinkman, K., Ruys, T. A., … Coutinho, R. A. (2007). Increase in HCV incidence among men who have sex with men in Amsterdam most likely caused by sexual transmission. *The Journal of Infectious Diseases*, *196*(2), 230–238. doi:10.1086/518796

Wagner, G. A., Pacold, M. E., Kosakovsky Pond, S. L., Caballero, G., Chaillon, A., Rudolph, A. E., … Smith, D. M. (2014). Incidence and Prevalence of Intrasubtype HIV-1 Dual Infection in At-Risk Men in the United States. *Journal of Infectious Diseases*, *209*(7), 1032–1038. doi:10.1093/infdis/jit633

Wang, Y. E., Li, B., Carlson, J. M., Streeck, H., Gladden, A. D., Goodman, R., … Allen, T. M. (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. *Journal of Virology*, *83*(4), 1845–1855. doi:10.1128/JVI.01061-08

Watson, S. J., Welkers, M. R. A., Depledge, D. P., Coulter, E., Breuer, J. M., de Jong, M. D., & Kellam, P. (2013). Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philosophical Transactions of the Royal Society*, *368*(1614), 20120205. doi:10.1098/rstb.2012.0205

Webster, D. P., Klenerman, P., & Dusheiko, G. M. (2015). Hepatitis C. *The Lancet*,

*385*, 1124–35. doi:10.1016/S0140-6736(14)62401-6

Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., … Shaw, G. M. (2003). Antibody neutralization and escape by HIV-1. *Nature*, *422*(6929), 307–312. doi:10.1038/nature01470

Whitney, J. B., Hill, A. L., Sanisetty, S., Penaloza-MacMaster, P., Liu, J., Shetty, M., … Barouch, D. H. (2014). Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature*, *512*(7512), 74–77. doi:10.1038/nature13594

Williams, I., Churchill, D., Anderson, J., Boffito, M., Bower, M., Cairns, G., … Youle, M. (2014). British HIV Association guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2012 (Updated November 2013). *HIV Medicine*, *15*(S1), 1–6. doi:10.1111/hiv.12119

Wolbers, M., Babiker, A., Sabin, C., Young, J., Dorrucci, M., Chêne, G., … Bucher, H. C. (2010). Pretreatment CD4 cell slope and progression to AIDS or death in HIV-infected patients initiating antiretroviral therapy—The CASCADE Collaboration: A collaboration of 23 cohort studies. *PLoS Medicine*, *7*(2), e1000239. doi:10.1371/journal.pmed.1000239

Yang, O. O. (2009). Candidate vaccine sequences to represent intra- and inter-clade HIV-1 variation. *PloS One*, *4*(10), e7388. doi:10.1371/journal.pone.0007388

Yang, X., Charlebois, P., Gnerre, S., Coole, M. G., Lennon, N. J., Levin, J. Z., … Henn, M. R. (2012). De novo assembly of highly diverse viral populations. *Biomed Central Genomics*, *13*(1), 475. doi:10.1186/1471-2164-13-475

Yang, X., Charlebois, P., Macalalad, A., Henn, M. R., & Zody, M. C. (2013). V-Phaser 2: variant inference for viral populations. *BMC Genomics*, *14*, 674. doi:10.1186/1471-2164-14-674

Yu, X. G., Lichterfeld, M., Chetty, S., Williams, K. L., Mui, S. K., Miura, T., … Walker, B. D. (2007). Mutually exclusive T-cell receptor induction and differential susceptibility to human immunodeficiency virus type 1 mutational escape associated with a two-amino-acid difference between HLA class I subtypes. *Journal of Virology*, *81*(4), 1619–1631. doi:10.1128/JVI.01580-06

Yue, L., Pfafferott, K. J., Baalwa, J., Conrod, K., Dong, C. C., Chui, C., … Hunter, E. (2015). Transmitted virus fitness and host T cell responses collectively define divergent infection outcomes in two HIV-1 recipients. *PLoS Pathogens*, *11*(1), e1004565. doi:10.1371/journal.ppat.1004565

Zhou, T., Georgiev, I., Wu, X., Yang, Z.-Y., Dai, K., Finzi, A., … Kwong, P. D. (2010). Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science*, *329*(5993), 811–817. doi:10.1126/science.1192819

Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M., & Ho, D. D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature*, *391*(6667), 594–597. doi:10.1038/35400

“It always seems impossible until it’s done.”

Nelson Mandela (1918-2013)