Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________ ____________________

Daniel Claiborne Date

New Insight into Host and Viral Factors that Influence HIV-1 Pathogenesis

By

Daniel Claiborne
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Immunology and Molecular Pathogenesis

____

Eric Hunter, Ph.D.
Advisor

____

Cynthia Derdeyn, Ph.D.
Committee Member

____

Paul Goepfert, M.D.
Committee Member

____

Guido Silvestri, M.D.
Committee Member

____

Brian Evavold, Ph.D.
Committee Member

____

Arash Grakoui, Ph.D.
Committee Member

Accepted:

____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____

Date

New Insight into Host and Viral Factors that Influence HIV-1 Pathogenesis

By

Daniel Claiborne
B.S., Florida State University, 2008

Advisor: Eric Hunter, Ph.D.

An abstract of
a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Immunology and Molecular Pathogenesis
2014

Abstract

New Insight into Host and Viral Factors that Influence HIV-1 Pathogenesis

By Daniel Claiborne

Human immunodeficiency virus type-1 (HIV-1), the etiologic agent of acquired immunodeficiency syndrome (AIDS), was responsible for 1.6 million deaths in 2012, and there are over 35 million people worldwide currently living with HIV infection. HIV is a life-long, chronic viral infection characterized by a steady decline in CD4+ T cells resulting in a state of overt immunodeficiency. Despite the fact that a majority of HIV-1 infected individuals eventually progress to AIDS, they do so at varying rates, implying that host or viral factors may alter the trajectory of disease. Here, we study acute HIV-1 infection in a cohort of Zambian volunteers in order to define the complex interplay between transmitted viral characteristics and the host immune response and their influence on the trajectory of HIV-1 pathogenesis.

Here, we demonstrate that viral replicative capacity (vRC), as defined by the viral Gag protein, has a dramatic impact on HIV-1 disease progression, in that high vRC is associated with elevated plasma viral loads and accelerated loss of CD4+ T cells. Furthermore, we highlight an integral role for vRC in driving multiple facets of HIV-1 immunopathology. High vRC initiates an exacerbated inflammatory state characterized by increased levels of inflammatory cytokines, aberrant T cell activation, exhaustion, and proliferation, and increased infection of central memory CD4+ T cells.

Events dictated by vRC can be further modulated by the host's cellular immune response. In this same cohort of acutely infected individuals we identify novel immunogenetic factors associated with significant protection from CD4+ T cell decline. Interestingly, these alleles exert a protective effect without significantly controlling plasma viral load, suggesting an alternate mechanism. Indeed, we find that in the earliest stages of infection, these protective immunogenetic factors are associated with reduced markers of gut damage and microbial translocation, which are known to contribute to chronic immune activation in HIV infection.

These data expand our current knowledge of the viral and host characteristics that influence the trajectory of HIV-1 pathogenesis and specifically highlight the importance of early events post transmission, which dramatically impact the course of disease. Furthermore, these results suggest that vaccine-induced immune responses or interventional therapeutics capable of attenuating early viral replication may have a significant, long-term benefit for the host.

New Insight into Host and Viral Factors that Influence HIV-1 Pathogenesis

By

Daniel Claiborne
B.S., Florida State University, 2008

Advisor: Eric Hunter, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Immunology and Molecular Pathogenesis
2014

Dedication

This body of work is dedicated to my parents, Edward and Patricia Claiborne. When two people believe in you so fiercely, it is impossible to fail.

Acknowledgements

I would like to thank:

- My wife, Jasmine Camps Claiborne, for her continuing love and support, and for having the courage to take this journey with me

- My parents, Ed and Patti, for being my biggest supporters from the very beginning

- Jessica Prince-Guerra, for her invaluable contributions to this body of work and for the opportunity to work with one of the finest scientists I have ever met

- All of the graduate students and post docs in the Hunter Lab for their contributions to this work, their spirit of teamwork and inclusiveness, and their thoughtful encouragement over the years

- Dr. Susan Allen and all of the staff at the Zambia-Emory HIV Research Project (ZEHRP) for having the vision to establish a cohort that has dramatically enhanced our understanding of a devastating disease

- The human volunteers enrolled in ZEHRP for their courage, and for their essential contributions to this body of work and the advancement of the field

- Dr. Eric Hunter, for being the type of mentor I needed in every way, and for being a man worth emulating

Table of Contents

Abstract

Dedication

Acknowledgements

Chapter I: Introduction ... 1

The identification of HIV as the cause of AIDS and the state of the current pandemic ... 1

Origins of HIV ... 3

Viral characteristics underlying HIV persistence ... 5

Pathogenic mechanisms of HIV-1 infection ... 14

Factors that influence the trajectory of HIV pathogenesis ... 19

Summary ... 28

Chapter II: Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis ... 31

Table 1. Cohort statistics generated from the 149 transmission pairs selected from the ZEHRP cohort ... 70

Figure 1. Insertion of the *gag* gene from newly infected individuals dramatically alters the replicative capacity of MJ4 ... 71

Figure 2. Replicative capacity is correlated to viral load in recipients and donors ... 72

Figure 3. Identification of polymorphisms in Gag that significantly affect RC, several of which can be linked to HLA-class I alleles ... 73

Figure 4. Rare polymorphisms have a significantly greater impact on RC ... 75

Figure 5. The balance of fitness increasing and decreasing HLA-associated polymorphisms strongly correlates with RC ... 76

Figure 6. The balance of HLA-associated fitness increasing and decreasing mutations strongly correlates with set point viral load in newly infected individuals ... 77

Figure 7. RC affects the rate of CD4 decline in a manner that may be independent of viral load ... 79

Table 2. Cox proportional hazard models demonstrate the independent effects of RC and VL on CD4 decline ... 81

Figure S1. Donor and recipient population *gag* sequences cluster with one another ... 82

Table S1. Amino acids in Gag associated with changes in replicative capacity ... 84

Figure S2. Gag sequences that are less like the Gag subtype C consensus sequence replicate more efficiently *in vitro* ... 86

Chapter III: Transmitted HIV-1 replicative capacity drives immune activation and CD4 proviral load ... 100

Figure 1. HIV-1 replicative capacity, when defined by the transmitted gag sequence, predicts CD4 T cell decline in ART-naïve, HIV-1 infected individuals ... 128

Figure 2. Low vRC is associated with a distinct cytokine profile early in infection, characterized by muted inflammatory cytokine levels ... 129

Figure 3. High vRC is associated with increased CD8 T cell activation and lower cytotoxic potential ... 131

Figure 4. vRC is associated with increased cellular activation and proliferation in CD4 T cell memory subsets ... 133

Figure 5. Inflammatory cytokine profiles associated with vRC correlate with T cell activation ... 135

Figure 6. Viral RC correlates with the burden of HIV-1 viral DNA in CD4+ T_{CM} and T_N ... 137

Figure S1. The *gag* gene chimera is representative of full-length HIV-1 ... 139

Figure S2. The effect of log_{10}-increases in early set point VL on longitudinal CD4+ T cell decline ... 140

Figure S3. The first two principal components are significantly correlated with vRC and set point VL, respectively ... 141

Figure S4. CD8+ T cell activation phenotypes early after infection are associated with CD4+ T cell decline ... 142

Figure S5. High vRC is associated with an increased level of activation and turnover in CD4+ T cells that is highly deleterious ... 144

Table S1. Host and viral characteristics independently predict CD4+ T cell decline ... 145

Table S2. High vRC significantly increases early inflammatory cytokine levels ... 146

Chapter IV: Protective immunogenetic factors reduce microbial translocation in acute HIV infection 156

Table 1. Immunogenetic factors significantly affecting CD4+ T cell decline or longitudinal control of plasma viral load in HIV-1 subtype C infection 171

Figure 1. Protective HLA-I alleles are associated with reduced plasma lipopolysaccharide (LPS) levels at seroconversion 173

Figure 2. Protective HLA-I alleles are associated with reduced plasma LPS, sCD14, I-FABP, and IL-10 at 6-months post seroconversion 174

Figure 3. Circulating plasma LPS at seroconversion predicts CD4+ T cell decline and time to initiation of ARV treatment 176

Table 2. Prevention of early microbial translocation is a common mechanism among protective host and viral characteristics. 178

Figure S1. Protective and deleterious HLA alleles are additive in nature represent distinct pathogenesis profiles 179

Table S1. Protective HLA alleles predict CD4+ T cell decline in manner distinct from set point VL 180

Figure S2. Accounting for batch effects further improves the association between protective HLA-I alleles and low levels of LPS at seroconversion 181

Table S2. Detectable LPS in the plasma at seroconversion drives CD4+ T decline via a mechanism distinct from set point VL 182

Table S3. In a multivariable generalized linear model, total LPS levels are significantly reduced by protective host and viral factors 183

Chapter V: Discussion 198

The impact and implications of heritable viral characteristics on HIV-1 pathogenesis 198

Unique mechanisms of HLA-mediated protection 202

The complex nature of sex-based differences in the immune response to HIV infection 204

In Conclusion 207

Literature Cited in Chapters I (Introduction) and V (Discussion) 210

Appendix 231

Figure 1. Females exhibit lower viral loads and higher CD4+ T cells counts in the first 2 years of HIV infection 231

Figure 2. Females present with reduced cellular immune activation early in infection, even when controlling for the effects of plasma viral load 232

Figure 3. Acute HIV infection in females is characterized by a distinct inflammatory cytokine profile 233

Figure 4. Transmitted viruses in females exhibit differential reversion kinetics in comparison to viruses transmitted to males...234

Chapter I: Introduction

The identification of HIV as the cause of AIDS and the state of the current pandemic

In 1981, the Morbidity and Mortality Weekly Report, an epidemiological report used to disseminate issues of public health interest by the Centers for Disease Control and Prevention (CDC), documented cases of *Pneumocystis carinii* in 5 young homosexual men in Los Angeles, California(1). A similar report soon followed, detailing the presence of Kaposi's sarcoma (a rare malignancy) and other opportunistic infections, including *Pneumocystis carinii*, in homosexual men in New York, Los Angeles, and San Francisco, all of which were found to be severely immunocompromised(2). These two reports would herald the emergence of a new global pandemic, subsequently termed acquired immunodeficiency syndrome (AIDS). In 1983, two years after the first reports of AIDS, the etiologic agent of this immune deficiency syndrome, a new human retrovirus, was isolated by French scientists at the *Institut Pasteur*(3). This retrovirus would later be known as "human immunodeficiency virus" (HIV), and would spread across the globe, causing one of the most devastating public health crises of our generation.

Though AIDS was first reported in men who have sex with men, followed by injection drug users, and those receiving contaminated blood transfusions, heterosexual intercourse is the main route of HIV transmission worldwide, proving that this is not a disease relegated to initially defined high risk groups(4, 5). Since the beginning of the pandemic, HIV has infected more than 75 million people worldwide and has caused at least 30 million deaths(4). In 2012, there were 35.3 million people living with HIV, and

HIV infection led to 1.6 million deaths(4). The inordinate majority of this global pandemic resides in sub-Saharan Africa (71% of all infections globally), where 1 in 20 adults are infected with HIV, and the highest number of HIV-related deaths is reported(4). Though the statistics regarding HIV infection and mortality are sobering, the scientific community has made great strides in the past 30 years that have transformed HIV infection from a death sentence into a manageable, chronic disease. This has been largely accomplished by the advent of anti-retroviral drugs (ARVs) that are now capable of suppressing HIV replication to undetectable levels(6-9). Suppression of viral loads by ARVs in conjunction with the implementation of public health measures such as screening and testing, male circumcision, and condom provision, have slowed the spread of new HIV infections(4, 5, 10-13). However, HIV continues to spread globally, and HIV is still the leading cause of death in individuals 25 – 44 years of age, worldwide(14).

Anti-retroviral therapy (ART) has dramatically reduced the mortality of HIV in resource-rich countries, but access to ART is not universal and many individuals in sub-Saharan Africa, where HIV is most prevalent, do not have access to life saving drugs. This, coupled with the fact that there were 2.3 million new infections in 2012, calls for a safe and effective HIV vaccine to stem the tide of the HIV pandemic(4). This is and should be a primary focus of the scientific community as it is unlikely that ART alone, though potent in reducing morbidity and mortality and slowing transmission, will be able to prevent the majority of new infections.

Origins of HIV

Though HIV was identified as a new human retrovirus, retroviruses in general have been a part of mammalian evolution, and more specifically primate evolution, for millennia(15). In order to design rational vaccines and therapies for a rapidly evolving virus, we must first understand the origins of HIV and how retroviruses, and more specifically lentiviruses (the genus of the family *Retroviridae,* capable of infecting non-dividing cells, of which HIV is a member) have co-evolved with their hosts throughout history.

Acquired immunodeficiency syndrome (AIDS) in humans is actually caused by two distinct lentiviruses, HIV-1 and HIV-2, which represent two separate zoonotic transmission events from African nonhuman primates to humans. HIV-1 was identified first in 1983(3), and is responsible for the vast majority of infections globally. Its origins remained uncertain until two clues lead the way. First, HIV-2 was identified in 1986 in western Africa, and was also found to cause AIDS in infected individuals(16). Second, HIV-2 was found to be related, but genetically distinct from HIV-1, and was, in fact, more closely related to a simian virus that was shown to cause immunodeficiency in captive Rhesus macaques(16). This lead to the identification of additional simian immunodeficiency viruses (SIVs) in several African nonhuman primates, and to date, at least 30 distinct viral lineages have been isolated from various monkeys(17-19). With the identification of these SIVs in their natural hosts, the zoonotic sources of HIV-1 and HIV-2 were phylogenetically traced to chimpanzees (SIV*cpz*) and sooty mangabeys (SIV*smm*), respectively(20, 21). Additionally, isolation of SIV*cpz* from wild chimpanzees traced the source of HIV-1 group M (the source of the pandemic) to apes living in

southeastern Cameroon, suggesting a geographical location for the cross-species transmission event that sparked the beginning of the HIV-1 pandemic(22-24).

Exactly how HIV-1 was transmitted to humans from apes will most likely forever remain a mystery. However, in the geographical regions where cross-species transmission events have been phylogenetically traced, there is extensive interaction between humans and nonhuman primates, with individuals keeping primates as pets and hunting them for food. Hunting primates and preparing bushmeat(25) can lead to the exchange of body fluids or blood products, which is necessary for transmission of these lentiviruses. Indeed, sampling of primates hunted and butchered for bushmeat has demonstrated the presence of SIV infection in at least 20% of these animals(26, 27). These findings present a reasonable hypothesis for the basis of zoonotic transmission events, and suggest that cross-species transmission may continue to occur in the future(19, 28).

Interestingly, SIVs do not generally cause immunodeficiency in their natural hosts(29, 30). Though the prevalence rate can be higher than 50% and the virus replicates to relatively robust titers in naturally infected nonhuman primates such as the sooty mangabey, these animals very rarely show signs of compromised immune function(18, 31-34) In contrast, HIV-1 group M and its zoonotic ancestor, SIV*cpz*, cause immunodeficiency in their respective hosts(35). These observations raise at least two important questions: (I) why is the primate, and more importantly human, immune system unable to clear lentiviral infections, and (II) why does HIV infection in humans cause overt pathogenesis leadings to severe immunodeficiency, which is for the most part absent in nonhuman primate hosts naturally infected with SIV?

Viral characteristics underlying HIV persistence

HIV is a life-long, chronic viral infection, and no known cases to date have shown complete viral clearance in the absence of ART. Even with ART administration, there have only been a handful of cases demonstrating eradication of the virus, and this has only been achieved by extremely early (30 hours), post partum administration of ART to the infant in a select case of vertical transmission(36), or, in an adult, by a stem cell transplant with engraftment of donor bone marrow homozygous for the CCR5-Δ32 mutation(37), which renders CD4 T cells refractory to infection with CCR5-tropic strains of HIV(38-40).

Persistence of HIV in the face a vigorous immune response or highly active anti-retroviral therapy can be linked to a handful of viral characteristics; (I) infection of CD4+ T cells which causes direct cytopathic effects, or their indirect clearance by host cellular immune responses, which leads to the loss of this lymphocyte population, (II) integration of the provirus into the host genome, which results in a stable, long-lived, latent viral reservoir capable of reactivation, and (III) extreme genetic variation due to a relatively error-prone reverse transcriptase, recombination of genetically distinct viral genomes, and the high replication rate of HIV, which all lead to the rapid accumulation of mutations, some of which may be beneficial to the virus. These viral characteristics of CD4+ T cell tropism, integration, and rapid mutation have complex downstream effects that contribute to the chronic nature of HIV infection.

Impact of CD4+ T cell targeting

Due to the cellular tropism of HIV, which preferentially infects and kills activated CD4+ T cells(41, 42), virtually all facets of the immune system can be crippled early in acute HIV infection(43). CD4+ T cells are an integral player in the initiation and maintenance of a productive immune response. CD4+ T cells, also termed "helper T cells", contribute to effective mobilization of both the cellular and humoral arms of the adaptive immune system(44).

In the case of the humoral immune response, B cells generate antibodies specific for the extracellular proteins of a given pathogen, and help from CD4+ T cells is required for B cells to make high-affinity antibodies(45). In the case of HIV these antibodies are directed against the viral Envelope (Env) protein, which is responsible for attachment and entry into host cells. Virtually all aspects of the B cell response are perturbed in HIV infection(46), and the B cell compartment is defined by both being aberrantly hyperactivated(47), yet ultimately dysfunctional and unable to generate antibodies capable of preventing viral spread early in infection(48-51). Most of the B cell abnormalities associated with HIV infection can be directly attributed to ongoing viral replication. This can induce B cell hyperactivation and dysfunction either directly (52) or indirectly through soluble mediators of inflammation such as TNFa and IL-6(46, 53-56).

Skewing of certain B cell populations can be directly attributed to CD4+ T cell depletion. In HIV infection, CD4+ T cell loss is associated with increased levels of IL-7(57, 58), a homeostatic cytokine responsible for maintaining normal T cell numbers, and high levels of this cytokine have been linked to an increased frequency of immature transitional B cells(59). A similar increase in immature transitional B cell numbers is seen

in individuals with non-HIV related idiopathic CD4+ T cell lymphopenia(60), demonstrating that this skewing of B cell populations to a more immature (and presumably less functional) phenotype is independent of viral antigen load(46).

The infection and killing of CD4+ T cells also has a profound effect on the cellular immune response, which is perhaps even now not completely elucidated. Cytotoxic T lymphocytes (CTLs) are CD8+ T cells that recognize intracellular pathogens when pathogen-derived peptides are presented in the binding pocket of the major histocompatibility complex (MHC) class I molecule. Upon recognition of their cognate antigen-MHC complex, CTLs release stored cytotoxic granules, containing cytolytic molecules that initiate cell death pathways in the target infected cell, leading to its elimination, thereby limiting the spread of the pathogen(61). During an immune response, antigen-specific CD4+ T cells provide "help" during the priming of CD8 T cells by interacting with antigen presenting cells (such as dendritic cells and macrophages) and providing activating signals via CD40-CD40L interactions(62-64). This enables the antigen-presenting cell to fully activate naïve CD8 T cells to become effector CTLs. Naïve CD8+ T cells can become activated in the absence of CD4+ T cell help when they recognize their cognate antigen in the context of some pathogenic infections(65, 66), but certain functions, such as establishing a stable memory CD8 T cell pool capable of responding to secondary infections, and the ability to travel to certain infected tissues in the periphery, are ablated in the absence of CD4 T cell help(67-69).

Finally, though antigen-specific CD4 T cells are required to generate robust antibody responses to many pathogens and to generate optimal CTLs, they can also carry out their own effector functions and can even kill HIV-infected cells, much in the same

way as CD8+ CTLs(70, 71). However, HIV-specific CD4 T cells are also ideal activated targets for the virus, and reports have shown that HIV-specific CD4 T cells are preferentially infected in all stages of infection. They are even preferentially re-infected after interruption of successful ART(72). This observation poignantly illustrates the complex nature of HIV infection, and how immune responses that would be protective in other pathogenic infections are ineffective in the context of HIV, serving even to foment HIV replication and, thus, exacerbate pathogenesis.

Impact of chromosomal integration of the HIV genome

Like all retroviruses, HIV integrates its genome into the host genetic material. This, like reverse transcription of the viral genomic RNA to a double stranded DNA, is a defining characteristic of all retroviruses(73, 74). It is accomplished via the integrase protein, and upon integration, HIV recruits the host transcriptional machinery in order to produce genomic and sub-genomic viral RNAs. The latter are necessary to generate the protein building blocks that are assembled into progeny virions, while the former represent the full-length viral RNAs that will be packaged into budding virions to serve as progeny viral genomes(75). In some cases, however, rather than being productively infected, the infected CD4+ T cell returns to a quiescent state in which viral genes are not actively transcribed, most likely due to the absence of activation-dependent host transcription factors(76, 77). In this case, the integrated viral genome is stable and transcriptionally silent, and the cell goes on to survive and replicate the viral genome, along with host genes, upon cell division(78). This forms a latent viral reservoir, which poses a significant barrier to HIV eradication(79-81).

Latently infected CD4+ T cells are problematic for at least two reasons. First, the latent viral reservoir is preferentially found in resting memory CD4 T cells(82, 83), which do not typically exhibit a significantly activated phenotype necessary for virus production. Additionally, these cells have a long half-life measured to be between 40 – 44 months. Calculations based on the size of the initial viral reservoir in HIV-infected individuals suggest that highly active ART, though effective in completely blocking infection of new cells, is unlikely to be adequate in eradicating this stable viral reservoir, even in the context of life-long treatment(84, 85). Furthermore, the host cellular immune response, which must recognize viral proteins presented on the cell surface by MHC complexes, cannot kill these latently infected cells, as they do not actively express viral proteins. To further complicate matters, the viral reservoir is found in at least two subsets of long-lived memory CD4 T cells, central memory (T_{CM}) and transitional memory CD4 T cells (T_{TM}), the latter of which persists due to IL-7 mediated homeostatic proliferation, which does not require the presence of viral antigen(83).

Second, latently infected cells harbor viral genomes capable of producing replication competent progeny. At first, it was proposed that latently infected cells harbored defective viral genomes that could not produce infectious virus. However, seminal studies showed that latently infected T cells, though relatively rare in the periphery (1 in 10^6 CD4+ T cells), did indeed harbor viral genomes capable of reactivation in the presence of mitogens(82, 86), and that after cessation or failure of an ART regimen, these latently infected reservoirs were capable of reactivation and responsible for rapid viral rebound(87), with reseeding of the viral reservoir. In the case of reinitiating ART, this further confounds viral control as the viral reservoir can be

reseeded with drug resistant variants, while continuing to harbor the initial infecting variant, thus maintaining viral diversity(88-90).

When analyzed together, these data suggested that after seeding of the initial viral reservoir, which occurs very early post infection(91, 92), ART or therapeutic vaccination strategies will be insufficient to achieve a functional cure for HIV-infected individuals, and that new strategies aimed at directly targeting the viral reservoir will be necessary to achieve this goal(78).

Impact of HIV genetic variation

Though integration into the host genome and the establishment of the latent reservoir is paramount for viral persistence in the context of highly active ART, perhaps the greatest obstacle to the control of viral replication by the host immune system is HIV's extreme genetic diversity(93). HIV, which contains two positive-sense RNA genomes per virion, replicates through a DNA intermediate. In order to generate the DNA intermediate suitable for integration into the host genome and necessary for expression of viral gene products that produce progeny virions, HIV uses the virally encoded reverse transcriptase (RT) enzyme and its RNA genome as a template in order to transcribe a single-stranded, and then double stranded, proviral genome(94). However, during the process of reverse transcription RT incorporates between .2 and 6 incorrect nucleotide bases per 10^4 nucleotide base pairs (or roughly .2 to 6 mistakes per replication cycle)(95-97). The disparity between measurements can be traced to their respective experimental conditions, and it is thought that the in vitro measurements of RT fidelity, which give rise to the lower bounds of fidelity measurements, are an underestimation of

HIV RT fidelity(94). In any event, this mutation rate is magnified by the formidable replication capacity of HIV, as it is estimated that 10.3×10^9 new virions are produced in a single HIV-infected individual per day(98).

The genetic diversity of HIV can be further increased by recombination events(99-101). During the first step of reverse transcription, when the negative sense DNA strand is transcribed from the positive sense RNA viral genome, HIV RT can switch between the two positive strand RNA genomes packaged in each virion in order to generate the full length negative sense DNA strand, and it is estimated that this occurs two to three times per replication cycle of each viral genome(99). If the two co-packaged RNA genomes are genetically distinct, which would most likely occur if the cell they originated from was at least dually infected with different viruses, the resulting proviral genome will be chimeric. Genetic diversity will then be increased if the integrated provirus generates replication competent progeny. Of note, recombination is perhaps an even greater force in shaping the genetic diversity of the epidemic than variation generated within a single infected individual, since it has been proposed that more than 20% of the current circulating HIV variants are recombinants between different subtypes(102).

Thus, the misincorporation of nucleotides by HIV RT is magnified many times on a daily basis within an HIV-infected individual, and is further diversified by recombination events, which give rise to a diverse viral quasispecies that can rapidly be selected for advantageous mutations. These mutations may benefit the virus in escaping cellular and humoral adaptive immune responses as well as developing resistance to anti-retroviral drugs(93).

While the cellular immune response is absolutely necessary to control acute viremia down to a stable set point viral load(103-107), the intense pressure exerted by this cellular immune response, coupled with the extreme genetic diversity of HIV, leads to the rapid selection of escape mutations that are advantageous to the virus(108-111). Amino acid changes in the viral sequence can become fixed in the viral quasispecies when they provide a distinct advantage in the context of a specific cellular immune response. These mutations can serve to evade the cytotoxic T lymphocyte (CTL) response, which usually recognizes and eliminates infected cells, in at least three distinct ways(112, 113). First, amino acid substitutions can interfere with proteasome processing(114, 115), necessary to generate the short viral peptides that can be loaded onto MHC class I molecules, which are then trafficked to the cell surface, and recognized by the T-cell receptor (TCR) of an HIV-specific CTL. These processing mutations can prevent the generation of immunogenic epitopes, effectively thwarting the effector function of CTLs specific for that epitope. Second, mutations at anchor residues in the viral peptides can prevent or attenuate viral peptide binding to MHC class I, which occurs in the endoplasmic reticulum, and can prevent viral peptide-MHC complexes from trafficking to the surface(116, 117). Third, escape mutations found in the viral epitope that do not prevent proteosomal processing or binding to MCH class I can further circumvent CTL recognition of infected cells if they occur at contact residues normally recognized by the T-cell receptor (TCR) of HIV-specific CTLs(118). Such mutations can reduce or abrogate the ability of the CTL to recognize its cognate viral peptide-MHC complex, which results in the survival of the infected cell and the continued production of virus.

Though a virus-inhibitory humoral immune response is delayed in HIV infection, presumably due the dysregulation of B cells and the disruption of lymphoid architecture(46), autologous neutralizing antibodies, which prevent viral attachment or entry and help to block viral spread, are generated within the first few weeks of infection. Antibodies specific for the viral envelope (Env) protein that bind neutralizing epitopes or bind non-neutralizing epitopes that can recruit effector cells to kill infected target cells collectively exert significant pressure on the virus(51, 119, 120). This can been evidenced by the rapid accumulation of escape mutations in the Env protein(50), which serve to abrogate antibody binding either by mutation of key contact residues in the epitope or by altering HIV gp120's formidable glycan shield(51), which can serve to conceal vulnerable epitopes. However, the antibody response evolves to target new epitopes after viral escape, and successive waves of escape followed by refocusing of the humoral immune response represent an "arms race" characteristic of HIV infection(121). Unfortunately, due to immune activation, disruption of lymph node architecture, and general dysfunction in the B cell compartment, the humoral immune response is unable to suppress viral replication.

In summary, direct infection and depletion of CD4+ T cells, integration into the host genome and creation of a latent reservoir, and rapid HIV evolution collectively represent significant barriers to the eradication of HIV. These viral characteristics severely abrogate the ability of the host immune system to adequately control and eliminate HIV, which in the majority of cases leads to progressive damage to the immune system, CD4+ T cell loss, and the development of AIDS.

Pathogenic mechanisms of HIV-1 infection

In the context of sexual transmission, infection is established as discrete foci in the genital mucosa when HIV encounters and infects a susceptible target cell, which can be a dendritic cell or activated CD4+ T cell. Local inflammation recruits additional target cells, and HIV subsequently traffics to the draining lymph node and from there disseminates to the gut, which is rich in the preferred cellular target for HIV replication, activated memory CD4+ T cells. This occurs concomitantly with a burst in viral replication leading to acute viremia, which can reach up to several million viral RNA copies per milliliter of blood(122-124). Acute viremia results in complete depletion of CD4+ T cells in the gut, and a transient depletion of CD4+ T cells in the periphery(125). Acute viremia is reduced by the induction of the cellular immune response(103, 104), in which HIV-specific CTLs kill infected cells, and effectively reduce acute viremia down to a stable set point viral load (SPVL). The magnitude of viral load set point was recognized early on as a powerful predictor of disease progression(126, 127), with low SPVL being advantageous and high SPVL being deleterious. At this stage of infection, there is a rebound in CD4 T cell counts in the periphery, which then steadily declines throughout the course of infection due to the immune system's inability to completely control viral replication, ultimately leading to the loss of CD4+ T cells down to less than $200/mm^3$, the clinical definition of AIDS, and the ingress of opportunistic infections.

Though the case for direct infection and killing of CD4+ T cells due to uncontrolled viral replication as the cause of HIV pathogenesis appears self evident, recent findings have cast doubt on this being the sole mechanism driving disease(128,

129). The first lesson is taken from studies of sooty mangabeys and African green monkeys, two natural SIV hosts that do not show signs of CD4+ T cell depletion or AIDS-like symptoms in the face of robust viral replication and appreciable plasma viral loads, suggesting that viral replication alone is not sufficient to cause disease(18, 32). These observations were followed by work in humans, showing that the number of HIV infected cells correlated with the number of cells undergoing apoptosis (programmed cell death), with the central caveat being that productively infected cells rarely became apoptotic themselves(130). This demonstrated that HIV could lead to CD4 T cell death by indirect mechanisms, and suggested that CD4 T cells might be depleted by "bystander" activation and apoptotic cell death in conjunction with direct infection. Apoptosis of bystander CD4 T cells was subsequently linked to a state of immune activation found in the chronic stage of HIV infection, at a time in which immune activation was emerging as a reliable predictor of CD4 T cell decline(131-133). Perturbations in the proliferation kinetics of CD4 as well as CD8 T cells suggested an infection-independent mechanism of immune activation(134, 135), as HIV does not infect CD8 T cells. These data refocused the field to explore a more complex mechanism for CD4 T cell depletion in HIV infection outside of direct infection and killing, and pointed to chronic immune activation as the major driver of disease(128). This refocusing of the field led to a detailed comparison of the pathogenic and nonpathogenic models of SIV infection, which yielded invaluable insights into the mechanisms of SIV pathogenesis, and by extension, HIV pathogenesis in humans(30).

Following pathogenic SIV infection of rhesus macaques, the virus readily infects the activated, tissue-resident CD4+ T cells found in the gastrointestinal tract, leading to

complete depletion of these effector cells that is never recovered(136). Acute depletion also occurs in nonpathogenic natural hosts, such as the sooty mangabey, but these animals normalize this loss, and gut-resident CD4+ T cells are at least partially restored(137, 138). The massive depletion of CD4+ T cells in the gastrointestinal tract is mirrored in HIV infection(139, 140), suggesting this early depletion without subsequent recovery is an early event specific to pathogenic infection. In support of this, depletion of CD4+ T cells in the gastrointestinal tract is associated with breaches in tight junctions of the gut epithelia and the translocation of microbial products, such as bacterial lipopolysaccharide (LPS) from the lumen of the gut to the periphery(141, 142). LPS is a potent activator of the immune system, and this observed microbial translocation drives immune activation in both the innate and adaptive immune system during the chronic stages of HIV infection. Furthermore, microbial translocation is absent in natural hosts, providing strong evidence for microbial translocation as a key component in the events driving pathogenesis(142).

Though microbial translocation is associated with increased levels of circulating inflammatory cytokines and cellular immune activation in the chronic stages of infection, early and robust expression of inflammatory cytokines is a hallmark of HIV that occurs very early post transmission(143). Acute HIV infection is associated with waves of inflammatory cytokines, the first as early as 6 days after the detection of viremia in the plasma and characterized most notably by elevated levels of interferon alpha (IFNa), inducible protein-10 (IP-10), tumor necrosis factor alpha (TNFa), and monocyte chemotactic protein-1 (MCP-1), indicative of an intense innate immune response. This is rapidly followed by elevated levels of interleukin-6 (IL-6), interleukin-10 (IL-10), and

interferon gamma (IFNg). However this early and robust innate immune response characterized by an intense "cytokine storm" is not a prerequisite for viral clearance, as both hepatitis C and hepatitis B virus, two chronic viruses that are cleared in 20% and 90% of infections respectively(144), exhibit delayed induction kinetics and muted levels of inflammatory cytokines when compared to HIV(143). Interestingly, natural SIV hosts also exhibit robust induction of IFNa and downstream interferon stimulated genes (ISGs) at similar levels as seen in the pathogenic infection model of rhesus macaques, demonstrating the initial innate immune response to SIV infection is not attenuated in natural hosts. However, natural hosts resolve this innate immune activation and the expression of ISGs in an active down-regulation process, mediated in part by increased expression of indoleamine 2,3-deoxygenase (IDO), an immunosuppressive protein that negatively regulates both T cell proliferation and IFN activity(145). This suggests that sustained levels of these inflammatory mediators that persist into the chronic stage of infection are what contribute to the unresolved immune activation associated with pathogenic SIV/HIV infection.

An additional distinct mechanism by which natural hosts appear to avoid disease is protection of central memory (T_{CM}) CD4+ T cells from direct infection and depletion(146). This is accomplished in sooty mangabeys by reduced CCR5 co-receptor expression on CD4+ T_{CM} after stimulation (as demonstrated in vitro), which effectively limits infection of this vulnerable population *in vivo*(147). In contrast, the progressive loss of CD4+ T_{CM} in pathogenic SIV infected rhesus macaques had been clearly associated with the development of AIDS(148). CD4+ T_{CM} reside in the secondary lymphoid organs and exhibit decreased effector functions but great proliferative potential when stimulated

via recognition of their cognate antigen(149). CD4+ T_{CM} also proliferate under homeostatic conditions in the absence of antigen and in response to IL-7(150). Thus, an intact central memory T cell pool is necessary for the preservation of T cell homeostasis, a process that is greatly perturbed in pathogenic SIV/HIV infection(83, 148).

In summary, mechanisms that are absent in nonpathogenic SIV infection, and by extension drive pathogenesis in HIV infection include (I) sustained depletion of gut-resident CD4+ T cells without reconstitution, (II) damage to the gut epithelia and translocation of microbial products to the periphery, (III) a rapidly induced "cytokine storm" that is unable to clear the virus and persists into the chronic stages of infection, and (IV) infection and depletion of CD4+ T_{CM}, a memory T cell pool necessary for proper T cell homeostasis. The end result of these distinct mechanistic differences in pathogenic SIV/HIV infection is the induction of sustained chronic immune activation concomitant with aberrant proliferation and perturbed CD4+ T cell homeostasis, which is completely absent in nonpathogenic SIV infection of natural hosts(30, 151).

The insights gained from comparative studies of natural and pathogenic SIV/HIV hosts are compelling and unanimously point to chronic immune activation as the major driver of disease progression and the development of AIDS. Indeed, in HIV infected individuals with viral loads successfully controlled by ART, residual immune activation predicts poor reconstitution of CD4+ T cells(152), and elevated levels of circulating inflammatory mediators are associated with an increased risk of all-cause mortality(153, 154). Of note, immune activation, though ubiquitous in HIV infection, varies in magnitude between individuals, and the rate at which individuals progress to AIDS also varies greatly(155, 156). This implies that various factors that differ between individuals,

whether of host or viral origin, significantly contribute to the trajectory of CD4+ T cell decline and the development of disease.

Factors that influence the trajectory of HIV pathogenesis

Though the majority of individuals infected with HIV will progress to AIDS in the absence of ART, they do so at varying rates. In fact, some individuals, termed "long-term non-progressors", maintain healthy CD4 T cell counts and exhibit low-level viral replication for many years post infection(157-159). Though not mutually exclusive, an even rarer population (about 1 in 300 HIV-infected persons), termed elite controllers (ECs), control viral replication to undetectable levels in the absence of ART. These observations suggest that there may be key factors that influence HIV pathogenesis, and further suggest that progression to AIDS is not necessarily an inescapable outcome of HIV infection(156). Uncovering these factors and the mechanisms by which they exert their effects will undoubtedly help us to define correlates of protection, which will be integral in the rational design of target based therapies and, ultimately, an effective HIV vaccine.

HIV infection marks the beginning of a constant battle between the virus and the host, and by extension, viral and host characteristics can significantly impact the kinetics of disease progression. To date, the preponderance of data has been collected in regard to host factors and their influence on viral control and disease progression, specifically focusing on the human leukocyte antigen (HLA) genes, which dictate the specificity and character of the cellular immune response(160, 161). The contribution of host factors to HIV-1 pathogenesis was first recognized in cohorts of homosexual men; in which

particular HLA class I (HLA-I) alleles were associated with time from infection to AIDS(162). Specific HLA haplotypes were soon thereafter correlated with accelerated disease progression(163). In contrast certain HLA alleles, such as HLA-B*27 and B*57, are protective and have been shown to be significantly enriched in ECs, demonstrating that HLA-I alleles can have a profound effect on viral control(164, 165). Further evidence for the influence of HLA-I comes from the rhesus macaque model of pathogenic SIV infection, where animals that spontaneously control viral loads are enriched for the Mamu*0801 allele(166), which encodes an MHC class I variant that binds and presents similar epitopes as the protective B*27 allele found in Caucasian ECs(167). In several comprehensive genome-wide association analyses (GWAS), only single-nucleotide polymorphsims (SNPs) found in the HLA-I gene locus were significantly associated with HIV-1 control(168-170), highlighting viral peptide presentation and the cellular immune response as a major player in modulating HIV disease. These data collectively offer compelling proof that the HLA locus, and by extension the cellular immune response, plays an integral role in the severity of HIV-1 pathogenesis.

From a mechanistic standpoint, carriage of these protective HLA alleles generally leads to a more potent and functional CD8+ T cell response. CD8+ T cells isolated from ECs produce more cytolitic molecules, such as granzyme and perforin(171-174), and have been shown to more potently inhibit viral replication in CD4+ T cells *ex vivo*(175). Perhaps most importantly, a key feature of protective alleles is the targeting of conserved epitopes, such as those derived from the relatively conserved Gag protein, which are less tolerant of escape mutations due to functional constraints(176, 177). These data demonstrate that CD8+ T cells, based on their binding to specific HLA presented

peptides, are a major component of the cellular immune responses' ability to control HIV replication. However, additional binding partners other than the TCR appear to contribute to the effect of HLA on control of HIV. Natural killer (NK) cells, a cytotoxic lymphocyte and innate component of the cellular immune response, can interact with HLA class I via their inhibitory and activating "killer cell immunoglobulin-like receptors" (KIR). Certain KIR/HLA haplotypes have been associated with slower progression to AIDS(178, 179), and NK cells expressing certain KIR alleles have potent antiviral activity and can productively inhibit HIV replication *in vitro*(180, 181).

Perhaps the best evidence in support of the profound effect of the cellular immune response on viral suppression is the presence HLA-associated viral escape mutations. Moore et al. were the first to show the presence of HLA-associated polymorphisms in HIV at the population level, and many subsequent studies have followed extending this analysis to elucidate HLA footprints in various populations and viral clades(182-185). Moreover, protective HLA alleles select for escape mutations that are deleterious for viral replication, especially when found in conserved viral proteins such as Gag(186-189). The presence of HLA-associated escape mutations illustrates that the cellular immune response puts sufficient pressure on the virus that it must adapt to survive. Escape mutations have also been associated with particular KIR/HLA allele combinations(181), and these data demonstrate that both CTLs and NK cells are potent effectors of the cellular immune response that are responsible for the influence of host immunogenetics on HIV disease course.

Though great strides have been made in elucidating the role of host immunogenetics, several important questions remain regarding the downstream

mechanisms by which they may act, and their ultimate influence on HIV-1 pathogenesis. It should be noted that the majority of protective HLA class I alleles are defined by their ability to control viral load (especially viral load set point). However, some recent data suggests that at least one HLA allele, HLA-B*81, is highly protective in terms of preventing CD4 decline, while only marginally impacting viral load(190). This implies that certain alleles may influence pathogenesis in a mechanism distinct from CTL killing of virus-infected cells, and that the influence of host HLA on HIV-mediated pathogenesis may be more complex. Additionally, HIV-specific CD4+ T cells have been associated with viral control(191), and the presence of cytolytic HIV-specific CD4+ T cells has been shown to delay disease progression(71). This suggest that HLA class II alleles, which present viral peptides recognized by CD4+ T cells, may have some influence on HIV control as well, and data to support this has recently emerged(192). The elucidation of additional protective HLA class II alleles will be important in order to broaden our knowledge of the immunogenetic underpinnings of HIV control. Of note, not all individuals with protective HLA alleles are guarded from disease, and conversely, many ECs do not carry any of the canonical protective HLA alleles. This suggests that there are additional contributing factors, whether derived from the host or the virus, and that much remains to be discovered concerning host control of HIV.

Though host factors, and more specifically host immunogenetics, have received the greatest attention in the area of researched aimed at elucidating the factors that modulate the severity of HIV-1 pathogenesis, it has been estimated that host factors can only account for roughly 13% of the observed variation in set point viral load (SPVL). This estimate can be extended to 22% when including more general host characteristics

such as age and gender(168, 193); however, this clearly demonstrates a gap in our knowledge and requires the elucidation of additional factors that significantly influence early SPVL. In a virus exhibiting such extreme genetic diversity as HIV, it is logical to speculate that HIV may carry a broad spectrum of viral characteristics and potentially, a broad spectrum of virulence. If this were true, then transmitted viral characteristics may play a considerable role in determining HIV pathogenesis and ultimate disease outcome. In support of this, it has been observed that viral load is a heritable characteristic, and viral loads in chronically infected donors are correlated to SPVLs in recipients in epidemiologically linked transmission pairs(193-195). This implies that certain viral characteristics are heritable and can have similar pathogenic effects in individuals with likely disparate immunogenetic backgrounds. However, the exact nature and quality of viral characteristics that explain the heritability of pathogenesis have remained more elusive, but significant recent progress has been made.

Two proposed viral characteristics that could explain the heritability of virulence are (I) the inherent ability of the virus to induce immune activation and (II) the replicative fitness of the transmitted virus(196). HIV-1 and is zoonotic precursor, SIV*cpz*, have lost the ability to down-regulate CD3 on T cells(197), which is a distinguishing feature of these viruses in comparison to other primate lentiviruses such as HIV-2 and its zoonotic ancestor, SIV*smm(198)*. This most likely occurred when SIVcpz acquired *vpu*, a viral gene encoding an accessory protein that antagonizes the action of tetherin, a host restriction factor that prevents the release of virions from infected cells(199). With the acquisition of vpu, another viral accessory protein, nef, subsequently lost the ability to down-regulate CD3, a cell surface molecule involved in the TCR-mediated activation of T

cells(197). This most likely lead to a virus that caused elevated levels of immune activation, a hallmark of HIV pathogenesis that is absent from SIVs infecting natural hosts. This observation, in a broad evolutionary sense, traces the heritability of pathogenesis between zoonotic transmissions. The nef protein is also responsible for additional host evasion mechanisms(200), which appear to be at least partially responsible for the pathogenic nature of HIV infection(201). Since the early days of the epidemic, some individuals who did not progress to AIDS were shown to be infected with viruses encoding defective *nef* genes(202, 203). This observation is mirrored in rhesus macaques infected with SIV lacking a functional *nef* gene, in which the majority of these animals spontaneous control viral load(204, 205). This establishes nef an important contributor to virulence, suggesting that heritability of nef function may play some role in the heritability of pathogenesis.

On the other hand, viral replicative fitness in vivo is a combination of at least two features, (I) the inherent replicative capacity of the virus as measured by the number of cells subsequently infected by each productively infected cell (typically measured in vitro and in the absence of an immune response) and (II) the ability of the host immune response to effectively suppress viral replication. The latter feature is modulated by viral escape, which occurs quickly after the advent of the cellular immune response during acute infection and is abetted by HIV's extreme genetic diversity. However, viral escape represents a trade-off between the need to evade adaptive immune pressure and the requirement to retain adequate replicative potential(113), as evidenced by the fact that many HLA-associated escape mutations, especially those associated with protective HLA alleles, can dramatically attenuate the replicative capacity of the virus(186-189). This

effectively illustrates the constant battle between the virus and the adaptive immune response, which often results in the imprinting of distinct features on the viral genotype that may subsequently have repercussions on viral phenotype. Since viral genotype is heritable, this suggests that viral escape mutations are also heritable and may represent an important transmitted viral characteristic. Indeed, Goepfert et al. reported that the number of transmitted escape mutations within well-defined CTL epitopes negatively correlated with SPVL in newly infected individuals, suggesting that viral escape in the donor, which is generally advantageous to the virus, can exert a protective effect when transmitted to a new host, especially a new host with a disparate immunogenetic background(206). Interestingly, this correlation was highly significant for escape mutations found in Gag, a structural protein that is functionally constrained, but not in Nef, an accessory viral protein more tolerant of sequence variation. These results point to viral adaptation as a heritable feature, and more importantly, implicate replicative capacity as the underlying viral characteristic based on the involvement of Gag. Further evidence supports replicative capacity, as defined by the *gag* gene, as an important viral characteristic influencing HIV pathogenesis, as gag-pro sequences isolated at chronic and acute time points from ECs display an attenuated replication phenotype(207, 208). Moreover, replication capacity defined by the *gag* gene also affects SPVL in non-controllers, suggesting replicative capacity has a broader influence on the spectrum of disease(209). However, the degree to which replicative capacity is heritable between transmission pairs, and the extent to which it affects pathogenesis remains to be fully elucidated. This is the subject of studies described in Chapter 2 and 3 of this thesis.

Gender effects present an additional layer of complexity that alters the severity of disease in many viral infections, including HIV. Men and women show a clear difference in their immune responses to infection. Females tend to exhibit more robust immune responses to both primary infection as well as vaccination, though they have also been shown to exhibit increased immune activation and exacerbated immunopathology(210-212). This is hypothesized to be the result of a more active and vigorous innate immune response, most likely due to the immunomodulatory effect of sex hormones such as estrogens(213, 214). These differences appear to play some role in the way the female and male immune responses interact with HIV. Several reports have conclusively shown that females exhibit significantly lower plasma VLs than males, both in the acute and chronic stages of HIV-1 infection, yet females are not protected from long-term CD4+ T cell decline or from developing AIDS(215-219). This is thought to be due to the increased innate immune activation in females, which, in turn, leads to elevated levels of cellular immune activation, a known risk factor for accelerated pathogenesis in HIV-1 infection(133, 155, 220-223). Plasmacytoid dendritic cells isolated from females produce significantly higher levels of IFNa than males when stimulated with HIV-derived toll-like receptor 7 (TLR7) agonists(224). Additionally, several ISGs were found to be expressed at significantly higher levels in females than males during the chronic stage of HIV-1 infection after correcting for plasma VL, consistent with a more activated innate phenotype in females given similar exposure to viral antigen(225). During the chronic stages of infection, CD8+ T cells isolated from females expressed higher levels of CD38 and HLA-DR(224), two markers of cellular immune activation that have been consistently associated with disease progression in chronic HIV-1 infection(133, 155,

223). These data establish a convincing causal link between differential innate immune responses and cellular activation in females during the chronic stages of infection, and may explain why females progress to AIDS with a similar trajectory as males despite having significantly lower VL.

Despite the accumulating data describing the sex differences inherent to chronic HIV-1 infection, much remains to be elucidated with regard to the acute stages of infection, particularly with respect to differential immune activation between the sexes. The preponderance of data generated to date relies on samples derived from cross-sectional cohorts, during the chronic stages of infection, with limited longitudinal follow up. Indeed, the bulk of the disparity in plasma viral load between males and females is observed during the acute stages of infection, and this disparity appears to wane over time(217), presumably again due to the increased levels of immune activation displayed by females in these stages of infection. It is therefore reasonable to postulate that early acute HIV-1 infection in females might be characterized by more favorable immune responses that initially control viremia, and that this immune response evolves over time to be less protective.

A recent study by Yue et al. elegantly demonstrates the cumulative nature of transmitted viral characteristics, host immunogenetics, and gender in determining SPVL in acutely infected Zambians. In a multivariable generalized linear model, donor viral load (a surrogate for transmitted viral characteristics), protective HLA alleles, and gender independently contribute to SPVL outcome(193). Taken together, it is clear that host immunogenetics, transmitted viral characteristics, and gender all play significant and independent roles in HIV-1 pathogenesis. However, it is also apparent that there is

significant space to advance our understanding regarding these three distinct factors. First, can we uncover novel HLA class I and II alleles that significantly affect HIV-1 pathogenesis, as specifically measured by CD4+ T cell decline, without using VL as a surrogate? Moreover, upon identification of these alleles, can we elucidate their underlying mechanism of action as well as the epitopes they target? Second, what is the full contribution of viral replicative capacity to HIV-1 pathogenesis, and how are viral characteristics dictated by host immunogenetics? Thirdly, what, if anything, is different in the early stages of HIV infection between men and women, and what events lead to the higher levels of chronic immune activation demonstrated in females? Answering these questions will significantly advance our understanding of HIV pathogenesis, both from the standpoint of the virus and the host, and will ultimately lead to the elucidation of correlates of protection, which will motivate the creation of novel therapeutics and inform rational vaccine design.

Summary

Though much work has been done to define the host immunogenetic characteristics that influence HIV-1 pathogenesis, much less is known regarding viral characteristics and their potential influence in disease progression. Additionally, the HIV field suffers from a lack of knowledge regarding the correlates of protection, which if defined could help to prevent the transmission of HIV or prevent the development of AIDS after HIV infection. Therefore, understanding the mechanisms by which HIV causes AIDS and the factors that influence this process will grant us valuable insight; insight that is desperately needed for the rational design of an HIV vaccine.

In chapter two, we first describe a new in vitro assay to measure the contribution of the Gag protein to HIV-1 replicative capacity. We subsequently employ this molecular cloning and in vitro replication assay to define amino acid polymorphisms that significantly affect Gag fitness. We furthermore, establish viral replicative capacity as a transmitted viral characteristic that, at least in part, explains the heritability of viral load between epidemiologically linked transmission pairs. Finally, we demonstrate that viral replicative capacity of the transmitted/founder virus significantly alters set point viral load as well as the trajectory of CD4+ T cell decline in newly infected individuals, demonstrating that transmission of highly replicating viruses leads to accelerated pathogenesis.

In chapter three, we delve further into the effects of viral replicative capacity in determining clinical outcome in newly infected individuals. We demonstrate that replicative capacity significantly alters the trajectory of CD4+ T cell decline in a manner independent of its association with set point viral load, and independent of protective HLA alleles, implying that transmitted viral characteristics are having an effect on HIV-1 pathogenesis before host factors play a significant role. Furthermore, we demonstrate that viral replicative capacity has a profound effect on virtually all facets of the early cellular immune response. High replicative capacity leads to increased immune activation, cellular exhaustion, and aberrant proliferation of T cells, all of which are associated with accelerated disease progression and are hallmarks of HIV pathogenesis. Moreover, we demonstrate that the magnitude of early viral replication impacts the size of the initial viral reservoir, specifically in vulnerable memory CD4+ T cell populations.

Finally, in chapter four, we describe novel advantageous HLA class I and class II alleles that protect individuals from CD4+ T cell decline without dramatically impacting early set point viral load. Additionally, we demonstrate that the protective HLA class I alleles exert their effect, at least in part, by preventing early gut damage and microbial translocation, which, as discussed above, has been shown to drive immune activation and exacerbate HIV pathogenesis and disease. This implies that gut damage occurs early, predisposes individuals to a more rapid disease course, and can be influenced earlier than expected by certain adaptive cellular immune responses.

This body of work as a whole provides new insights into the factors that modulate HIV-1 pathogenesis and, to some extent, the mechanisms by which they act, and opens new avenues for areas of study.

Chapter II

Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis

Jessica L. Prince*[1], Daniel T. Claiborne*[1], Jonathan M. Carlson[2], Malinda Schaefer[1], Tianwei Yu[3], Shabir Lakhi[4], Heather A. Prentice[5], Ling Yue[1], Sundaram A. Vishwanathan [1†], William Kilembe[4], Paul Goepfert[6], Matthew A. Price[7], Jill Gilmour[8], Joseph Mulenga[4], Paul Farmer[1], Cynthia A. Derdeyn[1,9], Jianming Tang[6], David Heckerman[2], Richard A. Kaslow[5], Susan A. Allen[4,9,10], and Eric Hunter[1,9]

1. Emory Vaccine Center at Yerkes National Primate Research Center, Emory University, Atlanta, Georgia, USA
2. Microsoft Research, Los Angeles, CA, USA
3. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA
4. Zambia-Emory HIV Research Project, Lusaka, Zambia
5. Department of Epidemiology, University of Alabama, Birmingham, USA
6. Department of Medicine, University of Alabama, Birmingham, USA
7. International AIDS Vaccine Initiative, San Francisco, CA, USA
8. International AIDS Vaccine Initiative, London, England
9. Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, USA

10. Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

* Daniel T. Claiborne and Jessica L. Prince contributed equally to this manuscript

Published in *PLoS Pathogens*, 2012, 8(11)

Abstract

Initial studies of 88 transmission pairs in the Zambia Emory HIV Research Project cohort demonstrated that the number of transmitted HLA-B associated polymorphisms in Gag, but not Nef, was negatively correlated to set point viral load (VL) in the newly infected partners. These results suggested that accumulation of CTL escape mutations in Gag might attenuate viral replication and provide a clinical benefit during early stages of infection. Using a novel approach, we have cloned *gag* sequences isolated from the earliest seroconversion plasma sample from the acutely infected recipient of 149 epidemiologically linked Zambian transmission pairs into a primary isolate, subtype C proviral vector, MJ4. We determined the replicative capacity (RC) of these Gag-MJ4 chimeras by infecting the GXR25 cell line and quantifying virion production in supernatants via a radiolabeled reverse transcriptase assay. We observed a statistically significant positive correlation between RC conferred by the transmitted Gag sequence and set point VL in newly infected individuals ($p = 0.02$). Furthermore, the RC of Gag-MJ4 chimeras also correlated with the VL of chronically infected donors near the estimated date of infection ($p = 0.01$), demonstrating that virus replication contributes to VL in both acute and chronic infection. These studies also allowed for the elucidation of novel sites in Gag associated with changes in RC, where rare mutations had the greatest effect on fitness. Although we observed both advantageous and deleterious rare mutations, the latter could point to vulnerable targets in the HIV-1 genome. Importantly, RC correlated significantly ($p = 0.029$) with the rate of CD4+ T cell decline over the first 3 years of infection in a manner that is partially independent of VL, suggesting that the

replication capacity of HIV-1 during the earliest stages of infection is a determinant of pathogenesis beyond what might be expected based on set point VL alone.

Introduction

Despite a diverse HIV-1 quasispecies within chronically infected individuals, a single variant establishes infection in the majority of heterosexual transmission cases, resulting in a severe genetic bottleneck [1-3]. A more profound understanding of the interaction between host and viral characteristics and how they shape early pathogenesis and disease progression will be integral for understanding the trajectory and impact of early events after heterosexual transmission. While it is well established that host factors such as HLA-class I alleles can play a major role in determining clinical progression in those individuals recently infected with HIV-1 [4-10] the role of transmitted viral characteristics has been understudied due to the lack of suitable cohorts in which virus from both the donor and linked recipient are available. Accordingly, studies using epidemiologically linked heterosexual transmission pairs are essential for understanding how viral evolution, adaptation, and the characteristics of the transmitted variant influence HIV-1 pathogenesis.

Previous studies in both heterosexual and homosexual transmission pairs have demonstrated that viral loads (VL) in the newly infected partners are correlated with VL in the transmitting partner [11-13]. This finding is intriguing as the majority of the linked couples have disparate HLA-I alleles, and would therefore have varying immune responses to the incoming virus. Thus, the characteristics of the virus in the donor quasispecies that impact replication can similarly impact the replication of the newly infecting virus even in the context of what is frequently a distinct immunogenetic environment. Indeed, when host factors known to modulate VL such as the presence of

protective or unfavorable HLA-I alleles, gender, age, and HLA-B sharing are taken into account in a multivariable analysis utilizing a general linearized model, the correlation between donor and recipient VL becomes highly significant (Yue *et al.*, manuscript in submission).

It is clear that both humoral [14] and cellular immune responses can drive virus evolution over the course of infection. In the case of the latter, escape mutations arise that abrogate the ability of cytotoxic T lymphocytes (CTL) to kill virus-infected cells [15-22]. While the selection of CTL escape mutations provides an *in vivo* advantage for the virus, if a mutation occurs within a functionally constrained region of the genome such as Gag, it could reduce *in vitro* replicative fitness [17,23-30]. This phenomenon has been demonstrated for several CTL escape mutations associated with protective alleles such as HLA-B*57, B*5801, B*27, and B*81 [31-36]. The ability of protective alleles to target conserved regions of the genome that escape with difficulty, due to the fitness costs incurred by mutations at these epitopes, may partially explain the mechanism of enhanced protection from disease progression in individuals with these alleles [37-41]. While evasion from the CTL response may result in such deleterious mutations, the *in vivo* fitness benefit outweighs that of the replication cost [42], and the ongoing selection of additional mutations may allow the virus to compensate for these defects [17,29,32,33,35,43-49]. The functional consequence of escape and compensatory mutations upon transmission to an HLA-mismatched individual has not been fully explored.

Initial studies in the ZEHRP cohort of 88 transmission pairs demonstrated that the number of transmitted HLA-B associated polymorphisms in Gag, but not Nef, was

negatively correlated to set point VL in the newly infected partners, suggesting that an accumulation of escape mutations might attenuate viral replication and provide a clinical benefit during early stages of infection [50]. In addition, in a smaller study of nine newly infected individuals infected by viruses with fitness reducing HLA-B*5703 associated epitope-escape mutations in p24, a lower set point VL was observed [24]. Previous studies have also shown that elite controllers can harbor viruses with *gag-pro* sequences that confer reduced *in vitro* replicative capacities (RC) [51-53]. In a series of studies, Brockman and colleagues have shown that *in vitro* RC conferred by *gag-pro* variants isolated from both subtype B and C chronically infected individuals correlates to VL, demonstrating the role of intrinsic viral characteristics in defining this marker of pathogenesis [43,47,54,55]. However, in these studies the recombination of population-based PCR amplified sequences into the lab adapted NL4.3 provirus required the outgrowth of virus in a CEM-CCR5 based cell line, potentially skewing the nature of the virus recovered. In contrast, studies of HIV-1 fitness in acute infection did not yield a statistically significant correlation between RC and VL, potentially due to small sample sizes and the limitations of the methodologies used.

The identification of 149 heterosexual epidemiologically linked transmission pairs from a discordant couple cohort in Lusaka, Zambia, provides a unique opportunity to investigate the role that HLA-mediated adaptation of Gag within a chronically infected individual plays in modulating the RC of the transmitted variant. We hypothesize that HLA-mediated adaptation of HIV-1 resulting in Gag variants conferring varying levels of RCs will be a major viral characteristic linking donor and recipient VLs, and that the *in*

vitro RC conferred by the transmitted Gag sequence defines early clinical parameters of HIV-1 pathogenesis.

To test this hypothesis and using a novel approach, we cloned *gag* sequences from the earliest seroconversion plasma sample from 149 newly infected recipients of epidemiologically linked Zambian transmission pairs into the MJ4 proviral backbone [56]. The RC of each Gag-MJ4 chimera was then determined and used to investigate how the RC conferred by the transmitted Gag sequence defines clinical parameters, such as early set point VL and CD4+ T cell decline in the newly infected individuals. These studies allowed us to identify novel residues in Gag that influence RC, and demonstrate a strong correlation between RC and early set point VL, as well as between RC and CD4 decline during the first three years of infection, which was also found to be independent of VL. Thus, the RC of the transmitted virus as defined by its *gag* gene influences critical aspects of HIV-1 pathogenesis.

Materials and Methods

Study subjects

All participants in the Zambia Emory HIV Research Project (ZEHRP) discordant couples cohort in Lusaka, Zambia were enrolled in human subjects protocols approved by both the University of Zambia Research Ethics Committee and the Emory University Institutional Review Board. Prior to enrollment, individuals received counseling and signed a written informed consent form agreeing to participate. The subjects selected from the cohort were initially HIV-1 serodiscordant partners in cohabiting heterosexual couples with subsequent intracouple (epidemiologically linked) HIV-1 transmission [57-59]. Epidemiological linkage was defined by phylogenetic analyses of HIV-1 *gp41* sequences from both partners [60]. Viral isolates from each partner in the transmission pair were closely related, with median and maximum nucleotide substitution rates of 1.5 and 4.0%, respectively. In contrast, median nucleotide substitution rate for unlinked HIV-1 C viruses from the Zambian cohort and elsewhere was 8.8% [60]. The algorithm used to determine the estimated date of infection (EDI) was previously described by Haaland *et al.* [2]. All patients in this cohort were antiretroviral therapy naïve. Zambian linked recipients were identified 45.5 days (median, IQR = 41.5 – 50.5) after the estimated date of infection, at which time plasma samples were obtained from both the transmitting partner (donor) and the seroconvertor (recipient). The vast majority (95%) of HIV-1 sequences derived from ZEHRP transmission pairs belonged to HIV-1 subtype C with subtypes A, D, G, and J being detected only occasionally [60]. All of the transmission pairs utilized in this study are infected with subtype C HIV-1.

Viral loads and CD4+ count measurements

Early set point VL for newly infected individuals was defined as the earliest stable nadir VL value measured between 3 and 9 months post infection and which did not show a significant increase in value within a 3-4 month window. HIV plasma VL was determined at the Emory Center for AIDS Research Virology Core Laboratory using the Amplicor HIV-1 Monitor Test (version 1.5; Roche). CD4+ T cell counts were based on T-cell immunophenotyping, with assays done using the FACScount System (Beckman Coulter Ltd., London, United Kingdom) in collaboration with the International AIDS Vaccine Initiative.

HLA-class I genotyping

Genomic DNA was extracted from whole blood or buffy coats (QIAamp blood kit; Qiagen). HLA class I genotyping relied on a combination of PCR-based techniques, involving sequence-specific primers (Invitrogen) and sequence-specific oligonucleotide probes (Innogenetics), as described previously [10]. Ambiguities were resolved by direct sequencing of three exons in each gene, using kits (Abbott Molecular, Inc.) designed for capillary electrophoresis and the ABI 3130*xl* DNA Analyzer (Applied Biosystems).

Amplification and sequencing of gag from donors and linked recipients

Viral RNA was extracted from 140 μL plasma samples using the Qiagen viral RNA extraction kit (Qiagen). *Gag-pol* population sequences were generated using nested gene specific primers. Combined RT-PCR and first round synthesis was performed using

SuperScript III Platinum One Step RT-PCR (Invitrogen) and 5 μL viral RNA template. RT-PCR and first round primers include GOF (forward) 5' ATTTGACTAGCGGAGGCTAGAA 3' and VifOR (RT-PCR and reverse) 5' TTCTACGGAGACTCCATGACCC 3'. Second round PCR was performed using Expand High Fidelity Enzyme (Roche) and 1 μL of the first round PCR product. Nested second round primers include GIF (forward) 5' TTTGACTAGCGGAGGCTAGAAGGA 3' and VifIR (reverse) 5' TCCTCTAATGGGATGTGTACTTCTGAAC 3'. Three positive amplicons per individual were pooled and purified via the Qiagen PCR purification kit (Qiagen). Purified products were sequenced by the University of Alabama at Birmingham DNA Sequencing Core. Sequence chromatograms were analyzed using Sequencher 5.0 (Gene Codes Corp.), and degenerate bases were denoted using the International Union of Pure and Applied Chemistry codes when minor peaks exceeded at least thirty percent of the major peak height.

The percent similarity between donor and recipient population *gag* sequences was determined by building a neighbor-joining tree using Geneious v5.5.7 (Biomatters Ltd.). The percent similarity between nucleotide and amino acid alignment was determined based on the output matrix from these neighbor-joining trees. In calculating the percent similarity between amino acid sequences, degenerate bases that resulted in non-synonymous changes and, thus, a mixture of amino acid residues, were translated as an "X". When one of the amino acids comprising a mixture in the donor was found in the recipient Gag sequence, this was counted as a mismatch, making the average percent similarity reported between donor and recipient *gag* sequences a maximal estimate of percent mismatch.

Generation of Gag-MJ4 chimeras

Viral RNA was extracted from linked recipients at the day of seroconversion time point using the Qiagen viral RNA extraction kit (Qiagen). First round PCR products were generated as was previously described for the *gag* sequencing of all 149 transmission pairs [50]. Second round PCR was performed to generate *gag* amplicons for Gag-MJ4 chimera generation using Phusion Hot Start II polymerase (Fisher) and 1 μL of the first round PCR product. Nested second round primers include GagInnerF1 (forward) 5' AGGCTAGAAGGAGAGAGATG 3' and BclIDegRev2 (reverse) 5' AGTATTTGATCATAYTGYYTYACTTTR 3', which generate a *gag* amplicon starting from the initiation codon of *gag* and extending 142 nucleotides after the *gag* stop codon and into *pro*. The 5' portion of the MJ4 long terminal repeat (LTR) was amplified using Phusion Hot Start II polymerase and the MJ4For1b (forward) 5' CGAAATCGGCAAAATCCC 3' and MJ4Rev (reverse) 5' CCCATCTCTCTCCTTCTAGC 3' primer set. In order to make the proper insert for cloning, the patient-specific *gag* and MJ4-LTR sequences were joined using a splice-overlap extension PCR utilizing the MJ4For1b (forward) and BclIRev (reverse) 5' TCTATAAGTATTTGATCATACTGTCTT 3' primer set. Joined splice-overlap-extension PCR products were gel purified using the Qiagen gel extraction kit (Qiagen). Purified Gag-LTR inserts and wild-type MJ4 vector (NIH AIDS Research and Reference Reagent Program, [56]) were digested with NgoMIV and BclI restriction enzymes (NEB) and ligated overnight at 4° C with T4 DNA ligase (Roche) at a 3:1 insert to vector ratio. JM109 competent cells were transformed with ligation products, plated onto LB/agar

plates supplemented with 100 μg/ml ampicillin and grown at 30° C. Gag-MJ4 chimeric DNA was isolated from cultures using the Qiagen miniprep kit (Qiagen). Gag-MJ4 chimeras were sequenced to confirm *gag* insert fidelity as compared to previously amplified population sequences. Two identical independent clones per linked recipient were chosen for replication assays in order to ensure backbone fidelity during the cloning process.

Generation and titration of viral stocks

Viral stocks were generated by transfecting 1.5 μg purified proviral plasmid DNA into 293T cells (American Type Culture Collection) using the Fugene HD transfection reagent (Roche) according the manufacturer's protocol. Viral stocks were collected 72 hrs post transfection, clarified by low-speed centrifugation, and frozen at -80° C. The titer of each viral stock was determined by infecting TZM-bl cells (NIH AIDS Research and Reference Reagent Program) with 5-fold serial dilutions of virus in a manner previously described [36,61].

Both 293T and TZM-bl cell lines were maintained in complete Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (HyClone Laboratories), 2mM L-glutamine, and 100 U/ml penicillin G sodium, and 100 μg/ml streptomycin sulfate (Gibco) at 37°C and 5% CO_2.

In vitro replication capacity (RC) assay

In order to assess the RC of Gag-MJ4 chimeras, 5×10^5 GXR25 cells [62] were infected at an MOI of 0.05, and 100 μl of viral supernatants were collected at 2 day

intervals. Briefly, GXR25 cells and virus were incubated with 5 μg/ml polybrene at 37°C for 3 hours, washed 5 times with complete Roswell Park Memorial Institute (RPMI) medium (Gibco) and plated into 24-well plates. Cells were split 1:2 to maintain confluency, replaced with an equal volume of fresh media, and viral supernatants were taken at days 2, 4, 6, and 8 as previously described [36,61]. Virion production was quantified using a ^{33}P-labeled reverse transcriptase assay. Based on values obtained for days 2-8, the optimal window for logarithmic growth for all viruses was determined to be between days 2 and 6, as by day 8 many high replicating viruses had exhausted target cells causing a flattening or decline of the replication curve. Therefore, $\log_{10}$-transformed slopes were calculated based on days 2, 4, and 6 for all viruses. Replication scores were generated by dividing the $\log_{10}$-transformed slope of the replication curve for each Gag-MJ4 chimera by the $\log_{10}$-transformed slope of wild-type MJ4. Two independent Gag-MJ4 chimera clones per acutely infected linked recipient were run in duplicate to confirm cloning fidelity. After both independent clones were confirmed to have identical replicative capacities, one clone was subsequently run in triplicate in two independent experiments in order to generate consistent replicative capacity scores. GXR25 cells were maintained in complete RPMI medium supplemented with 10% fetal bovine serum (HyClone Laboratories), 100 U/ml penicillin G sodium, 100 μg/ml streptomycin sulfate (Gibco), and 10mM HEPES buffer at 37°C and 5% CO_2.

Quantification of HIV-1 reverse transcriptase

Aliquots of culture supernatants from infected cells were added to an RT-PCR master mix [63] and incubated at 37°C for 2 hours; then the RT-PCR product was blotted

onto DE-81 paper, and allowed to dry. Blots were washed 5 times with 1X SSC (0.15 M NaCl, 0.015 M sodium citrate, pH 7.0) and 3 times with 90% ethanol, allowed to dry, and exposed to a phosphoscreen (Perkin Elmer) overnight. Counts were read using a Cyclone PhosphorImager [36,61].

HIV-1 Gag polymorphism scores

HLA-associated polymorphisms were defined as any non-consensus polymorphism that occurred at an amino acid position having known escape mutations adapted to specific HLA-class I alleles as defined by a list of associations generated in a manner similar to that described previously [44,64] from 1899 subtype C *gag* sequences from Zambia and South Africa (Carlson, *et al.*, manuscript in preparation). Polymorphisms that increased or decreased replicative capacity were defined based on amino acid associations with RC derived from an exploratory pair-wise analysis detailed in the experimental results. To generate a summed polymorphism score, the number of HLA-associated fitness-decreasing polymorphisms was subtracted from the number of HLA-associated fitness-increasing polymorphisms.

Statistical Analysis

The relationships between RC and set point VL, donor VL, and average CD4+ counts; RC and the number and quality of HLA-associated polymorphisms; and VL and the number and quality of HLA-associated polymorphisms were analyzed using the Spearman rank correlation. Linear regression analyses were utilized to generate trend

lines to facilitate visualization of correlation graphs. Mann-Whitney tests were used to compare the differences in RC between rare and more common polymorphisms. Mann-Whitney tests were used to analyze the difference in median set point VLs between different RC groups (RC<1, RC=1-2, RC>2). All Spearman correlations, Mann-Whitney tests, and linear regression analyses were performed using Prism GraphPad v5.0 (GraphPad Software, Inc.).

The Mann-Whitney U test was used to identify statistically significant differences in RC between two groups (e.g. sequences with or without a given HIV polymorphism). Multiple tests were addressed using q-values [65], which estimates the expected proportion of significant tests that are false positives. To limit the number of tests, we considered only groups containing at least 3 individuals.

A subset of volunteers with longitudinal CD4+ T cell counts (n=63) was analyzed to characterize the relationship between RC and T cell decline. Survival analysis was used to estimate the association between replicative capacity (RC) and the drop of CD4+ cell count. The endpoint is defined as the time before CD4+ T cell counts drop below a threshold, e.g. 300 or 350 cells/mm^3. The set point VL is another factor being considered. Kaplan Meier curve and log-rank test were used to compare the survival between the groups with RC<1 and RC>2. Cox proportional hazard regression was used to assess risk associated with high RC with or without adjustment for the confounding factor of set point VL. The sample size (n=66) was inadequate for a more complete analysis with additional covariates.

Results

Selection and characterization of transmitted gag sequences

Studying heterosexual transmission of HIV-1 within the context of discordant couples allows for full characterization of viruses from both donor and recipient. From a total of 294 epidemiologically linked transmission pairs identified from 1998-2010 in the Zambia Emory HIV Research Project (ZEHRP) discordant couple cohort [66], we selected 149 individuals that had been enrolled for at least 1 year and had at least 9 months of follow up post serconversion. The median time post estimated date of infection (EDI) for all 149 individuals was 45 days. Thus, this represents a unique early infection linked transmission pair cohort.

The median $\log_{10}$ VL for donors near the time of transmission and the median $\log_{10}$ set point VL for linked recipients was 5.02 (IQR = 4.51 – 5.45) and 4.39 (IQR = 3.91 – 4.99) respectively for all participants included in this study. Figure S1 depicts the phylogenetic clustering of population sequences for the *gag* gene of the 149 epidemiologically linked transmission pairs and highlights the high degree of sequence similarity between donor and recipient viruses.

The majority of population *gag* sequences isolated from acute/early time points of linked recipients were homogeneous and in most cases were identical to the donor population *gag* sequence isolated near the EDI. Overall, donor and linked recipient sequences differed in amino acid composition by only 2.7%. Mutations were counted when a mixture of nucleotides (amino acids) in the donor population sequence resolved to a single residue in the recipient and thus represent maximal values. Of the 149 pairs

only 37 recipient sequences had evidence of potential early escape and a majority of these individuals (28/37) had only a single amino acid change. Therefore, we can conclude that the majority of the sequence polymorphisms present at the seroconversion time point are derived from the chronically infected donor. Additional characteristics of the cohort including set point VL and CD4+ counts are listed in Table 1.

Construction of Gag-MJ4 chimeric viruses

Previous studies investigating the role of Gag viral fitness have employed a recombination approach in which sequences are PCR amplified as a bulk population, allowed to recombine into a *gag-deleted* NL4-3, and resulting viruses propagated in permissive cells [43,47,51,53-55]. This method has three distinct disadvantages: there is no control over the sites of recombination, it requires the outgrowth of virus which may select for the most fit virus in the population and could also select for sequence changes, and the introduction of subtype C sequences into a lab adapted subtype B proviral backbone may introduce biases due to the interaction of subtype B proteins with subtype C Gag. In order to avoid these limitations and because we are studying individuals recently infected with HIV, where the population is generally homogeneous, we have employed a direct cloning method that allows for the introduction of the entire *gag* gene into a replication competent, CCR5 tropic, clade C provirus, MJ4 [56].

A splice-overlap-extension PCR was employed to fuse the MJ4-LTR-U5 sequence with the transmitted *gag* sequence. This ensures that the cis-acting sequences upstream of Gag, which may influence expression levels, are constant throughout all constructs. The resulting chimeras include the entire transmitted *gag* sequence from the initiation codon

to the end of Gag and extend into conserved region of protease by 47 amino acids. For each newly infected individual, at least two independent Gag-MJ4 chimeras were sequence confirmed and assayed for replicative capacity (RC). An analysis of variation of RC between the two independent clones derived from each newly infected linked recipient was 8.5%. Testing two independent clones, therefore, ensures that the observed RC is not due to the confounding effect of backbone mutations that might have arisen during the cloning process and provides an estimate of experimentally induced variation.

Overall, a low amount of heterogeneity was detected in *gag* population sequences isolated from linked recipients, with only 28% having one or more mixed bases resulting in amino acid changes in the sequences from which the Gag-MJ4 chimeras were generated. When this was the case, the *gag* clone with the sequence closest to the donor *gag* sequence was chosen in order to avoid sampling *gags* containing *de novo* escape or reversion. In some cases in which multiple variants appeared to be transmitted, several *gag* variants were assayed for RC as described in the materials and methods section. In each case, these minor variants were found to have similar or identical RC values (data not shown).

Introduction of gag sequences from newly infected individuals into MJ4 drastically alters replicative capacity

In initial replication assays performed in order to test assay precision, wild-type MJ4 exhibited an intra-assay variability of 10.4% and an inter-assay variability of 8.7%. Figure 1 shows the results of a typical experiment for all 149 Gag-MJ4 chimeras, with wild-type MJ4 depicted in red. The normalized RC values of the chimeras ranged from

less than 0.01 to greater than 3.5. Some viruses replicated more than 100 times more efficiently than MJ4, demonstrating that substitution of Gag can have a profound impact on the ability of the virus to replicate in cells.

Correlation between the replicative capacities conferred by transmitted gag sequences and viral loads in newly infected individuals and their transmitting partners

An examination of the RC of transmitted viruses allows us to determine the role of viral replication in defining set point VL in acutely infected individuals before significant viral adaptation to immune pressure of the host has taken place, which might confound the relationship of RC to VL. We observed a statistically significant positive correlation between the replicative capacities of Gag-MJ4 chimeras and set point VLs in newly infected individuals (Figure 2A; Spearman correlation $r = 0.17$, $p = 0.02$), a correlation that persists when conditioning on the presence of B*57 in, and the sex of, the linked recipient ($p = 0.009$). This finding indicates that the RC conferred by the transmitted Gag sequence clearly plays a role in defining early set point VL of newly infected Zambian linked recipients.

In several cohorts VL in the transmitting partner and that in the linked seroconvertor have been shown to be correlated [11-13]. In order to more fully explore the possible contribution of RC in explaining this phenomenon, we compared VLs of the transmitting partner at the time of transmission to the RC defined by the transmitted Gag sequence. Despite both a higher maximum and wider range of VLs within transmitting partners, we observed a statistically significant positive correlation between RC and the set point VL of the donors, similar to that of their newly infected partners (Figure 2B;

Spearman correlation r = 0.18, p = 0.01). This supports the concept that RC, defined by Gag, is a viral characteristic contributing to the positive correlation between donor and recipient VLs that has been previously reported [11-13].

Several amino acids in Gag significantly correlate to changes in replicative capacity

Uncovering sites of vulnerability in HIV-1 is a high priority for the informed design of an effective HIV vaccine [21]. Therefore, we examined all 149 Gag sequences and their RC using an exploratory pairwise analysis described previously [43,54], to uncover residues that significantly affect the virus' ability to replicate *in vitro*. We found 49 residues at 31 unique positions that had a statistically significant effect on RC at $p < 0.05$ ($q < 0.51$) and 4 residues at 3 unique positions that were significant at $p<0.002$ ($q<0.2$) (Table S1). In what follows, we will use $q<0.2$ as the significance threshold when individual sites of significance are considered, and $q<0.51$ ($p<0.05$) as the significance threshold when we are testing broad trends, in which we are willing to increase our expected false positive rate as a tradeoff to substantially reduce our expected false negative rate.

The locations of all statistically significant polymorphisms ($p < 0.05$), along with their effects on RC as compared to the median RC of all viruses, are plotted linearly on a graphical representation of the Gag protein (Figure 3A.). Residues that dramatically modulate RC were enriched in p17 and p2 (Fisher's exact test, $p<0.0001$). In addition, roughly two-thirds of the non-consensus residues with $p<0.05$ increase fitness relative to the median RC for the entire population.

An expanded data set of 1899 subtype C *gag* sequences from Zambia and South Africa (Carlson, Schaefer et. al., manuscript in preparation) was utilized to identify residues that affected RC and were also HLA-associated. Within this data set, HLA-associated polymorphisms are classified as being either adapted or non-adapted. An adapted residue is one that is escaped relative to the HLA-allele in question. In contrast, a residue that is non-adapted is the susceptible form and may render the virus vulnerable to immunological targeting by the HLA-allele in question. This new dataset has identified a total of 199 HLA-linked polymorphisms ($q<0.2$, $p<0.0007$) vs. 59 associations utilized previously from a smaller subset of *gag* sequences [27,50]. Within the 49 residues associated with changes in RC, 7 polymorphisms were found to be adapted to specific HLA class I alleles, clearly demonstrating the impact of the cellular immune response in affecting viral fitness (Figure 3B, * denotes $q< 0.2$). Six polymorphisms were found to be non-adapted to specific HLA class I alleles (Figure 3C, * denotes $q < 0.2$). Since these are non-consensus polymorphisms, it is possible that consensus at these residues is escaped relative to these HLAs, potentially explaining why an adapted consensus residue at this position is the less fit variant. Indeed, 5 consensus residues (62K, 451S, 488S, 85L, and 309A) with $p<0.05$ were found to be adapted to HLA-I alleles, demonstrating that the cellular immune response can drive selection for consensus residues.

Rare polymorphisms have the greatest effect on replicative capacity

During our analysis of amino acid polymorphisms linked to changes in RC ($p < 0.05$), we observed a negative correlation between the frequency of polymorphisms and the magnitude of their effect on RC (Spearman correlation, $r = -0.89$, $p < 0.0001$). Indeed,

rare polymorphisms, those occurring in less than 10 of the 149 individuals studied, had significantly greater impact (both negative and positive) on RC than polymorphisms that occurred more frequently (Figure 4A and 4B). This finding is especially intriguing in the case of rare deleterious mutations, as these residues may highlight epitopes at which HIV escapes or compensates for fitness defects with great difficulty, similar to those described for elite controllers [52], and may, therefore, be attractive targets for a cellular-based vaccine.

The cumulative and qualitative effect of HLA associated polymorphisms in Gag on replicative capacity

In order to investigate whether there is a cumulative effect of viral escape from cellular immune pressure in Gag on RC, the expanded dataset of HLA-associated polymorphisms generated from an analysis of 1899 *gag* sequences from Zambia and South Africa (Carlson, Schaefer *et al.*, manuscript in preparation), described above, was employed. The number of non-consensus polymorphisms located at HLA-associated positions was determined for each MJ4 chimera and then correlated with the RC defined by those Gag sequences. Surprisingly, we found a positive association between the number of HLA-associated polymorphisms and RC (Fig 5A; r = 0.14, p = 0.05). Although counterintuitive, this is consistent with the fact that not all HLA-associated polymorphisms within a particular Gag sequence will necessarily reduce fitness. We have shown in the previous sections that several non-adapted (or "susceptible" to HLA pressure) HLA-associated polymorphisms increase fitness relative to the median of all sequences. Indeed, we observe a highly statistically significant positive correlation

between the number of non-adapted HLA-associated polymorphisms and RC (Spearman correlation, $r = 0.23$, $p = 0.003$; data not shown). Thus, the inclusion of both adapted (or escaped with respect to specific HLA alleles) and non-adapted polymorphisms within this expanded HLA-associated dataset may explain the observed positive association between numbers of HLA-associated polymorphisms and RC. Therefore, we hypothesize that it is the balance and interaction of both fitness increasing and fitness decreasing polymorphisms within a particular sequence that ultimately determines the RC of the virus.

In order to more accurately determine how the number and quality of HLA-associated polymorphisms affects RC and to correct for the opposing influence of both increasing and decreasing polymorphisms within a particular sequence, a summed polymorphism score was calculated by assigning fitness increasing polymorphisms a score of +1, fitness decreasing polymorphisms a score of -1, and neutral polymorphisms a score of 0. HLA-associated polymorphisms were defined as being positive, negative, or neutral based on the previously described univariate analysis that correlated specific residues within our 149 sequences with changes in RC. In this modified analysis, we observed a highly statistically significant correlation between the summed polymorphism score and RC (Figure 5B: Spearman rank correlation; $r = 0.6$, $p = <0.0001$), confirming that the sequence features are approximately independent of each other and suggesting that the offsetting influence of fitness decreasing and increasing polymorphisms is a strong contributor to RC. This finding may explain the observation that, in general, the most-fit viruses are less like the consensus subtype C Gag sequence, consistent with a majority of polymorphisms increasing fitness (Figure S2, [55]).

The cumulative and qualitative effect of HLA-associated polymorphisms in Gag on set point viral load

In a previous report using 88 Zambian linked seroconverters, we reported that increasing numbers of transmitted HLA-B associated polymorphisms within or adjacent to well defined epitopes were associated with lower set point VLs [50]. When we expand this analysis to include all 149 Zambian linked recipients and use the same dataset of HLA-linked polymorphisms used by Goepfert *et al.* [50] we observe the same correlation ($r = -0.15$, $p = 0.03$, Figure 6A). However, when we use the expanded HLA-associated data set (199 associations) to define HLA-associated polymorphisms, we no longer observe a statistically significant negative association between the number of transmitted HLA-associated polymorphisms in Gag and set point VL (Figure 6B).

We therefore hypothesized that, as with RC, this correlation between the total number of transmitted HLA-associated polymorphisms in Gag and set point VL in newly infected individuals may be confounded by not taking into account whether polymorphisms increase or decrease fitness. Indeed, using the summed polymorphism score, we observe a highly significant correlation between the summed score of HLA-associated polymorphisms and set point VL (Figure 6C: $r = 0.21$, $p = 0.006$). This demonstrates that it is not merely the quantity of HLA-associated polymorphisms present in the transmitted Gag sequence that ultimately defines set point VL, but it is the influence of both fitness increasing and decreasing polymorphisms that contributes to RC and in turn set point VL in newly infected individuals.

Transmission of viruses with low replicative capacities provides recipients with a longer-term clinical benefit

Though set point VL has been shown to be a relevant marker for disease progression [67,68], CD4+ T cell counts are traditionally used to define those individuals that have progressed to AIDS and are at a higher risk for opportunistic infections [69,70]. Therefore, we analyzed a subset of individuals (n=66) for whom longitudinal CD4+ T cells counts for at least one-year post-infection are available. We observed a statistically significant correlation between the average CD4+ T cell counts and the replicative capacities of Gag-MJ4 chimeras (Figure 7A, Spearman correlation, $r = -0.24$, $p = 0.02$). This demonstrates that infection with attenuated viruses may impart some survival benefit to newly infected individuals, at least within the first year of infection.

Subsequently, we determined whether individuals infected with poorly replicating viruses exhibit differential pathogenesis over the first three years of infection or whether this early benefit is transient and quickly lost. To answer this question, we studied a subset of the linked recipients (n=63) for whom CD4+ T cell counts were available at regular three month intervals for greater than one year post-infection.

In a Kaplan-Meier survival analysis, in which we defined the endpoint as having a CD4+ T cell count >350 cells/mm^3 (WHO recommendation for initiation of anti-retroviral therapy [71]), we observed a statistically significant difference in the number of individuals that maintain CD4+ counts >350 cells/mm^3 between those infected with viruses that replicate very poorly (RC<1) and those infected with highly replicating viruses (RC>2), within the first 3 years of infection (Figure 7B, Mantel-Cox test p =

0.029). This disparity in disease progression was even more pronounced when the endpoint was defined as having CD4+ T cell counts >300 cells/mm^3, demonstrating a median difference of 896 days before falling below the CD4+ count cut off between individuals infected with low and high replicating viruses (Figure 7C, Mantel-Cox test p = 0.014). Using a Cox proportional hazard model, we demonstrate a significantly increased risk of CD4+ T cell counts falling below 350 (Hazard Ratio (HR) 2.36; p = 0.034) or 300 (HR 3.80; p = 0.021) over the first three years of infection for individuals whose Gag conferred an RC>2 vs. RC<1.

Interestingly, the benefit conferred by low replicating viruses could not be wholly explained by differences in set point VL within this smaller data set. Although there was a trend towards higher VLs between the two most disparate groups, with a 2.5 fold difference in median VLs (Figure 7D), we observed no statistically significant differences in median set point VLs between individuals infected with low (RC<1), medium (RC= 1-2), and highly (RC>2) replicating viruses. Further, in Cox proportional hazard models that take into account VL, the HR remained high (2.18 and 3.12 respectively) and p values continued to trend or remain borderline significant (p = 0.093 and 0.051) (Table 2), indicating that both VL and RC can independently affect CD4 loss. Moreover, the HR associated with log10 increases in set point VL alone was lower than that for RC alone (HR = 1.75 versus 2.62, and 2.09 versus 3.80; CD4 <350 and 300 respectively; Table 2). These results suggest that infection with a low replicating virus confers clinical benefit outside of the effect of RC on set point VL, and that the kinetics of viral replication early in infection can ultimately dictate long-term pathogenesis.

Discussion

In this study of 149 linked Zambian heterosexual transmission pairs from the ZEHRP discordant couple cohort we have more fully characterized the role that HIV-1 viral characteristics, in particular the replicative capacity (RC) conferred by the transmitted *gag* sequence, plays in defining parameters of early HIV-1 pathogenesis. Identification and comparison of both donor and recipient *gag* sequences for all 149 transmission pairs revealed a high degree of similarity (97.6%) within each linked pair, allowing us to conclude that the majority of polymorphisms in Gag present at the seroconversion time point were transmitted from the chronically infected donor.

Transmitted gag sequences from newly infected individuals engineered into MJ4 drastically alters replicative capacity

Since all of the transmission pairs in this study were infected with subtype C viruses, our approach of precisely cloning *gag* genes from acutely-infected recipients into a primary isolate (MJ4) provirus has many important advantages over previously employed methods. MJ4 is a CCR5 tropic infectious molecular clone derived from a subtype C clinical isolate from Botswana, providing greater homology to viruses circulating within the Zambian population than other previously used subtype B lab-adapted strains [33,47,54,55]. Additionally, this cloning method for generating Gag-MJ4 chimeric viruses does not rely on recombination based technologies that require the outgrowth of viral quasispecies, which may select for the most fit virus, and in some cases, amino acid changes in the viral stocks that are not present in the individual from which

they were derived [32,47,53-55]. The use of a common BclI restriction site located 137 nucleotides after the *gag* stop codon in MJ4 does result in a chimeric protease, however, this region is 96.5% conserved in this cohort and we did not observe a high prevalence of dead or inactive Gag-MJ4 chimeras.

The impact of engineering foreign *gag* sequences into MJ4 on virus replication was highly significant, with many of the chimeras exhibiting RC values greater than a hundred-fold higher than wild-type MJ4, which in this assay is one of the poorer replicators. This indicates that substitution of Gag can drastically alter the *in vitro* RC of the virus when all other viral components are constant. Multiple intra-molecular contacts as well as host protein interactions in Gag are necessary for effective intracellular Gag trafficking [72,73], particle formation [74], budding [75,76], maturation [77] and disassembly [78]. Therefore, immune mediated adaptation of this functionally constrained protein could have clear consequences for viral replication through disruption of these many interactions.

Correlation between the replicative capacities conferred by transmitted gag sequences and viral loads in newly infected individuals and their transmitting partners

It has been well established that the set point VL in those recently infected with HIV-1 is correlated to disease progression and clinical outcome [67,68]. Previous data from our group demonstrated that transmission of sequences with increasing numbers of CTL escape mutations in Gag resulted in lower set point VLs in newly infected individuals, a finding that suggested that transmitted HLA-associated polymorphisms in Gag might negatively affect viral replication [50]. We have confirmed this association in

the current study after increasing the number of transmission pairs analyzed from 88 to 149. This result is consistent with studies by Brockman *et al.*, which have demonstrated a statistically significant link between the RC conferred by *gag-pro* genes in subtype B and C chronically infected individuals to VL [43,47,54,55]. However, a statistically significant correlation between Gag RC derived from acutely infected individuals and set point VL has not previously been definitively reported in a subtype C cohort.

In contrast, in this large group of very early ZEHRP seroconvertors (with samples drawn a median of 45 days post-EDI) we observed a clear statistically significant correlation (p=0.02) between the RC conferred by the transmitted *gag* sequence and the early set point VL in newly infected individuals. This result implies that RC plays a role in defining the overall level of virus replication during the first year of infection. Moreover, in multivariable analyses that take into account the early viral control imposed by the B*57 allele and by gender, the impact of RC on set point VL was found to be independent of these two host factors (p=0.009). Other factors such as NK cells and restriction factors such as TRIM or APOBEC may potentially affect RC and VL, however little is currently known regarding these potential effects, and future efforts should evaluate the role of such factors. While we observed a statistically significant positive correlation between RC and set point VL, outliers in the data exist that do not fit the trend, and in some cases can be explained by the presence of protective HLA-alleles or by a large number of escape mutations present in the transmitted sequence that are relevant to the HLA background of the newly infected individual. Set point VL is clearly determined by a combination of both host factors, including HLA-alleles, and viral factors such as RC, and this may explain the differences in the absolute correlation for each individual.

The RC of Gag-MJ4 chimeras also correlated with VLs near the estimated date of infection in chronically infected donors, consistent with the previously reported observation that donor and recipient VLs are correlated within epidemiologically linked transmission pairs [11-13]. The data presented here would suggest that the RC conferred by the transmitted Gag sequence is a contributing viral characteristic of that donor virus responsible for influencing early set point VL in the newly infected partner.

Several amino acids in Gag significantly correlate to changes in replicative capacity

In a pair-wise analysis, a large number of residues were associated with changes in fitness ($p<0.05$, $q<0.51$), with 4 residues at 3 unique positions at $q<0.2$. These residues included the polymorphisms 30R and 31I in p17 (MA), and 309S in p24 (CA). However, in an exploratory analysis of those residues associated with changes in RC with a p value <0.05, it was clear that associated polymorphisms were noticeably enriched on a per residue basis in p17 and p2 (Fisher's exact test, $p<0.001$). The former plays critical roles in intracellular trafficking, and membrane association of Gag [73,79,80], while the latter is an important structural element involved in formation of the immature protein shell [81,82] and a target of the novel drug Bevirimat during maturational cleavage of the Gag precursor [83,84]. Surprisingly, only one third of the associated mutations negatively affected virus replication, while nearly two-thirds of the associations increased fitness. Some of these fitness-increasing mutations represent adapted polymorphisms (i.e. selected as immune escape) and in terms of vaccine design it may be important to avoid the inclusion of such epitopes.

Polymorphisms positively or negatively affecting replication in the p24 region of Gag were limited to just six residues (4 positive, 2 negative), in accordance with the conserved nature of this protein. Surprisingly, none of the canonical B*57/B*5801 associated escape mutations within p24, whose fitness defects have been well documented [24,31-34], were found to be significantly associated with decreases in RC in our present study. This may be due to the high prevalence of B*57/B*5801 positive individuals within this cohort (25%), which could promote viral adaptation to these alleles over time through compensatory mutations [22]. It is also possible that some fitness defects such as those associated with T242N within the TW10 epitope might be missed in the current study, as a previous study has shown that it is cell-type dependent [32].

The most deleterious HLA-associated mutation that we observed was K12E, which reduced RC by almost 10-fold relative to the median RC of the cohort. This polymorphism is found quite rarely in the population (3 out of 149), and is statistically associated with HLA-A*74, an allele found to be highly protective in both this Zambian subtype C cohort as well as others [9,85]. The protective effect of A*74 has recently been demonstrated to be independent of HLA-B*57 [86]. The negative *in vitro* impact of mutations at residue 12 on replication is supported by a longitudinal study of a subset of this seroconvertor cohort (n=81), in whom polymorphisms at residue 12 were found to revert at a high rate (25%/yr), over the first two years of infection (Schaefer *et al.*, manuscript in preparation). Furthermore, in this same study, escape at position 12 occurred only once and at 24 months post-infection in a total of ten A*74 positive individuals, confirming the high fitness cost associated with CTL-induced escape mutations at this position. We hypothesize that the targeting of this putative epitope,

KR9 [86], may account for part of the protective effect conferred by A*74 and indicates that protective immune responses can target regions of Gag outside of p24. While the nature of the replication defect in viruses encoding K12E remains to be determined, this residue does lie in the highly basic region at the N-terminus of p17 (MA), which is involved in membrane targeting and membrane association of Gag [80,87,88].

Rare polymorphisms have the greatest effect on replicative capacity

Rare mutations, such as K12E, which occur in a small subset of the population studied here (less than 10 individuals of the 149), affected fitness to a statistically greater degree than more common polymorphisms. Rare fitness decreasing mutations are likely unique to specific circumstances such as those where a considerable decrease in RC is warranted in the face of a very effective cellular immune response that is largely abrogated upon mutation. Such mutations have been found to subsequently revert after transmission to individuals lacking the selecting HLA-allele [17,42,89] and in whom they now confer a fitness deficit for the virus. These sites of rare fitness reducing polymorphisms may emphasize vulnerable epitopes at which HIV-1 escapes from immune pressure with great difficulty. Alternatively, it is possible that, when escape occurs, it is consistently associated with a decrease in RC that cannot be completely compensated.

A similar observation was made for rare mutations that greatly increase RC. Global compensatory mutations do exist that can compensate multiple deleterious mutations, such as those within the cyclophilin binding loop [42,46]. Some of the rare fitness increasing mutations may be of this type, although those reported previously have

generally been quite common in the population. Compensatory mutations can also be secondary site-suppressors of deleterious mutations [90]. Frequently, such mutations are only conditionally beneficial and can be deleterious in a different context, which could explain why some fitness increasing mutations are rare. It is also possible that these mutations do carry some unrecognized *in vivo* fitness cost that cannot be captured in the *in vitro* replication system used here. Due to the fact that these mutations are rare, they are difficult to statistically link to HLA alleles or to link to other residues with which they may covary, making the potential fitness defects that these mutations mitigate difficult to elucidate.

The cumulative and qualitative effect of HLA associated polymorphisms in Gag on replicative capacity and VL

A key goal of this study was to understand how the cellular immune response might select for mutations in Gag that reduce viral RC, and while identification of specific amino acid polymorphisms that either increase or decrease fitness can be informative, it is equally important to elucidate how the accumulation of specific HLA-associated polymorphisms in Gag affects both RC of the virus and VL in the newly infected person. Previous efforts to correlate the total number of HLA-associated polymorphisms in Gag to RC have yielded inconclusive results [32,54], perhaps because the quality of the polymorphisms in question was not considered. Using an expanded list of HLA-associated polymorphisms (Carlson, Schaefer *et al.*, manuscript in preparation) we report a weak positive correlation between the total number of HLA-associated polymorphisms in Gag and RC. The fact that this correlation was positive is consistent with our

observation that a large fraction of the non-consensus HLA-associated polymorphisms increased RC. In particular, in the expanded data set of HLA-associated polymorphisms, we observed that non-adapted residues, which would be predicted to render the virus susceptible to the linked HLA allele, were statistically associated with increased fitness.

These findings suggest that CTL escape mutations, which decrease the overall RC of the virus, are being driven to consensus as a result of population level immune pressure. In the absence of immune pressure, the non-escaped (non-adapted) residues would be expected to predominate, but if they render the virus susceptible to a large portion of the population, then the consensus residue will be escaped rather than susceptible, despite reducing *in vitro* fitness. This is consistent with the findings of Kawashima *et al.* [22] that the frequency of certain HLA-class I alleles within a particular population can influence the fixation of escape mutations in the overall population. Moreover, Wright *et al.* [55] showed that Gag-NL43 recombinant viruses encoding *gag-pro* sequences most disparate from the subtype C consensus *gag-pro* sequence had statistically higher replicative capacities than their more consensus-like counterparts, and this finding has been recapitulated in this current study. Taken together, these data suggest that overall, HLA-mediated adaptation is driving the fixation of consensus residues that are less fit than their susceptible counter-parts.

When we account for this ability of HLA-associated polymorphisms to either increase or decrease fitness by assigning a summed polymorphism score, which subtracts the number of fitness decreasing polymorphisms from the number of fitness increasing polymorphisms in a particular sequence, we find a highly statistically significant correlation between RC and the summed polymorphism score ($p<0.0001$). Although this

p-value should be interpreted cautiously, since it reflects the summation of features previously identified to be correlated with RC, the data do suggest that the effect of polymorphisms is cumulative, and that as a Gag sequence accumulates an excess of fitness-reducing polymorphisms, the RC decreases proportionally. Similarly, utilization of a summed polymorphism score improved previously reported correlations between the total number of HLA-associated polymorphisms in Gag and set point VL in newly infected individuals [50]. We observed a highly statistically significantly correlation (p = 0.006) between the summed polymorphism score and set point VL in newly infected individuals. Just as this balance of fitness increasing and decreasing polymorphisms impacts RC, it simultaneously influences the set point VL of the newly infected individual.

Transmission of viruses with low RCs provides recipients with a longer-term clinical benefit

While VL has been demonstrated to influence the rate of disease progression in HIV-1 infected individuals [68,69], it is possible that, during the very earliest stages of infection and before host immune control, the replication rate of the virus may affect the rate of future damage to the immune system. Indeed, we observed a statistically significant negative correlation between RC and average CD4 counts for the first year post infection, suggesting a role for RC in defining this important parameter of pathogenesis at early stages after infection. However, it is possible that this early benefit could be quickly lost due to further adaptation of the virus to the new host's immunogenetic background and further compensation for *de novo* escape mutations.

Consequently, we analyzed individuals with longitudinal CD4 counts out to three-years post infection in order to determine if the observed early benefit was sustained in early chronic stages of infection. Using Kaplan-Meier survival analyses to examine the relative time for individuals infected with viruses encoding *gag* genes conferring RC values of <1 and >2 to reach CD4 T cell counts of 350 after 3 years of infection, we observed a clear and statistically significant difference. This was even more striking when CD4 counts less than 300 were used as the endpoint. Moreover a Cox proportional hazard model demonstrated a significantly increased risk of CD4 counts falling below both 350 (HR 2.36) or 300 (HR 3.80) over the first three years of infection for individuals whose *gag* gene conferred an RC>2 vs. RC<1. These findings indicate that the RC conferred by the transmitted Gag sequence may have profound and prolonged effects on HIV-1 pathogenesis from acute to early chronic stages of infection.

While RC and VL are correlated in the full data set, set point VL does not fully explain the effect of RC on CD4, because we did not observe any statistically significant differences in set point VL between the two groups (RC<1 and RC>2) for the subset of individuals with CD4+T cell counts (n=63). Moreover, in Cox proportional hazard models which take into account VL, the HR remained high (2.17 and 3.11 respectively) and p values continued to trend or remain borderline significant (p=0.093 and 0.051). This suggests that both VL and RC have independent effects on CD4 decline, however, because this analysis was conducted on a subset of less than half of our initial cohort, additional work is underway to further confirm and extend these results.

It seems possible, therefore, that the RC of the transmitted variant may initiate crucial events, early in infection and dissemination, that dictate both acute and later stage

pathogenesis regardless of the ability of the immune system to control viral replication down to set point. Infection with highly replicating variants could lead to a more complete depletion of central memory CD4+ T cell pools at this early time that could predispose an individual to more rapid CD4+ T cell loss, irrespective of adequate control of viral replication. This is evidenced in a few individuals infected with highly replicating Gag variants (RC>2), who go on to control VL to a low set point, but whose CD4+ T cells counts rapidly drop below 300 (data not shown). Additionally, a high level of peak viremia or initial high antigen loads could establish an inflammatory environment that leads to sustained immune activation, which has been implicated as a more reliable marker for disease progression [91]. These possibilities are the focus of ongoing work.

In summary, using an *in vitro* approach to define the impact of polymorphisms in Gag on transmitted virus RC has clearly shown that this property of the virus is a significant contributor to early set point VL in a newly infected individual. More importantly, however, these studies suggest a critical role for RC in defining the trajectory of immune depletion and pathogenesis, beyond simply its impact on VL, and highlight the importance of the very earliest events in virus-host interactions. It also raises the possibility that a vaccine that can attenuate early virus replication would have a positive impact both on vaccinated individuals, as well as non-vaccinated individuals by weakening the transmitted/founder virus and increasing the likelihood of transmission of low replicating variants.

Acknowledgements

The investigators thank all the volunteers in Zambia who participated in this study and all the staff at the Zambia Emory HIV Research Project in Lusaka who made this study possible. The investigators would like to thank Jon Allen, Smita Chavan, and Mackenzie Hurlston for technical assistance and sample management. We would also like to thank Dr. Mark Brockman for his discussions and generous donation of the GXR25 cells.

Table 1. Cohort statistics generated from the 149 transmission pairs selected from the ZEHRP cohort.

Parameters	Linked Seroconverter (SC)
Number of individuals	149
Number of Females (%)	77 (52%)
Age of SC at time of seroconversion	28 (24 – 35)[a]
Estimated days post infection	45.5 (41.5 – 50.5)[a]
Log_{10} Set-point VL of SC	4.39 (3.91 – 4.99)[a]
Log_{10} VL of Donor Partner[b]	5.02 (4.51 – 5.45)[a]
CD4 cell count 1yr post SC[c]	383 (308 – 503)[a]

[a] Median (Inter-Quartile Range)
[b] Measured at or within one month of seroconversion of LR
[c] LR for whom we have longitudinal CD4+ T cell counts (n=63)

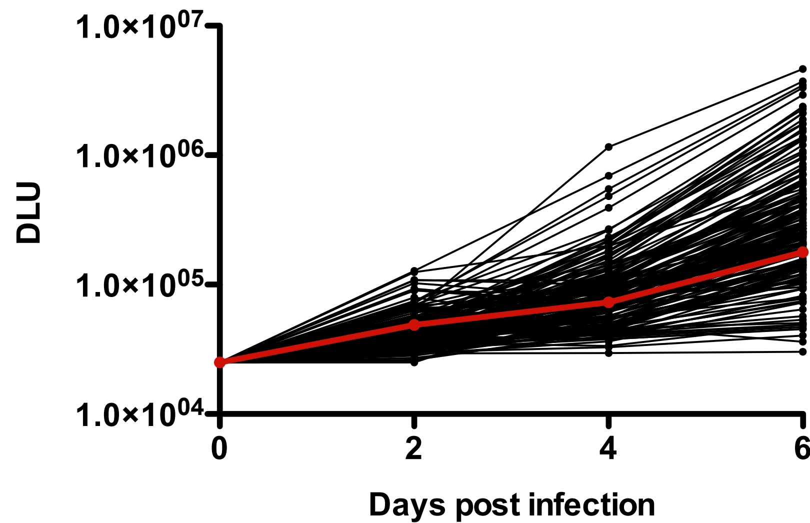


Figure 1. Insertion of the *gag* gene from newly infected individuals dramatically alters the replicative capacity of MJ4 In order to generate replication curves, 5 x 10^5 GXR25 cells were infected with Gag-MJ4 chimeras at an MOI of 0.05. Viral supernatants (100 µL) were collected on days 2, 4, and 6, and the amount of virus in supernatants was quantified using a radiolabeled reverse transcriptase assay. Replicative capacity scores were generated by dividing the $\log_{10}$-transformed slope of replication of each Gag-MJ4 chimera by the $\log_{10}$-transformed slope of MJ4. It is clear that the insertion of the *gag* gene alone can dramatically alter the *in vitro* RC, generating viruses with up to 1000-fold different replication curves. DLU, digital light units.

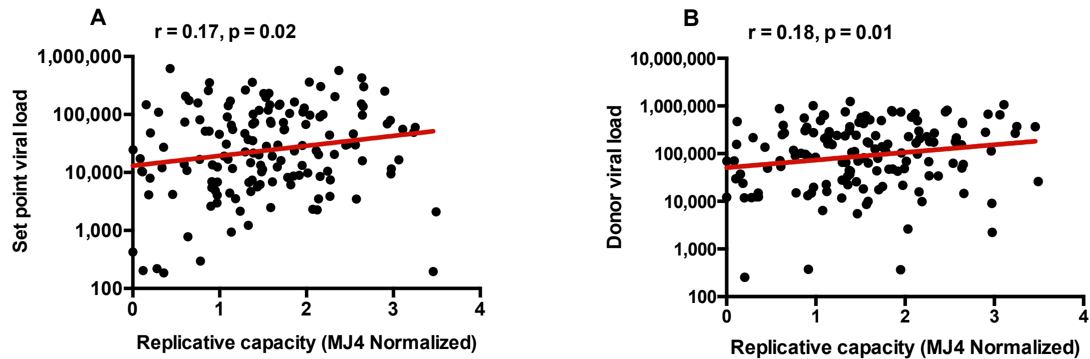


Figure 2. Replicative capacity is correlated to viral load in recipients and donors

(A) The RC of Gag-MJ4 chimeras generated from *gag* sequences isolated from epidemiologically linked recipients at acute time points correlates to early set point VL in the same acutely infected recipients (Spearman correlation, $r = 0.17$, $p = 0.02$). Replicative capacity scores were generated by normalizing the $\log_{10}$-transformed slopes of replication curves from days 2 through 6 for each Gag-MJ4 chimeric virus to the $\log_{10}$-transformed slope of wild-type MJ4. (B) The RC of Gag-MJ4 chimeric viruses also correlates to the VL near the estimated date of infection in chronically infected donors (Spearman correlation, $r = 0.18$, $p = 0.01$). Trend lines were generated using linear regression analysis, and are shown in order to facilitate visualization of correlations.

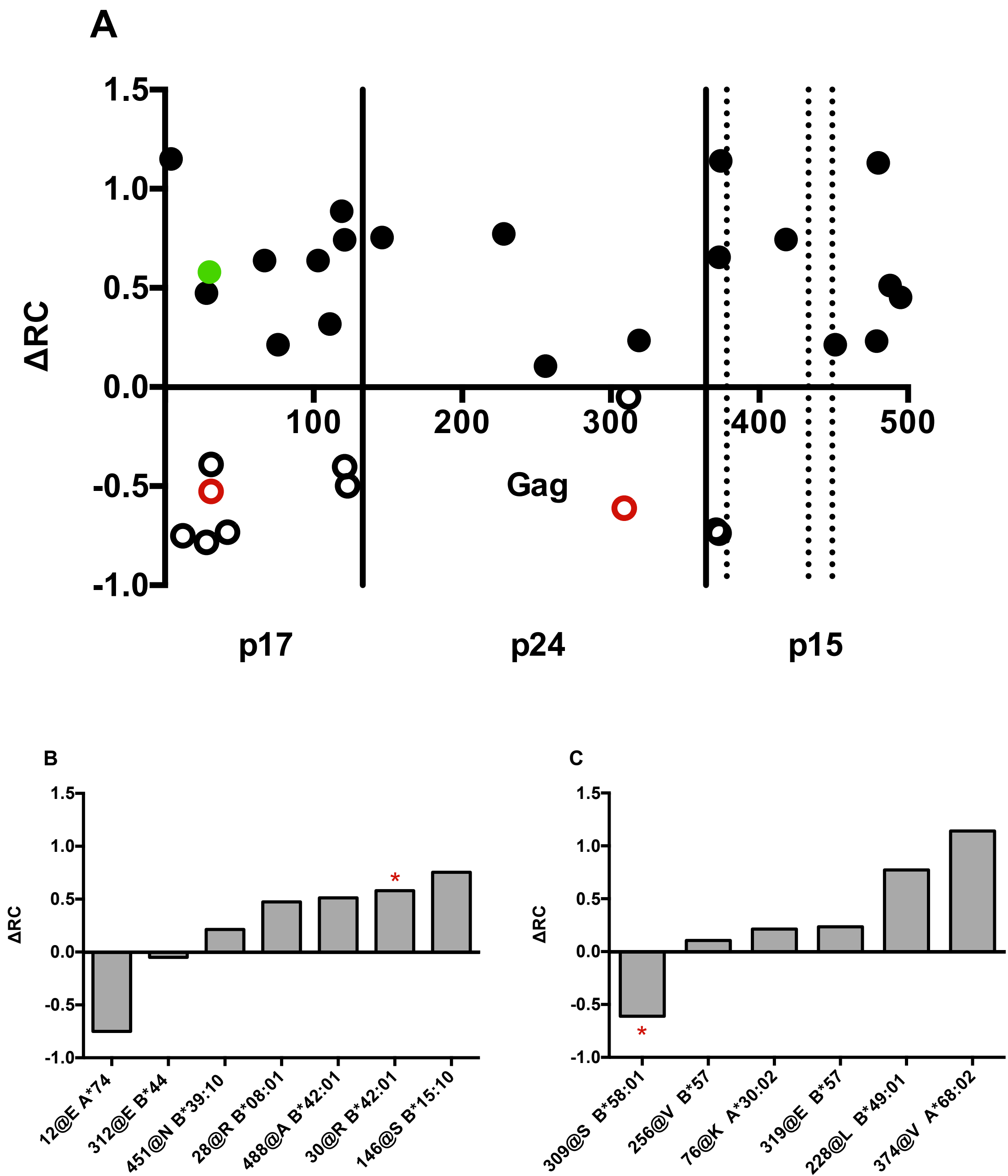


Figure 3. Identification of polymorphisms in Gag that significantly affect RC, several of which can be linked to HLA-class I alleles (A) In a pair-wise analysis of all amino acids represented in the 149 *gag* sequences tested, we observed 49 amino acids at 31

unique positions that were statistically associated (p<0.05) with changes in RC (open circles RC decreasing, filled circles RC increasing), 3 of which were significant at (q<0.2) when adjusted for multiple comparisons (red for RC reducing and green for RC increasing). The x-axis shows the polymorphism position in the primary Gag sequence (HXB2 numbering), and the y-axis depicts the impact of the polymorphism on RC relative to the median RC of all viruses (~1.5). (B) In a separate study analyzing 1899 subtype C gag sequences from Zambia and South Africa, 199 residues were linked to HLA-I alleles (q< 0.2, Carlson, Schaefer *et al.,* manuscript in preparation). From this, a total of 7 polymorphisms associated with changes in RC (p< 0.05) were found to be adapted to specific HLA-I alleles, adapted (amino acid is present when HLA-I allele is also present.) (* denotes q<0.2). (C) Six polymorphisms associated with changes in RC (p<0.05) were found to be non-adapted (amino acid is present only when HLA-I allele is absent) to specific HLA-I alleles (* denotes q<0.2).

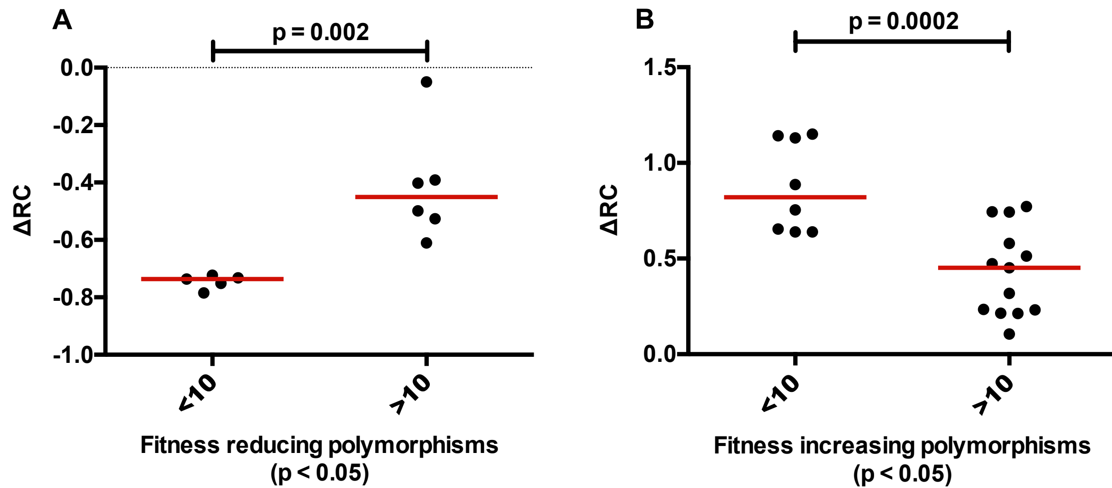


Figure 4. Rare polymorphisms have a significantly greater impact on RC (A) In an exploratory analysis, rare fitness reducing polymorphisms, occurring in less than 10 out of 149 of the sequences tested, were found to decrease fitness to a significantly greater degree than more common fitness decreasing mutations (Mann Whitney U test, $p = 0.002$). The y-axis depicts the impact of the polymorphism on RC relative to the median RC of all viruses (~1.5) (B) Rare fitness increasing polymorphisms, occurring in less than 10 out of 149 of the sequences tested, were found to increase fitness to a significantly greater degree than more common ones (Mann Whitney U test, $p = 0.0002$).

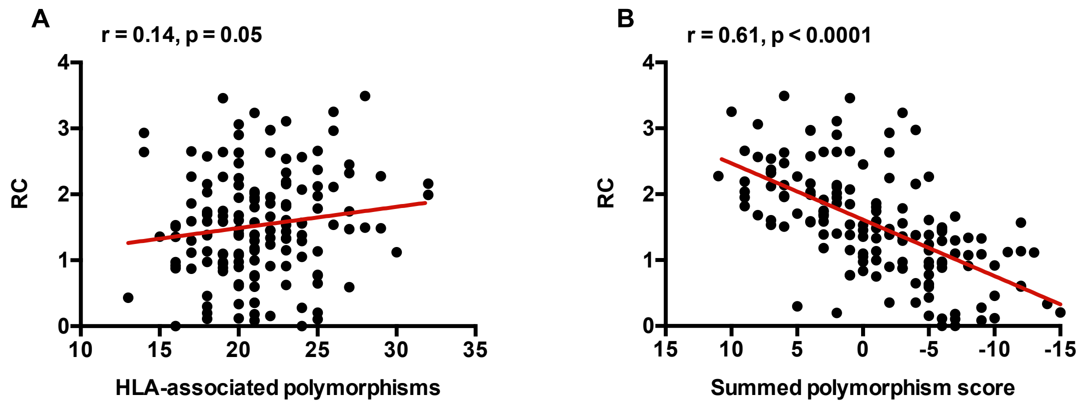


Figure 5. The balance of fitness increasing and decreasing HLA-associated polymorphisms strongly correlates with RC (A) The total number of HLA-associated polymorphisms positively correlates with RC. HLA-associated polymorphisms were defined as non-consensus residues at any amino acid position known to have polymorphisms statistically linked to HLA-I alleles. (B) RC of Gag-MJ4 chimeras is highly correlated to the summed polymorphism score in Gag. Each amino acid polymorphism was given a score of 1 (fitness increasing) or -1 (fitness decreasing). Summed polymorphism scores were generated by summing these scores for each Gag protein. Trend lines were generated using linear regression analysis, and are shown in order to facilitate visualization of correlations.

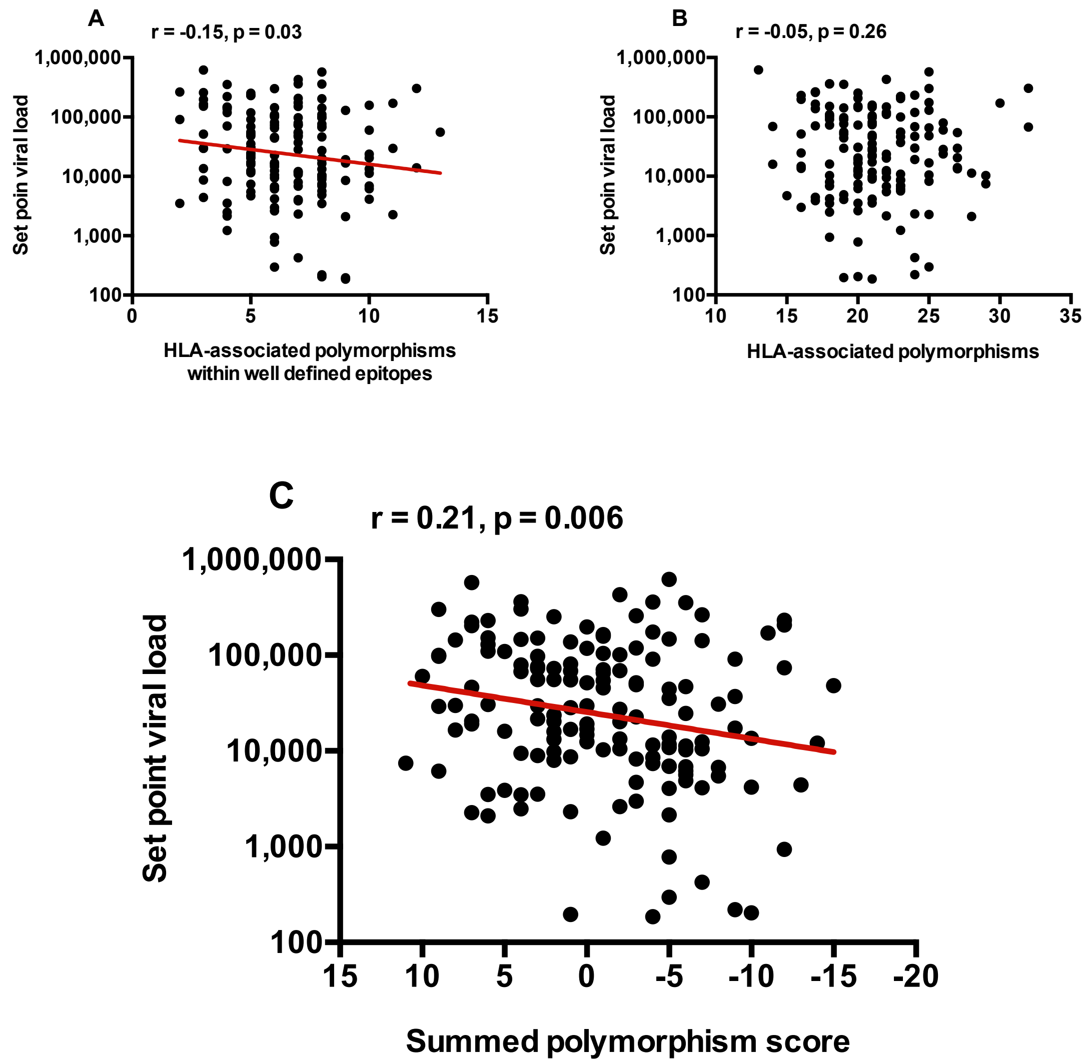


Figure 6. The balance of HLA-associated fitness increasing and decreasing mutations strongly correlates with set point viral load in newly infected individuals (A) The number of HLA-associated amino acid polymorphisms within well-defined epitopes for each Gag protein negatively correlates with set point VL. (B) The total number of HLA-associated polymorphisms (including those outside well-defined CTL epitopes) in Gag does not correlate to set point VL in newly infected individuals. (C) When the quality of HLA-associated polymorphisms is considered, a strong correlation between the summed

polymorphism score (as defined in Figure 5) and set point VL is observed. Trend lines were generated using linear regression analysis, and are shown in order to facilitate visualization of correlations.

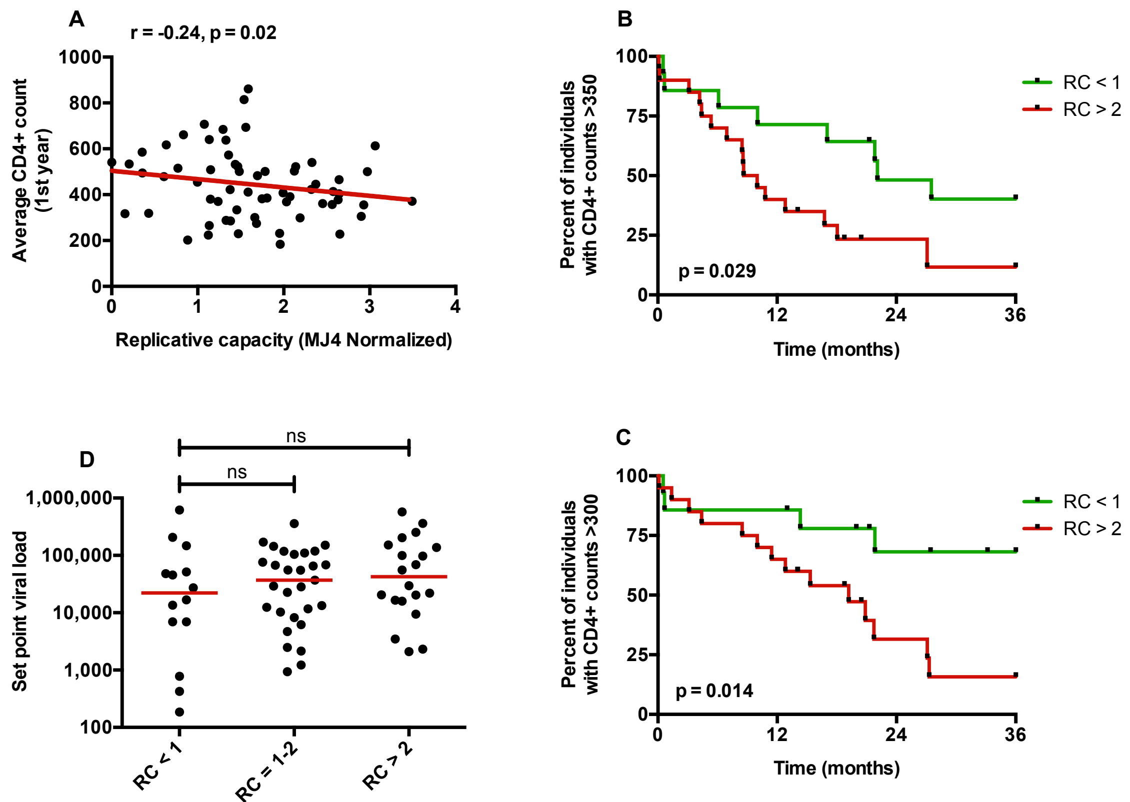


Figure 7. RC affects the rate of CD4 decline in a manner that may be independent of viral load (A) In a subset of individuals for which longitudinal CD4+ counts were available for at least 1 year post infection (n=63), RC of Gag-MJ4 chimeras negatively correlated with the average CD4+ counts of linked recipients within the first year (Spearman rank correlation, r = -0.24, p = 0.02). (B) Kaplan-Meier plots in which the endpoint was defined as the first CD4+ count below 350. Interval cut-off for endpoint was set to 36 months. The difference in median time to endpoint between those receiving viruses with RC<1 (n = 14) and RC>2 (n = 20) was 384 days (Log-rank test, p = 0.029). (C) Kaplan-Meier plots in which the endpoint was defined as the first CD4+ count below 300. Interval cut-off for endpoint was set to 36 months. The difference in median time to endpoint between those receiving viruses with RC<1 (n = 14) and RC>2 (n = 20) was

>800 days (Log-rank test, p = 0.014). (D) The difference in median VL between RC groups RC<1, RC 1-2, and RC >2 were not significantly different (Mann Whitney U test), consistent with the effect of RC on CD4 decline being independent of VL.

Table 2. Cox proportional hazard models demonstrate the independent effects of RC and VL on CD4 decline

Risk of CD4 below 350 in first 3 years	p	Hazard Ratio[c]	lower .95[d]	upper .95[d]
RC>2[a]	0.034	2.62	1.077	6.374
Log10(SetPoint VL)[b]	0.04	1.75	1.014	3.004
RC>2 + Log10(SetPointVL)	0.093	2.18	0.878	5.395
Risk of CD4 below 300 in first 3 years	**p**	**Hazard Ratio[c]**	**lower .95[d]**	**upper .95[d]**
RC>2[a]	0.021	3.80	1.226	11.801
Log10(SetPoint VL)[b]	0.023	2.09	1.105	3.94
RC>2 + log10(SetPointVL)	0.051	3.12	0.995	9.751

[a] as compared to RC<1

[b] Model in which $\log_{10}$ set point viral load in the linked recipient is taken into account

[c] The ratio of hazard rates between the variables RC>2 and RC<1

[d] The upper and lower bounds of confidence

0.001

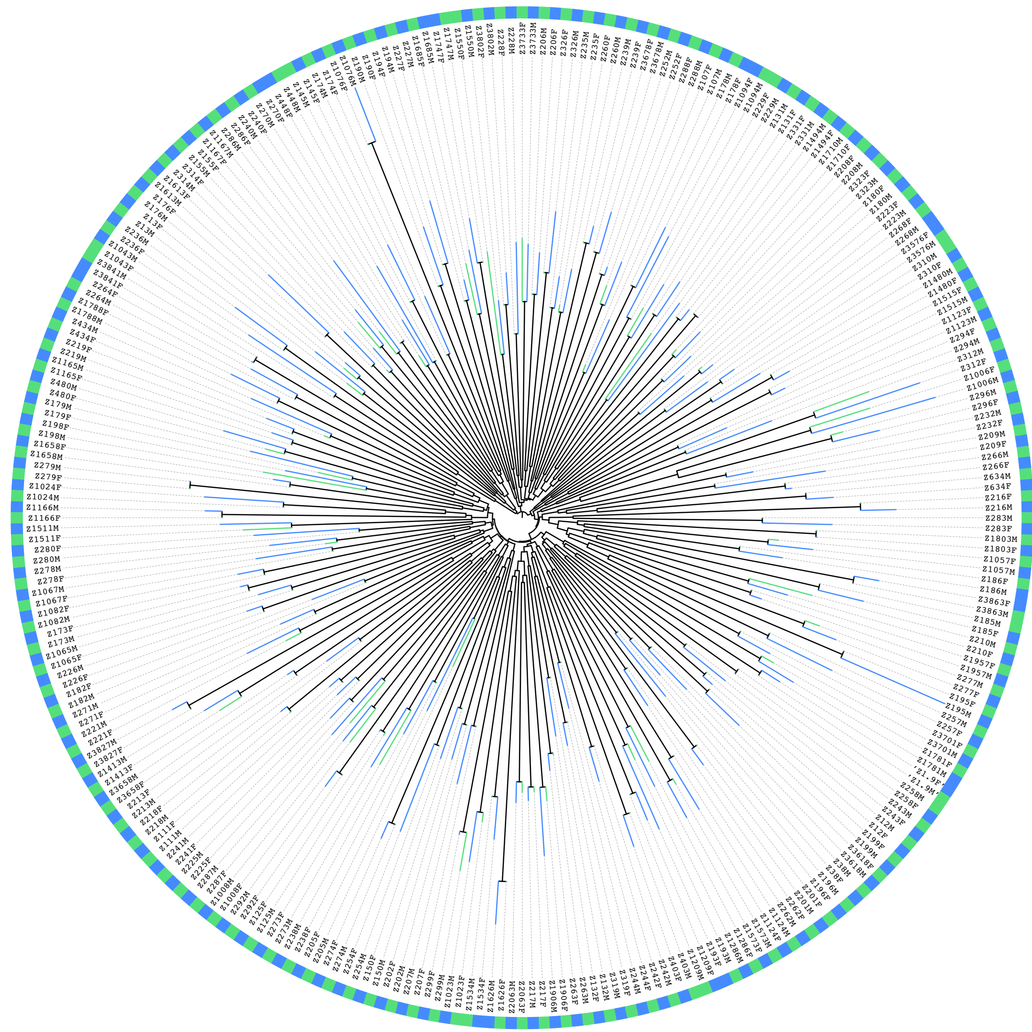


Figure S1. Donor and recipient population *gag* sequences cluster with one another

Gag population sequences from donors and linked recipients were amplified and sequenced as described in the methods section. Nucleotide *gag* sequences were aligned using the Gene Cutter tool accessible on the Los Alamos Nation Lab HIV Sequence Database (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html) and a neighbor-joining tree was generated using the Geneious sequence analysis software v5.5.7 (Biomatters Ltd.). The radial tree was annotated using the Interactive Tree of Life (iTOL)

online tool for phylogenetic tree display and annotation (Letunic and Bork, Bioinformatics, 2006). Blue denotes transmitting partners (donors) and green denotes linked recipients. This tree demonstrates the high degree of similarity between donor and linked recipient *gag* population sequences within epidemiologically linked transmission pairs.

Codon	Amino Acid	Consensus Amino Acid	Sequences w/o residue	Sequences with residue	Median RC w/o residue	Median RC with residue	p value	q value	ΔRC[a]	Within Epitope[b]	Epitope HLA Restriction[c]
31	L	L	49	103	1.051613	1.694342	0.00001	0.00545	0.19326	YES	A*2402; B*0801
31	I	L	119	33	1.606183	0.974471	0.00016	0.04360	-0.52661	YES	A*2402; B*0801
30	R	M	130	22	1.442259	2.080627	0.00107	0.18394	0.57955	YES	A*2402; B*0801
309	S	A	126	26	1.572069	0.890472	0.00135	0.18394	-0.61061	YES	B*4402; B*5301; B*5701; Cw5
28	R	H	125	27	1.433357	1.974499	0.00326	0.32077	0.47342	YES	A*0301; A*2402; B*0801
309	A	A	29	123	0.910234	1.567691	0.00397	0.32077	0.06661	YES	B*4402; B*5301; B*5701; Cw5
370	A	A	48	104	1.797895	1.380252	0.00412	0.32077	-0.12083	YES	B*4501
4	S	R	149	3	1.484227	2.651962	0.00965	0.40790	1.15088	NO	
4	R	R	3	149	2.651962	1.484227	0.01005	0.40790	-0.01685	NO	
451	N	S	96	56	1.417936	1.713877	0.01179	0.40790	0.21280	NO	
111	S	S	61	90	1.74016	1.383489	0.01199	0.40790	-0.11759	NO	
418	K	K	22	130	2.0575	1.468355	0.01217	0.40790	-0.03272	NO	
373	A	N	125	4	1.534145	0.764576	0.01258	0.40790	-0.73650	YES	B*4501
146	S	A	142	10	1.468355	2.255673	0.01380	0.40790	0.75460	YES	A*2501; B*1510; B*5701
119	A	E	145	4	1.472093	2.387669	0.01388	0.40790	0.88659	NO	
121	A	D	137	15	1.472093	2.244058	0.01418	0.40790	0.74298	NO	
42	D	E	145	7	1.534145	0.769036	0.01449	0.40790	-0.73204	YES	A*30; B*3501
451	S	S	63	89	1.68293	1.386771	0.01499	0.40790	-0.11431	NO	
69	Q	Q	26	126	1.822397	1.442259	0.01663	0.40790	-0.05882	NO	
12	E	K	149	3	1.532977	0.750097	0.01901	0.40790	-0.75098	YES	B*4002
418	R	K	136	16	1.47816	2.245238	0.01981	0.40790	0.74416	NO	
319	E	D	125	27	1.464616	1.735435	0.02014	0.40790	0.23436	NO	
319	D	D	27	125	1.735435	1.464616	0.02014	0.40790	-0.03646	NO	
85	L	L	4	148	2.364864	1.47816	0.02077	0.40790	-0.02292	YES	A*0201; A*1101; A*2902; A*3002; B*4403; B*58; B*63; Cw14
111	C	S	110	41	1.417936	1.819739	0.02115	0.40790	0.31866	NO	
480	G	D	145	3	1.49407	2.632309	0.02118	0.40790	1.13123	NO	
62	K	K	61	91	1.68293	1.386563	0.02145	0.40790	-0.11451	NO	
312	E	D	79	73	1.588149	1.451162	0.02178	0.40790	-0.04992	YES	B*4402; B*5301; B*5701; Cw5
312	D	D	73	79	1.451162	1.588149	0.02178	0.40790	0.08707	YES	B*4402; B*5301; B*5701; Cw5
123	E	G	133	13	1.535189	1.002543	0.02266	0.40790	-0.49853	NO	
495	N	S	117	35	1.451162	1.952761	0.02395	0.40790	0.45168	NO	
495	S	S	35	117	1.952761	1.451162	0.02395	0.40790	-0.04992	NO	
228	L	M	141	11	1.472093	2.273388	0.02650	0.43765	0.77231	YES	B*13

103	K	K	10	142	2.086015	1.474422	0.03202	0.48755	-0.02666	NO	
371	S	N	148	3	1.520532	0.778439	0.03300	0.48755	-0.72264	YES	B*4501
67	S	A	145	7	1.472093	2.139988	0.03310	0.48755	0.63891	NO	
67	A	A	7	145	2.139988	1.472093	0.03310	0.48755	-0.02898	NO	
103	R	K	143	9	1.484227	2.139988	0.03564	0.50140	0.63891	NO	
373	Q	N	124	5	1.47816	2.15603	0.03588	0.50140	0.65495	YES	B*4501
42	E	E	10	142	0.903087	1.534667	0.03691	0.50290	0.03359	YES	A*30; B*3501
121	G	D	135	17	1.538821	1.099119	0.03823	0.50552	-0.40196	NO	
76	K	R	110	42	1.394643	1.714888	0.04012	0.50552	0.21381	YES	A*3002; B*0801; B*4403; B*58; B*63
28	T	H	148	4	1.533561	0.716807	0.04064	0.50552	-0.78427	YES	A*0301; A*2402; B*0801
488	A	S	141	11	1.472093	2.013763	0.04174	0.50552	0.51269	YES	B*4001
488	S	S	11	141	2.013763	1.472093	0.04174	0.50552	-0.02898	YES	B*4001
479	R	K	128	22	1.459385	1.733306	0.04370	0.51775	0.23223	NO	
31	M	L	136	16	1.534083	1.109909	0.04710	0.54616	-0.39117	YES	A*2402; B*0801
256	V	I	99	53	1.454154	1.606183	0.04825	0.54784	0.10511	YES	B*3501
374	V	T	149	3	1.484227	2.642477	0.04981	0.55401	1.14140	YES	B*4501

Table S1. Amino acids in Gag associated with changes in replicative capacity

This table lists all amino acids associated with changes in RC. Residues that remain significantly associated with changes in RC after correction for multiple comparisons (q<0.2) are depicted in green. A total of 152 sequences and RC values were available for association analysis, with 149 of these with sufficient clinical follow-up for inclusion in the broader study.

[a] ΔRC is defined as median RC of all viruses tested (~1.5) subtracted from the median RC of all viruses with the particular polymorphism.

[b] The location of epitopes was defined by the compendium of "A-list" epitopes available in the LANL Immunology Database (*HIV Molecular Immunology 2009).*

[c] HLA class I alleles restricting epitopes harboring these polymorphisms that affect RC were also defined base on the LANL Immunology Database compilation of "A-list" epitopes.

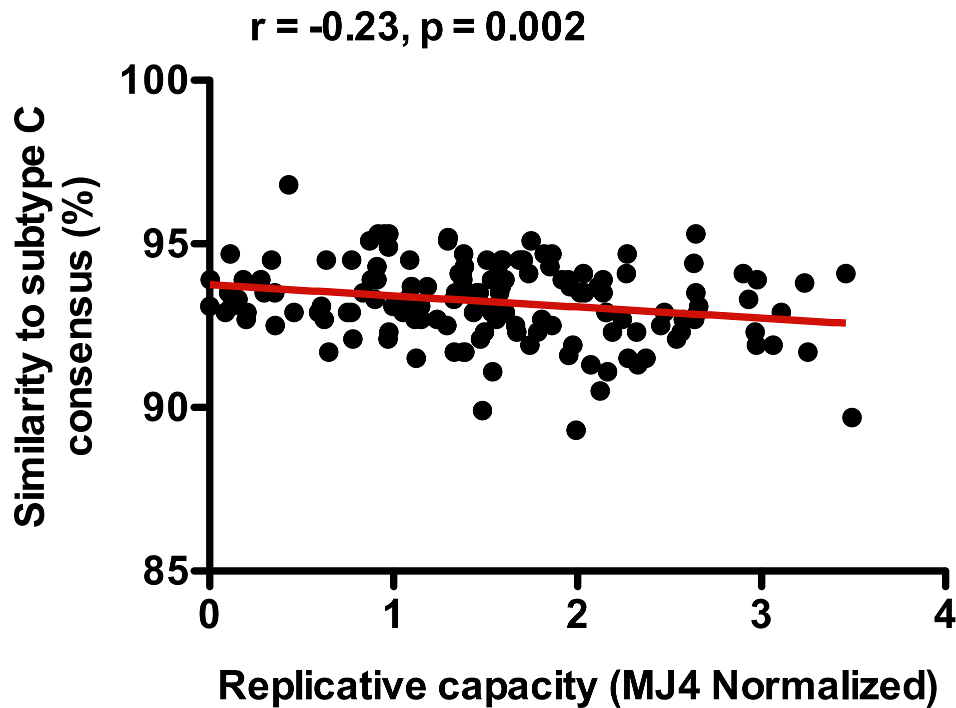


Figure S2. Gag sequences that are less like the Gag subtype C consensus sequence replicate more efficiently *in vitro* The Gag amino acid sequences of all Gag-MJ4 chimeras were compared to a Zambian subtype C consensus Gag sequence (generated using the LANL Consensus Maker tool; http://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html) by building a neighbor-joining tree using the Geneious sequence analysis software v5.5.7 (Biomatters Ltd.) and the percent similarity to consensus was determined for each sequence. Notably, viruses encoding Gag sequences most disparate from the subtype C consensus Gag sequence replicated to higher levels, and we observed a statistically significant negative correlation between RC and the percent similarity of Gag to consensus (Pearson correlation, $p = 0.002$, $r = -0.23$).

References Cited

1. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, et al. (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105: 7552-7557.

2. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, et al. (2009) Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS Pathog 5: e1000274.

3. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, et al. (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science 303: 2019-2022.

4. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317: 944-947.

5. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, et al. (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 2: 405-411.

6. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, et al. (2004) Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature 432: 769-775.

7. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330: 1551-1557.

8. Tang J, Kaslow RA (2003) The impact of host genetics on HIV infection and disease progression in the era of highly active antiretroviral therapy. AIDS 17 Suppl 4: S51-60.

9. Tang J, Malhotra R, Song W, Brill I, Hu L, et al. (2010) Human leukocyte antigens and HIV type 1 viral load in early and chronic infection: predominance of evolving relationships. PLoS One 5: e9629.

10. Tang J, Tang S, Lobashevsky E, Myracle AD, Fideli U, et al. (2002) Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. J Virol 76: 8276-8284.

11. Tang J, Tang S, Lobashevsky E, Zulu I, Aldrovandi G, et al. (2004) HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. AIDS Res Hum Retroviruses 20: 19-25.

12. Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly CA, Serwadda D, et al. (2010) HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. PLoS Pathog 6: e1000876.

13. Hecht FM, Hartogensis W, Bragg L, Bacchetti P, Atchison R, et al. (2010) HIV RNA level in early infection is predicted by viral load in the transmission source. AIDS 24: 941-945.

14. Rong R, Li B, Lynch RM, Haaland RE, Murphy MK, et al. (2009) Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. PLoS Pathog 5: e1000594.

15. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, et al. (1997) Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nat Med 3: 205-211.

16. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, et al. (1997) Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 3: 212-217.

17. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, et al. (2004) HIV evolution: CTL escape mutation and reversion after transmission. Nat Med 10: 282-289.

18. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, et al. (1991) Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature 354: 453-459.

19. Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, et al. (1997) Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. Proc Natl Acad Sci U S A 94: 1890-1895.

20. Jones NA, Wei X, Flower DR, Wong M, Michor F, et al. (2004) Determinants of human immunodeficiency virus type 1 escape from the primary CD8+ cytotoxic T lymphocyte response. J Exp Med 200: 1243-1256.

21. Goulder PJ, Watkins DI (2004) HIV and SIV CTL escape: implications for vaccine design. Nat Rev Immunol 4: 630-640.

22. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, et al. (2009) Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458: 641-645.

23. Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichterfeld M, et al. (2004) Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. J Virol 78: 7069-7078.

24. Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, et al. (2008) Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS Pathog 4: e1000033.

25. Friedrich TC, Dodds EJ, Yant LJ, Vojnov L, Rudersdorf R, et al. (2004) Reversion of CTL escape-variant immunodeficiency viruses in vivo. Nat Med 10: 275-281.

26. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, et al. (2007) Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. J Virol 81: 193-201.

27. Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, et al. (2008) Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. J Virol 82: 8548-8559.

28. Rousseau CM, Daniels MG, Carlson JM, Kadie C, Crawford H, et al. (2008) HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. J Virol 82: 6434-6446.

29. Schneidewind A, Brockman MA, Sidney J, Wang YE, Chen H, et al. (2008) Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. J Virol 82: 5594-5605.

30. Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, et al. (2011) Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proc Natl Acad Sci U S A 108: 11530-11535.

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

32. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, et al. (2007) Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. J Virol 81: 12608-12618.

33. Crawford H, Prado JG, Leslie A, Hue S, Honeyborne I, et al. (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. J Virol 81: 8346-8351.

34. Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, et al. (2006) Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J Virol 80: 3617-3623.

35. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, et al. (2007) Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. J Virol 81: 12382-12393.

36. Wright JK, Naidoo VL, Brumme ZL, Prince JL, Claiborne DT, et al. (2012) Impact of HLA-B*81-associated mutations in HIV-1 Gag on viral replication capacity. J Virol 86: 3193-3199.

37. Mothe B, Llano A, Ibarrondo J, Daniels M, Miranda C, et al. (2011) Definition of the viral targets of protective HIV-1-specific T cell responses. J Transl Med 9: 208.

38. Rolland M, Heckerman D, Deng W, Rousseau CM, Coovadia H, et al. (2008) Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads. PLoS One 3: e1424.

39. Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, et al. (2009) Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. J Virol 83: 1845-1855.

40. Wagner R, Leschonsky B, Harrer E, Paulus C, Weber C, et al. (1999) Molecular and functional analysis of a conserved CTL epitope in HIV-1 p24 recognized from a long-term nonprogressor: constraints on immune escape associated with targeting a sequence essential for viral replication. J Immunol 162: 3727-3734.

41. Peyerl FW, Barouch DH, Letvin NL (2004) Structural constraints on viral escape from HIV- and SIV-specific cytotoxic T-lymphocytes. Viral Immunol 17: 144-151.

42. Crawford H, Lumm W, Leslie A, Schaefer M, Boeras D, et al. (2009) Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. J Exp Med 206: 909-921.

43. Brockman MA, Brumme ZL, Brumme CJ, Miura T, Sela J, et al. (2010) Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication

capacity that may be largely compensated for in chronic infection. J Virol 84: 11937-11949.

44. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, et al. (2008) Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol 4: e1000225.

45. Chopera DR, Mlotshwa M, Woodman Z, Mlisana K, de Assis Rosa D, et al. (2011) Virological and immunological factors associated with HIV-1 differential disease progression in HLA-B 58:01-positive individuals. J Virol 85: 7070-7080.

46. Crawford H, Matthews PC, Schaefer M, Carlson JM, Leslie A, et al. (2011) The hypervariable HIV-1 capsid protein residues comprise HLA-driven CD8+ T-cell escape mutations and covarying HLA-independent polymorphisms. J Virol 85: 1384-1390.

47. Huang KH, Goedhals D, Carlson JM, Brockman MA, Mishra S, et al. (2011) Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations. PLoS One 6: e19018.

48. Rolland M, Carlson JM, Manocheewa S, Swain JV, Lanxon-Cookson E, et al. (2010) Amino-acid co-variation in HIV-1 Gag subtype C: HLA-mediated selection pressure and compensatory dynamics. PLoS One 5.

49. Schneidewind A, Brumme ZL, Brumme CJ, Power KA, Reyor LL, et al. (2009) Transmission and long-term stability of compensated CD8 escape mutations. J Virol 83: 3993-3997.

50. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, et al. (2008) Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J Exp Med 205: 1009-1017.

51. Miura T, Brumme ZL, Brockman MA, Rosato P, Sela J, et al. (2010) Impaired replication capacity of acute/early viruses in persons who become HIV controllers. J Virol 84: 7581-7591.

52. Miura T, Brockman MA, Schneidewind A, Lobritz M, Pereyra F, et al. (2009) HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition. J Virol 83: 2743-2755.

53. Miura T, Brockman MA, Brumme ZL, Brumme CJ, Pereyra F, et al. (2009) HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. J Virol 83: 140-149.

54. Wright JK, Novitsky V, Brockman MA, Brumme ZL, Brumme CJ, et al. (2011) Influence of Gag-protease-mediated replication capacity on disease progression in individuals recently infected with HIV-1 subtype C. J Virol 85: 3996-4006.

55. Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, et al. (2010) Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. J Virol 84: 10820-10831.

56. Ndung'u T, Renjifo B, Essex M (2001) Construction and analysis of an infectious human Immunodeficiency virus type 1 subtype C molecular clone. J Virol 75: 4964-4972.

57. Allen S, Karita E, Chomba E, Roth DL, Telfair J, et al. (2007) Promotion of couples' voluntary counselling and testing for HIV through influential networks in two African capital cities. BMC Public Health 7: 349.
58. Kempf MC, Allen S, Zulu I, Kancheya N, Stephenson R, et al. (2008) Enrollment and retention of HIV discordant couples in Lusaka, Zambia. J Acquir Immune Defic Syndr 47: 116-125.
59. McKenna SL, Muyinda GK, Roth D, Mwali M, Ng'andu N, et al. (1997) Rapid HIV testing and counseling for voluntary testing centers in Africa. AIDS 11 Suppl 1: S103-110.
60. Trask SA, Derdeyn CA, Fideli U, Chen Y, Meleth S, et al. (2002) Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. J Virol 76: 397-405.
61. Bhakta SJ, Shang L, Prince JL, Claiborne DT, Hunter E (2011) Mutagenesis of tyrosine and di-leucine motifs in the HIV-1 envelope cytoplasmic domain results in a loss of Env-mediated fusion and infectivity. Retrovirology 8: 37.
62. Brockman MA, Tanzi GO, Walker BD, Allen TM (2006) Use of a novel GFP reporter cell line to examine replication capacity of CXCR4- and CCR5-tropic HIV-1 by flow cytometry. J Virol Methods 131: 134-142.
63. Ostrowski MA, Chun TW, Cheseboro B, Stanley SK, Tremblay M (2006) Detection assays for HIV proteins. Curr Protoc Immunol Chapter 12: Unit 12 15.
64. Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, et al. (2012) Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. J Virol 86: 5230-5243.

65. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100: 9440-9445.

66. Fideli US, Allen SA, Musonda R, Trask S, Hahn BH, et al. (2001) Virologic and immunologic determinants of heterosexual transmission of human immunodeficiency virus type 1 in Africa. AIDS Res Hum Retroviruses 17: 901-910.

67. Saag MS, Holodniy M, Kuritzkes DR, O'Brien WA, Coombs R, et al. (1996) HIV viral load markers in clinical practice. Nat Med 2: 625-629.

68. Mellors JW, Rinaldo CR, Jr., Gupta P, White RM, Todd JA, et al. (1996) Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science 272: 1167-1170.

69. Rodriguez B, Sethi AK, Cheruvu VK, Mackay W, Bosch RJ, et al. (2006) Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection. JAMA 296: 1498-1506.

70. Fahey JL, Taylor JM, Detels R, Hofmann B, Melmed R, et al. (1990) The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. N Engl J Med 322: 166-172.

71. Crowley S, Rollins N, Shaffer N, Guerma T, Vitoria M, et al. (2010) New WHO HIV treatment and prevention guidelines. Lancet 375: 874-875.

72. Freed EO, Orenstein JM, Buckler-White AJ, Martin MA (1994) Single amino acid changes in the human immunodeficiency virus type 1 matrix protein block virus particle production. J Virol 68: 5311-5320.

73. Spearman P, Horton R, Ratner L, Kuli-Zade I (1997) Membrane binding of human immunodeficiency virus type 1 matrix protein in vivo supports a conformational myristyl switch mechanism. J Virol 71: 6582-6592.

74. Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI (2004) Assembly properties of the human immunodeficiency virus type 1 CA protein. J Virol 78: 2545-2552.

75. Weiss ER, Gottlinger H (2011) The role of cellular factors in promoting HIV budding. J Mol Biol 410: 525-533.

76. Gottlinger HG, Dorfman T, Sodroski JG, Haseltine WA (1991) Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. Proc Natl Acad Sci U S A 88: 3195-3199.

77. von Schwedler UK, Stemmler TL, Klishko VY, Li S, Albertine KH, et al. (1998) Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. EMBO J 17: 1555-1568.

78. Forshey BM, von Schwedler U, Sundquist WI, Aiken C (2002) Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. J Virol 76: 5667-5677.

79. Dorfman T, Mammano F, Haseltine WA, Gottlinger HG (1994) Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. J Virol 68: 1689-1696.

80. Freed EO, Martin MA (1996) Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. J Virol 70: 341-351.

81. Krausslich HG, Facke M, Heuser AM, Konvalinka J, Zentgraf H (1995) The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. J Virol 69: 3407-3419.

82. Wright ER, Schooler JB, Ding HJ, Kieffer C, Fillmore C, et al. (2007) Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. EMBO J 26: 2218-2226.

83. Li F, Goila-Gaur R, Salzwedel K, Kilgore NR, Reddick M, et al. (2003) PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. Proc Natl Acad Sci U S A 100: 13555-13560.

84. Lu W, Salzwedel K, Wang D, Chakravarty S, Freed EO, et al. (2011) A single polymorphism in HIV-1 subtype C SP1 is sufficient to confer natural resistance to the maturation inhibitor bevirimat. Antimicrob Agents Chemother 55: 3324-3329.

85. Lazaryan A, Song W, Lobashevsky E, Tang J, Shrestha S, et al. (2011) The influence of human leukocyte antigen class I alleles and their population frequencies on human immunodeficiency virus type 1 control among African Americans. Hum Immunol 72: 312-318.

86. Matthews PC, Adland E, Listgarten J, Leslie A, Mkhwanazi N, et al. (2011) HLA-A*7401-mediated control of HIV viremia is independent of its linkage disequilibrium with HLA-B*5703. J Immunol 186: 5675-5686.

87. Fledderman EL, Fujii K, Ghanam RH, Waki K, Prevelige PE, et al. (2010) Myristate exposure in the human immunodeficiency virus type 1 matrix protein is modulated by pH. Biochemistry 49: 9551-9562.

88. Zhou W, Parent LJ, Wills JW, Resh MD (1994) Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. J Virol 68: 2556-2569.

89. Navis M, Matas DE, Rachinger A, Koning FA, van Swieten P, et al. (2008) Molecular evolution of human immunodeficiency virus type 1 upon transmission between human leukocyte antigen disparate donor-recipient pairs. PLoS One 3: e2422.

90. Poon A, Chao L (2005) The rate of compensatory mutation in the DNA bacteriophage phiX174. Genetics 170: 989-999.

91. Hunt PW, Martin JN, Sinclair E, Bredt B, Hagos E, et al. (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. J Infect Dis 187: 1534-1543.

Chapter III

Transmitted HIV-1 replicative capacity drives immune activation and CD4 proviral load

D. Claiborne[1†], J. Prince[1†], E. Scully[2], G. Macharia[3], L. Micci[1], B. Lawson[1], J. Kopycinski[3,4], M. Deymier[1], T. Vanderford[1], K. Nganou-Makamdop[5], Z. Ende[1], J. Tang[6], T. Yu[7], S. Lakhi[8], W. Kilembe[8], G. Silvestri[1], D. Douek[5], P. Goepfert[6], M. A. Price[9,10], S. Allen[8,11,12], M. Paiardini[1], M. Altfeld[2,13], J. Gilmour[3,4], E. Hunter[1,11*]

1. Emory Vaccine Center at Yerkes National Primate Research Center, Emory University, Atlanta, Georgia, USA
2. Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA
3. International AIDS Vaccine Initiative, London, England
4. Imperial College, London, England
5. Vaccine Research Center (VRC) at the National Institutes of Health (NIH), Bethesda, Maryland, USA
6. Department of Medicine, University of Alabama at Birmingham (UAB), Birmingham, Alabama, USA
7. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA
8. Zambia-Emory Research Project, Lusaka, Zambia
9. International AIDS Vaccine Initiative, San Francisco, California, USA
10. Department of Epidemiology and Biostatistics, UCSF, SF, CA, USA
11. Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, USA

12. Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
13. Heinrich-Pette-Institut, Hamburg, Germany

† Contributed equally to this manuscript

Manuscript submitted

Abstract

Characteristics of HIV-1 transmitted/founder viruses, especially viral replicative capacity (vRC), are known to influence set point viral load (SPVL) and CD4+ T cell decline in HIV-1 seroconvertors. In a cohort of 127 acutely infected Zambian volunteers we show that this association of vRC with CD4+ T cell decline is largely independent of protective HLA alleles (B*57/5801) and SPVL. Viral RC drives an early inflammatory cytokine response that is associated with profound T cell activation, exhaustion, and proliferation, all strong predictors of HIV 1 disease progression in this cohort. Moreover, vRC is positively correlated with the magnitude of viral burden in naïve and central memory CD4+ T cells during early infection. Taken together, these new findings support an unprecedented role for vRC in driving multiple facets of HIV-1-related immunopathology.

Introduction

From the start of the AIDS epidemic, HIV-1 infection has been characterized by a steady decline in CD4+ T cells resulting in a state of overt immunodeficiency marked by an increased susceptibility to opportunistic infections and malignancies (*1, 2*). Yet, despite the fact that a majority of HIV-1 infected individuals eventually progress to AIDS, they do so at varying rates (*3*). To date, research has primarily focused on identifying host factors that contribute to viral load and disease progression (*4*) from the earliest stages of infection. Despite studies highlighting the importance of several HLA-class I alleles, it has been estimated that host genetic factors explain only 22% of the variability in viral load (*5*).

In general a single virus establishes HIV infection (*6, 7*), but the extent to which viral characteristics, and in particular vRC, explain variations in disease severity have not been determined. It is known that viral loads in epidemiologically linked transmission pairs are correlated between chronically infected donors and acutely infected recipients, implying that transmitted viral characteristics can impact viral control in individuals with disparate immunogenetic backgrounds (*8-10*). This is supported by our observation that the number of transmitted cytotoxic T lymphocyte (CTL) induced escape mutations in the structural protein precursor, Gag, but not the accessory protein, Nef, correlated with a reduction in early set point viral load (SPVL). These results are consistent with escape mutations within Gag attenuating viral replication and thereby impacting viral control (*11*).

We showed in a previous study that low viral replicative capacity (vRC), defined *in vitro* by the transmitted Gag sequence in chimeric viruses, was associated with a

delayed loss of CD4+ T cells in individuals recently infected with HIV-1 subtype C. Because this effect appeared to be partially independent of SPVL, we hypothesized that high levels of replication of the transmitted/founder virus might initiate irreversible pathogenic events early in infection that dictate the kinetics of subsequent disease progression (*12*). Here, in a study of 127 acutely HIV-1 infected Zambian volunteers, identified a median of 46 days post estimated date of infection, we test this hypothesis and show that infection by high vRC HIV-1, as defined by Gag, is associated with increased levels of inflammatory cytokines, greater cellular immune activation and exhaustion, and higher levels of proviral burden in naïve and central memory T cell subsets at early time points after infection. Thus, the replicative capacity conferred by the *gag* gene of transmitted HIV-1 is a critical factor in defining the initial inflammatory state, the preservation or loss of CD4+ T cell homeostasis, and the trajectory of disease progression. Interventions including early antiretroviral therapy or vaccine-induced immunity that impact these early events may have a significant effect on the development of clinical disease.

Materials and Methods

Study subjects

All participants in the Zambia Emory HIV Research Project (ZEHRP) discordant couples cohort in Lusaka, Zambia were enrolled in human subjects protocols approved by both the University of Zambia Research Ethics Committee and the Emory University Institutional Review Board. Prior to enrollment, individuals received counseling and signed a written informed consent form agreeing to participate.

The subjects included in this study were selected from the ZEHRP cohort based on being recently infected with HIV-1. All subjects were initially seronegative partners within serodiscordant cohabitating heterosexual couples that subsequently seroconverted. All subjects were antiretroviral therapy naïve and were identified a median of 46 (IQR = 33-49) days after the estimated date of infection (EDI). The algorithm used to determine the EDI has been previously described (*50*). All subjects were infected by HIV-1 subtype C viruses.

Generation of Gag-MJ4 chimeras and transmitted/founder full-length infectious molecular clones

Gag-MJ4 chimeras were generated from frozen plasma isolated at the seroconversion time point for 127 subjects as previously described (*12*). Briefly, viral RNA was extracted from 140 ul of plasma using the Qiagen Viral RNA extraction kit (Qiagen). Combined RT-PCR and first round PCR were performed in a single reaction, and *gag* genes were amplified using a nested second round PCR. Patient-derived *gag*

genes were joined with the MJ4 long terminal repeat portion via splice-overlap-extension PCR, and *gag*-LTR amplicons were cloned into the MJ4 proviral vector using NgoMIV and HpaI endonuclease restriction enzymes.

Full-length genome cDNA synthesis and near full-length single genome amplification was performed as described previously (Deymier et al. Submitted 2014) on seroconversion plasma samples for five individuals where Gag-chimeras vRC had been measured. Briefly, viral RNA was extracted from 140μl of plasma using the Qiagen Viral RNA extraction kit, which was immediately used to generate full-length cDNA using an anchor oligo-dT primer with the Superscript III Reverse Transcriptase Enzyme (Life Technologies). cDNA was used at limiting dilutions to amplify near full-length genomes with previously described primers (*51*) using the Q5 hot-start enzyme (NEB). Sequencing of multiple amplicons allowed for confirmation of single variant transmission by starlike phylogeny, and to infer the near full-length sequence of the Transmitted/Founder virus (*52*).

Using patient specific primers derived from the TF sequence, viral LTR sequence was amplified from genomic DNA extracted from white cell pellets using the Qiagen DNeasy Blood Kit. The LTR was cloned into a pBluescript vector and used to generate two LTR PCR products which when ligated in combination with the near full-length TF amplicon, using the Clontech Infusion-HD Cloning kit, generates a full-length genome infectious molecular clone of the entire TF sequence (Deymier et al. submitted 2014).

Evaluation of plasma cytokines

Plasma levels of cytokines and chemokines were measured using MILLIPLEX Human Cytokine/Chemokine detection kits (Millipore). High sensitivity kits were used for measurement of IFNγ, IL-1β, IL-6 and IL-7 and regular sensitivity kits were used for IFNα2, IL-10, IP-10, MCP-1, MIP-1β and TNFα and were used according to the manufacturer's instructions. Samples were run in duplicate with all individuals on the same plate and wells with low bead count or coefficient of variance >30% were excluded from subsequent analysis. Plates were read on the Bio-Plex® 3D Suspension Array System (Bio-Rad). Levels of sCD14 (R&D systems), CRP (Millipore), sCD163 (Trillium Diagnostics) and D-dimer (American Diagnostica) were all measured using standard ELISA based assays according to the manufacturer's instructions. Measurement of LPS levels was performed using the LAL Chromogenic Endotoxin Quantification kit (American Diagnostica). Intestinal fatty acid binding protein (I-FABP) was measured using a commercially available ELISA DuoSet assay (R&D Systems, Minneapolis, MN, USA) according to the manufacturer's instructions with minor adjustments. Plasma samples were diluted to 10% in diluent from the R&D Systems soluble CD14 ELISA kit (DC140) and plates were blocked with Sigma Blocking Buffer.

Flow cytometry analysis

Cryopreserved peripheral blood mononuclear cells (PBMCs) isolated from 35 HIV-1 infected Zambians at a median of 49 days post estimated date of infection were analyzed by flow cytometry for levels of activation (CD38 and HLA-DR), exhaustion (PD1 and CD57) and cytotoxicity (granzyme B and perforin) on CD8+ T cells. Predetermined optimal concentrations of the following monoclonal antibodies were

utilized in a twelve-parameter cytometric staining panel to assess CD8+ T cell activation and exhaustion: anti-CD3-Qdot 655 (Clone S4.1, Invitrogen), anti-CD8-Qdot-705 (Clone 3B5, Invitrogen), anti-CD4-Qdot-605 (Clone S3.5, Invitrogen), anti-PD1-BV-421 (Clone EH12.1, BD Biosciences), anti-CCR5-PE-CY7 (Clone 2D7, BD Biosciences), anti-CCR7-Alexa700 (Clone 150503, BD Biosciences), anti-HLA-DR-PE (Clone G46-6, BD Biosciences), anti-CD57-FITC (NK-1, BD), anti-CD45RO-PECF594 (Clone UCHL1, BD Biosciences), anti-CD38-APC (Clone HB7,BD Biosciences), anti-CD27-APC-CY7 (Clone MT271, BD Biosciences), Am Cyan Live dead (Synthetic, Invitrogen). Frozen PBMCs were thawed, washed once with R20 (RPMI 1640 (Sigma, UK) containing 20% Fetal Calf Serum (Sigma, UK) 1% 1M HEPES buffer (Sigma, UK), 1%L-glutamine (Sigma, UK), 1%Penicillin-streptomycin (Sigma, UK) and 1% Sodium Pyruvate (Sigma, UK)), and then rested overnight in 4 mL at 37^0C, 5% CO_2, 95% humidity.

For each individual, 10^6 PBMCs per well washed once with Phosphate Buffered Saline (PBS) (Sigma, UK) before staining. For the intracellular markers Granzyme B and Perforin, Cytofix/Cytoperm (BD, UK) was utilized for permeabilization and staining per the manufacturer's protocol. To correct for spectral overlap between fluorochromes, compensation controls for each antibody were prepared using anti mouse Igk BD™ Compensation Particles (BD, UK) according to the manufacture's instructions. Flow cytometry was performed on a LSR II flow cytometer (BD, UK) with data acquisition using BD FACSDiva software.

CD4+ T cell memory populations and markers of activation and proliferation were analyzed in a separate flow analysis. Cryopreserved PBMCs isolated from a subset

of 19 HIV-1 infected Zambians at 3-months post seroconversion as well as 14 uninfected healthy Zambians were assessed for activation and proliferation using twelve-parameter flow cytometric analysis. The individuals were chosen based on replication capacity phenotype, overlap with other parameters measured, and on the sample availability at this time point, as 20 million cells were needed for the sorting experiments. The following monoclonal antibodies were utilized: Aqua Live/Dead amine dye-AmCyan (Invitrogen), anti-CD3-APC-CY7 (clone SP34-2, BD Biosciences), anti-CD4-APC (clone OKT4, Biolegend), anti-CD8-Brilliant Violet-605 (clone RPA-T8, Biolegend), anti-CD45RO-PE Texas Red (clone UCHL1, Beckman Coulter), anti-CD27-PE-Cy5 (clone 1A4CD27, Beckman Coulter), anti-CCR7-PE-CY7 (clone 3D12, BD Biosciences), anti-Ki67-Alexa700 (clone B56, BD Biosciences), anti-CD38-FITC (clone AT-1, Stemcell), anti-HLADR-PerCp-Cy5.5 (clone G46-6, BD Biosciences), anti-PD1-Brilliant Violet-421 (clone EH12.2H7, Biolegend), and anti-CCR5-PE (clone 3A9, BD Biosciences).

Approximately twenty million PBMCs from each individual was first thawed, washed twice with 10 mL of complete RPMI supplemented with 2 uL of DNase. Once the cells were counted using an automated cell counter, approximately 2 million cells were set aside for staining with the above twelve-parameter *ex vivo* staining panel, while the rest of the cells were utilized for sorting if they were derived from an HIV-1 infected individual (see below for further methods on sorting). Cells were then washed with 3mL of Dulbeco's Phosphate Buffered Saline without Ca+/Mg+ (Invitrogen) and stained for 5 min at room temperature with Aqua Live/Dead amine dye-AmCyan (Invitrogen). Anti-CCR7-PE-CY7 was added to the cells, and incubated at 37^{0}C for fifteen minutes. The rest of the monoclonal antibodies were then added and incubated at room temperature for

thirty minutes (minus the intracellular Ki67). The cells were then permeabilized and stained for the intracellular marker Ki67 for 30 minutes at room temperature using the BD perm/wash kit (Fischer, BDB554723) following the manufacturer's directions. Flow cytometric acquisition was performed on at least 100, 000 $CD3^+$ T cells on an LSRII cytometer with FACSDiVa Version 6.1.3 software. Compensation controls were prepared for each fluorphore using the AbC™ Anti-Mouse Bead Kit (Life Technologies, A-10344). The instrument was calibrated before acquisition using the Ultra Rainbow Calibration Kit – 6 peaks (Spherotec, URCP-38-2K). Analysis of all flow cytometric data was performed using FlowJo Version 9.7.5 software (TreeStar).

Detection of cell-associated HIV-1 viral DNA in CD4+ T cell compartments

Cryopreserved PBMCs isolated from a subset of 21 HIV-1 infected individuals at 3-months post seroconversion were stained for T cell subsets: Viability – Live/Dead Fixable Aqua (Invitrogen); CD3 Alexa Fluor 700, CD4 PE, CD8 Pacific Blue, CCR7 PE-Cy7 (BD Biosciences); CD27 PE-Cy6, and CD45RO PE Texas Red (Beckman-Coulter). Samples were sorted into naïve CD4 T cells (CD3+, CD4+, CD8-, CD27+, CD45RO-), central memory CD4 T cells (CD3+, CD4+, CD8-, CD27+/-, CD45RO+, CCR7+), and effector memory CD4 T cells (CD3+, CD4+, CD8-, CD27+/-, CD45RO+, CCR7-) using a FACSAria II flow cytometer (Becton Dickinson, San Jose, CA). Sorted populations were checked for purity and were found to be between 94.5% - 99.4% specific for the population desired.

Live, sorted CD4+ T cell subsets were immediately lysed and genomic DNA extracted using the DNeasy Blood & Tissue Kit (Qiagen). Samples were eluted in nuclease-free water and analyzed for DNA concentration. Live, sorted CD4+ T cell

subsets were immediately lysed and genomic DNA extracted using the DNeasy Blood & Tissue Kit (Qiagen). Samples were eluted in nuclease-free water and analyzed by quantitative real-time PCR (qPCR) for total cell-associated HIV DNA. Approximately 10,000 cell equivalents of genomic DNA were loaded in a 50a 5qPCR reaction using a custom designed HIV clade C primer and probe set designed to detect HIV *int* sequences derived from the patient population. Albumin was used as an internal control to quantify the number of genomes present. Both HIV and Albumin absolute copies were determined using an external standard curve. HIV Clade C primer and probe sequences are: Fwd 5’-GTTATYCCAGCAGARACAGG-3', Rev 5’-TGACTTTGRGGATTGTAGGG-3', probe 5'-RGCAGCCTGYTGGTGGGC-3'. Human albumin primer and probe sequences are: Fwd 5’-TGCATGAGAAAACGCCAGTAA-3’, Rev 5’-ATGGTCGCCTGTTCACCAA-3’, probe 5'-FAM-TGACAGAGTCACCAAATGCTGCACAGAA-3'. qPCRs were performed using the Taqman Universal master mix (Life Technologies), 0.2 µU of each primer, and 0.125 µo of probe. All assays were performed on the ABI 7500 systems (Life Technologies).

Statistical Analyses

All statistical analysis was performed using JMP, version 11 (SAS Institute Inc., Cary, NC). All bivariate continuous correlations were performed using standard linear regression. One-way comparison of means was performed using the Student’s t-test, and one-tailed p-values are reported. Kaplan-Meier survival curves and Cox proportional hazards models were performed using an endpoint defined as a single CD4+ T cell count reading less than 300, unless otherwise specified, and statistics reported for survival

analyses are generated from the log-rank test. This endpoint for CD4 loss provided the strongest predictive value for progression to CD4 counts <200, while maintaining the largest number of events for statistical power.

Principal Component Analysis (PCA) was performed using the JMP version 11 statistical package. PCA was used in place of multiple linear regression to evaluate inflammatory marker levels and their relationship to vRC for at least two reasons. First, PCA represents an unsupervised approach, where the data is not fit to a particular outcome (such as vRC), but is used to look for variation in the entire data set and to extract latent variables that consist of analytes that co-vary and that may be related to biological phenotypes. Second, the protein expression levels of many of these inflammatory markers are highly correlated, leading to instability in multiple regression models based on a high degree of collinearity. PCA avoids this problem of collinearity by extracting combinations of linear variables (principal components) that are themselves, by definition, uncorrelated with one another. Before PCA was performed, raw data was preprocessed. Extreme positive values for each of the 16 analytes measured in plasma were Winsorized to the 90th percentile. Missing values were imputed using multiple linear regression models, and individuals for whom more than 3 cytokine values were missing were excluded. All 16 analytes that were measured in 33 individuals were used to extract latent variables. The appropriate number of latent variables to extract (6 latent variables) was determined via a Scree plot, and each latent variable was required to have an eigenvalue >1 (encompassing the effects of at least 1 analyte measured). Significance of each latent variable was evaluated using the Bartlett test.

Depiction of cytokine loadings for principal components 1 and 2 (PC1, PC2), Figure 2C, was generated using Cytoscape v3.0. Briefly, the loading matrix (list of pairwise correlations between each latent variable and each of the 16 analytes tested) was extracted from the PCA, and used to represent the strength of contribution of each analyte to either PC1 or PC2. Dark red arrows depict the strongest positive correlations (r >0.75), while blue arrows depict the negative correlations (r<-0.35). The intensity of the color depicts the strength of individual contributions to each PC.

Results

Viral characteristics determine HIV-1 pathogenesis

Previous work by our group and others has shown that transmitted viral characteristics significantly correlate with early SPVL (*10*) as well as CD4+ T cell decline up to three years post infection (*12*). Here, we sought to determine if the vRC of HIV-1, conferred by the *gag* gene alone, durably affects the trajectory of CD4+ T cell decline even in the chronic stages of infection, and if this effect is independent of both host factors and SPVL that are known to modulate disease severity.

To assess the replicative capacity conferred by the transmitted *gag* sequence, we amplified the *gag* gene from plasma virus during acute infection time points (median 46 days post estimated date of infection), generated replication competent virus by cloning the *gag* gene into a common proviral backbone, and assessed vRC in an *in vitro* cell culture assay as described previously (*12*). In comparing five Gag-MJ4 chimeric viruses to transmitted/founder full-length infectious molecular clones derived from the same individuals, we find a strong positive correlation between the RC of the chimeric viruses and the particle to infectivity ratio of the full-length infectious molecular clones (**Supplementary Figure 1**). This indicates that although other genes undoubtedly play a role in defining *in vitro* HIV-1 replicative capacity, the contributions of Gag are a significant component of the replicative capacity of the full-length virus.

In this cohort of 127 acutely infected individuals from Zambia, low vRC significantly delayed the time to CD4+ T cell counts <300 for up to 5 years post infection **(Figure 1A)**. The clearest benefit is observed with the lowest vRC tercile compared to the

middle and highest tercile. A significant benefit remained even down to CD4 counts of <200, the clinical definition of AIDS, when individuals infected with intermediate and highly replicating viruses were combined into one group **(Figure 1B)**.

We have previously shown an association between early SPVL and vRC (*12*). Thus, we sought to more definitively determine if the replication capacity defined by the *gag* gene affected CD4 decline in a manner linked to, or independent of, the well-documented effect of early SPVL on subsequent disease progression. In this cohort, we found SPVLs $>10^5$ RNA copies/ml to be associated with poor outcomes for all volunteers **(Supplemental Figure 2A)**; however, vRC significantly dichotomized the trajectory of CD4 decline ($p<0.0001$) in individuals with SPVLs $<10^5$ **(Figure 1C)**, a majority (77%) of the cohort. This suggested independent, but additive effects, of both vRC and early SPVL on HIV disease progression. This is confirmed in a multivariable Cox proportional hazards model assessing the relative risk of vRC in the context of other well-established predictors of HIV disease progression. We find that low vRC, early SPVL, and canonical protective HLA class I alleles (B*57, B*5801) were each independent predictors of CD4 decline **(Supplemental Table 1)**. Moreover, vRC significantly dichotomizes disease trajectories of those with these protective HLA alleles **(Figure 1D)**. Taken together, these data firmly establish vRC as a distinct contributor to HIV disease progression. Moreover, they suggest that vRC may modulate innate immune events very early after infection, which could alter both the establishment of an inflammatory state and the development of an effective adaptive immune response capable of controlling viremia, that together may alter the trajectory of disease. To further test this hypothesis, we assessed early levels

of circulating inflammatory cytokines, immune activation and exhaustion in T cell compartments, as well as viral burden in memory CD4+ T cells.

Viral replicative capacity alters early inflammatory cytokine profiles

Acute HIV infection is characterized by a rapid and robust expression of type I interferons (IFN-I), IFN-I stimulated genes, and inflammatory cytokines (*13*). Disruption of the Gut Associated Lymphoid Tissue (GALT) and subsequent microbial translocation have also been shown to contribute significantly to this inflammatory state, possibly through a positive feedback loop (*14*). This inflammatory response, particularly during chronic infection, contributes to disease progression (*15-18*). Therefore, we analyzed the levels of 16 inflammatory cytokines, chemokines, and markers of gut damage and microbial translocation at or before seroconversion to assess the effect of vRC on the early inflammatory milieu (n = 33; previously dichotomized into low and high vRC phenotypes (*12*)). We found that vRC was positively correlated with a number of inflammatory cytokines **(Supplemental Table 2)**, most notably IL-6 and IL-1β, two pro-inflammatory cytokines previously implicated in driving aberrant CD4+ T cell turnover and impairing homeostatic proliferation (*19*). Of note, vRC was most strongly correlated with elevated levels of IL-10, an important anti-inflammatory cytokine linked to T cell dysfunction in HIV infection (*20, 21*).

In an effort to define distinct inflammatory “profiles”, we employed an unsupervised data reduction tool, Principal Component Analysis (PCA), which groups linear variables into combinations, termed principal components (PCs) **(Figure 2A)**. Strikingly, principal component 1 (PC1), which describes the greatest variation in the

data set, significantly correlates with vRC, in that individuals with positive loadings for PC1 (elevated levels of inflammation) tend to have higher vRC, while those with negative loadings (low levels of inflammation) are significantly enriched for poorly replicating viruses **(Supplemental Figure 3A)**. Moreover, principal component 2 (PC2), which by definition is uncorrelated with PC1, describes the second greatest variation in the data and is significantly correlated with SPVL **(Supplemental Figure 3B)** but not vRC. The differences in analyte loadings between PC1 and PC2 are shown in **Figure 2B**, and the inflammatory cytokines that substantially contribute to PC 1 and 2 are depicted schematically in **Figure 2C**. This further highlights the independence of factors associated with initial viral replication and subsequent adaptive immune control of SPVL. Moreover, it demonstrates that viruses with high vRC are correlated with a distinct inflammatory cytokine profile characterized by a heightened type I and type II interferon response and elevated levels of key inflammatory cytokines such as IL-6 and IL-1β.

Replicative capacity is associated with levels of CD8+T cell activation and exhaustion

Chronic immune activation is a hallmark of HIV-1 infection; it often persists following ART and is a more reliable predictor of disease progression than viral load (*22-26*). Therefore, we assessed the impact of vRC on levels of cellular immune activation by measuring the co-expression of CD38 and HLA-DR on CD8+ T cells isolated within three months post-infection for a subset of individuals under study (n=35). We found that vRC is positively correlated with the expression of CD38 and HLA-DR on CD8+ T cells (**Figure 3A**). Consistent with previous studies (*22, 24, 26*), we observe that higher CD8+ T cell activation in this cohort is associated with faster CD4+ T cell decline

(**Supplementary Figure 4A**), thus positioning T cell activation as a link between vRC and subsequent CD4+ T cell decline.

In addition to cellular immune activation, CD8+ T cell exhaustion is characteristic of pathogenic HIV/SIV infection. Exhaustion of CD8+ T cells is marked by the increased expression of the inhibitory receptor Programmed Death 1 (PD-1), and levels of PD-1 expression predict the rate of disease progression (*27, 28*). PD-1^{hi} CD8+ T cells are typically $CD57^{low}$ (*27, 28*), while, in contrast, Petrovas et al. showed that PD-$1^{low}CD57^{high}$ CD8+T cells are more resistant to apoptosis (*29*). We find that a greater percentage of CD8+ T cells isolated from individuals infected with low RC viruses displayed high levels of CD57 while maintaining low levels of PD-1 (**Figure 3B**) relative to individuals infected by high RC viruses. In CD8+ T cells, markers of exhaustion are often associated with impaired cytotoxic function (*30*). Indeed, dual expression of granzyme B and perforin was positively correlated with the frequency of PD-$1^{-}CD57^{+}$ $CD8^{+}$ T cells (**Figure 3C**) and was inversely correlated with vRC (**Figure 3D**). Consequently, CD57 expression is protective in terms of CD4+ T cell decline while PD-1 expression is associated with faster disease progression in this cohort (**Supplementary Figure 4B and 4C**). Taken together, these data suggest that individuals infected with low RC viruses mount a more functional and less exhausted CTL response early in infection that may provide an extended clinical benefit for the individual.

Replicative capacity predicts levels of CD4+ T cell activation, proliferation, and exhaustion

Immune activation in HIV-1 infection is associated with global immunological dysfunction, characterized by increased cellular-turnover, and ultimately the disruption

of critical CD4+ T cell homeostasis, indicative of progressive disease (*31-33*). The percentage of CD4+ T cells expressing CD38/HLA-DR and Ki67 were measured in PBMCs isolated less than 3-months post infection for a subset of individuals (n=19, see methods for sample selection). Expression of these markers was assessed on total CD4+ T cells as well as in the context of different CD4+ T cell subsets: naïve (T_N, CD27+, CD45RO-, CCR7+), central memory (T_{CM}, CD27+/-, CD45RO+, CCR7+), and effector memory (T_{EM}, CD27 +/-, CD45RO+, CCR7-).

High vRC was associated with increased expression of CD38 and HLA-DR on total (**Supplementary Figure 5A,** P = 0.02) and T_{EM} (**Figure 4A,** P = 0.02) CD4+ T cells, however this association was most pronounced in the T_{CM} (**Figure 4B,** P = 0.006) compartment. Consistent with previous reports (*22, 24*), high levels of CD4+ T cell activation (CD38+/HLA-DR+) are also associated with faster CD4+ T cell decline (**Figure 4C**). Furthermore, individuals infected with high RC viruses exhibited a significantly greater percentage of CD4+ T cells expressing Ki67. This association was most striking in the effector memory T cell subset (**Figure 4D,** P = 0.003), but was also significant for total (**Supplementary figure 5B,** P = 0.006) and central memory CD4+ T cells (**Figure 4E,** P = 0.008). Increased CD4+ T cell proliferation, demonstrated by elevated expression of Ki-67 on CD4+ T cell subsets, was also associated with faster CD4+T cell decline, as has been reported previously **(Figure 4F)** (*32, 34*). Furthermore, individuals infected with poorly replicating viruses displayed expression levels of CD38/HLA-DR and Ki67 on their CD4+ T cells, which, although somewhat elevated, for the most part did not differ significantly when compared to HIV-uninfected Zambians **(Figure 4 A-D)**. These results suggest that, shortly after transmission, individuals infected with low vRC viruses preserve an immune

system more similar to that of a healthy individual.

Early inflammatory cytokine profiles associated with vRC are linked to activated T cell phenotype

To determine whether changes in T cell activation are associated with the distinct inflammatory cytokine profiles associated with vRC, we compared individuals with positive (increased inflammatory cytokine levels) and negative loadings for Principal Component 1 (PC1; **shown in Figure 2**). Individuals with positive loadings for PC1 presented with significantly increased CD8+ T cell activation (**Figure 5A**) and exhaustion (**Figure 5B and 5C**). Positive loadings were further associated with increased frequency of $PD1^{+}$ $CD4^{+}$ T_{CM} cells (**Figure 5D**), as well as higher levels of CD4+ T_{EM} cell proliferation, as measured by Ki67 expression (**Figure 5E**). Thus, higher levels of inflammatory cytokines are closely linked to the activated, exhausted T cell phenotype observed in individuals infected with high RC viruses.

High vRC is associated with increased proviral burden in CD4+ T cell subsets

Establishment of the latent HIV reservoir occurs early during acute infection and sets the stage for viral persistence (*35-38*). Thus, even in the context of suppressive antiretroviral therapy, HIV-1 cannot be fully eradicated. To determine if the replicative capacity of the transmitted/founder virus influences the proviral DNA burden during early infection, levels of cell-associated viral DNA were measured in naïve, central memory, and effector memory CD4+ T cell subsets sorted from PBMCs isolated 3-months post-infection (n=21). We observed that high vRC was associated with a

significant increase in the amount of cell-associated viral DNA in both T_{CM} (P = 0.01) and T_N (P = 0.001) CD4+ subsets (**Figure 6A**). Because high vRC was associated with increased levels of immune activation and proliferation, we sought to determine whether a direct association exists between these markers and the magnitude of viral burden in $T_{CM,}$ a key population in CD4+ T cell homeostasis. We find that HIV-1 DNA in T_{CM} positively correlates with expression of CD38+/HLADR+ on T_{EM} (**Figure 6B**, P<0.0001) and, to a lesser extent, on T_{CM} ($R^2 = 0.3$, P = 0.01, data not shown). Similarly, HIV-1 DNA in T_{CM} was positively correlated with Ki67 expression on T_{EM} CD4+ cells (**Figure 6C**). Moreover, higher levels of cell associated HIV DNA in central memory CD4+ T cells was associated with an accelerated loss of CD4+ T cells (**Figure 6D**).

Discussion

Untreated HIV-1 infection is typically characterized by a progressive depletion of CD4+ T cells and eventual development of fatal immunodeficiency. However, the disease course and kinetics vary between individuals and different HIV-1 subtypes (*3, 39, 40*). Although numerous studies have implicated host immunogenetic factors in governing the tempo of disease progression (*41*), viral characteristics have received less scrutiny. Recent studies have implicated transmitted viral characteristics in explaining the heritability of HIV-1 pathogenesis (*42, 43*). In support of this, we previously showed that the replicative capacity of the transmitted virus, as conferred by the transmitted *gag* sequence, is correlated both with chronic donor viral loads near the time of transmission and with SPVL in recently infected recipients, establishing vRC as a heritable viral trait impacting HIV-1 disease (*12*).

Here we have highlighted an integral role for replicative capacity of the transmitted virus, as defined by Gag, in driving multiple facets of HIV-1 immunopathology. In a cohort of Zambian seroconvertors acutely infected with HIV-1 subtype C we demonstrate that vRC is (I) an independent contributor to CD4+ T cell decline in conjunction with host immunogenetic factors that dictate SPVL, (II) linked to an inflammatory state early in infection that is characterized by elevated levels of key inflammatory cytokines known to drive pathogenesis, (III) linked to aberrant CD8+ and CD4+ T cell phenotypes characterized by increased levels of cellular activation, exhaustion, and proliferation, and correlated with CD4+ T cell decline within this cohort and (IV) associated with increased viral burden in naïve CD4+ T cells and CD4+ T_{CM}, the

latter being integral in establishing the latent viral reservoir (*36*) and preferentially spared in non-pathogenic SIV infection of natural hosts (*44*). Taken together, these results support an unprecedented role for transmitted viral characteristics, specifically viral replicative capacity, in determining the early inflammatory state of HIV-1 infected individuals and their subsequent disease trajectory.

A critical role for viral replicative capacity in disease pathogenesis is consistent with studies of the lymphocytic choriomeningitis virus (LCMV) infection model. In the case of this virus, the higher replicative capacity the LCMV clone 13 strain goes on to induce chronic infection while the more attenuated Armstrong strain is quickly contained by the immune system (*45*). Although with this cohort we were unable to accurately measure peak viral load, it seems likely that viruses with high replicative capacity have the potential to induce much higher levels of peak viremia, which then induces exacerbated immunopathology that cannot be reversed by the immune response.

Previous studies show that the level of immune activation is established very early after HIV-1 infection and varies greatly between individuals, but remains quite stable over time (*22*). Furthermore, immune activation predicts disease progression even when viral loads are suppressed either immunologically or by ART (*22, 24-26*). For the first time, we demonstrate that levels of immune activation established during early infection can be significantly attributed to the replicative capacity of the transmitted virus, despite only measuring the contribution of the *gag* gene to this process.

We furthermore demonstrated that the amount of HIV-1 DNA harbored by both T_{CM} and naïve (CD27+/CD45RO-) CD4+ T cells is highly associated with viral replicative capacity. It is possible that the increased levels of HIV-1 DNA found in these “naïve”

CD4+ T cells could be due in part to the presence of T memory stem cells (T_{SCM}), that would be found within our sorted naïve T cell populations (*46, 47*). These long-lived memory T cells with stem cell-like properties are susceptible to HIV-1 or SIV in pathogenic infections but spared in non-pathogenic SIV disease (*48*).

T_{CM} CD4+ cells, which are characterized as being more stable and longer-lived, are preferentially spared in non-pathogenic SIV infection and infection of this subset has been linked to the immunopathology of HIV-1 infection (*32, 44, 49*). The results of the current study provide further evidence that HIV-1 infection of this subset occurs early, is linked to increased CD4+ T cell activation, and significantly predicts disease progression, perhaps by disrupting the capacity to renew the CD4+ T_{EM} population (*32*).

T_{CM} CD4+ cell have also been highlighted as an integral population for the maintenance of latency and viral persistence (*36*). Our findings suggest that the extent to which HIV infects T_{CM} CD4+ T cells and establishes a latent reservoir might be influenced by characteristics of the transmitted/founder virus. This has important therapeutic implications, and suggests that individuals infected with attenuated (low vRC) viruses may be better candidates for proof-of-concept cure strategies.

The data presented here suggests that viral characteristics can provide independent information about an individual's risk for disease progression and point to a novel target for interventions to reduce immune activation and viral burden. Vaccine-induced immune responses or interventions that effectively attenuate vRC in the earliest stages of infection could not only have a dramatic impact on viral control and disease progression, but, based on our recent findings (*43*) could also impact the efficiency of subsequent transmission to other partners. Moreover, reducing the size of the viral

burden in key T cell populations prior to ARV treatment could significantly augment cure strategies aimed at eliminating the latent reservoir.

Acknowledgments

The investigators thank all the volunteers in Zambia who participated in this study and all the staff at the Zambia Emory HIV Research Project in Lusaka who made this study possible. The investigators would like to thank Emmanuel Cormier, Jon Allen, Sheng Luo, and Paul Farmer for technical assistance, sample management, and database management. We thank Kiran Gill and Barbara Cervasi at the Emory Vaccine Center/ Emory Center for AIDS Research Flow Core (grant P30 AI050409) for performing cell sorting and for assisting with flow cytometry experiments. We would also like to thank Rafick-Pierre Sékaly and Ali Filali for helpful discussions and important biostatistics support. This study was funded by R01 AI64060 and R37 AI51231 (E.H.). This work was also supported, in part, by the Virology Core at the Emory Center for AIDS Research (grant P30 AI050409); the Yerkes National Primate Research Center base grant (2P51RR000165-51) through the National Center for Research Resources P51RR165 and by the Office of Research Infrastructure Programs/OD P51OD11132. This study is supported in part by IAVI (SA), whose work is made possible by generous support from many donors including: the Bill & Melinda Gates Foundation; the Ministry of Foreign Affairs of Denmark; Irish Aid; the Ministry of Finance of Japan; the Ministry of Foreign Affairs of the Netherlands; the Norwegian Agency for Development Cooperation (NORAD); the United Kingdom Department for International Development (DFID), and the United States Agency for International Development (USAID). The full list of IAVI donors is available www.iavi.org. The contents are the responsibility of the study authors and do not necessarily reflect the views of USAID or the United States Government. J.P.

and D.T.C., were supported in part by Action Cycling Fellowships. E.H. is a Georgia Eminent Scholar.

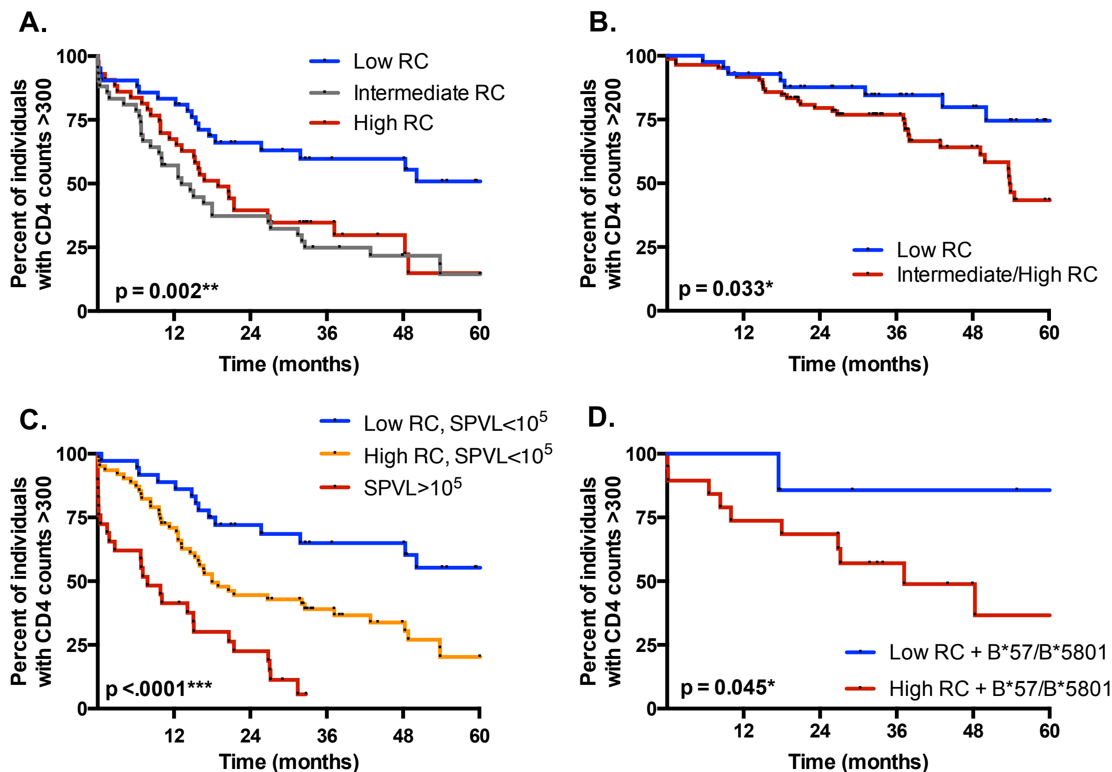


Fig. 1. HIV-1 replicative capacity, when defined by the transmitted Gag sequence, predicts CD4 T cell decline in ART-naïve, HIV-1 infected individuals. (**A**) In a Kaplan-Meier survival analysis with an endpoint defined as CD4⁺ T cells counts falling below 300, low vRC, defined as the lowest tercile of vRC values in 127 acutely infected individuals, provides a significant protective benefit from CD4 T cell decline. (**B**) When the middle and upper terciles of vRC groups are combined, low vRC provides significant protection from CD4⁺ T cell counts falling below 200. (**C**) vRC has an independent, but additive, effect with early set point VL in determining CD4 decline. High vRC is defined as the upper 67% of vRC scores, and low vRC is defined as the lowest 33% of vRC scores. (**D**) In individuals with protective HLA class I alleles (*n* = 26), vRC significantly dichotomizes CD4⁺ T cell decline trajectory.

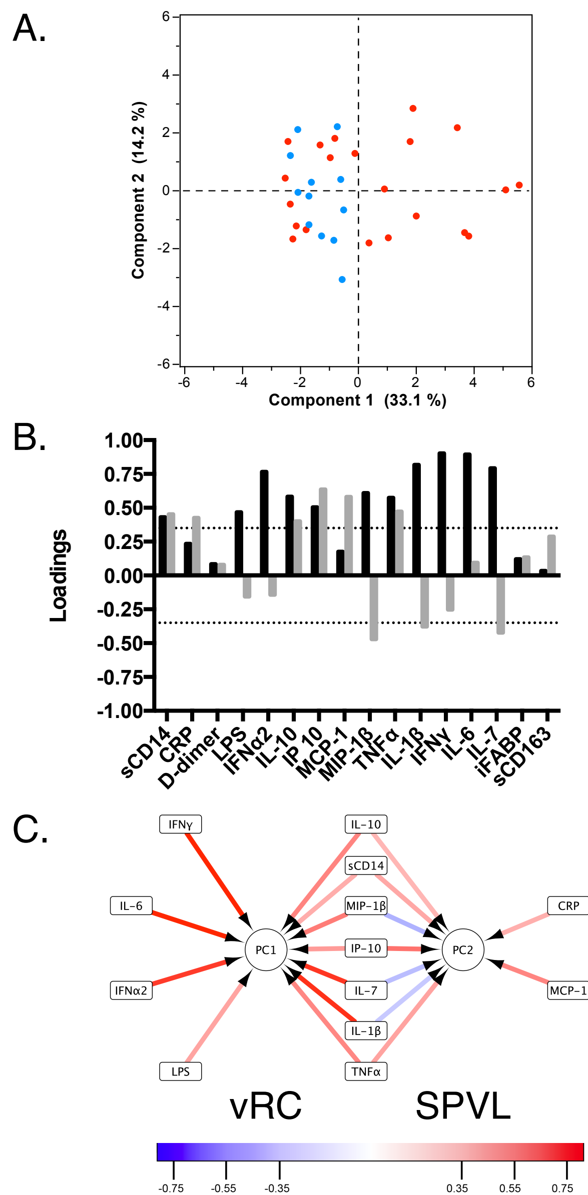


Fig. 2. Low vRC is associated with a distinct cytokine profile early in infection, characterized by muted inflammatory cytokine levels. Principal Component Analysis (PCA) was used to reduce dimensionality and extract latent variables comprised of linear combinations of 16 different analytes measured at the seroconversion time point (n = 33).

(**A**) Principal component 1 (PC1) and principal component 2 (PC2) scores are depicted in a 2-dimensional scatter plot for each individual. The first principal component (x-axis), which describes the greatest variation in the data set, has positive loadings associated with high vRC. Red = individuals with high vRC; Blue = individuals with low vRC. (**B**) Bar graph depicting analyte loadings for PC1 (black) and PC2 (gray). Loadings were extracted from the PCA correlation matrix and represent the degree of contribution of each analyte to both latent variables. Dotted lines represent pair-wise association thresholds ($r > 0.35$ or $r < -0.35$). (**C**) A relational network depicts the analytes most positively (red) and negatively (blue) associated with the first two principal components, which themselves are also correlated with two distinct biological phenotypes, vRC (PC1) and set point VL (PC2). The color key describes how the strength of linear dependence, based on the pair-wise (Pearson) correlation coefficient (r), corresponds to the color of connecting arrows (the threshold for visible connections was set to a pair-wise (Pearson) correlation coefficient of $r > 0.35$ or < -0.35).

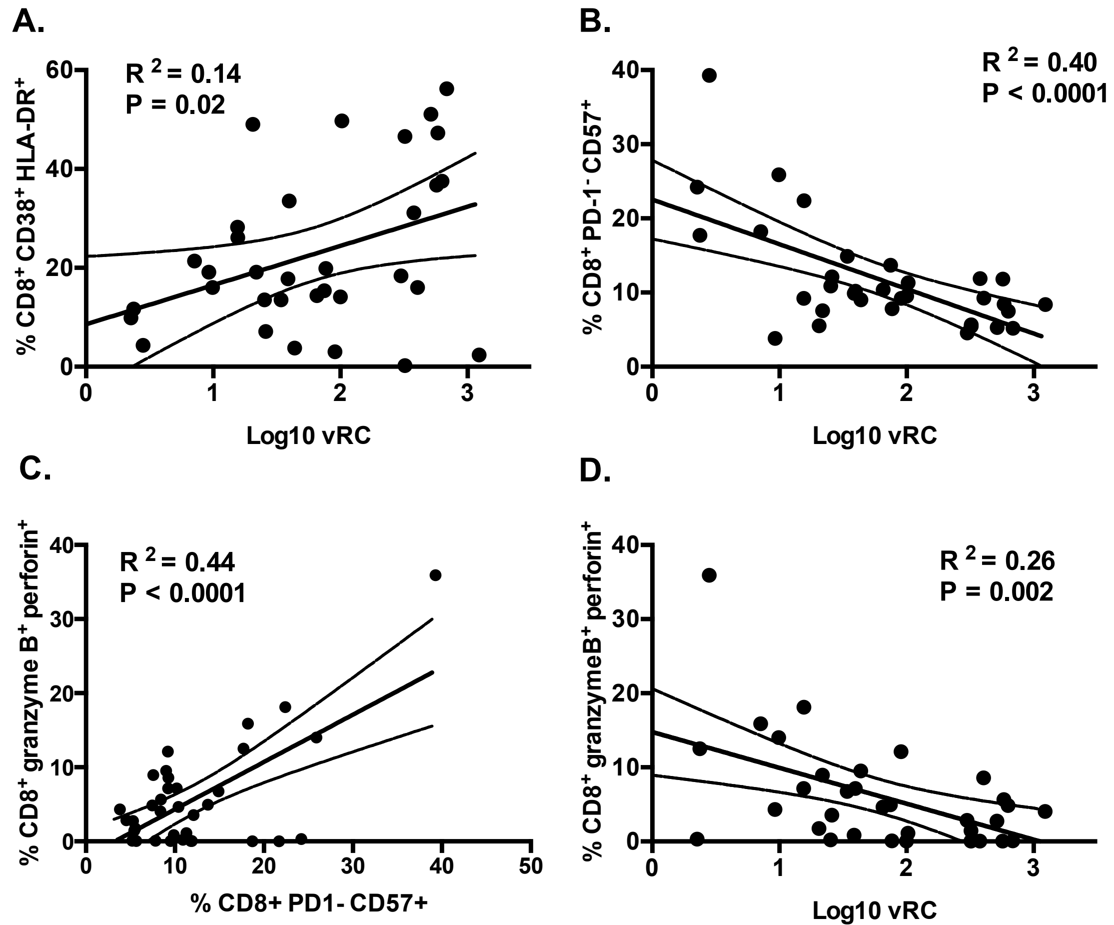


Fig. 3. High vRC is associated with increased CD8+ T cell activation and lower cytotoxic potential. (**A-D**) Cryopreserved PBMCs were collected less than 3-months post estimated date of infection (*n* = 35) and stained with two multi-color flow cytometry panels evaluating either memory subsets and markers of cellular activation and exhaustion (**A, B**) or markers of T cell cytotoxicity (**C, D**). (**A**) Correlation between vRC and the percentage of total CD8+ T cells co-expressing CD38 and HLA-DR. (**B**) Correlation between vRC and the percentage of CD8+ T cells that are $CD57^+$ but remain $PD\text{-}1^-$. (**C**) Individuals with a greater percentage of $CD57^+/PD\text{-}1^-$ CD8 T cells also have CD8+ T cells with greater cytotoxic potential, as measured by the percentage of cells

expressing both granzyme B and perforin. (**D**) Correlation between cytotoxic potential in CD8+ T cells, as measured by co-expression of granzyme B and perforin, and vRC. Correlation statistics were generated using linear regression. Solid lines indicate trend lines, and dashed lines represent 95% confidence bands.

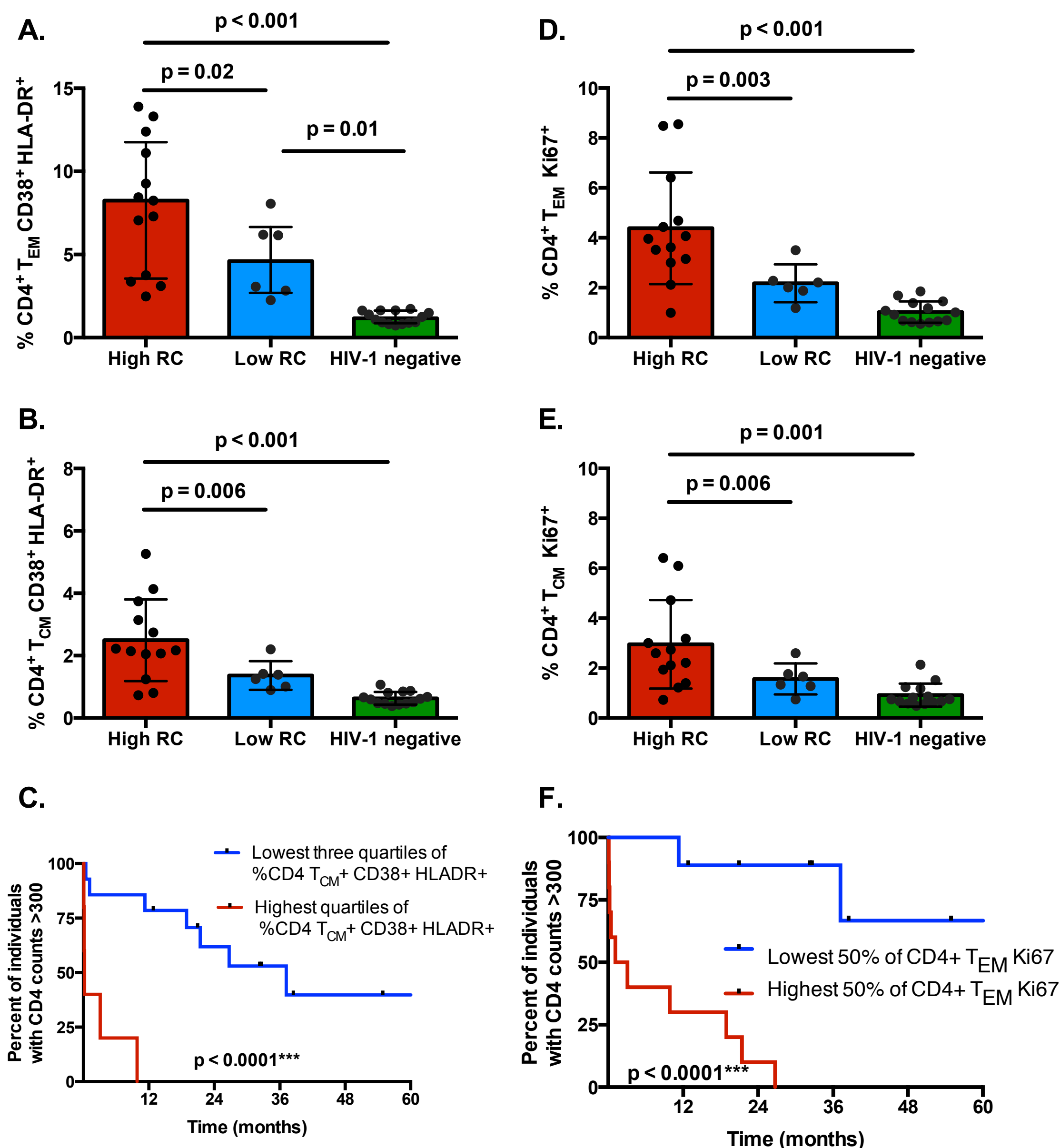


Fig. 4. Viral RC is associated with increased cellular activation and proliferation in CD4+ T cell memory subsets. (**A-D**) Cryopreserved PBMCs from individuals with high vRC (*n* = 13), low vRC (*n* = 6), and from HIV-negative Zambian individuals (*n* = 14) isolated at 3-months post estimated date of infection were stained for markers of activation, cellular turnover, and delineation of memory T cell subsets. (**A-B**) Co-expression of the activation markers CD38 and HLA-DR on CD4+ T_{EM} and T_{CM} cells was

significantly increased in individuals infected with highly replicating viruses as compared to individuals infected with attenuated viruses or HIV-negative individuals. The percentage of CD4+ T_{CM} cells co-expressing CD38 and HLA-DR did not differ significantly between individuals infected with poorly replicating viruses and those that were HIV-negative. (**D-E**) CD4+ T cell proliferation, as measured by Ki67 expression, was significantly higher in T_{CM} and T_{EM} of individuals infected with highly replicating viruses. There was no statistically significant difference in the percentage of T_{CM} and T_{EM} expressing Ki67 between individuals infected with attenuated viruses and HIV-negative individuals (statistical comparisons were made using the Student's t-test). (**C, F**) In a Kaplan-Meier survival analysis with an endpoint defined as CD4 T cell counts falling below 300, high percentages of CD4+ T_{CM} cells co-expressing CD38 and HLA-DR or CD4+ T_{EM} expressing Ki67 was highly deleterious and lead to rapid disease progression.

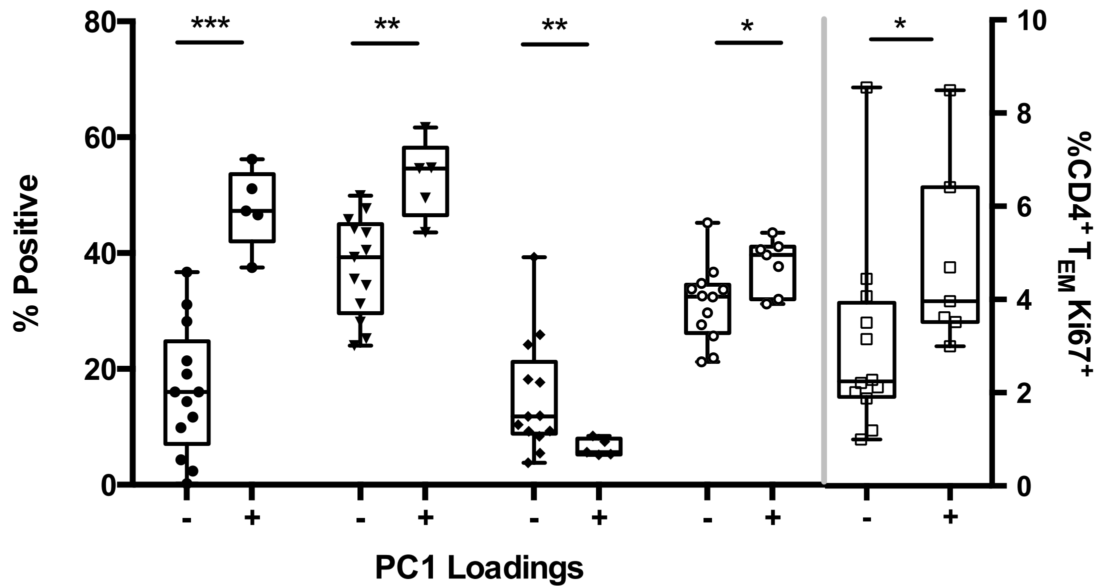


Fig. 5. Inflammatory cytokine profiles associated with vRC correlate with T cell activation. Inflammatory cytokines measured at seroconversion in 33 acutely infected individuals were used to define distinct cytokine profiles via Principal Component Analysis (**Figure 2**). Of this group, a subset of 18 and 19 individuals had also been immunophenotyped for CD8+ and CD4+ T cell activation, respectively. Positive loadings (increased inflammatory cytokines) were defined as a PC1 score >0 and negative loadings were defined as a PC1 score <0. Positive loadings for PC1 are associated with increased CD8+ T cell activation as demonstrated by increased percentages of CD38+/HLA-DR+ cells (**A**); and increased markers of CD8+ T cell exhaustion as demonstrated by increased percentages of PD-1+/CD57- cells (**B**) and decreased percentages of PD-1-/CD57+ cells (**C**). Positive loadings for PC1 are also associated with increased percentages of PD-1+ CD4+ T_{CM} cells (**D**) and increased percentages of Ki67+ CD4+ T_{EM} cells (**E**). The left hand y-axis displays percent positive cells for panels **A**-**D**; the right hand y-axis displays percent

Ki67+ cells for panel **E**. (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; statistical comparisons were made using the Student's t-test).

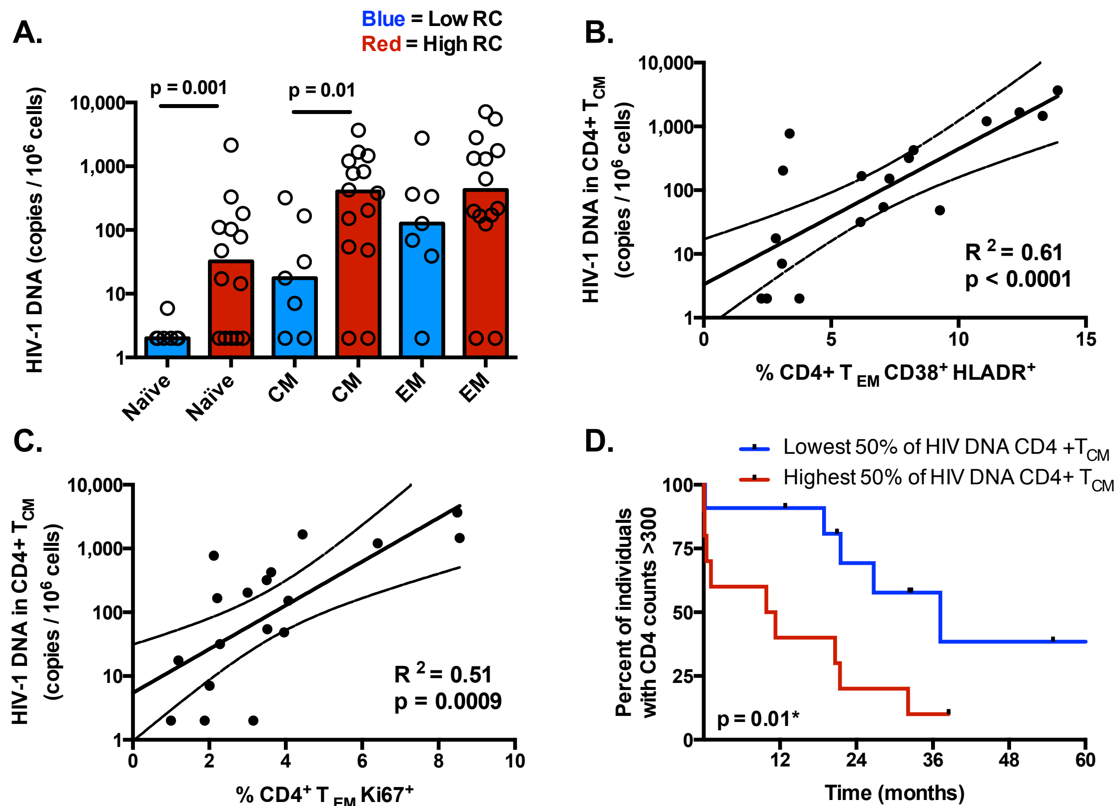
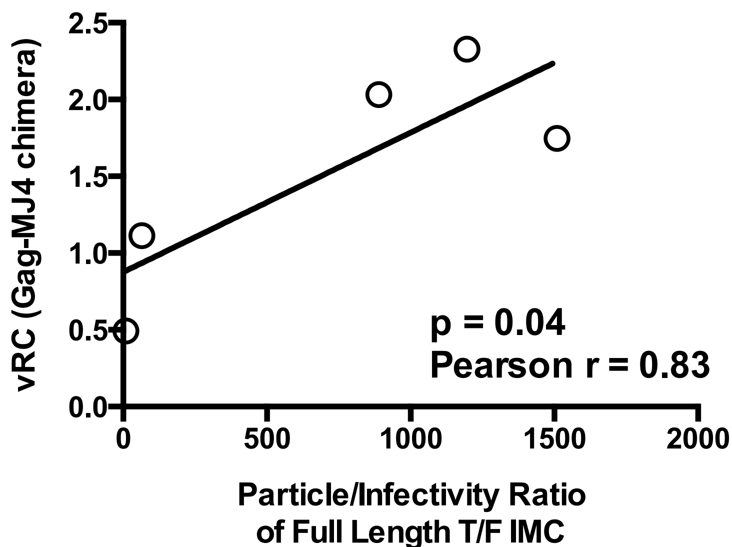


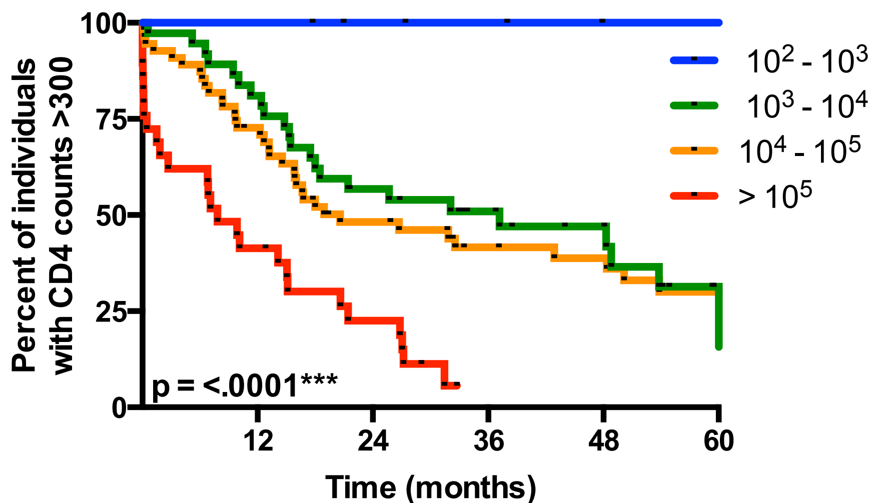
Fig. 6. vRC correlates with the burden of HIV-1 viral DNA in CD4+ T_{CM} and T_N.

(**A**) Cryopreserved PBMCs isolated 3-months post infection (*n* = 21) were stained and CD4+ T cells were sorted into 3 distinct populations, T_N ($CD27^+$, $CD45RO^-$, $CCR7^+$), T_{EM} ($CD27^{+/-}$, $CD45RO^+$, $CCR7^-$), and T_{CM} ($CD27^{+/-}$, $CD45RO^+$, $CCR7^+$). Cell-associated HIV-1 DNA was quantified by real-time PCR amplification of the HIV-1 subtype C *integrase* gene. Cell copy number was normalized based on real-time PCR amplification of the human *albumin* gene. Low vRC was associated with significantly fewer CD4+ T_N and T_{CM} cells harboring HIV-1 DNA. (**B-C**) In a majority of the same individuals (*n* = 18), a fraction of unsorted PBMCs was stained for markers of cellular activation and proliferation. The viral burden in CD4+ T_{CM} was positively correlated with levels of proliferation (**C**) and activation (**B**) in CD4+ T_{EM} cells. Correlation statistics were

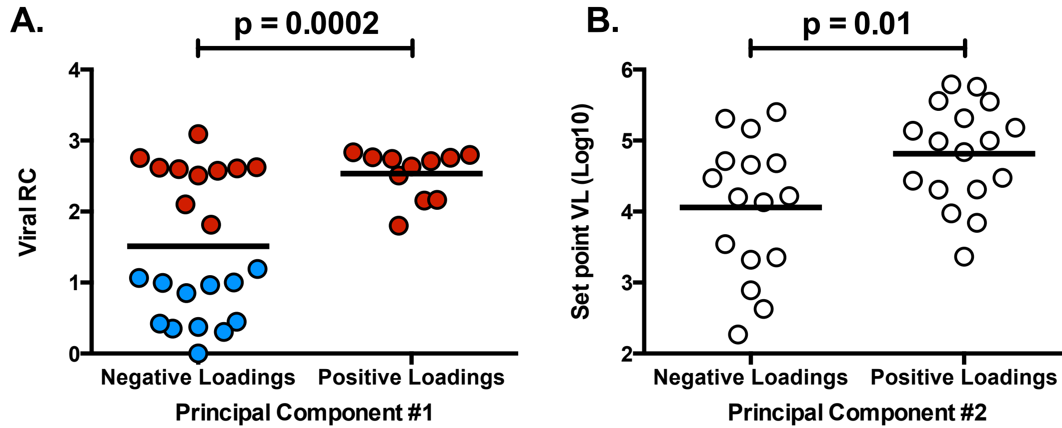
generated using linear regression and $\log_{10}$-transformed viral DNA copies/10^6 cells. (**D**) In a Kaplan-Meier survival analysis with an endpoint defined as CD4 T cell counts falling below 300, high viral burden in CD4+ T_{CM} was associated with an increased risk of CD4+ T cell decline.



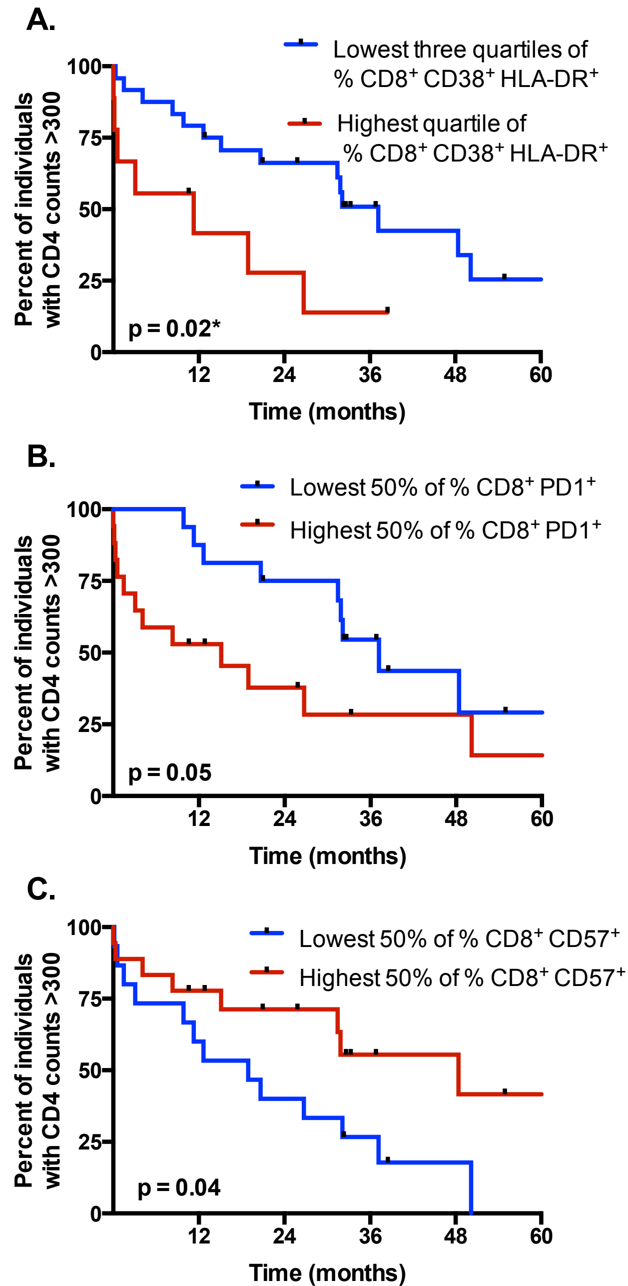
Supplementary Figure 1. The *gag* gene chimera is representative of full-length HIV-1. Gag-MJ4 chimeras were generated by amplifying the *gag* gene from plasma collected at seroconversion and cloning *gag* into the subtype C, primary isolate, MJ4. Full-length infectious molecular clones were generated by amplifying viral RNA-derived near full-length genomes from the same seroconversion plasma sample and cloning these viral genomes into a plasmid backbone containing patient-derived long terminal repeats in order to generate full-length HIV-1 infectious molecular clones. Particle to infectivity ratios were calculated by dividing the infectious units per ul of virus stock by the reverse-transcriptase (RT) activity per ul of virus stock. Statistics were generated using the Pearson correlation, and the reported p-value is one-tailed.



Supplementary Figure 2. The effect of $\log_{10}$-increases in early set point VL on longitudinal CD4 T cell decline post seroconversion for the ZEHRP cohort. In a Kaplan-Meier survival analysis with an endpoint defined as reaching CD4 T cell counts <300, individuals with set point VLs greater than 1 x 10^5 were at the highest risk of rapid CD4 T cell decline (n = 127). Groups were defined as the $\log_{10}$-transformed set point VL measurement rounded down to the nearest power.

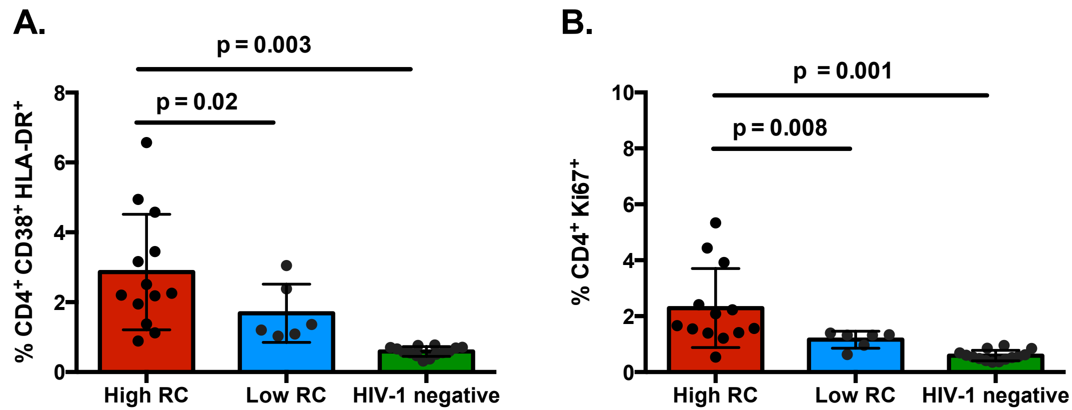


Supplementary Figure 3. The first two principal components are significantly correlated with vRC and set point VL, respectively (n = 33; statistical comparisons were made using the Student's t-test). **A.** Positive loadings for PC#1, which are associated with high levels of inflammatory cytokines, are significantly enriched for individuals with high vRC. **B.** Positive loadings for PC#2, which are associated with high levels of IP-10 and MCP-1, are significantly enriched for individuals with higher set point VLs.



Supplementary Figure 4. CD8 T cell activation phenotypes early after infection are associated with the rate of CD4 T cell decline. **A-C.** Kaplan-Meier survival analysis with an endpoint defined as CD4 T cell counts <300 and statistics generated from the log-rank test (n=35). **A.** The highest quartile of CD38 and HLA-DR co-expression on CD8 T cells is significantly correlated with accelerated CD4 T cell decline. **B.** High levels of PD-1

expression on CD8 T cells show a trend for accelerated CD4 T cell decline. **C.** Individuals with a high percentage of $CD57^+$ CD8 T cells exhibit delayed CD4 T cell decline.



Supplementary Figure 5. High vRC is associated with an increased level of activation and proliferation in CD4 T cells that is highly deleterious. **A.** High vRC is associated with an increased percentage of total CD4 T cells that co-express CD38 and HLA-DR as compared to individuals infected with attenuated viruses or to HIV-negative individuals. Individuals infected with attenuated viruses exhibit CD4 activation phenotypes that more closely resemble that of uninfected individuals. **B.** High vRC is associated with increased proliferation of total CD4 T cells as evidenced by intracellular Ki67 expression. Individuals infected with attenuated viruses more closely resemble uninfected individuals in terms of the percentage of total CD4 T cells expressing Ki67.

	Cox Proportional Hazards Model (Time to CD4<300)		
Factors Tested	**HR**	**95% CI**	***P*-value**
Female	1.10	0.67 – 1.70	.78
Low vRC (lowest tercile)	0.48	0.28 – 0.80	0.004
B*57/5801	0.45	0.23 – 0.81	0.006
Set point VL	10.00	2.86 – 44.1	0.0004

Supplementary Table 1. Host and viral characteristics independently predict CD4 T cell decline. In a multivariable Cox proportional hazards model with an endpoint defined as CD4 T cell counts <300, low vRC, carriage of HLA-B*57 or B*5801, and set point VL were independent predictors of CD4 decline (n = 127). Transmission of a virus exhibiting low vRC (defined as the lowest tercile of vRC scores) and carriage of protective HLA-I alleles were found to be independently and similarly protective, while high set point VLs were hazardous.

Analyte	Low vRC (mean pg/mL)	High vRC (mean pg/mL)	p-value
IL-10	5.5	10.73	0.004
IL-6	1.88	3.94	0.004
IL-1β	0.21	0.57	0.008
IFNγ	4.38	10.26	0.014
IP-10	639.56	1108.43	0.018
TNFα	10.76	13.87	0.028
IL-7	1.81	2.65	0.046
IFNα2	21.09	31.48	0.048

Supplementary Table 2 High vRC significantly increases early inflammatory cytokine levels. In an analysis of 16 inflammatory cytokines, chemokines, and markers of gut integrity and microbial translocation in plasma at the time of seroconversion (median 46 days post estimated date of infection), 8 analytes were found to be significantly increased in individuals infected with highly replicating viruses (n = 33; statistical comparisons were made using the Student's t-test, p-values are one-tailed). Low vRC is defined as the lowest tercile of vRC values in the full data set (n = 127).

References

1. H. Masur, F. P. Ognibene, R. Yarchoan, J. H. Shelhamer, B. F. Baird *et al.*, CD4 counts as predictors of opportunistic pneumonias in human immunodeficiency virus (HIV) infection. *Annals of internal medicine* **111**, 223-231 (1989).

2. C. B. Small, R. S. Klein, G. H. Friedland, B. Moll, E. E. Emeson *et al.*, Community-acquired opportunistic infections and defective cellular immunity in heterosexual drug abusers and homosexual men. *The American journal of medicine* **74**, 433-441 (1983).

3. A. Munoz, M. C. Wang, S. Bass, J. M. Taylor, L. A. Kingsley *et al.*, Acquired immunodeficiency syndrome (AIDS)-free time after human immunodeficiency virus type 1 (HIV-1) seroconversion in homosexual men. Multicenter AIDS Cohort Study Group. *American journal of epidemiology* **130**, 530-539 (1989).

4. M. P. Martin, M. Carrington, Immunogenetics of HIV disease. *Immunological reviews* **254**, 245-264 (2013).

5. J. Fellay, D. Ge, K. V. Shianna, S. Colombo, B. Ledergerber *et al.*, Common genetic variation and the control of HIV-1 in humans. *PLoS genetics* **5**, e1000791 (2009).

6. C. A. Derdeyn, J. M. Decker, F. Bibollet-Ruche, J. L. Mokili, M. Muldoon *et al.*, Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**, 2019-2022 (2004).

7. B. F. Keele, E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham *et al.*, Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7552-7557 (2008).

8. F. M. Hecht, W. Hartogensis, L. Bragg, P. Bacchetti, R. Atchison *et al.*, HIV RNA level in early infection is predicted by viral load in the transmission source. *AIDS* **24**, 941-945 (2010).

9. T. D. Hollingsworth, O. Laeyendecker, G. Shirreff, C. A. Donnelly, D. Serwadda *et al.*, HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS pathogens* **6**, e1000876 (2010).

10. L. Yue, H. A. Prentice, P. Farmer, W. Song, D. He *et al.*, Cumulative impact of host and viral factors on HIV-1 viral-load control during early infection. *Journal of virology* **87**, 708-715 (2013).

11. P. A. Goepfert, W. Lumm, P. Farmer, P. Matthews, A. Prendergast *et al.*, Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *The Journal of experimental medicine* **205**, 1009-1017 (2008).

12. J. L. Prince, D. T. Claiborne, J. M. Carlson, M. Schaefer, T. Yu *et al.*, Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. *PLoS pathogens* **8**, e1003041 (2012).

13. A. R. Stacey, P. J. Norris, L. Qin, E. A. Haygreen, E. Taylor *et al.*, Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *J Virol* **83**, 3719-3733 (2009).

14. J. M. Brenchley, D. A. Price, T. W. Schacker, T. E. Asher, G. Silvestri *et al.*, Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nature medicine* **12**, 1365-1371 (2006).

15. S. M. Keating, E. S. Jacobs, P. J. Norris, Soluble mediators of inflammation in HIV and their implications for therapeutics and vaccine development. *Cytokine & growth factor reviews* **23**, 193-206 (2012).

16. L. H. Kuller, R. Tracy, W. Belloso, S. De Wit, F. Drummond *et al.*, Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. *PLoS medicine* **5**, e203 (2008).

17. N. G. Sandler, H. Wand, A. Roque, M. Law, M. C. Nason *et al.*, Plasma levels of soluble CD14 independently predict mortality in HIV infection. *The Journal of infectious diseases* **203**, 780-790 (2011).

18. S. A. Vaidya, C. Korner, M. N. Sirignano, M. Amero, S. Bazner *et al.*, Tumor Necrosis Factor alpha Is Associated With Viral Control and Early Disease Progression in Patients With HIV Type 1 Infection. *The Journal of infectious diseases*, (2014).

19. C. L. Shive, J. C. Mudd, N. T. Funderburg, S. F. Sieg, B. Kyi *et al.*, Inflammatory Cytokines Drive CD4+ T-Cell Cycling and Impaired Responsiveness to Interleukin 7: Implications for Immune Failure in HIV Disease. *The Journal of infectious diseases*, (2014).

20. M. A. Brockman, D. S. Kwon, D. P. Tighe, D. F. Pavlik, P. C. Rosato *et al.*, IL-10 is up-regulated in multiple cell types during viremic HIV infection and reversibly inhibits virus-specific T cells. *Blood* **114**, 346-356 (2009).

21. E. A. Said, F. P. Dupuy, L. Trautmann, Y. Zhang, Y. Shi *et al.*, Programmed death-1-induced interleukin-10 production by monocytes impairs CD4+ T cell activation during HIV infection. *Nature medicine* **16**, 452-459 (2010).

22. S. G. Deeks, C. M. Kitchen, L. Liu, H. Guo, R. Gascon *et al.*, Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. *Blood* **104**, 942-947 (2004).

23. J. V. Giorgi, H. N. Ho, K. Hirji, C. C. Chou, L. E. Hultin *et al.*, CD8+ lymphocyte activation at human immunodeficiency virus type 1 seroconversion: development of HLA-DR+ CD38- CD8+ cells is associated with subsequent stable CD4+ cell levels. The Multicenter AIDS Cohort Study Group. *The Journal of infectious diseases* **170**, 775-781 (1994).

24. P. W. Hunt, J. Brenchley, E. Sinclair, J. M. McCune, M. Roland *et al.*, Relationship between T cell activation and CD4+ T cell count in HIV-seropositive individuals with

undetectable plasma HIV RNA levels in the absence of therapy. *The Journal of infectious diseases* **197**, 126-133 (2008).

25. P. W. Hunt, J. N. Martin, E. Sinclair, B. Bredt, E. Hagos *et al.*, T cell activation is associated with lower CD4+ T cell gains in human immunodeficiency virus-infected patients with sustained viral suppression during antiretroviral therapy. *J Infect Dis* **187**, 1534-1543 (2003).

26. Z. Liu, W. G. Cumberland, L. E. Hultin, H. E. Prince, R. Detels *et al.*, Elevated CD38 antigen expression on CD8+ T cells is a stronger marker for the risk of chronic HIV disease progression to AIDS and death in the Multicenter AIDS Cohort Study than CD4+ cell count, soluble immune activation markers, or combinations of HLA-DR and CD38 expression. *Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association* **16**, 83-92 (1997).

27. C. L. Day, D. E. Kaufmann, P. Kiepiela, J. A. Brown, E. S. Moodley *et al.*, PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature* **443**, 350-354 (2006).

28. L. Trautmann, L. Janbazian, N. Chomont, E. A. Said, S. Gimmig *et al.*, Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. *Nature medicine* **12**, 1198-1202 (2006).

29. C. Petrovas, B. Chaon, D. R. Ambrozak, D. A. Price, J. J. Melenhorst *et al.*, Differential association of programmed death-1 and CD57 with ex vivo survival of CD8+ T cells in HIV infection. *J Immunol* **183**, 1120-1132 (2009).

30. D. Zhang, P. Shankar, Z. Xu, B. Harnisch, G. Chen *et al.*, Most antiviral CD8 T cells during chronic viral infection do not express high levels of perforin and are not directly cytotoxic. *Blood* **101**, 226-235 (2003).

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

32. A. Okoye, M. Meier-Schellersheim, J. M. Brenchley, S. I. Hagen, J. M. Walker *et al.*, Progressive CD4+ central memory T cell decline results in CD4+ effector memory insufficiency and overt disease in chronic SIV infection. *The Journal of experimental medicine* **204**, 2171-2185 (2007).

33. A. A. Okoye, L. J. Picker, CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunological reviews* **254**, 54-64 (2013).

34. J. M. Orendi, A. C. Bloem, J. C. Borleffs, F. J. Wijnholds, N. M. de Vos *et al.*, Activation and cell cycle antigens in CD4+ and CD8+ T cells correlate with plasma human immunodeficiency virus (HIV-1) RNA level in HIV-1 infection. *The Journal of infectious diseases* **178**, 1279-1287 (1998).

35. J. Ananworanich, A. Schuetz, C. Vandergeeten, I. Sereti, M. de Souza *et al.*, Impact of multi-targeted antiretroviral treatment on gut T cell depletion and HIV reservoir seeding during acute HIV infection. *PloS one* **7**, e33948 (2012).

36. N. Chomont, M. El-Far, P. Ancuta, L. Trautmann, F. A. Procopio *et al.*, HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature medicine* **15**, 893-900 (2009).

37. D. Finzi, M. Hermankova, T. Pierson, L. M. Carruth, C. Buck *et al.*, Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295-1300 (1997).

38. M. C. Strain, S. J. Little, E. S. Daar, D. V. Havlir, H. F. Gunthard *et al.*, Effect of treatment, during primary infection, on establishment and clearance of cellular reservoirs of HIV-1. *The Journal of infectious diseases* **191**, 1410-1418 (2005).

39. O. T. Ng, O. Laeyendecker, A. D. Redd, S. Munshaw, M. K. Grabowski *et al.*, HIV type 1 polymerase gene polymorphisms are associated with phenotypic differences in replication capacity and disease progression. *The Journal of infectious diseases* **209**, 66-73 (2014).

40. P. N. Amornkul, E. Karita, A. Kamali, W. N. Rida, E. J. Sanders *et al.*, Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. *AIDS* **27**, 2775-2786 (2013).

41. M. Carrington, B. D. Walker, Immunogenetics of spontaneous control of HIV. *Annual review of medicine* **63**, 131-145 (2012).

42. C. Fraser, K. Lythgoe, G. E. Leventhal, G. Shirreff, T. D. Hollingsworth *et al.*, Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science* **343**, 1243727 (2014).

43. J. M. Carlson, M. Schaefer, D. C. Monaco, R. Batorsky, D. T. Claiborne *et al.*, HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science* **345**, 1254031 (2014).

44. M. Paiardini, B. Cervasi, E. Reyes-Aviles, L. Micci, A. M. Ortiz *et al.*, Low levels of SIV infection in sooty mangabey central memory CD(4)(+) T cells are associated with limited CCR5 expression. *Nature medicine* **17**, 830-836 (2011).

45. Q. Li, P. J. Skinner, S. J. Ha, L. Duan, T. L. Mattila *et al.*, Visualizing antigen-specific and infected cells in situ predicts outcomes in early viral infection. *Science* **323**, 1726-1729 (2009).

46. L. Gattinoni, E. Lugli, Y. Ji, Z. Pos, C. M. Paulos *et al.*, A human memory T cell subset with stem cell-like properties. *Nature medicine* **17**, 1290-1297 (2011).

47. E. Lugli, M. H. Dominguez, L. Gattinoni, P. K. Chattopadhyay, D. L. Bolton *et al.*, Superior T memory stem cell persistence supports long-lived T cell memory. *The Journal of clinical investigation* **123**, 594-599 (2013).

48. E. K. Cartwright, C. S. McGary, B. Cervasi, L. Micci, B. Lawson *et al.*, Divergent CD4+ T Memory Stem Cell Dynamics in Pathogenic and Nonpathogenic Simian Immunodeficiency Virus Infections. *J Immunol* **192**, 4666-4673 (2014).

49. N. L. Letvin, J. R. Mascola, Y. Sun, D. A. Gorgone, A. P. Buzby *et al.*, Preserved CD4+ central memory T cells and survival in vaccinated SIV-challenged monkeys. *Science* **312**, 1530-1533 (2006).

50. R. E. Haaland, P. A. Hawkins, J. Salazar-Gonzalez, A. Johnson, A. Tichacek *et al.*, Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS pathogens* **5**, e1000274 (2009).

51. C. M. Rousseau, B. A. Birditt, A. R. McKay, J. N. Stoddard, T. C. Lee *et al.*, Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *Journal of virological methods* **136**, 118-125 (2006).

52. J. F. Salazar-Gonzalez, M. G. Salazar, B. F. Keele, G. H. Learn, E. E. Giorgi *et al.*, Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *The Journal of experimental medicine* **206**, 1273-1289 (2009).

Chapter IV

Protective immunogenetic factors reduce microbial translocation in acute HIV infection

D. Claiborne[1], J. Prince[1], E. Scully[2], K. Nganou-Makamdop[3], J. Tang[4], T. Yu[5], S. Lakhi[6], W. Kilembe[6], D. Douek[3], P. Goepfert[4], M. A. Price[7,8], S. Allen[6,9,10], J. Gilmour[11], M. Altfeld[2,12], E. Hunter[1,9]*

1. Emory Vaccine Center at Yerkes National Primate Research Center, Emory University, Atlanta, Georgia, USA
2. Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA
3. Vaccine Research Center (VRC) at the National Institutes of Health (NIH), Bethesda, Maryland, USA
4. Department of Medicine, University of Alabama at Birmingham (UAB), Birmingham, Alabama, USA
5. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA
6. Zambia-Emory Research Project, Lusaka, Zambia
7. International AIDS Vaccine Initiative, San Francisco, California, USA
8. Department of Epidemiology and Biostatistics, UCSF, SF, CA, USA
9. Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, USA
10. Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

11. International AIDS Vaccine Initiative, London, England

12. Heinrich-Pette-Institut, Hamburg, Germany

Manuscript in preparation

Abstract

In HIV-1 infection, the cellular immune response plays an integral role in restricting early viral replication during the acute phase of viremia in order to establish set point viral load (SPVL)(1-6). Canonical protective HLA class I alleles, such as HLA-B*27 and B*57, exert their protective effects largely by reducing plasma viral loads(7-11). In contrast, HLA-B*81 has been associated with a profound protective benefit in regard to CD4+ T cell decline without dramatically reducing plasma viremia(12). To determine whether additional host immunogenetic factors can impact the immunopathogenesis of disease, we have studied a cohort of HIV-1 subtype C infected seroconvertors for whom the replicative capacity of the virus has been defined(13, 14). We show that several HLA class I, HLA class II, and KIR alleles are capable of delaying CD4+ T cell loss without significantly controlling early SPVL or longitudinal plasma viremia. Protective HLA class I alleles of this nature were associated with a significant reduction in plasma LPS at seroconversion, and reduced markers of gut damage and microbial translocation at 6-months post infection. Moreover, plasma LPS levels at seroconversion significantly predicted CD4+ T cell decline and the time to ART initiation. This data expands upon the mechanisms by which immunogenetic factors can exert their protective effects, in the context of viral replicative capacity, and provides further evidence to suggest that interventions which limit viral replication early after infection or preserve gut epithelial integrity can result in significant positive alterations in HIV-1 disease course.

Introduction

Elucidating the components of a protective immune response is paramount for defining the correlates of protection needed to develop an effective HIV vaccine. The importance of CD8+ cytotoxic T lymphocytes (CTLs) in HIV infection was first suggested by the observation that HIV-specific CD8+ T cells arise concomitantly with the reduction of acute viremia to set point(1, 2) and this was confirmed in the SIV model, where depletion of CD8+ cells prevented viral control(15-17). However, not all CTLs are created equal, and specific HLA class I alleles, which determine the repertoire of viral peptides presented to CD8+ T cells and thus the specificity of the cellular immune response after viral infection, are associated with significantly enhanced viral control(7, 18). In particular, HLA-B*27 and B*57 are consistently associated with low viral loads (VL) (8, 10, 19, 20), and are statistically enriched in individuals who control virus to undetectable levels (7, 8, 20). Moreover, in a series of genome-wide association studies (GWAS), the HLA I locus has been shown to be the primary site of polymorphisms in the human genome significantly affecting HIV-1 disease outcome(18, 21, 22).

Though plasma viremia has been consistently proven to be a strong predictor of HIV-1 disease progression(23-26), the true clinical endpoint of HIV-1 pathogenesis is CD4+ T cell count as measured in the blood, and this has been shown to be more accurately predicted by levels of chronic immune activation(27-30). Moreover, the protective HLA class I allele, B*81, is associated with protection from CD4+ T cell decline without significant control of plasma viremia(12). This suggests that there may be additional mechanisms of HLA-mediated protection distinct from simply controlling plasma viral load. We propose that additional immunogenetic factors with this phenotype

exist, and furthermore, that they may exert their protective effect by altering early events post infection, before the establishment of a viral load set point.

Methods

Study subjects

All participants in the Zambia Emory HIV Research Project (ZEHRP) discordant couples cohort in Lusaka, Zambia were enrolled in human subjects protocols approved by both the University of Zambia Research Ethics Committee and the Emory University Institutional Review Board. Prior to enrollment, individuals received counseling and signed a written informed consent form agreeing to participate.

The subjects included in this study were selected from the ZEHRP cohort based on being recently infected with HIV-1. All subjects were initially seronegative partners within serodiscordant cohabitating heterosexual couples that subsequently seroconverted. All subjects were antiretroviral therapy naïve and were identified a median of 46 (IQR = 33-49) days after the estimated date of infection (EDI). The algorithm used to determine the EDI has been previously described (45). All subjects were infected by HIV-1 subtype C viruses.

Viral loads and CD4+ count measurements

Early set point VL for newly infected individuals was defined as the earliest stable nadir VL value measured between 3 and 9 months post infection and which did not show a significant increase in value within a 3-4 month window. HIV plasma VL was determined at the Emory Center for AIDS Research Virology Core Laboratory using the Amplicor HIV-1 Monitor Test (version 1.5; Roche). CD4+ T cell counts were based on T-cell immunophenotyping, with assays done using the FACScount System (Beckman

Coulter Ltd., London, United Kingdom) in collaboration with the International AIDS Vaccine Initiative.

Generation of Gag-MJ4 chimeras

Gag-MJ4 chimeras were generated from frozen plasma isolated at the seroconversion time point for 127 subjects as previously described (14). Briefly, viral RNA was extracted from 140 ul of plasma using the Qiagen Viral RNA extraction kit (Qiagen). Combined RT-PCR and first round PCR were performed in a single reaction, and *gag* genes were amplified using a nested second round PCR. Patient-derived *gag* genes were joined with the MJ4 long terminal repeat portion via splice-overlap-extension PCR, and *gag*-LTR amplicons were cloned into the MJ4 proviral vector using NgoMIV and HpaI endonuclease restriction enzymes.

HLA and KIR typing

Genomic DNA was extracted from whole blood or buffy coats (QIAamp blood kit; Qiagen). HLA class I genotyping relied on a combination of PCR-based techniques, involving sequence-specific primers (Invitrogen) and sequence-specific oligonucleotide probes (Innogenetics), as described previously(46). Ambiguities were resolved by direct sequencing of three exons in each gene, using kits (Abbott Molecular, Inc.) designed for capillary electrophoresis and the ABI 3130*xl* DNA Analyzer (Applied Biosystems).

Evaluation of plasma cytokines

Plasma levels of cytokines and chemokines were measured using MILLIPLEX Human Cytokine/Chemokine detection kits (Millipore). High sensitivity kits were used for measurement of IFNγ, IL-1β, IL-6 and IL-7 and regular sensitivity kits were used for IFNα2, IL-10, IP-10, MCP-1, MIP-1β and TNFα and were used according to the manufacturer’s instructions. Samples were run in duplicate with all individuals on the same plate and wells with low bead count or coefficient of variance >30% were excluded from subsequent analysis. Plates were read on the Bio-Plex® 3D Suspension Array System (Bio-Rad). Levels of sCD14 (R&D systems), CRP (Millipore), sCD163 (Trillium Diagnostics) and D-dimer (American Diagnostica) were all measured using standard ELISA based assays according to the manufacturer’s instructions. Measurement of LPS levels was performed using the LAL Chromogenic Endotoxin Quantification kit (American Diagnostica).

Intestinal fatty acid binding protein (I-FABP) was measured using a commercially available ELISA DuoSet assay (R&D Systems, Minneapolis, MN, USA) according to the manufacturer's instructions with minor adjustments. Plasma samples were diluted to 10% in diluent from the R&D Systems soluble CD14 ELISA kit (DC140) and plates were blocked with Sigma Blocking Buffer.

Data Processing

Data was pre-processed in order to account for missing data and to increased statistical power before protective alleles were screened for associations with levels of plasma cytokines, chemokines, and markers of microbial translocation and gut integrity. Extreme

positive values for each of the 16 analytes measured in plasma were Winsorized to the 90th percentile. Missing values were imputed using multiple linear regression models, and individuals for whom more than 3 cytokine values were missing were excluded.

Statistical Analyses

All statistical analysis was performed using JMP, version 11 (SAS Institute Inc., Cary, NC). All bivariate continuous correlations were performed using standard linear regression. One-way comparison of means was performed using the Student's t-test, and one-tailed p-values are reported. Kaplan-Meier survival curves and Cox proportional hazards models were performed using an endpoint defined as a single CD4+ T cell count reading less than 300, unless otherwise specified, and statistics reported for survival analyses are generated from the log-rank test. This endpoint for CD4 loss provided the strongest predictive value for progression to CD4 counts <200, while maintaining the largest number of events for statistical power.

Results

In order to identify additional protective immunogenetic factors, we studied a well-characterized Zambian HIV-1subtype C heterosexual acute infection cohort(13, 14). We analyzed a group of 127 acutely infected Zambians for whom detailed clinical, immunogenetic, and virological data were available including (i) longitudinal CD4+ T cell counts and plasma viral loads starting at seroconversion (median of 46 days post estimated date of infection, EDI) and continuing for 3-month intervals thereafter up to 6 years post infection, (ii) 4-digit resolution HLA class I and HLA class II typing, (iii) KIR allele typing, and (iv) viral replicative capacity (vRC) as measured by an *in vitro* replicative fitness assay (14).

We employed the use of Cox proportional hazards models coupled with a stepwise backward variable selection approach in order to identify HLA class I, HLA class II, and KIR alleles associated with significant alterations in CD4+ T cell decline or control of plasma viral load. CD4+ T cell decline was defined as the time to CD4+ T cell counts less than 300, and control of viremia was defined as the time to plasma viral loads > $1x10^5$. Gender of the recipient and vRC of the transmitted virus represent additional host and viral characteristics previously shown to significantly affect CD4+ T cell decline and plasma viremia(14, 25, 31-36), particularly early in HIV infection, and were thus added to all models as covariates.

We found 4 HLA class I alleles (B*1401, B*57,B*5801, and B*81), 2 HLA class II alleles (DQB1*02 and DRB1*15), and 1 KIR allele (KIR 2DS4) to be independent protective factors delaying the loss of CD4+ T cells (**Table 1A**). As has been described previously, vRC was also a significant predictor of CD4+ T cell decline(14). In contrast,

we identified a distinct group of 5 HLA class I alleles (A*03, A*33, B*15, B*44, and B*57) significantly affecting control of plasma viral load. HLA-A*03 and A*33 were deleterious, while HLA-B*15, B*44, and B*57 were protective in terms of viral control, as has been described previously for this and other cohorts(37-40). No HLA class II or KIR alleles were significantly associated with viral control in multivariable models (**Table 1B**) Alleles affecting CD4+ T cell decline or longitudinal viral control were additive in nature and formed distinct profiles even in the absence of gender and vRC covariates (**Supplementary Figure 1**).

Since alleles conferring distinct protective functions appear to be unique (with the exception of B*57, which both robustly controls plasma viremia and protects individuals from rapid CD4+ T cell decline in this cohort), we sought to determine if the protective effects of HLA-B*1401, B*57, B*5801, B*81, DQB1*02, and DRB1*15 were independent of the effects of set point VL. In a multivariable Cox proportional hazards model including set point VL as a covariate, these alleles remained statistically significant predictors of delayed CD4+ T cell decline (**Supplementary Table 1**). This implies that these alleles exert their protective effects apart from control of plasma viral loads and perhaps may influence events very early after infection.

In order to determine how these alleles might reduce CD4 decline independent of VL, we examined retrospectively a subset of these individuals (n=33) for whom early inflammatory cytokine profiles had been evaluated at seroconversion (median of 46 days post EDI), 3-months post seroconversion, and 6-month post seroconversion. We observed a striking association between protective HLA class I (HLA-I) alleles and reduced levels of LPS in plasma at seroconversion (**Figure 1A**). This was not the case for

protective HLA class II or KIR alleles. The early reduction of plasma LPS occurred at a time point in which these protective HLA-I alleles showed no enhanced control of plasma viral load when compared to other non-protective HLA-I alleles (**Figure 1B**). This suggests that these alleles control infection early and at initial sites of viral replication in the tissues, leading to preservation of gut epithelial integrity and reduced translocation of microbial products. While this effect was significant at seroconversion, HLA-I alleles were also associated with reduced levels of LPS, soluble CD14, and intestinal fatty acid-binding protein (I-FABP) at 6-months post seroconversion (**Figure 2A-C**), indicating that this protective effect is durable and directly linked to markers of gut integrity. Furthermore, individuals carrying these protective HLA-I alleles demonstrated a significant reduction in IL-10 levels over time (**Figure 2D**), which is in agreement with published literature linking the presence of circulating microbial products with increased production of IL-10 by monocytes(41, 42).

Due to the relatively small sample size (n=33; 8 with protective HLA-I) represented in this initial observation, we extended this study to evaluate plasma LPS levels at seroconversion in a total of 59 acutely infected individuals, 17 of whom carried HLA-B*1401, B*57, B*5801, or B*81. Plasma LPS levels at seroconversion remained significantly reduced in the 17 individuals with protective HLA-I (**Figure 3A**), again without a significant reduction in plasma viral loads at seroconversion (**Figure 3B**). This reduction in LPS became even more striking when batch effects between the two sample runs were accounted for (**Supplementary Figure 2**). Additionally, the absence of detectable LPS in the plasma at seroconversion significantly predicted CD4+ T cell decline and the time to ARV treatment (**Figures 3C and 3D**). This association was found,

in a generalized linear regression model, to be independent of set point VL (**Supplementary Table 2**), consistent with the observation that the increased hazard of immune activation is generally independent of plasma viral load(43). Though shown to be a strong predictor of disease in chronic HIV-1 infection (44), this is the first report linking levels of circulating LPS during the acute stage of HIV-1 infection with subsequent disease progression.

Since we observed a strong association between protective HLA-I alleles and acute LPS levels, as well as between acute LPS levels and longitudinal CD4+ T cell decline, we hypothesize that early viral replication in the gut and subsequent microbial translocation may represent a fulcrum by which several protective host and viral factors may affect disease progression. In a multivariable generalized linear model predicting the absence or presence of LPS at seroconversion, we found protective HLA-I alleles, female gender, low vRC (as conferred by the transmitted Gag), and KIR 2DS4 to be significant predictors (**Table 2**). Female gender, low vRC, and KIR 2DS4 reduced early LPS levels and are thus protective in nature. This was also true when LPS was treated as a continuous outcome, demonstrating that these protective factors reduce the total levels of circulating LPS (**Supplementary Table 3**). These results demonstrate that both host and viral factors are linked to translocation of microbial products in acute HIV infection, and serve to further highlight the importance of microbial translocation as a potent factor affecting HIV-1 disease course.

Discussion

In this study we have identified 4 HLA class I alleles, 2 HLA class II alleles, and 1 KIR allele that significantly delay CD4+ T cell decline in a cohort of Zambians acutely infected with HIV-1 subtype C. Importantly, with the exception of B*57, these alleles did not substantially reduce plasma viral loads, demonstrating that HLA-mediated protection can occur through additional mechanisms. Although no associations between protective HLA class II alleles and early inflammatory cytokine profiles were observed, a significant reduction in the levels of plasma LPS at seroconversion in individuals carrying these protective HLA class I alleles was. This describes a new mechanism of action regarding HLA-mediated protection and suggests that certain HLA class I alleles may control viral replication early in the tissues, perhaps via a unique mechanism.

Furthermore, these results demonstrate that translocation of microbial products into the blood stream, likely due to HIV-1 induced damage to the gut epithelium, is an event that occurs early after HIV-1 transmission and is not relegated only to the chronic stage of infection. Moreover, we show that LPS levels in acute infection are associated with disease progression and represent a fulcrum by which other host and viral factors exert their protective effects.

The association of LPS levels in acute HIV-1 infection with subsequent disease course provides further support to the concept that the earliest events after HIV-1 transmission can determine the trajectory of disease. Taken together, these results provide further impetus for development of vaccine-mediated therapeutic interventions that limit early viral replication in the tissues, and maintain the integrity of the gut epithelia. Through the prevention of microbial product translocation to the periphery,

and/or inactivation of LPS in the plasma, it may be possible to reduce levels of systemic immune activation in acute HIV infection and significantly alter disease course.

Table 1A

Cox Proportional Hazards Model (Time to CD4 < 300)

	Univariable			Multivariable		
Factors Tested	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
Gender (female)	0.79	0.51 – 1.21	0.29	0.65	0.41 – 1.03	0.07
vRC (lowest tercile)	0.42	0.25 – 0.69	0.0004	0.38	0.21 – 0.63	0.0001
B*1401	0.18	0.01 – 0.83	0.02	0.18	0.01 – 0.83	0.02
B*57/5801	0.43	0.22 – 0.77	0.003	0.39	0.20 – 0.70	0.001
B*81	<.01	0.19 – 0.19	<.0001	<.01	0.00 – 0.14	<.0001
DQB1*02	0.69	0.43 – 1.08	0.10	0.41	0.24 – 0.67	0.0004
DRB1*15	0.74	0.46 – 1.17	0.20	0.47	0.27 – 0.78	0.003
KIR2DS4	0.43	0.20 – 1.12	0.08	0.30	0.13 – 0.79	0.02

Table 1B

Cox Proportional Hazards Model (Time to VL > $1x10^5$)

	Univariable			Multivariable		
Factors Tested	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
Gender (female)	0.52	0.32 – 0.82	0.006	0.43	0.26 – 0.69	0.0005
vRC (lowest tercile)	0.54	0.31 – 0.91	0.02	0.52	0.28 – 0.90	0.02
A*03	2.32	0.97 – 4.74	0.06	3.67	1.38 – 8.83	0.01
A*33	3.46	1.65 – 6.52	0.002	2.92	1.33 – 5.87	0.009
B*15	0.63	0.38 – 1.03	0.07	0.45	0.24 – 0.80	0.006
B*44	0.98	0.50 – 1.76	0.96	0.49	0.24 – 0.93	0.03
B*57	0.08	0.00 – 0.38	0.0001	0.05	0.00 – 0.24	<.0001

Table 1. Immunogenetic factors significantly affecting CD4+ T cell decline or longitudinal control of plasma viral load in HIV-1 subtype C infection. We used Cox proportional hazards models to define HLA and KIR alleles associated with longitudinal preservation of CD4+ T cell counts or longitudinal control of plasma viral load in a cohort of 127 Zambian volunteers acutely infected with HIV-1 subtype C. Endpoints for

Cox proportional hazards models were defined as the time to CD4+ T cell counts <300 (**A**) or the time to plasma viral loads > $1x10^5$ (**B**). Step-wise backward analysis was used to create multivariable models, and gender and vRC of the transmitted Gag sequence were included as covariates. The left side includes statistics for factors analyzed in singularity (univariable). The right side includes statistics for the full multivariable model, including all factors tested.

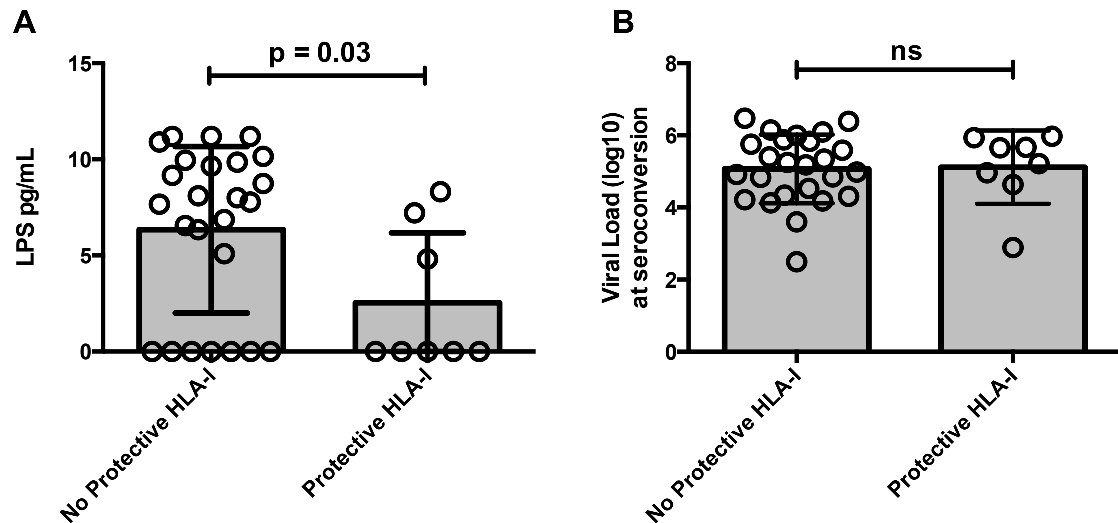


Figure 1. Protective HLA-I alleles are associated with reduced plasma lipopolysaccharide (LPS) levels at seroconversion. B*1401, B*57, B*5801, and B*81 are HLA class I (HLA-I) alleles defined as protective in this cohort (n=127) based on significant associations with reduced time to CD4+ T cell counts <300 loss in the first 5 years of infection as depicted in Table 1A. In a subset of 33 individual, levels of 16 inflammatory cytokines, chemokines, and markers of gut integrity and microbial translocation (see Methods) were evaluated in the plasma at seroconversion (median of 46 days post EDI). (**A**) The difference in plasma LPS levels at seroconversion between individuals with protective HLA-I alleles (n=8) and those without protective HLA-I alleles (n=25). (**B**) The difference in $\log_{10}$ viral RNA copies/mL of plasma at seroconversion between individuals with protective HLA-I alleles (n=8) and those without protective HLA-I alleles (n=25). Statistical comparisons were made using the Student's t-test and reported p-values are two-tailed. Error bars represent the standard deviation. The heights of bar graphs represent the mean. LPS, lipopolysaccharide.

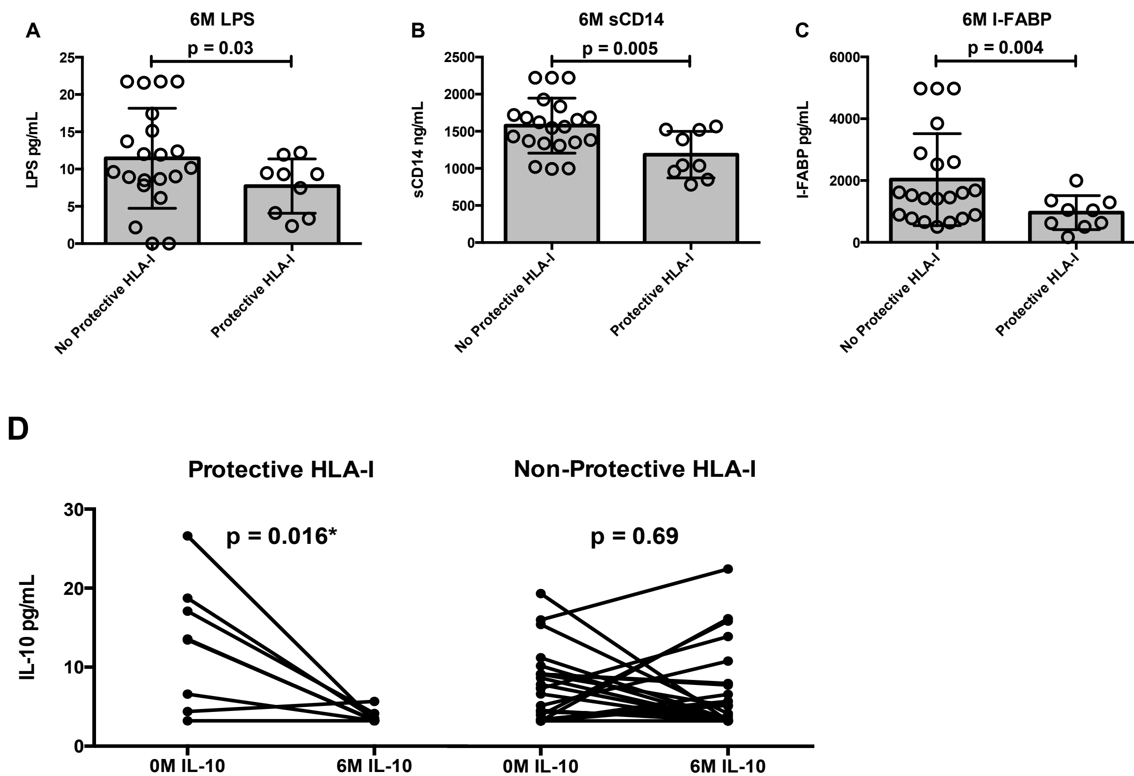


Figure 2. Protective HLA-I alleles are associated with reduced plasma LPS, sCD14, I-FABP, and IL-10 at 6-months post seroconversion. Protective HLA-I alleles (B*1401, B*57, B*5801, and B*81) are defined in Table 1A. In a subset of 29 individual, levels of 16 inflammatory cytokines, chemokines, and markers of gut integrity and microbial translocation (see Methods) were evaluated in the plasma at both seroconversion (median of 46 days post EDI) and 6-months post seroconversion. (**A-C**) The difference in LPS, sCD14, and I-FABP levels in the plasma at 6-months after seroconversion between individuals with protective HLA-I alleles (n=9) and those without protective HLA-I alleles (n=21). Individuals with protective HLA-I alleles exhibit significantly lower levels of LPS, sCD14, and I-FABP at 6-months post seroconversion. Statistical comparisons were made using the Student's t-test and p-values reported are one-tailed. Error bars

represent the standard deviation. The heights of bar graphs represent the mean. (**D**) The reduction in IL-10 levels measured in the plasma at seroconversion and 6-months after seroconversion in those with protective HLA-I (left, n=7) and individuals without protective HLA-I (right, n=21) Statistical comparisons were made using a paired t-test, and p-values reported are two-tailed. I-FABP, intestinal fatty acid-binding protein.

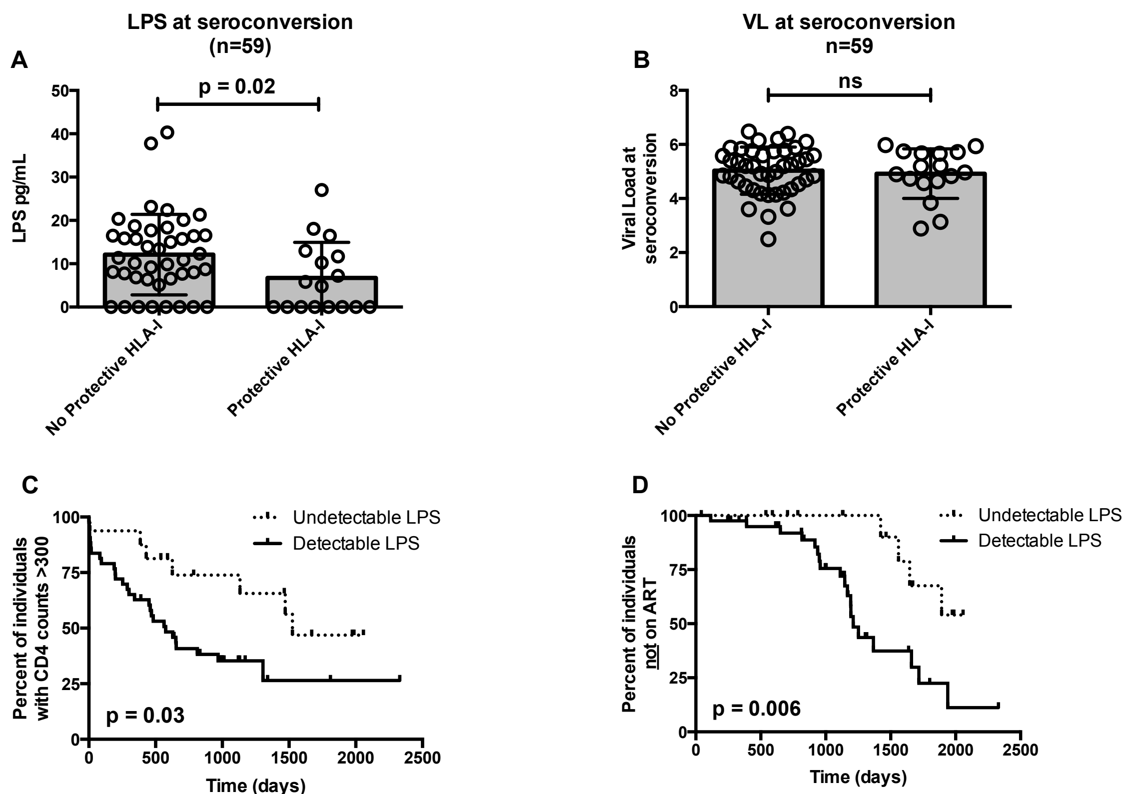


Figure 3. Circulating plasma LPS at seroconversion predicts CD4+ T cell decline and time to initiation of ARV treatment. Initial evaluation of inflammatory cytokines at seroconversion in 33 acutely infected individuals was expanded to include an additional 26 seroconvertors (n=59). (**A**) The difference in plasma LPS levels between those with protective HLA-I (n=17) and those without protective HLA-I (n=42) in a larger cohort of individuals (n=59). (**B**) The number of log10 viral RNA copies/mL in the plasma at seroconversion between those with and without protective HLA-I. Stastistical comparisons were performed using the Student's t-test, and reported p-values are one-tailed. Error bars represent the standard deviation. The heights of bar graphs represent the mean. (**C,D**) Kaplan-Meier survival models assessing the effect of detectable LPS in the plasma at seroconversion on the time to CD4+ T cell counts <300 (**C**) or the time to

ART initiation (**D**). Statistics for Kaplan-Meier survival analyses were generated using the Log-rank test. ART, anti-retroviral therapy.

Generalized Linear Model
(Undetectable LPS as outcome)

	Univariable			Multivariable		
Factors Tested	**β**	**Conf. Limit**	**p-value**	**β**	**Conf. Limit**	**p-value**
Gender (female)	-0.56	-1.21 – 0.03	0.06	-1.02	-1.98 – -0.25	0.004
vRC (lowest tercile)	-0.48	-1.10 – 0.09	0.09	-0.91	-1.78 – -0.18	0.009
Protective HLA-I	-0.64	-1.26 – -0.05	0.03	-0.94	-1.81 – -0.21	0.006
KIR 2DS4	-0.66	-3.12 – 0.50	0.18	-1.22	-2.43 – -0.44	0.056
Group 1 vs Group 2	-0.39	-1.01 – 0.18	0.17	-1.30	-3.84 – 0.25	0.0007

Table 2. Prevention of early microbial translocation is a common mechanism among protective host and viral characteristics. Generalized linear models were used to predict the absence of LPS (binomial distribution) in the plasma at seroconversion. The left panel includes statistics for univariable analyses, and protective HLA class I was the only significant predictive variable in singularity. The right hand panel includes statistics for a multivariable model including all factors tested. Female gender, low vRC, protective HLA-I, and KIR 2DS4 were all found to be associated with undetectable LPS levels in the plasma at seroconversion, and are thus considered protective. Because LPS levels for the second set of 26 individuals were universally higher, a group 1 vs 2 comparison was added as a normalization variable in order to account for batch effects between the two groups of tests for plasma LPS levels. Group 1 (n=33); Group 2 (n=26).

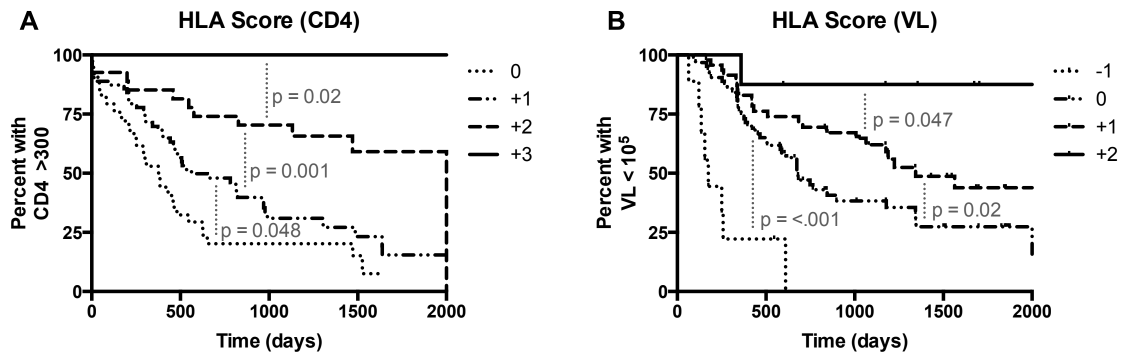


Figure S1. Protective and deleterious HLA alleles are additive in nature represent distinct pathogenesis profiles. In two distinct multivariable Cox proportional hazards models, 6 HLA alleles (B*1401, B*57, B*5801, B*81, DQB1*02, and DRB1*15) were found to be significantly correlated with the time to CD4+ T cell counts <300, and 4 HLA alleles (A*03, A*33, B*15, B*44, and B*57) were significantly associated with the time to plasma viral loads > $1x10^5$. For each group of alleles evaluating either CD4+ T cell decline (**A**) or time to ART initiation (**B**), HLA "scores" were made for each individual (n=127) by subtracting the number of deleterious HLA alleles from the number of favorable HLA alleles. These HLA scores describe distinct profiles, and demonstrate that the effects of these alleles are additive in nature. Statistics were generated using Cox proportional hazards models with multilevel variables. Vertical dotted lines represent statistical comparisons.

Cox Proportional Hazards Model
(Time to CD4 < 300)

	Multivariable		
Factors Tested	**HR**	**95% CI**	**p-value**
Gender (female)	0.83	0.51 – 1.34	0.45
vRC (lowest tercile)	0.48	0.27 – 0.82	0.007
Set point VL	6.35	1.64 – 26.1	0.007
B*1401	0.19	0.01 – 0.87	0.029
B*57/5801	0.44	0.22 – 0.81	0.008
B*81	<0.01	0.00 – 0.16	<.0001
DQB1*02	0.45	0.27 – 0.74	0.001
DRB1*15	0.45	0.26 – 0.75	0.002
KIR 2DS4	0.37	0.16 – 1.00	0.051

Table S1. Protective HLA alleles predict CD4+ T cell decline in manner distinct from set point VL. The $\log_{10}$ set point viral load of each individual (n=127) was added to a multivariable Cox proportional hazard model evaluating the effects of host and viral characteristics on CD4+ T cell decline. The endpoint was defined as CD4+ T cell counts <300. Set point viral load was defined as the earliest nadir viral load reading between 3 and 9 months post estimated date of infection, where immediately subsequent viral load measurements remained relatively stable.

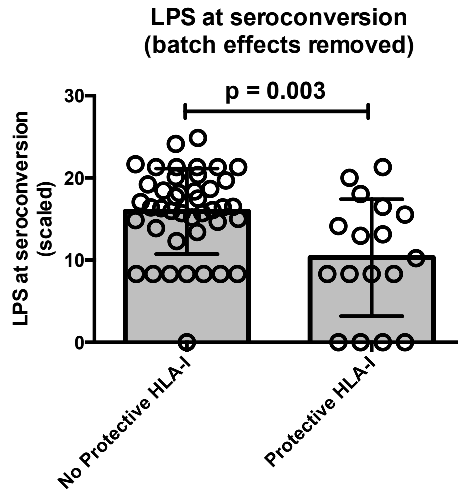


Figure S2. Accounting for batch effects further improves the association between protective HLA-I alleles and low levels of LPS at seroconversion. Levels of 16 inflammatory markers in the plasma at seroconversion were measured in two groups in separate experiments (group 1 n = 33; group 2 n = 26). Due to the significant increase in global signal for analytes measured in the second group, the "pamr" R package was used to adjust for batch effects. Data sets were scaled using a one-way ANOVA adjustment using all 16 analytes.

A

Cox Proportional Hazards Model
(Time to CD4 < 300)

	Multivariable		
Factors Tested	**HR**	**95% CI**	**p-value**
Undetectable LPS	0.39	0.15 – 0.88	0.02
Set Point VL	9.66	2.01 – 50.4	0.004

B

Cox Proportional Hazards Model
(Time to ARV treatment)

	Multivariable		
Factors Tested	**HR**	**95% CI**	**p-value**
Undetectable LPS	0.22	0.06 – 0.60	0.002
Set Point VL	5.77	0.87 – 40.6	0.07

Table S2. Detectable LPS in the plasma at seroconversion drives CD4+ T decline via a mechanism distinct from set point VL. The set point viral load of each individual (n=59) was added as a covariate to a multivariable Cox proportional hazards model evaluating the effect of undetectable LPS in the plasma at seroconversion on the time to CD4+ T cell counts <300 (**A**) or the time to initiation of ART (**B**).

	Generalized Linear Model (LPS levels at seroconversion as outcome)					
	Univariable			Multivariable		
Factors Tested	**β**	**Conf. Limit**	**p-value**	**β**	**Conf. Limit**	**p-value**
Gender (female)	-0.37	-2.31 – 1.58	0.71	-1.87	-3.29 – -0.44	0.01
vRC (lowest tercile)	0.10	-1.85 – 2.04	0.92	-1.41	-2.80 – -0.02	0.047
Protective HLA-I	-2.40	-4.45 – -0.34	0.02	-2.72	-4.21 – -1.24	0.0005
KIR 2DS4	-2.75	-6.55 – 1.05	0.15	-2.90	-5.57 – -0.23	0.03
Group 1 vs Group 2	-4.16	-5.78 – -2.53	<.0001	-5.36	-6.83 – 3.89	<.0001

Table S3. In a multivariable generalized linear model, total LPS levels are significantly reduced by protective host and viral factors. Univariable (left) and multivariable (right) generalized linear models were used to illustrate the effects of gender, vRC, and protective immunogenetic factors on plasma LPS levels at seroconversion. A "Group 1 vs Group 2" normalization variable was included in order to account for batch effects. Group 2 (n=26) have universally higher LPS levels at seroconversion, but accounting for these batch effects revealed the significant contributions of female gender, vRC, protective HLA-I, and KIR 2DS4 in reducing plasma LPS levels at seroconversion.

References

1. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, Farthing C, Ho DD. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. Journal of virology. 1994;68(7):4650-5. PubMed PMID: 8207839; PubMed Central PMCID: PMC236393.

2. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. Journal of virology. 1994;68(9):6103-10. PubMed PMID: 8057491; PubMed Central PMCID: PMC237022.

3. Walker CM, Moody DJ, Stites DP, Levy JA. CD8+ lymphocytes can control HIV infection in vitro by suppressing virus replication. Science. 1986;234(4783):1563-6. PubMed PMID: 2431484.

4. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, Elvin JG, Rothbard JA, Bangham CR, Rizza CR, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature. 1991;354(6353):453-9. doi: 10.1038/354453a0. PubMed PMID: 1721107.

5. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, Keele BF, Learn GH, Turnbull EL, Salazar MG, Weinhold KJ, Moore S, B CCC, Letvin N, Haynes BF, Cohen MS, Hraber P, Bhattacharya T, Borrow P, Perelson AS, Hahn BH,

Shaw GM, Korber BT, McMichael AJ. 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. 2009;206(6):1253-72. doi: 10.1084/jem.20090365. PubMed PMID: 19487423; PubMed Central PMCID: PMC2715063.

6. Walker B, McMichael A. The T-cell response to HIV. Cold Spring Harbor perspectives in medicine. 2012;2(11). doi: 10.1101/cshperspect.a007054. PubMed PMID: 23002014.

7. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, Detels R, Blattner W, Phair J, Erlich H, Mann DL. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nature medicine. 1996;2(4):405-11. PubMed PMID: 8597949.

8. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(6):2709-14. doi: 10.1073/pnas.050567397. PubMed PMID: 10694578; PubMed Central PMCID: PMC15994.

9. Altfeld M, Addo MM, Rosenberg ES, Hecht FM, Lee PK, Vogel M, Yu XG, Draenert R, Johnston MN, Strick D, Allen TM, Feeney ME, Kahn JO, Sekaly RP, Levy JA, Rockstroh JK, Goulder PJ, Walker BD. Influence of HLA-B57 on clinical presentation

and viral control during acute HIV-1 infection. Aids. 2003;17(18):2581-91. doi: 10.1097/01.aids.0000096870.36052.b6. PubMed PMID: 14685052.

10. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, Zimbwa P, Moore S, Allen T, Brander C, Addo MM, Altfeld M, James I, Mallal S, Bunce M, Barber LD, Szinger J, Day C, Klenerman P, Mullins J, Korber B, Coovadia HM, Walker BD, Goulder PJ. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature. 2004;432(7018):769-75. doi: 10.1038/nature03113. PubMed PMID: 15592417.

11. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, Altfeld M, O'Brien SJ, Carrington M. AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. Nature medicine. 2005;11(12):1290-2. doi: 10.1038/nm1333. PubMed PMID: 16288280.

12. Prentice HA, Porter TR, Price MA, Cormier E, He D, Farmer PK, Kamali A, Karita E, Lakhi S, Sanders EJ, Anzala O, Amornkul PN, Allen S, Hunter E, Kaslow RA, Gilmour J, Tang J, Network IAHR. HLA-B*57 versus HLA-B*81 in HIV-1 infection: slow and steady wins the race? Journal of virology. 2013;87(7):4043-51. doi: 10.1128/JVI.03302-12. PubMed PMID: 23365442; PubMed Central PMCID: PMC3624227.

13. Amornkul PN, Karita E, Kamali A, Rida WN, Sanders EJ, Lakhi S, Price MA, Kilembe W, Cormier E, Anzala O, Latka MH, Bekker LG, Allen SA, Gilmour J, Fast PE, Partnership IAHP. Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. Aids. 2013;27(17):2775-86. doi: 10.1097/QAD.0000000000000012. PubMed PMID: 24113395; PubMed Central PMCID: PMC3815107.

14. Prince JL, Claiborne DT, Carlson JM, Schaefer M, Yu T, Lahki S, Prentice HA, Yue L, Vishwanathan SA, Kilembe W, Goepfert P, Price MA, Gilmour J, Mulenga J, Farmer P, Derdeyn CA, Tang J, Heckerman D, Kaslow RA, Allen SA, Hunter E. Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. PLoS pathogens. 2012;8(11):e1003041. doi: 10.1371/journal.ppat.1003041. PubMed PMID: 23209412; PubMed Central PMCID: PMC3510241.

15. Matano T, Shibata R, Siemon C, Connors M, Lane HC, Martin MA. 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. 1998;72(1):164-9. PubMed PMID: 9420212; PubMed Central PMCID: PMC109361.

16. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallon BJ, Ghrayeb J, Forman MA, Montefiori DC, Rieber EP, Letvin NL, Reimann KA. Control of viremia in simian immunodeficiency

virus infection by CD8+ lymphocytes. Science. 1999;283(5403):857-60. PubMed PMID: 9933172.

17. Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J, Irwin CE, Safrit JT, Mittler J, Weinberger L, Kostrikis LG, Zhang L, Perelson AS, Ho DD. Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. The Journal of experimental medicine. 1999;189(6):991-8. PubMed PMID: 10075982; PubMed Central PMCID: PMC2193038.

18. Martin MP, Carrington M. Immunogenetics of HIV disease. Immunological reviews. 2013;254(1):245-64. doi: 10.1111/imr.12071. PubMed PMID: 23772624; PubMed Central PMCID: PMC3703621.

19. Magierowska M, Theodorou I, Debre P, Sanson F, Autran B, Riviere Y, Charron D, Costagliola D. Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. Blood. 1999;93(3):936-41. PubMed PMID: 9920843.

20. Pereyra F, Addo MM, Kaufmann DE, Liu Y, Miura T, Rathod A, Baker B, Trocha A, Rosenberg R, Mackey E, Ueda P, Lu Z, Cohen D, Wrin T, Petropoulos CJ, Rosenberg ES, Walker BD. Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. The Journal of infectious diseases. 2008;197(4):563-71. doi: 10.1086/526786. PubMed PMID: 18275276.

21. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB. A whole-genome association study of major determinants for host control of HIV-1. Science. 2007;317(5840):944-7. doi: 10.1126/science.1143767. PubMed PMID: 17641165; PubMed Central PMCID: PMC1991296.

22. International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Ripke S, Brumme CJ, Pulit SL, Carrington M, Kadie CM, Carlson JM, Heckerman D, Graham RR, Plenge RM, Deeks SG, Gianniny L, Crawford G, Sullivan J, Gonzalez E, Davies L, Camargo A, Moore JM, Beattie N, Gupta S, Crenshaw A, Burtt NP, Guiducci C, Gupta N, Gao X, Qi Y, Yuki Y, Piechocka-Trocha A, Cutrell E, Rosenberg R, Moss KL, Lemay P, O'Leary J, Schaefer T, Verma P, Toth I, Block B, Baker B, Rothchild A, Lian J, Proudfoot J, Alvino DM, Vine S, Addo MM, Allen TM, Altfeld M, Henn MR, Le Gall S, Streeck H, Haas DW, Kuritzkes DR, Robbins GK, Shafer RW, Gulick RM, Shikuma CM, Haubrich R, Riddler S, Sax PE, Daar ES, Ribaudo HJ, Agan B, Agarwal S, Ahern RL, Allen BL, Altidor S, Altschuler EL, Ambardar S, Anastos K, Anderson B, Anderson V, Andrady U, Antoniskis D, Bangsberg D, Barbaro D, Barrie W, Bartczak J, Barton S, Basden P, Basgoz N, Bazner S, Bellos NC, Benson AM, Berger J, Bernard NF, Bernard AM, Birch C, Bodner SJ, Bolan RK, Boudreaux ET, Bradley M, Braun JF, Brndjar JE, Brown SJ, Brown K, Brown ST, Burack J, Bush LM, Cafaro V, Campbell O, Campbell

J, Carlson RH, Carmichael JK, Casey KK, Cavacuiti C, Celestin G, Chambers ST, Chez N, Chirch LM, Cimoch PJ, Cohen D, Cohn LE, Conway B, Cooper DA, Cornelson B, Cox DT, Cristofano MV, Cuchural G, Jr., Czartoski JL, Dahman JM, Daly JS, Davis BT, Davis K, Davod SM, DeJesus E, Dietz CA, Dunham E, Dunn ME, Ellerin TB, Eron JJ, Fangman JJ, Farel CE, Ferlazzo H, Fidler S, Fleenor-Ford A, Frankel R, Freedberg KA, French NK, Fuchs JD, Fuller JD, Gaberman J, Gallant JE, Gandhi RT, Garcia E, Garmon D, Gathe JC, Jr., Gaultier CR, Gebre W, Gilman FD, Gilson I, Goepfert PA, Gottlieb MS, Goulston C, Groger RK, Gurley TD, Haber S, Hardwicke R, Hardy WD, Harrigan PR, Hawkins TN, Heath S, Hecht FM, Henry WK, Hladek M, Hoffman RP, Horton JM, Hsu RK, Huhn GD, Hunt P, Hupert MJ, Illeman ML, Jaeger H, Jellinger RM, John M, Johnson JA, Johnson KL, Johnson H, Johnson K, Joly J, Jordan WC, Kauffman CA, Khanlou H, Killian RK, Kim AY, Kim DD, Kinder CA, Kirchner JT, Kogelman L, Kojic EM, Korthuis PT, Kurisu W, Kwon DS, LaMar M, Lampiris H, Lanzafame M, Lederman MM, Lee DM, Lee JM, Lee MJ, Lee ET, Lemoine J, Levy JA, Llibre JM, Liguori MA, Little SJ, Liu AY, Lopez AJ, Loutfy MR, Loy D, Mohammed DY, Man A, Mansour MK, Marconi VC, Markowitz M, Marques R, Martin JN, Martin HL, Jr., Mayer KH, McElrath MJ, McGhee TA, McGovern BH, McGowan K, McIntyre D, McLeod GX, Menezes P, Mesa G, Metroka CE, Meyer-Olson D, Miller AO, Montgomery K, Mounzer KC, Nagami EH, Nagin I, Nahass RG, Nelson MO, Nielsen C, Norene DL, O'Connor DH, Ojikutu BO, Okulicz J, Oladehin OO, Oldfield EC, 3rd, Olender SA, Ostrowski M, Owen WF, Jr., Pae E, Parsonnet J, Pavlatos AM, Perlmutter AM, Pierce MN, Pincus JM, Pisani L, Price LJ, Proia L, Prokesch RC, Pujet HC, Ramgopal M, Rathod A, Rausch M, Ravishankar J, Rhame FS, Richards CS, Richman DD, Rodes B, Rodriguez M, Rose RC, 3rd, Rosenberg ES, Rosenthal D, Ross PE,

Rubin DS, Rumbaugh E, Saenz L, Salvaggio MR, Sanchez WC, Sanjana VM, Santiago S, Schmidt W, Schuitemaker H, Sestak PM, Shalit P, Shay W, Shirvani VN, Silebi VI, Sizemore JM, Jr., Skolnik PR, Sokol-Anderson M, Sosman JM, Stabile P, Stapleton JT, Starrett S, Stein F, Stellbrink HJ, Sterman FL, Stone VE, Stone DR, Tambussi G, Taplitz RA, Tedaldi EM, Telenti A, Theisen W, Torres R, Tosiello L, Tremblay C, Tribble MA, Trinh PD, Tsao A, Ueda P, Vaccaro A, Valadas E, Vanig TJ, Vecino I, Vega VM, Veikley W, Wade BH, Walworth C, Wanidworanun C, Ward DJ, Warner DA, Weber RD, Webster D, Weis S, Wheeler DA, White DJ, Wilkins E, Winston A, Wlodaver CG, van't Wout A, Wright DP, Yang OO, Yurdin DL, Zabukovic BW, Zachary KC, Zeeman B, Zhao M. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010;330(6010):1551-7. doi: 10.1126/science.1195271. PubMed PMID: 21051598; PubMed Central PMCID: PMC3235490.

23. Semple M, Loveday C, Weller I, Tedder R. Direct measurement of viraemia in patients infected with HIV-1 and its relationship to disease progression and zidovudine therapy. Journal of medical virology. 1991;35(1):38-45. PubMed PMID: 1940882.

24. Mellors JW, Kingsley LA, Rinaldo CR, Jr., Todd JA, Hoo BS, Kokka RP, Gupta P. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Annals of internal medicine. 1995;122(8):573-9. PubMed PMID: 7887550.

25. Katzenstein DA, Hammer SM, Hughes MD, Gundacker H, Jackson JB, Fiscus S, Rasheed S, Elbeik T, Reichman R, Japour A, Merigan TC, Hirsch MS. The relation of

virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. AIDS Clinical Trials Group Study 175 Virology Study Team. The New England journal of medicine. 1996;335(15):1091-8. Epub 1996/10/10. doi: 10.1056/NEJM199610103351502. PubMed PMID: 8813039.

26. O'Brien WA, Hartigan PM, Martin D, Esinhart J, Hill A, Benoit S, Rubin M, Simberkoff MS, Hamilton JD. Changes in plasma HIV-1 RNA and CD4+ lymphocyte counts and the risk of progression to AIDS. Veterans Affairs Cooperative Study Group on AIDS. The New England journal of medicine. 1996;334(7):426-31. doi: 10.1056/NEJM199602153340703. PubMed PMID: 8552144.

27. Bofill M, Mocroft A, Lipman M, Medina E, Borthwick NJ, Sabin CA, Timms A, Winter M, Baptista L, Johnson MA, Lee CA, Phillips AN, Janossy G. Increased numbers of primed activated CD8+CD38+CD45RO+ T cells predict the decline of CD4+ T cells in HIV-1-infected patients. Aids. 1996;10(8):827-34. PubMed PMID: 8828739.

28. Gougeon ML, Lecoeur H, Dulioust A, Enouf MG, Crouvoiser M, Goujard C, Debord T, Montagnier L. Programmed cell death in peripheral lymphocytes from HIV-infected persons: increased susceptibility to apoptosis of CD4 and CD8 T cells correlates with lymphocyte activation and with disease progression. Journal of immunology. 1996;156(9):3509-20. PubMed PMID: 8617980.

29. Giorgi JV, Hultin LE, McKeating JA, Johnson TD, Owens B, Jacobson LP, Shih R, Lewis J, Wiley DJ, Phair JP, Wolinsky SM, Detels R. Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. The Journal of infectious diseases. 1999;179(4):859-70. doi: 10.1086/314660. PubMed PMID: 10068581.

30. McCune JM. The dynamics of CD4+ T-cell depletion in HIV disease. Nature. 2001;410(6831):974-9. doi: 10.1038/35073648. PubMed PMID: 11309627.

31. Farzadegan H, Hoover DR, Astemborski J, Lyles CM, Margolick JB, Markham RB, Quinn TC, Vlahov D. Sex differences in HIV-1 viral load and progression to AIDS. Lancet. 1998;352(9139):1510-4. Epub 1998/11/20. doi: 10.1016/S0140-6736(98)02372-1. PubMed PMID: 9820299.

32. Sterling TR, Lyles CM, Vlahov D, Astemborski J, Margolick JB, Quinn TC. Sex differences in longitudinal human immunodeficiency virus type 1 RNA levels among seroconverters. The Journal of infectious diseases. 1999;180(3):666-72. Epub 1999/08/07. doi: 10.1086/314967. PubMed PMID: 10438353.

33. Sterling TR, Vlahov D, Astemborski J, Hoover DR, Margolick JB, Quinn TC. Initial plasma HIV-1 RNA levels and progression to AIDS in women and men. The New

England journal of medicine. 2001;344(10):720-5. Epub 2001/03/10. doi: 10.1056/NEJM200103083441003. PubMed PMID: 11236775.

34. Gandhi M, Bacchetti P, Miotti P, Quinn TC, Veronese F, Greenblatt RM. Does patient sex affect human immunodeficiency virus levels? Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2002;35(3):313-22. Epub 2002/07/13. doi: 10.1086/341249. PubMed PMID: 12115098.

35. Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, Brumme CJ, Wang B, Losina E, Miura T, Chonco F, van der Stok M, Mncube Z, Bishop K, Goulder PJ, Walker BD, Brockman MA, Ndung'u T. Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. Journal of virology. 2010;84(20):10820-31. doi: 10.1128/JVI.01084-10. PubMed PMID: 20702636; PubMed Central PMCID: PMC2950592.

36. Wright JK, Novitsky V, Brockman MA, Brumme ZL, Brumme CJ, Carlson JM, Heckerman D, Wang B, Losina E, Leshwedi M, van der Stok M, Maphumulo L, Mkhwanazi N, Chonco F, Goulder PJ, Essex M, Walker BD, Ndung'u T. Influence of Gag-protease-mediated replication capacity on disease progression in individuals recently infected with HIV-1 subtype C. Journal of virology. 2011;85(8):3996-4006. doi: 10.1128/JVI.02520-10. PubMed PMID: 21289112; PubMed Central PMCID: PMC3126116.

37. Lazaryan A, Song W, Lobashevsky E, Tang J, Shrestha S, Zhang K, McNicholl JM, Gardner LI, Wilson CM, Klein RS, Rompalo A, Mayer K, Sobel J, Kaslow RA, Group HIVERS, Reaching for Excellence in Adolescent C, Health Study G. The influence of human leukocyte antigen class I alleles and their population frequencies on human immunodeficiency virus type 1 control among African Americans. Human immunology. 2011;72(4):312-8. doi: 10.1016/j.humimm.2011.01.003. PubMed PMID: 21262311; PubMed Central PMCID: PMC3778654.

38. Tang J, Cormier E, Gilmour J, Price MA, Prentice HA, Song W, Kamali A, Karita E, Lakhi S, Sanders EJ, Anzala O, Amornkul PN, Allen S, Hunter E, Kaslow RA, Network IAHR. Human leukocyte antigen variants B*44 and B*57 are consistently favorable during two distinct phases of primary HIV-1 infection in sub-Saharan Africans with several viral subtypes. Journal of virology. 2011;85(17):8894-902. doi: 10.1128/JVI.00439-11. PubMed PMID: 21715491; PubMed Central PMCID: PMC3165830.

39. Zhang X, Huang X, Xia W, Li W, Zhang T, Wu H, Xu X, Yan H. HLA-B*44 is associated with a lower viral set point and slow CD4 decline in a cohort of Chinese homosexual men acutely infected with HIV-1. Clinical and vaccine immunology : CVI. 2013;20(7):1048-54. doi: 10.1128/CVI.00015-13. PubMed PMID: 23677320; PubMed Central PMCID: PMC3697455.

40. Li X, Price MA, He D, Kamali A, Karita E, Lakhi S, Sanders EJ, Anzala O, Amornkul PN, Allen S, Hunter E, Kaslow RA, Gilmour J, Tang J, Partnership IAHP. Host

genetics and viral load in primary HIV-1 infection: clear evidence for gene by sex interactions. Human genetics. 2014. doi: 10.1007/s00439-014-1465-x. PubMed PMID: 24969460.

41. Brockman MA, Kwon DS, Tighe DP, Pavlik DF, Rosato PC, Sela J, Porichis F, Le Gall S, Waring MT, Moss K, Jessen H, Pereyra F, Kavanagh DG, Walker BD, Kaufmann DE. IL-10 is up-regulated in multiple cell types during viremic HIV infection and reversibly inhibits virus-specific T cells. Blood. 2009;114(2):346-56. doi: 10.1182/blood-2008-12-191296. PubMed PMID: 19365081; PubMed Central PMCID: PMC2714209.

42. Said EA, Dupuy FP, Trautmann L, Zhang Y, Shi Y, El-Far M, Hill BJ, Noto A, Ancuta P, Peretz Y, Fonseca SG, Van Grevenynghe J, Boulassel MR, Bruneau J, Shoukry NH, Routy JP, Douek DC, Haddad EK, Sekaly RP. Programmed death-1-induced interleukin-10 production by monocytes impairs CD4+ T cell activation during HIV infection. Nature medicine. 2010;16(4):452-9. doi: 10.1038/nm.2106. PubMed PMID: 20208540.

43. Deeks SG, Kitchen CM, Liu L, Guo H, Gascon R, Narvaez AB, Hunt P, Martin JN, Kahn JO, Levy J, McGrath MS, Hecht FM. Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. Blood. 2004;104(4):942-7. doi: 10.1182/blood-2003-09-3333. PubMed PMID: 15117761.

44. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, Kazzaz Z, Bornstein E, Lambotte O, Altmann D, Blazar BR, Rodriguez B, Teixeira-Johnson L, Landay A, Martin JN, Hecht FM, Picker LJ, Lederman MM, Deeks SG, Douek DC.

Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nature medicine. 2006;12(12):1365-71. doi: 10.1038/nm1511. PubMed PMID: 17115046.

45. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, Manigart O, Mulenga J, Keele BF, Shaw GM, Hahn BH, Allen SA, Derdeyn CA, Hunter E. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS pathogens. 2009;5(1):e1000274. Epub 2009/01/24. doi: 10.1371/journal.ppat.1000274. PubMed PMID: 19165325; PubMed Central PMCID: PMC2621345.

46. Tang J, Tang S, Lobashevsky E, Myracle AD, Fideli U, Aldrovandi G, Allen S, Musonda R, Kaslow RA, Zambia UABHIVRP. Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. Journal of virology. 2002;76(16):8276-84. PubMed PMID: 12134033; PubMed Central PMCID: PMC155130.

Chapter V: Discussion

The impact and implications of heritable viral characteristics on HIV-1 pathogenesis

As has been discussed in Chapter I, HIV-1 pathogenesis contains a heritable component, as evidenced by the fact that viral loads are correlated between epidemiologically linked transmission pairs[193-195]. These observations implicate transmitted viral characteristics, dictated by the viral genotype, as the basis for this heritability. However, no study to date had definitively described the nature of the viral characteristic(s) responsible. In Chapter II, we show a positive correlation between viral replicative capacity (vRC), as conferred by the transmitted Gag sequence, and early set point viral load (SPVL) in individuals recently infected with subtype C HIV-1. Intriguingly, vRC, based on the transmitted Gag sequence, was also positively correlated with viral load in the epidemiologically linked, chronically infected donor near the time of transmission. This establishes vRC as a heritable viral characteristic that can drive similar pathogenic effects in both the chronically infected donor and the acutely infected linked recipient.

Moreover, this work provided a preliminary demonstration that individuals infected with low RC viruses were protected from rapid CD4+ T cell decline. This initial link between vRC and CD4+ T cell decline appeared to be both stronger than and independent of vRC's association with early SPVL. In Chapter III, we expand on the data initially presented in Chapter II, and unequivocally demonstrate that low vRC is strongly correlated with delayed CD4+ T cell decline in a manner independent of the impact of protective HLA-I alleles and vRC's association with SPVL. This suggested that vRC

influenced events very early after transmission that significantly affected the trajectory of subsequent CD4+ T cell decline. We hypothesized that increased vRC may drive (I) elevated inflammatory cytokine profiles, (II) exacerbated gut damage and microbial translocation, (III) increased T cell activation, exhaustion, and proliferation, and (IV) increased magnitude of viral burden in memory CD4+ T cell subsets. The data generated in the examination of these hypotheses established vRC as an integral player in the forces driving HIV pathogenesis, and will hopefully serve to refocus the HIV pathogenesis field to give equal importance to the study of transmitted viral characteristics along with (and in the context of) host immunogenetics.

However, much remains to be discovered with respect to the underlying mechanism(s) by which vRC drives immune activation, increased viral burden in memory CD4+ T cell subsets, and subsequent pathogenesis. Identical viral epitopes have been shown to drive differential CD8+ T cell exhaustion phenotypes based on their relative abundance and persistence during infection[226]. Reasonably, increased vRC would most likely result in elevated levels of circulating HIV antigens, and thus HIV antigen-derived epitopes that can be recognized by cytotoxic T lymphocytes (CTLs), which in turn could lead to aberrant T cell activation via T-cell receptor (TCR) overstimulation. In support of this hypothesis, we found vRC to be positively correlated with PD-1 expression (a negative regulator of T cell activation linked to cellular exhaustion and decreased effector function) on CD8+ T cells[227,228]. As PD-1 expression in chronic HIV infection was shown to be most pronounced on HIV-specific CD8+ T cell populations[227], this suggests that increased circulating antigen may be a

factor linking vRC to increased markers of immune activation and exhaustion in the CD8+ T cell compartment.

It has been definitively shown in both the lymphocytic choriomeningitis virus (LCMV) model system[229,230], as well as in HIV infection, that persistent antigen-driven T cell stimulation leads to cellular exhaustion and functional impairment of CD8+ T cell responses in the chronic stage of infection[227,228]. However, higher circulating epitope levels in acute infection can also lead to deletion of otherwise dominant virus-specific CD8+ T cell lineages[226]. Furthermore, we propose that vRC might most dramatically affect the magnitude of acute viral loads, before the advent of a strong cellular immune response necessary to reduce acute viremia down to a stable viral load set point. This could lead to skewing of the CD8+ T cell epitope dominance hierarchy by the deletion of dominant, and potentially protective, CD8+ T cell responses due to the increased levels of circulating antigen. This may be a contributing factor driving the association between vRC and CD8+ T cell activation, exhaustion, and an attenuated cytotoxicity phenotype (where cells express low levels of perforin/granzyme).

Alternatively, low vRC may influence the quality of the CD8+ T cells response in a manner independent of antigen load. In Chapter II, we demonstrated that low vRC was significantly associated with Gag sequences more closely related to the viral population consensus sequence. That is, viruses with low in vitro vRC had Gag sequences containing fewer polymorphisms and more consensus amino acid residues. This may lead to the preservation of targetable CTL epitopes, which could then lead to a more broad CTL response. Broad CTL responses, especially those targeting Gag, have been linked to

enhance immune control of HIV[231], and evaluating whether low vRC is associated with broader Gag-specific CTL responses should be the focus of future investigation.

In conjunction with antigen-driven T cell stimulation, the innate immune response undoubtedly contributes to the immune activation associated with vRC. The innate immune response is responsible for generating large amounts of soluble inflammatory mediators during acute viral infections when innate immune cells recognize pathogen-associated molecular patterns via innate immune sensors, such as toll-like receptors (TLRs). Plasmacytoid dendritic cells (pDCs), the immune system's primary producers of type I interferon, are activated after HIV endocytosis. Following endocytosis and endosome acidification, HIV genomic RNA sequences can signal through TLR7, which leads to the production of interferon alpha (IFNa) and tumor necrosis factor alpha (TNFa)[232]. Other uridine-rich HIV-derived RNA sequences can signal monocytes to produce interleukin-6 (IL-6) and IFNa[233]. We show that IFNa, TNFa, and IL-6 levels in the plasma are increased at seroconversion in individuals infected with highly replicating viruses, and the increased abundance of viral RNA-derived TLR7-agonists may contribute to this elevated interferon and inflammatory cytokine production. Additionally, experiments using purified T cell populations have shown that CD4+ T cells enter cell cycle upon direct stimulation with TLR agonists[234]. This suggests that HIV-derived TLR agonists, which may be increased in individuals infected with viruses exhibiting high vRC, could be either directly or indirectly responsible for the observed correlation between vRC and CD4+ T cell proliferation (an indicator of accelerated pathogenesis in HIV infection) as demonstrated in Chapter III.

One of the most striking and perhaps most clinically relevant findings presented in Chapter III is the observation that increased vRC is associated with elevated cell-associated viral DNA in central memory CD4+ T cells. Since total cell-associated viral DNA levels are more closely correlated with integrated HIV DNA rather than non-functional 2-LTR circles[83], we interpret this finding to suggest that high vRC drives, through some unknown mechanism, the size of the initial latent viral reservoir. This observation has at least two important clinical implications. First, proof-of-concept cure strategies aimed at reactivating and eliminating the viral reservoir might benefit from targeting individuals infected with poorly replicating viruses as our results suggest these individuals may harbor diminished viral reservoirs and thus may be more amenable candidates. Second, this suggests that vaccination strategies that do not provide sterilizing immunity, but that do induce functional CTL responses capable of suppressing early viral replication or driving early viral attenuation through CTL escape could provide a benefit for the vaccinated individual as well as new hosts upon transmission, since viruses with an attenuated replication phenotype may seed smaller latent reservoirs. Studies aimed at elucidating the underlying mechanism(s) by which increased vRC drives the size of the initial viral reservoir should be a priority as these studies may yield valuable insight into the development of targeted therapeutics aimed at eliminating the latent viral reservoir.

Unique mechanisms of HLA-mediated protection

In Chapter IV, we reported the identification of several HLA class I, HLA class II, and KIR alleles (HLA-B*1401, B*57, B*5801, B*81, DQB1*02, DRB1*15, and KIR 2DS4) that were capable of protecting individuals from rapid CD4+ T cell decline, but that,

paradoxically, did not substantially reduce plasma viral loads. Furthermore, HLA class I alleles exhibiting this protective phenotype were correlated with reduced microbial translocation during acute infection and the subsequent maintenance of gut integrity. This data describes a novel action by which particular HLA specificities can reduce the pathogenic effects of HIV infection. We speculate that these HLA alleles engender a CTL response capable of suppressing viral replication at a critical time point (early in acute infection) and at a critical location (the gastrointestinal tract). Additional studies will be required to define the viral epitopes targeted by these alleles and to determine if CTLs from these individuals exhibit unique phenotypes, such as enhanced polyfunctionality, cytotoxicity, or ability to inhibit viral replication ex vivo. These studies will be imperative in order to define the correlate(s) of protection from CD4+ T cell decline associated with these protective HLA alleles, and to determine if these correlates are distinct from those of HLA alleles that robustly suppress viral loads.

An alternative hypothesis that may explain the protective effect of these HLA class I alleles in the absence of early viral control could be due to differences in pre-infection phenotypes. We interpret reduced microbial translocation to be the result of early and robust control of viral replication, specifically in the gastrointestinal tract. However, these alleles may additionally (or alternatively) be associated with reduced microbial translocation, enhanced gut integrity, and reduced general immune activation pre-infection, perhaps due to their control of or influence on an individual's gut microbiome. Experiments evaluating the levels of pre-infection LPS in the plasma of individuals with protective HLA will serve to help definitively answer this question.

Though we observed a striking phenotype associated with protective HLA class I alleles, the protective HLA class II alleles identified in Chapter IV (DQB1*02 and DRB1*15) were not associated with a reduction in microbial translocation. However, we observed an enrichment in CD4+ T cells that were positive for both perforin and granzyme B in individuals carrying either DQB1*02 or DRB1*15. This suggests that individuals with these protective HLA class II alleles generated CD4+ T cell responses with augmented cytotoxic potential, a phenotype that has been associated with more favorable disease outcome in studies by other groups[71]. Additionally, CD4+ T cells from individuals carrying these protective alleles exhibited a less exhausted phenotype as demonstrated by significantly lower percentages of PD-1 positive CD4+ T cells in both the central and effector memory compartments. This implies that the epitopes targeted by these HLA class II alleles may be unique in their ability to elicit a more functional CD4+ T cell response with enhanced cytotoxic potential, and the characterization of these epitopes will be an important prospective goal. In defining new HLA class I and class II alleles capable of generating protective immune responses in the absence of control of plasma viral load, we have opened up new avenues of research into the determinants of a protective cellular immune response and the underlying mechanisms responsible for this control. A more in depth understanding of the correlates of protective cellular immunity will be integral in the design of HIV vaccines that contain a T cell component.

The complex nature of sex-based differences in the immune response to HIV infection

Though HIV infection in females is consistently characterized by plasma viral loads at least 60% lower than males[193], they progress to AIDS at a similar rate[215-

219]. Furthermore, females exhibit higher CD8+ T cell activation in chronic HIV infection, and this has been linked to increased IFNa production by pDCs isolated from females[224], which in turn leads to elevated expression of interferon stimulated genes[225]. Thus, it is proposed that the protective effect generally associated with reduced viral loads is mitigated in females by elevated levels of innate and adaptive immune activation.

However, in Chapter IV, we demonstrate that female gender (in conjunction with vRC and protective HLA class I alleles) is associated with reduced plasma LPS at seroconversion and is thus a generally protective phenotype. Therefore, we hypothesize that the female immune response to HIV infection may be bimodal, in that early events are characterized by a more favorable immune response that possesses a greater capacity to suppress viral replication, while during the chronic stage of infection, females present with elevated immune activation and accelerated pathogenesis. This is illustrated in Appendix Figure 1, where, in a group of 127 acutely infected Zambians with longitudinal follow up, females have significantly lower plasma viral loads up to two years post infection (Appendix Figure 1A). Indeed, females receive some clinical benefit, as they have significantly higher CD4+ T cell counts up to twelve months post infection (Appendix Figure 1B). This suggests that early events in HIV infection may be radically different between males and females, but that this protective benefit for females wanes with time.

In a preliminary analysis, we looked retrospectively at inflammatory cytokine profiles and cellular immune activation levels evaluated at the seroconversion time point (an average of 45 days after the estimated date of infection) in 11 females and 23 males.

Strikingly, females showed significantly lower CD8+ T cell activation, as evidenced by fewer cells expressing the CD38 and HLA-DR activation markers. This remained significant when controlling for plasma viral load in a multivariable generalized linear model (Appendix Figure 2). This is in direct contrast to the cellular activation phenotype shown in females during the chronic stage of infection[224], where females exhibit significantly higher percentages of CD8+ T cells expressing CD38 and HLA-DR. Additionally, females are characterized by distinct inflammatory cytokine profiles at seroconversion when compared to males (Appendix Figure 3). These preliminary data provide clear evidence that the immune response to HIV infection is significantly different between men and women, and that this response is initially more favorable in females. However, subsequent events must reverse this protective phenotype in women, and experiments to elucidate the underlying cause of this deleterious transition will be the focus of future investigation.

One potential contribution to the observed differences in early immune profiles between men and women may be the nature of the transmitted virus. Recently, in an analysis of the *gag*, *pol*, and *nef* gene sequences for 135 transmission pairs, we have shown that a highly significant selection bias occurs during transmission, resulting in preferential transmission of consensus amino acid residues[235]. This bias appears to be related to the transmission of viruses with greater overall fitness. Strikingly, this selection bias is different during infection of men and women, with women being infected with viruses that have significantly more non-consensus amino acids than the men, and suggests that the barrier to infection is less in women. That this in turn represents the transmission of less fit viruses to women is supported by the distinct kinetics of reversion

of non-consensus polymorphisms to consensus between men and women following acquisition (Appendix Figure 4). During a 24-month follow up period in which *gag*, *pol*, and *nef* population sequences derived from plasma were amplified at regular 3 month intervals up to 24 months post infection, females exhibited significantly faster rates of reversion. This implies that the fitness of the transmitted/founder virus in females may be disproportionately reduced in comparison to viruses transmitted to males, which in turn could explain the early clinical advantage of women. The loss of this clinical benefit for females, as evidenced by a lack of significantly different CD4+ T cells counts between men and women at 24 months post infection, coincides with equivalent levels of reversion at this time-point. This intriguing result provides impetus to study the interplay of immunologic parameters, inflammatory profiles and viral characteristics in defining the differences in HIV-1 disease progression between men and women post seroconversion.

In Conclusion

The data presented here significantly expand our current knowledge of the viral and host characteristics that influence the trajectory of HIV-1 pathogenesis, and specifically highlight the importance of early events post transmission that dramatically impact the course of disease. We have demonstrated that HIV-1 replicative capacity, as defined by the *gag* gene, affects early events after transmission and, furthermore, that high viral replicative capacity (vRC) initiates an inflammatory state characterized by increased cellular immune activation, exhaustion, and aberrant proliferation; all of which are well-defined markers of HIV pathogenesis. High vRC is also associated with an increased viral

burden in central memory CD4+ T cells, a cell type integral for the maintenance of the latent viral reservoir. However, events dictated by vRC can be further modulated by the adaptive cellular immune response, the specificity of which is dictated by host immunogenetics. We have elucidated protective HLA alleles that significantly protect individuals from rapid CD4+ T cell decline in a manner distinct from the effects of viral replicative capacity. Interestingly, these protective HLA alleles appear not to delay CD4+ T cell decline by suppressing plasma viremia, as has been reported in the past, but appear to exert their protective effect by limiting early microbial translocation, presumably by controlling viral replication in the gastrointestinal tract and thus maintaining the integrity of the gut epithelia. Together, this collection of findings illustrates the complex interplay of transmitted viral characteristics and host immune responses in determining the trajectory of disease in HIV infection, and has at least 3 important implications for the development of interventional therapeutics and the rational design of an HIV vaccine.

First, the elucidation of new favorable HLA class I and II alleles may provide new viral epitope targets relevant for inclusion in a cellular-based HIV vaccine. Furthermore, the protective effect of these alleles, which reduces microbial translocation early after infection, points to an amenable target for therapeutics that could be aimed at reducing gut damage and maintaining integrity of epithelial tight junctions or eliminating circulating LPS from the plasma, in order to mimic the protective effects of these HLA alleles. Second, since attenuated viral replicative capacity results in a more favorable disease outcome, our results argue that an HIV vaccine able to attenuate early viral replication, either through targeting vulnerable epitopes or by forcing viral escape and the selection of polymorphisms that reduce replicative capacity, would provide a clear benefit

to the vaccinated individual as well as a protective benefit to non-vaccinated individuals by increasing the chances of transmitting a low replicating variant. Third, attenuated vRC is associated with a diminished viral reservoir in vulnerable T cell populations, demonstrating that therapeutic interventions that reduce viral fitness pre-transmission, or quickly control viral replication post transmission, may reduce the size of the latent viral reservoir and provide more tractable candidates for cure strategies.

Bibliography

1. Centers for Disease C (1981) Pneumocystis pneumonia--Los Angeles. MMWR Morb Mortal Wkly Rep 30: 250-252.

2. Centers for Disease C (1981) Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. MMWR Morb Mortal Wkly Rep 30: 305-308.

3. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-871.

4. (WHO) WHO (2013) Global report: UNAIDS report on the global AIDS epidemic 2013. WHO Library Cataloguing-in-Publication Data: UNAIDS.

5. Dunkle KL, Stephenson R, Karita E, Chomba E, Kayitenkore K, et al. (2008) New heterosexually transmitted HIV infections in married or cohabiting couples in urban Zambia and Rwanda: an analysis of survey and clinical data. Lancet 371: 2183-2191.

6. Collier AC, Coombs RW, Schoenfeld DA, Bassett RL, Timpone J, et al. (1996) Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. N Engl J Med 334: 1011-1017.

7. D'Aquila RT, Hughes MD, Johnson VA, Fischl MA, Sommadossi JP, et al. (1996) Nevirapine, zidovudine, and didanosine compared with zidovudine and didanosine in patients with HIV-1 infection. A randomized, double-blind, placebo-controlled trial. National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators. Ann Intern Med 124: 1019-1030.

8. Staszewski S, Miller V, Rehmet S, Stark T, De Cree J, et al. (1996) Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients. AIDS 10: F1-7.

9. Arts EJ, Hazuda DJ (2012) HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med 2: a007161.

10. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, et al. (2005) Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. PLoS Med 2: e298.

11. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, et al. (2007) Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. Lancet 369: 643-656.

12. Gray RH, Kigozi G, Serwadda D, Makumbi F, Watya S, et al. (2007) Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. Lancet 369: 657-666.

13. Shaw GM, Hunter E (2012) HIV transmission. Cold Spring Harb Perspect Med 2.

14. (IHME) IfHMaE (2013) GBD Arrow Diagram.

15. Worobey M, Telfer P, Souquiere S, Hunter M, Coleman CA, et al. (2010) Island biogeography reveals the deep history of SIV. Science 329: 1487.

16. Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, et al. (1986) Isolation of a new human retrovirus from West African patients with AIDS. Science 233: 343-346.

17. Apetrei C, Robertson DL, Marx PA (2004) The history of SIVS and AIDS: epidemiology, phylogeny and biology of isolates from naturally SIV infected non-human primates (NHP) in Africa. Front Biosci 9: 225-254.

18. Klatt NR, Silvestri G, Hirsch V (2012) Nonpathogenic simian immunodeficiency virus infections. Cold Spring Harb Perspect Med 2: a007153.

19. Sharp PM, Hahn BH (2011) Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med 1: a006841.

20. Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S (1990) Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature 345: 356-359.

21. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR (1989) An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339: 389-392.

22. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, et al. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397: 436-441.

23. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, et al. (2006) Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313: 523-526.

24. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, et al. (2008) Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455: 661-664.

25. John G. Robinson KHR, Elizabeth L. Bennett (1999) Wildlife Harvest in Logged Tropical Forests. Science 284.

26. Peeters M, Courgnaud V, Abela B, Auzel P, Pourrut X, et al. (2002) Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. Emerg Infect Dis 8: 451-457.

27. Aghokeng AF, Liu W, Bibollet-Ruche F, Loul S, Mpoudi-Ngole E, et al. (2006) Widely varying SIV prevalence rates in naturally infected primate species from Cameroon. Virology 345: 174-189.

28. Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. Science 287: 607-614.

29. Sodora DL, Allan JS, Apetrei C, Brenchley JM, Douek DC, et al. (2009) Toward an AIDS vaccine: lessons from natural simian immunodeficiency virus infections of African nonhuman primate hosts. Nat Med 15: 861-865.

30. Chahroudi A, Bosinger SE, Vanderford TH, Paiardini M, Silvestri G (2012) Natural SIV hosts: showing AIDS the door. Science 335: 1188-1193.

31. Phillips-Conroy JE, Jolly CJ, Petros B, Allan JS, Desrosiers RC (1994) Sexual transmission of SIVagm in wild grivet monkeys. J Med Primatol 23: 1-7.

32. Rey-Cuille MA, Berthier JL, Bomsel-Demontoy MC, Chaduc Y, Montagnier L, et al. (1998) Simian immunodeficiency virus replicates to high levels in sooty mangabeys without inducing disease. J Virol 72: 3872-3886.

33. Silvestri G, Sodora DL, Koup RA, Paiardini M, O'Neil SP, et al. (2003) Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. Immunity 18: 441-452.

34. Santiago ML, Range F, Keele BF, Li Y, Bailes E, et al. (2005) Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. J Virol 79: 12515-12527.

35. Keele BF, Jones JH, Terio KA, Estes JD, Rudicell RS, et al. (2009) Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460: 515-519.

36. Persaud D, Gay H, Ziemniak C, Chen YH, Piatak M, Jr., et al. (2013) Absence of detectable HIV-1 viremia after treatment cessation in an infant. N Engl J Med 369: 1828-1835.

37. Hutter G, Nowak D, Mossner M, Ganepola S, Mussig A, et al. (2009) Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med 360: 692-698.

38. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, et al. (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 273: 1856-1862.

39. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, et al. (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382: 722-725.

40. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, et al. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86: 367-377.

41. Klatzmann D, Barre-Sinoussi F, Nugeyre MT, Danquet C, Vilmer E, et al. (1984) Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. Science 225: 59-63.

42. Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, et al. (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312: 767-768.

43. Munier ML, Kelleher AD (2007) Acutely dysregulated, chronically disabled by the enemy within: T-cell responses to HIV-1 infection. Immunol Cell Biol 85: 6-15.

44. Swain SL, McKinstry KK, Strutt TM (2012) Expanding roles for CD4(+) T cells in immunity to viruses. Nat Rev Immunol 12: 136-148.

45. Crotty S (2011) Follicular helper CD4 T cells (TFH). Annu Rev Immunol 29: 621-663.

46. Moir S, Fauci AS (2009) B cells in HIV infection and disease. Nat Rev Immunol 9: 235-245.

47. Lane HC, Masur H, Edgar LC, Whalen G, Rook AH, et al. (1983) Abnormalities of B-cell activation and immunoregulation in patients with the acquired immunodeficiency syndrome. N Engl J Med 309: 453-458.

48. Ammann AJ, Schiffman G, Abrams D, Volberding P, Ziegler J, et al. (1984) B-cell immunodeficiency in acquired immune deficiency syndrome. JAMA 251: 1447-1449.

49. Pahwa SG, Quilop MT, Lange M, Pahwa RN, Grieco MH (1984) Defective B-lymphocyte function in homosexual men in relation to the acquired immunodeficiency syndrome. Ann Intern Med 101: 757-763.

50. Richman DD, Wrin T, Little SJ, Petropoulos CJ (2003) Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 100: 4144-4149.

51. Wei X, Decker JM, Wang S, Hui H, Kappes JC, et al. (2003) Antibody neutralization and escape by HIV-1. Nature 422: 307-312.

52. Moir S, Malaspina A, Li Y, Chun TW, Lowe T, et al. (2000) B cells of HIV-1-infected patients bind virions through CD21-complement interactions and transmit infectious virus to activated T cells. J Exp Med 192: 637-646.

53. Mandl JN, Barry AP, Vanderford TH, Kozyr N, Chavan R, et al. (2008) Divergent TLR7 and TLR9 signaling and type I interferon production distinguish pathogenic and nonpathogenic AIDS virus infections. Nat Med 14: 1077-1087.

54. Rieckmann P, Poli G, Fox CH, Kehrl JH, Fauci AS (1991) Recombinant gp120 specifically enhances tumor necrosis factor-alpha production and Ig secretion in B lymphocytes from HIV-infected individuals but not from seronegative donors. J Immunol 147: 2922-2927.

55. Weimer R, Zipperle S, Daniel V, Zimmermann R, Schimpf K, et al. (1998) HIV-induced IL-6/IL-10 dysregulation of CD4 cells is associated with defective B cell help and autoantibody formation against CD4 cells. Clin Exp Immunol 111: 20-29.

56. Muller F, Aukrust P, Nordoy I, Froland SS (1998) Possible role of interleukin-10 (IL-10) and CD40 ligand expression in the pathogenesis of hypergammaglobulinemia in human immunodeficiency virus infection: modulation of IL-10 and Ig production after intravenous Ig infusion. Blood 92: 3721-3729.

57. Napolitano LA, Grant RM, Deeks SG, Schmidt D, De Rosa SC, et al. (2001) Increased production of IL-7 accompanies HIV-1-mediated T-cell depletion: implications for T-cell homeostasis. Nat Med 7: 73-79.

58. Fry TJ, Connick E, Falloon J, Lederman MM, Liewehr DJ, et al. (2001) A potential role for interleukin-7 in T-cell homeostasis. Blood 97: 2983-2990.

59. Malaspina A, Moir S, Ho J, Wang W, Howell ML, et al. (2006) Appearance of immature/transitional B cells in HIV-infected individuals with advanced disease: correlation with increased IL-7. Proc Natl Acad Sci U S A 103: 2262-2267.

60. Malaspina A, Moir S, Chaitt DG, Rehm CA, Kottilil S, et al. (2007) Idiopathic CD4+ T lymphocytopenia is associated with increases in immature/transitional B cells and serum levels of IL-7. Blood 109: 2086-2088.

61. Yang OO, Kalams SA, Rosenzweig M, Trocha A, Jones N, et al. (1996) Efficient lysis of human immunodeficiency virus type 1-infected cells by cytotoxic T lymphocytes. J Virol 70: 5799-5806.

62. Ridge JP, Di Rosa F, Matzinger P (1998) A conditioned dendritic cell can be a temporal bridge between a CD4+ T-helper and a T-killer cell. Nature 393: 474-478.

63. Bennett SR, Carbone FR, Karamalis F, Flavell RA, Miller JF, et al. (1998) Help for cytotoxic-T-cell responses is mediated by CD40 signalling. Nature 393: 478-480.

64. Schoenberger SP, Toes RE, van der Voort EI, Offringa R, Melief CJ (1998) T-cell help for cytotoxic T lymphocytes is mediated by CD40-CD40L interactions. Nature 393: 480-483.

65. Sun JC, Bevan MJ (2003) Defective CD8 T cell memory following acute infection without CD4 T cell help. Science 300: 339-342.

66. Mercado R, Vijh S, Allen SE, Kerksiek K, Pilip IM, et al. (2000) Early programming of T cell populations responding to bacterial infection. J Immunol 165: 6833-6839.

67. Janssen EM, Lemmens EE, Wolfe T, Christen U, von Herrath MG, et al. (2003) CD4+ T cells are required for secondary expansion and memory in CD8+ T lymphocytes. Nature 421: 852-856.

68. Shedlock DJ, Shen H (2003) Requirement for CD4 T cell help in generating functional CD8 T cell memory. Science 300: 337-339.

69. Nakanishi Y, Lu B, Gerard C, Iwasaki A (2009) CD8(+) T lymphocyte mobilization to virus-infected tissue requires CD4(+) T-cell help. Nature 462: 510-513.

70. Soghoian DZ, Streeck H (2010) Cytolytic CD4(+) T cells in viral immunity. Expert Rev Vaccines 9: 1453-1463.

71. Soghoian DZ, Jessen H, Flanders M, Sierra-Davidson K, Cutler S, et al. (2012) HIV-specific cytolytic CD4 T cell responses during acute HIV infection predict disease outcome. Sci Transl Med 4: 123ra125.

72. Douek DC, Brenchley JM, Betts MR, Ambrozak DR, Hill BJ, et al. (2002) HIV preferentially infects HIV-specific CD4+ T cells. Nature 417: 95-98.

73. Brown PO, Bowerman B, Varmus HE, Bishop JM (1989) Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc Natl Acad Sci U S A 86: 2525-2529.

74. Coffin JM, Hughes SH, Varmus HE (1997) The Interactions of Retroviruses and their Hosts. In: Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Cold Spring Harbor (NY).

75. Craigie R, Bushman FD (2012) HIV DNA integration. Cold Spring Harb Perspect Med 2: a006890.

76. Nabel G, Baltimore D (1987) An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature 326: 711-713.

77. Kinoshita S, Su L, Amano M, Timmerman LA, Kaneshima H, et al. (1997) The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. Immunity 6: 235-244.

78. Siliciano RF, Greene WC (2011) HIV latency. Cold Spring Harb Perspect Med 1: a007096.

79. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, et al. (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278: 1295-1300.

80. Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, et al. (1997) Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 278: 1291-1295.

81. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, et al. (1997) Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94: 13193-13197.

82. Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, et al. (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature 387: 183-188.

83. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, et al. (2009) HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat Med 15: 893-900.

84. Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, et al. (1999) Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med 5: 512-517.

85. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, et al. (2003) Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nat Med 9: 727-728.

86. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, et al. (2013) Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell 155: 540-551.

87. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, et al. (2008) HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A 105: 16725-16730.

88. Deeks SG, Wrin T, Liegler T, Hoh R, Hayden M, et al. (2001) Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. N Engl J Med 344: 472-480.

89. Ruff CT, Ray SC, Kwon P, Zinn R, Pendleton A, et al. (2002) Persistence of wild-type virus and lack of temporal structure in the latent reservoir for human immunodeficiency virus type 1 in pediatric patients with extensive antiretroviral exposure. J Virol 76: 9481-9492.

90. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, et al. (2006) Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. J Virol 80: 6441-6457.

91. Luzuriaga K, McManus M, Catalina M, Mayack S, Sharkey M, et al. (2000) Early therapy of vertical human immunodeficiency virus type 1 (HIV-1) infection: control of viral replication and absence of persistent HIV-1-specific immune responses. J Virol 74: 6984-6991.

92. Persaud D, Siberry GK, Ahonkhai A, Kajdas J, Monie D, et al. (2004) Continued production of drug-sensitive human immunodeficiency virus type 1 in children on combination antiretroviral therapy who have undetectable viral loads. J Virol 78: 968-979.

93. Bailey J, Blankson JN, Wind-Rotolo M, Siliciano RF (2004) Mechanisms of HIV-1 escape from immune responses and antiretroviral drugs. Curr Opin Immunol 16: 470-476.

94. Hu WS, Hughes SH (2012) HIV-1 reverse transcription. Cold Spring Harb Perspect Med 2.

95. Roberts JD, Bebenek K, Kunkel TA (1988) The accuracy of reverse transcriptase from HIV-1. Science 242: 1171-1173.

96. Mansky LM, Temin HM (1995) Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol 69: 5087-5094.

97. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH (2010) Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J Virol 84: 9864-9878.

98. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271: 1582-1586.

99. Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, et al. (2000) High rate of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 74: 1234-1240.

100. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, et al. (2002) Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J Virol 76: 11273-11282.

101. Rhodes T, Wargo H, Hu WS (2003) High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. J Virol 77: 11193-11200.

102. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2006) Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20: W13-23.

103. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, et al. (1994) Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. J Virol 68: 4650-4655.

104. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB (1994) Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. J Virol 68: 6103-6110.

105. Matano T, Shibata R, Siemon C, Connors M, Lane HC, et al. (1998) Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. J Virol 72: 164-169.

106. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, et al. (1999) Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. Science 283: 857-860.

107. Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, et al. (1999) Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. J Exp Med 189: 991-998.

108. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, et al. (1991) Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature 354: 453-459.

109. Koenig S, Conley AJ, Brewah YA, Jones GM, Leath S, et al. (1995) Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. Nat Med 1: 330-336.

110. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, et al. (1997) Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nat Med 3: 205-211.

111. Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, et al. (1997) Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. Proc Natl Acad Sci U S A 94: 1890-1895.

112. Goulder PJ, Watkins DI (2004) HIV and SIV CTL escape: implications for vaccine design. Nat Rev Immunol 4: 630-640.

113. Goulder PJ, Watkins DI (2008) Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat Rev Immunol 8: 619-630.

114. Draenert R, Le Gall S, Pfafferott KJ, Leslie AJ, Chetty P, et al. (2004) Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. J Exp Med 199: 905-915.

115. Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichterfeld M, et al. (2004) Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. J Virol 78: 7069-7078.

116. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, et al. (1997) Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 3: 212-217.

117. Goulder PJ, Brander C, Tang Y, Tremblay C, Colbert RA, et al. (2001) Evolution and transmission of stable CTL escape mutations in HIV infection. Nature 412: 334-338.

118. Minang JT, Trivett MT, Coren LV, Barsov EV, Piatak M, Jr., et al. (2008) The Mamu B 17-restricted SIV Nef IW9 to TW9 mutation abrogates correct epitope processing and presentation without loss of replicative fitness. Virology 375: 307-314.

119. Forthal DN, Landucci G, Daar ES (2001) Antibody from patients with acute human immunodeficiency virus (HIV) infection inhibits primary strains of HIV type 1 in the presence of natural-killer effector cells. J Virol 75: 6953-6961.

120. Aasa-Chapman MM, Holuigue S, Aubin K, Wong M, Jones NA, et al. (2005) Detection of antibody-dependent complement-mediated inactivation of both autologous and heterologous virus in primary human immunodeficiency virus type 1 infection. J Virol 79: 2823-2830.

121. Liao HX, Lynch R, Zhou T, Gao F, Alam SM, et al. (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496: 469-476.

122. Clark SJ, Saag MS, Decker WD, Campbell-Hill S, Roberson JL, et al. (1991) High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. N Engl J Med 324: 954-960.

123. Daar ES, Moudgil T, Meyer RD, Ho DD (1991) Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. N Engl J Med 324: 961-964.

124. Little SJ, McLean AR, Spina CA, Richman DD, Havlir DV (1999) Viral dynamics of acute HIV-1 infection. J Exp Med 190: 841-850.

125. Haase AT (2010) Targeting early infection to prevent HIV-1 mucosal transmission. Nature 464: 217-223.

126. Mellors JW, Kingsley LA, Rinaldo CR, Jr., Todd JA, Hoo BS, et al. (1995) Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Ann Intern Med 122: 573-579.

127. O'Brien TR, Blattner WA, Waters D, Eyster E, Hilgartner MW, et al. (1996) Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. JAMA 276: 105-110.

128. McCune JM (2001) The dynamics of CD4+ T-cell depletion in HIV disease. Nature 410: 974-979.

129. Rowland-Jones S (1999) HIV infection: where have all the T cells gone? Lancet 354: 5-7.

130. Finkel TH, Tudor-Williams G, Banda NK, Cotton MF, Curiel T, et al. (1995) Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes. Nat Med 1: 129-134.

131. Gougeon ML, Lecoeur H, Dulioust A, Enouf MG, Crouvoiser M, et al. (1996) Programmed cell death in peripheral lymphocytes from HIV-infected persons:

increased susceptibility to apoptosis of CD4 and CD8 T cells correlates with lymphocyte activation and with disease progression. J Immunol 156: 3509-3520.

132. Bofill M, Mocroft A, Lipman M, Medina E, Borthwick NJ, et al. (1996) Increased numbers of primed activated CD8+CD38+CD45RO+ T cells predict the decline of CD4+ T cells in HIV-1-infected patients. AIDS 10: 827-834.

133. Giorgi JV, Hultin LE, McKeating JA, Johnson TD, Owens B, et al. (1999) Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. J Infect Dis 179: 859-870.

134. Mohri H, Bonhoeffer S, Monard S, Perelson AS, Ho DD (1998) Rapid turnover of T lymphocytes in SIV-infected rhesus macaques. Science 279: 1223-1227.

135. Hellerstein M, Hanley MB, Cesar D, Siler S, Papageorgopoulos C, et al. (1999) Directly measured kinetics of circulating T lymphocytes in normal and HIV-1-infected humans. Nat Med 5: 83-89.

136. Veazey RS, DeMaria M, Chalifoux LV, Shvetz DE, Pauley DR, et al. (1998) Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science 280: 427-431.

137. Gordon SN, Klatt NR, Bosinger SE, Brenchley JM, Milush JM, et al. (2007) Severe depletion of mucosal CD4+ T cells in AIDS-free simian immunodeficiency virus-infected sooty mangabeys. J Immunol 179: 3026-3034.

138. Pandrea IV, Gautam R, Ribeiro RM, Brenchley JM, Butler IF, et al. (2007) Acute loss of intestinal CD4+ T cells is not predictive of simian immunodeficiency virus virulence. J Immunol 179: 3035-3046.

139. Brenchley JM, Schacker TW, Ruff LE, Price DA, Taylor JH, et al. (2004) CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. J Exp Med 200: 749-759.

140. Mehandru S, Poles MA, Tenner-Racz K, Horowitz A, Hurley A, et al. (2004) Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. J Exp Med 200: 761-770.

141. Estes JD, Harris LD, Klatt NR, Tabb B, Pittaluga S, et al. (2010) Damaged intestinal epithelial integrity linked to microbial translocation in pathogenic simian immunodeficiency virus infections. PLoS Pathog 6: e1001052.

142. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, et al. (2006) Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nat Med 12: 1365-1371.

143. Stacey AR, Norris PJ, Qin L, Haygreen EA, Taylor E, et al. (2009) Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. J Virol 83: 3719-3733.

144. Rehermann B, Nascimbeni M (2005) Immunology of hepatitis B virus and hepatitis C virus infection. Nat Rev Immunol 5: 215-229.

145. Bosinger SE, Li Q, Gordon SN, Klatt NR, Duan L, et al. (2009) Global genomic analysis reveals rapid control of a robust innate response in SIV-infected sooty mangabeys. J Clin Invest 119: 3556-3572.

146. Beaumier CM, Harris LD, Goldstein S, Klatt NR, Whitted S, et al. (2009) CD4 downregulation by memory CD4+ T cells in vivo renders African green monkeys resistant to progressive SIVagm infection. Nat Med 15: 879-885.

147. Paiardini M, Cervasi B, Reyes-Aviles E, Micci L, Ortiz AM, et al. (2011) Low levels of SIV infection in sooty mangabey central memory CD(4)(+) T cells are associated with limited CCR5 expression. Nat Med 17: 830-836.

148. Okoye A, Meier-Schellersheim M, Brenchley JM, Hagen SI, Walker JM, et al. (2007) Progressive CD4+ central memory T cell decline results in CD4+ effector memory insufficiency and overt disease in chronic SIV infection. J Exp Med 204: 2171-2185.

149. Sallusto F, Lenig D, Forster R, Lipp M, Lanzavecchia A (1999) Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature 401: 708-712.

150. Seddon B, Tomlinson P, Zamoyska R (2003) Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells. Nat Immunol 4: 680-686.

151. Lackner AA, Lederman MM, Rodriguez B (2012) HIV pathogenesis: the host. Cold Spring Harb Perspect Med 2: a007005.

152. Hunt PW, Martin JN, Sinclair E, Bredt B, Hagos E, et al. (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. J Infect Dis 187: 1534-1543.

153. Kuller LH, Tracy R, Belloso W, De Wit S, Drummond F, et al. (2008) Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. PLoS Med 5: e203.

154. Sandler NG, Wand H, Roque A, Law M, Nason MC, et al. (2011) Plasma levels of soluble CD14 independently predict mortality in HIV infection. J Infect Dis 203: 780-790.

155. Deeks SG, Kitchen CM, Liu L, Guo H, Gascon R, et al. (2004) Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. Blood 104: 942-947.

156. Walker BD, Yu XG (2013) Unravelling the mechanisms of durable control of HIV-1. Nat Rev Immunol 13: 487-498.

157. Cao Y, Qin L, Zhang L, Safrit J, Ho DD (1995) Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. N Engl J Med 332: 201-208.

158. Harrer T, Harrer E, Kalams SA, Elbeik T, Staprans SI, et al. (1996) Strong cytotoxic T cell and weak neutralizing antibody responses in a subset of persons with stable nonprogressing HIV type 1 infection. AIDS Res Hum Retroviruses 12: 585-592.

159. Lambotte O, Boufassa F, Madec Y, Nguyen A, Goujard C, et al. (2005) HIV controllers: a homogeneous group of HIV-1-infected patients with spontaneous control of viral replication. Clin Infect Dis 41: 1053-1056.

160. Doherty PC, Zinkernagel RM (1975) A biological role for the major histocompatibility antigens. Lancet 1: 1406-1409.

161. Becker Y (1994) HIV-1 proteins in infected cells determine the presentation of viral peptides by HLA class I and class II molecules and the nature of the cellular and humoral antiviral immune responses--a review. Virus Genes 8: 249-270.

162. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, et al. (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 2: 405-411.

163. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, et al. (1999) HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science 283: 1748-1752.

164. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, et al. (2000) HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc Natl Acad Sci U S A 97: 2709-2714.

165. Pereyra F, Addo MM, Kaufmann DE, Liu Y, Miura T, et al. (2008) Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. J Infect Dis 197: 563-571.

166. Loffredo JT, Maxwell J, Qi Y, Glidden CE, Borchardt GJ, et al. (2007) Mamu-B*08-positive macaques control simian immunodeficiency virus replication. J Virol 81: 8827-8832.

167. Loffredo JT, Sidney J, Bean AT, Beal DR, Bardet W, et al. (2009) Two MHC class I molecules associated with elite control of immunodeficiency virus replication, Mamu-B*08 and HLA-B*2705, bind peptides with sequence similarity. J Immunol 182: 7763-7775.

168. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, et al. (2009) Common genetic variation and the control of HIV-1 in humans. PLoS Genet 5: e1000791.

169. International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, et al. (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330: 1551-1557.

170. Martin MP, Carrington M (2013) Immunogenetics of HIV disease. Immunol Rev 254: 245-264.

171. Migueles SA, Laborico AC, Shupert WL, Sabbaghian MS, Rabin R, et al. (2002) HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. Nat Immunol 3: 1061-1068.

172. Migueles SA, Osborne CM, Royce C, Compton AA, Joshi RP, et al. (2008) Lytic granule loading of CD8+ T cells is required for HIV-infected cell elimination associated with immune control. Immunity 29: 1009-1021.

173. Hersperger AR, Pereyra F, Nason M, Demers K, Sheth P, et al. (2010) Perforin expression directly ex vivo by HIV-specific CD8 T-cells is a correlate of HIV elite control. PLoS Pathog 6: e1000917.

174. Chen H, Ndhlovu ZM, Liu D, Porter LC, Fang JW, et al. (2012) TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. Nat Immunol 13: 691-700.

175. Saez-Cirion A, Lacabaratz C, Lambotte O, Versmisse P, Urrutia A, et al. (2007) HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection ex vivo and peculiar cytotoxic T lymphocyte activation phenotype. Proc Natl Acad Sci U S A 104: 6776-6781.

176. Klein MR, van Baalen CA, Holwerda AM, Kerkhof Garde SR, Bende RJ, et al. (1995) Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. J Exp Med 181: 1365-1372.

177. Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, et al. (2011) Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proc Natl Acad Sci U S A 108: 11530-11535.

178. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, et al. (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet 31: 429-434.

179. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, et al. (2007) Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. Nat Genet 39: 733-740.

180. Alter G, Martin MP, Teigen N, Carr WH, Suscovich TJ, et al. (2007) Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. J Exp Med 204: 3027-3036.

181. Alter G, Heckerman D, Schneidewind A, Fadda L, Kadie CM, et al. (2011) HIV-1 adaptation to NK-cell-mediated immune pressure. Nature 476: 96-100.

182. Moore CB, John M, James IR, Christiansen FT, Witt CS, et al. (2002) Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 296: 1439-1443.

183. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, et al. (2008) Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol 4: e1000225.

184. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, et al. (2009) Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458: 641-645.

185. Cotton LA, Kuang XT, Le AQ, Carlson JM, Chan B, et al. (2014) Genotypic and functional impact of HIV-1 adaptation to its host population during the North American epidemic. PLoS Genet 10: e1004295.

186. Bailey JR, Zhang H, Wegweiser BW, Yang HC, Herrera L, et al. (2007) Evolution of HIV-1 in an HLA-B*57-positive patient during virologic escape. J Infect Dis 196: 50-55.

187. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, et al. (2007) Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. J Virol 81: 12608-12618.

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

189. Miura T, Brockman MA, Schneidewind A, Lobritz M, Pereyra F, et al. (2009) HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition. J Virol 83: 2743-2755.

190. Prentice HA, Porter TR, Price MA, Cormier E, He D, et al. (2013) HLA-B*57 versus HLA-B*81 in HIV-1 infection: slow and steady wins the race? J Virol 87: 4043-4051.

191. Rosenberg ES, Billingsley JM, Caliendo AM, Boswell SL, Sax PE, et al. (1997) Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. Science 278: 1447-1450.

192. Ranasinghe S, Cutler S, Davis I, Lu R, Soghoian DZ, et al. (2013) Association of HLA-DRB1-restricted CD4(+) T cell responses with HIV immune control. Nat Med 19: 930-933.

193. Yue L, Prentice HA, Farmer P, Song W, He D, et al. (2013) Cumulative impact of host and viral factors on HIV-1 viral-load control during early infection. J Virol 87: 708-715.

194. Tang J, Tang S, Lobashevsky E, Zulu I, Aldrovandi G, et al. (2004) HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. AIDS Res Hum Retroviruses 20: 19-25.

195. Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly CA, Serwadda D, et al. (2010) HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. PLoS Pathog 6: e1000876.

196. Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, et al. (2014) Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. Science 343: 1243727.

197. Schindler M, Munch J, Kutsch O, Li H, Santiago ML, et al. (2006) Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. Cell 125: 1055-1067.

198. Kirchhoff F (2009) Is the high virulence of HIV-1 an unfortunate coincidence of primate lentiviral evolution? Nat Rev Microbiol 7: 467-476.

199. Neil SJ, Zang T, Bieniasz PD (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451: 425-430.

200. Mwimanzi P, Markle TJ, Ueno T, Brockman MA (2012) Human leukocyte antigen (HLA) class I down-regulation by human immunodeficiency virus type 1 negative

factor (HIV-1 Nef): what might we learn from natural sequence variants? Viruses 4: 1711-1730.

201. Kuang XT LX, Anmole G, Mwimanzi P, Shahid A, Le AQ, Chong L, Qian H, Miura T, Markle T, Baraki B, Connick E, Daar ES, Jessen H, Kelleher AD, Little S, Markowitz M, Pereyra F, Rosenberg ES, Walker BD, Ueno T, Brumme ZL, Brockman MA (2014) Impaired Nef function is associated with early control of HIV-1 viremia. Journal of Virology.

202. Deacon NJ, Tsykin A, Solomon A, Smith K, Ludford-Menting M, et al. (1995) Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. Science 270: 988-991.

203. Kirchhoff F, Greenough TC, Brettler DB, Sullivan JL, Desrosiers RC (1995) Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. N Engl J Med 332: 228-232.

204. Kestler HW, 3rd, Ringler DJ, Mori K, Panicali DL, Sehgal PK, et al. (1991) Importance of the nef gene for maintenance of high virus loads and for development of AIDS. Cell 65: 651-662.

205. Daniel MD, Kirchhoff F, Czajak SC, Sehgal PK, Desrosiers RC (1992) Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene. Science 258: 1938-1941.

206. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, et al. (2008) Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J Exp Med 205: 1009-1017.

207. Miura T, Brumme CJ, Brockman MA, Brumme ZL, Pereyra F, et al. (2009) HLA-associated viral mutations are common in human immunodeficiency virus type 1 elite controllers. J Virol 83: 3407-3412.

208. Miura T, Brumme ZL, Brockman MA, Rosato P, Sela J, et al. (2010) Impaired replication capacity of acute/early viruses in persons who become HIV controllers. J Virol 84: 7581-7591.

209. Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, et al. (2010) Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. J Virol 84: 10820-10831.

210. Fish EN (2008) The X-files in immunity: sex-based differences predispose immune responses. Nat Rev Immunol 8: 737-744.

211. Klein SL, Jedlicka A, Pekosz A (2010) The Xs and Y of immune responses to viral vaccines. Lancet Infect Dis 10: 338-349.

212. McClelland EE, Smith JM (2011) Gender specific differences in the immune response to infection. Arch Immunol Ther Exp (Warsz) 59: 203-213.

213. Bengtsson AK, Ryan EJ, Giordano D, Magaletti DM, Clark EA (2004) 17beta-estradiol (E2) modulates cytokine and chemokine expression in human monocyte-derived dendritic cells. Blood 104: 1404-1410.

214. Kramer PR, Kramer SF, Guan G (2004) 17 beta-estradiol regulates cytokine release through modulation of CD16 expression in monocytes and monocyte-derived macrophages. Arthritis Rheum 50: 1967-1975.

215. Katzenstein DA, Hammer SM, Hughes MD, Gundacker H, Jackson JB, et al. (1996) The relation of virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. AIDS Clinical Trials Group Study 175 Virology Study Team. N Engl J Med 335: 1091-1098.

216. Farzadegan H, Hoover DR, Astemborski J, Lyles CM, Margolick JB, et al. (1998) Sex differences in HIV-1 viral load and progression to AIDS. Lancet 352: 1510-1514.

217. Sterling TR, Lyles CM, Vlahov D, Astemborski J, Margolick JB, et al. (1999) Sex differences in longitudinal human immunodeficiency virus type 1 RNA levels among seroconverters. J Infect Dis 180: 666-672.

218. Sterling TR, Vlahov D, Astemborski J, Hoover DR, Margolick JB, et al. (2001) Initial plasma HIV-1 RNA levels and progression to AIDS in women and men. N Engl J Med 344: 720-725.

219. Gandhi M, Bacchetti P, Miotti P, Quinn TC, Veronese F, et al. (2002) Does patient sex affect human immunodeficiency virus levels? Clin Infect Dis 35: 313-322.

220. Fahey JL, Taylor JM, Detels R, Hofmann B, Melmed R, et al. (1990) The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. N Engl J Med 322: 166-172.

221. Fahey JL, Taylor JM, Manna B, Nishanian P, Aziz N, et al. (1998) Prognostic significance of plasma markers of immune activation, HIV viral load and CD4 T-cell measurements. AIDS 12: 1581-1590.

222. Boasso A, Shearer GM (2008) Chronic innate immune activation as a cause of HIV-1 immunopathogenesis. Clin Immunol 126: 235-242.

223. Hunt PW, Brenchley J, Sinclair E, McCune JM, Roland M, et al. (2008) Relationship between T cell activation and CD4+ T cell count in HIV-seropositive individuals with undetectable plasma HIV RNA levels in the absence of therapy. J Infect Dis 197: 126-133.

224. Meier A, Chang JJ, Chan ES, Pollard RB, Sidhu HK, et al. (2009) Sex differences in the Toll-like receptor-mediated response of plasmacytoid dendritic cells to HIV-1. Nat Med 15: 955-959.

225. Chang JJ, Woods M, Lindsay RJ, Doyle EH, Griesbeck M, et al. (2013) Higher expression of several interferon-stimulated genes in HIV-1-infected females after adjusting for the level of viral replication. J Infect Dis 208: 830-838.

226. Wherry EJ, Blattman JN, Murali-Krishna K, van der Most R, Ahmed R (2003) Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. J Virol 77: 4911-4927.

227. Day CL, Kaufmann DE, Kiepiela P, Brown JA, Moodley ES, et al. (2006) PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. Nature 443: 350-354.

228. Trautmann L, Janbazian L, Chomont N, Said EA, Gimmig S, et al. (2006) Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. Nat Med 12: 1198-1202.

229. Ahmed R, Salmi A, Butler LD, Chiller JM, Oldstone MB (1984) Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice. Role in suppression of cytotoxic T lymphocyte response and viral persistence. J Exp Med 160: 521-540.

230. Li Q, Skinner PJ, Ha SJ, Duan L, Mattila TL, et al. (2009) Visualizing antigen-specific and infected cells in situ predicts outcomes in early viral infection. Science 323: 1726-1729.

231. Julg B, Williams KL, Reddy S, Bishop K, Qi Y, et al. (2010) Enhanced anti-HIV functional activity associated with Gag-specific CD8 T-cell responses. J Virol 84: 5540-5549.

232. Beignon AS, McKenna K, Skoberne M, Manches O, DaSilva I, et al. (2005) Endocytosis of HIV-1 activates plasmacytoid dendritic cells via Toll-like receptor-viral RNA interactions. J Clin Invest 115: 3265-3275.

233. Meier A, Alter G, Frahm N, Sidhu H, Li B, et al. (2007) MyD88-dependent immune activation mediated by human immunodeficiency virus type 1-encoded Toll-like receptor ligands. J Virol 81: 8180-8191.

234. Funderburg N, Luciano AA, Jiang W, Rodriguez B, Sieg SF, et al. (2008) Toll-like receptor ligands induce human T cell activation and death, a model for HIV pathogenesis. PLoS One 3: e1915.

235. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, et al. (2014) HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science 345: 1254031.

Appendix

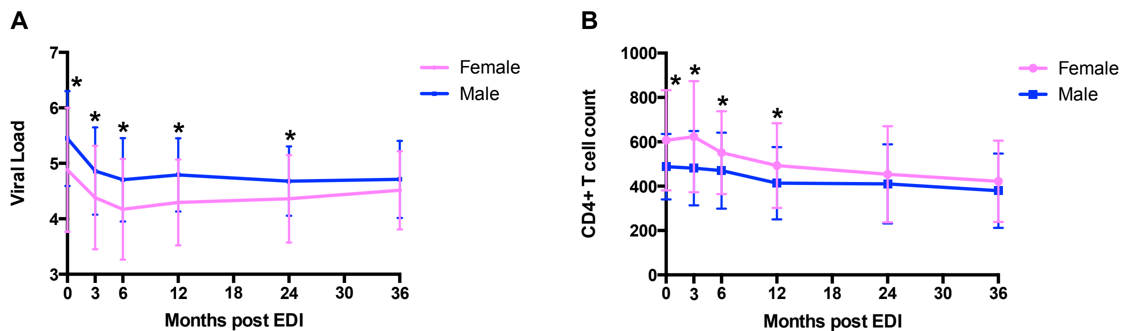


Figure 1. Females exhibit lower viral loads and higher CD4+ T cells counts in the first 2 years of HIV infection. Longitudinal plasma viral loads and CD4+ T cells counts were collected at 3-month intervals for 127 Zambian seroconvertors acutely infected with HIV-1 subtype C (55 females, 72 males). The time of initial sampling was a median of 46 days after the estimated date of infection (EDI). (**A**) Females (n=55) exhibit significantly lower plasma viral loads than males (n=72) up to 24 months post infection. (**B**) Females also have significantly higher CD4+ T cell counts up to 12 months post infection. This illustrates an early but transient benefit for females that appears to wane with time. (* $p < 0.05$; statistical comparisons were made using the the Student's t-test, p-values are two-tailed)

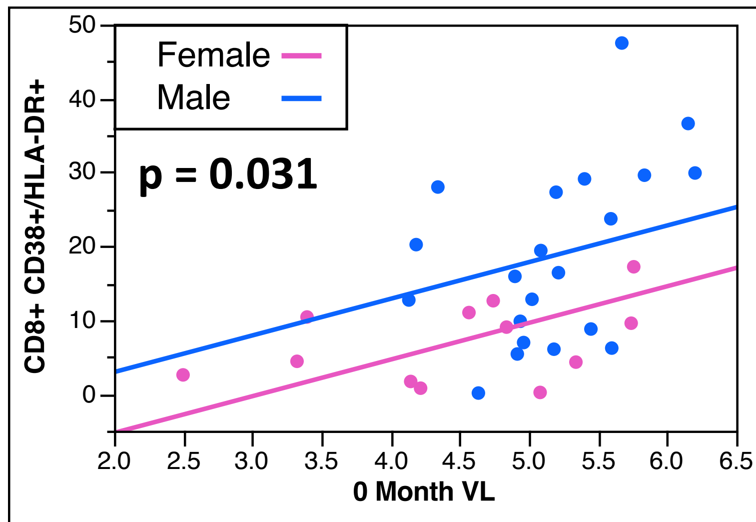


Figure 2. Females present with reduced cellular immune activation early in infection, even when controlling for the effects of plasma viral load. In 35 individuals (13 females, 22 males) cryopreserved peripheral blood mononuclear cells (PBMCs) isolated at the seroconversion time point (median of 46 days after EDI) were stained with antibodies specific for activations markers and analyzed via flow cytometry. In a multivariable generalized linear model, both viral load (p=0.03) at seroconversion (the time of PBMC isolation) and gender (p=0.004) significantly predicted the percentage of CD38+/HLA-DR+ CD8+ T cells.

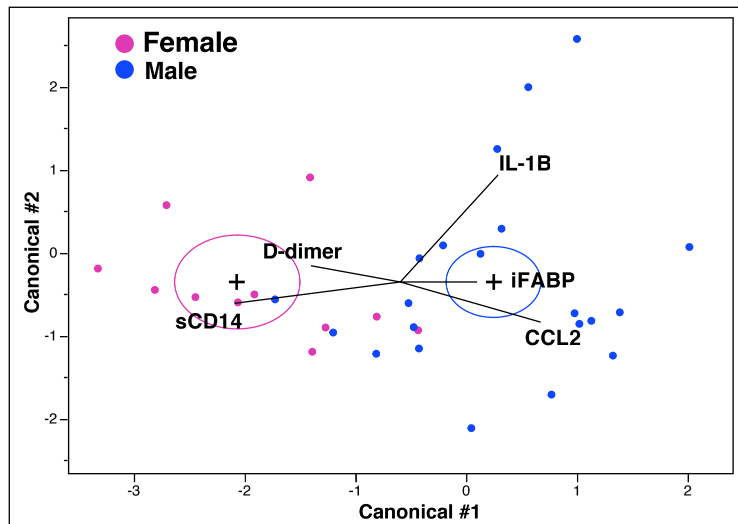


Figure 3. Acute HIV infection in females is characterized by a distinct inflammatory cytokine profile. The levels of 16 inflammatory cytokines, chemokines, and markers of gut damage and microbial translocation were evaluated in a subset of 33 individuals (12 females, 21 males) via multiplexed Luminex assay at the seroconversion time point (median of 46 days post EDI). Linear Discriminant Analysis (LDA) was use to generate distinct cytokine profiles capable of significantly differentiating between males and females (Pillai's trace = 0.57; p = 0.0002). Females are characterized by lower levels of IL-1β, MCP-1 (CCL2), and iFABP (intestinal fatty-acid binding protein) and higher levels of D-dimer and sCD14.

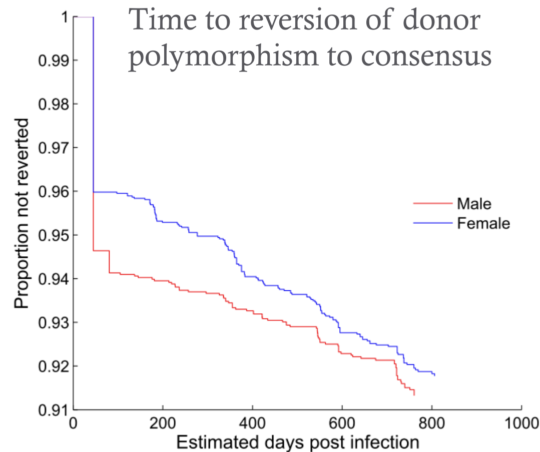


Figure 4. Transmitted viruses in females exhibit differential reversion kinetics in comparison to viruses transmitted to males. In a group of 88 Zambian seroconvertors, *gag*, *pol*, and *nef* genes were amplified and sequenced from the seroconversion time point and subsequent 3-months intervals up to 24 months post infection. The number of non-consensus amino acid residues was significantly greater in viruses transmitted to females. In a time course analysis measuring the rate at which transmitted polymorphisms reverted to consensus residues, females exhibit distinct reversion kinetics. Though the selection for consensus residues at transmission is less severe in females, these polymorphisms revert more quickly than those transmitted to males (p=0.03), implying these polymorphisms represent less fit viral variants. These observations may explain the early, but transient, clinical benefit seen in females that wanes with time.