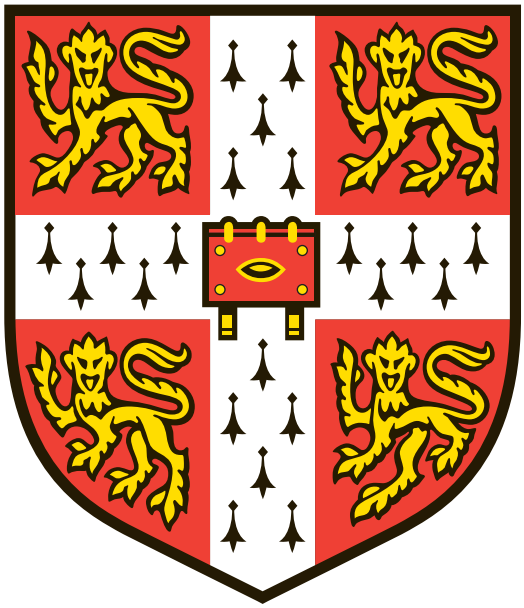
Simian Immunodeficiency Virus Infection of CD4+ T Cell Subsets and Its Impact on B Cell Responses in Early Infection

Eva Jacqueline Archer



University of Cambridge
Pembroke College

This dissertation is submitted for the degree of Doctor of Philosophy

July 2016

Abstract

Simian Immunodeficiency Virus Infection of CD4+ T Cell Subsets and Its Impact on B Cell Responses in Early Infection

Eva Jacqueline Archer

Human immunodeficiency virus (HIV) disease is marked by infection and loss of CD4+ T cells and by a gradual crippling of the immune system including the absence of an effective humoral response in a majority of individuals. The model system of simian immunodeficiency virus (SIV) infection of rhesus macaques offers an opportunity to study both the immunology of early infection and immune responses in secondary lymphoid organs. In this thesis, I examine SIV evolution and infection of CD4+ T cell subsets and its impact on the SIV-specific immunoglobulin response.

Follicular helper T cells (TFH), located in secondary lymphoid organs, are thought to be a major target cell for HIV/SIV infection. We showed that during acute infection, multiple subpopulations of memory CD4+ T cells are equally infected, but productively infected cells are largely restricted to follicular T cells in chronic infection. We measured SIV infection using DNA qPCR to quantify proviral infection, and RNA qRT-PCR to quantify spliced and unspliced cell-associated RNA. Next, we used RNA probes in conjunction with flow cytometric phenotyping panels to measure the frequency of individually infected cells in memory CD4+ T cell compartments. The viral protein nef down-regulates CD4 in the early phase of cellular infection, and we show that CD3+ T cells with dim CD4 staining have the highest copy numbers of SIV RNA. Finally, we performed phenotypic analysis of CD4+ T cell subsets in uninfected, acutely infected, and chronically infected macaques using flow cytometry panels of cytokine markers.

HIV/SIV infection of TFH may impact the development of the humoral response, as early antibodies to HIV/SIV are non- or weakly neutralizing. To characterize the development of the antibody response, we performed deep sequencing of immunoglobulin heavy chains from longitudinal blood, lymph node, and bone marrow biopsies from four to 24 weeks post infection. We used a trimeric SIV gp140 protein probe to isolate SIV-specific IgG and IgM memory B cells from total memory B cells. We found that SIV protein probe specific IgG B cells increase as a percentage of lymph node and PBMC B cells over the course of infection, while SIV-specific IgM B cells are detected at lower frequencies throughout infection. SIV-specific IgG heavy chain repertoires had low scores on measures of diversity in early infection, and showed accumulation of mutations by later infection relative to total memory B cells. Common SIV-specific sequences — immunoglobulins with identical CDR3 regions and VDJ gene assignments found in different animals — represented a small but significant percentage of the total SIV-specific response.

The lifespan and location of infected cells varies greatly, and only a fraction of cells with an integrated proviral genome go on to produce replication competent virus. We examined the compartmentalization of SIV by sequencing cell-associated RNA and proviral DNA from memory CD4+ T cell subsets. We compared tissue-resident sequences to circulating plasma virus, and measure the frequency of mutations across the genomes. Finally, we examined the phylogenies within individual animals to identify the spread and evolution of viral lineages chronologically and anatomically throughout infection.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text

It does not exceed the word limit of 60,000 words (excluding bibliography, figures, and appendixes) as prescribed by the Degree Committee for the Faculty of Biology at the University of Cambridge.

Dr. Kristin Boswell and Dr. Takuya Yamamoto assisted with biopsy and necropsy tissue collection. David Ambrozak performed all infectious cell sorting. All sequencing was performed either at the Wellcome Trust Sanger Institute sequencing core facility or at the Vaccine Research Center Genome Analysis Core. SIV sequencing was primers were designed with help from Dr. Shelby O'Connor at the University of Wisconsin. Dr. Swee Hoe Ong performed SIV sequence assembly.

Eva Jacqueline Archer

July 2016

Acknowledgements

First and foremost, thanks to my supervisors: Dr. Richard Koup, Dr. Daniel Douek and Prof. Paul Kellam for agreeing to take on a student they would only see 50% of the time, and for your support, advice and encouragement over the last four years. I am very grateful for the opportunity to do this work and for your mentorship. Thanks to all the members of the labs in both places, for training and helping me, and for endless discussion and feedback, especially Kristin Boswell, Sara Ferrando-Martinez, Takuya Yamamoto, David Ambrozak, and Joe Casazza in the Koup Lab; Sam Darko, Amy Ransier, Brenna Hill and Eli Boritz in the Douek Lab; and Rachael Bashford-Rogers, Swee Hoe Ong, Simon Watson, and Astrid Gall in the Kellam Lab. Thanks also to Rebecca Lynch, Jason Hataye, Adam Wheatley, Pratip Chattopadhyay, Rosie Mason, and to the animal facility (especially JP Todd) at the VRC, and the sequencing cores at the Sanger Institute. Thanks to the NIH Oxford Cambridge Scholars Program, especially Katie Soucy and Liz McIntyre, and to Annabel and Christina at the Sanger Institute for keeping track of me as I came back and forth. Thanks also to my college, Pembroke, for support for travel to conferences.

When not in the lab (!) I am grateful for the friendships, both new and old, that have enriched my time on both sides of the Atlantic. Thanks to all of the UNC group in DC (Katherine Novinski, Emma Din, Genevieve Kelly, Hogan Medlin, Clayton Thomas, Brad Waters, and many others); fellow OxCammers Monica Kasbekar, Geoff Lynn, Ian Goldlust and Bennett Waxse; the VRC sorority; and in Cambridge Pinky Langat, Neneh Sallah, James Hadfield, Beverly McCann, and James Hamp.

My biggest thanks go to my family, whose love and support I have relied on from day one. Dad, you are the reason I wanted to become a scientist, and I would not have made it this far without your advice and encouragement. Mom, your unflagging support and belief in me has helped me through highs and lows. Isabella, Helena, and Norman, you are the best siblings and the best friends anyone could ask for- thanks for listening to me complain, celebrate, and everything in between.

Eva Jacqueline Archer

July 2016

Table of Contents

1 Introduction

Since human immunodeficiency virus (HIV) was discovered as the causative agent of AIDS, (acquired immunodeficiency syndrome) in 1983[1,2], understanding the pathogenesis of the virus has been the focus of countless research groups around the world. The discovery of antiretroviral drugs and the success of public health based interventions such as needle exchanges, treatment as prevention, and pre-exposure prophylaxis have reduced morbidity, mortality, and the number of new infections[3], but both an HIV cure and a vaccine have proved elusive. In spite of this, over 30 years of research have provided insights and breakthroughs in areas far beyond HIV, in understanding the details of the human immune system in the presence of a persistence viral infection and inflammation.

1.1 Introduction to HIV Virology

1.1.1 HIV subtypes and geographic distribution

HIV, human immunodeficiency virus, is a retrovirus of the family *retroviridae* and genus *lentivirus*. The primate lentiviruses are generally divided into the viruses that infect humans (HIVs) and over forty non-human primate tropic viruses (simian immunodeficiency virus, SIV)[4]. HIV has two subtypes, HIV-1 and HIV-2. HIV-1 is responsible for the majority of infections and AIDS, whereas HIV-2 is predominantly found in West Africa, is more closely related to the SIVs, and is generally less pathogenic than HIV-1 [5]. There is evidence from historical sampling and phylogenetic inference that HIV-1 has passed to humans from chimpanzees (*pan troglodytes*) and HIV-2 from sooty mangabeys (*cercocebus atys)* at least three times independently, as the M (main), N (non-M, non-O), and O (outlier) groups [5–7]. As its name suggests, the majority of infections are M type viruses, with an estimated 90% of all known infections [8]. Within the M group, viruses are further divided into clades (A-K) that vary in their frequency in different parts of the world. B is the most common clade in North America and Western Europe, where A and C are dominant in sub-Saharan Africa and A in the former Soviet Union (Figure 1.1)[9]. The major difference between groups and clades (also known as subtypes) is in the envelope gene, where groups differ by about 30% and clades by 15-20% [10,11]. However, accurate sub-typing can be assigned using either *gag*, *env*, or *pol* genes[12]. Estimates of the age of the HIV common ancestor date as far back as 1915-1941, with the M group common ancestor originating from the Kinshasa region of the Democratic Republic of the Congo [13].

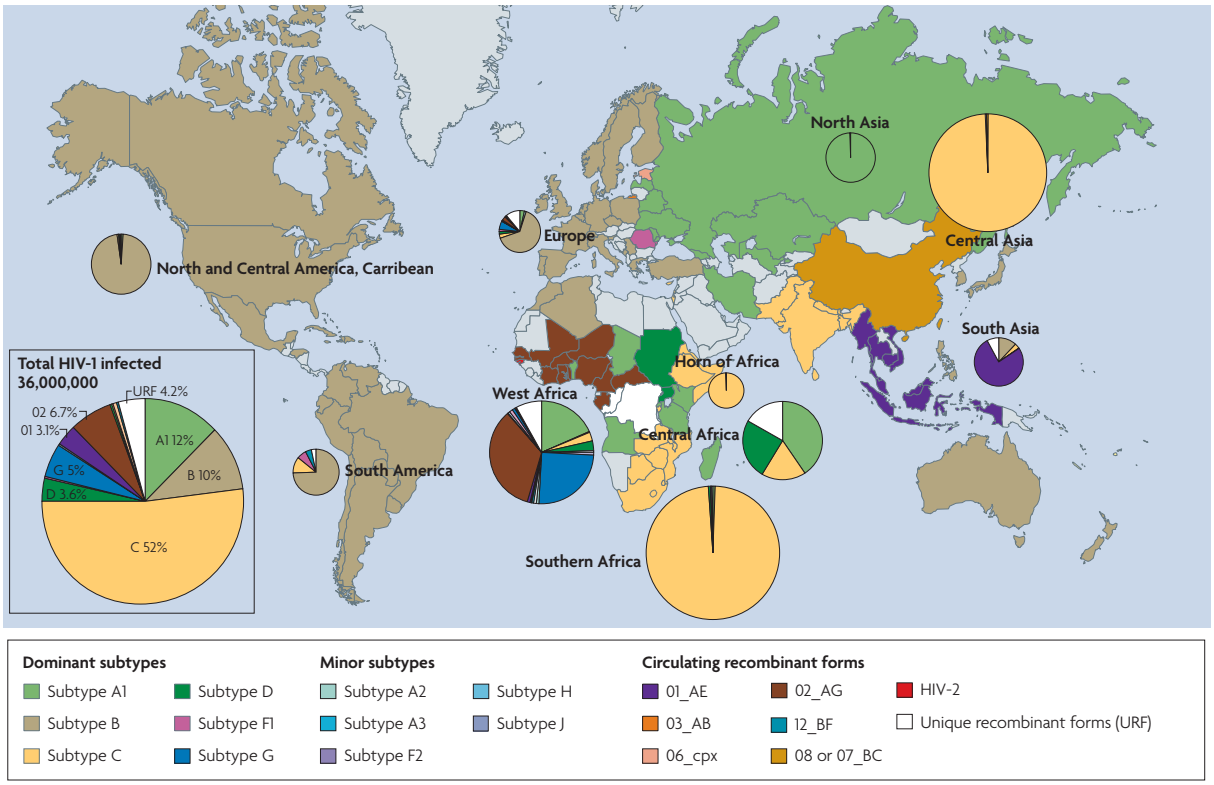


Figure 1.1. Global prevalence and distribution of HIV-1 subtypes. From Ariën, Vanham, and Arts 2007 Nature Reviews Microbiology[14]. Subtype C is the most common worldwide and is largely found in southern and eastern Africa and in the Indian subcontinent. The majority of infections in the Americas and Western Europe are subtype B. North Asia and parts of central Africa have a majority of subtype A infections, while West Africa contains a mix of A1, G and recombinant AG

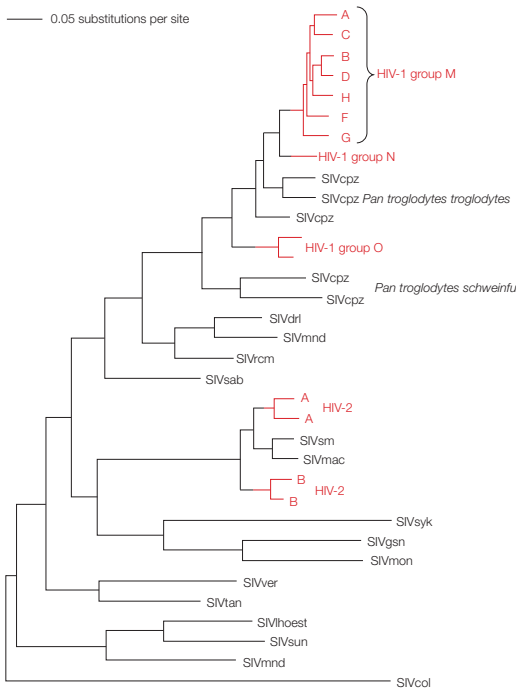


Fig 1.2. Phylogeny and ancestry of primate lentiviruses. From Pybus and Rambaut, 2009 Nature Reviews Genetics[4] The two HIV lineages (HIV-1 and HIV-2) fall between SIV branches and represent independent cross-species transmission events rather than a single transmission event. HIV-2 is closest to SIV found in sooty mangabeys and that is pathogenic in macaques (SIVsm and SIVmac), while HIV-1 is most closely related to chimpanzee-derived SIV (SIVcpz).

1.1.2 Virus features and lifecycle

In HIV/SIV, two identical copies of the positive-sense, single-stranded RNA genome are packaged in an enveloped viral capsid approximately 145 nm in diameter[15]. The viral genome is 9.7 kb in length and contains three structural genes- Gag, Pol, and Env - and six accessory genes - Tat, Rev, Vif, Vpr, Vpu, and Nef [15]. In HIV-2 and some SIVs, Vpx replaces Vpr. Major structural elements include the LTR, long terminal repeats on the 5' and 3' ends of the genome; and RRE, the rev response element that is critical for Rev function.

HIV binds to lymphocytes expressing CD4 and a co-receptor, typically CCR5 or CXCR4[16,1718–21]. CD4 is primarily expressed on helper T cells, but is also found on dendritic cells, macrophages, and a subset of neural cells. Viruses are sub-classified based on their CCR5/CXCR4 tropism as R5 or X4 viruses respectively [22]. Most viruses are CCR5-tropic, and it is typically R5 viruses that are transmitted. However, over the course of infection, the viral population within a single host typically diversifies to include X4-tropic viruses [18,22]. The use of seven transmembrane protein G-protein-coupled receptors (CCR5 and CXCR5) means that beta chemokines (such as MIP-1α, MIP-1β, and RANTES) that block these receptors can inhibit HIV infection *in vitro* [18,20]. In rhesus macaques and in humans, X4 viruses are associated with rapid and irreversible loss of CD4 T cells in the periphery [3,23]. In clinical cases, it has been shown that the shift within the host from R5 to X4 tropism marks the onset of AIDS and the rapid decrease in CD4 T cells [24].

Table 1.1 – HIV/SIV genes [15]

Gene name	Function
Gag	Structural proteins: nucleocapsid (NC/p7), core capsid (CA/p24), matrix protein (MA/p17)
Pol	Encodes enzymes integrase, reverse transcriptase, protease, and RNAse H
Env	gp120 and gp41 external glycoproteins that bind CD4 and coreceptors (CCR5 & CXCR5)
Tat	Trans-activator of transcription, increases transcription of HIV double stranded DNA
Rev	Activates nuclear export of unspliced viral RNA
Vif	Inhibits APOBEC activity to promote infectivity
Vpr	Immunosuppressant, regulates pre-integration complex and localization to nucleus.
Vpu	Release of virions, degradation of CD4 in the endoplasmic reticulum
Nef	Downregulates cell surface receptors to promote infectivity and survival
Vpx	HIV-2/SIV homolog of Vpr. Not in HIV-1

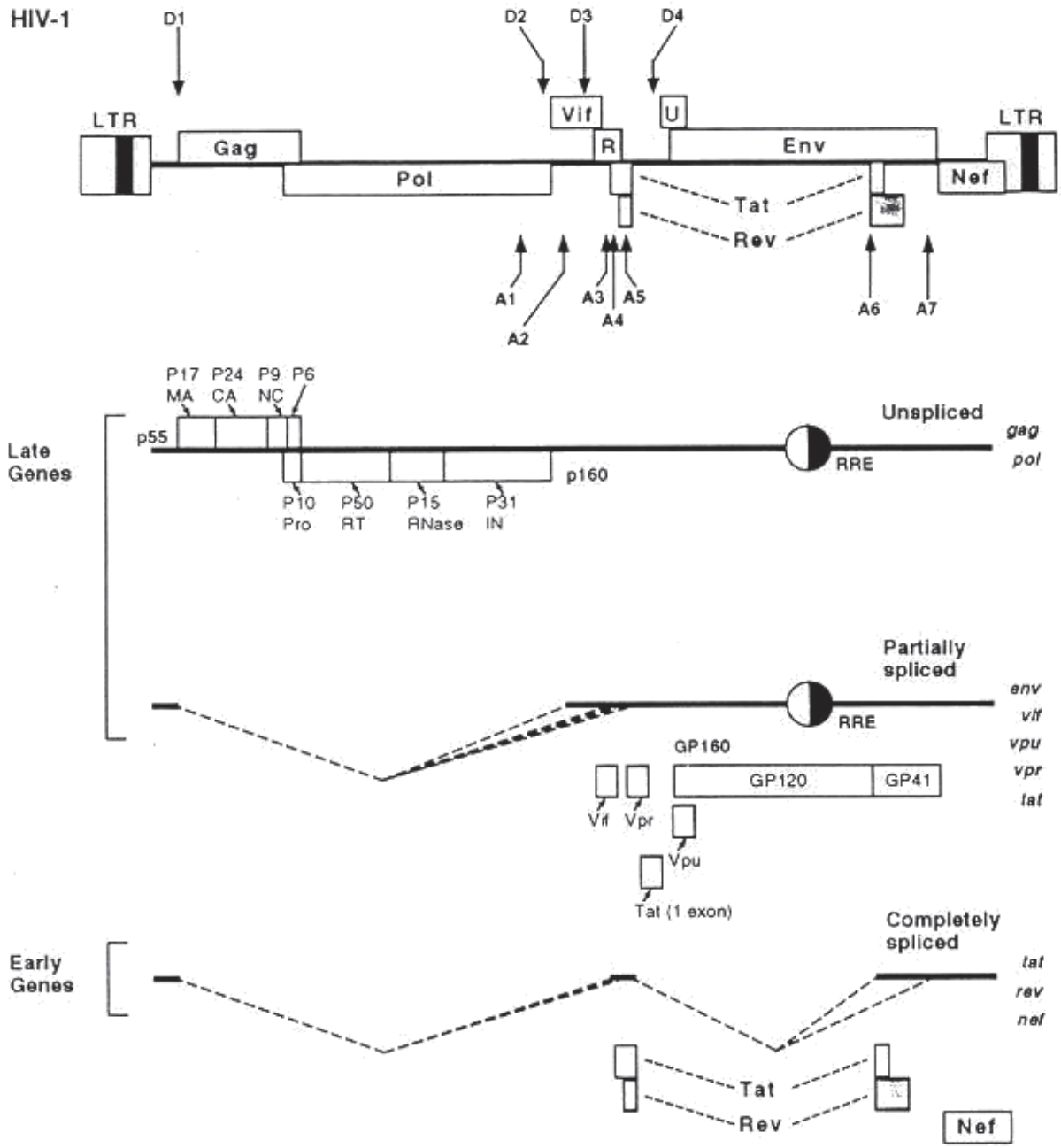


Fig 1.3 [25] **HIV/SIV genome organisation and splicing.** From http://www2.gsu.edu/~biotkf/bio475/475lecture8.htm . The HIV/SIV genome contains nine genes: three structural genes and six accessory genes. The 2kb RNAs are multiply spliced and are produced first in viral infection (Tat, rev, nef), followed by singly spliced 4kb RNAs are produced next (Vif, Vpr/Vpx, Env, Vpu) and finally by full-length genomic transcripts.

1.1.3 HIV Transmission and cellular replication

Most infections are caused by a single virus species (called the transmitted/founder virus) rather than by the swarm of related viruses in the infecting host [26,27]. This bottleneck of virus diversity at transmission is in contrast to the massive viral diversity present in the blood of infected individuals in chronic infection [26]. The initial virus undergoes multiple successive rounds of infection, replication, and expansion in host CD4 T cells at the site of infection and in dendritic cells and CD4 T cells that migrate to lymph nodes [3]. During this phase, plasma

viremia increases rapidly, to approximately 10^6 copies/mL at 3-4 weeks post infection in humans and 2-3 weeks post infection in SIV infection of rhesus macaques[28]. CD4 T cells are the primary target of HIV/SIV, but dendritic cells, macrophages, and cells in the central nervous system can also be infected[21,29–32]. The trafficking of infected cells to lymph nodes where there is an abundance new targets for infection is responsible for the rapid increase in plasma viremia and systematic infection and loss of CD4 T cells, which reach a nadir around the time of peak plasma viremia[33–35]. Up to 60% of memory CD4 T cells in all tissues and compartments are infected in the acute phase, with the vast majority of infected cell dying within days[36].

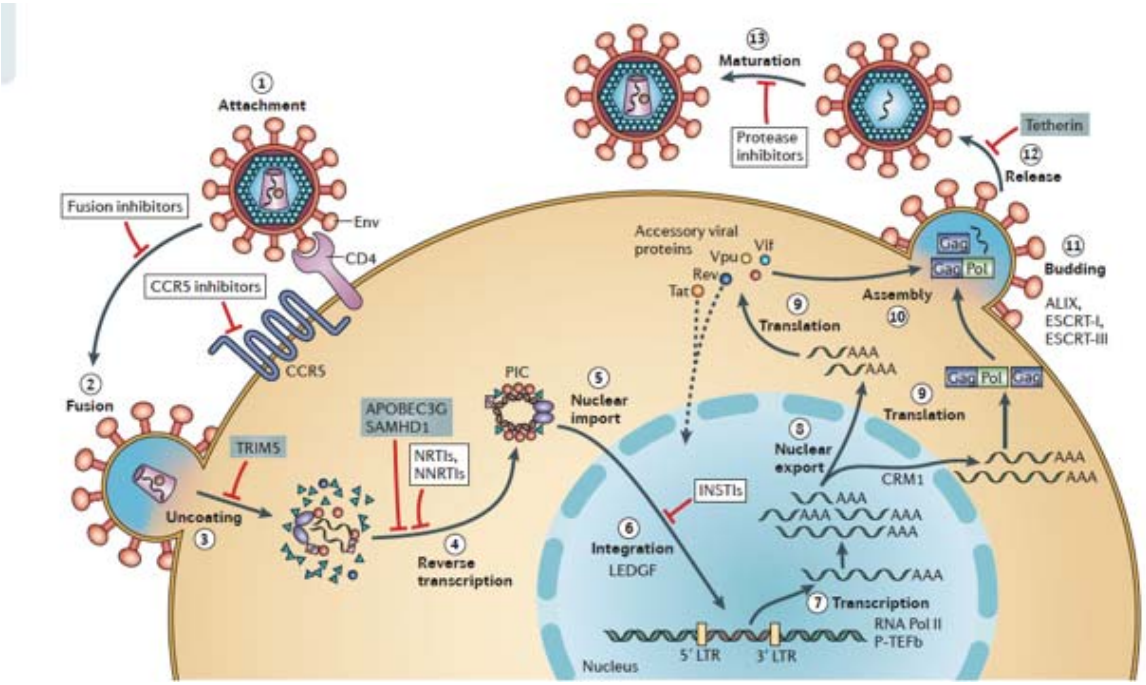


Fig 1.5 HIV/SIV Infection and lifecycle. From Engelmen and Cherepanov, 2012 Nature Reviews Microbiology[37]. HIV binds to the cell surface via CD4 and a co-receptor (CCR5 or CXCR4), and undergoes membrane fusion and release of the viral capsid into the cytoplasm. After capsid uncoating and reverse transcription of the viral RNA genome, the newly formed proviral DNA genome translocates to the nucleus where it is incorporated into the cellular genome. Upon activation, viral genes are transcribed and viral proteins are synthesized, and newly formed virions bud from the cell surface. The entire replication cycle takes place in around 48 hours.

Infection at the cellular level begins when HIV attaches to the cell surface and binds CD4 and a coreceptor, either CCR5 or CXCR4[16,17,19–21]. HIV is endocytosed via a series of conformational changes in the envelope trimer proteins, with gp41 and gp120 re-orientating to allow fusion with the cellular membrane[38,39]. Reverse transcription of viral RNA is the canonical feature of retroviruses and occurs shortly after virus entry, as the virus transits the cytoplasm en route to the nuclear pore. Once the capsid has localized to the nuclear pore, the encapsulated RNA and accessory proteins are released after the internal capsid is uncoated and drives the newly synthesized proviral DNA into the nucleus[40]. The viral pre-integration

complex migrates to the nucleus via nuclear localization signals on viral integrase and matrix proteins[41–43]. The linear proviral DNA is incorporated into the cellular genome after it enters the nucleus and requires viral protein integrase, and can be integrated into a variety of sites in the cellular genome. Subsequent expression levels of viral RNA are affected by integration site [44], and actively transcribed genes are preferential sites for HIV genome integration[45]. The higher error rate of the reverse transcriptase (compared with DNA polymerase) is responsible for much of the genetic variation that is observed, as well as the rapid evolution of HIV in response to the host response [46,47]. As a diploid virus, recombination of viral genomes when a cell is infected with multiple viruses also plays a role in expanding the genetic diversity of newly synthesized viruses.

Different viral RNAs are produced at different stages in the viral lifecycle. Early on, multiply spliced (2kb) RNAs such as *rev*, *tat*, and *nef* are produced in large quantities. Tat protein binds a 5' RNA loop structure on viral RNAs to promote transcription elongation, while Rev is required for the subsequent production of HIV structural proteins[48]. Rev protein binds a target sequence of 233 bases called the Rev Response Element (RRE), and transcripts with this sequence (most of the singly spliced (4 kb) and unspliced full length (10kb) RNAs) can be exported from the nucleus upon binding of Rev[49,50].

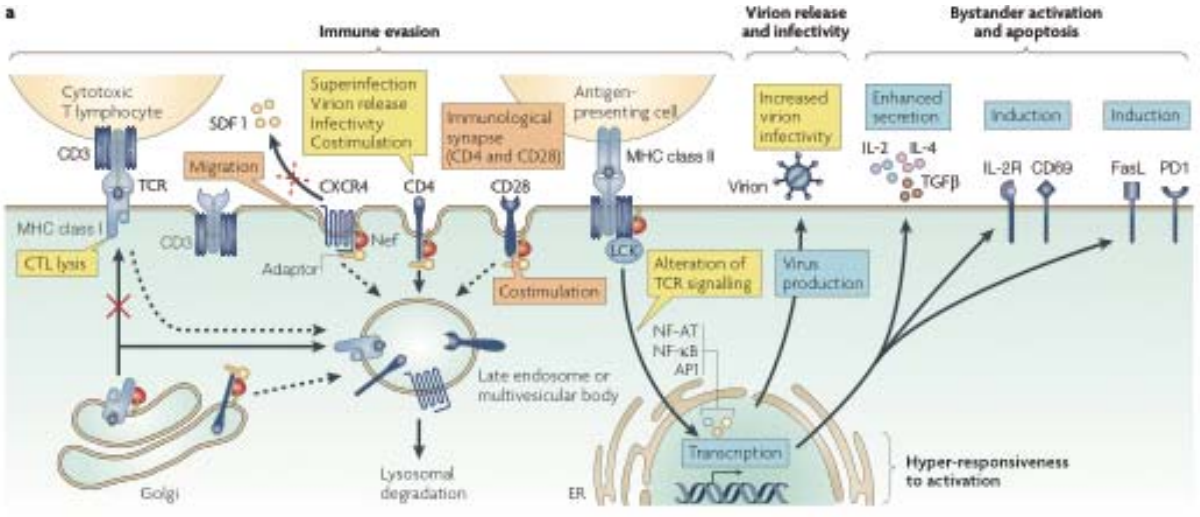


Fig 1.4 Immune evasion by nef. From Kirchoff, 2009 Nature Reviews Microbiology[51]. Nef downregulates CD4, CD28, MHC-1, MHC-II and in the case of SIV, CD3 from the cell surface. This allows the infected cell to avoid adaptive immune responses from CTL and antigen-presenting cells. Nef also promotes virus production and release of virions, and increases secretion of chemokines including IL-2, IL-4, and TGFβ.

Virus protein Nef in particular has many functions that aid HIV/SIV replication and immune evasion. One of the most important is the down-regulation of cell surface receptors, most prominently CD4 and CD8, and MHC I and II [52]. CD1, CD28, CD80, and CD86 are also

down-regulated in HIV infection [52]. Nef is produced early in the viral lifecycle, as its multiply spliced mRNA transcript is translated before the rev response element promotes the export of singly spliced and unspliced HIV RNAs [53,54]. Nef binds CD4 and clathrin adaptor AP-2 to promote endocytosis of CD4 [52,55,56]. It has been reported that Nef can also utilize the multivesicular body pathway, though the exact pathway by which Nef decreases CD4 is not clear [52]. Down-regulation of MHC-I by Nef prevents infected cells from being recognized by CTL, although they remain susceptible to natural killer (NK) cells[57,58]. CD28 is critical to activation of T cells via engagement with B7, and down-regulation of CD28 further impairs the ability of CD4 T cells to respond to infection[59]. Both SIV and HIV-2 Nef have been shown to down-regulate CD3 - this has not been show in HIV-1 Nef[60]. Thus, Nef dampens the immune response by reducing or eliminating infected cells from engaging in the immune response. The importance of nef is further demonstrated by nef-deletion mutant viruses, which show reduced pathogenicity and lower viral loads[61–64]. Env and Vpu (Vpx in SIVmac) are also capable of down-modulating CD4, by restricting newly translated CD4 molecules to the endoplasmic reticulum (Env) or linking gp160-CD4 complexes to the ubiquitin ligase degradation pathway (Vpu)[65,66].

In the late phase of HIV replication, structural viral genes are expressed once the export of longer RNAs is permitted by Rev binding the Rev response element contained in their transcripts[53]. Late-stage viral RNAs are translated to produce Gag polyprotein precursors, regulatory and accessory proteins, and viral envelope glycoproteins. Gag, GagPol, and Env glycoproteins are trafficked to the plasma membrane, where they are assembled [40]. Two copies of the viral RNA genome are packaged into the nucleocapsid, and the newly formed virion buds from the surface of the cell. In HIV infection of activated CD4+ T cells, thousands of copies of HIV are generated from a single cell. A complete lifecycle, from initial binding to new virus release into the cytoplasm takes approximately 48 hours.

The Gag precursor protein is 55 kDa and contains structural proteins: matrix, capsid, nucleocapsid, and p6 domains in addition to two spacer peptides[15]. Gag precursor picks up viral genomic RNA in the cytoplasm (via the nucleocapsid domain) and multimerizes as the matrix domain targets the polyprotein to the plasma membrane. It anchors to the plasma membrane via the myristylated amino terminus and the capsid begins to assemble into virions. GagPol precursor protein is produced at the same time as the result of a ribosomal frameshift, contains viral enzymes (protease, reverse transcriptase, and integrase) and expressed at 20x lower concentrations than Gag [67].

Envelope proteins are synthesized in the rough endoplasmic reticulum (RER) after the gp160 precursor protein targets the newly formed protein to the RER. The transmembrane domain keeps the envelope securely in the lipid membrane, with the extracellular components of Env in the RER lumen and the cytoplasmic tail (gp41 domain) in the cytoplasm [68]. The gp160 domain is glycosylated and monomers oligomerize into trimers, and the envelope trimers subsequently migrate to the Golgi complex, where they are cleaved into gp120 and gp41 [68]. The trimer is subsequently trafficked to the plasma membrane, where it associated with the lipid raft microdomains that contain the Gag proteins and is incorporated into virion assembly[67]. After virions are released, viral protease that has been packaged into the virion cleaves the Gag precursor protein into its four components, and the morphology of the virion changes into the mature virus[40].

1.2 Pathogenesis of HIV infection

While most cells are infected when they are in an activated state[45], there is evidence that HIV can be incorporated into resting and naïve cells, albeit at much lower levels than in activated T cells [69]. The establishment of the latent reservoir requires that infected cells are in or return to a resting state, so that the T cell can persist undetected by any CTL or antibodies specific to HIV [70]. Latently infected cells are also immune to current forms of antiretroviral therapy, as existing anti-HIV drugs rely on blocking viral fusion or interaction with co-receptors, integration into the host DNA, reverse transcription of viral RNA, or protease inhibitors - none of which are present in latently infected cells [71]. The virus can be reactivated at a later time and will be vulnerable to anti-HIV defences, but re-activating latent cells has proved to be difficult [70]. Estimates of the latent reservoir differ based on the method used to detect and quantify latently infected cells, but a common estimate is one in a million memory T cells[72]. The size and persistence of the reservoir is complicated by the unknowns surrounding the homeostatic proliferation (contributing to the maintenance) and the rate of decay of this population by non-HIV specific cell death and normal cell turnover [73–76]. Central memory and transitional memory are the major reservoirs of HIV/SIV[77], which persists by two mechanisms. In central memory cells, low-level antigen-driven proliferation maintains the infected reservoir, which is slowly depleted. In transitional memory cells, homeostatic proliferation driven by IL-7 maintains the reservoir [77]. There is some evidence that low-level *de novo* infection of CD4 T cells replenishes the latent reservoir even in the presence of highly active anti-retroviral therapy, particularly in lymphoid tissues[78]. Though undetectable in blood, this study cites on-going evolution of virus and trafficking between lymph nodes.

However, other studies have refuted this, finding no evidence of evolution of viral genomes within the reservoir[79,80].

HIV is most commonly spread via sexual transmission, with rates of infection differing between penile-vaginal, penile-anal, and oral sex[81]. The viral load of the transmitting partner plays a major role in infection rates, with higher rates around seroconversion and in late-stage infection[82]. The presence of other STIs increases the risk of transmission, most notably with herpes simplex virus, syphilis, and human papilloma virus[83]. Male circumcision lowers infection rates by providing some protection from infection compared to uncircumcised men[84]. Other transmission routes include via contaminated needles used for intravenous drug injections, and in the 1980s there was transmission through HIV-infected blood donation and platelet infusion for haemophiliacs[85]. Mother to child transmission (vertical transmission) can occur during pregnancy, during labour and delivery, or during breastfeeding[85–88]. HIV can be transmitted either as free virus in body fluids or as cell-bound virus[389]. Transmission across mucosal membranes can be restricted by innate immune responses, and it typically takes multiple exposures for transmission to occur[89,90]. Estimates of infection frequency vary based on the viral load of the infecting partner and any co-infections of the newly infected host, the type of sexual contact, and the presence of certain STIs and genital ulceration[90].

1.2.1 Acute HIV infection

Acute HIV infection is divided into five Fiebig stages, based on how HIV can be detected clinically in the blood – by viral PCR (Stage I), p24 ELISA (Stage II), HIV specific antibody ELISA (Stage III), and HIV specific antibody Western blots (indeterminate -Stage IV and clearly positive – Stage V)[91,92]. The eclipse phase is from transmission and undetectable virus to Fiebig Stage I[28]. The early chronic phase of infection is classified as Fiebig stage VI, when Western blot and p31 staining are both positive[91]. At this point, viral load has decreased from a peak of up to 10^7 copies/mL in blood to approximately 3-5 x 10^5 copies /mL[28]. This viral set point remains steady throughout infection until the onset of AIDS. Lower viral set-points are associated with more favourable (delayed and less severe) disease progression[93]. In early SIV infection, memory CD4 T cells in the lamina propria that are highly activated (defined as CD69+ CD38+ HLD-DR+) are depleted rapidly[94]. CD4 T cells in the intestine show higher expression of CCR5 compared to spleen, blood, or lymph nodes, and are eliminated within two weeks of SIV infection in macaques[95]. The central memory and naïve CD4 T cells are not as drastically depleted as effector memory cells[33,96].

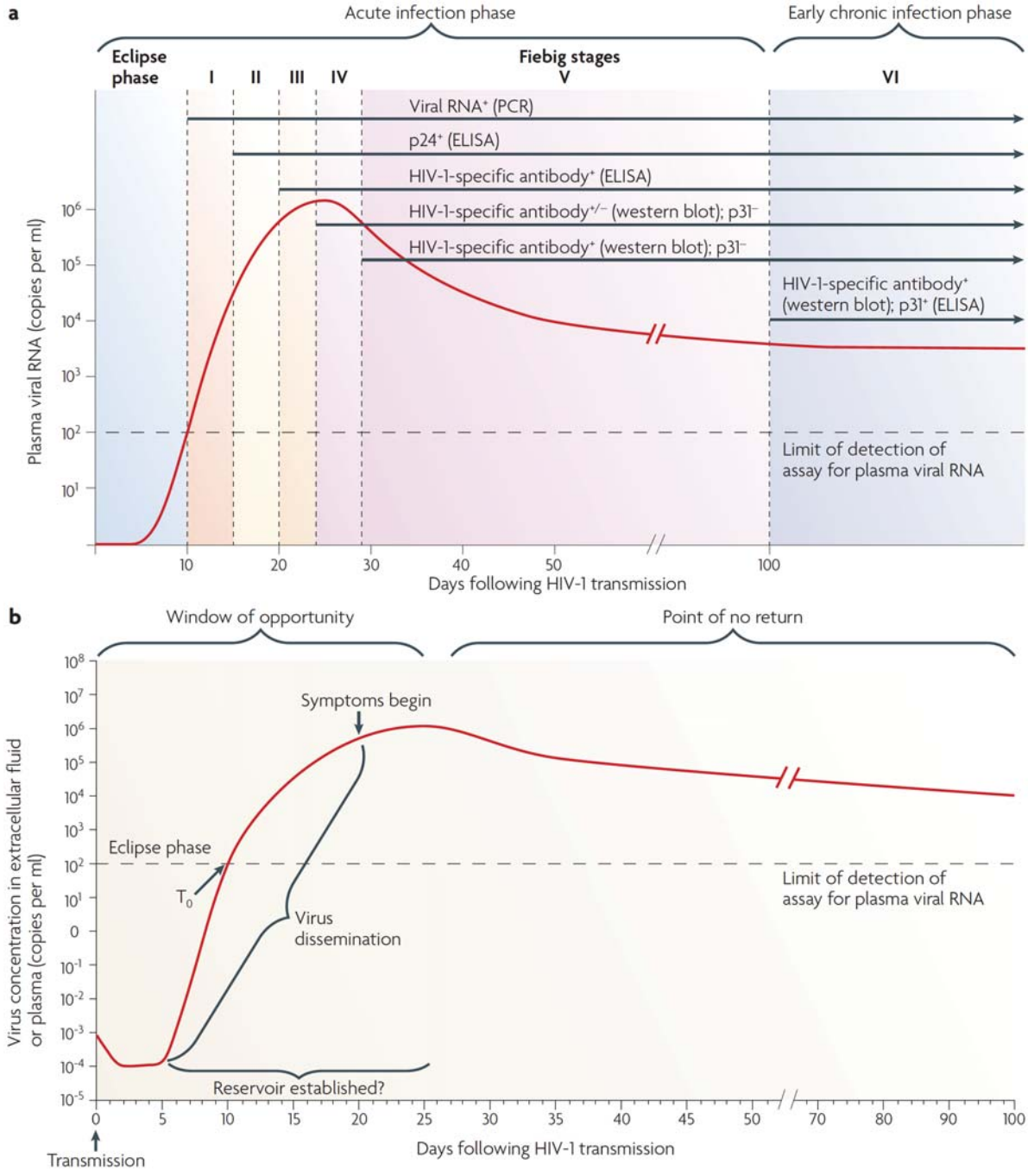


Fig 1.6. Stages of HIV infection. From McMichael, Borrow, Tomaras, Goonetilleke, and Haynes, 2010 Nature Reviews Immunology[28]**.** Fiebig stages of infection are determined by the type of test required to diagnose infection (PCR, ELISA, Western blot). The eclipse phase of infection refers to the period after transmission but before infection can be detected. The reservoir is established within the first few weeks of infection.

1.2.2 Chronic HIV infection

Within six months of infection, the host immune response has regained control of the infection and virus levels are reduced to a steady level in plasma – typically below $2x10^4$ copies/mL[3]. Most individuals do not have any symptoms during this period, although there is an increase in circulating pro-inflammatory cytokines and in activation of T cells, and

increased turnover of NK, B and T cells[23,97]. CD4 T cells gradually decline over the course of infection, and in the absence of antiretroviral therapy, there is on-going viral replication and evolution[98]. The immune response is active during this time, with on-going evolution of the humoral response in response to changes in viral envelope proteins[98,99]. Immune activation is driven at least in part by on-going HIV replication, as virus control under ART reduced immune activation in studies of HIV-infected individuals[100,101].

Approximately 5% of infected individuals will not develop any signs of immunodeficiency (called long term non progressors)[102]. They are typically grouped into two categories: those that spontaneously control virus to low or undetectable levels in blood, called elite controllers; and those who maintain plasma viremia but do not progress to AIDS-like symptoms or succumb to opportunistic infections[103,104]. Certain HLA types are strongly associated with delayed disease progression and viral control, most notably HLA-B*27 and HLA-B*57[3,105,106]. Elite controllers, long-term non progressors, and viremic controllers have been the subject of intense study as they represent outcomes that a therapeutic vaccine or functional cure could achieve in absence of sterilizing immunity – control of virus and limited inflammation and damage to the immune system.

By ten years post infection, approximately 50% of infected individuals will develop signs of immunodeficiency, most significantly a drop in blood CD4 T cell counts (below 200 cells/mm^3) and the emergence of opportunistic infections including Kaposi's sarcoma, *Pneumocystis carinii* pneumonia, and candidiasis[107,108]. At this point significant damage has been done to lymphatic tissues, with extensive fibrosis observed and loss of healthy lymph node architecture[109]. Analysis of lymph nodes of chronically HIV-infected patients show increased immune activation, loss of architecture in lymph node structures including B cell follicles and germinal centres, and deposition of collagen resulting in extensive fibrosis [110,111]. Bone marrow also shows signs of dysfunction, from anaemia, thrombocytopenia, and neutropenia in chronic AIDS cases [5,112].

1.2.3 Virus Evolution and Phylogenetics

HIV/SIV has high genetic diversity, due in part to the low fidelity of HIV reverse transcriptase, which has an error rate of approximately 1/2000-1/4000 (equating to roughly 5-10 mutations in a new synthesized HIV genome)[46]. This coupled with a high rate of virus production in a viremic host (on the order of 10^9 virions per day[113]) and complete replication cycle in roughly 48 hours means that HIV can effectively escape host immune responses, and is engaged in a constant arms race with CTL and humoral responses. Within a given viral

population, patterns of genetic diversity vary across the regions of the genome, with highly conserved regions and hotspots of great diversity[47]. The greatest variability is found in the 3 kb envelope gene, which is targeted by both T and B cell host defences and must constantly adapt to the pressures from the host adaptive immune response[114,115]. HIV also has an unusually high recombination rate of 2-3 times per replication cycle that contributes to the genetic diversity[116]. Viral recombination occurs when a cell is infected with two different proviruses and during subsequent transcription and viral packaging, one copy of each unique genome is packaged into a virion instead of two identical copies of the same genome[116].

1.2.4 The innate immune response to HIV

The innate immune system is the first line of defence against systemic HIV/SIV infection. Natural killer (NK) cells, natural killer T (NKT) cells, and dendritic cells (DC) all play a role in preventing HIV infection at mucosal surfaces and in other forms of infection[117]. The innate immune response also plays an important role in the persistent immune activation that drives progression to AIDS and chronic inflammation [118].

HIV, like many other viral infections, triggers the production of type I interferons (IFNα and IFNβ) by leukocytes (particularly plasmacytoid dendritic cells, pDCs) and fibroblasts[66,119]. Plasmacytoid dendritic cells respond to double stranded viral DNA via Toll-like receptor 9 and single stranded RNA through Toll-like receptor 7/8 by producing IFNα, which binds to interferon receptors and initiates a cascade of signalling and downstream effects[28,66,120]. The interferons have been shown to be a double edged sword in HIV infection, with both protective and destructive qualities[119]. In chronic infection, interferon and the resulting expression of interferon stimulated genes (ISG)s have been linked to more rapid disease progression[66]. In SIV infection of macaques, administration of IFNα2a before exposure was protective from infection, and treatment with an interferon antagonist caused rapid infection with high viremia and major CD4 T cell depletion[119]. Long-term exposure of infected animals to IFNα increased viral reservoir size and loss of CD4 T cells[119]. Furthermore, in natural hosts of SIV such as sooty mangabeys, show greatly reduced levels of innate immune activation and lower production of IFNα via TLR 7 and 9 stimulation[120].

Beyond interferons, the driving factor behind chronic immune activation in HIV infection is translocation of microbial products from the gut into general circulation. Lipopolysaccharide (LPS) and bacterial 16SRNA are among the many gut-derived circulating factors that bind pattern recognition receptors and trigger generalized immune activation. Markers of immune activation and of the response to microbial translocation, such as soluble CD14, predict

disease progression[121]. Even in the presence of suppressive ART, immune activation persists[101]. Acute infection is marked by a cytokine storm: with levels of serum amyloid A, IL-1, IL-15, CXCL10, IL-18, TNF, and IL-22 increasing within 2-3 weeks of infection[28].

Interferons are among the cellular factors that recruit NK cells, which have both direct and antibody-mediated cytotoxic capacities. They are capable of killing HIV-infected cells, and secrete large amounts of chemokines that bind the HIV co-receptor CCR5 (MIP-1α, MIP-1β, and RANTES) which inadvertently competes to reduce infection of CD4 T cells[118]. Disease progression and control of SIV infection in macaques has been linked to NK effector cell function in cooperation with antigen-specific CD4 T cells[122]. NK cell subsets are skewed in HIV infection, and dysfunctional CD16+CD56- NK cells accumulate[123]. Long-lived antigen-specific NK cells that are capable of killing autologous dendritic cells have been found in SIV and SHIV-infected macaques[124].

Restriction factors also play a role in the host defence from HIV infection, with APOBEC3G, TRIM5α, and tetherin as the most prominent targets of investigation. The APOBEC3 (apolipoprotein B messenger RNA-editing enzyme catalytic polypeptide-like 3) family of proteins, with APOBEC3G in particular, are a form of innate host antiviral defences that are expressed in T lymphocytes, and inhibit HIV infection by interfering with reverse transcription of HIV unless viral protein Vif is present[125]. APOBECs edit newly transcribed HIV genomes and mutate cytidine residues to uridines on the negative strand of HIV DNA, rendering the HIV genome incapable of replication[126]. Tetherin (specifically CD317) retains newly formed viral particles on the cell surface, preventing their release and is targeted by viral protein Vpu[127]. TRIM5α is located in the cytoplasm and blocks reverse transcription[128]. Human TRIM5α is less effective than rhesus TRIM5α in blocking HIV-1, but rhesus TRIM5α is less effective against SIV[128,129].

1.3 The adaptive immune response to HIV

In early infection, cytotoxic T cells (CTL) control viremia and reduce infection, and help restore CD4 T cells[28]. Humoral immunity also plays a major role in controlling virus, with gp41 non-neutralizing IgM antibodies first detected around 3-4 weeks post infection, and autologous virus-specific neutralizing antibodies appearing at about 3 months after infection[130,131]. Broad and potent neutralization of virus occurs only in a fraction of individuals but does not keep up with virus escape[98,132,133].

1.3.1 T Follicular Helper Cells

In the adaptive immune response, one particular group of CD4 T cells has emerged as a critical population in HIV infection and pathogenesis. A subset of T cells, called T follicular helper cells (TFH), play an important role in the humoral immune response by providing B cell help that is critical to somatic hypermutation and affinity maturation[134,135]. TFH were first named and identified as a distinct T cell subpopulation by their high expression of CXCR5 in 2000, although the requirement of T cell help to antigen-experienced B cells had been known for some time without a precise definition or identification of the properties of the T cells involved in this process[136–138]. CXCR5 had been previously identified as a homing receptor expressed on B cells to guide them to B cell follicles in lymph nodes via BCA-1/CXCL13 gradient (the ligand for CXCR5)[139,140]. Antigen-specific CD4 T cells in lymph nodes are primed on dendritic cells in the T cell zone, recognizing peptide presented on MHC-II complexes[141]. TFH then upregulate CXCR5 and downregulate CCR7 and migrate into the B cell follicle and into germinal centers[142]. There, they provide activation and co-stimulatory signalling to antigen-experienced B cells and promote survival of affinity-matured B cells[137,138]. TFH are a critical component of germinal center formation and maintenance via secretion of IL-21 and help maintain Bcl-6 expression on B cells[143,144].

TFH have high surface expression of CXCR5, PD-1, ICOS, SLAM-associated protein (SAP), BLTA, and OX40[145–148]. They also express high levels of IL-21 and transcription factors Bcl6 and cMaf[143,149,150]. Microarrays of TFH have strongly associated CD84 and CD200, and additional genes also found in B cells, with a TFH phenotype[151]. The defining transcription factor for TFH is Bcl6 (B cell lymphoma 6 protein), which acts as a repressor on more than 3,000 genes[152,153]. IL-21, which promotes immunoglobulin secretion and differentiation of B cells, is suppressed by IL-4, which in turn reduces expression of Blimp-1[150]. An ICOS-deficient mouse model showed reduced frequency of TFH, with defective IL-17 production and high expression of transcription factor c-Maf[154]. Other transcription factors include IRF4, AP-1, Batf, and members of the STAT family[134]. IL-2 signalling is a potent inhibitor of TFH differentiation, and STAT5 negatively regulates TFH development[155,156].

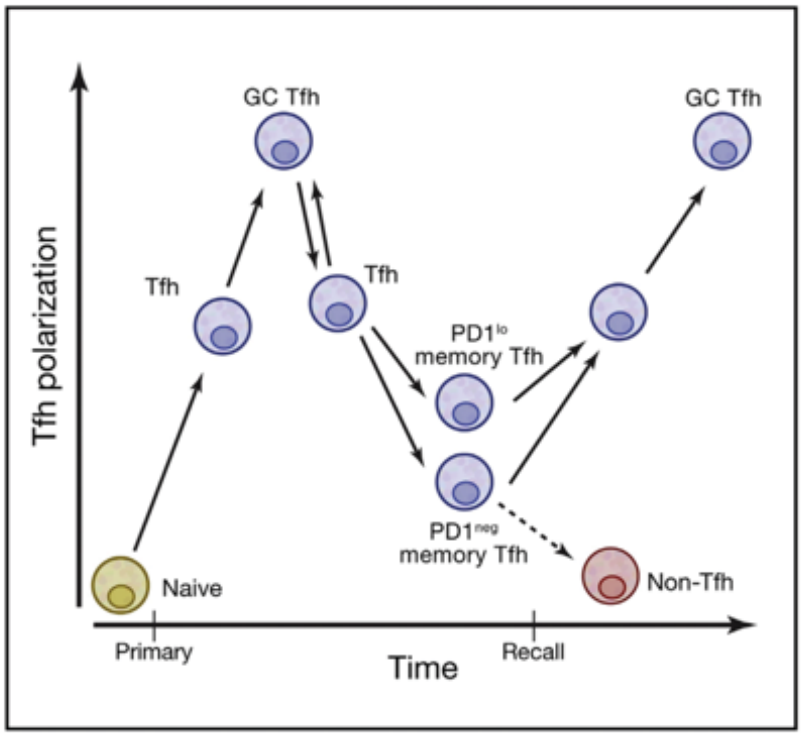


Fig 1.7 Differentiation and plasticity of T follicular helper cells in lymph nodes. From Crotty 2014, Immunity[134]. TFH develop from naïve T cells and gain TFH characteristics through interactions with dendritic cells and B cells, and can enter the germinal center upon up-regulation of CXCR5 and PD1 to promote somatic hypermutation and affinity maturation of antigen-exposed B cells.

The differentiation pathway to generate TFH from naïve CD4 T cells has multiple stages. It starts with priming of naïve T cells on dendritic cells in the T cell zone, and if the T cell receives the appropriate cytokine signals and receptor-ligand interactions (in mice, IL-6, ICOS, IL-2 and TCR[157]), it will up-regulate CXCR5 and migrate to B cell follicle[135]. Further antigen presentation on memory B cells along with ICOS-ICOSL signalling is critical for the second stage of TFH development[158]. These "early" TFH which reside in the T/B cell border and in the follicle, while "late" TFH have higher surface expression of CXCR5 and PD-1 and reside in the germinal center[158–160]. This movement into the germinal centers is marked by further polarization, with high expression of Maf, Bcl6, SAP, and BTLA on nearly all GC TFH[134]. However, GC TFH are not restricted to germinal centers: they can migrate from the B cell follicle and develop a resting memory phenotype upon upregulation of IL7Rα[158,160]. While this pathway is the most widely accepted, there are multiple alternative TFH differentiation pathways, and as of yet there are no well-defined in vitro differentiation conditions that can reliably polarize cells to a TFH phenotype[135].

Control of the number of TFH within germinal centers helps prevent the generation and selection of autoantibodies[161]. Within the population of CXCR5++PD1++ CD4+ T cells

found in germinal centers, 10-25% are Foxp3+Blimp-1+ T follicular regulatory cells (TFR), and play a regulatory role in maintaining the number of TFH and B cells in the germinal center [162,163]. In mouse models, follicular regulatory T cells were inhibited by PD-1 and PD-L1[164]. The ratio of TFH to TFR is moderated by IL-21, which inhibits TFR but not TFH [165]. The mechanisms of TFR suppression are not well defined, but CTLA-4 regulation of CD28 ligand on B cells is likely a major suppressor of signalling received by TFH[166]. Though they share some characteristics of TFH (including some transcriptional profiles and cell surface markers), TFR are derived from Foxp3+ regulatory precursors instead of naïve T cells or existing TFH. The ratio of TFH to TFR can vary with infection and anatomical location[167], and the degree of expansion/contraction and function of TFR is not well understood in the scenario of overall TFH expansion such as in HIV/SIV infection.

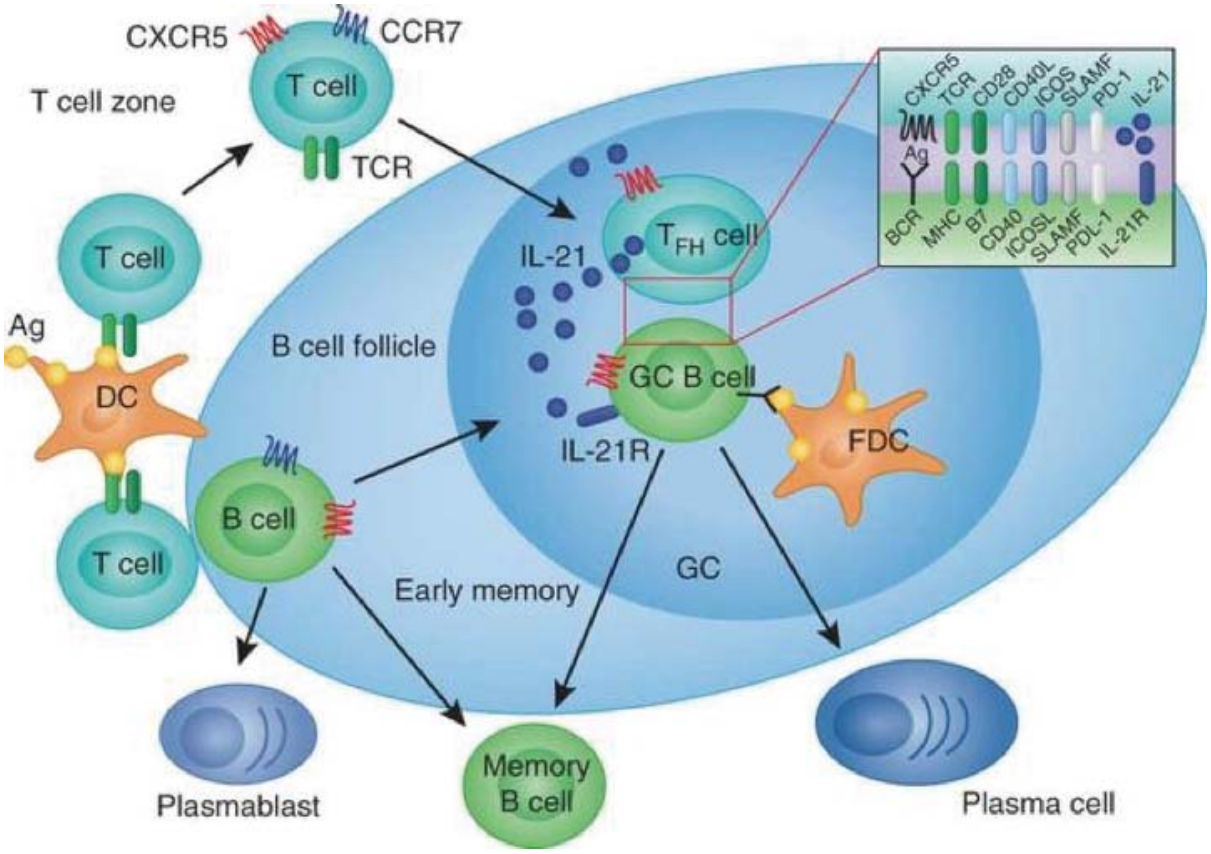


Fig 1.8. TFH localisation and migration. From Nutt and Tarlinton 2011, Nature Immunology[168]. T cells are primed on dendritic cells in the T cell zone, upregulate CXCR5 and downregulate CCR7 and migrate to the B cell follicle and into germinal centers. There, they interact with B cells and signal using TCR, CD28, CD40L, ICOS, SLAMF, and PD-1 to promote affinity maturation and somatic hypermutation.

1.3.2 TFH in HIV/SIV

TFH were first identified as a target of HIV/SIV infection when studies showed that TFH were expanded in chronic HIV and SIV infection[169,170]. Infection of macaques with

SIVmac239 showed an increase in both number of germinal centers and accumulation of Ki67+ cells within GCs, with CD4 T cells accumulating within and CD8 T cells largely excluded from GCs[171170]. The CD4 T cells that accumulated in chronic infection were PD-1^{hi} and correlated with CD27+ memory B cells and antibody production[171]. TFH were susceptible to SIV infection, with higher copy numbers of SIV gag DNA than in non-TFH, and increased as a proportion of central memory T cells in chronically infected macaques[170]. In lymph nodes from chronically HIV infected individuals, bulk and HIV-specific TFH were expanded compared to healthy controls and were associated with plasma viremia, with Gag-specific TFH preferentially secreting IL-21[169]. TFH were enriched for HIV-specific T cells in another study, and overall TFH were higher in viremic than in cART-suppressed donors[172]. Productive infection in an in vitro system and copy numbers of gag DNA from ex vivo sorted cells were both highest in TFH, and TFH again correlated with plasma viremia[172]. However, expansion of TFH does not lead to improved B cell help: TFH isolated from HIV-infected individuals did not support high levels of IgG production and led to reduced B cell survival in co-culture assays[173]. In elite controller rhesus macaques, productive SIV infection is restricted to TFH and was attributed to highly effective CTL that clear infected cells except in "sanctuary" B cell follicles[174].

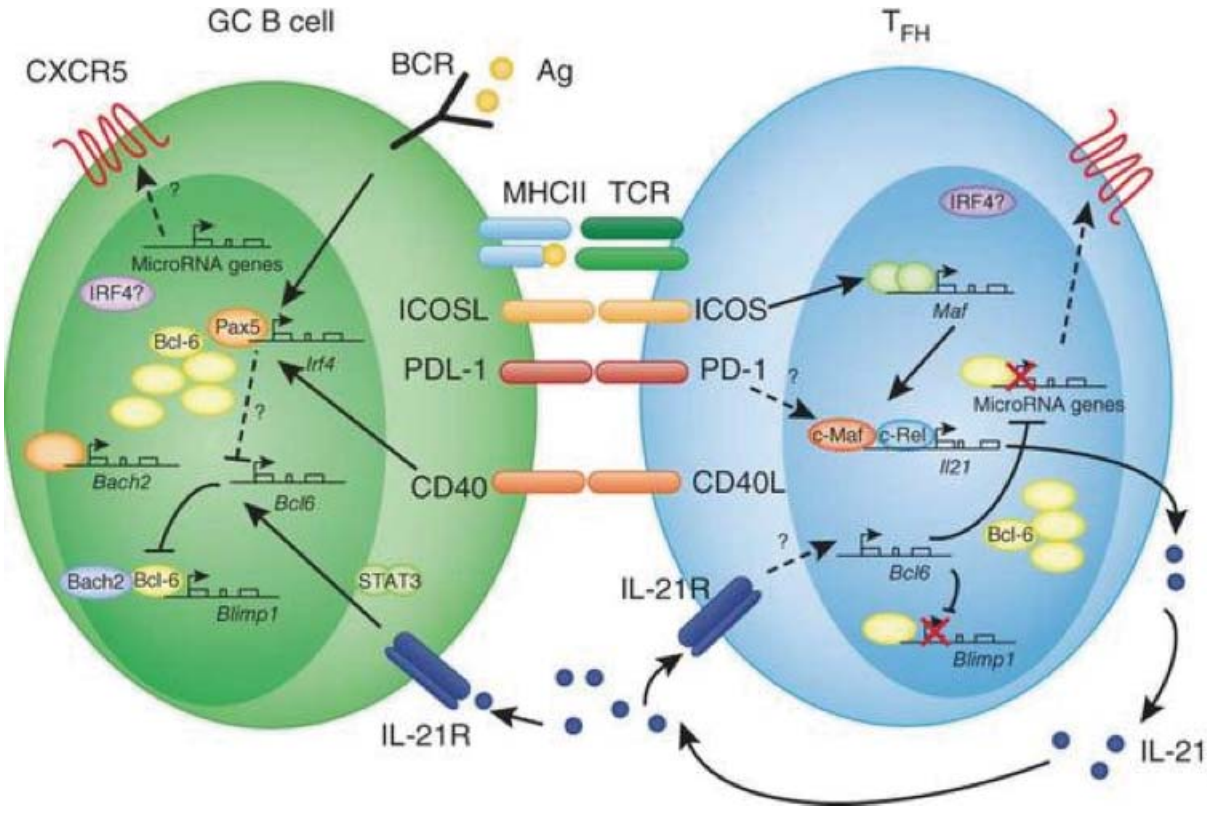


Fig 1.9. TFH and B cell interactions and transcriptional networks. From Nutt and Tarlinton 2011, Nature Immunology[168]

Table 1.2 – Markers of Follicular helper T cells

CXCR5	CD185, BLR1	GPCR, CXC chemokine receptor. Migration to GC. Receptor for CXCL13
PD1	CD279	Binds PD-L1 and PD-L2. Immune checkpoint, promotes self-tolerance and reduces autoimmunity.
ICOS	CD278	Inducible T cell co-stimulator, expressed on activated T cells.
BTLA	CD272	B- and T-lymphocyte attenuator, expressed on Th1 cells, negatively regulates T cell responses.
CD84		Member of the SLAM family
CD40L	CD154	TNF superfamily, expressed on activated T cells. Binds CD40, α5β1. Promotes B cell maturation
OX40	CD134, TNFRSF4	Secondary co-stimulatory immune checkpoint molecule
SAP	SLAM-associated protein	
IL-21	Interleukin 21	Induces cell division and proliferation in targets : NK cells and CTL
Bcl6	Transcription factor	Master regulator of TFH
cMaf	Transcription factor	

1.3.3 Peripheral TFH-like cells in HIV/SIV

While the majority of TFH are located in lymph nodes and secondary lymphoid tissues, circulating CD4 T cells that have a TFH-like phenotype have been described in multiple infections[175–177]. These cells have high expression of PD-1 and CXCR5, but it is unclear what their exact relationship is to lymph node resident TFH. There is evidence that CXCR5+ CD4 T cells in the blood have a memory phenotype that can be Th1, Th2, or Th17-like, but can still provide B cell help[178,179]. Circulating CXCR5+ CD4 T cells from healthy donors provide B cell help in co-culture experiments, and a subset of circulating TFH that were $CCR7^{hi}CXCR5^{hi}CCR6^{hi}PD1^{hi}$ enhanced isotype switching and secreted IL-21[177]. However, in HIV-infection there was no correlation between the frequency of circulating TFH and memory B cells or plasma neutralization, and gene expression profiling of the circulating TFH were more akin to memory than lymph node TFH[177]. Another study showed that in spite of no correlation between total circulating CXCR5+ CD4 and HIV bnAb development, a PD1+CXCR3- subset shared a transcriptional profile with germinal center TFH, were functional for B cell help and were associated with HIV bnAb[180]. Higher frequency of peripheral TFH was associated with protective antibody responses in the RV144 HIV vaccine trial[181]. PD1+ CXCR5+ peripheral CD4 T cells showed higher frequency of inducible HIV and showed reduced activation under cART in another study[182]. Finally, circulating TFH from another study showed an IL-2 gene signature and TH1 polarization, with reduced B cell

responses[183]. This impairment was reprogrammable, with interference in the IL-2 signalling leading to restored antibody responses[183].

1.3.4 Humoral immunity

B cell follicles and germinal centers are dynamic: imaging of germinal centers shows repeated movement of B cells between the dark and light zones as they undergo somatic hypermutation of immunoglobulin genes in the dark zone and return to light zone to form contacts with cognate TFH[184]. Some studies have indicated that B cell division is restricted to the dark zone[185], although there is some evidence that B cells can divide in both zones[184]. TFH provide strong selection signals for high affinity B cells, with B cells that have undergone more rounds of hypermutation and cell division capturing and presenting higher levels of antigen to TFH[186,187].

B cell follicles and germinal centers contain follicular dendritic cells (FDC), which have a mesenchymal origin (unlike plasmacytoid or myeloid dendritic cells)[30]. FDC play an important role in organization lymph node architecture and produce CXCL13, thereby recruiting CXCR5+ B and T cells [188]. Though they are not productively infected by HIV/SIV, the FDC network traps circulating virus and other pathogens and maintain them for long periods as immune complexes. These immune complexes, made up of antigen with specific antibodies and/or complement, and are retained via low affinity immunoglobulin gamma Fc region receptor II-B (FcγRIIB, CD32)[189]. This contributes to the selection of high-affinity BCRs that can crosslink the HIV/SIV antigens deposited on the surfaces of FDCs [190]. At some point, high affinity B cells need signalling to exit the B cell follicle and develop a memory B cell phenotype. The B cells that engage antigen on FDCs can also have their FcγRII engaged, and this crosslinking of FcγRII to the BCR adds a negative signalling pathway that could help GC cells exit the SHM-affinity maturation cycle[190].

Diversity of the immunoglobulin repertoire is generated on two levels: the recombination of V, D, and J segments (heavy chain) and V and J segments (light chain) from the germline, and through the processes of somatic hypermutation and affinity maturation in the adaptive immune response. The human immunoglobulin locus is split between chromosome 14 (heavy chain, chromosome 7 in macaques), chromosome 2 (kappa chain, chromosome 13 in macaques) and chromosome 22 (lambda chain, chromosome 10 in macaques)[191,192]. As B cells develop in the bone marrow, they recombine at immunoglobulin locus, first joining a D (diversity) segment with a J (joining) segment (eliminating all other D and J segments) and then joining the DJ to a V segment (again eliminating all other V segments)[193]. The VDJ

segment is joined with the M constant region, and the D constant region is added later on and spliced to express either M or D constant region. There are 129 total IgH V genes (including 38-46 functional genes plus ORFs and pseudogenes), 27 IgH D genes (25 functional, 23 unique)[191], and 9 IgH J genes (6 functional), though it is unclear what the distribution of individual genes is among the general population[193,194]. For the light chain genes, there are 76 IgK V genes (34-38 functional), 5 IgK J genes (all functional), 74 IgL V genes (29-33 functional genes), and 11 IgK J genes (4-5 functional genes)[195–197]. This gives a combinatorial diversity of 5244-6348 V(D)J heavy chains, 170-190 kappa chains and 116-165 lambda chains (assuming perfect recombination), yielding millions of possible pairings of heavy and lights chains[197].

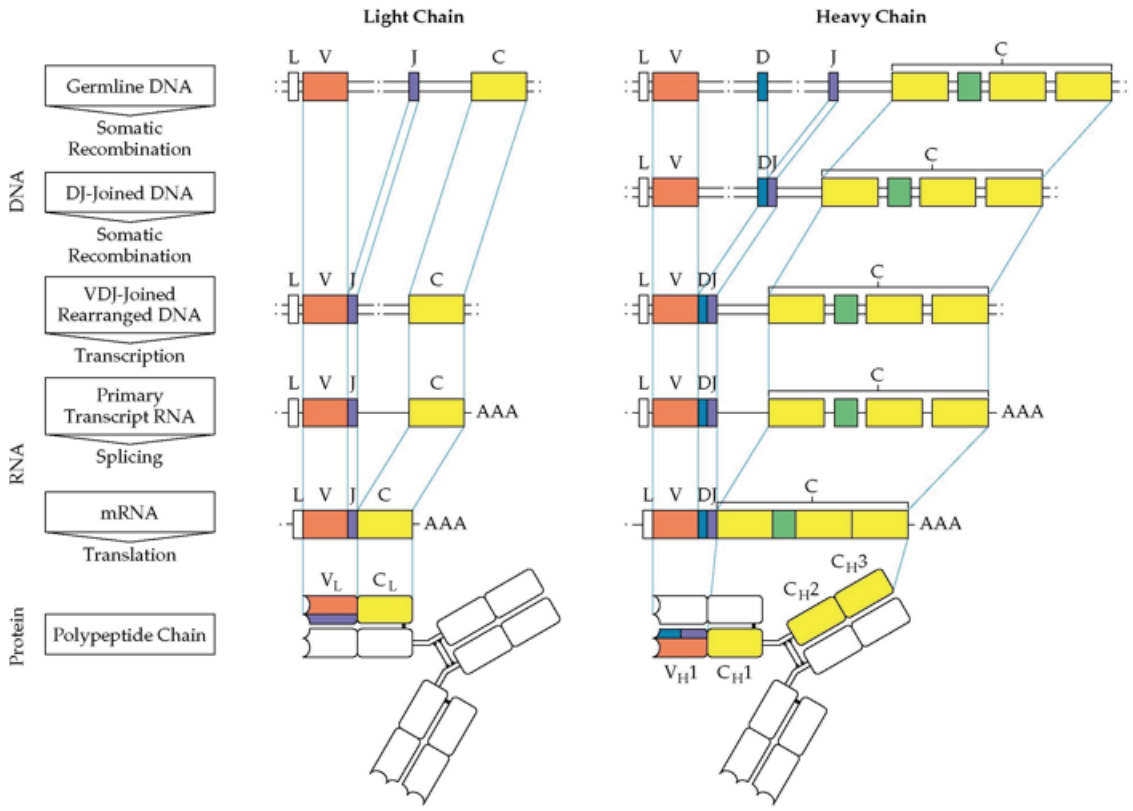


Fig 1.10. Germline rearrangement of heavy and light chains. Adapted from Schatz and Swanson 2011 Annual review of genetics[198]. At the germline level, chromosomal DNA is cut so that a single V, D, and J gene are joined to form a heavy chain exon (in light chains, a single D gene and J gene are joined to form a light chain exon), with the remaining gene segments discarded. The constant regions, located downstream, are spliced to the heavy or light chain exons during RNA transcript processing, and synthesized into the VL/CL or VH/CH polypeptide.

At each of the VD and DJ join sites, a terminal deoxynucleotidyltransferase can randomly add additional nucleotides, such that two B cells with the same VDJ segments may differ in their amino acid sequences by several residues[199]. The kappa and lambda chains do not have a D segment, but similarly combine a single V and J segment at the genomic level, and the kappa or lambda constant region is added at the level of mRNA transcription[200]. The addition or

deletion of several nucleotides at each junction in the process of generating functional BCRs from germline genes increases the diversity in the pool of naïve BCRs beyond the thousands of VDJ and VJ combinations by several orders of magnitudes. Some naïve BCRs are initially autoreactive, and undergo receptor editing to render them non-autoreactive[201,202]. Receptor editing, not clonal deletion, is the primary mechanism by which the immune system regulates autoreactive B cells and maintains tolerance[203]. Autoreactive B cells in the bone marrow will die due to neglect, be rendered anergic, or undergo receptor editing and rescue[203]. In receptor editing, a secondary gene rearrangement using the original recombination machinery and occurs first at the IgK, and if unsuccessful, IgL locus (and infrequently at the IgH locus)[204]. Introduction of nucleotides in the CDR3 is biased towards tyrosine and glycine, and away from highly charged/polar and hydrophobic amino acids[205,206].

D gene segment rearrangement has additional features that can increase the diversity of immunoglobulin heavy chains from germline. While the specificity of VDJ recombination largely relies on the 12/23 rule (that recombination signal sequences (RSS) with 12 base pair spacers can only join to an RSS with a 23 base pair spacer) to preserve the architecture of heavy chains, D genes can violate this rule and form D-D segments in approximately 5% of antibodies[200]. There is also evidence, though controversial, that D gene segments can also be inverted or include irregular spacers[207,208]. D gene heavy chains segments can be used in six reading frames (three forward and three reverse), and in the murine model, one reading frame was preferentially selected while other reading frames were selected against on the basis of stop codons or truncated proteins[209]. This selection occurs at the point of D-J rearrangement and before antigen selection can influence the repertoire. A high-throughput sequencing study using human PBMC showed that for most D genes a single reading frame dominated the sample, and there was only rare usage of inverted reading frames[210].

The heavy chain is divided into four framework regions (FR1-FR4) and three complementarity-determining regions (CDR1-CDR3, also called hypervariability regions). The FR are typically more highly conserved than the CDRs. FR2 and FR4 form the hydrophobic core of the VH-VL dimer; FR1 forms the joint between VH and the constant region; and FR3 interacts with some of the CDRs[211]. The heavy chain CDRs, with CDR3 in particular, contribute heavily to antigen binding and sensitivity, with the heavy chain alone sufficient for binding irrespective of light chain pairing in many antibodies[212,213].

1.3.5 Genetics and genomics of antibodies

The diversity generated after recombination at the germline level is on the order of 10^6 unique BCRs made up of single heavy chain and light chain (either kappa or lambda). The uniformity of BCRs decorating a single cell is governed by two properties: allelic exclusion and isotypic exclusion. Allelic exclusion ensure that each B cell expresses the heavy chain from only one chromosome – if the first rearrangement is out of frame, recombination of the VDJ regions proceeds to the second chromosome to produce a functional heavy chain[214]. In a similar fashion, isotypic exclusion permits only one light chain, either kappa or lambda, to be expressed by the cell[200]. In both cases, once a successful (in-frame) heavy or light chain is expressed, further rearrangement of the remaining heavy or light chains is suppressed so that the cell expresses a single heavy-light chain combination.

B cells initially express BCRs with constant region M, but after exposure to antigen can undergo class switch recombination to express D ,G, A, or E constant regions[215]. Each class of antibody (determined by the constant region) has different properties and different polypeptide structures, and confers a unique set of functional attributes that specialize isotypes to different types of pathogens[214]. IgM antibodies activate the complement system and are made of a complex of five units. IgA antibodies are found largely in mucosal surfaces and are secreted into fluids and can be a single unit or multimerized. Single-unit IgE antibodies are involved in allergic responses and in responses to parasites and bind mast cells and basophils. IgD antibodies exclusively act as membrane-bound surface receptors for antigen. IgG antibodies are divided into four subclasses, have a wide variety of functions, and are the most abundant antibody in serum. All four classes are involved in responses to pathogens and can, to varying degrees, can cross the placenta, activate the complement pathway, and bind the Fc receptor on macrophages and phagocytic cells.

The repertoire of naïve BCRs is vast enough to bind most antigens but with relatively low affinity. The second step in the generation of antibody diversity occurs with the onset of the adaptive immune response, where the initial B cells that bind antigen undergo a process of refinement and maturation to generate a highly specific and potent immune response. After exposure to an antigen, naïve B cells differentiate to become antibody-secreting plasma cells or memory B cells. Memory B cells migrate to lymph nodes and secondary lymphoid tissues and interact with antigen presented on TCRs of specialized T cells called TFH. If they are signalled appropriately, they form germinal centers within B cell follicles and undergo rounds of somatic hypermutation and affinity maturation. In the dark zone of the germinal center, B

cells undergo clonal expansion and somatic hypermutation[216]. After rearranging their heavy and light chains, B cells exit the dark zone and enter the light zone, where they undergo selection by interacting with T cells and follicular dendritic cells. They also undergo class switch recombination in the light zone. If they receive adequate signalling from TFH and DCs, they can re-enter the dark zone and continue to undergo somatic hypermutation, or exit the germinal center as high-affinity memory B cells, differentiate into plasmablasts and plasma cells. The rate of somatic hypermutation is estimated to be around 1 in 1000 mutations per base per generation, equating to about one mutation in the heavy or light chain per cell division[214].

The diversity generated by multiple rounds of somatic hypermutation and affinity maturation is considerable, and based on the distance in sequences between highly mutated antibodies discovered in longitudinally infected patients and putative germline sequences, it is estimated that up to 10^{11} possible unique BCRs could be generated – more than the number of B cells in an adult human at a given time. The pool of naïve B cells is constantly replenished from the bone marrow in healthy individuals, but declines with age[217].

1.3.6 Broadly neutralizing antibodies in HIV

The first antibodies to HIV that were capable of neutralizing multiple diverse strains of HIV-1 were isolated in the early 1990s[218,219] using phage display libraries and cell immortalization and hybridomas[220]. In recent years, more high-throughput methods have been developed, using single-cell sorting and cell culture to screen hundreds to thousands of individual cells, and proceeding with antibody characterization after screening for neutralization[221,222]. Neutralization is measured in terms of breadth of viruses neutralized as well as potency (concentration of antibody needed to neutralize 50 or 80% of virus) in a luciferase-based assay[223,224]. Neutralization panels contain molecularly cloned pseudoviruses from HIV-1 isolates and include viruses of high (tier 1a/1b), moderate (tier 2) or low (tier 3) sensitivity[225]. Standardized panels can be subtype specific and contain a wide range of geographic and genetic diversity to accurately characterize the potential of monoclonal antibodies.[131,226,227]

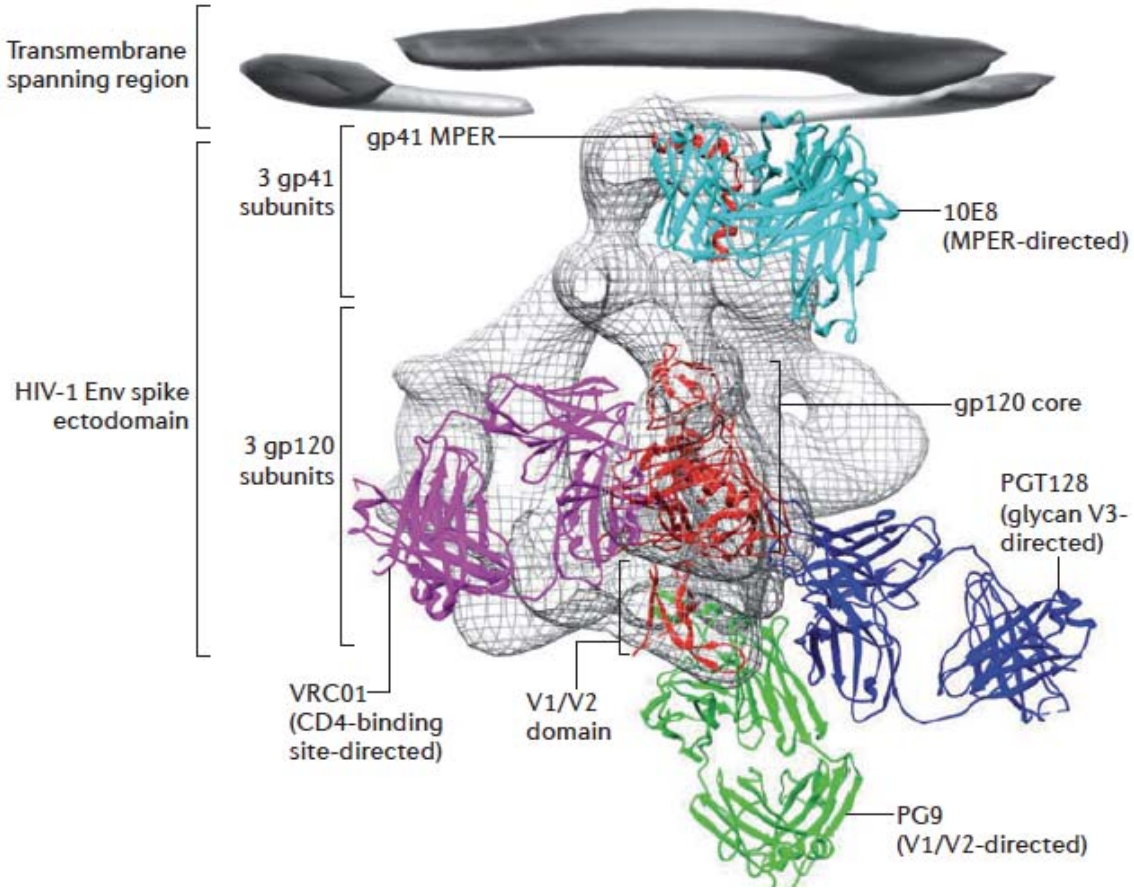


Fig 1.11. Sites of HIV broadly neutralizing antibody binding to envelope. From Kwong, Mascola, and Nabel 2013 Nature Reviews Immunology[219]. From HIV bnAbs are grouped based on the area where their class of antibodies bind, and include CD4 binding site, V1/V2 directed, glycan/V3 directed, membrane proximal external region (MPER), and V3/CD4 binding site.

Broadly neutralizing antibodies are grouped into five groups based on their binding position to the HIV envelope spike. Families of bnAbs bind to the CD4 binding site (CD4bs), variable regions V1/V2, variable region V3, the glycans bridging the gp120/gp41 region, and the membrane proximal external region (MPER)[228]. CD4 binding site antibodies mimic the interactions between CD4 and HIV envelope, particularly on gp120[99]. Some antibodies bind both individual subunits of the env spike and the entire assembled envelope complex, while others only bind one or the other[218]. The role of glycans on the surface of HIV envelope is critical to bNAb binding and neutralization. Many bNabs target, partially or exclusively, carbohydrates decorating the amino acid surface of V1/V2 and V3 regions[229]. In regions where bNAbs target only amino acid sequences, their targets can be masked by glycans to reduce or eliminate binding. Thus, the issue of accessibility to target sites and the dynamics of glycan decorations impacts the binding and neutralization of mAbs.

Table 1.3 – Notable HIV broadly neutralizing antibodies

Antibody	Region	VH gene	Notable Features
VRC01	CD4bs	IGHV1-02*02	14 AA CDR3, 87% neutralization
NIH45-46	CD4bs	IGHV1-2	18 AA CDR3, 96-98 mutations
10E8	MPER	IGHV3-15*05	22 AA, 64-69 somatic mutations
4E10	MPER	IGHV1-69	18 AA CDR3
2F5	MPER	IGHV2-5	22 AA CDR3
2G12	Glycan	IGHV3-21	13 AA CDR3
3BC117	V3/CD4i	IGHV1-2	12 AA, 76-77 somatic mutations
PGT120s	V3/V4 glycan	IGHV4-39*07	19-24 AA, 58-78 somatic mutations
PGT130s	V3/V4 glycan	IGHV4-39*07	18-19 AA, 60-84 somatic mutations
PG9	V2 glycan	IGHV3-20	30 AA, 41 somatic mutations
CH01-04	V2 glycan	IGHV3-20	26 AA, 48-49 somatic mutations
PG16	V1/V2	IGHV3-33	30 AA CDR3, 43 somatic mutations (na)

As more broadly neutralizing antibodies were discovered, common features that contributed to neutralization began to be elucidated. MPER-directed antibodies frequently have germline VH1-69, while CD4 binding site antibodies frequently have germline VH1-2 and VH1-46[230]. These VH genes likely make a series of key contacts with HIV envelope from the germline sequence, and these contacts and binding are subsequently refined and develop into highly potent antibodies after multiple rounds of somatic hypermutation[231]. Many bnAbs also have unusually long CDR3 regions- whereas a typical naive CDR3 is 12-15 amino acids in length[232], and in the memory B cell population less than 10% of CDR3s are greater than 25 amino acids, many broadly neutralizing antibodies have CDR3s 20-30 amino acids long[232,233]. The advantage of long CDR3 regions is related to the ability of long loops to penetrate the glycans surrounding the V1V2 site, with crystal structures of PG9 (28 AA CDR3) bound to HIV envelope showing the long CDR3 loop entering the CD4bs pocket[234]. Finally, broadly neutralizing antibodies show evidence of extensive mutation (calculated by measuring the divergence of sequence from the putative germline using BLAST or other alignment algorithms), with 30% or higher mutation from germline[228]. However, the calculated putative ancestors of many broadly neutralizing antibodies do not bind to HIV-1 envelope glycoproteins, raising the question of how reliable these inferences are. [98,230,233,235]. While high-affinity antibodies require successive rounds of mutation to acquire the key residues and contacts for better neutralization, without initial binding and the induction of the humoral response, no antibody response to HIV can be generated. Finally, many bnAbs are

polyreactive, which may be disadvantageous in terms of normal B cell development checkpoints against autoreactivity but may be essential for some HIV neutralization[236].

While much attention has been focused on the induction of broadly neutralizing antibodies, the role of non-neutralizing or weakly neutralizing antibodies is also of interest, in particular because the RV-144 vaccine trial showed a modest correlation between protection from infection and titers of non-neutralizing antibodies to the V2 region[237]. IgG3 and IgG4 antibodies to p24 and gp120 and highly glycosylated HIV-specific antibodies have been associated with HIV controllers, even though the antibodies isolated were not considered broad and potent neutralizers[238–240]. Non-neutralizing antibodies may still play a role in ADCC, activating the complement system, and in recruiting phagocytes[241].

1.3.7 Antibody repertoires and sequencing of the humoral response

With the advent of next-generation sequencing, it became feasible to sequence a greater sample of the immunoglobulin repertoire – resulting in a more detailed picture of the breadth of healthy and pathogen-specific humoral immunity[242]. Repertoire sequencing has been used to inform understanding of the evolution of HIV bnAbs, from germline (putatively weakly or non binding) to early mAbs (that bind but do not neutralize, or not with high affinity) to broad neutralization with high affinity. This has required developing new assays and bioinformatics tools to accurately sample and sequence BCRs[243,244]. In some techniques for high throughput antibody sequencing, pairing of heavy and light chains is lost in favour of bulk sorting and sequencing of populations of B cells, with nucleic acids extracted in bulk and pairing information lost. However, in one study, pairing of heavy and light chains was re-created using phylogenetic inference and subsequently validated by targeted single-cell sequencing[245]. Paired sequencing of heavy and light chains requires either single-cell sorting into individual wells of 96 or 384 well plates and individual amplification of heavy and light chains, or custom microfluidic devices to capture and lyse individual cells in droplets with barcoded primers so that heavy and light chains pairing can be re-joined in the informatics stage of analysis[221,243,246].

One of the challenges in tracking the humoral response to HIV/SIV is in isolating HIV/SIV specific B cells from the total B cell pool. There are a number of techniques to do this, but the most common are the use of HIV envelope probes in flow cytometry based sorting to identify B cells that bind HIV/SIV, or screening of B cell culture supernatants for neutralization[230,247]. Different HIV/SIV probes have been generated, including monomeric gp120 subunits, soluble

trimeric gp140 proteins that have the hydrophobic gp41 domain removed, resurfaced stabilised core proteins (RSC), and cleaved trimers that have covalently linked gp120/gp41 (SOSIP probes)[248–251]. Early generations of probes, namely monomeric and unstabilized gp140 trimers, yielded mixed results due to the instability of the probes and the binding of B cells to non-biologically relevant epitopes[248–250]. However, stabilised probes have been used to isolate many broad and potent neutralizing antibodies, and studies using mutated probes to select for antibodies that target particular epitopes has been successful[99,252]. With any probe-based sorting, the quality of the B cells and antibodies generated is only as good as the probes themselves. Probes are also unable to capture plasmablasts and plasma cells that secrete high levels of antibodies but no longer display antibodies on the cell surface.

Phylogenetic analysis of immunoglobulins is not as advanced as other phylogenetic areas, and presents unique challenges in adapting the tools and methods used to the evolution of human and pathogen genomes. A typical sample of antibody heavy and light chain sequences from a memory B cell population can contain both newly generated memory B cells from a primary response as well as long-lived memory B cells that have undergone multiple rounds of affinity maturation – all with minimal supporting information on the evolutionary history of an individual cell. It can be difficult to distinguish mutations from germline that are generated by deoxynucleotidyltransferase at VDJ and VJ junctions from somatic hypermutations that arose as part of the antigen-specific immune response. At any given time, the memory B cell pool contains memory B cells that are highly specific for pathogens from the infectious history of the host but without any reliable way to map them back to a specific antigen. Within the B cells specific to an on-going infection like HIV/SIV, multiple generations of closely related B cells are present, including parent, daughter, and granddaughter sequences co-existing without any robust methods for determining the chronology of sequences. Each round of somatic hypermutation produces random changes that are subject to selection but contain no information on which mutations arose first in the absence of sequencing from serial samples. Furthermore, the naïve pool is constantly being replenished and in on-going infection, “new” memory responses are constantly generated. In patients under cART, it is unclear if there is on-going SHM and affinity maturation in the absence of viremia and circulating antigen. Furthermore, because SHM changes are indistinguishable from PCR errors, stringency is critical in designing and analysing sequencing experiments to ensure that biological diversity is not confounded with technical margins of error. The phylogeny and ontogeny of broadly neutralizing antibodies is of clinical relevance, as sequential immunization with targeted immunogens to push the humoral immunity on a path to neutralization is a major HIV vaccine

strategies[233]. In these schemes, a series of carefully designed immunogens first target a specific germline B cell receptor, and subsequent immunogens target the daughter generations of BCRs derived from that germline[233]. In this fashion the mutation process is directed to wards a family of BCRs that can neutralize HIV. Teasing apart the evolution of broadly neutralizing antibodies is a key component of this strategy, and efforts to find the putative ancestors of broadly neutralizing antibodies rely on accurate modelling of the evolution of antibody lineages.

1.4 Animal Models of HIV infection

The history of SIV infection in non-natural hosts begins before the discovery of HIV, with two outbreaks of lymphoma in captive rhesus macaques and stump-tailed macaques in the 1970s at the California National Primate Research Center[253]. SIV was not known to be the infectious agent at the time, though both groups had previous contact with sooty mangabeys infected with SIV prior to the lymphoma outbreak, and the infection was eventually traced from the original sooty mangabeys to the two California primate groups[254]. The California primates were later found to have spread the virus to other primate colonies when healthy macaques were transferred in subsequent years[255]. SIV infection has been found in forty different simian species and subspecies, with different epidemiological prevalence in different species: ranging from elevated rates in African Green Monkeys (estimated 40-50%) and sooty mangabeys to much lower rates in apes and none detected in gorillas, though accurate prevalence data is difficult due to the difficulty and restrictions on collecting serological data on dispersed and endangered species[255]. There are six classical lineages of SIV (table 4) that are all approximately the same genetic distance apart (up to 40% in the pol gene) and when visualized on a phylogenetic tree, are evenly interspersed with different lineages of HIV-1 and HIV-2[255].

SIV infection of non-natural hosts, mostly rhesus, cynomolgus, and pigtailed macaques, are some of the best characterized of the animal models of infection, with hundreds of studies to date[256,257]. Natural hosts of SIV are primarily of African origin, such as African Green monkeys and sooty mangabeys, and have coexisted with various strains of SIV for thousands of years. These animals can be infected but do not show disease progression or immune activation, and have been useful in understanding how virus is controlled[258–260]. Rhesus macaques, pig-tailed macaques and cynomolgus macaques develop pathogenic infection, with the onset of AIDS-like symptoms typically within two years (more rapidly than humans, with an average onset in untreated individuals in 5-10 years)[261]. These animal models are particularly useful in understanding acute infection, pathogenesis in secondary lymphoid

tissues and other compartments that are inaccessible in human subjects, and timing, dosage, and route of infection can be controlled. Nonhuman primates have proven invaluable in understanding HIV-related diseases and in testing novel drugs and interventions, from vaccines (prophylactic and therapeutic) to cure strategies and novel drugs.

Table 1.4 – SIV classic lineages

SIV	Animal Origin	Features
SIVcpz	Chimpanzees (*Pan troglodytes)*	Contains HIV-1, best defined lineage
SIVsm	Sooty mangabey (*Cercocebus atys*)	Contains SIVmac and HIV-2
SIVagm	African Green monkeys (*Chlorocebus*)	Diverse lineage with evidence of long history of co-evolution with host
SIVsyk	Syke's monkey (*Cercopithecus albogularis)*	Restricted *in vitro* cellular tropism to Syke's monkey PBMC (not in human, mangabey, or rhesus macaque)
SIVlhoest	L'Hoesti supergroup and mandrils	
SIVcol	*Colobus guereza*	

The human HLA types that are associated with altered HIV disease progression: either accelerating the onset of AIDS (B*08, A*01-B*08-DR3) or delaying onset (B*27, B*51, B*57, B*1503) have parallels in the macaque model: MHC alleles *Mamu A*01*, *B*17*, and *B*08* are all associated with control of viral replication, and DQB1 associated with rapid disease progression[262–266]. Epitopes in SIV gag, tat, vif, and nef have been identified as targets of CTL, with Gag epitopes TW10, KF11, and KP9 particularly well characterized[267]. Mamu-A*01 macaques have a CD8 T cell response dominated by Gag epitope CM9, with viral escape mutants occurring only late in infection[263,268,269]. Macaques also differ from humans in their ability to restrict viral infection and replication using TRIM5-α, APOBEC3, and tetherin. Indian and some Chinese origin rhesus macaques can restrict SIV and HIV infection with TRIM5- α, and certain APOBEC variants (F,G,H) are effective against HIV-1[270]. Indian rhesus and pig-tailed macaques have tetherin that can restrict HIV-1[270].

1.4.1 SIV Strains used in infection models

Because HIV does not naturally infect macaques, different strains of SIV have been used in infection and vaccine studies to recapitulate the disease pathogenesis and model acquisition and protection from infection. The molecular clone mac329 and the corresponding viral swarm mac251 were derived from colony-bred rhesus macaques and are R5-tropic, replicate preferentially in CD4 T cells, and are largely resistant to neutralizing antibodies[270,271]. SIVmac251 has a median viral peak of 7.3-7.5 log10 RNA copies/mL at 10-17 days post

infection, and a median setpoint of 4.6-5.7 log 10 RNA copies/mL in rhesus macaques[270]. Similarly to HIV, primary stocks of SIVmac251 are more resistant to neutralization than laboratory-adapted virus stocks[272]. Neutralizing antibodies are typically observed 5 to 8 months after infection and at low titers, but are effective against escape mutants particularly in the V2, V3, and V4 regions of Env[273]. ADCC responses are detected 3 weeks post infection, and gp140 antibody titers are associated with the NK cell-mediated ADCC responses[274]. There is heterogeneity even within SIVmac251 challenge stocks, with 10-368 unique Env sequences with an average nucleotide diversity of 0.3-1.0% recorded in one study of available viral swarms[275]. Passage of SIVmac251 through primary cells led to the isolation of SIVmac32H and its infectious molecular clones SIVmac32H(pJ5) and SIVmac32H(pC8)[276]. The latter mutant is attenuated in cynomolgus and rhesus macaques due to a 12-nuclotide deletion in nef[270]. The other major strain used is SIVsmE660, a swarm challenge stock derived from molecular clone SIVsmE543. SIVsmE660 is R5-tropic, and has less resistance to neutralizing antibodies than SIVmac251[277]. This is thought to be due to less masking of conserved Env epitopes in E660, although a subset of viruses in the SIVsmE660 swarm are highly resistant to antibody neutralization[277].

SIVmac251 and SIVsmE660 induce different humoral responses and different patterns of neutralization in response to infection. SIVmac251 is more difficult to neutralize than SIVsmE660, and most SIVmac251 envelopes are considered Tier 1b and Tier 2 (moderately resistant to neutralization) whereas SIVsmE660 envelopes are Tier 1a (highly sensitive to neutralization). One study of cloned SIVsmE660 envelopes from primary isolates and transmitted/founder viruses found that most envelopes were highly sensitive to antibodies targeting V3, CD4 induced epitopes, CD4 binding site, and V4 regions, whereas envelope clones from SIVmac239 and SIVmac251 were largely resistant to the panel tested[277].

Another study of animals infected with SIVsmE660 found that they were unable to control viremia despite developing high levels of autologous neutralizing antibodies within a month of infection[278]. Neutralization escape virus variants were associated with substitutions and insertion/deletion polymorphisms in the V1 and V4 envelope domains[278], and neutralizing antibodies were subsequently found in plasma, indicating a selective pressure-driven escape of SIV envelope. In contrast, neutralizing antibodies to SIVmac251 arise at 5 to eight months post infection and are initially focused on the V4 epitope[273]. Antibodies in the same animals at 16 months post infection broadened to target V1/V2 as well as V4, and analysis of viral escape variants showed selection of mutants with substitutions and insertion/deletions in the

same regions. Another detailed study of a macaque infected with SIVmac239 that responded with unusually high neutralizing antibody titers despite progressing to AIDS-like symptoms had evidence of viral escape targeted to the V1 and V4 regions of envelope[279]. Finally, an investigation of several mutations in SIVmac329 envelope (substitutions, elimination of N-glycan attachment sites, and deletions in V1/V2 loops) showed that these mutations increased sensitivity to neutralization from multiple pools of SIV-positive plasma, indicating global increases in neutralization sensitivity instead of merely exposing localized epitopes [280].

SIVmac251 and SIVsmE660 are frequently used together in heterologous vaccination-challenge studies, particularly in live attenuated virus (LAV) vaccines. These LAV vaccines have shown in numerous studies the ability to provide complete or partial protection from infection with homologous s and heterologous challenge, usually by deleting nef (Δnef), nef and vpr (Δ3), or the V1 and V2 regions of envelope (ΔV1V2)[281–284]. In a study of macaques infected with a live attenuated SIVmac239 (nef deletion) virus and subsequently challenged with SIVsmE660, vaccinated animals showed reduced acquisition of SIVsmE660 and better control of virus replication in the cases of infection[285].

1.4.2 Simian-human hybrid viruses

While the SIV-macaque model continues to provide valuable insights into disease pathogenesis and progression, problems arise in testing HIV vaccine candidates in this model because HIV-1 does not establish pathogenic infection in macaques. The differences in HIV and SIV envelope antigenicity means that neutralizing antibodies to SIV-derived immunogens do not clearly correlate to HIV immunogenicity, and thus SIV models cannot directly predict the immunogenicity of HIV vaccines but instead are used to test vaccine concepts [286,287]. There is also relatively limited diversity in the available SIV strains for testing broad and potent neutralization [288]. Simian-Human Immunodeficiency Virus hybrids (SHIVs), containing *tat*, *rev*, and *env* from HIV-1 on an SIV genetic scaffold, were developed to work around these problems as challenge viruses for HIV-1 vaccine candidates directly in macaques[289,290].

Several SHIV strains, with varying degrees of pathogenicity and in-host replication, were developed and have been used widely in vaccine studies and studies of passive immunization with neutralizing antibodies[291–293]. SHIV-89.6P is widely used in vaccine preclinical trials, as it causes very rapid loss of nearly all CD4 T cells, has high levels of viral replication, and infected animals develop simian AIDS quickly[294,295]. However, it is sensitive to early autologous neutralizing antibodies and is CXCR4-tropic, unlike most HIV infections that are

CCR5-tropic and largely escape and are resistant to early antibodies[295]. This is likely due to the depletion of naïve CXCR4+CCR5- CD4 T cells, which in turn drastically reduces the effector CD4 T cell response and causes rapid disease progression[287]. Efforts to develop R5-tropic SHIVs have been moderately successful, with SHIVSF162P3, SHIV1157ipd3N4 and SHIVAD8 all used in immunotherapy and/or antiviral resistance studies with pathogenesis more closely approximating human disease[289,290,296–298].

Broadly neutralizing antibodies to SIV have recently been isolated from chronically infected rhesus macaques. Rhesus macaques infected with SHIVSF162P3N yielded tier-2 neutralizing antibodies after 1-2 years of infection, including gp130-reative neutralizing antibodies[299]. In another study, an envelope construct of soluble human CD4 and HIV envelope yielded antibodies that could neutralize a wide range of HIV-1 isolates[300].

The main alternative animal models to nonhuman primates are humanized mice, which offer a unique set of advantages and drawbacks. The most commonly used humanized mice have a *scid* background and are transplanted with human fetal thymus and liver or peripheral blood lymphocytes[270]. *scid* mice can be crossed with NOD (non obese diabetic) mice, and then crossed again with *Il2rg*$^{-/-}$ mice and then engrafted with cord blood, fetal liver, or adult PBMC for use as humanized mice[301]. *Rag*$^{-/-}$*Il2rg*$^{-/-}$ mice have also been used, but like the aforementioned models, T cell development occurs in the mouse thymus instead of a human organ[302]. Finally, BLT mice have a *scid*/NOD/*Il2rg* background and have human fetal liver and thymus plus bone marrow transplantation after sub-lethal irradiation, and offer perhaps the best of the mouse models in terms of recapitulating the human immune system[303]. Nonetheless, the mouse models have major drawbacks compared to nonhuman primates: infection by mucosal transmission is not well established, humoral responses to HIV are nigh undetectable, and the durability of the engrafted human hematopoietic system, as well as the interaction between murine and human immune compartments, adds additional uncertainty[270].

1.5 HIV Vaccines

In 1983, the United States Secretary of Health and Human Services stated that a vaccine for HIV would be available within two years – over thirty years later, with more than 30 vaccine products tested nearly 95 clinical trials, that goal has not yet been achieved[304]. There have been many unforeseen challenges in developing an HIV vaccine, from the rapid evolution and escape of the virus itself, to the lack of correlates of protection and immunological barriers in human hosts. HIV-1 is incredibly diverse and mutates rapidly[305,306], meaning a successful vaccine would need to provide broad protection against a wide array of viral epitopes. Two

major targets for a vaccine, the env and gag genes, differ by up to 35% and 15% between viral clades, and more highly conserved viral epitopes are hidden within the envelope trimer and are exposed only upon binding to surface receptors[307]. HIV is decorated by glycoproteins that further camouflage and hinder binding of antibodies[239]. While elite controllers and long-term viremic non-progressors offer models of control of infection without AIDS, there are no natural models of viral clearance[308]. The early establishment of the latent reservoir, and its location within secondary lymphoid organs makes it difficult to target and eliminate all infected cells[74,75,309].

Table 1.5 Major HIV Vaccine Trials and results.

Efficacy trial (# participants)	Vaccine	Efficacy and significant results	Immune response	Immune correlates of risk
VAX004 (N=5403)	AIDSVAX B/B (rgp120 immunogens)	None	Weak nAb response	N/A
VAX003 (N=2546)	AIDSVAX B/E (rgp120 immunogens)	None	Weak nAb response	N/A
Step (N=3000)	MRKAd5 HIV-1 (rAd5 vector expressing Gag, Pol, and Nef)	None; trial halted after meeting prespecified futility boundaries. Significantly increased risk of HIV infection in men who were both Ad5-seropositive and uncircumcised, which waned with time since vaccination	$CD8^+$ T-cell response detected in the majority of vaccinees, although weak and of narrow breadth	N/A
HVTN 503/Phambili (N=801)	Same as Step	None; enrollment halted after lack of efficacy seen in Step	Similar to Step	N/A
RV144 (N=16,402)	ALVAC (canarypox vector expressing Env, Gag, and Pol) prime followed by AIDSVAX B/E boost	31.2% overall efficacy for prevention of HIV-1 infection in the modified intention-to-treat analysis. No subsequent effect on viremia or CD4 count in vaccinees who were infected . 68% efficacy for low or medium risk participants, no efficacy in the high-risk group; efficacy was highest over the first 12 months and then fell rapidly	Weak nAb response .Moderate $CD8^+$ and $CD4^+$ T-cell response; the $CD4^+$ T-cell response was directed against the V2 region of Env	Binding of IgG to the V1 and V2 regions of Env correlated with protection; protection was mitigated by the presence of plasma IgA directed against Env
HVTN 505 (N=2504)	VRC-HIVDNA016-00-VP (DNA expressing Gag, Pol, Nef, and Env) prime followed by VRC-HIVADV014-00-VP (rAd5 expressing Gag, Pol, and Env) boost	None; trial halted after meeting prespecified futility boundaries	Awaiting final trial results	N/A
Adapted from Cohen & Dolin, Novel HIV vaccine strategies: overview and perspective. Ther Adv Vaccines (2013) 1(3) 99 –112				

Several types of candidate vaccines have been tested in animal and human trials, the most prominent of which have been summarised in Table 1.5. Live attenuated vaccines, which had proved successful in protecting against measles and polio, were created using *nef* mutants to limit replication[310]. While initially successful in some animal trials, some strains reverted to pathogenic viruses and caused AIDS, making them too dangerous to move into clinical trials[311]. A killed whole virus vaccine was tested alongside anti-retroviral therapy but showed no effect on survival, and a later trial as a therapeutic vaccine showed no effect on viral reservoir[312]. Most vaccine trials have used recombinant viral proteins, delivery of plasmids with viral genes, or alternative viral vectors to attempt to induce a robust immune response, with varying degrees of success.

Vaccines consisting of HIV-1 envelope gp160 and gp120 proteins were well-tolerated and induced high antibody titers within 3 doses, but did not provide protection against primary isolates[313,314]. The AIDSVAX B/B and B/E trials used bivalent gp120 vaccines in two large phase III trials, but had no difference in rate of infection between the vaccinated and unvaccinated arms, showing that vaccination with envelope alone was not sufficient to protect against infection[313]. DNA vaccines that expressed HIV structural proteins from several clades had good immunogenicity in animal trials, but performed poorly in human phase I trials unless co-administered with adenoviral vector boosts[315,316]. Early trials of vaccinia vectors with HIV *env*, *gag*, and *pol* genes showed low antibody and T cell responses in humans and did not protect non-human primates from HIV infection[317,318]. Modified vaccinia virus Ankara (MVA), which has a more favourable safety profile, is currently being studied in on-going trials[319]. Finally, adenovirus and canarypox vectors have been used in large phase IIb trials, with RV-144 (using ALVAC canarypox with a subunit AIDSVAX B/E boost) showing 31% efficacy. The STEP trial used an adenovirus vector made up of three adenoviruses with clade B *gag*, *pol*, and *nef* in three doses, but was halted early due to vaccine futility. The subsequent HVTN 505 used a DNA prime with a different adenovirus vector boost (containing *gag* and *pol* from clade B and *env* from clades A, B, and C) with the hopes of inducing both cellular and humoral immunity, but was also stopped early due to insufficient protection from infection.

Each trial informs the next, and current vaccine products aim to induce broadly neutralizing antibodies along with a strong cell-mediated response. The discovery of broadly neutralizing antibodies has trigged studies into how they might be induced using stepwise vaccination with a series of HIV envelope proteins that guide and focus the humoral response[235,320]. This B cell

lineage vaccine design utilizes existing knowledge of pathways to maturation of broadly neutralizing antibodies and tries to replicate them using rationally designed immunogens. Trials using cytomegalovirus vectors in rhesus macaques has shown persistent, highly effective CTL that control pathogenic SIV infection and show potential to clear latent reservoirs[321,322]. Finally, in an immune correlates of risk analysis of RV-144, binding of IgG to V1 and V2 regions of envelope showed that *ex vivo* non-neutralizing antibodies may play an important role in *in vivo* protection[323]. Though many past HIV vaccine trials have been disappointing, there are future vaccine candidates and strategies that are promising and build upon the lessons learned in the last 30 years of research.

1.6 Summary and Aims

In this thesis, I investigate three aspects of SIV infection of rhesus macaques. The SIV-rhesus macaque model of HIV infection offers two key features: the ability to serially sample secondary lymphoid tissues, where the bulk of HIV/SIV infection takes place; and the opportunity to study early immunological events (before clinical detection of infection) that are critically important in disease pathogenesis. First, we wanted to ask what are the features of the early immunoglobulin response to SIV infection, and how do those features change over the course of infection. Next, we asked what the markers are that define productively infected CD4 T cells and what the relative contributions of different subsets of CD4 T cells are to virus replication. Finally, we asked if viral evolution and diversity are consistent in subsets of CD4 T cells in lymph nodes throughout infection.

In the first results chapter, I use a gp140 envelope protein probe to isolate SIV-specific memory B cell from blood, lymph nodes, and bone marrow throughout early infection. I sequence immunoglobulin repertoires to track the evolution of the humoral response to infection and identify features of IgM and IgG SIV-specific responses. In the second results chapter, I examine the infection of subsets of CD4 T cells, using cell surface markers to identify features of productively infected cells, and use RNA probes to identify individually infected cells. I quantify viral infection to show the relative contributions of CD4 T cell populations at different stages of infection. Finally, I use viral sequencing of plasma, cell-associated RNA and proviral DNA to examine the history of infection and on-going viral evolution in in secondary lymphoid tissues.

2 Materials and Methods

2.1 Animal Study Protocol

Animal study protocol VRC-12-417 (Tracking SIV Infection of T cells and its Impact on B Cell Responses in Early Infection), was approved by the Animal Care and Use Committee (ACUC) at the Vaccine Research Center, NIAID/NIH. Six rhesus macaques (*macaca mulatta*, of Indian origin) were infected intravenously with a 1:3000 dilution of SIVmac251. This dose had been shown in previous studies to be infectious with a single IV injection. Each animal was sampled four times (up to three biopsies and euthanasia) in early infection. Blood and tissue samples (inguinal lymph node, bone marrow, and gut) were taken at each biopsy, with additional tissues taken at necropsy (mesenteric and axillary lymph nodes, jejunum and ilium, and spleen). Each animal was sampled on a different schedule to obtain an n=3 or n=4 at each time point while maintaining an appropriate interval between biopsies of individual animals.

Table 2.1: Bleed Schedule

Day/Week/ Month	**Procedure**						**Bleed Volume**
	RM-1	**RM-2**	**RM-3**	**RM-4**	**RM-5**	**RM-6**	
Day 0	Infection						
Week 1	Biopsy Bleed	Biopsy Bleed	Biopsy Bleed				5-10ml
Week 2	Euthanasia	Biopsy Bleed		Biopsy Bleed			5-10ml
Week 3		Euthanasia			Biopsy Bleed	Biopsy Bleed	5-10ml
Week 4			Biopsy Bleed	Biopsy Bleed	Biopsy Bleed	Biopsy Bleed	5-10ml
Month 2			Biopsy Bleed	Biopsy Bleed	Biopsy Bleed	Biopsy Bleed	5-10ml
Month 6			Euthanasia	Euthanasia	Euthanasia	Euthanasia	5-10ml

Additional samples were from previously completed studies (approved by the NIAID/VRC ACUC) of natural infection, vaccines, and healthy controls and are listed in a table at the end of this chapter.

2.2 Sample Processing

Blood draws were received in 15 mL EDTA anticoagulation tubes and spun at 1800 rpm for 15 minutes with braking and acceleration set to the lowest level. Plasma was removed from the surface and stored at -80C. Lymphocytes were removed and diluted in PBS, layered on top of Ficoll-Paque PLUS (GE Healthcare Life Sciences), and spun for 25 min at 1800 rpm with no braking or acceleration. The buffy coat containing lymphocytes was removed, washed 2x with 10 mL R10, counted and cryopreserved using sterile filtered foetal bovine serum (FBS) with 10% DMSO. R10 was prepared by sterile filtration of RPMI with 10% FBS and 1X Penicillin-Streptomycin-Glutamine (100x from Thermo Fisher).

Bone marrow biopsies were diluted twofold with PBS and layered over Ficoll-Paque PLUS and centrifuged fro 25 minute at 1800 RPM with no braking or acceleration.. The buffy coat containing lymphocytes was removed, washed 2x with 10 mL R10 by re-suspending cells and centrifuging at 1500 rpm for 7 minutes, counted and either immediately stained for FACS or cryopreserved using sterile filtered FBS with 10% DMSO.

For all tissues, fat and connective tissues were removed using surgical scalpels before tissues were minced and strained into single cell suspensions using a 70uM filter. Gut tissues were incubated with Collagenase-D (1mg/mL in RPMI+ PSG) shaking for 30 min at 37°C before washing twice in R10 and commencing fat/connective tissue removal and cell dissociation. After dissociation, cells were washed twice in R10 and counted before cryopreservation.

Single cell suspensions of splenocytes were incubated in 10 mL ACK Lysing Buffer (Thermo Fisher) for 7 minutes at room temperature, diluted with 40 mL PBS, washed (1500 rpm 5 minutes) and resuspended in 30 mL PBS. Diluted splenocytes were layered on 15 mL Ficoll-Paque and spun for 25 min at 1800 rpm with no braking or acceleration.. The buffy coat containing lymphocytes was removed, washed 2x with 10 mL R10, counted and cryopreserved using sterile filtered foetal bovine serum (FBS) with 10% DMSO.

2.3 Flow cytometry and fluorescence-activated cell sorting

For conventional flow cytometry, live cells were washed in 1mL PBS, centrifuged at 1500rpm for 5 minutes and resuspended in the residual volume after removing the supernatant. Samples were stained with 5ul of Aqua (1:40 dilution) for 5 minutes at room temperature in the dark and subsequently with an antibody cocktail (Table 2.6-Table 2.12 for different experiments) for 25 minutes at room temperature in the dark. Samples were washed with PBS and incubated with 200 uL of 2% paraformaldehyde for 30 minutes at 4°C in the dark before acquisition on a custom-built LSR II (BD Biosciences). For intracellular staining (ICS), cells were washed with 1 mL Perm/Wash Buffer (BD Biosciences) after surface staining, and incubated with CytoFix/CytoPerm Solution (BD Biosciences) for 15 min in the dark at room temperature. Cells were washed again in 1 mL Perm/Wash and intracellularly stained for 20 min at room temperature, then washed with 2 mL PBS and resuspended in 2% PFA. Alternatively, cells were fixed with FACSjuice (1X BD FACS Lysing solution, Tween 20 in PBS) on ice for 10 minutes, and washed with 1mL FACSwash (2% FBS in PBS) before ICS.

For fluorescence-activated cell sorting (FACS), cells were either sorted immediately after isolation or thawed from cryopreservation, washed in 10 mL R10, and resuspended in R10 at <10 million cells/mL. Cells were rested at 37℃ for 30-60 minutes before washing with PBS. Cells were gently resuspended and 5uL of 1:40 dilution of Aqua live/dead cell dye was added for 5 minutes at the dark at room temperature. The appropriate antibody cocktail was added and cells were stained for 25 minutes at room temperature in the dark. Cells were then washed in R10, filtered using 70uM sterile filters and sorted on a custom-built FACSAria II (BD Biosciences).

2.4 DNA and RNA Extraction

Quantitative polymerase chain reaction (qPCR) was used to measure SIV infection in sorted T cell subsets. Live cells were sorted into 300 uL R10 in 1.5 mL Eppendorf tubes, pelleted at maximum speed for 1 minute before supernatant was aspirated from the pellet. The pellet was resuspended in Proteinase K (Boehringer) at 100ul/mL in Tris-HCL (10mM, pH8)[324]. The volume of lysis solution varied by experiment but was a minimum of 10ul per 100,000 cells sorted. The lysis solution was incubated for a minimum of one hour at 56°C before inactivation at 95℃ for 10 minutes and storage at -20°C.

SIV RNA was measured using quantitative reverse transcription polymerase chain reaction (qRT-PCR). Two methods were used to extract RNA for qRT-PCR. In the first, cells were sorted into 200-300 uL R10, and resuspended in 1 mL RNAlater (Thermo Fisher) and frozen at -80C until RNA extraction. RNA was extracted using RNAqueous-Micro Total RNA isolation kits (Thermo Fisher) according to the manufacturer's instructions with the addition of an internal standard derived from a replication competent avian retrovirus (RCAS). Each sample had 5 ul of titred RCAS stock added at the lysis step, and recovery of RCAS was measured alongside SIV RNA to quantify loss of RNA during the extraction process [325].

In the second method, live cells were sorted into 300 uL FBS, and pelleted by centrifugation at 400rpm for 10 minutes. Supernatants were removed and cell pellets were resuspended in 300uL cold RNAzol (Molecular Research Center) before cryopreservation at -80℃. For RNA extraction, RNA lysates were thawed, and sterile water (0.4 volume of lysate) was added before shaking and incubating for 15 min at room temperature. Samples were centrifuged at 16,000 rpm for 15 minutes, and RNA was extracted from the top (liquid) phase. DNA in the bottom phase was preserved for later extraction and stored at -80℃. For RNA extraction, 1 uL glycogen (20 ug/ul) was added, and RNA was precipitated by adding an equal volume of isopropanol to the RNA lysate/water mixture and vortexing, followed by centrifugation at full

speed for 15 minutes. Supernatant was removed and the pellet was washed 2x with 750 ul 75% ethanol before drying the pellet and dissolving the RNA in sterile water. DNA was extracted from the bottom phase after thawing, solubilizing the pelleted with 900 ul DNAzol (Molecular Research Center) and precipitation with 2 ul glycogen (20 ug/ul) and 500 ul 100% ethanol. The solution was vortexed, incubated for 10 minutes at room temperature and centrifuged for 15 at full speed in a bench top microcentrifuge. The supernatant was discarded and the DNA pellet was washed 2x with 75% ethanol before drying and dissolution in 300 uL NaOH (8 mM) and neutralization by 24 uL HEPES (0.1M).

2.5 In Vitro infections of cell lines and primary cells

Virus stocks were obtained from the NIH AIDS Reagent Program (SIVmac251 32H) and from existing lab stocks of culture supernatants (SIVmac251, SIVsmE660). Two cell lines were used for SIV propagation and virus expansion: H9 and CEMx174, obtained from the NIH AIDS Reagent Program. Both cell lines were maintained at 37°C with 5% CO_2 and cultured in R10 at a density of $1x10^5$-$1x10^6$ cells/mL and split at 1:2 twice weekly. For infection of cell lines, 1x106 cells were pelleted (5 minutes 1500 rpm), resuspended in 1 mL viral supernatant, and spinoculated for 1 hour at 1200g at 32°C. Cells were then washed with 10mL R10 and resuspended in 10 mL R10 in a 25 mL culture flask at 37°C. After 2-3 days based on confluence of cells and observation of syncytia, 5 mL fresh R10 was added to the culture. Infection was confirmed by intracellular staining with p24 (Beckman Coulter), and supernatants from peak infection (3-5 days) were stored at -80°C.

Infection of primary cells was done using fresh PBMC. After isolation of lymphocytes using Ficoll-Paque, CD4 T cells were positively selected using the CD4+ T cell Isolation kit for nonhuman primates (Miltenyi Biotec), and washed in R10. Cells at $1x10^5$ cells/mL were activated using either Concanavalin A at 25ul/mL (Sigma-Aldrich) or phytohemaglutainin at 10 ug/mL (Fisher Scientific) in R10 overnight. The next morning cells were washed and resuspended in R10 + IL2 (50U/mL) and incubated for 24-48 hours. Frozen virus stocks were concentrated using 0.5 mL 100,000 MWCO filters and spun for 10 minutes at 14,000 rpm. Cells were gently pelleted and had supernatant was removed before re-suspension in the residual volume. Concentrated virus was layered on top of the cells and incubated for 2 hours at 37°C. Cells were washed with R10 and cultured in R10+IL2. Cultures were checked daily for syncytia, and Infection was measured by intracellular p24 staining.

2.6 Construction of absolute standards for qRT-PCR

RNA standards for qRT-PCR were created to measure absolute quantities of spliced and unspliced SIV RNA in infected cells. For gag RNA, previously published primers and probes were used [326] and for tat/rev RNA, primers and probes were designed across the D4/A7 splice site to uniquely amplify multiply spliced SIV RNA.

PBMCs were isolated from fresh blood draws and infected with SIVmac251 as described above, and total RNA was extracted and purified after 5 days of infection using the RNAqueous-Micro Total RNA isolation kits (Thermo Fisher) according to the manufacturer's instructions. cDNA synthesis was performed on the purified RNA by first incubating 1 uL of reverse primers for tat/rev (2 uM), 1 ug total RNA, 1 uL dNTP (10 mM) and 8 uL PCR-grade water at 65°C for 5 minutes followed by 1 minute on ice. Next, 4 ul 5x First Strand buffer (Invitrogen), 1 uL 0.1M DTT, and 1 uL RNAseout (Invitrogen) were added to the reaction mix and incubated at 42°C for 2 minutes. Next, 1 uL Superscript III (Invitrogen) was added along with 3uL PCR-grade water for a total volume of 20 uL, and the reaction mixture was incubated at 45°C for 50 minutes followed by 70°C for 15 minutes. The resulting cDNA was amplified by adding 5 uL 10X PCR Buffer (TRIS pH 8.4, KCL, Thermo Fisher), 1.5 uL 50 mM MgCl2, 1 uL 10 mM dNTPs, 1 uL each of the forward and reverse primers (10uM), 0.4 uL Platinum *Taq* DNA Polymerase (Thermo Fisher), and 38.1 uL PCR-grade water to 2 uL cDNA, and amplifying using the following thermocycling program: 94°C for 5 minutes, 35 cycles [94°C 30 seconds, 55°C 30 seconds, 72°C 45 seconds], 72°C 10 minutes, 4°C hold. Two microliters of the reaction mixture were run on a 2% agarose gel with a 100bp ladder in 1xTAE at 70 volts, and stained with 10uL SYBRGold to confirm amplification. The remaining reaction mixture was purified using the Wizard SV Gel and PCR Clean-up System (Promega) according manufacturer's instructions. DNA was quantified using a Nanodrop 1000 (Nanodrop).

Ligation was performed using p-Gem-T Easy Vector (Promega) according to manufacturer's instructions, including positive and negative controls. The ligation mixture (7.5 uL) was incubated with 50ul of DH5-a cells for 30 minutes on ice before each tube was heat shocked for 50 seconds at 42°C. 950 uL of S.O.C. media (Thermo Fisher) was added to each tube and incubated shaking at 37°C for 2 hours. LB-agar plates with ampicillin (100ug/mL), X-gal (100ug/mL) and IPTG (100 ug/mL) were prepared (Invitrogen) and 100 ul or 200 uL of transformed cells were spread over plates before an overnight incubation. The following day, colonies were picked and amplified with 2.5 uL HiFi buffer (Fisher Scientific), 1 uL 50 mM

$MgSO_4$, 0.5 uL 10 mM dNTPs, 1 uL M13 forward and reverse primers (20 mM), and 0.14 uL HiFi Taq DNA polymerase (Thermo Fisher) with the following thermocycling conditions: 94°C for 5 minutes, 35 cycles [94°C 30 seconds, 57 30 seconds], 68°C 3 minutes, 4°C hold. Samples were sent to Agencourt Bioscience for Sanger sequencing, and plasmids with the appropriate amplicons were grown to 100mL cultures in LB. Plasmids were purified using Plasmid Midiprep Kits (Qiagen) according to the manufacturer's instructions, linearized with restriction enzyme PstI (NEB), run on a 1% agarose gel and purified using the Wizard SV Gel and PCR Clean-up System (Promega). RNA synthesis was performed using the MEGAscript T7 Transcription Kit (Thermo Fisher) according to the manufacturer's instructions, and purified RNA was quantified on a Bioanalyzer High Sensitivity RNA chip. RNA was diluted in PCR-grade water to create standards of 2 copies to $2x10^7$ copies/uL and stored at -89°C.

2.7 PrimeFlow

All RNA flow cytometry experiments were done using a modified version of the eBiosciences PrimeFlow protocol. All wash steps were centrifuged for 800g for 5 minutes in a swinging bucket centrifuge unless otherwise noted. Cryopreserved cells were thawed in a 37°C water bath, resuspended in R10, washed, and rested for 30 min at 37C. In cases where the viability of the cells was below 40%, a dead cell removal was performed using the Dead Cell Removal Kit (Miltenyi Biotec). After centrifugation of cells at 300g for 5 minutes and incubation for 15 minutes at room temperature with Dead Cell Removal Microbeads at a ratio of 100 uL beads per 10^7 total cells, cells were added to an LS MACS column and eluted using 3 x 3 mL Binding Buffer. After elution, cells were washed in R10 and re-counted. All samples were split into three aliquots, with up to 5 million cells per aliquot: surface stain only, RNA probe, and no probe control. Cells were stained with 10 uL of a 1:40 dilution of LIVE/DEAD Fixable Aqua (Thermo Fisher) for five minutes and subsequently with the surface antibody panel for 25 minutes at 2-8°C (Table 2). Surface stain only samples were additionally stained with CD3 at this time before washing in PBS and fixation with 2% paraformaldehyde (PFA) and storage at 2-8°C. Probe samples were washed in 1 mL PBS and resuspended in Fixation Buffer 1 for 30 minutes at 2-8°C, washed two times in 1X Permeabilization Buffer and stained with CD3 for 25 minutes at 2-8°C. Following CD3 staining, samples were washed with Permeabilization Buffer, fixed using 1mL Fixation Buffer 2, and incubated at room temperature in the dark for 60 minutes. Cells were washed in RNA Wash buffer and transferred to low-bind 1.5 mL Eppendorf tubes included in the kit.

Target probes were thawed and diluted 20x in RNA Target Probe diluent pre-warmed to 40°C and 100 uL were added to cell suspensions and incubated for 2 hours at 40°C. No-probe

controls had only 100 uL Target Probe Diluent added. Samples were washed 2x with RNA Wash Buffer and gently resuspended in 100 uL residual buffer before addition of 100 uL RNA PreAmp mix pre-warmed to 40°C. Cells were briefly vortexed and incubated at 40°C for 1.5 hours, washed three times with RNA Wash Buffer, and resuspended in 100uL residual buffer before adding 100 uL RNA Amp Mix. Each sample was briefly vortexed and incubated at 40°C for 1.5 hours and subsequently washed twice with RNA Wash Buffer. RNA Label probes were diluted 100x in RNA Label Probe Diluent and 100uL of the probe-diluent mix was added to each sample before incubation for 1 hour at 40°C. Samples were then washed twice with RNA wash buffer and washed and resuspended in 1mL PBS before acquiring on the flow cytometer.

For all RNA flow samples, the surface-only stain was used to confirm gating and frequencies of T cell subsets in the full probe staining protocol samples. No-probe controls that underwent the full RNAFlow staining protocol were used to set the FMO gates for the probe-positive samples. Compensation on the flow cytometer was performed using antibody-bound beads and the PrimeFlow Compensation Kit (eBiosciences)

To compare the staining of the ACD RNAscope probes and amplifiers with the eBiosciences PrimeFlow kit, some samples were incubated ACD reagents for the target probe and amplification steps (all samples underwent identical surface and ICS staining and fixation/permeabilization). ACD samples were incubated with ACD target probes (2 drops) for 2 hours at 40°C, washed, and subsequently incubated with 2 drops each of Amp1 (30 minutes), Amp2 (15 minutes), Amp3 (30 minutes), and Amp4 (15 minutes) with two washes in between each incubation. Samples were washed, resuspended in PBS, and acquired on the flow cytometer.

2.8 SIV sequencing

Full length SIV genomes were sequenced from plasma RNA and from DNA and RNA extracted from sorted lymph node T cell subsets. Lymph node samples were stained with the panel in Table 2.8, and seven populations were sorted: live singlet CD14- CD16- CD20- CD3+ CD4+ lymphocytes were divided into naïve (CCR7+ CD45RA+ CD28+ CD95-) and central memory (CD45RA- CD28+ CD95+), while two populations were sorted from CD3+ CD4- cells: CXCR5+ and CXCR5-. Central memory T cells were further divided based on expression of CXCR5 and PD1, with four populations (CXCR5++PD1++, CXCR5+PD1+, CXCR5+PD1-, and CXCR5-PD1-). Each of the seven populations was sorted from lymph

node biopsies at 2 weeks 4 weeks, 10 weeks, and 24 weeks post infection and both DNA and RNA were extracted using protocols described above.

Plasma samples were diluted to 10 mL with PBS and clarified by centrifugation at 200g for 10 minutes with no braking. Clarified plasma was overlaid onto 25 mL of a 1:10 dilution of OptiPrep (Sigma Aldrich) in PBS and centrifuged at 20,000 rpm for 3 hours at 4°C with no braking. Plasma and Optiprep were removed and the viral pellet was resuspended in 400 uL cold RNAzol RT (Molecular Research Center) and frozen at -80°C. For RNA extraction, lysates were thawed, and sterile water (0.4 volume of lysate) was added before shaking and incubating for 15 min at room temperature. Samples were centrifuged at 16,000 rpm for 15 minutes, and 85% of the RNA-containing supernatant was removed to a fresh tube. 0.5% volumes of BAN (Molecular Research Center) was added to the supernatant, mixed, and incubated for 5 minutes at room temperature. The mixture was centrifuged at 12,000g for 10 min at room temperature. 90% of this supernatant was added to a fresh tube (containing 1ul glycogen and 4 uL water), and RNA was precipitated from this mixture by the addition of 1 volume of isopropanol, vortexing, and incubation for 10 minutes at room temperature. RNA was pelleted by centrifugation at full speed for 15 minutes, and the pellet was washed twice with 750 ul 75% ethanol. The pellet was dried and dissolved in sterile water and frozen at -80°C.

The protocol for amplification, sequencing, and assembly was adapted from Gall et al[327], with primers designed to generate four overlapping amplicons covering nearly the full SIV genome, ranging from 1791 bases to 2859 bases. RNA was amplified using Superscript III One-Step RT-PCR with Platinum Taq High Fidelity kit (Invitrogen), with 12.5 uL 2X Reaction Mix, 0.5 ul SuperScriptIII/HiFI Platinum Taq, 1 ul primer mix (5 uM), 1.5 uL 5 mM $MgSO_4$, 4 uL template RNA, and 5.5 uL PCR-grade water, and thermocycling conditions of 50°C 60 minutes, 94°C 2 minutes, 2 cycles [94°C 15 seconds, 60°C 60 seconds, 68°C 4 minutes], 2 cycles [94°C 15 seconds, 58°C 60 seconds, 68°C 4 minutes], 41 cycles [94°C 15 seconds, 55°C 60 seconds, 68°C 4 minutes], 68°C 10 minutes, 4°C hold. DNA was amplified using Platinum Taq High Fidelity kit (Invitrogen), with 2.5 ul 10x HiFI PCR buffer, 0.5 ul 10 mM dNTP, 0.1 ul HiFi Platinum Taq, 1 ul primer mix (5 uM), 1 ul 5 mM $MgSO_4$, 5 ul template DNA, and 14.9 uL PCR-grade water, and thermocycling conditions of 94°C 2 minutes, 30 cycles [94°C 15 seconds, 55°C 60 seconds, 68°C 4 minutes], 10°C hold. After amplification, samples were run on a 1% agarose gel for 60 minutes at 100V and stained with SYBRGOld to visualise bands. Samples were multiplexed and sequenced using Illumina

HiSeq 2500 Rapid Run with 250 base pair paired end reads by the Sequencing Core at the Wellcome Trust Sanger Institute.

2.9 B cell sorting and 5' RACE sequencing

A SIV mac239 gp140 foldon trimer probe (molecular weight 140,000) containing an Avitag (Avidity) and a His tag was provided by Rosie Mason and Mario Roederer[328] and was conjugated to Streptavidin-APC (Invitrogen). After diluting 10 ug of the probe into PBS and protease inhibitor to a final volume of 117 uL, 8 uL of Streptavidin-APC was added in 5 aliquots, incubating the probe-SA-APC mix for 10 minutes shaking at 4C between each addition. Conjugated protein was stored at 4C until use.

For immunoglobulin heavy chain repertoire sequencing, serial biopsies of three animals (ZF61, ZF76, and ZG13) were used to analyse the evolution of the humoral response. PBMC, bone marrow, inguinal lymph node biopsies from 4 weeks, 10 weeks, and 24 weeks post infection were sorted and sequenced. For each tissue, samples were thawed (with the exception of bone marrow, which was sorted fresh), stained with the B cell sorting panel, and five populations of B cells were sorted into 200ul fresh R10., After excluding T cells and monocytes, 19+20+ B cells were sorted on expression of IgM, IgD, IgG, CD27, and gp140 to obtain naïve, IgG memory, IgM memory, antigen-specific IgG memory, and antigen-specific IgM memory B cells. Cells were pelleted, resuspended and stored in RNAlater until RNA extraction. RNA was extracted using µMACS mRNA Isolation Kit (Miltenyi Biotec) and cDNA was synthesized by first incubating 8 ul RNA with 1uL 5' CDS Oligo dT primers (12 uM) for 1 minute at 72°C and cooling on ice, before the addition of 1uL Superscript II (Invitrogen), 1uL SMARTer oligo (12uM), ul dNTP (1mM), 1 ul DTT (20 mM), 1 ul RNAseOUT (Invitrogen), and 3.5 ul 5X RT Buffer (Clontech) and incubating at 42°C for 120 minutes followed by 70°C for 120 minutes. cDNA synthesis was cleaned up with using Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions. PCR amplification of BCR heavy chains was performed using Kapa HiFi Real-Time PCR Library Amplification kits (Kapa) with Ig constant region primers (5 uL of 2uM) and 5PIIA primer (0.7 ul of 12uM). Samples were sequenced on the Illumina MiSeq platform with paired end reads of 300 base pairs.

2.10 Single cell B cell sequencing

The protocol for single cell sorts for B cell heavy chain sequencing was adapted from Tiller *et al* [329]. Individual cells were sorted into 96 well plates, either dry or containing 0.5 uL RNAse Out (Thermo Fisher), 5 ul 5X buffer (Invitrogen, for Superscript III), 1.25 ul DTT (Thermo Fisher), 0.0625 uL Igepal (Sigma), 0.26 uL *S. cerevisiae* carrier RNA (Sigma) and 13 uL PCR

qualified water (Thermo Fisher); and immediately frozen at -80. For cDNA synthesis, dry plates had the above mixture added to each well after thawing on ice immediately before cDNA synthesis. Reverse transcription was performed after adding 0.3 uL Superscript III (Invitrogen), 3ul random hexamers (GeneLink, 150 ng/uL), and 2 uL dNTP (10mM, Thermo Fisher) to each well, with a PCR program of 10 minutes at 42°C, 10 minutes at 25°C, 60 minutes at 50°C, 5 minutes at 94°C, and a hold at 4°C. 26 uL PCR-grade water was added to each well.

Nested PCR was used to amplify the resulting cDNA, using 4uL cDNA, 2.5 ul 10X Buffer (Qiagen), 0.5 uL dNTP (10mM, Thermo Fisher), 0.5 uL $MgCl_2$ (25 mM, Qiagen), 5 uL Q-solution (Qiagen), 0.5 uL Forward Outer Primer mixture (25 mM), 0.5 uL Reverse Outer Primer mixture (20 mM), 0.2 uL HotStart Taq Plus (Qiagen), and 11.3 uL PCR-grade water (Qiagen) per well. Thermocycling conditions were: 94°C for 5 minutes, 50 cycles of [94°C for 30 seconds, 55°C for 30 seconds, 70°C for 60 seconds], 70°C for 10 minutes, and hold at 4°C. The second PCR used 3 ul of the first round PCR with a mastermix of 2.5 ul 10X Buffer (Qiagen), 0.5 uL dNTP (10mM, Thermo Fisher), 5 uL Q-solution (Qiagen), 0.5 uL Forward Inner Primer mixture (25 mM), 0.5 uL Reverse Inner Primer mixture (20 mM), 0.2 uL HotStart Taq Plus (Qiagen), and 12.8 uL PCR-grade water (Qiagen) per well. Thermocycling conditions were: 94°C for 5 minutes, 50 cycles of [94°C for 30 seconds, 60°C for 30 seconds, 70°C for 60 seconds], 70°C for 10 minutes, and hold at 4°C. Amplification of heavy chains was verified using gel electrophoresis, with 5 uL of second-round PCR product diluted with 1uL 6X Orange Gel Loading Dye (NEB) run on a 2% agarose gel alongside a 100 bp ladder (NEB) at 100 volts for 20-30 minutes. Gels were visualised using SYBR Gold Gel Stain (Thermo Fisher) diluted 10,000X in TAE for 30 minutes and positive wells were sent for Sanger sequencing to GATC.

2.11 Bioinformatics analysis

For each sequencing run of pooled immunoglobulin heavy chain samples, sequences were demultiplexed into sample groups. FASTQ files were generated from raw BAM files and quality controlled using QUASR7.01[330] with a median sequence quality cutoff of 34. Paired reads that passed the quality control filtering and overlapped by at least 30 base pairs were joined, and primers were trimmed from the full-length sequences. Each sequence was aligned using BLAST[331] against a database of IgJ genes and removed if it did not match a reference J sequence with an e-value of 0.1 or less. Open reading frames were calculated for each sequence and sequences passed if the length was above a given threshold and had a V gene

match to the reference database with an e-value of 1. Sequences were clustered based on sequence similarity with CD-HIT with a sequence identity threshold of 0.85. Diversity statistics including Gini Index and Renyi index for the clusters and vertices were calculated as in Bashford-Rogers *et al* [332], as were network plots.

Virus sequences were assembled *de novo* using IVA [333] and underwent IVA's quality control pipeline using SIV reference genomes (accession numbers KC522232, M33262 and M76764). All contigs mapping to reference genomes underwent hypermutation analysis using Hypermut [334], and contigs that were significantly hypermutated ($p<0.01$) were removed from the assemblies. In cases where overlapping contigs mapped to the genome, sequences were manually merged to create a single consensus sequence for each sample. All sequences were first aligned using MAFFT [335] and manually edited. Pairwise distances were calculated in MEGA [336]. Maximum likelihood trees were generated using FastTree [337] with a generalized time-reversible (GTR) model of nucleotide evolution and rapid bootstrapping with 1,000 resamples, and trees were rooted on reference sequence KC522232.

2.12 Statistical Analysis

All statistics were calculated in GraphPad Prism. In Chapter 3, the Wilcoxon matched-pairs test was used to evaluate differences in population proportions between paired samples from individual animals. Multiple T-tests were used to analyze differences between germline divergence, CDR3 length, diversity measures (Gini and Renyi index, vertices in largest cluster), and proportion of shared sequences between animals and populations. Chi-squared tests were used to calculated differences in V and J gene distribution proportions. T-tests were used to analyze differences in computational derived isoelectric point, hydrophobicity, aliphatic index, and charge in CDR3 regions. In chapter 4, one-way Anova tests were used to analyze expression patterns of surface receptors on T cell subpopulations, copy numbers of SIV DNA and RNA in T cell subpopulations, and the proportion of SIV RNA probe-positive cells. In chapter 5, one-way Anova tests were used to analyze pairwise distances between sets of sequences from different animals, T cell subpopulations, and nucleic acid origins (DNA, RNA or plasma).

2.13 PCR Primers

Table 2.2: SIV DNA/RNA quantification primers

Primers	Sequence
M13 Forward	GTAAAACGACGGCCAG
M13 Reverse	CAGGAAACAGCTATGAC
SIV tatrev_F04a	GAACTCCGAAAAAGGCTAAGGCTAATACA
SIV tatrev_F04c	GAACTCCGAARAAGRCTAAGRCTAATMCA
SIV tatrev Left D4A7	TCCGAAAAAGGCTAAGGCTA
SIV tatrev_R04d	CCGTCTCRTTCTTTGCCTTCTCTGGTT
SIV tatrev_R04c	CCKTCTCCTTCTTCTCCTTCTTTG GTT
SIV tatrev Right D4A7	CTCCACCGTCTCCTTCTTTG
Gag Forward	GTCTGCGTCATPTGGTGCATTC
Gag Reverse	CACTAGKTGTCTCTGCACTATPTGTTTTG
Probes (for qRT-PCR)	5' FAM 3' BHQ-1
SIV tatrev_P04a	CTGCATCAAACAAACCCATATCCAACAGGACC
SIV tatrev_P04c	CTGCATCAAACAAATCCCTATCCACAAGGRCC
Gag Probe	CTTCPTCAGTKTGTTTCACTTTCTCTTCTGCG

Table 2.3: SIV full genome sequencing primers

Primer Mixes	Forward	Reverse
Amp1	TGTCTTTTATCCAGGAAGGGGTA	CTCTAATTAACCTACAGAGATGTTTGT, CTCTAGTTAACCTACAGAGATGTTTGT
Amp2	AAAATTGAAGCAGTGGCCATTAT	TTCTTATGAGCTCTCGGGAACCT
Amp3	GCTTTACAGCGGGAGAAGTG	ATTGCAGAACCTGCCGTTG
Amp4	CAGTCACCATTATGTCTGGATTG	GAATACAGAGCGAAATGCAGTG

Table 2.4: 5' RACE BCR sequencing primers

CDS Oligo dT	TTTTTTTTTTTTTTTTTTTTTTTTVN
SMARTER IIA	AAGCAGTGGTATCAACGCAGAGTACATrGrGrG
5PIIA	AAGCAGTGGTATCAACGCAGAGT
RhHu IgM	GAGCGAGGGGGAAAAGGGTTGGGGCGGATGCA
RhHu IgG	GCCAGGGGGAAGACCGATGGGCCCTTGGTGGA
RhHu IgD	CTGATATGATGGGGAACACATCCGGAGCCTTGG

Table 2.5 B cell single cell sorting/sequencing primers

Outer Forward L1 Primers - Mix @ equal volumes to 25uM		Inner Forward SE Primers - Mix @ equal volumes to 25uM	
5'VH1.L1	ATGGACTKGACCTGGAGG	5'VH1A.SE	TGGCAGCAGCTACAGGTGC
5'VH2.L1	ATGGACACGCTTTGCTCC	5'VH1B.SE	TGACAGCAGCTACAGGCGC
5'VH3A.L1	ATGGAGTTKGGGCTGAGC TG	5'VH1C.SE	TGGCAGCAGCAACAGGCAC
5'VH3B.L1	ATGGAGTTTGKRCTGAGC TGG	5'VH2.SE	GTCCCGTCCTGGGTCTTGTC
5'VH3C.L1	ATGGAGTCRTGGCTGAGC TGG	5'VH3A.SE	GCTGTTTGGAGAGGTGTCCAGT GTG
5'VH3D.L1	ATGGAGTTTGTGCTGAGTT TGG	5'VH3B.SE	GCCATATTAGAAGGTGTCCAGT GTG
5'VH4.L1	ATGAAGCACCTGTGGTTC	5'VH3C.SE	GCTCTTTTGAAAGGTGTCCAGT GTG
5'VH5A.L1	ATGGGGTCAACTGCCATC	5'VH3D.SE	GCTATTTTAAGAGGTGTCCAGT GTG
5'VH5B.L1	ATGGGGTCCACCGTCACC	5'VH3E.SE	GCTATTTTAAAAGGTGTCCAGT GTG
5'VH6.L1	ATGTCTGTCTCCTTCCTCA	5'VH4.SE	AGCTCCCAGATGGGTCYTGTCC
5'VH7.L1	ATGGACCTCACCTGGAGC	5'VH5.SE	GCTGTTCTCCARGGAGTCTGTG
5'VH2.L1	ATGGACACGCTTTGCTCC	5'VH6.SE	GGCCTCCCATGGGGTGTC
		5'VH7A.SE	GCAGCAACAGGTGCCCACTC
		5'VH7B.SE	GCAGCAACAGGCACCCACTC
Reverse Outer Primer - Use @ 20uM		Reverse Inner Primer - Use @ 20uM	
3'IgG Outer	GGAAGGTGTGCACGCCGCT GGTC	3'IgG Inner	GTTCAGGGAAGTAGTCCTTGA C

2.14 Flow Cytometry Panels

Table 2.6: In vitro infection panel

Surface marker	Fluorophore	µl per test	Clone	Company	Notes
Live/Dead	Aqua	5		Thermo Fisher	
CD4	FITC	2.5	L200	BD Biosciences	
CD3	Cy7APC	3	SP34-2	BD Biosciences	
CD8	BV785	2	RPA-T8	BioLegend	
p24	PE	0.5	KC57	BioLegend	intracellular

Table 2.7: RNA Flow panel and probes

Surface marker	*Fluorophore*	*µl per test*	*Clone*	*Company*	*Notes*
CD4	BV605	4	L200	BD Biosciences	
CD8	BV785	2	RPA-T8	BioLegend	
PD1	BV711	2	EH12.2H7	BioLegend	
CXCR5	PE	5	MU5UBEE	eBiosciences	
CD28	TRPE	7	CD28.2	Beckman Coulter	
CD95	Cy5PE	7	DX2	BD Biosciences	
CD3	Cy7APC	3	SP34-2	BD Biosciences	ICS in probe samples
RNA Probe		*Channel*	*Probe Type*	*Company*	
SIVmac239 gag, pol, env		Ax647	Type 1	eBiosciences	
B2M macaca Mulatta		Ax488	Type 4	eBiosciences	
SIVmac239 SENSE C1		Ax488	C1	ACD	
DAPB		Ax647	C2	ACD	

Table 2.8: SIV full genome sequencing panel

Surface marker	Fluorophore	µl per test	Clone	Company	
CCR7	Ax700	2	150503	BD Biosciences	
CD14	BV510	1.5	M5E2	BioLegend	
CD16	BV510	1	3G8	BioLegend	
CD8	BV510	2	RPA-T8	BioLegend	
CD20	BV570	2.5	2H7	BD Biosciences	
CD45RA	Qd655	1	MEM-56	Invitrogen	
PD-1	BV785	2	EH12.2H7	BioLegend	
gdT	APC	2	B1	BioLegend	
CD3	Cy7APC	2	SP34-2	BD Biosciences	
CD4	FITC	4	L200	BD Biosciences	
CXCR5	PE	5	MU5UBEE	eBiosciences	
CD28	TRPE	7	CD28.2	Beckman Coulter	
CD95	Cy5PE	7	DX2	BD Biosciences	

Table 2.9: Lymph node Tscm staining panel

Surface marker	Fluorophore	µl per test	Clone	Company	
CCR7	BV421		150503	BD Biosciences	
CD14	BV510	1.5	M5E2	BioLegend	
CD16	BV510	1	3G8	BioLegend	
CD4	BV605		L200	BD Biosciences	
CD8	BV785	2.5	RPA-T8	BioLegend	
gdT	APC	2	B1	BioLegend	
CXCR3	Ax700	3	1C6	BD Biosciences	
CD3	Cy7APC	2	SP34-2	BD Biosciences	
CD45RA	FITC		MEM-56	Invitrogen	
CXCR5	PE	5	MU5UBEE	eBiosciences	
CD127	Cy7PE	5	R34.34	Beckman Coulter	
CD28	TRPE	7	CD28.2	Beckman Coulter	
CD95	Cy5PE	7	DX2	BD Biosciences	

Table 2.10 Lymph node Chemokine staining panel

Surface marker	Fluorophore	µl per test	Clone	Company	
ICOS	PacBlue	2.5	C398.4A	BioLegend	
CD14	BV510	1.5	M5E2	BioLegend	
CD16	BV510	1	3G8	BioLegend	
CD8	Qd565	0.5	3B5	Invitrogen	
CD4	BV605	2.5	L200	BD Biosciences	
PD-1	Biotin	0.25	EH12.2H7	eBioscience	20 min 37c
Streptavidin	BV650	0.5		BD Biosciences	
CCR6	Ax647	2.5	G034E3	BioLegend	
CXCR3	Ax700	3	1C6	BD Biosciences	
CD3	Cy7APC	2	SP34-2	BD Biosciences	
SLAM	FITC	5	A12 (7D4)	BioLegend	
CXCR5	PE	5	MU5UBEE	eBiosciences	
CCR4	Cy7PE	3	TG6	BioLegend	
CD28	TRPE	7	CD28.2	Beckman Coulter	
CD95	Cy5PE	7	DX2	BD Biosciences	

Table 2.11: Lymph node Activation staining panel

Surface marker	Fluorophore	µl per test	Clone	Company	
ICOS	PacBlue	2.5	C398.4A	BioLegend	
CD14	BV510	1.5	M5E2	BioLegend	
CD16	BV510	1	3G8	BioLegend	
CD20	BV570	2.5	2H7	BioLegend	
CD25	BV605	5	BC96	BioLegend	
PD-1	Biotin	0.25	EH12.2H7	eBioscience	20 min 37c
Streptavidin	BV650	0.5		BD Biosciences	
CD8	BV785	1.5	RPA-T8	BioLegend	
gdT	APC	2	B1	BioLegend	
CCR7	Ax700	2	150503	BD Biosciences	20 min 37c
CD3	Cy7APC	2	SP34-2	BD Biosciences	
CD4	FITC	4	L200	BD Biosciences	
CXCR5	eFluor710	5	MU5UBEE	eBiosciences	
SLAM	PE	2	A12 (7D4)	BioLegend	
CD127	Cy7PE	5	R34.34	Beckman Coulter	
CD28	TRPE	7	CD28.2	Beckman Coulter	
CD95	Cy5PE	7	DX2	BD Biosciences	

Table 2.12: B cell sorting Panel

Surface marker	Fluorophore	µl per test	Clone	Company
CD3	Cy7APC	2	SP34-2	BD Biosciences
CD14	Qd800	1	Tuk4	Invitrogen
CD19	ECD	10	J3-119	Beckman Coulter
CD20	BV570	2.5	2H7	BD Biosciences
CD27	Cy5PE	3	CLB 27/1	Invitrogen
CD138	FITC	20	MI15	BD Biosciences
IgD	PE	0.5	polyclonal	Southern Biotech
IgM	V450	2.5	G20-127	BD Biosciences
IgG	Ax680	1.5	G18-145	VRC
Gp140	APC	1		VRC

2.15 Animal Specimens

Table 2.13: Animal Study Protocols and Samples

ASP	Animals	Tissues	Infection time	Virus
238.2	8-23, 08D213, 8-46, ZG08, ZG32, ZG72	Axillary LN, PBMC	6 months	SIVsmE660
206	8E-9, DBME, 7F-8, AV33X	Inguinal LN, PBMC	2 years	SIVsmE660
238.3	8-40, 8-76, 8-118, 8-130, 08D0236, 8-8, 8C-8, 8-160, 8-88, ZC26, ZG38, 8-28, 8-160	Axillary LN, Inguinal LN, PBMC	Various	SIVmac251
274	4140, FLI, CK3P, 4131, IAC	PBMC	Uninfected	N/A
417	ZF76, ZG13, ZF61, ZH29, ZF32, ZF60	Mesenteric, Axillary, Inguinal lymph nodes; ilium/jejenum, bone marrow, PBMC	5 days – 6 months	SIVmac251
	6049, 4975. 4939, 4755, 4875, 4931	Axillary LN, PBMC	acute	SIVmac251
	7-20, FA4L, 4359, 08D272, 0G-4, 4359,	Axillary LN, Inguinal LN, PBMC	uninfected	N/A

2.15 IGH Sequencing Results – cells sorted and sequencing reads retained

Table 2.14: IgG cells sorted, reads retained in initial QC

Sample ID	cells sorted	% reads retained	N raw reads (1)	N raw reads (2)	N joined reads	N reads primer matched	N reads w/t ORF
ZF61_4_BM_gp140_IgG_A	400	0.001884037	119977	106155	66358	681	2
ZF61_4_BM_gp140_IgG_B	400	1.232868655	110509	85005	8965	1232	1048
ZF61_4_BM_gp140_IgG_0	400	58.01973777	741355	709300	413969	412846	411534
ZF61_4_BM_mem_IgG_0	35000	45.65305531	683304	656070	316233	312567	299516
ZF61_4_INGLN_gp140_IgG_0	660	39.4418844	702418	672298	294667	290508	265167
ZF61_4_INGLN_mem_IgG_0	46200	42.26670078	492954	472987	204526	203691	199916
ZF61_4_LN_gp140_IgG_A	660	0.04477126	100976	73708	40712	5626	33
ZF61_4_PBMC_gp140_IgG_A	925	3.501358608	109804	93478	39263	3532	3273
ZF61_4_PBMC_gp140_IgG_0	925	0.007790975	144337	141189	124166	34752	11
ZF61_4_PBMC_gp140_IgG_B	925	0.003057668	112359	98114	73251	72090	3
ZF61_4_PBMC_mem_IgG_0	81000	45.36098407	974316	938227	795788	529735	425589
ZF61_10_BM_gp140_IgG_A	800	0.56186009	137670	99135	5976	2733	557
ZF61_10_BM_gp140_IgG_B	800	0.892450991	118893	92330	15752	3404	824
ZF61_10_BM_gp140_IgG_0	800	2.025189468	701671	675838	57448	14476	13687
ZF61_10_BM_mem_IgG_0	32000	61.08007179	773744	759968	479731	475077	464189
ZF61_10_INGLN_gp140_IgG_0	810	56.51454317	855255	815675	494905	474658	460975
ZF61_10_INGLN_mem_IgG_0	50000	51.4240943	807108	775370	411241	409566	398727
ZF61_10_LN_gp140_IgG_A	810	1.740061571	132649	104594	12140	4464	1820
ZF61_10_PBMC_gp140_IgG_A	9500	0.370091304	64622	45124	6969	2950	167
ZF61_10_PBMC_gp140_IgG_0	9500	0.021945028	201659	200501	197998	19299	44
ZF61_10_PBMC_gp140_IgG_B	9500	0.769555715	125848	108894	65890	41921	838
ZF61_10_PBMC_mem_IgG_0	188000	41.67404232	1348768	1309489	1062072	733686	545717
ZF61_24_BM_gp140_IgG_A	136	3.254085305	108154	84755	5294	3692	2758
ZF61_24_BM_gp140_IgG_B	136	0.006732258	114908	103977	89585	15920	7
ZF61_24_INGLN_gp140_IgG_0	14100	68.18606818	641237	626738	434448	429513	427348
ZF61_24_INGLN_mem_IgG_0	65000	56.85390583	601522	579633	342120	339165	329544
ZF61_24_LN_gp140_IgG_0	14100	11.76154276	118432	94843	26441	18238	11155
ZF61_24_PBMC_gp140_IgG_A	32000	16.45421771	118508	108732	81016	61264	17891
ZF61_24_PBMC_gp140_IgG_B	32000	0.002799186	115235	107174	95853	84959	3
ZF61_24_PBMC_gp140_IgG_0	32000	47.18404473	471801	459986	230870	227254	217040
ZF61_24_PBMC_mem_IgG_0	100000	53.63241337	414706	405956	231709	225740	217724
ZF76_4_BM_gp140_IgG_A	500	9.389990734	107048	85261	10998	9038	8006
ZF76_4_BM_gp140_IgG_B	500	16.6828574	104763	88528	38613	37921	14769
ZF76_4_BM_gp140_IgG_0	500	70.45376107	423952	414337	299380	295839	291916

ZF76_4_BM_mem_IgG_0	4000	32.91392335	491113	478132	179278	177828	157372
ZF76_4_INGLN_gp140_IgG_0	450	29.11383823	500083	486717	243324	242354	141702
ZF76_4_INGLN_mem_IgG_0	50000	68.79591918	537592	527149	373652	371491	362657
ZF76_4_LN_gp140_IgG_A	450	13.23685527	102042	81953	14548	12249	10848
ZF76_4_PBMC_gp140_IgG_A	2100	5.297656646	123582	96571	14271	9855	5116
ZF76_4_PBMC_gp140_IgG_B	2100	6.571236926	173712	135667	27772	18750	8915
ZF76_4_PBMC_gp140_IgG_0	2100	47.20007315	112920	109360	93108	64564	51618
ZF76_4_PBMC_mem_IgG_0	55600	48.48522424	513786	500008	410303	289601	242430
ZF76_10_BM_gp140_IgG_A	184	0.043593563	46236	27527	5230	205	12
ZF76_10_BM_gp140_IgG_B	184	3.912032593	115539	93762	28397	10159	3668
ZF76_10_BM_mem_IgG_0	6000	53.04104973	672534	658835	365650	361358	349453
ZF76_10_INGLN_gp140_IgG_0	725	57.06777349	529120	517518	298854	296230	295336
ZF76_10_INGLN_mem_IgG_0	50000	67.62548418	505729	495680	348797	346978	335206
ZF76_10_LN_gp140_IgG_A	725	8.961356972	118561	95153	17448	10635	8527
ZF76_10_PBMC_gp140_IgG_A	12000	0.005880969	117750	102024	58592	20385	6
ZF76_10_PBMC_gp140_IgG_0	12000	0.064155594	60458	59231	55209	26905	38
ZF76_10_PBMC_gp140_IgG_B	12000	2.01740865	129938	110290	56517	40234	2225
ZF76_24_BM_gp140_IgG_B	70	4.909163605	102127	80034	5900	4796	3929
ZF76_24_BM_gp140_IgG_A	70	3.521197307	126854	99824	10010	6151	3515
ZF76_24_BM_gp140_IgG_0	70	63.62772105	441900	431268	282637	280395	274406
ZF76_24_BM_mem_IgG_0	800	59.1310805	474951	464600	278631	276939	274723
ZF76_24_INGLN_gp140_IgG_0	17000	31.65349207	164618	52247	17065	16608	16538
ZF76_24_INGLN_gp140_IgG_1	1254	7.731235338	99975	97617	81093	48864	7547
ZF76_24_INGLN_mem_IgG_0	55000	46.32120699	1105623	1063800	512170	505928	492765
ZF76_24_LN_gp140_IgG_A	1254	1.960541954	136024	105175	12858	3355	2062
ZF76_24_LN_gp140_IgG_0	1254	11.53451181	123547	100299	15820	13040	11569
ZF76_24_PBMC_gp140_IgG_A	1974	0.009918077	152339	100826	20087	39	10
ZF76_24_PBMC_gp140_IgG_C	1974	9.330800764	119346	93186	19483	13869	8695
ZF76_24_PBMC_gp140_IgG_B	1974	1.405850211	125156	106697	54250	38232	1500
ZF76_24_PBMC_gp140_IgG_2	1974	0.045345348	106041	103649	89705	62083	47
ZF76_24_PBMC_gp140_IgG_1	1974	44.28984793	100019	97458	84068	64396	43164
ZF76_24_PBMC_gp140_IgG_0	1974	58.87757475	992251	958781	581353	571875	564507
ZF76_24_PBMC_mem_IgG_1	54700	0.043983647	403441	393328	331640	242733	173
ZF76_24_PBMC_mem_IgG_2	54700	52.32840263	544841	520142	414829	308589	272182
ZF76_24_PBMC_mem_IgG_0	54700	44.5326107	1179843	1140270	528254	522679	507792
ZG13_4_BM_gp140_IgG_B	1700	3.702078003	113751	93407	14999	12618	3458
ZG13_4_BM_gp140_IgG_A	1700	0.108841945	99347	85445	54853	39434	93
ZG13_4_BM_gp140_IgG_0	1700	45.47953232	570796	546105	257202	249367	248366
ZG13_4_BM_mem_IgG_0	5000	30.75053452	599593	571073	186402	179925	175608
ZG13_4_INGLN_gp140_IgG_0	970	47.6935442	825715	790065	387750	380081	376810
ZG13_4_INGLN_mem_IgG_0	60000	41.96192362	535690	512181	217656	216167	214921
ZG13_4_LN_gp140_IgG_A	970	0.461759885	66946	50026	2904	247	231
ZG13_4_PBMC_gp140_IgG_A	2900	1.985192417	132811	98328	6254	5451	1952
ZG13_4_PBMC_gp140_IgG_B	2900	5.460049944	136495	103717	18933	14086	5663
ZG13_4_PBMC_gp140_IgG_0	2900	51.5819399	101041	98139	86358	55491	50622
ZG13_4_PBMC_mem_IgG_0	67000	48.45189811	747244	717976	598099	414839	347873
ZG13_10_BM_gp140_IgG_A	5500	6.484749617	126118	101145	11887	7272	6559
ZG13_10_BM_gp140_IgG_B	5500	7.579800647	135972	109053	18256	14768	8266
ZG13_10_BM_gp140_IgG_0	5500	47.81176324	1322477	1287315	721398	696443	615488
ZG13_10_BM_mem_IgG_0	212000	43.55508872	1278845	1248132	740165	725606	543625
ZG13_10_INGLN_gp140_IgG_B	1000	58.75039813	96930	94190	78574	57497	55337
ZG13_10_INGLN_gp140_IgG_0	1000	54.64746791	1209590	1191466	727343	720974	651106
ZG13_10_INGLN_mem_IgG_0	100000	54.94394739	1210253	1186653	738513	730876	651994
ZG13_10_LN_gp140_IgG_A	1000	2.087242521	122367	94862	7928	2249	1980
ZG13_10_PBMC_gp140_IgG_B	37000	0.004169272	34951	23985	6333	21	1
ZG13_10_PBMC_gp140_IgG_A	37000	6.017580879	121510	94307	13108	12335	5675
ZG13_10_PBMC_gp140_IgG_0	37000	45.73507917	269423	262033	210227	141328	119841
ZG13_10_PBMC_mem_IgG_0	200000	46.34787112	1333097	1272477	1047959	774583	589766
ZG13_10_PBMC_mem_IgG_B	200000	38.41656485	1704079	1643528	1406317	928315	631387
ZG13_24_AXLN_mem_IgG_10KA	10000	50.55769213	228356	218938	176574	120548	110690
ZG13_24_BM_gp140_IgG_B	300	7.569688464	121254	97613	11543	7530	7389
ZG13_24_BM_gp140_IgG_A	300	7.064758725	133472	106642	12728	10116	7534
ZG13_24_BM_gp140_IgG_0	300	46.99070108	1332411	1301549	685296	678571	611607
ZG13_24_BM_mem_IgG_0	4150	51.13825587	1326928	1280951	745934	723801	655056
ZG13_24_INGLN_gp140_IgG_0	7750	38.25261325	671930	639721	264991	246063	244710
ZG13_24_INGLN_mem_IgG_0	100000	41.73335327	553406	528467	227611	225031	220547
ZG13_24_LN_gp140_IgG_0	7750	3.802326647	137393	109170	11861	6690	4151
ZG13_24_PBMC_gp140_IgG_B	70000	0.022375432	99974	71507	13399	45	16
ZG13_24_PBMC_gp140_IgG_A	70000	2.383507186	64100	56849	22005	15901	1355
ZG13_24_PBMC_gp140_IgG_0	70000	26.66268896	456347	434086	123347	119083	115739
ZG13_24_PBMC_mem_IgG_0	100000	25.76331385	415257	390796	124791	112353	100682

Table 2.15: IgM cells sorted, reads retained in initial QC

Sample ID	cells sorted	% reads retained	N raw reads (1)	N raw reads (2)	N joined reads	N reads primer matched	N reads w/t ORF
ZF61_4_INGLN_gp140_IgM_0	34	0.01807764	129439	127229	122091	104517	23
ZF61_4_LN_gp140_IgM_A	34	0.430322322	119015	107129	84101	39556	461
ZF61_4_LN_mem_IgM_A	15000	4.529062949	118067	91233	19254	17181	4132
ZF61_4_PBMC_gp140_IgM_A	1000	2.450521227	130296	99285	10284	3720	2433
ZF61_4_PBMC_gp140_IgM_0	1000	52.83587416	103975	100886	85005	61428	53304
ZF61_4_PBMC_mem_IgM_A	40000	1.075792788	131251	106712	40715	37959	1148
ZF61_4_PBMC_mem_IgM_0	40000	53.63846837	416788	400828	332064	239733	214998
ZF61_10_INGLN_gp140_IgM_0	1870	22.521391	79147	76434	39339	17258	17214
ZF61_10_LN_gp140_IgM_A	1870	7.00964865	140725	109549	24439	19136	7679
ZF61_10_LN_mem_IgM_A	24000	3.636905522	150122	117435	25889	21557	4271
ZF61_10_PBMC_gp140_IgM_A	335	0.011114591	105516	98969	87842	82842	11
ZF61_10_PBMC_mem_IgM_A	100000	5.939079987	130843	104629	31663	20279	6214
ZF61_10_PBMC_mem_IgM_0	100000	52.42990417	855963	827831	681478	508541	434031
ZF61_24_LN_gp140_IgM_B	5600	10.08725974	145989	116434	32254	25899	11745
ZF61_24_PBMC_gp140_IgM_A	8800	0.527262223	131716	122899	103828	30932	648
ZF61_24_PBMC_gp140_IgM_0	8800	44.2447732	205874	200218	172336	116366	88586
ZF61_24_PBMC_mem_IgM_A	100000	0.970246542	107099	86679	20855	14042	841
ZF76_4_INGLN_gp140_IgM_0	600	0.059412912	138400	136334	126834	33071	81
ZF76_4_LN_gp140_IgM_A	600	2.148417754	98172	79128	30224	25305	1700
ZF76_4_LN_mem_IgM_A	22000	9.167259958	105737	82871	24241	18152	7597
ZF76_4_PBMC_gp140_IgM_A	1900	6.037209651	132807	106854	16932	7413	6451
ZF76_4_PBMC_gp140_IgM_0	1900	53.51379949	93472	90728	77759	57463	48552
ZF76_4_PBMC_mem_IgM_A	26000	5.014683505	135048	103177	16019	11979	5174
ZF76_4_PBMC_mem_IgM_0	26000	0.116352867	109081	103994	84859	51047	121
ZF76_10_INGLN_gp140_IgM_0	4400	2.009672339	279470	268601	215527	96834	5398
ZF76_10_LN_gp140_IgM_A	4400	0.107061605	123046	112085	94350	88911	120
ZF76_10_LN_mem_IgM_A	22000	7.807399843	144139	109678	21660	17777	8563
ZF76_10_PBMC_gp140_IgM_A	280	2.471986236	147233	112177	17548	3322	2773
ZF76_10_PBMC_gp140_IgM_0	280	55.50623645	224393	215828	182209	132002	119798
ZF76_10_PBMC_mem_IgM_A	30000	0.998697351	144503	85211	46932	16607	851
ZF76_24_LN_gp140_IgM_B	17000	4.553755523	115376	84875	11227	9541	3865
ZF76_24_LN_gp140_IgM_A	17000	0.005794462	135291	120805	97818	71493	7
ZF76_24_MESLN_gp140_IgM_0	17000	36.27283463	384022	375143	338785	235368	136075
ZF76_24_PBMC_gp140_IgM_A	22000	1.496638351	144045	99356	9487	3970	1487
ZF76_24_PBMC_mem_IgM_A	100000	0.056174515	146926	106810	47140	813	60
ZG13_4_INGLN_gp140_IgM_0	500	0.042529256	159022	145782	120777	30500	62
ZG13_4_LN_gp140_IgM_A	500	2.686193768	129151	109523	60092	46481	2942
ZG13_4_LN_mem_IgM_A	32400	0.008619649	81546	58007	15905	29	5
ZG13_4_PBMC_gp140_IgM_A	1080	6.64446936	110843	88299	12344	6415	5867
ZG13_4_PBMC_gp140_IgM_0	1080	37.02056217	119161	115552	97176	53926	42778
ZG13_4_PBMC_mem_IgM_A	59200	7.055342406	142352	110512	22040	17719	7797
ZG13_4_PBMC_mem_IgM_0	59200	52.7107682	537816	516883	418865	304753	272453
ZG13_10_LN_gp140_IgM_A	3600	6.238246198	149277	111666	27174	20842	6966
ZG13_10_LN_mem_IgM_A	50000	5.619294705	134564	104070	18982	15012	5848
ZG13_10_PBMC_gp140_IgM_A	1150	2.227893923	127588	98344	21526	11018	2191
ZG13_10_PBMC_gp140_IgM_0	1150	39.57855669	456437	437022	350074	238211	172967
ZG13_10_PBMC_mem_IgM_A	200000	7.717391304	113517	90160	22737	16991	6958
ZG13_10_PBMC_mem_IgM_0	200000	51.39466952	1454324	1405315	1158894	871178	722257
ZG13_24_LN_gp140_IgM_B	1620	5.350380433	110683	86349	17148	11779	4620
ZG13_24_LN_gp140_IgM_A	1620	8.686488893	130737	101065	21311	17139	8779
ZG13_24_LN_mem_IgM_0	100000	7.576042162	118001	90509	18602	14203	6857
ZG13_24_MESLN_gp140_IgM_0	1620	47.49855615	210381	202583	166904	118045	96224
ZG13_24_PBMC_gp140_IgM_A	1500	6.065806512	245762	192291	45280	31622	11664
ZG13_24_PBMC_gp140_IgM_0	1500	49.55568653	91203	88226	72143	53090	43721
ZG13_24_PBMC_mem_IgM_A	100000	9.070920457	142108	115975	40517	14620	10520

3 Evolution of immunoglobulin heavy chain repertoires throughout SIV infection

3.1 Introduction and aims

The antigen-specific humoral response to a pathogen develops rapidly after infection, and through the iterative processes of somatic hypermutation and affinity maturation, generates a highly specific pool of antibody secreting cells and memory B cells. Next-generation sequencing of antibodies offers the ability to sample and sequence tens of thousands of B cells throughout infection[194], and combined with serial sampling across tissues yields a detailed view of the evolution of the antigen-specific antibody repertoire. Since the discovery of the first broadly neutralizing antibodies (bnAbs) to HIV in the early 1990s, hundreds of HIV-specific antibodies ranging in breadth and potency have been catalogued and studied. Certain features are highly represented within the bnAbs: prominent use of germline V genes including VH1-2, VH1-69 and VH1-4, long HCDR3 regions (24-32 residues), and high levels of mutation from germline V genes ranging from 20-30% or higher)[132,219,228]. High-throughput sequencing of the antibody repertoire and phylogenetic analysis have also been used to identify VRC01 class antibodies that bind epitopes in the CD4 binding site of HIV envelope[338]. Here, we use the SIV-rhesus macaque model to identify features of the maturing memory B cell response to SIV during on-going viral infection, examining V-D-J gene usage, mutations and features of heavy chain CDR3 regions, and the diversity and clonality of the memory response.

The aims of this chapter are to:

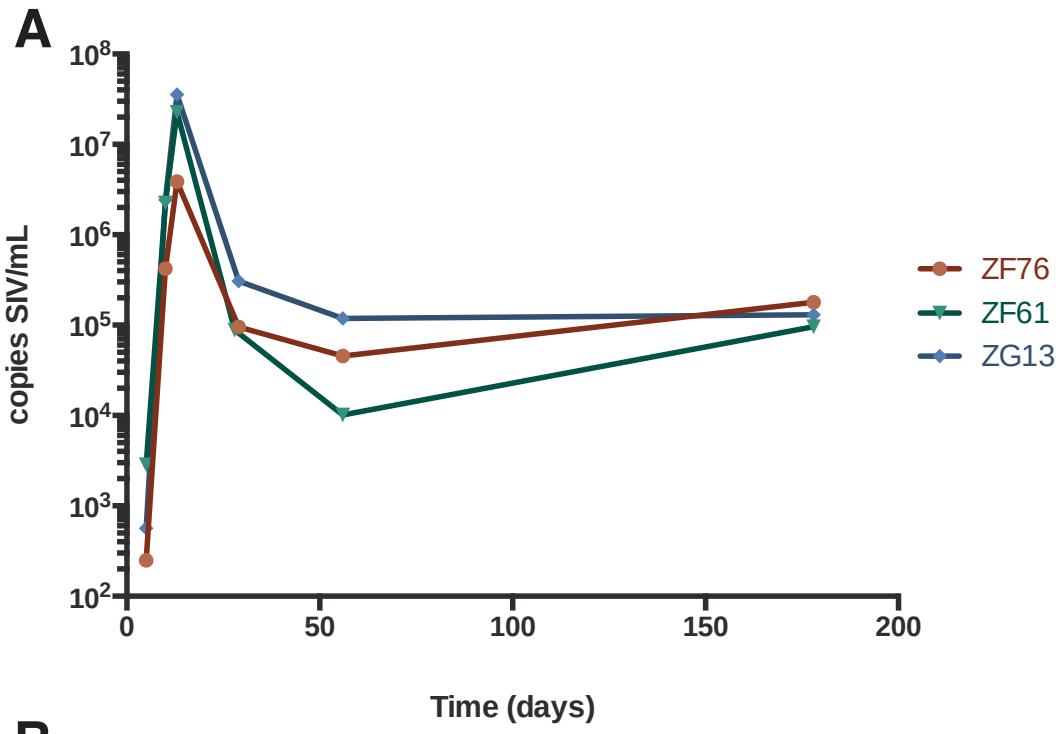
- Measure the proportion of SIV-specific B cells in lymph nodes, bone marrow, and peripheral blood throughout infection
- Track the changes in genetic properties of SIV-specific B cells change over the course of infection, and compare IgG and IgM SIV-specific antibodies
- Compare the physiochemical features of SIV-specific and bulk memory B cells
- Trace antibody linages across tissues and over time, and measure public versus private repertoires
- Compare 5’ RACE repertoire sequencing to single cell primer based methods

3.2 Results

3.2.1 Infection Model and plasma neutralization

Three rhesus macaques (Indian, colony-bred) were infected with intravenous dose of SIV mac251 at a 1:3000 dilution. Blood draws and lymph node and bone marrow biopsies were taken at 4 weeks, 10 weeks, and 24 weeks post infection. Plasma viremia peaked at 13 days post infection ($3.9x10^6$, $2.3x10^7$, and $3.5x10^7$ copies/mL) (Fig 3.1 A). Viral loads were steady at the three time points sampled, with $9.5x10^4$, $8.8x10^4$, and $3.1x10^5$ copies/mL at 4 weeks; $4.5x10^4$, $1.0x10^4$, and $1.2x10^5$ copies/mL at 10 weeks, $1.8x10^5$, $9.6x10^4$, and $1.3x10^5$ copies/mL at 24 weeks. These values are within the expected ranges of plasma viremia during peak infection, viral setpoint, and on-going replication in the absence of antiretroviral therapy[339].

Plasma neutralization was measured using a TZM-BL assay[223] with serial dilutions of plasma, with ID_{50} values representing reciprocal inhibitory dilution (Fig. 3.1 B). The SIVsmE660 isolate used here is a tier 1A envelope (meaning it is highly sensitive to neutralization) while the SIVmac251 envelope in this assay is Tier 2 (moderate resistance to neutralization). Three animals used for immunoglobulin repertoire analysis showed moderate neutralization of SIVsmE660 at 4 weeks post infection (ID_{50} of $5.8x10^3$ -$1.9x10^4$) and no neutralization of SIVmac251 (the virus used for infection) at 4 weeks. By 24 weeks post infection, all animals showed higher neutralization of SIVsmE660 (ID_{50} of $1.7x10^5$ to above $3.3x10^5$, the limit of detection in this assay) and low-level neutralization of SIVmac251 (ID_{50} of 58-746). Previous studies have shown that SIVmac251 founder viruses induce autologous neutralizing antibodies at 5 to 8 months post infection[273], while neutralization of sensitive viruses such as smE660 typically develops earlier[340]. Thus, the levels of plasma neutralization of SIVmac251 and smE660 at 24 weeks are evidence of a typical humoral response in the animals in this study.



B

Animal ID	Infection period (weeks)	ID_{50} SIVsmE660.CP3C-P-A8 (tier 1A)	ID_{50} SIVmac251.30 (tier 2)
ZF61 4 weeks	4	5,888	<20
ZF76 4 weeks	4	6,010	<20
ZG13 4 weeks	4	19,493	<20
ZF61 24 weeks	24	>327,680	273
ZF76 24 weeks	24	176,279	58
ZG13 24 weeks	24	>327,680	746

Values represent reciprocal inhibitory dilution at 50% (ID_{50}) or 80% (ID_{80}) neutralization

Note: Neutralization curves for SIVmac251.30 exhibited a plateau at approx. 40% neutralization

Fig. 3.1 Plasma viremia and neutralization. (A) Plasma viral loads for three animals (ZF76, ZF61, and GZ13) throughout infection. (B) Plasma neutralization of SIVsmE660 and SIVmac251 at 4 weeks and 24 weeks post infection. Good neutralization (high reciprocal inhibitory dilution) indicated in red, moderate neutralization in yellow, and no detectable neutralization in white.

3.2.2 Sorting of SIV-specific B cell throughout infection

To evaluate the SIV-specific repertoire of immunoglobulins generated in response to infection, we used a trimeric SIVgp140 protein probe to sort SIV-specific B cells from live cell samples of PBMCs, lymph nodes, spleen, and bone marrow cells[328]. The SIVmac239 gp140 foldon trimer probe used in this assay to isolate SIV-specific B cells with FACS has been previously validated and published [328]. All cells were sorted based on staining as Aqua (live/dead cell marker) CD14 (monocytes) CD3 (T cells) negative and CD19 CD20 positive B cells (Fig. 3.2 A). SIV gp140 protein probe-positive and protein probe-negative class switched IgG memory B cells (IgM- IgD- IgG+ CD27+) and probe-positive and probe-negative memory IgM B cells (IgM+ IgG- CD27+) were sorted at 4, 10 and 24 weeks post infection from all tissues (Fig 3.2 B).

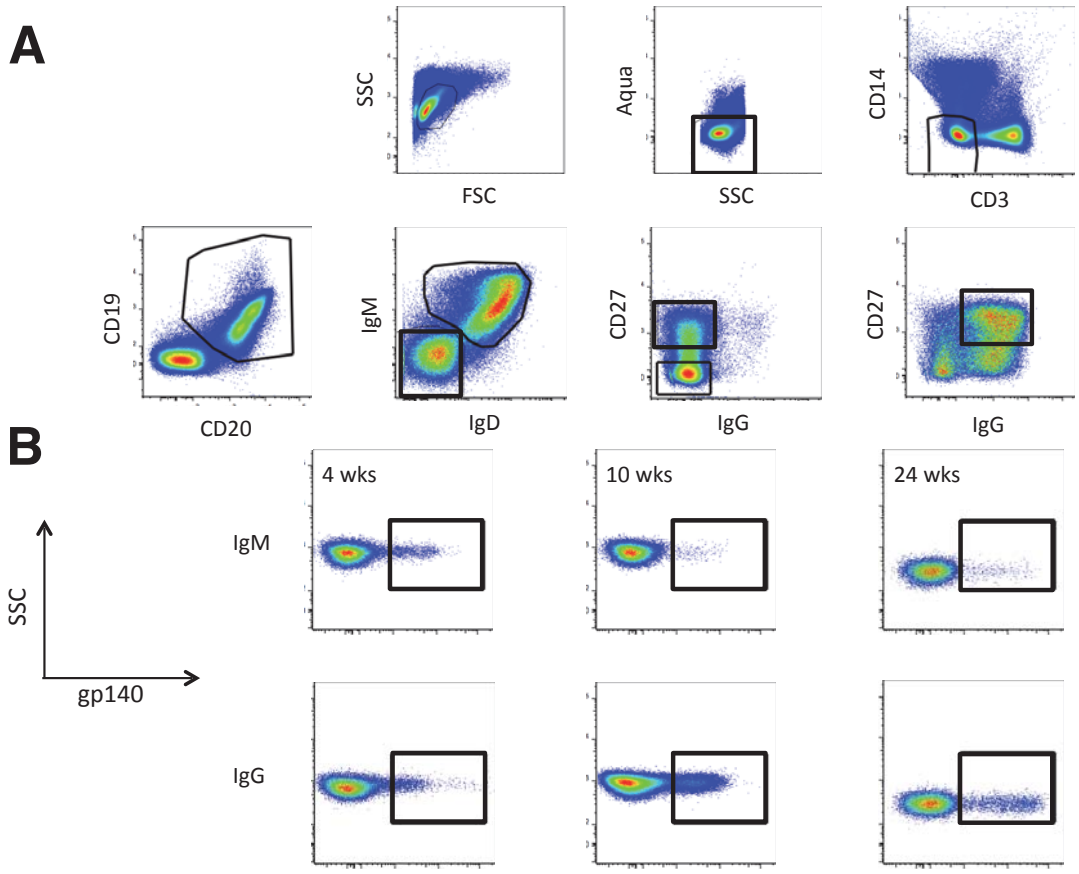


Fig 3.2 Sorting schematic and SIVgp140 probe staining (A) Gating strategy for isolating B cells from lymph nodes, PBMC, bone marrow, and spleen. SIV-specific and non-specific B cells were sorted after gating on lymphocytes, live (Aqua negative), CD14- CD3- CD19+ CD20+ B cells, and either IgM+IgD+CD27+ (memory IgM) or IgM-IgD-CD27+IgG+ (memory IgG). (B) SIVgp140 probe staining on memory IgM (IgM+IgD+IgG-CD27+) top trow, and class switched IgG memory B cells (IgM-IgD-IgG+CD27) bottom row, at 4 weeks, 10 weeks, and 24 weeks post infection.

The frequencies of naïve B cells measured as a percentage of total B cells were unchanged over the 24 weeks of infection studied here in bone marrow and in PBMC (Fig. 3.3 A, B), but in lymph nodes dropped significantly in all animals at 10 weeks post infection to a mean of 27% before increasing at 24 weeks to 44% (Fig. 3.3). IgM CD27+ memory cells in PBMC decreased in all animals, from a mean of 29% to 15% by 24 weeks (Fig. 3.3 A). The percentage of IgM CD27+ memory cells was unchanged in lymph nodes (7%-15%) and declined, though not statistically significantly, in bone marrow from 5.6% to 0.39% of total B cells. The proportion of IgG CD27+ B cells increased in PBMC at 10 weeks post infection of 2 of the three animals (to a mean of 45% of total B cells), but declined by 24 weeks post infection (Fig. 3.3 A). Lymph node IgG CD27+ B cells were statistically unchanged over the infection period studied here, while bone marrow IgG CD27+ cells declined from 7.3% to 0.15% of all B cells.

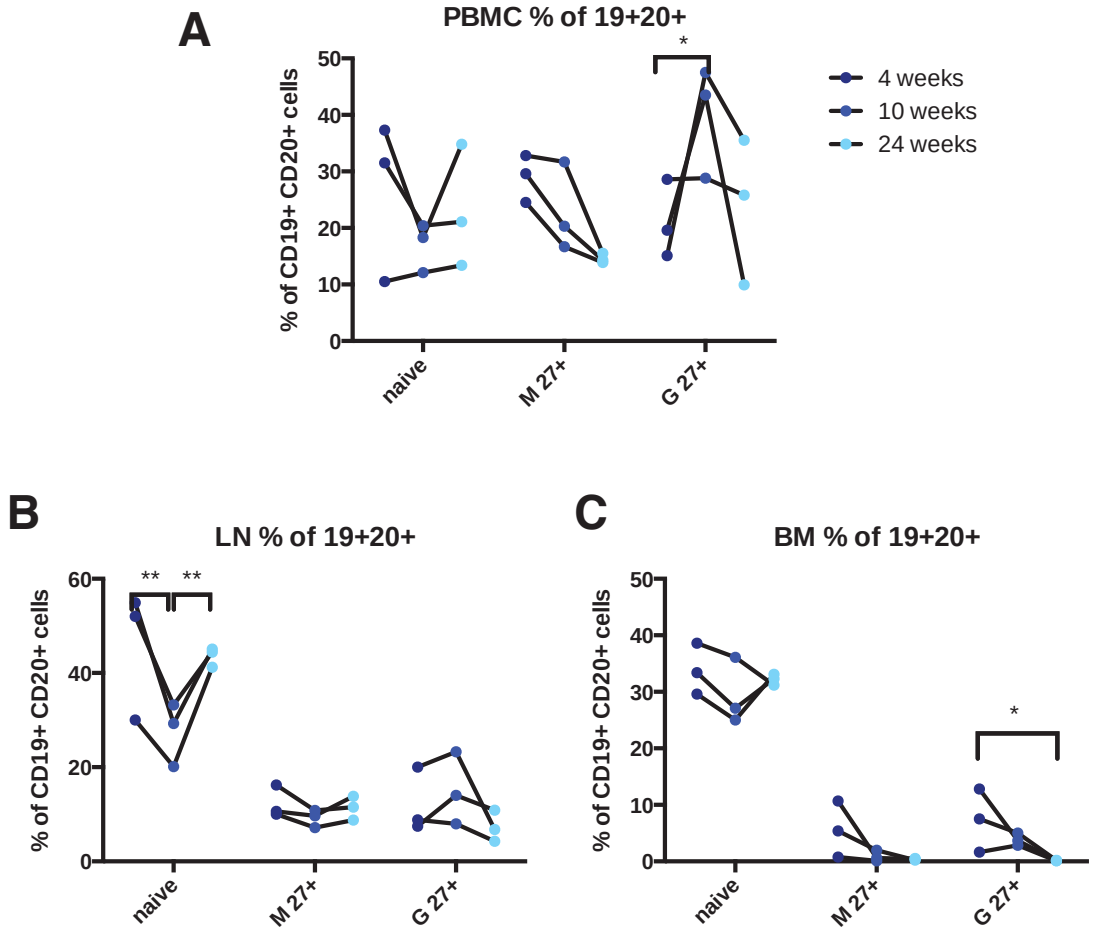


Fig 3.3 Proportions of B cell subpopulations. (A) Naïve, CD27+ IgM memory, and CD27+ IgG memory B cells in PBMC at 4 weeks, 10 weeks, and 24 weeks post infection as a proportion of all B cells. (B) Naïve, CD27+ IgM memory, and CD27+ IgG memory B cells in inguinal lymph nodes at 4 weeks, 10 weeks, and 24 weeks post infection as a proportion of all B cells (C) Naïve, CD27+ IgM memory, and CD27+ IgG memory B cells in bone marrow at 4 weeks, 10 weeks, and 24 weeks post infection as a proportion of all B cells. Significant differences were determined using a paired Student's T-test.

The proportion of SIV-specific B cells increased throughout infection. Measured as a percentage of IgG memory B cells, SIV protein probe positive cells increased (though not statistically significantly) in PBMC and lymph nodes in all three animals, from a mean of 2.5% to 8.6% in PBMC and 5.3% to 12.3% in lymph nodes (Fig. 3.4 A). In bone marrow, one animal had an increase in IgG SIV-specific cells, one animal showed no change, and another animal had an increase in SIV-specific cells at 10 weeks but a lower proportion at 24 weeks than at 4 weeks. The proportion of IgM SIV-specific cells was unchanged in lymph nodes and PBMC (ranging from 0.2% to 1.2% in PBM and 0.9%-1.8% in lymph nodes), while in bone marrow one animal had an increase in SIV-specific cells at 10 weeks followed by a decrease at 24 weeks (the same animal that had a peak in IgG SIV-specific cells) (Fig. 3.4 B)

When measured as a percentage of total B cells, SIV-specific IgG cells were measured at a significantly higher proportion of PBMC in two of three animals (Fig 3.4 C), with a mean of 0.5% at 4 weeks and 2.5% at 24 weeks. In lymph nodes, the pattern of the proportion of SIV-specific cells was the same in all three animals, increasing from a mean of 0.5% at 4 weeks to 1.7% at 10 weeks before decreasing to 0.8% at 24 weeks. The proportion of SIV specific IgG cells in bone marrow declined in all animals, from a mean of 0.6% at 4 weeks to 0.02% at 24 weeks. IgM SIV-specific cells declined slightly in all animals, from a mean of 0.31% at 4 weeks to 0.17% at 24 weeks (with a dip at 10 weeks to 0.04%) in PBMC (Fig. 3.4 D). Lymph node ISV-specific IgM cells declined slightly in two animals (from 0.1% to 0.06%) but increased in one animal to 0.6% at 24 weeks. Bone marrow SIV-specific IgM B cells similarly declined in two animals to 0.004%, but increased to 0.05% in one animal (the same animal that had an increase in lymph node IgM SIV-specific cells. Overall, the SIV gp140-specific response is dominated by IgG B cells when measured by absolute quantity of B cells and proportion of memory populations but has contributions, especially in bone marrow, from IgM memory cells.

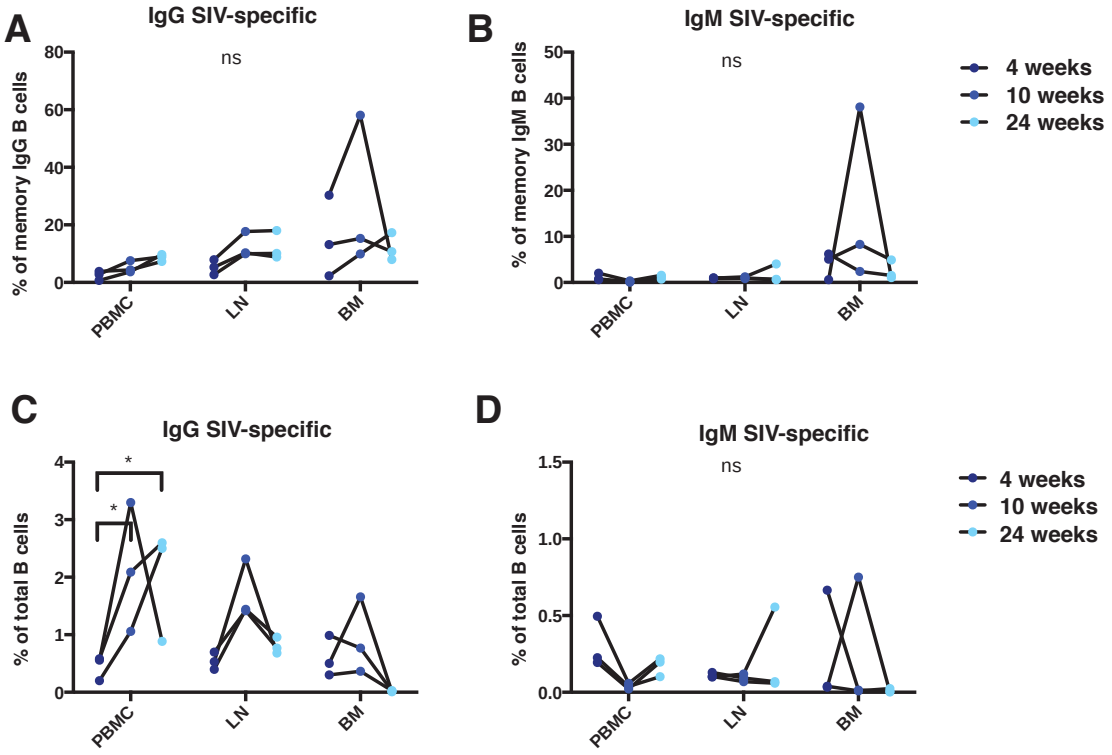


Fig 3.4 Antigen-specific B cells throughout infection. (A) Antigen-specific IgG B cells in lymph nodes, bone marrow and PBMC as a proportion of memory IgG B cells at 4, 10, and 24 weeks post infection. (B) Antigen-specific IgM B cells in lymph nodes, bone marrow and PBMC as a percentage of memory IgM B cells at 4, 10, and 24 weeks post infection. (C) Antigen-specific IgG B cells in lymph nodes, bone marrow and PBMC as a percentage of total B cells at 4, 10, and 24 weeks post infection. (D) Antigen-specific IgM B cells in lymph nodes, bone marrow and PBMC as a percentage of total B cells at 4, 10, and 24 weeks post infection. Significant differences were determined using a paired Student's T-test.

3.2.3 Next-generation sequencing of memory, naïve, and antigen-specific B cells

Quantity alone of antigen-specific cells does not determine the quality (including the breath, specificity, and diversity of epitopes seen) of the humoral response. To further evaluate the immunoglobulin heavy chain repertoire generated in response to infection and to compare the SIV-specific response to the non-SIV-specific memory cells, we extracted RNA from sorted cell populations and amplified immunoglobulin heavy chains using 5' RACE PCR. Briefly, after total cDNA synthesis from mRNA and IgH-specific amplification with constant region primers (either IgM, or IgG), samples were sequenced on the Illumina MiSeq platform using 2x300 base paired end reads. To analyse the sequencing results, a bioinformatics pipeline was adapted from a published pipeline[332] used for BCR analysis of blood disorders and adapted for 5' RACE sequencing methods and rhesus macaque genes (Fig 3.5). After quality control with QUASR7.0[330] and primer trimming, reads were joined using overlapping regions to obtain full length heavy chain sequences. Joined sequences of minimum length 200 underwent filtering for V-gene and J-genes identification, productive open reading frames, and were subsequently collapsed into identical reads to yield a final set of unique heavy chain sequences. Reads were only retained if they had a median read quality (Phred score) of 34, overlapped with the matched paired read by at least 30 identical base pairs, matched a reference database of V genes by nucleotide BLAST with a e-value of 1 and J genes with an e-value of 0.1, and had a productive open reading frame. V, D, and J genes were mapped to a recently generated draft rhesus macaque Ig heavy chain reference database [341]. E-values describe random background noise, with higher e-values (closer to 1) allowing a lower threshold for a match, while high e-values (close to 0) ensure more stringent matches. In this case, higher e-values (1 and 0.1) were used to allow highly mutated J genes to be matched to the nearest reference sequence without discarding highly mutated sequences. A median of 45% of reads retained after quality control, joining, primer matching, and open reading frame filtering, though there was some variability in the quality of the MiSeq runs due to low-quality kits. In high-quality sequencing runs, samples had a median of 45% of reads retained (interquartile range of 38%-51%), while in low-quality sequencing runs, a median of 3.1% of reads were retained (IQR of 0.5%-6.6%). If the number of retained reads did not exceed the number of cells sorted, samples were re-amplified from cDNA or mRNA and re-sequenced to obtain coverage of the entire sample. A list of the samples sorted, the number of cells from each sample, the raw number of sequencing reads obtained for that sample, and the number of reads retained after each step of filtering is detailed in the Materials and Methods Tables 2.14 and 2.15.

The number of cells sorted from each population varied widely, with several hundred SIV-specific IgG and IgM B cells isolated from acute PBMC and lymph node samples, and over 200,000 cells sorted from certain non-specific memory B cell samples. Low cell number samples often required additional PCR cycles to meet the DNA threshold for sequencing, and in many cases had a lower percentage of reads retained in the final analysis. In calibrating the read processing and quality control pipeline, there is a balance between retaining what could be genuine reads that happen to fall below a strict quality threshold and discarding lower-quality reads that may in fact be biologically relevant. Even with strict quality control measures, there were still samples from which the final number of unique reads exceeded the absolute input number of cells. Other groups have noted this in their immunoglobulin sequencing[342], which could be attributed to contamination, chimeric PCR products formed during amplification, or imprecision in the cell sorter counting. However, the vast majority of samples had equal or fewer unique heavy chain sequencing after filtering and quality control. Nonetheless, caution must be used in interpreting the results when only small numbers of cells are isolated from a given population, with the knowledge that the sequences represent a small subset of the total repertoire.

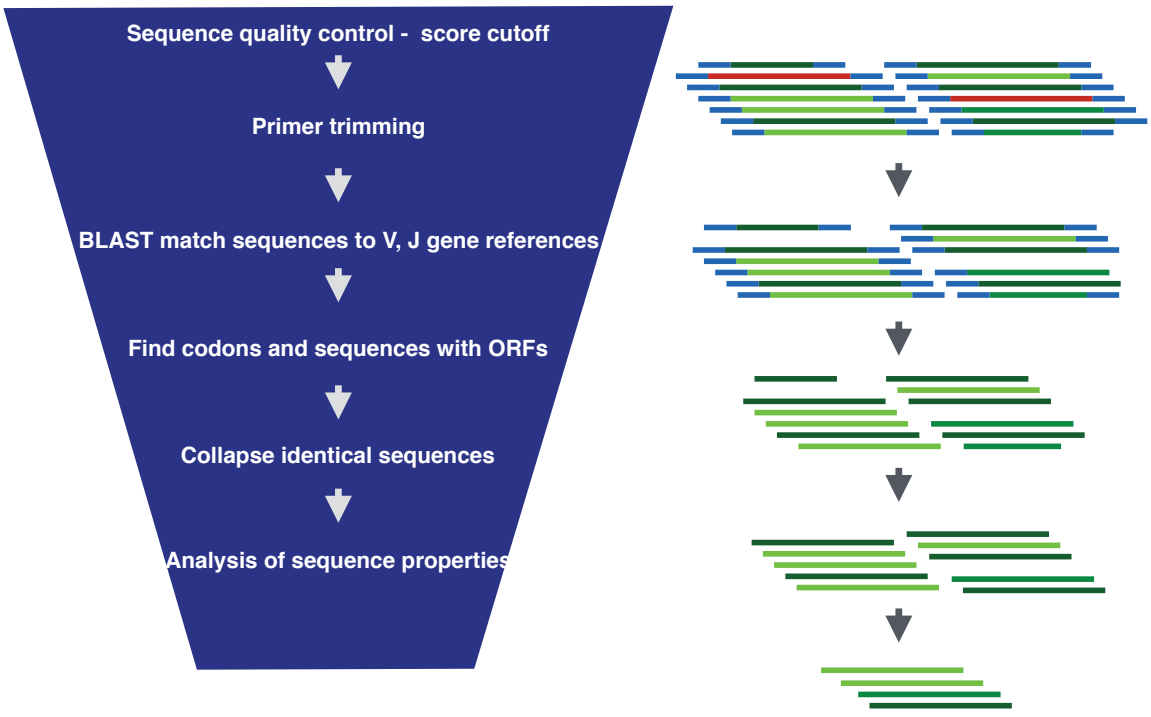


Fig. 3.5 Schematic of BCR Bioinformatics Pipeline for Sequence Analysis. Sequencing reads were filtered from the analysis if they did not meet a minimum PHRED (quality of sequencing) score, had primers trimmed, were BLAST matched to reference genomes (and eliminated if there was not a match, allowing for mutation from germline), and checked for open reading frames (with truncated sequences and sequences with early stop codons eliminated). Identical sequences were collapsed to yield a set of unique, high quality immunoglobulin heavy chains from each sample.

3.2.4 Diversity and evolution of IgG B cells

To examine the diversity and breadth of the response to infection, we analysed several properties of heavy chain immunoglobulin sequences. Long heavy chain CDR3 regions have been strongly associated with several broadly neutralizing antibodies to HIV, interacting with the binding pocket in the CD4 binding site and penetrating the glycan shield to access conserved epitopes[343]. In total IgG memory B cells, median CDR3 length was the statistically unchanged at all three time points, with median CDR3 length of 12-14 amino acids (standard deviation of 5.2-5.7) in bone marrow, 11-12 amino acids (standard deviation of 5.0-5.4) in lymph nodes, and 12 amino acids (standard deviation of 4.7-5.2) in PBMC (Fig. 3.6 A,B, C). In PBMC, there was no difference between SIV-specific and total CDR3 sequences at 4 weeks ($p = 0.09$), while at 10 weeks and 24 weeks SIV-specific and total memory median CDR3 lengths were significantly different ($p<0.0001$) (Table 3.1). Median length in SIV-specific sequences in PBMC was 11 amino acids at 4 weeks and 10 weeks (standard deviation of 4.6 and 4.8, respectively) but increased significantly ($p<0.0001$) to 13 amino acids (standard deviation 4.9) by 24 weeks (Fig. 3.6 F). In lymph nodes and in bone marrow, CDR3 lengths of SIV-specific sequences were significantly different from total IgG memory sequences at all three time points ($p <0.0001$) (Fig. 3.6 D, E; Table 3.1).

In comparing SIV-specific sequences at 4, 10, and 24 weeks, we measured the proportion of long CDR3s (defined as 18 amino acids or longer). In lymph nodes, PBMC, and bone marrow, long CDR3s (18 AA and longer) made up 0.5%-1.1% of all sequences with no significant difference between time points, while antigen-specific sequences had anywhere from 0.2%-1.9% of CDR3s 18 AA or longer (Table 3.1). Previous studies have shown that there is no difference between the total memory and Env-specific median CDR3 length, which is in accordance with the data presented here[344].

Table 3.1 CDR3 Lengths and VH mutation in IgG samples

Time Point (weeks)	Tissue	Population	IGH	**Median CDR3**	**CDR3 standard deviation**	**Median VH mutation**	**VH mutation standard deviation**	**Percentage of CDR3s 18AA or longer**
4	BM	gp140	IgG	11	4.70	3.47	4.30	0.23
10	BM	gp140	IgG	12	4.85	3.78	3.60	1.12
24	BM	gp140	IgG	12	4.77	6.94	3.09	0.48
4	LN	gp140	IgG	11	5.09	3.44	3.01	0.93
10	LN	gp140	IgG	11	5.31	2.75	2.32	1.92
24	LN	gp140	IgG	11	4.76	6.8	2.32	0.34
4	PBMC	gp140	IgG	11	4.47	3.82	2.90	0.46
10	PBMC	gp140	IgG	11	4.81	3.82	2.58	1.02
24	PBMC	gp140	IgG	13	4.89	6.25	2.60	0.61
4	BM	mem	IgG	12	5.27	6.6	4.49	0.67
10	BM	mem	IgG	13	5.19	4.83	4.08	1.11
24	BM	mem	IgG	12	5.67	6.87	2.30	0.78
4	LN	mem	IgG	12	5.32	5.17	3.39	0.80
10	LN	mem	IgG	11	5.02	5.15	3.78	0.57
24	LN	mem	IgG	12	5.36	5.9	3.71	0.55
4	PBMC	mem	IgG	12	4.77	3.74	4.09	0.69
10	PBMC	mem	IgG	12	4.92	3.18	3.49	0.88
24	PBMC	mem	IgG	12	5.21	5.14	2.74	0.91

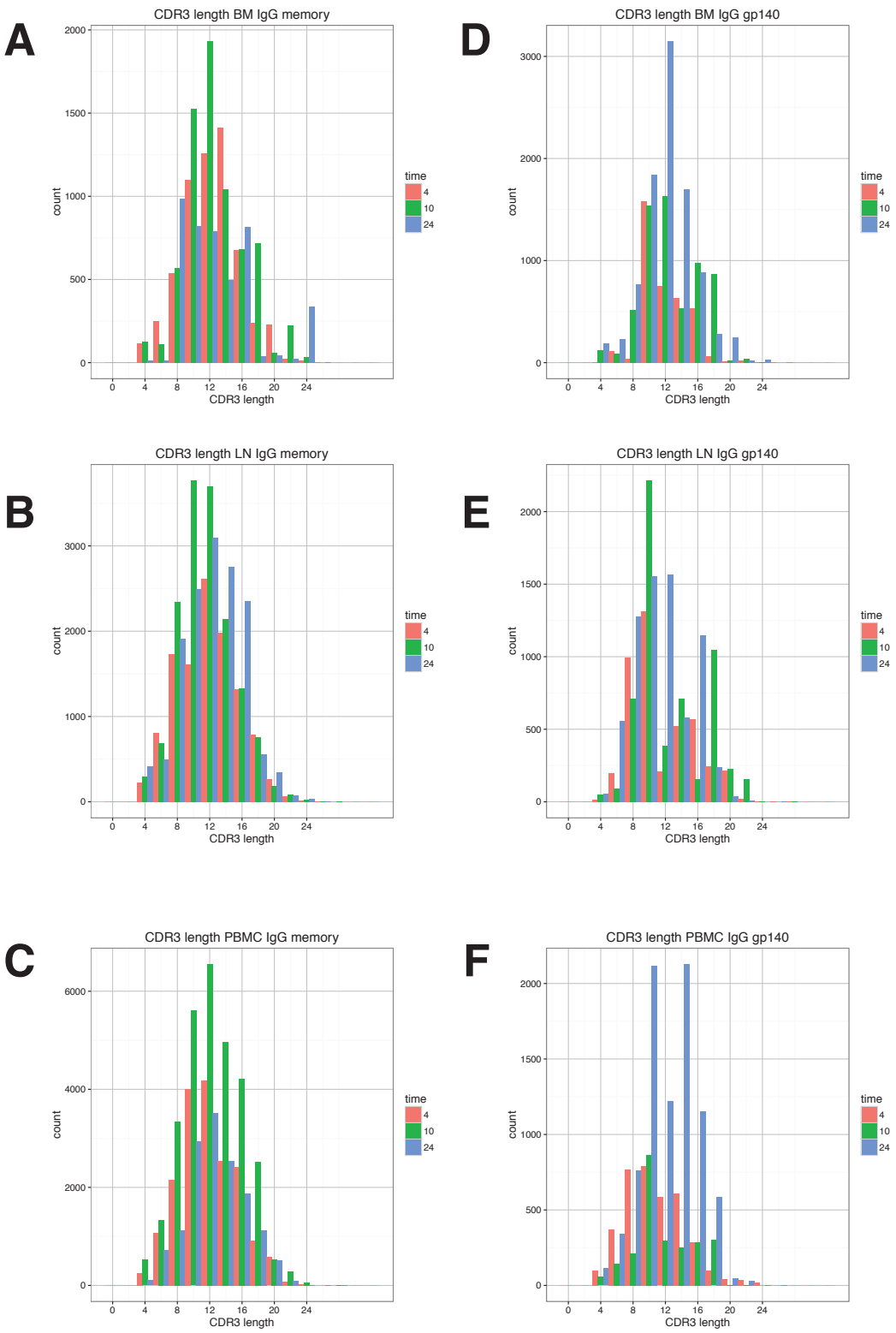


Fig 3.6 CDR3 length in SIV-specific and non-specific IgG heavy chain sequences. Histogram of CDR3 length in heavy chain sequences from non-specific IgG memory B cells by time point (4, 10, 24 weeks) in bone marrow (A), lymph nodes (B), and PBMC (C). Histogram of CDR3 length in heavy chain sequences from SIV-specific IgG memory B cells by time point (4, 10, 24 weeks) in bone marrow (D), lymph nodes (E), and PBMC (F).

Many neutralizing antibodies to HIV are also heavily mutated, with a subset of HIV bnAbs with 60-90 mutations in the V gene (10-30% mutation)[345]. Divergence from germline, measured using the genetic distance between the closest germline reference sequence of the variable region of the heavy chains, was unchanged in total lymph node and PBMC memory B cells, with a divergence of 4.8% to 6.9% in bone marrow, 5.2%-5.9% in lymph nodes, and 3.7%-5.1% in PBMC (Fig. 3.7 A-C). In all three compartments, median V mutation of SIV-specific sequences was unchanged at 4 and 10 weeks (from 3.5% to 3.8% in bone marrow, from 3.4% to 2.8% in lymph nodes, and 3.8% at both time points in PBMC) but increased at 24 weeks to 6.9%, 6.8%, and 6.3% in each tissue (Fig. 3.7 D-F). This was true in VH mutation in individual animals, with a statistically significant increase in the mean VH mutation between 4 weeks and 24 weeks in lymph nodes and PBMC in all three animals (Table 3.2). Two of the three animals had higher VH mutation at 24 weeks than at 4 weeks in bone marrow, while one animal had a slight decrease in the mean VH mutation between 4 weeks and 24 weeks. This shift in median divergence from germline likely reflects the on-going mutation and adaptation of SIV-specific B cells in the milieu of chronic infection and persistent viremia.

Table 3.2 VH mutation and CDR3 length in SIV-specific IgG sequences within individual animals

			CDR3 Length				VH mutation			
			4 weeks		24 weeks		4 weeks		24 weeks	
			median	SD	median	SD	median	SD	median	SD
ZF61	IgG gp140	PBMC	8	0.77	14	4.02	5.63	0.95	6.97	2.75
ZF76	IgG gp140	PBMC	12	4.63	10	3.52	4.51	2.73	4.86	2.39
ZG13	IgG gp140	PBMC	10	3.94	13	3.91	1.76	2.67	5.56	2.29
ZF61	IgG gp140	LN	15	3.76	10	3.86	7.72	2.38	6.8	2.16
ZF76	IgG gp140	LN	9	2.95	11	3.99	2.41	0.69	6.6	2.04
ZG13	IgG gp140	LN	10	2.87	12	3.27	1.72	2.45	6.94	2.23
ZF61	IgG gp140	BM	11	1.49	13	3.38	1.37	0.85	6.53	3.09
ZF76	IgG gp140	BM	14	3.07	12	3.85	11.5	2.60	6.87	3.03
ZG13	IgG gp140	BM	11	2.78	12	2.54	1.04	1.86	7.56	1.42

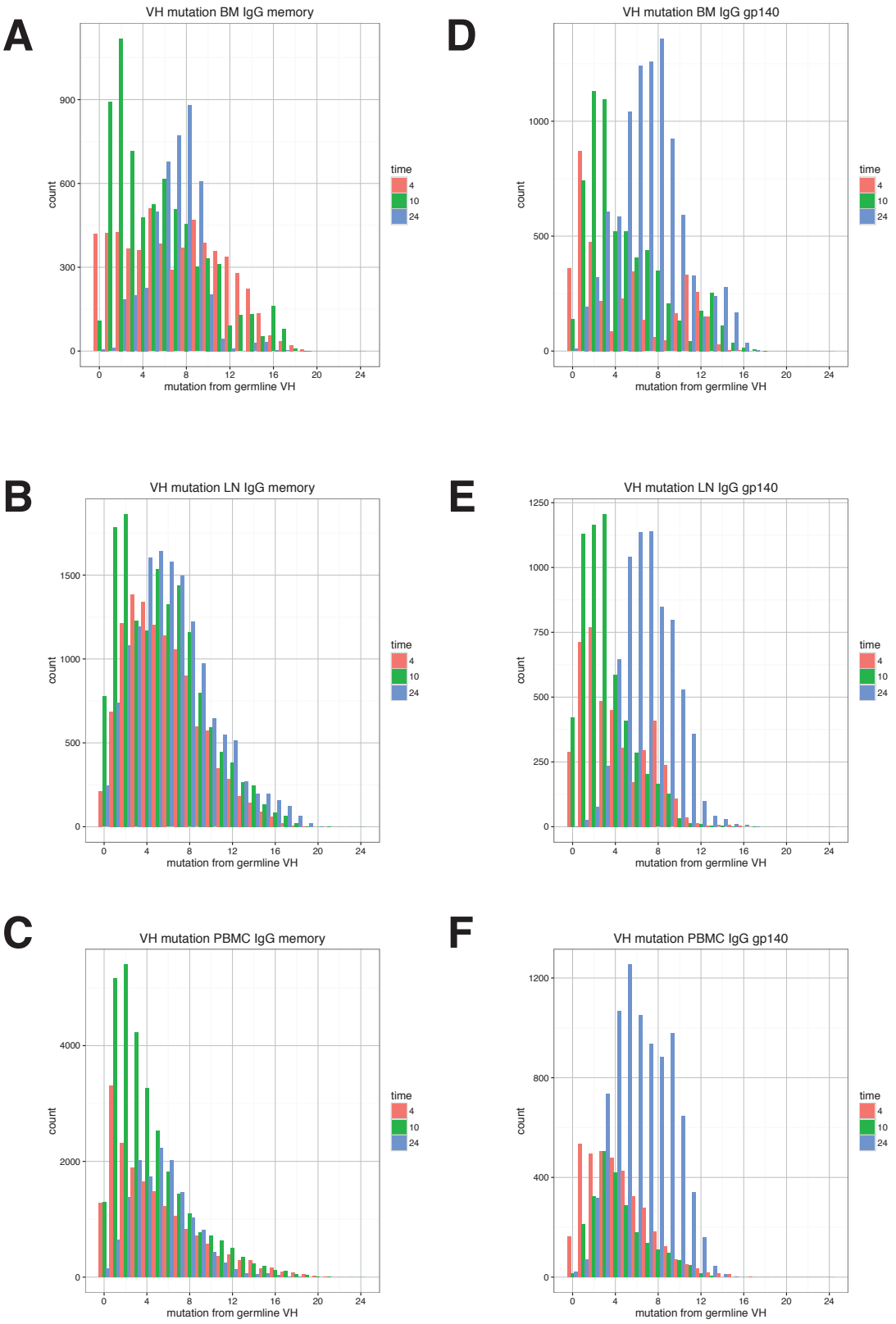


Fig 3.7 Percent mutation from predicted VH germline in SIV-specific and non-specific IgG heavy chain sequences. Percent divergence from V gene germline in heavy chains from non-specific IgG memory B cells in bone marrow (A), lymph nodes (B), and PBMC (C) by time point (4 weeks, 10 weeks, 24 weeks). Percent divergence from V gene germline in heavy chains from SIV-specific IgG memory B cells in bone marrow (D), lymph nodes (E), and PBMC (F) by time point (4 weeks, 10 weeks, 24 weeks).

3.2.5 Gene usage in IgG B cells

Of the hundreds of known human broadly neutralizing antibodies, certain V gene families are over-represented relative to their proportion in the pool of memory B cells[346]. To examine the early usage of V and J gene families in the SIV-specific response, each sequence was assigned a V and J gene based on BLAST nucleotide specificity, and the relative frequencies of each family were compared. While there was some variation in the proportion of the seven V gene families within total IgG memory bone marrow, lymph node, and PBMC cells at the three time points measured (Fig 3.8 A), the distributions largely followed previously established proportional distributions, with VH1, VH3, and VH4 dominating the total V gene usage[346]. IGHV2 made up only a small fraction of bone marrow and lymph node IgG memory sequences (up to 3.5%), but was slightly higher in PBMC (up to 10%). VH5, VH6, and VH7 generally represented less than 10% combined of all VH gene sequences in total memory IgG cells, but in lymph node SIV-specific sequences at four weeks and ten weeks, make up over 25% of all heavy chains (Fig. 3.8 B). Lymph node and PBMC VH gene distributions differed significantly ($p<0.0001$) from the total memory VH gene distributions at all time points. SIV-specific IgG bone marrow sequences at four weeks are dominated by VH4 and are significantly different from total IgG memory ($p<0.00001$) but by 24 weeks have frequencies of VH1, VH2, and VH3 that not significantly different from total bone marrow IgG memory ($p=0.10$). The dominance of VH4, VH3, and VH1 we observe is in agreement with data published by Sundling et al (2014) of IgG repertoires after immunization of macaques with HIV-1 Env[347]. Our work is also in accordance with studies showing no bias toward particular gene segments for Env-specificity[348].

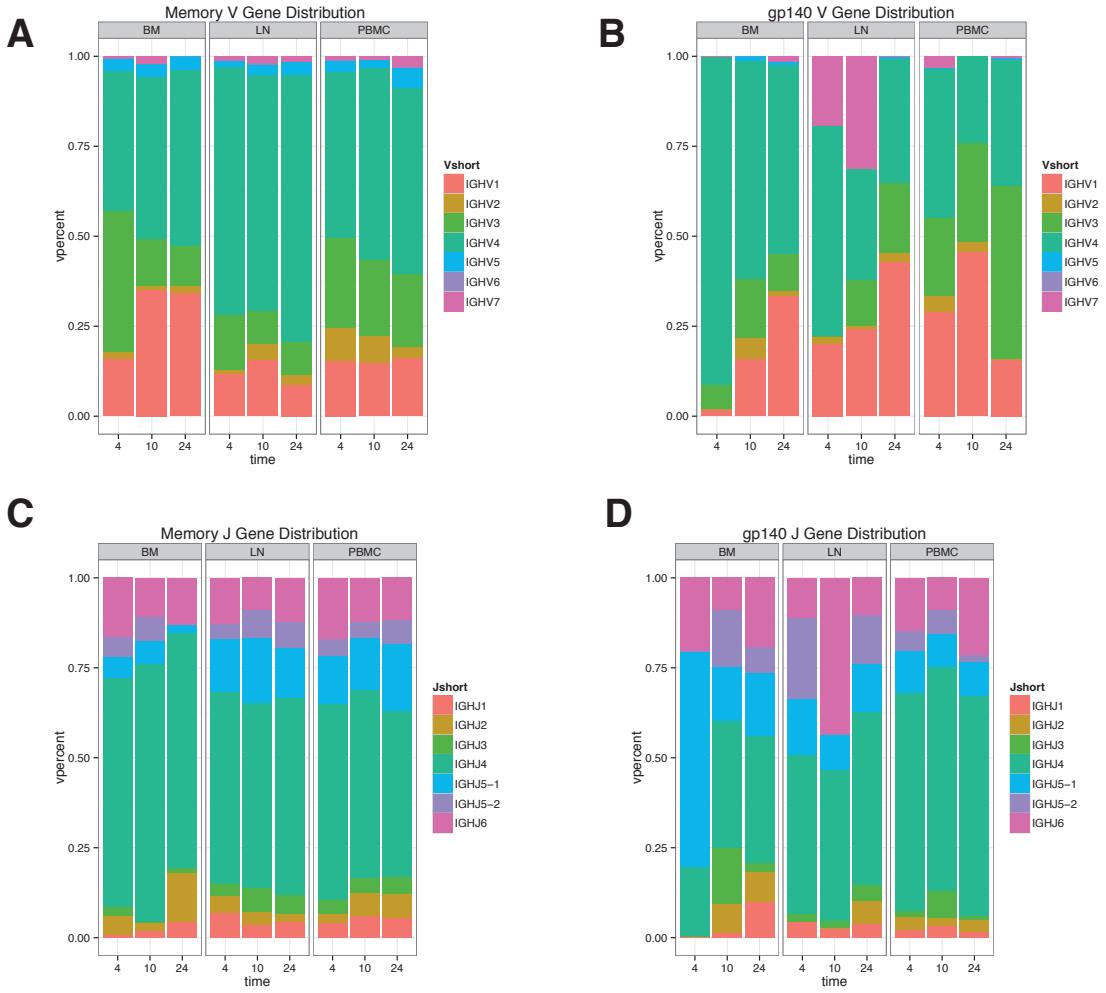


Fig 3.8 V and J gene distributions in IgG B cells. (A) Proportions of heavy chain V gene families in non-specific IgG memory B cells in bone marrow, lymph nodes, and PBMC. (B) Proportions of heavy chain V gene families in SIV-specific IgG memory B in bone marrow, lymph nodes, and PBMC. (C) Proportions of heavy chain J gene families in non-specific IgG memory B cells in bone marrow, lymph nodes, and PBMC. (D) Proportions of heavy chain J gene families in SIV-specific IgG memory B cells in bone marrow, lymph nodes, and PBMC.

IgG memory sequence J genes were dominated by JH4 (45%-60%), in agreement with published data[346]. In total IgG B cells, bone marrow sequences had fewer JH1 and JH5 genes than lymph node or PBMC, but all non-SIV specific sequences had similar frequencies of J genes (Fig. 3.8 C). J genes in SIV-specific sequences in lymph nodes were significantly different than total IgG B cells at all time points ($p = 0.004$ at 4 weeks, $p=0.02$ at 10 weeks and $p<0.0001$ at 24 weeks), largely due to fewer JH4 genes and more JH5 and JH6 genes (Fig 3.8 B). The SIV-specific PBMC J gene frequencies were not significantly different from total PBMC J gene frequencies at 4 weeks ($p=0.64$) but by 10 weeks ($p=0.0006$) and 24 weeks ($p<0.0001$) were significantly different, with increases in the proportion of JH4 genes in SIV-specific cells. Bone marrow J genes showed the greatest differences between SIV-specific and total memory, with significantly more JH5-1 genes at 4 weeks in SIV-specific sequences

(p<0.0001) and fewer JH4 sequences at 10 weeks and 24 weeks (both p<0.0001). However, there did not appear to be a unique J gene signature of SIV-specific sequences.

Of the long CDR3 sequences (18 amino acids or greater), the V and J gene distributions were significantly altered when compared to the total distribution in SIV-specific sequences (Fig. 3.9). Long CDR3 sequences in bone marrow were almost entirely from VH1 (95% at 4 weeks) and VH4 (77% at 10 weeks and 78% at 24 weeks) (Fig. 3. A). Lymph node sequences with CDR3s of 18 amino acids or longer had significantly more VH1 genes at 4 and 10 weeks, and significantly fewer VH7 genes at all time points. PBMC sequences with CDR3s of 18 amino acids or longer contained higher frequencies of VH2 genes at 4 weeks and 10 weeks and of VH3 and VH5 at 24 weeks, and fewer VH4 genes at 10 weeks. J genes were also significantly different in sequences with long CDR3s (Fig. 3.9 B). JH1 genes were overrepresented in bone marrow at four weeks, and JH2 were present at a significantly higher proportion at 24 weeks. In lymph nodes, there were higher frequencies of JH6 genes at 4 and 10 weeks in long CDR3 samples, and higher JH5-1 at 24 weeks. PBMC samples also had higher frequencies of JH1 (4 weeks and 10 weeks) and JH6 (24 weeks) than total SIV-specific sequences.

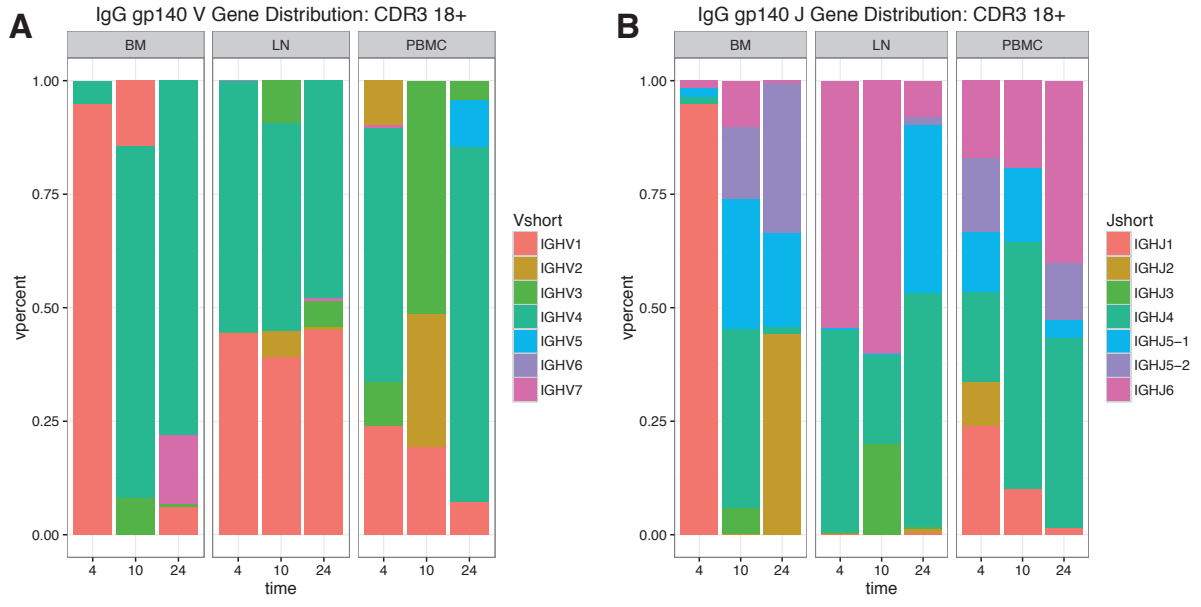


Fig 3.9 V and J gene distributions in SIV-specific IgG B cells with CDR3 length 18 amino acids or longer. (A) Proportions of heavy chain V gene families in bone marrow, lymph nodes, and PBMC at 4, 10, and 24 weeks post infection. (B) Proportions of heavy chain J gene families in bone marrow, lymph nodes, and PBMC at 4, 10, and 24 weeks post infection.

3.2.6 IgM memory B cells: CDR3, mutation, and V/J gene usage

While the humoral response to HIV/SIV is dominated by IgG, there are low numbers of antigen binding IgM memory B cells circulating and in lymphoid tissues[349]. Nonspecific IgM memory B cells in PBMC and lymph nodes had median CDR3 lengths of 10-12 amino acids (standard deviation 3.8-4.9), while SIV-specific IgM memory B cells had median CDR3 lengths of 11-12 amino acids (SD 3.2-5.0) (Fig 3.10). Although the median CDR3 lengths were similar, there was a significant difference between the distributions of SIV-specific and total lymph node and PBMC IgM CDR3 lengths at each time point ($p<0.0001$) (Table 3.3). No significant differences were observed in the percentage of long CDR3s (over 18 AA), ranging from 0% to 0.26% of SIV-specific sequences and 0.09%-0.44% of bulk IgM memory sequences (Table 3.4). However, IgM sequences had fewer long CDR3s than their IgG counterparts, with an average of 3.2 times fewer CDR3s of length 18 AA or longer.

Table 3.3 Median CDR3 length and VH gene mutation in IgM B cells throughout infection.

Time Point (weeks)	Tissue	Population	IGH	Median CDR3	CDR3 standard deviation	Median VH mutation	VH mutation standard deviation	Percentage of CDR3s 18AA or longer
4	**LN**	**gp140**	**IgM**	11	3.3	5.9	2.83	0.00
10	**LN**	**gp140**	**IgM**	11	4.0	2.79	3.15	0.24
24	**LN**	**gp140**	**IgM**	11	4.1	4.53	3.32	0.18
4	**PBMC**	**gp140**	**IgM**	10	4.9	4.48	3.42	0.29
10	**PBMC**	**gp140**	**IgM**	11	5.0	5.78	2.24	0.11
24	**PBMC**	**gp140**	**IgM**	11	4.0	3.33	3.09	0.27
4	**LN**	**mem**	**IgM**	10	4.3	4.81	2.87	0.13
10	**LN**	**mem**	**IgM**	10	3.8	4.23	3.38	0.09
24	**LN**	**mem**	**IgM**	11	4.9	5.54	2.80	0.44
4	**PBMC**	**mem**	**IgM**	11	4.2	3.94	3.06	0.23
10	**PBMC**	**mem**	**IgM**	11	4.1	3.44	2.87	0.33
24	**PBMC**	**mem**	**IgM**	12	3.8	4.53	2.91	0.18

Table 3.4 VH mutation and CDR3 length in SIV-specific IM sequences within individual animals

			CDR3 Length				**VH mutation**			
			4 weeks		**24 weeks**		**4 weeks**		**24 weeks**	
			median	**SD**	**median**	**SD**	**median**	**SD**	**median**	**SD**
ZF61	IgM gp140	PBMC	10	5.8	10	4.00	5.44	3.62	3.78	3.03
ZF76	IgM gp140	PBMC	11	4.0	11	3.73	4.51	2.91	3.04	2.93
ZG13	IgM gp140	PBMC	9	3.8	17	3.14	2.38	3.73	11.81	1.87
ZF61	IgM gp140	LN	13	2.0	10	3.33	7.22	1.61	6.97	2.87
ZF76	IgM gp140	LN	12	2.9	11	4.13	5.84	3.48	3.47	3.25
ZG13	IgM gp140	LN	10	2.1	10	4.74	5.84	1.58	5.26	3.51

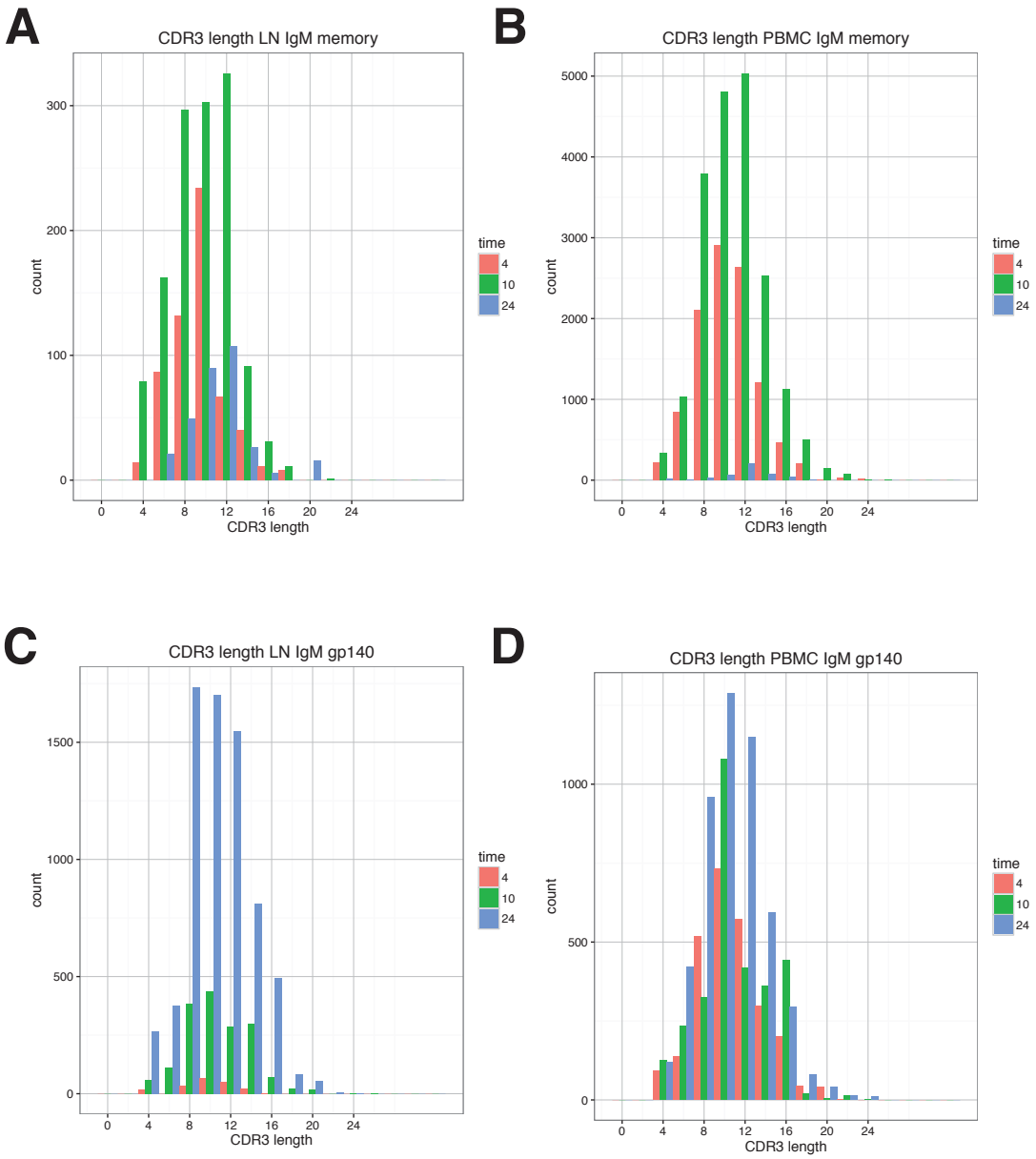


Fig 3.10 CDR3 length in SIV-specific and non-specific IgM heavy chain sequences. (A) CDR3 length in heavy chains from non-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (B) CDR3 length in heavy chains from non-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks). (C) CDR3 length in heavy chains from SIV-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (D) CDR3 length in heavy chains from SIV-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks).

SIV-specific IgM memory V genes did not show accumulation of mutations over the course of infection. Sequences from 4 weeks post infection in lymph nodes had a median VH gene mutation of 5.9% and standard deviation of 2.8 but by 24 weeks was 4.5% with a standard deviation of 3.3 (Fig 3.11 C). In PBMC, median VH gene mutation similarly declined from 4.5% at 4 weeks to 3.3% at 24 weeks (Fig. 3.11 D). In the non-SIV specific IgM memory cells, median VH gene mutation ranged from a median of 4.2% to 5.5% in lymph nodes and 3.9%-4.5% in PBMC (Fig 3.11 A, B). Thus, in contrast to IgG SIV-specific memory B cells,

SIV-specific IgM cells did not show the accumulation of mutations over the course of infection.

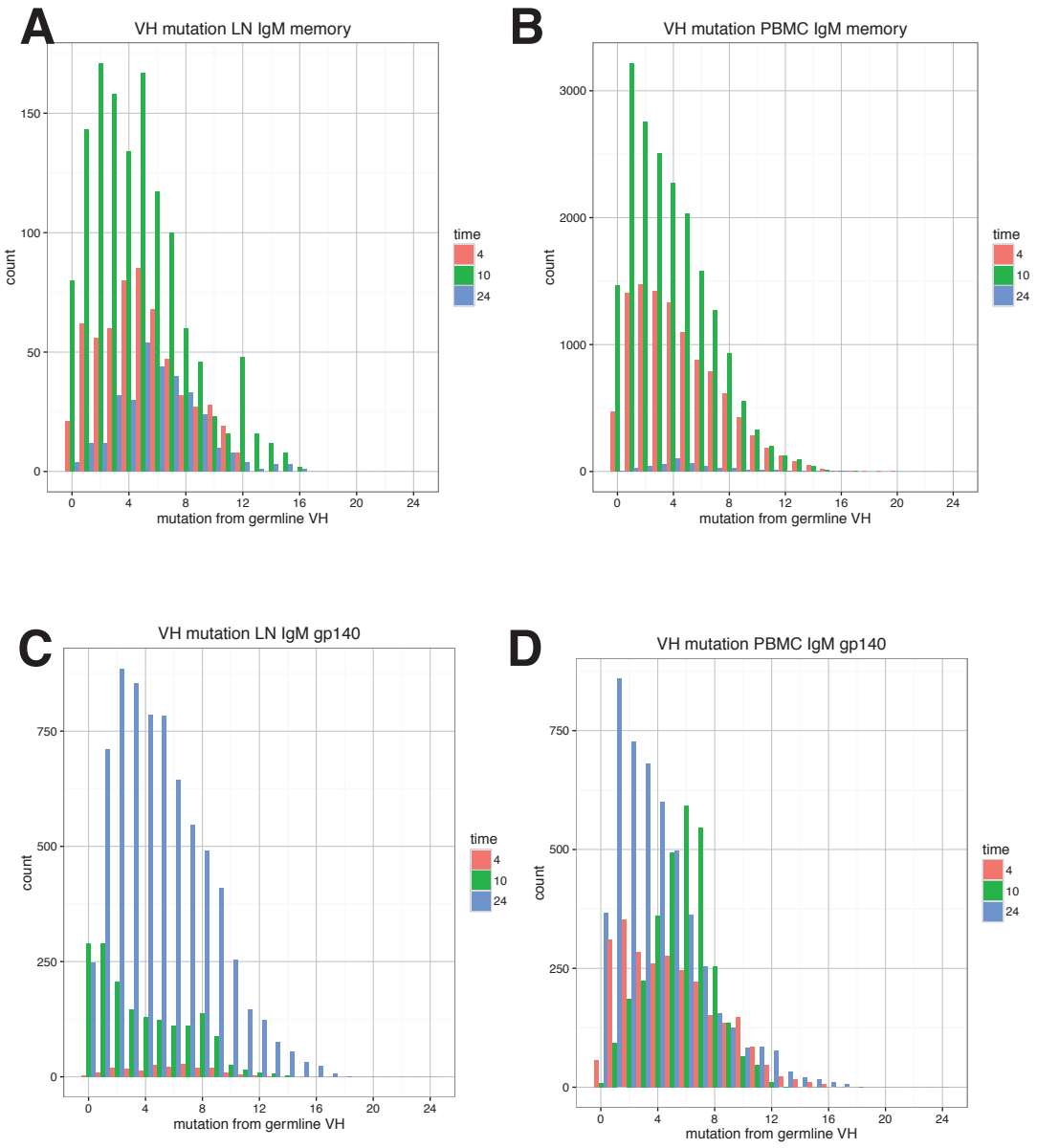


Fig 3.11 Percent mutation from predicted V Gene germline in SIV-specific and non-specific IgM heavy chain sequences. (A) Percent divergence from V gene germline in heavy chains from non-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (B) Percent divergence from V gene germline in heavy chains from non-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks). (C) Percent divergence from V gene germline in heavy chains from SIV-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (D) Percent divergence from V gene germline in heavy chains from SIV-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks).

VH and JH gene distributions in the non-specific IgM memory sequences in lymph node and PBMC showed greater variation between time points than the IgG sequences. VH gene frequencies were dominated by VH4 and VH3, except in PBMC at 24 weeks post infection when the majority of sequences had a VH1 gene assignment (Fig 3.12). There was also a

substantially higher fraction of VH2 sequences in total IgM lymph node samples at all time points, in contrast to total IgG sequences where VH2 sequences were only detected in as a small fraction of sequences. The SIV-specific IgM sequences had higher frequencies of VH5 and VH7 in PBMC at 4 weeks and 10 weeks, and in both tissues had low to undetectable fractions of VH1, in contrast to the IgG SIV-specific sequences (Fig 3.12 B). Both lymph node and PBMC sequences had significantly different distributions of V genes between SIV-specific and total IgM cells at 4, 10, and 24 weeks (all $p<0.0001$). Lymph node SIV-specific sequences were dominated by VH2 and VH3 at 4 weeks and 10 weeks (65% and 67% of all J genes). J gene sequences in non-specific IgM memory B cells were similar to the IgG memory distributions with the exception of both lymph node and bone marrow at 24 weeks post infection, where JH6 genes dominated and made up nearly 75% of the sequences (Fig. 3.12 C). This bias was not observed as strongly in the SIV-specific J gene sequences, which had nearly 40% JH6 genes in lymph nodes at 24 weeks and PBMC at 10 weeks (Fig 3.12 D).

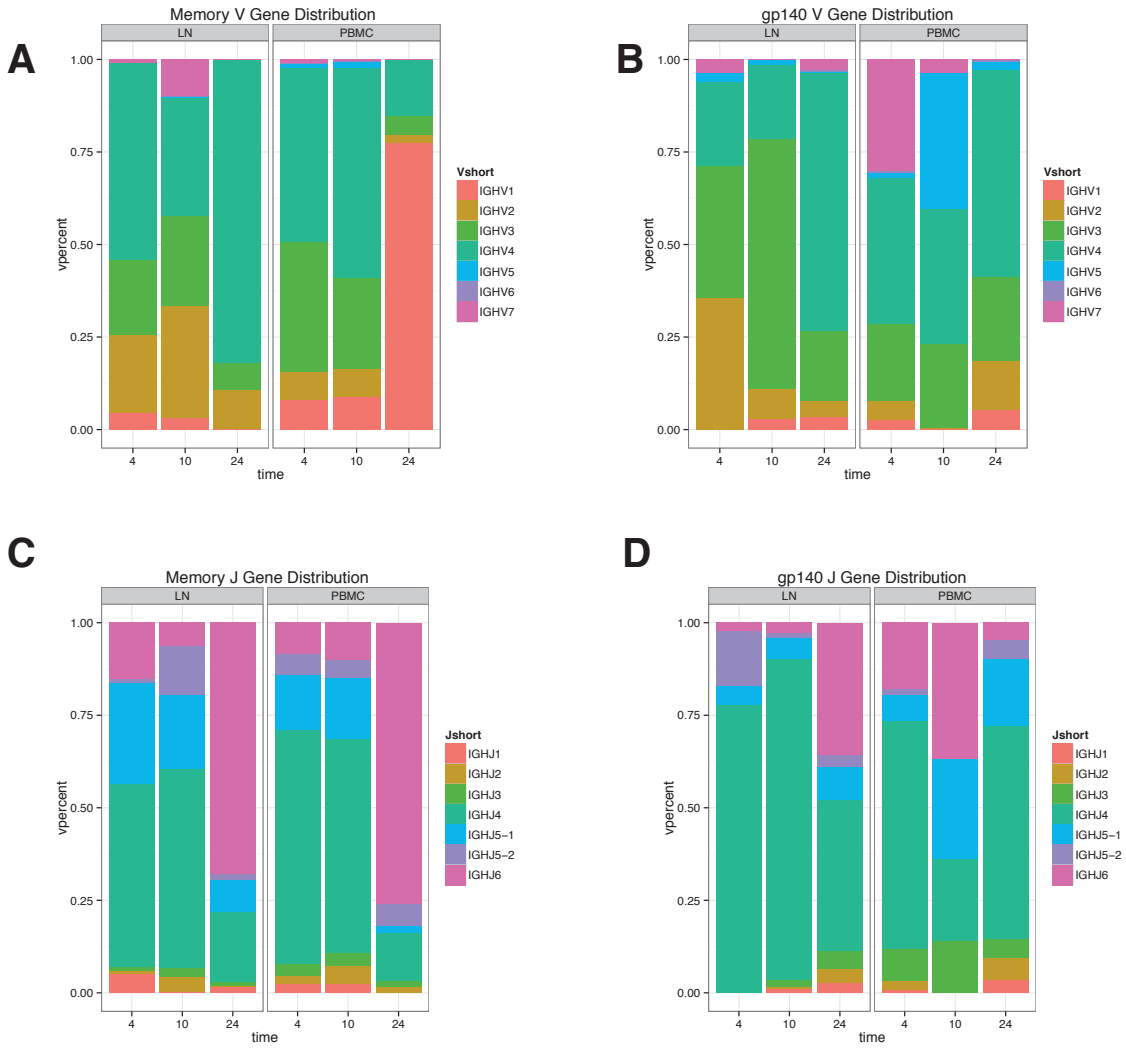


Fig 3.12 V and J gene distributions in IgM B cells. (A) Proportion of V gene families in non-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection. (B) Proportion of V gene families in SIV-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection. (C) Proportion of V gene families in non-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection. (D) Proportion of V gene families in SIV-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection.

3.2.7 SIV-specific heavy chains become more diverse throughout infection

While the number of sequencing reads per unique heavy chain in next-generation bulk sequencing is not a direct measure of biological clone size, it nonetheless offer a qualitative insight into the distribution of clones and expansion of lineages. We used a clustering algorithm, CD-HIT[350], to analyse closely related sequences and study the size of clusters and the connectivity of sequences within a sample. These network plots show expanded clones and the diversity of sequences within a sample, and allow the calculation of several diversity measures using the size and number of connections within networks[332]. Non-specific memory populations typically show a large number of unique sequences with some smaller networks of expanded clones, while the SIV-specific samples are generally dominated by a small number of clones that are closely related to many other sequences within the sample (Fig. 3.13). This is reflected in the percentage of vertices in the largest cluster, where memory

samples have an average of 2.1% of all vertices (unique sequences) in the largest cluster in the sample, whereas SIV-specific samples have an average of 10% of vertices in the largest cluster (Fig 3.14). While there was not a statistically significant difference between the total and SIV-specific samples, early samples (4 weeks) had the greatest mean percentage of vertices in the largest cluster with 20%, indicating that the early response was dominated by a small number of clones. In contrast, in later infection (24 weeks) only 4.6% of vertices were in the largest cluster, meaning a single clonal family did not dominate the SIV-specific response. In the IgM sequences, the non-specific sequences had a mean 2.6% of vertices in the largest cluster and did not differ significantly across time-points (Fig 3.14 B). The SIV-specific IgM sequences had an average of 6.6% of vertices in the largest cluster, and did not show the dominance of a small number of clones in the early time point as the IgG sequences did.

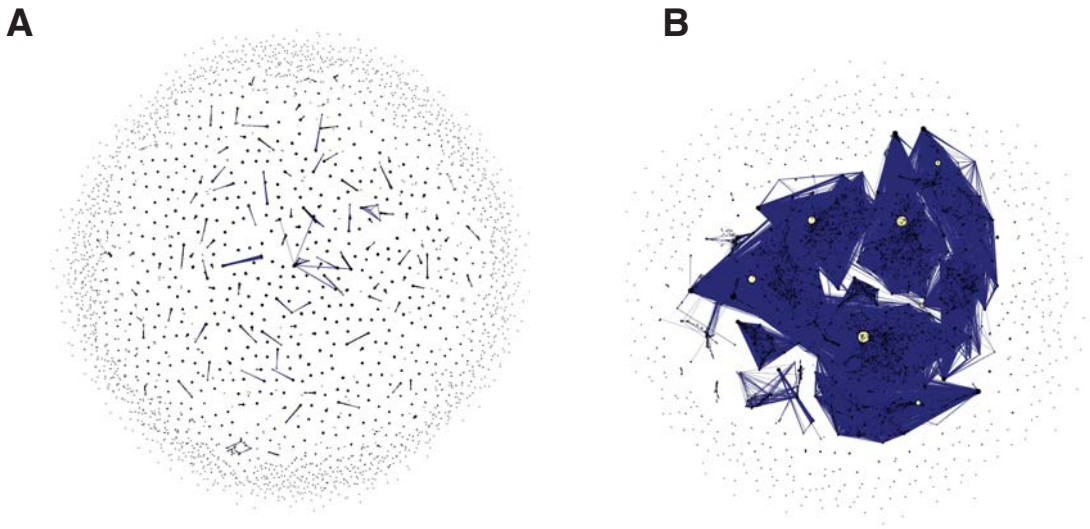


Fig 3.13 Representative network plots of non-specific memory and SIV-specific heavy chain sequences. Network plots of non-SIV specific (A) and SIV-specific (B) IgG memory heavy chain sequences. Network plots visualize the relatedness of sequences in a sample by drawing each unique sequence as a point and lines connecting closely related sequences. The size of each point is determined by the number of identical sequencing reads for that sequence.

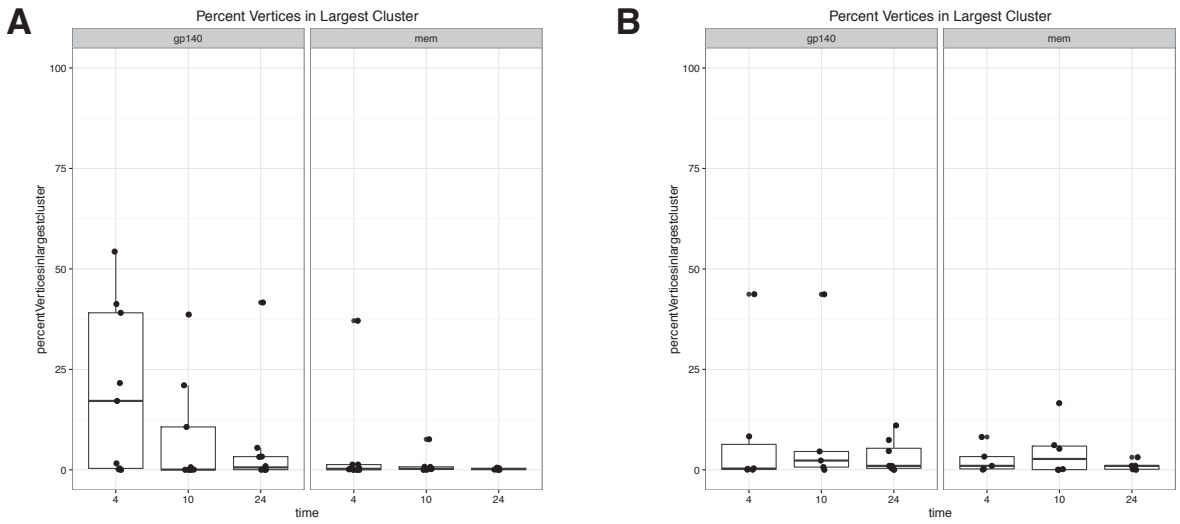


Fig 3.14 Vertices in largest cluster (A) Percentage of vertices in the largest cluster in the network plots of bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgG B cells at 4, 10, and 24 weeks. (B) Percentage of vertices in the largest cluster in the network plots of bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgM B cells at 4, 10, and 24 weeks.

We use two diversity measures with two parameters to further examine the patterns in the SIV-specific and nonspecific memory heavy chains. The Gini index is a measure of the inequality of a system, while the Renyi index is a measure of evenness in a system, both on a scale of 0 to 1[332]. When applied to BCR repertoires, inequality and evenness can be used to quantify whether a group of sequences contain equal numbers of diverse sequences (low inequality and high evenness), or whether the sample is dominated by a small number of closely related sequences (high inequality and low evenness). A low Gini index (close to 0) indicates low inequality, and a low Renyi index indicates high evenness. In this system, the Gini and Renyi indices can be calculated both for the clusters distributions (based on the total size of each cluster) and the vertices (how many related sequences are in each cluster). The cluster Gini index indicates the level of somatic hypermutation and families of related sequences, while the vertex Gini index describes how clonal, or dominated by a sequence or small group of sequences, the sample is. The vertex Gini for non-specific IgG B cells had a median of 0.67 and a range of 0.41 to 0.79 (Fig 3.15 A), and was statistically unchanged between 4, 10, and 24 weeks (p=0.15, p=0.38, and p=.90). The SIV-specific IgG B cells had a median of 0.39 and range 0.13 of to 0.74 and was statistically unchanged between 4, 10, and 24 weeks (p=0.33, p=0.53, and p=0.70). However, in comparing SIV-specific to non-specific vertex Gini indices, SIV-specific samples were significantly lower at all three time points (p=0.02, p=0.0003, and p=0.005). In nonspecific memory IgG B cells, the cluster Gini index was not significantly different (p=0.36, p=0.84, p=0.35) between sampling time points, with a median of 0.84 and range of 0.71 to 0.88 (Fig. 3.15 B). The SIV-specific IgG sequences had a

lower median Gini index of 0.61, and a range of 0.41 to 0.83, and were not significantly different between time points. In comparing SIV-specific to non-specific cluster Gini indices, SIV-specific samples were significantly lower at all three time points ($p=0.003$, $p=0.0001$, and $p=0.0006$). Thus, the SIV-specific sequences were more clonal (a small fraction of clones made up a large proportion of the sample) than the non-specific IgG memory sequences. Thus, the SIV-specific IgG B cells tend to have more somatic hypermutation within clonally related families, indicated by the higher number of vertices connecting related sequences within clusters and representing closely related sequences.

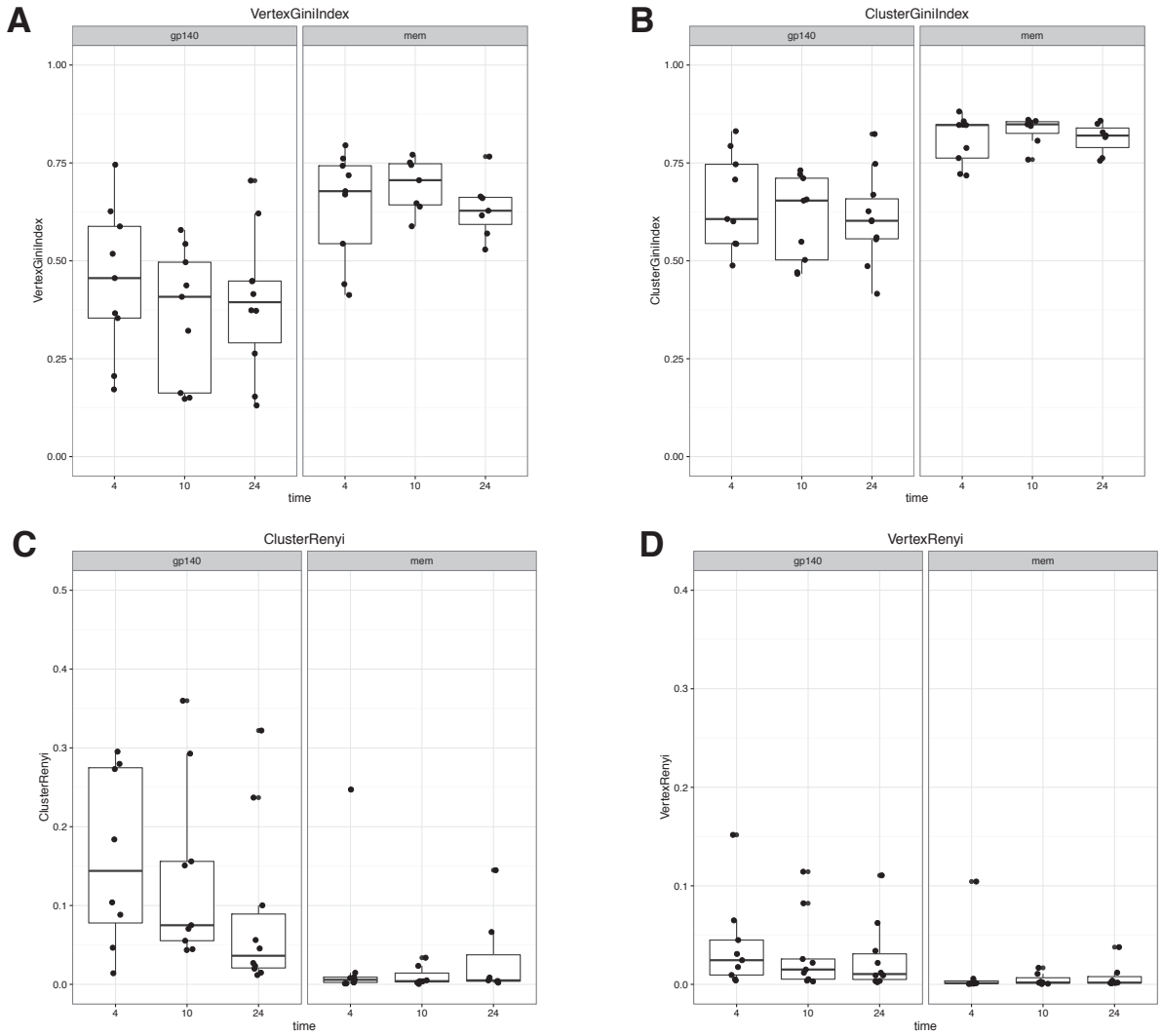


Fig 3.15 Diversity measures in IgG B cells (A) Vertex and (B) cluster Gini index for bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgG B cells at 4, 10, and 24 weeks. (C) Vertex and (D) cluster Renyi index for bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgG B cells at 4, 10, and 24 weeks.

The Renyi index measures evenness, and can similarly be applied to the clusters and vertices of heavy chain sequences. A low cluster Renyi index indicates an even sample, with most clones and families of similar size and read count. A low vertex Renyi indicates relatively low levels of somatic hypermutation and less expansion of closely related B cells. The IgG non-specific B cells had a median cluster Renyi index of $5.8x10^{-3}$, with a range of $5.5x10^{-4}$ to 0.24 (although only 2 values were over 0.07), and no significant difference between Renyi index at 4, 10, and 24 weeks (p=0.29-0.97) (Fig. 3.15 C). SIV-specific B cells had a median cluster Renyi index of 0.073, with a range of 0.012 to 0.57, and were unchanged between 4, 10, and 24 weeks (p=0.08-034). SIV-specific B cell cluster Renyi indices were significantly lower at 4 weeks (p=0.015) and 10 weeks (p=0.012), but not at 24 weeks (p=0.25). This indicates that total IgG memory B cell heavy chain sequences had a more even distribution of clone sizes, while the SIV-specific samples had an uneven distribution of clone sizes at 4 and 10 weeks. The vertex Renyi values for non-specific sequences had a median of $1.6x10^{-3}$ and range of $3.6x10^{-4}$ to 0.104 and none were significantly different between 4, 10, and 24 weeks post infection (p=0.54-0.73). (Fig 3.15 D). SIV-specific IgG heavy chains had a median vertex Renyi index of 0.18 and a range of $2.2x10^{-3}$to 0.015 and were not significantly different between 4, 10, and 24 weeks post infection (p=0.26-0.70). Though the vertex Renyi indices were higher in SIV-specific memory B cells (indicating greater unevenness of somatic hypermutation and expanded clonal families), there was not a statistically significant difference between total and SIV-specific B cells. Lymph node samples had a higher median vertex and cluster Gini Index (0.50 and 0.71) than bone marrow and PBMC (0.37 and 0.60, 0.34 and 0.54). Vertex Renyi index was also highest in lymph nodes, 0.022) but not cluster Renyi index (0.075).

Unlike the IgG samples, the differences between Gini and Renyi diversity indices in SIV-specific and non-specific IgM memory sequences were not statistically significant (Fig 3.16 A-D). The median cluster Gini index for SIV-specific sequences was 0.52 (range 0.40 to 0.68) whereas for non-specific IgM sequences it was 0.54 (range 0.39 to 0.73). The median vertex Gini index for SIV-specific sequences was 0.43 (range 0.10 to 0.60), while the median non-specific IgM vertex Gini index was 0.34 (range 0.14 to 0.56). PBMC sequences had consistently higher Gini indices in SIV-specific sequences at 4 weeks (0.51 and 0.66 for vertex and cluster) than lymph nodes (0.16 and 0.50), but by 24 weeks were not substantially different (0.44 and 0.53 for PBMC, 0.43 and 0.59 for lymph nodes). This indicates early IgM SIV-specific heavy chains in PBMC are less diverse than later sequences, but overall the IgM

sequences do not show the same expansion and diversification as IgG SIV-specific sequences. With the exception of two samples, the vertex Renyi indices for IgM samples were virtually identical across tissue types and time points, with median values of $7.3x10^{-3}$ and $9.7x10^{-3}$for SIV-specific and nonspecific and ranges of $1.4x10^{-3}$- 0.17 and $1.3x10^{-3}$-0.11 respectively (Fig. 3.16 C). Cluster Renyi indices were similarly low, with median values of 0.058 and 0.029 for SIV-specific and non-specific IgM heavy chains (Fig. 3.16 D).

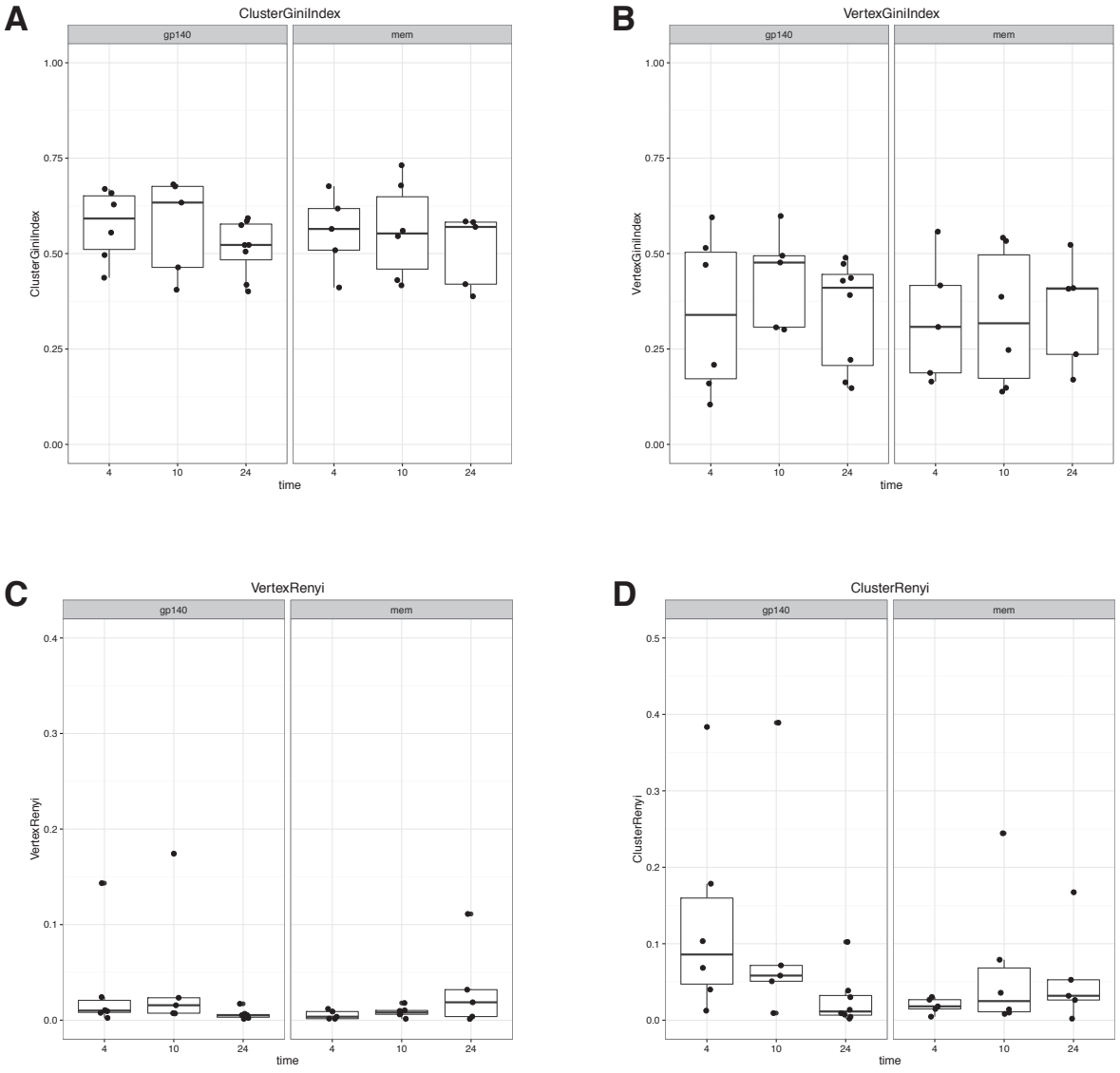


Fig 3.16 Diversity measures of IgM heavy chains. (A) Cluster and (B) vertex Gini index for lymph node, PBMC, and spleen SIV-specific and total memory IgM B cells at 4, 10, and 24 weeks. (C) Cluster and (D) vertex Renyi index for lymph node, PBMC, and spleen SIV-specific and total memory IgM B cells at 4, 10, and 24 weeks.

3.2.8 Shared sequences across time and space

Antibody lineages are derived from a single common ancestor and represent generations of antibodies that target specific epitopes and have undergone successive rounds of somatic

hypermutation. To track the persistence of related sequences over time and in different tissue compartments, we defined sequences as identical if they had the same V and J genes and identical CDR3 regions. We also looked for sequences in different compartments and time points that had were closely (same V and J genes but CDR3 regions that differed by one or two amino acids, either insertions, deletions or substitutions). We wanted to look at the features of the gp140-specific response within individual animals as between animals: identical or similar heavy chain sequences in multiple animals.

Table 3.5 Sharing of CDR3 sequences between tissues.

SIV-specific			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LNvsPBMC	**IgG**	**gp140**	3.1%	1.8%	14.2%	3.3%	25.6%	6.4%	34.4%	9.2%
LNvsBM	**IgG**	**gp140**	1.8%	2.5%	18.4%	14.2%	22.7%	15.6%	25.7%	16.8%
BMvsPBMC	**IgG**	**gp140**	0.8%	1.5%	3.4%	4.2%	6.7%	7.7%	12.3%	11.2%
LNvsPBMC	**IgM**	**gp140**	0.3%	0.4%	1.7%	1.3%	5.8%	3.4%	17.6%	7.4%
Total Memory			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LNvsPBMC	**IgG**	**gp140**	1.0%	0.4%	3.5%	1.6%	7.2%	2.4%	16.3%	3.3%
LNvsBM	**IgG**	**gp140**	0.5%	0.3%	8.2%	12.5%	11.1%	14.4%	17.5%	14.1%
BMvsPBMC	**IgG**	**gp140**	0.4%	0.3%	3.5%	5.1%	5.3%	5.3%	12.0%	6.1%
LNvsPBMC	**IgM**	**gp140**	0.5%	0.6%	1.8%	1.5%	6.7%	6.2%	20.0%	11.2%

Of all the SIV-specific IgG sequences within an individual animal, an average of 3.1% of sequences were identical between lymph nodes and PBMC from all time points (Table 3.5), though not significantly different than the proportion of identical sequences in total memory (1.0%). An average of 14.2% of all lymph node and PBMC sequences were within one amino acid apart, 25.6% were within 2 amino acids, and 34.4% were within three amino acids from a sequence in the other compartment – all significantly more than in total memory lymph node vs. PBMC sequences ($p= 0.007$, $p=0.009$, $p=0.03$). Between lymph nodes and bone marrow, 1.8% of sequences were identical, rising to 25.7% of sequences within a 3 amino acid difference in the CDR3 region, though there were not significantly more sequences shared than in total memory. Bone marrow and PBMC shared fewer sequences, with an average of 0.8% identical sequences, 3.4% of sequences within a single amino acid difference in the CDR3 region, and 12.3% of sequences within three amino acids, again with no significant difference between SIV-specific and total memory sharing. For IgM SIV-specific sequences, only lymph node compared to PBMC had sufficient sequences for statistical comparison.

Between these compartments, an average of 0.3% of sequences were identical, rising to 17.6% of sequences within a 3 amino acid difference in the CDR3, but were not significantly different from total memory Thus, the greatest sharing of sequences is between lymph nodes and PBMC and there is significantly more sharing within SIV-specific sequences than in total memory, though shared sequences are observed across all compartments throughout infection.

Table 3.6 Sharing of CDR3 sequences across time points

SIV-specific			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
4vs10	**IgG**	**LN**	7.9%	6.5%	31.6%	33.1%	32.8%	33.3%	38.6%	30.9%
10vs24	**IgG**	**LN**	0.5%	0.5%	2.3%	2.9%	5.8%	4.6%	13.7%	10.3%
4vs24	**IgG**	**LN**	0.2%	0.2%	5.5%	6.4%	6.3%	7.5%	9.4%	9.7%
4vs10	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	2.3%	5.0%	6.3%	7.4%
10vs24	**IgM**	**LN**	0.3%	0.6%	1.1%	1.3%	3.6%	3.6%	13.3%	8.4%
4vs24	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	3.8%	5.5%	9.8%	12.7%
Total Memory			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
4vs10	**IgG**	**LN**	0.3%	0.3%	1.3%	0.8%	3.4%	1.5%	9.1%	3.0%
10vs24	**IgG**	**LN**	0.2%	0.2%	1.0%	1.1%	2.7%	1.6%	8.4%	3.9%
4vs24	**IgG**	**LN**	0.6%	0.2%	2.1%	1.1%	4.9%	2.1%	11.4%	3.5%
4vs10	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	1.2%	1.5%	9.4%	6.4%
10vs24	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	0.6%	0.8%	5.0%	5.0%
4vs24	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	0.8%	1.4%	6.3%	8.1%

Within the lymph nodes, certain heavy chain sequences persisted over the twenty-four weeks of infection studied here (Table 3.6). There was no statistically significant difference between the percentages of sequences shared over time in SIV-specific versus total memory at any time point. Between 4 weeks and 10 weeks, an average of 7.9% of SIV-specific IgG sequences were identical, while between 10 weeks and 24 weeks 0.5% of sequences were identical, and between 4 weeks and 24 weeks only 0.2% of sequences were identical. There were more closely related sequences shared between 4 and 10 weeks (an average of 38.6% of sequences were within 3 amino acids) than there were between 10 to 24 weeks (13.7% of sequences within 3 amino acids), or 4 to 24 weeks (9.4% of sequences within 3 amino acids). This likely reflects the different time intervals between sampling (6 weeks vs. 14 weeks), however, the persistence of identical and closely related sequences indicates the on-going evolution of antibody lineages in addition to the constant generation of new SIV-specific B cells from naïve B cells. There were fewer IgM unique sequences isolated from lymph nodes, however sharing of closely related sequences between time-points was observed in many of the samples. Identical sequences were recorded only between 10 and 24 weeks post infection

(average of 0.3%), but there were an average of 6.3% of sequences within a three amino acid difference between 4 and 10 weeks, 13.3% between 10 and 24 weeks, and 9.8% between 4 and 24 weeks. Even in cases where few sequences were recorded in a sample, a fraction of those sequences or their close relatives were recorded at different time points. Thus, there is greater persistence of lineages in IgG memory B cells, but IgM lineages are also observed over the time frame of this study.

Table 3.7 Sharing of CDR3 sequences between animals

SIV-specific			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LN	**IgG**	**gp140**	2.2%	1.7%	22.8%	23.4%	25.8%	25.9%	29.4%	26.0%
LN	**IgM**	**gp140**	5.8%	9.0%	6.7%	9.4%	10.5%	9.7%	23.4%	9.5%
PBMC	**IgG**	**gp140**	1.7%	2.6%	13.4%	27.6%	16.2%	27.0%	23.2%	24.9%
PBMC	**IgM**	**gp140**	2.5%	4.8%	3.9%	5.3%	6.5%	6.8%	16.0%	9.3%
Total memory			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LN	**IgG**	**gp140**	14.2%	8.0%	28.9%	7.2%	32.1%	7.6%	39.5%	7.5%
LN	**IgM**	**gp140**	0.0%	0.0%	0.1%	0.2%	4.2%	1.9%	18.1%	6.0%
PBMC	**IgG**	**gp140**	1.5%	1.8%	6.1%	8.5%	8.8%	8.8%	15.5%	8.4%
PBMC	**IgG**	**gp140**	2.3%	2.6%	4.4%	4.4%	9.6%	5.6%	23.9%	9.6%

Sequencing studies of both IGH and TCR have shown that there are both private, i.e. specific to an individual; and public, i.e. shared between individual, repertoires[351–356]. In spite of the small number of animals sampled here, we were able to identify numerous sequences within the SIV-specific IgG and IgM heavy chains that were similar between individual animals (Table 3.7). An average of 2.2% of all IgG lymph nodes sequences were identical between animals, and an average of 5.8% of IgM lymph node sequences were identical between animals at all time points. The proportion of identical or closely related sequences between animals did not differ significantly between SIV-specific and total memory repertoires. The mean CDR3 length of IgG similar sequences was 13.6, significantly higher ($p<0.0001$) than the total CDR3 length of 12.5 (Fig 3.17 A). However, IgM similar CDR3s were significantly shorter ($p<0.0001$), with a mean of 10.0 versus 11.2 in total IgM sequences (Fig 3.17 B). The mean VH mutation in similar IgG sequences was slightly lower than in total sequences (5.24 vs. 5.54, $p=0.003$). IgM similar sequences had significantly higher ($p<0.0001$) VH mutation than total IgM sequences, with a mean of 5.9% compared to 4.9% in total IgM. Similar IgG sequences showed accumulation of mutations in V genes that mirrored the total sequences, increasing from a mean of 5.25% at 4 weeks to 6.5% at 24 week. However, similar IgM sequences also showed an increase in VH mutation, from 5.5% at 4 weeks to 6.6% at 24 weeks in IgM. Total IgM sequences did not show this increase, with an average of 5.1%

mutation t 4 weeks and 4.7% at 24 weeks. The V and J gene proportional distributions were significantly different in the common repertoires for both IgM and IgG sequences (Fig 3.18). IgG similar sequences had a higher proportion of VH1 and VH5 genes, and JH1 and JH6 genes, than the total IgG repertoire (Fig 3.18) IgM similar sequences has significantly higher VH4, JH6, and fewer V2, V5, and J5-1 than total IgM (Fig 3.18). When we expanded our filtering to include sequences within a three amino acid difference in the CDR3 region, an average of 29.4% of IgG and 23.4% of IgM sequences were similar between animals. In comparing sequences in the PBMC, an average of 1.7% of IgG and 2.5% of IgM sequences were identical, expanding to 23.2% and 16.0% of sequences within 3 amino acids. These data indicate common features of the SIV-specific response between animals.

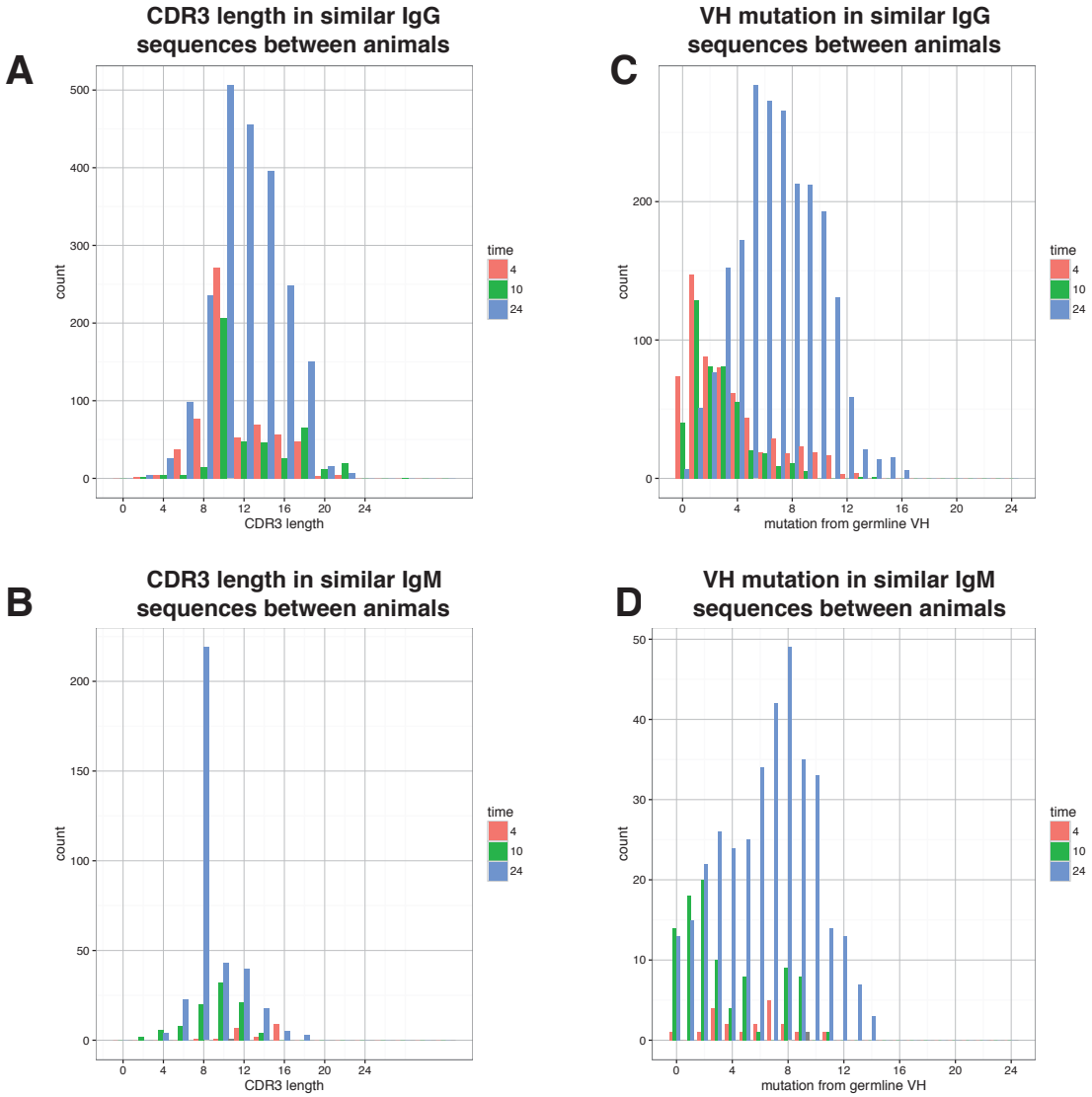


Fig 3.17 CDR3 Length and VH mutation in similar sequences between animals. (A) CDR3 length in similar IgG sequences (B) CDR3 length in similar IgM sequences (C) Percent divergence from V gene germline in heavy chains in similar IgG sequences (D) Percent divergence from V gene germline in heavy chains in similar IgM sequences

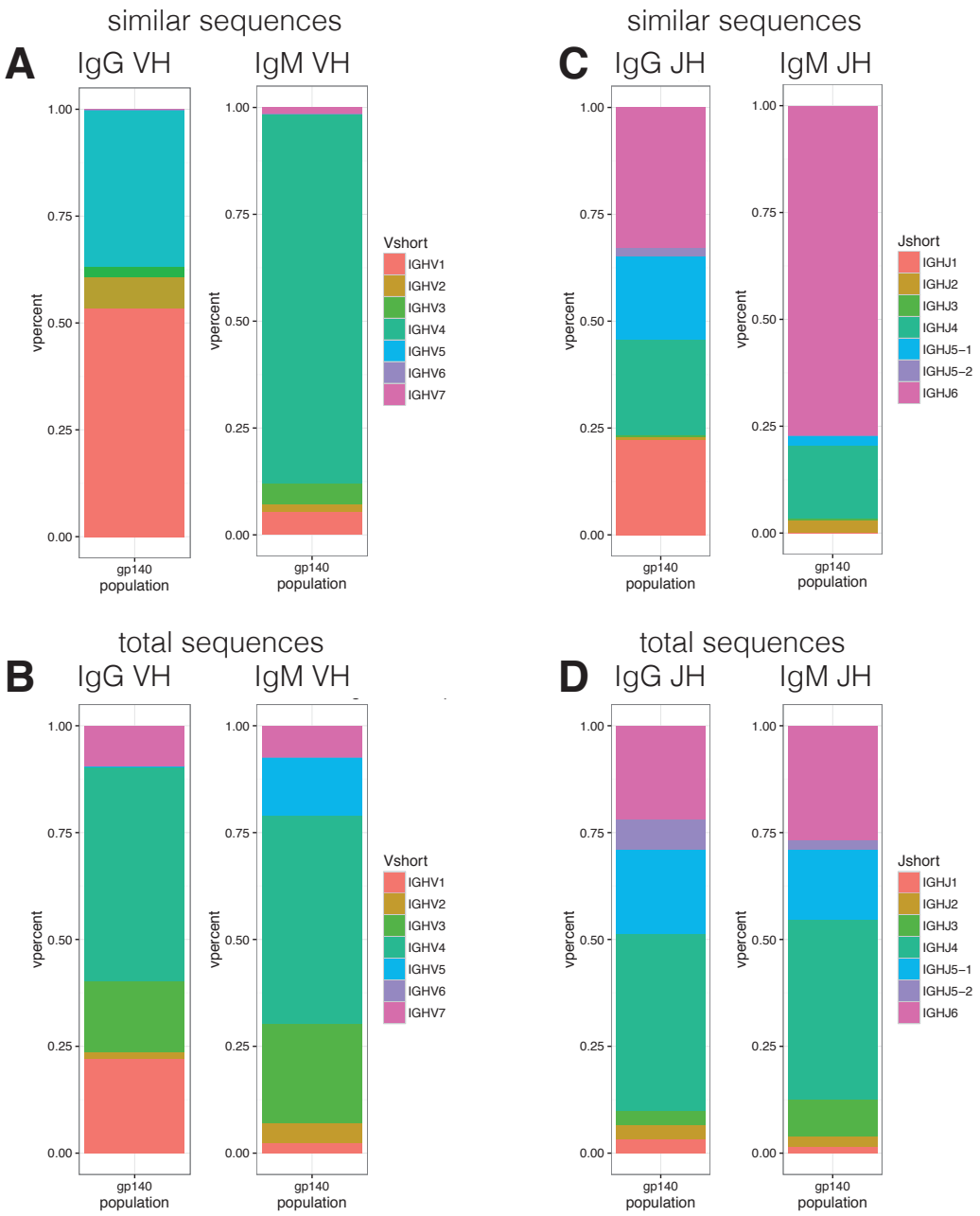


Fig 3.18– VH and JH gene frequencies in similar sequences. (A) VH gene frequencies in similar IgG and IgM sequences. (B) JH gene frequencies in similar IgG and IgM sequences. (B) VH gene frequencies in total IgG and IgM sequences. (D) JH gene frequencies in total IgG and IgM sequences.

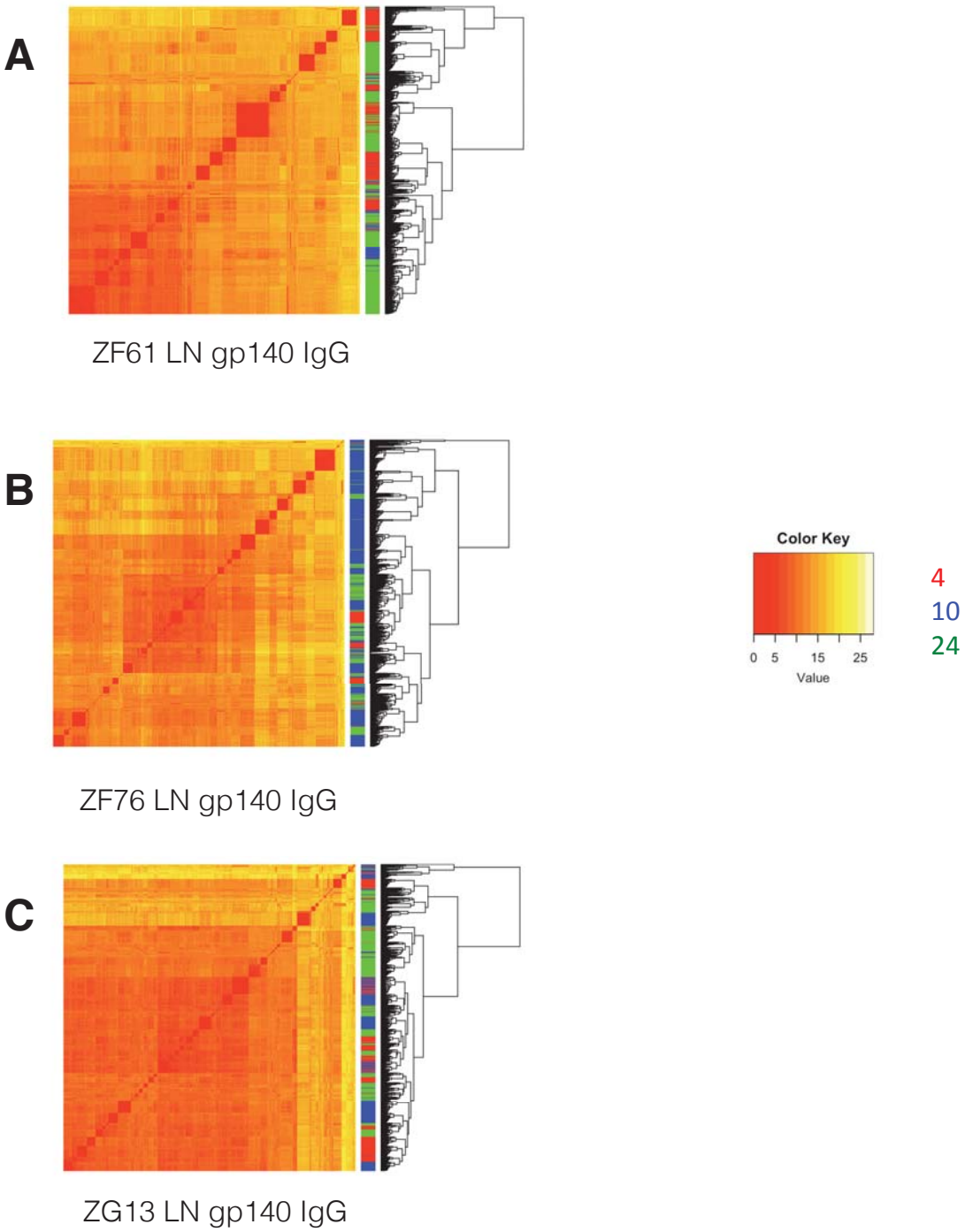


Fig 3.19 Sharing of CDR3 and V-J gene identical sequences between samples in lymph nodes. Hierarchical clustering of lymph node SIV-specific IgG heavy chain sequences from 4, 10, and 24 weeks post infection in animals (A) ZF61, (B) ZF76, and (C) ZG13. Colored bar on right indicates whether sequences are from week 4 (red), 10 (blue), or 24 (green).

Hierarchical clustering of IgG CDR3 sequences from lymph node samples taken at 4, 10, and 24 weeks shows the relative genetic distance between sequences. In all three animals, clusters of closely related lymph node sequences (indicated by blocks of red and orange) are made up of sequences from all three time points, indicated by sidebars of red, blue, and green (Fig 3.19). These clusters represent lineages of heavy chain sequences that persist and evolve over

the course of infection, and that related sequences are located at different times and not restricted to a single sampling point. PBMC IgG sequences showed similar clustering, except in cases where there were significantly lower numbers of heavy chain sequences recovered from a sample (eg.ZF61 IgG PBMC week 10, Fig. 3.20 A). In the remaining two animals, sequences from all three time points were evenly distributed in small clusters, with no blocks of sequences restricted to a single time point (Fig 3.20 B, C). These plots illustrate the presence of individual lineages that persist over the period of sampling, and show that the sequences sampled here are not primarily clustered by time point. Most samples cluster with similar sequences from other time points and indicate closely related B cell linages. While these samples represent only a small fraction of the circulating SIV-specific B cells, these plots show that we are still able to detect the presence of related B cells that persist throughout infection.

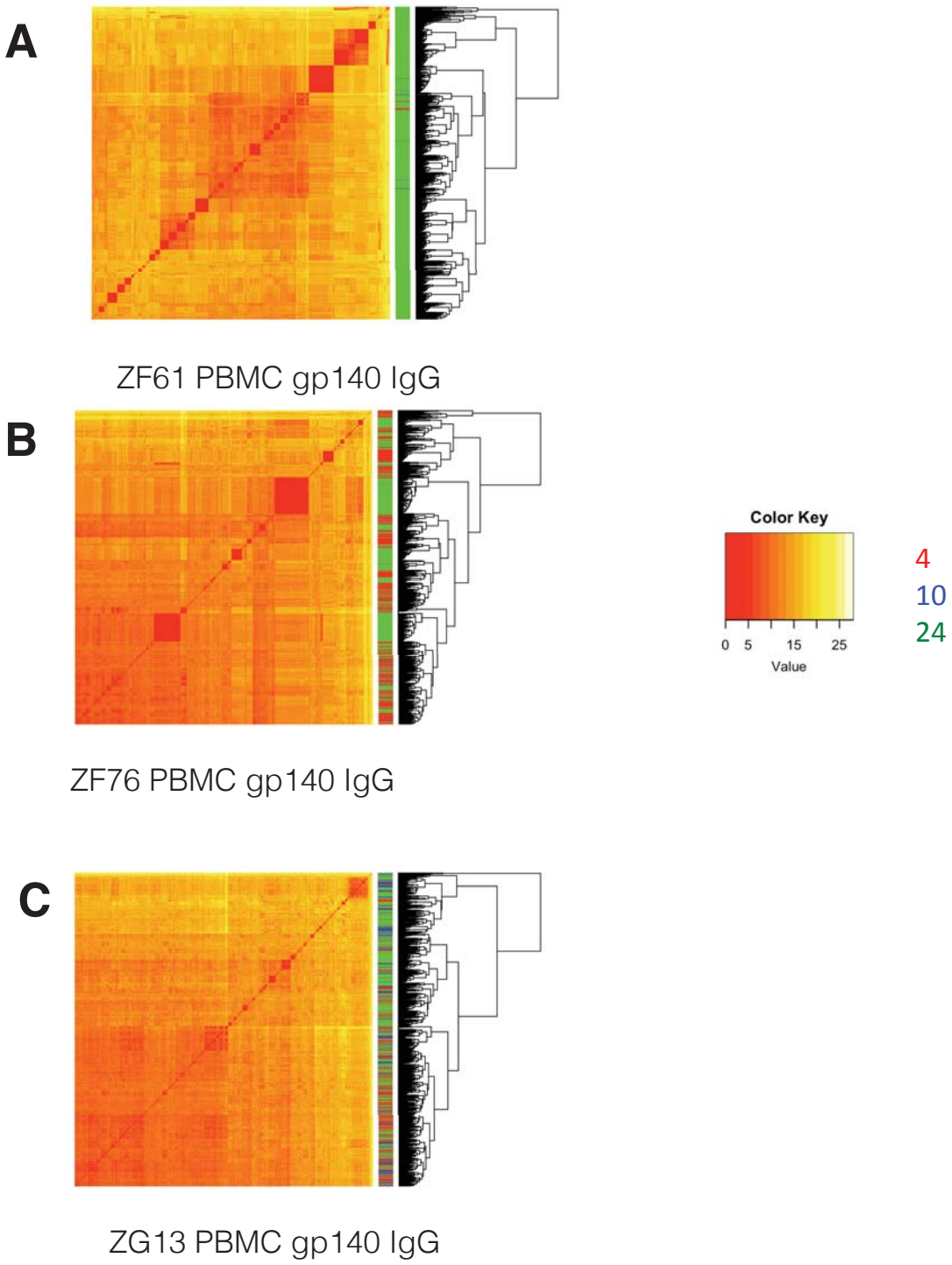


Fig 3.20 Sharing of CDR3 and V-J gene identical sequences between samples in PBMC. Hierarchical clustering of PBMC SIV-specific IgG heavy chain sequences from 4, 10, and 24 weeks post infection in animals (A) ZF61, (B) ZF76, and (C) ZG13. Colored bar on right indicates whether sequences are from week 4 (red), 10 (blue), or 24 (green).

3.2.9 Isoelectric point and charge distinguish HCDR3 in antigen-specific memory cells

The physicochemical properties of the CDR3 regions can influence the binding of the antibody to its target[357]. Furthermore, they can be used to indicate if the antibody response to

an antigen is targeted beyond engaging a set of VDJ germline genes, and distinguish IgG and IgM memory responses[235,358]. We performed a computational analysis of four summary properties of the HCDR3 amino acid sequences in SIV-specific and non-specific IgG and IgM memory B cells in PBMC, LN and bone marrow. Hydrophobicity is computed using the GRAVY (grand average of hydropathy) using the Kidera method[359], where scores above zero are hydrophobic and scores below zero are more hydrophilic. Isoelectric point is the pH at which the CDR3 region carries no net charge: proteins with basic amino acids will have a high pI, whereas proteins with many acidic amino acids will have a low pI. It is calculated using the pKa of individual amino acids. The net charge of the CDR3 region was calculated using the EMBOSS method[360]. Finally, the aliphatic index calculates the relative volume occupied by alanine, valine, leucine, and isoleucine[361]. High aliphatic index indicates higher thermostability of globular proteins.

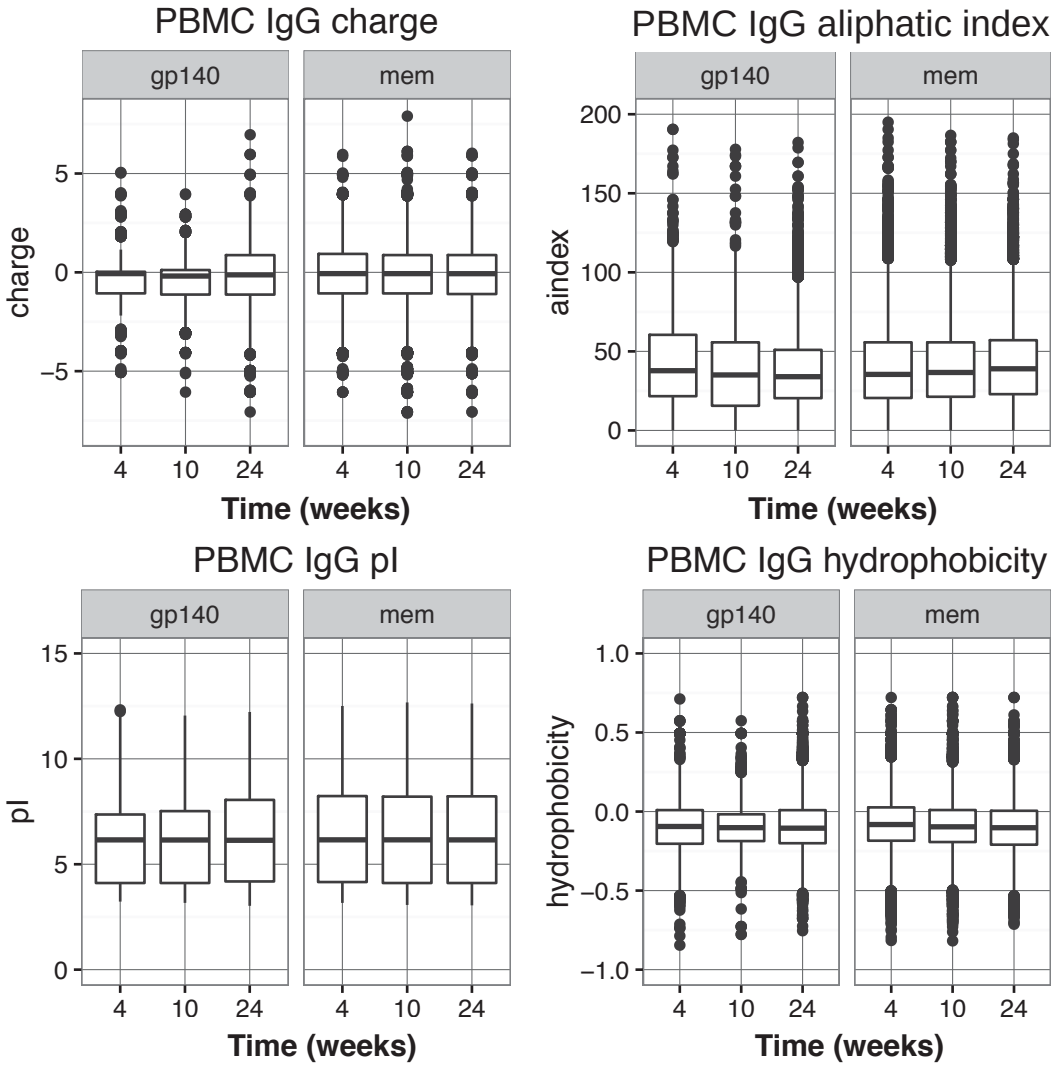


Fig 3.21 Physiochemical features of CDR3 sequences in PBMC IgG B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from PBMC IgG samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

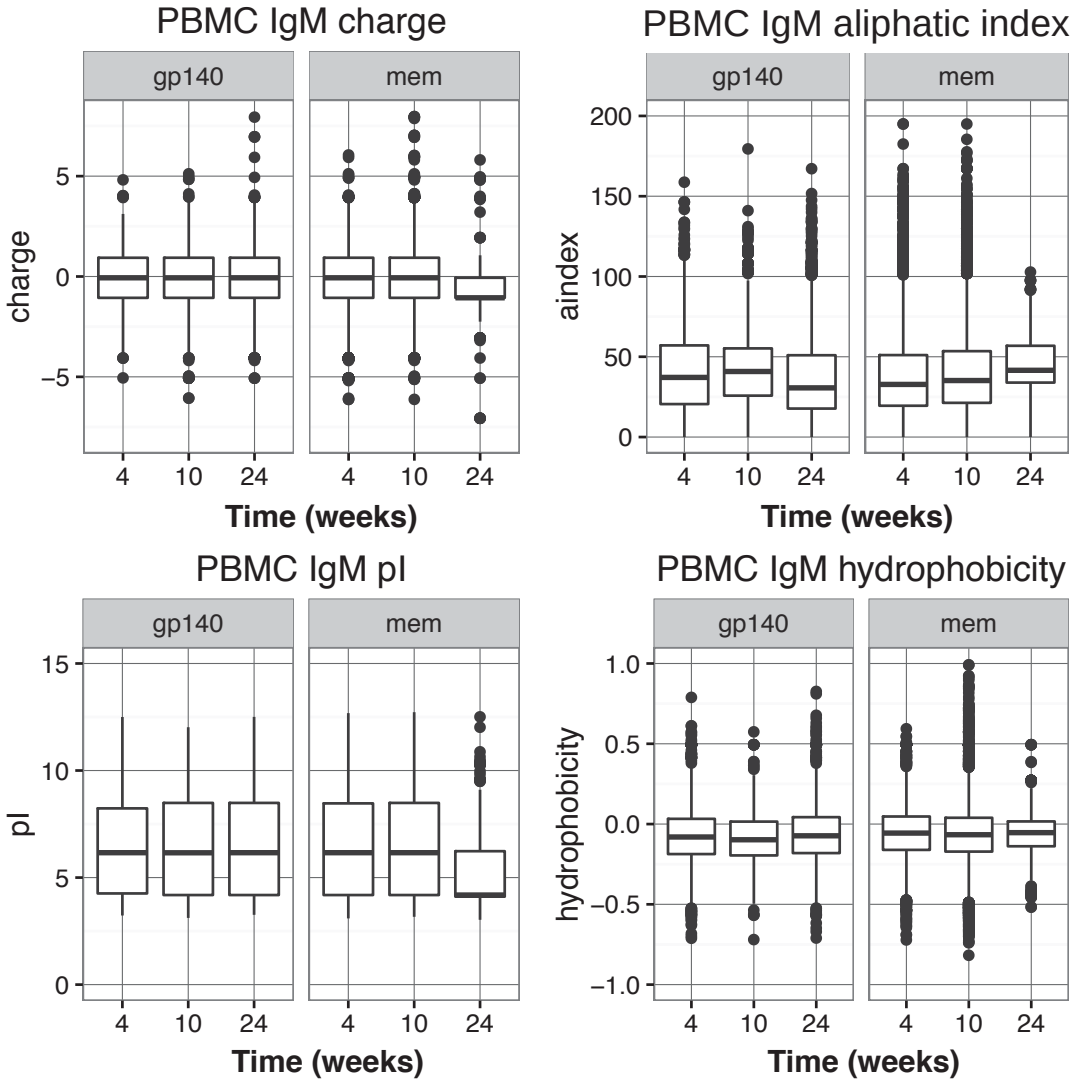


Fig 3.22 Physiochemical features of CDR3 sequences in PBMC IgM B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from PBMC IgM samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

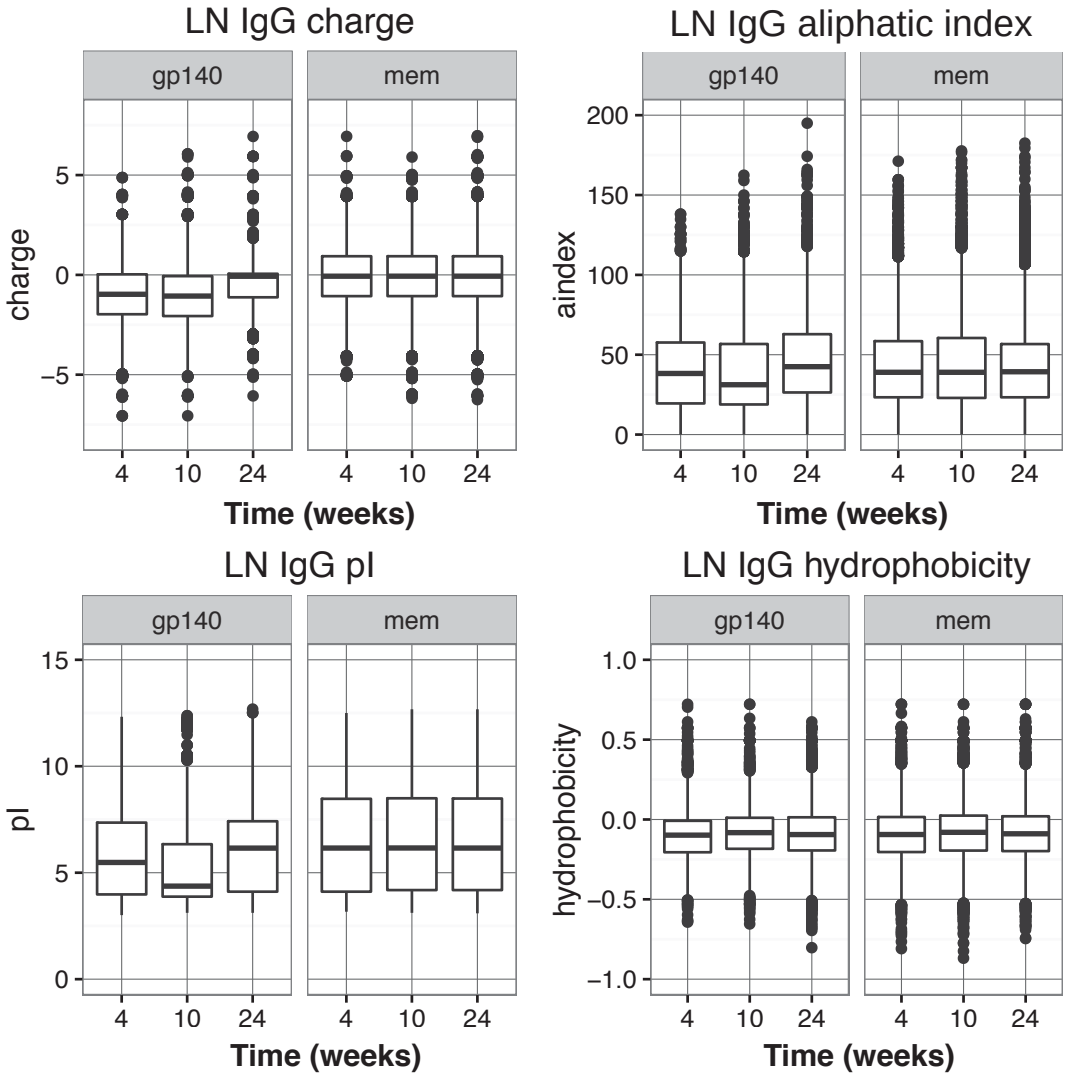


Fig 3.23 Physiochemical features of CDR3 sequences in lymph node IgG B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from lymph node IgG samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

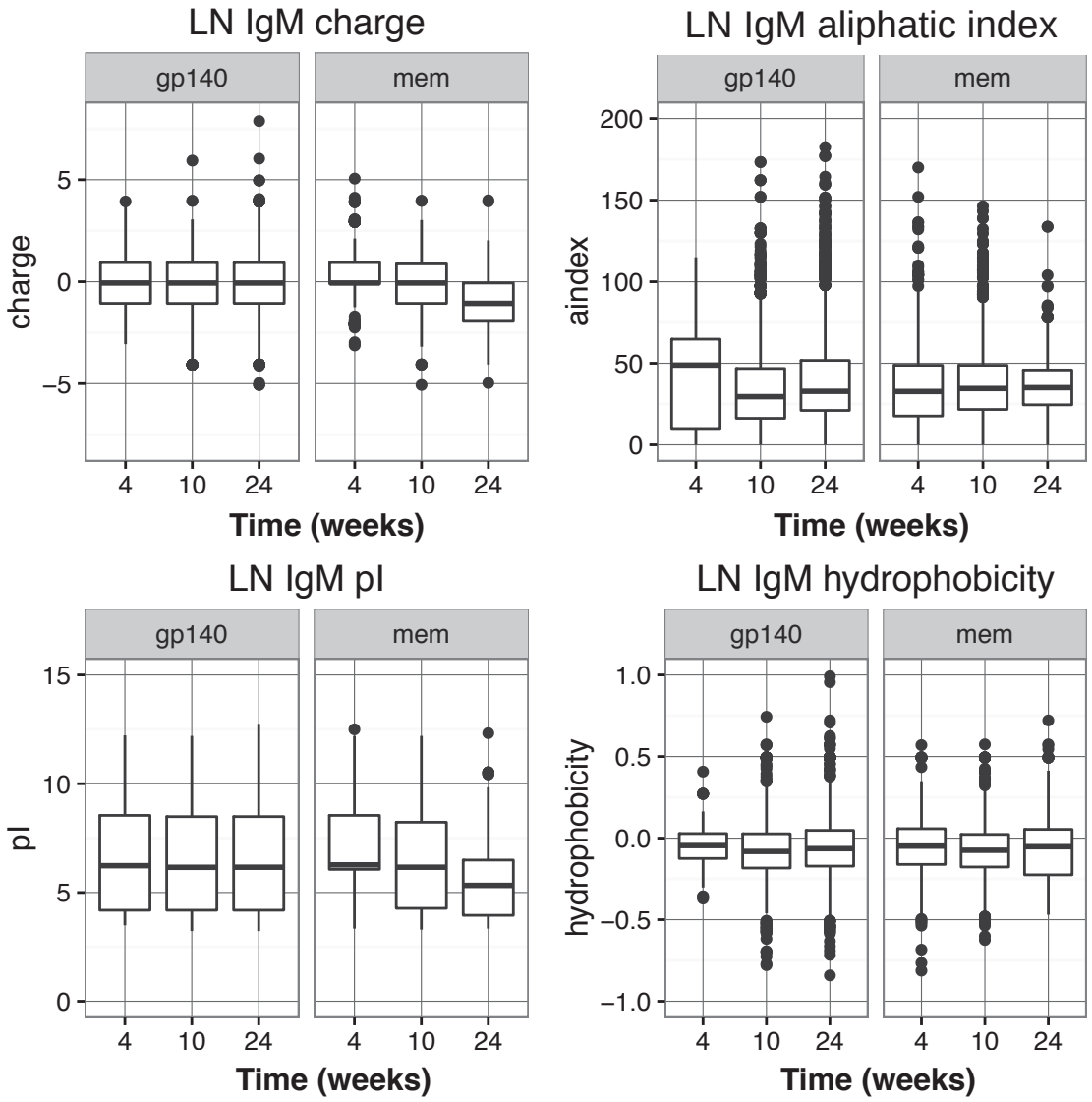


Fig 3.24 Physiochemical features of CDR3 sequences in lymph node IgM B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from lymph node IgM samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

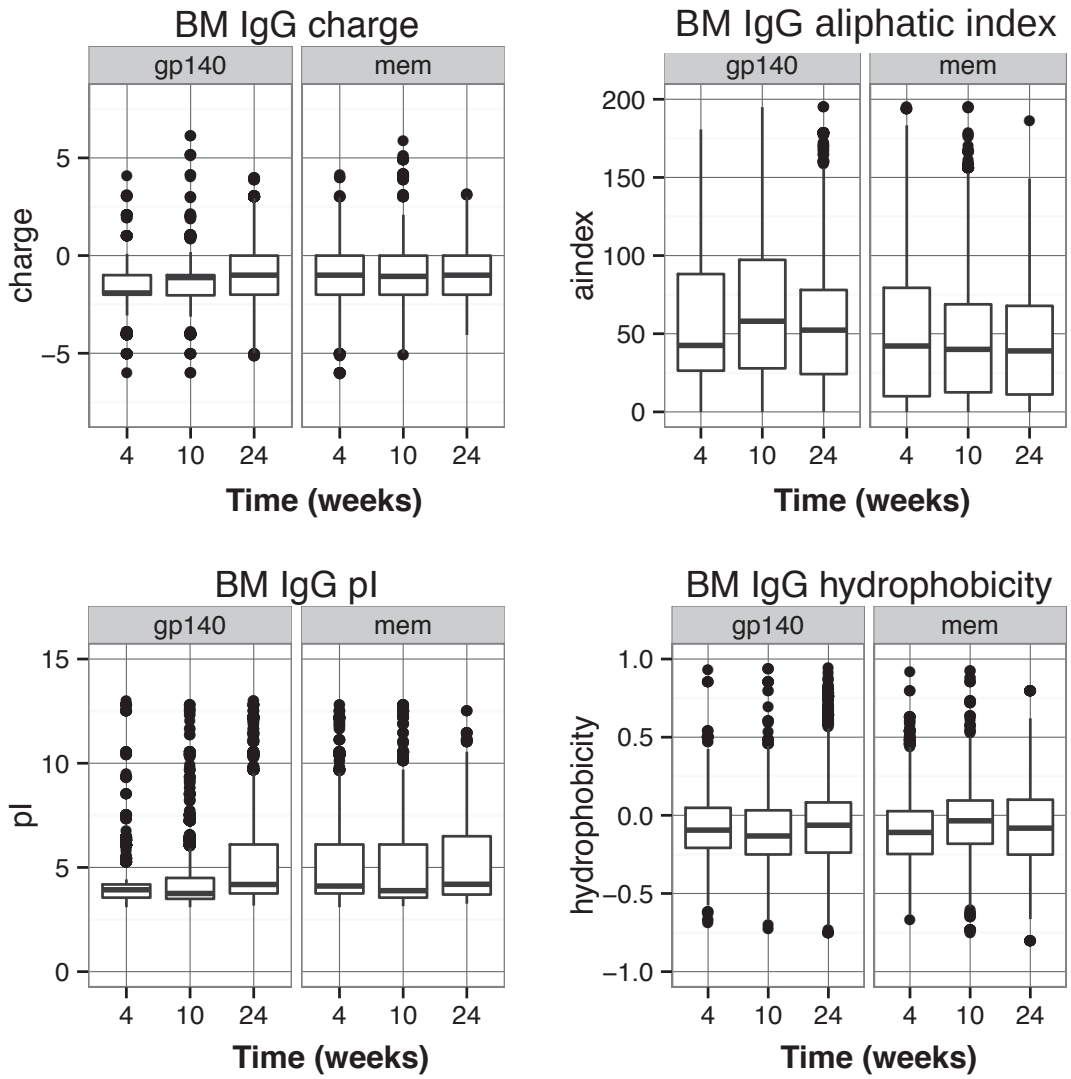


Fig 3.25 Physiochemical features of antigen-specific CDR3 sequences in antigen-specific B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from bone marrow IgG samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

Sequences from non-specific IgG PBMC were unchanged in hydrophobicity, net charge, and isoelectric point. There was a small shift in aliphatic index (from 35 at 4 weeks to 39 at 24 weeks). SIV-specific IgG PBMC sequences had no different in hydrophobicity and isoelectric point at the three time points sampled (and were equivalent to non-specific sequences), but had an increase in net charge over the course of infection (from -0.065 at 4 weeks to -0.13 at 24 weeks) and a decrease in median aliphatic index (from 38 at 4 weeks to 34 at 24 weeks) (Fig. 3.21) Median aliphatic index in IgM memory PBMC sequences increased over the course of infection, from 33 at 4 weeks to 42 at 24 weeks, while in SIV-specific sequences aliphatic index decreased from 37 at 4 weeks to 31 at 24 weeks (Fig. 3.22). SIV specific sequences had slightly lower hydrophobicity than non-specific sequences, and net charge and isoelectric point were largely unchanged between the two groups (except at 24 week non-

specific IgM sequences, which had lower median isoelectric point and lower median net charge).

The aliphatic index of IgG sequences in lymph nodes was unchanged in non-specific sequences (median 39) but varied in SIV-specific IgG sequences, with a median of 38 at 4 weeks, 31 at 10 weeks, and 43 at 24 weeks (Fig. 3.23). Hydrophobicity was unchanged between SIV-specific and non-specific sequences at all time points, with medians between -0.080 and -0.098. No difference in hydrophobicity between Env-specific and total memory IgG B cells was previously reported by Wang et al (2016)[362] and our analysis confirms this finding. Net charge was unchanged in non-specific sequences, but increased in SIV-specific sequences from -0.97 at 4 weeks to -0.068 at 24 weeks and was indistinguishable from non-specific sequences at 24 weeks (-0.065). Isoelectric point was similar unchanged in non-specific sequences but varied in SIV-specific sequences, with a median of 5.5 at 4 weeks, 4.4 at 10 weeks and 6.2 at 24 weeks (compared to 6.2 in all non-specific sequences). IgM lymph node sequences showed no variation in aliphatic index or hydrophobicity (Fig. 3.24), but at 24 weeks lower median net charge (-1.06 vs. -0.064 at 4 weeks and 10 weeks) and median isoelectric point (5.3 vs. 6.3 and 6.2). In contrast, SIV-specific IgM lymph node heavy chain sequences, aliphatic index decreased from 49 at 4 weeks to 30 at 10 weeks and 33 at 24 weeks, while hydrophobicity, charge and isoelectric points remained constant throughout infection.

In bone marrow IgG sequences, isoelectric point was unchanged in non-specific heavy chains but increased from a median of 4.4 to median 6.1 in SIV-specific sequences (Fig 3.25). Median charge also increased in SIV-specific sequences, from median -1.06 at 4 weeks to -0.19 at 24 weeks. Hydrophobicity and aliphatic index were unchanged from SIV-specific to non-specific and throughout infection. Both SIV-specific and non-specific IgM bone marrow samples had a lower isoelectric point than IgM and IgG lymph node and PBMC samples at all time points. Bone marrow IgM SIV-specific samples also had a lower charge than both lymph node samples and PBMC samples, with IgM samples (both PBMC and lymph node) having higher charge than IgG samples.

3.2.10 Single cell sorting of B cells and comparison with bulk sorting

To evaluate the efficiency of the 5’ RACE bulk PCR, we re-sorted SIV-specific IgG memory B single cells from a PBMC sample and used a primer-based method to sequence their heavy chains. After sorting four plates (384 cells) and amplifying heavy chains using a nested PCR

with V-gene specific primers, 181 wells showed amplified sequences by gel electrophoresis. These wells were Sanger sequenced, and 165 had full-length sequences, of which 154 were productive rearrangements. There were 102 unique V-D-J combinations and 146 unique CDR3s for a total of 147 unique VDJ-CDR3 combinations. Cells from the same sample were sorted in bulk and amplified using the 5' RACE method, with a total of 1974 cells sorted yielding 145 unique VDJ combinations and 608 unique CDR3s (Table 3.8).

Table 3.8 Shared sequences between bulk and single cell sequencing.

Sample	Cells Sorted	Unique numseqs	Unique V	Unique J	Unique D	Unique VDJ	Number of CDR3lengths	Unique CDR3aaseqs	Unique vdj-CDR3
PBMCsc	288	165	12	8	26	102	19	146	147
PBMCbulk	1974	9651	15	9	30	144	18	603	697

Though the numbers of cells sequenced here represent only a small fraction of the SIV-specific repertoire, the distributions of V, D, and J genes not significantly different between the two methods of sequencing (Fig 3.15). Sequences were dominated by VH1, VH3, and VH4, with a small number of VH2, VH5, and VH7 sequences. This is in accordance with published comparisons of bulk and single cells sequencing[347]. The JH gene distribution had a bias toward J4 in the single cell that was not observed in the bulk samples. DH genes had high frequencies of families DH3, DH4, and DH6 in all samples, with single cell sequences also having higher frequencies of DH1 and DH2 than the bulk samples. There was a good correlation (p-0.0079 and p=0.0085) between the frequencies of the VH genes and DH genes in the two samples.

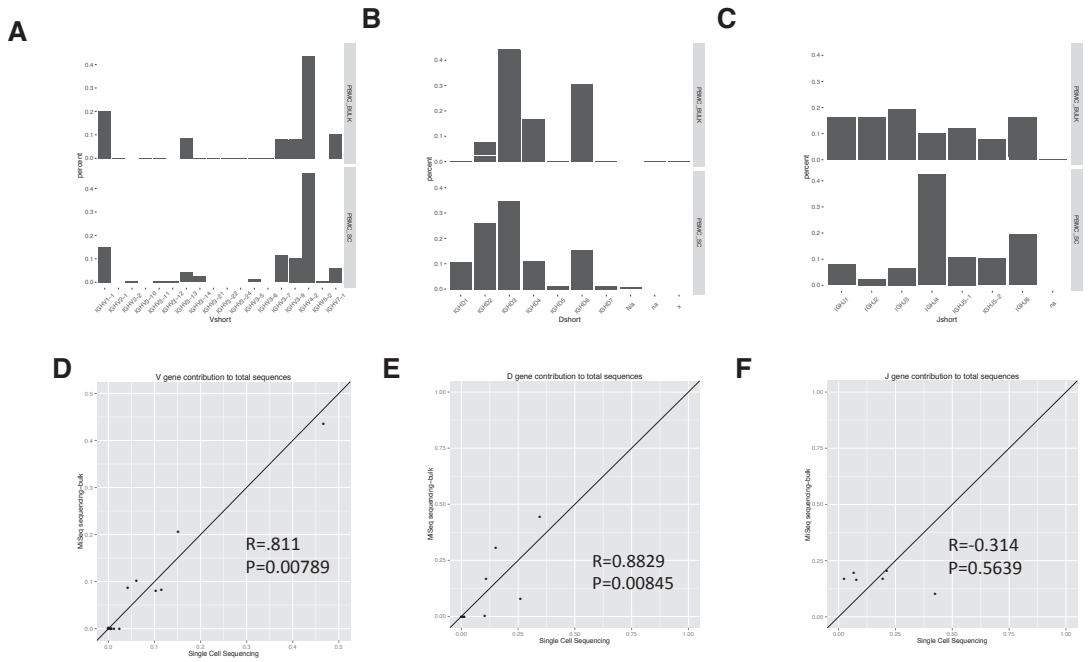


Fig. 3.26 V, D, and J gene frequencies in bulk and single cell sequencing. (A) VH gene proportion distribution from bulk (top) and single cell (bottom) SIV-specific PBMC. (B) DH gene proportion distribution from bulk (top) and single cell (bottom) SIV-specific PBMC. (C) JH gene proportion distribution from bulk (top) and single cell (bottom) SIV-specific PBMC. (D) Correlation between proportion distributions of VH genes in bulk and single cell sequencing. (E) Correlation between proportion distributions of DH genes in bulk and single cell sequencing. (F) Correlation between proportion distributions of JH genes in bulk and single cell sequencing.

The depth of sequencing in the bulk 5' RACE method yields data on the relative frequencies of individual clones, with identical sequences from multiple cells occurring at higher frequencies than reads from unique single cell BCR heavy chains. The sequences with the highest read counts showed some overlap with the single cell sequences, although there were some sequences that were detected in the bulk that were not detected in the single cell, and vice versa. This was partially due to mismatches between the V gene primers (used in single cell sequencing) and sequences in the bulk samples, with the result that the V gene primers were unable to properly amplify BCR heavy chains that were detected in the bulk samples (Fig 3.27). Three of the heavy chain sequences present at high frequencies in the bulk samples (16%, 10% and 8%) had at least one mismatch between the VH sequence and the inner and/or outer VH primers used in the single cell method. Thus, while the single cell method has the advantage of downstream pairing of heavy and light chains, primer mismatches can result in mutated sequences not being detected.

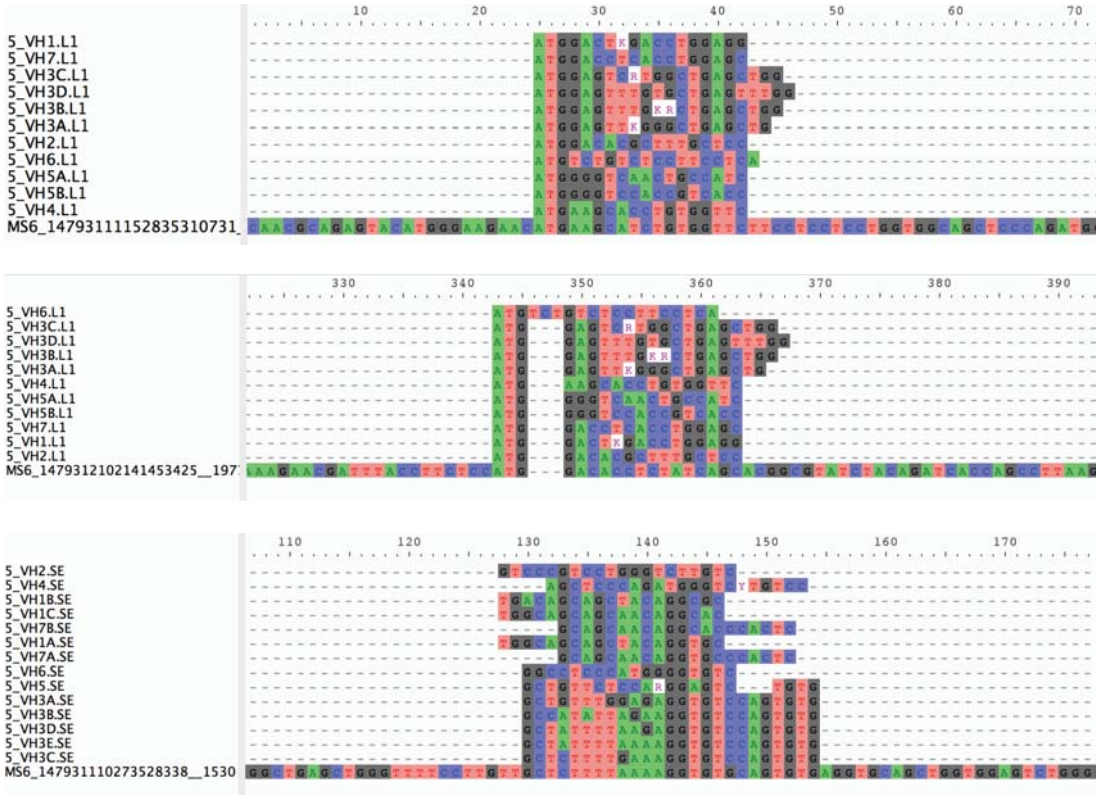


Fig 3.23 Single cell primer mismatches in sequences from bulk sequencing. Mismatches in the VH primer sets and the V gene sequences from heavy chains present at high frequencies in the bulk sequencing data.

3.3 Conclusions

This chapter summarizes the sequencing of SIV-specific and non-specific memory B cells throughout SIV infection of three rhesus macaques. One of the advantages of the macaque SIV infection model is the ability to serially sample compartments that are inaccessible in human studies, most prominently bone marrow and lymph nodes, to better capture the development of the humoral response to SIV. We identified IgG and IgM memory B cells that bound an SIV gp140 envelope protein probe and sequenced the heavy chain repertoires in blood, lymph nodes, and bone marrow. SIV-specific IgG B cells increased as fraction of all memory B cells over the course of infection in blood and lymph nodes, while SIV-specific IgM B cells were present at a steady and lower proportion. Median CDR3 length was unchanged in SIV-specific IgG and IgM heavy chains over the course of infection and was not significantly different than non-SIV specific memory B cells. However, the average mutation in the VH gene of SIV-specific IgG heavy chains shifted over the course of infection with an accumulation of more highly mutated heavy chains later in infection. While we observed some differences in the proportion distributions of VH and JH genes, we did not identify a particular signature of the SIV-specific response. We used several diversity measures to track the proportion and size of related antibody lineages, finding that SIV-specific IgG B cells were dominated by a small number of clones and had low diversity and low evenness early in infection, but resembled the non-specific memory repertoire later in

infection. Identical and closely related sequences were found across tissues and over the three time points sampled in this study, and a small percentage of similar IgG SIV-specific sequences were found in multiple animals. CDR3 sequences from SIV-specific IgG heavy chains in lymph nodes had small differences in net charge and isoelectric point when compared with non-SIV specific sequences. Finally, in a limited comparison of bulk 5' RACE PCR based sequencing versus single cell VH-primer directed sequencing, we identified efficiency of the single cell method and primer mismatches as primary reasons for differences in repertoires sequenced using the two methods.

4 Dynamics of SIV infection of follicular T cells

4.1 Introduction and aims

HIV/SIV infects primate cells using CD4 alongside CCR5 and CXCR4 as co-receptors[16,20,21]. However, not all CD4 T cells are equally infected: it is now well established that HIV preferentially infects memory T cells[363] and HIV-specific CD4+ T cells. Efforts to further define the features of infected cells have focused on chemokine, activation, and localization markers in search of signatures of preferentially infected T cell subsets. Resting memory CD4 T cells are highly infected[364], with memory CD57- CD4+ T cells containing more HIV gag DNA than terminally differentiated CD57+ T cells. CCR6+RORγt+ Th17 cells were shown to be preferentially infected in very acute SIV infection of macaques in vaginal transmission of macaques[365], and CCR6+CCR4+ and CCR6+CXCR3+ CD4 T cells were found to be highly permissive to HIV-1 infection [366].

Recently follicular helper T cells (TFH), located in secondary lymphoid organs, are cited as a major compartment of HIV/SIV infection in CD4+ T cells[170,172,174]. Preferential infection of TFH, which are critical to antibody development and maturation, presents an additional impairment to the immune response to HIV/SIV. TFH provide B cell help in germinal centers to promote somatic hypermutation and class switch recombination, key components of the development of an effective humoral immune response. In a study by Yamamoto et al [367], the quality and quantity of TFH were closely related to the induction of broadly neutralizing antibody responses in a SHIV model of infection. Work by Cubas et al[173]. showed that TFH in chronic HIV infection are unable to provide adequate B cell help. Finally, recent work has shown that the B cell follicle may represent a sanctuary of productive infection in elite controller macaques that is protected from effective CTL, with on-going viral replication in TFH located in the follicle and germinal centers[174]. Questions remain about the timing of infection of TFH and how big of a role they play in supporting HIV/SIV replication throughout infection. Here, we aim to measure the contributions of follicular T cells and other CD4+ T cell subsets to the overall pool of virus and infected cells, from initial infection through peak viremia to viral setpoint. We profile changes in activation and chemokine markers on CD4+ T cells throughout early infection and measure SIV DNA and RNA in memory T cell subsets to determine the phenotype and characteristics of productively and persistently infected CD4 T cells in SIV infection.

The aims of this chapter are to:

- Profile changes phenotypic changes in chronically infected HIV PBMC and acutely and chronically SIV-infected PBMC and lymph nodes
- Design primers, probes, and standards to quantify spliced SIV RNA
- Investigate CD4 downregulation as a phenotypic marker of active SIV infection
- Identify productively infected CD4 T cells using qPCR and SIV RNA probes in conjunction with flow cytometry phenotyping panels

4.2 Results

4.2.1 Chemokine and activation markers on CD4 T cell subsets throughout HIV/SIV infection

HIV/SIV infection induces massive changes in the inflammation and activation of T cells in lymph nodes, including alterations in T cell trafficking; changes in architecture of lymph nodes, bone marrow, and thymus; and shifts in the relative proportions of naïve and memory T cells [23]. We first wanted to examine, using flow cytometry profiling of surface markers, changes in the activation and chemokine markers on subsets of CD4 T cells in acute and chronic infection. Flow cytometry panels were designed and tested to ensure distinct separation of markers for longevity (CD127/ILR-a), activation (ICOS, CD25, PD-1), and trafficking (CCR6, CCR4, CXCR3). CD127 plays an important role in T cell development and antigen-specific responses, and is down-regulated upon expansion[368,369]. ICOS is expressed on activated T cells and enhances proliferation, cytokine secretion, and secretion of antibodies by B cells[370]. PD-1 is a negative regulator of T cells and plays an important role in infectious immunity, tumor immunity, and autoimmunity[371]. CD25 is part of the IL-2 receptor and marks a variety of activated cells[372]. CCR6 is the receptor for MIP3-α, is expressed on Th17 cells, and plays a role in their migration into inflamed tissues[373]. CCR4 is associated with Th2 cells while CXCR3 is associated with Th1 cells and both play a role in trafficking activated cells to sites of inflammation[374,375].

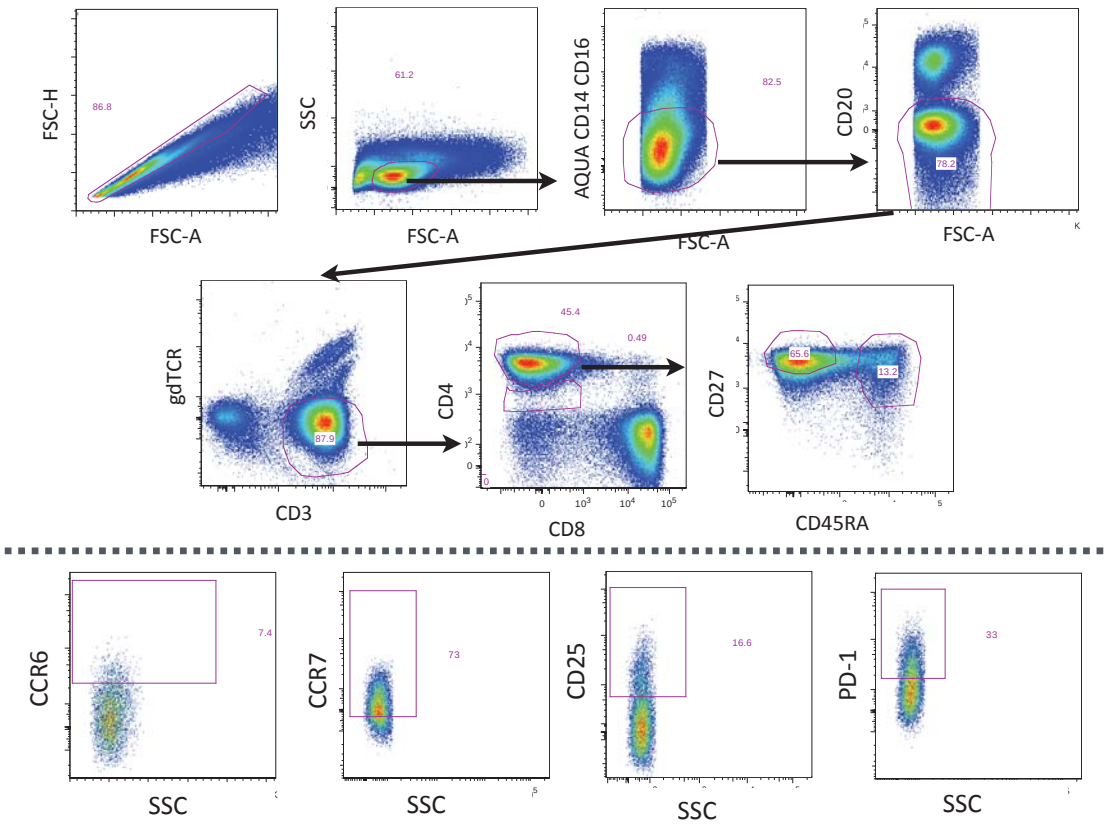


Fig. 4.1 Representative gating schematic for phenotyping of HIV-infected PBMC. PBMC were stained and acquired on a flow cytometer, gating on forward/side scatter, lymphocytes, live CD14-CD16-CD20-gdTCR-CD3+ T cells, and subsequently on CD8- CD4 bright and CD8- CD4 dim expression. Expression of CCR6, CCR7, CD25, and PD-1 was measured on CD45RA- CD27+ memory T cells.

4.2.1.1 Expression patterns in HIV-positive PBMC

In an initial assessment of T cell phenotype in infection, we profiled five PBMC samples from chronically HIV-infected, virally suppressed donors on combination antiretroviral therapy and compared them to five uninfected controls. CD4 is down-regulated during active infection[52], and by examining CD4 dim cells (that are CD3+ CD8-) separately from CD4 bright cells we hypothesized that we might distinguish features of infected cells by examining an enriched population (Fig 4.1). There were fewer CCR6+ CD4 bright T cells in HIV infection than in uninfected controls (means of 6.1% in HIV positive samples and 17.1% in uninfected controls), but no difference in expression on CD4 dim cells (Fig 4.2). CCR7 expression was unchanged between CD4 high and dim cells in HIV positive samples but in HIV negative samples, was lower on CD4 high cells (43.2%) than on CD4 dim cells (66.1%). Though it did not reach statistical significance, we observed lower expression of CD25 in both populations in HIV infected samples than in uninfected samples (means of 13.7% and 14.3% compared to 19.4% and 23.0%). This is in agreement with published work showing depletion of CD25+ T cell subsets in HIV infection [376]. PD-1 expression was significantly higher in CD4 high HIV positive samples (34.4%) than CD4 dim uninfected samples

(11.9%), and higher than CD4 high uninfected (17.9%) and HIV infected CD4 dim (20.0%). Overall, HIV positive samples were characterized by higher expression of CCR7 and PD1 and lower expression of CCR6.

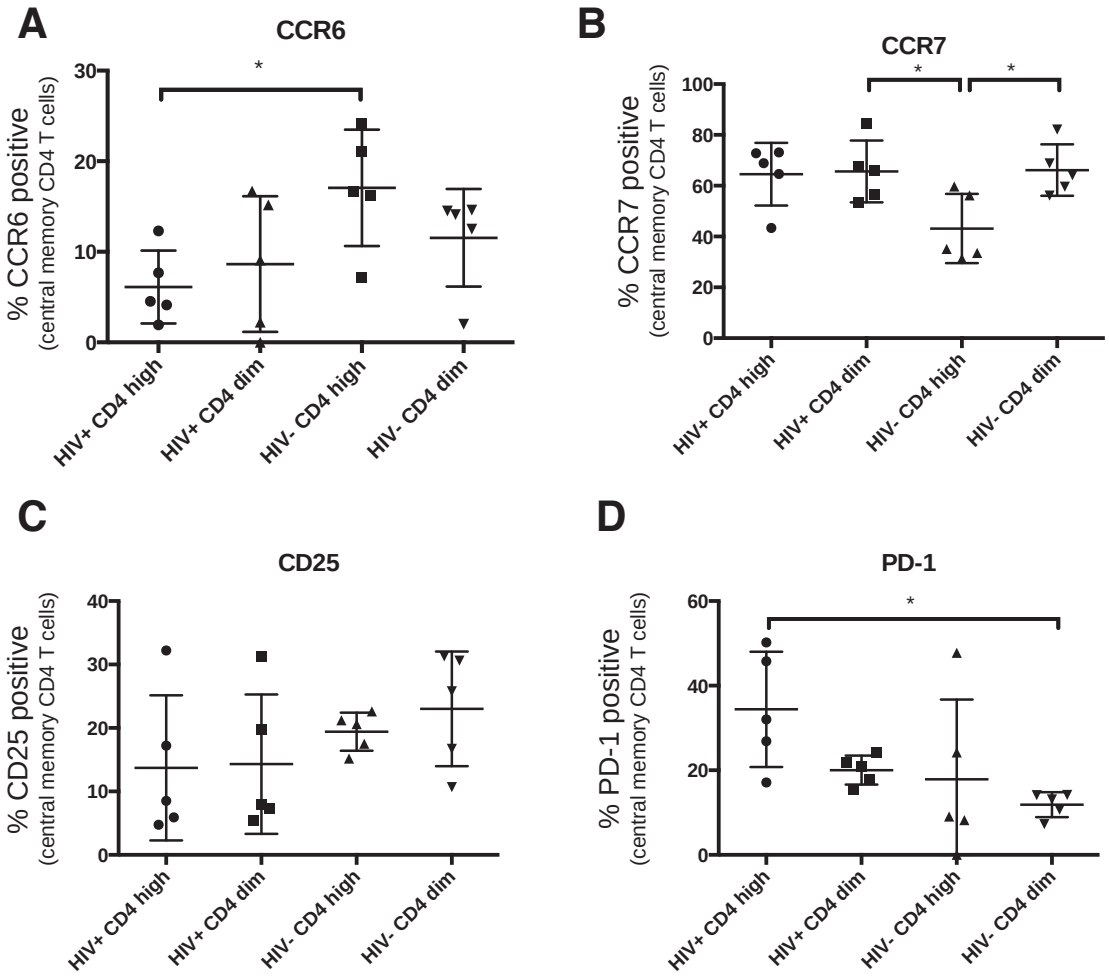


Fig 4.2 Expression patterns of cell surface receptors in HIV positive and negative PBMC. Percentage of central memory CD4 T cells expressing (A) CCR6, (B) CCR7, (C) CD25, and (D) PD-1 on CD4 high and dim populations in PBMC from HIV infected and uninfected donors. Significant differences were determined using a one way ANOVA test.

4.2.1.2 Expression patterns on uninfected, acutely SIV-infected, and chronically SIV-infected lymph node samples

To further investigate changes in T cell phenotypes during SIV infection, we stained and analysed lymph nodes from rhesus macaques infected with SIVmac251, with biopsies taken at the acute stage (5-14 days), chronic stage (6 months to 2 years), or uninfected controls. We looked at expression of chemokine and activation markers on CD4 T cells and CD3+ CD4 dim cells (to capture actively infected cells), and within the central memory compartment at three populations: TFH (defined here as PD1++CXCR5++ cells), CXCR5+ cells, and CXCR5- cells (gating strategy detailed in Fig 4.3). CXCR5 and PD-1 together identify CD4+ T cells located in B cell follicles, but without additional markers it is impossible to segregate follicular helper T cells (TFH) from follicular regulatory T cells (TFR), which do not directly

provide B cell help. In this work, the term “TFH” is used to refer to both CXCR5+PD1+ and CXCR5++PD1++ CD4+ T cells and is inclusive of TFR.

First, we looked at expression of CCR4 was lower on CD4 T cells from chronically infected macaques than uninfected controls (mean of 24% vs. 48%), with acute samples slightly higher than chronic samples with mean expression of 36% (Fig 4.4A). Uninfected samples had higher expression of CCR6 (21%) than acutely infected (10%) or chronically infected (10%) samples (Fig 4.4B). CXCR3 expression was lower on acutely infected samples (28%) than on chronically infected samples (44%) (Fig 4.4C). SLAM expression was higher in chronic samples than uninfected controls (32% vs. 21%) and acute samples had a wide range of SLAM expression (from 10% to 35%) (Fig 4.4D). There was no significant difference in expression of CD25 between samples, but acutely infected samples had significantly higher expression of CD127 (85%) than chronically infected samples (55%) or uninfected samples (66%) (Fig 4.4 E-F) Finally, ICOS expression was lower on acute samples (26%) than chronic samples (54%) (Fig 4.4G). In CD4 dim cells, there was no difference in expression of CCR6, CXCR3, or SLAM between acute, chronic, or uninfected samples (Fig 4.5B-D). CCR4 expression and CD127 expression were significantly lower in chronic infection on CD4 dim cells than in acute or uninfected samples (Fig 4.5 A and F), but ICOS expression was significantly higher in chronically infected samples (Fig 4.5 G). ICOS, CCR4, and CD127 expression on CD4 dim cells mirrored expression patterns in CD4 high cells, but expression of CD127 was lower overall in CD4 dim cells. Overall, lymph node CD4 T cells in acute infection had higher CCR4, CD127, and ICOS expression and lower CXCR3 expression than chronic infection. Chronically infected lymph node samples had lower CCR4, CCR6, and CD127 than uninfected samples.

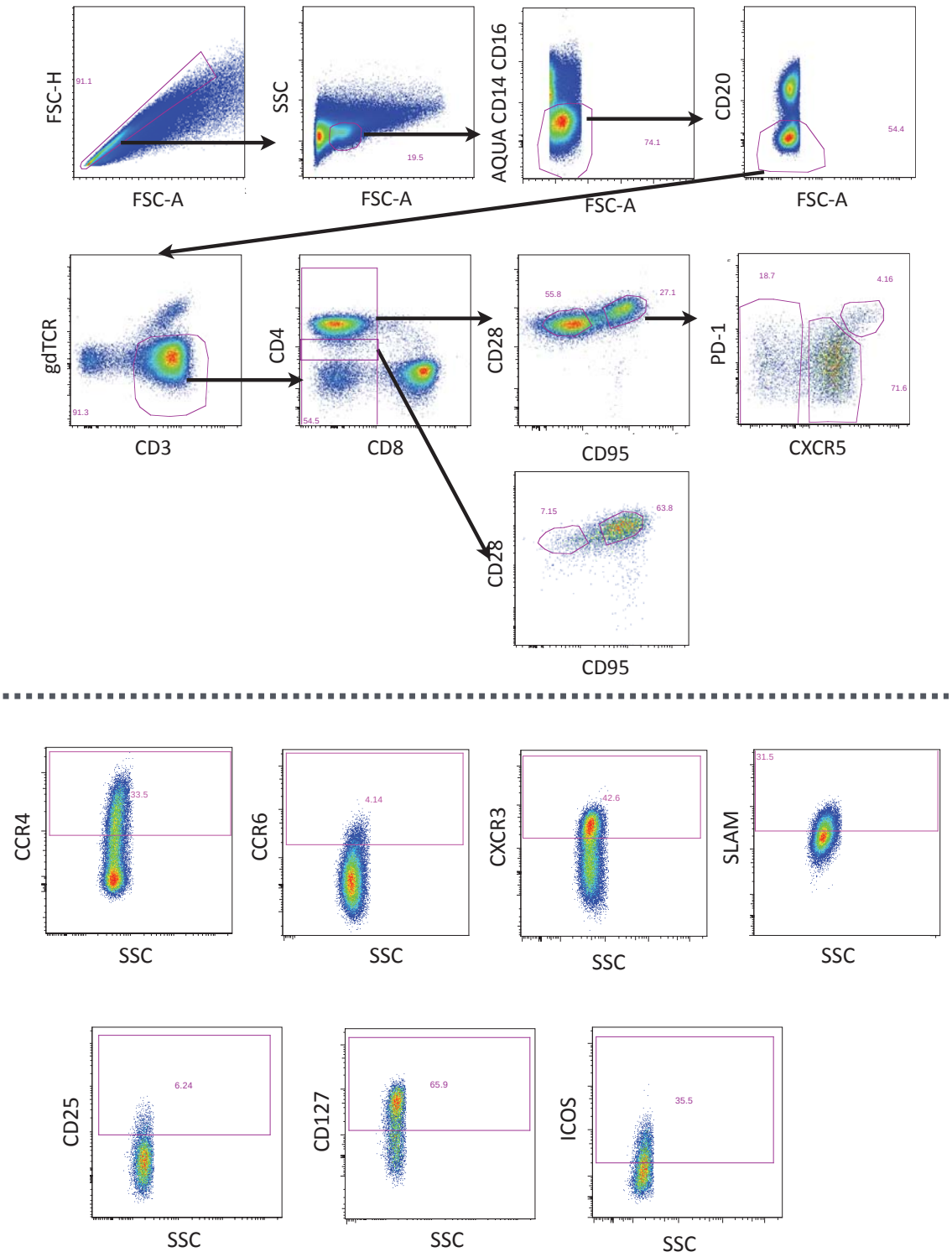


Fig 4.3 Representative gating schematic for phenotyping of SIV-infected and uninfected lymph node biopsies. Lymph node single cell suspensions were stained and acquired on a flow cytometer, gating on forward/side scatter, lymphocytes, live CD14-CD16-CD20-gdTCR- CD3+ T cells, and subsequently on CD8- CD4 bright and CD8- CD4 dim expression. Expression of CCR4, CCR6, CXCR3, SLAM, CD25, CD127, and ICOS was measured on CD28+CD95+ memory T cells.

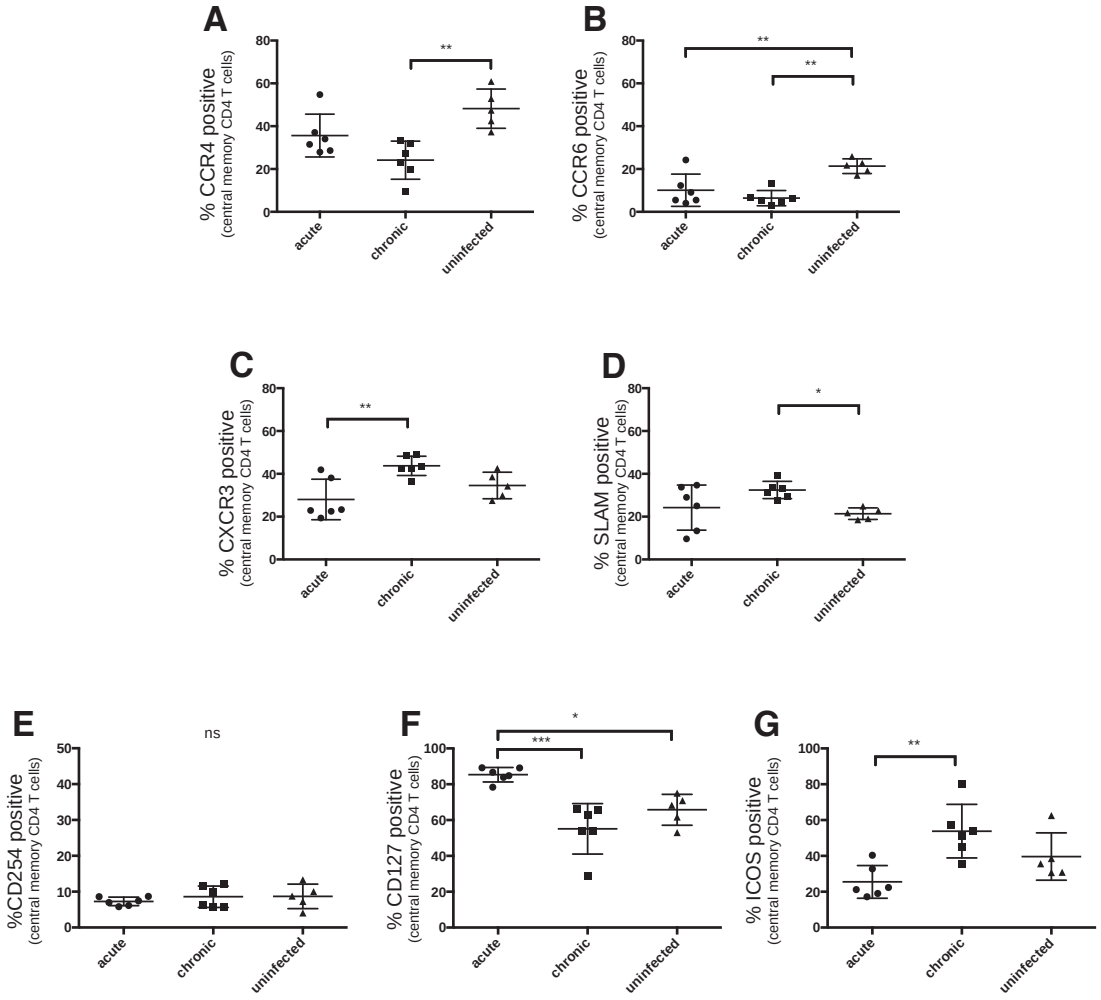


Fig 4.4 Expression patterns of cell surface receptors on lymph node memory CD4 high T cells in SIV infection. Percentage of central memory CD4 high T cells expressing (A) CCR4, (B) CCR6, (C) CXCR3, (D) SLAM, (E) CD25, (F) CD127, and (G) ICOS in acutely infected, chronically infected, and uninfected lymph nodes. Significant differences were determined using a one way ANOVA test.

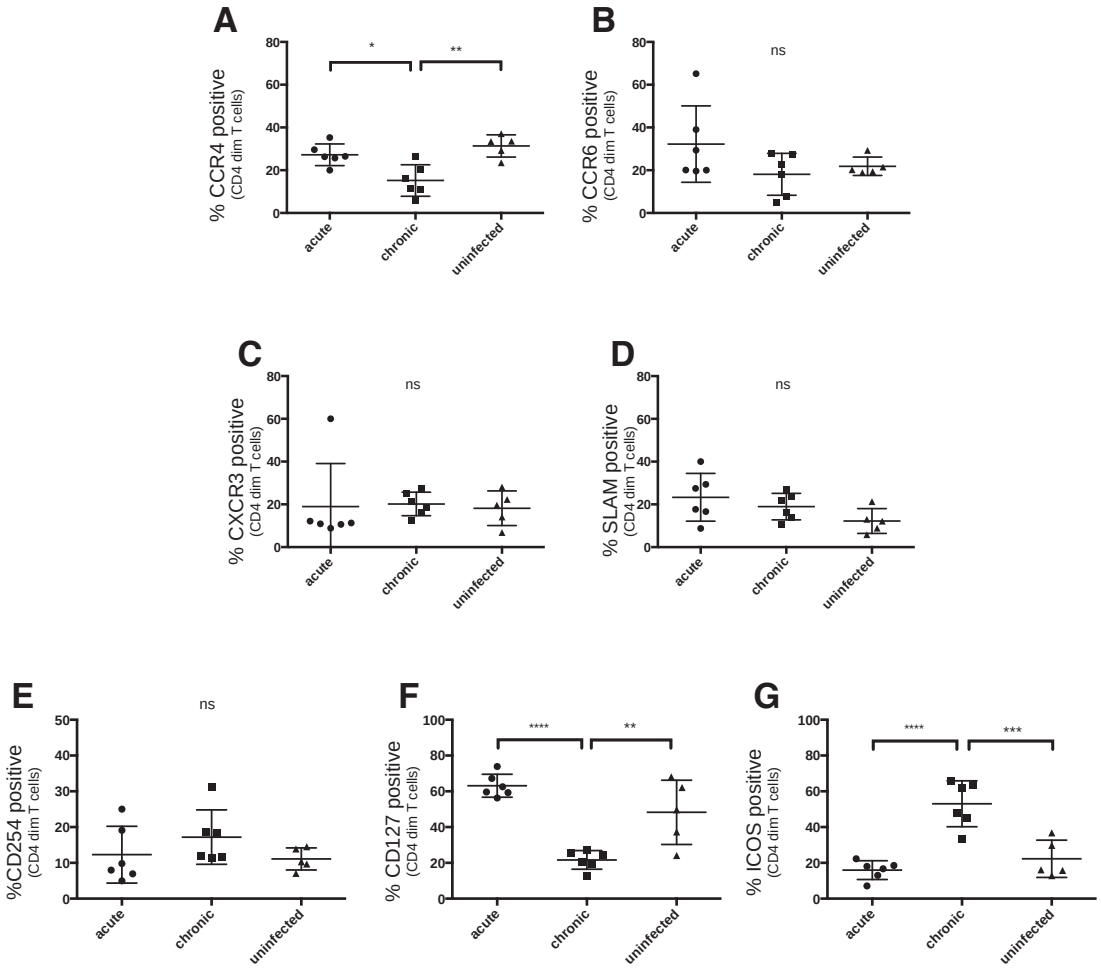


Fig 4.5 Expression patterns of cell surface receptors on lymph node CD4 dim T cells in SIV infection. Percentage of central memory CD4 dim T cells expressing (A) CCR4, (B) CCR6, (C) CXCR3, (D) SLAM, (E) CD25, (F) CD127, and (G) ICOS in acutely infected, chronically infected, and uninfected lymph nodes. Significant differences were determined using a one way ANOVA test.

Central memory T cells are not homogenous but rather represent a diverse population of cells with different activation and localization histories. As CD4 T cells migrate from the T cell zone into the B cell follicle and into germinal centers, they down-regulate CCR7 and up-regulate CXCR5. We divided CD95+CD28+ central memory CD4+ T cells into three populations based on expression of PD-1 and CXCR5 to distinguish between cells located in the T cell zone (CXCR5-), B cell follicle (CXCR5+PD1+) and within germinal centers (CXCR5++PD1++), and measured expression of several chemokine and activation markers within each population. The term TFH is used to refer to CXCR5+PD1+ and CXCR5++PD1++ cells and is inclusive of TFR, which are not distinguished from TFH in this panel

In CXCR5- cells, expression of CCR4, CCR6, CXCR3, SLAM, ICOS, and CD127 followed the same patterns as in overall central memory cells (Fig 4.6). CCR4 expression was higher in

acute CXCR5- samples than in overall central memory (49% vs. 35%, Fig 4.6A) and SLAM expression was higher in chronic CXCR5- samples than in overall central memory (55% vs. 32%, Fig 4.6D). However, CD25 expression was significantly lower on acute samples (mean of 6.7%) than on chronic samples (13.8%) or uninfected samples (13.1%) (Fig 4.6 E). In CXCR5+ cells, patterns of CCR6, CXCR3, CD127, and ICOS expression were unchanged from central memory cells (Fig. 4.7), but CCR4 expression was lower in acute, chronic, and uninfected than in the corresponding samples in central memory. CCR4 expression in TFH was lower than central memory or CXCR5+ cells, with a mean of 7.5% in acute samples and 4.2% in chronic samples. CCR6 expression was lowest in CXCR5+ cells in acute and chronic infection, and SLAM expression was highest in CXCR5+ chronically infected cells. However, there was no difference in SLAM staining on TFH between acute, chronic, and uninfected samples. ICOS expression was highest on TFH in all samples, but in chronic infection most TFH (92%) expressed ICOS compared to 47% of CXCR5+ and 46% of CXCR5- cells. TFH also had lower expression of CD127 in all samples, but acute samples still maintained higher expression of CD127 than in chronic or uninfected samples. Thus, TFH exhibit distinct patterns of expression of chemokine and activation markers in acute and chronic infection from central memory T cells and from CXCR5+ and CXCR5- memory subsets.

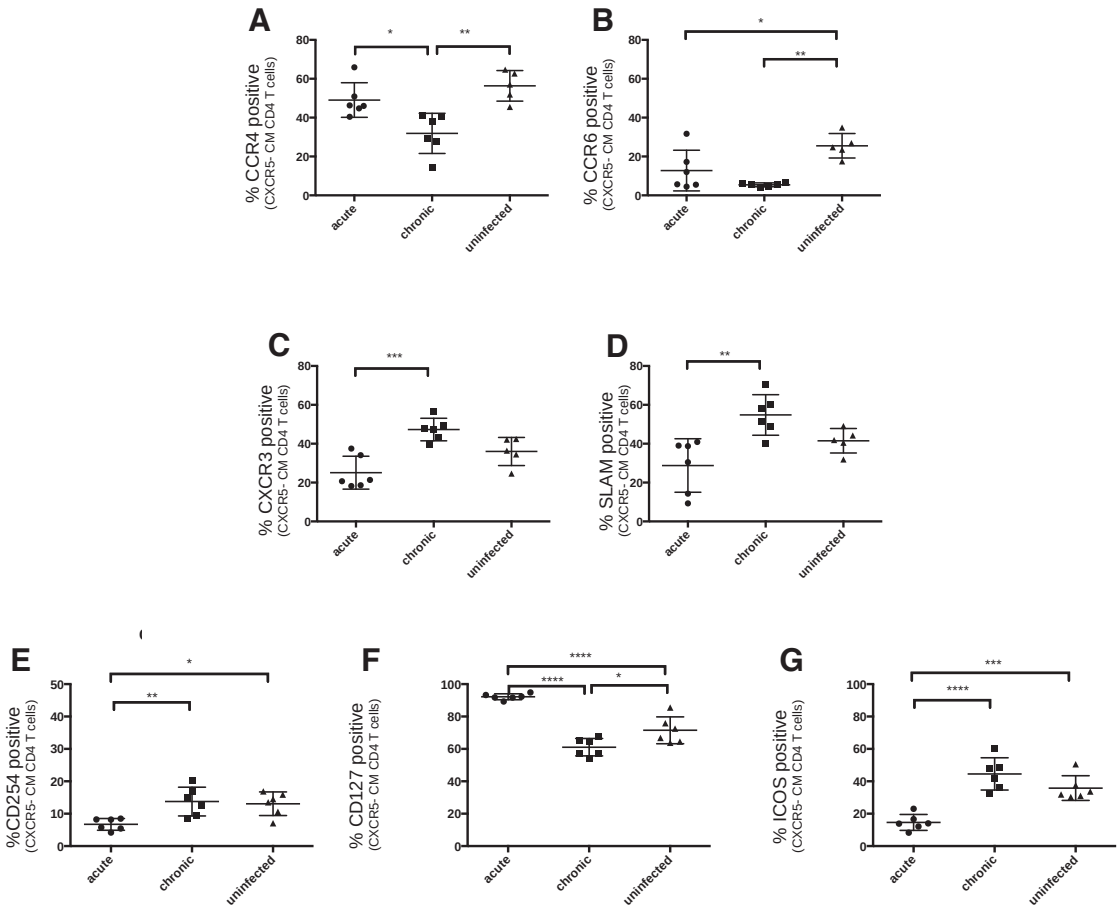


Fig 4.6 Expression patterns of cell surface receptors on lymph node CXCR5- memory CD4 T cells in SIV infection. Percentage of CXCR5- central memory CD4 T cells expressing (A) CCR4, (B) CCR6, (C) CXCR3, (D) SLAM, (E) CD25, (F) CD127, and (G) ICOS in acutely infected, chronically infected, and uninfected lymph nodes. Significant differences were determined using a one way ANOVA test.

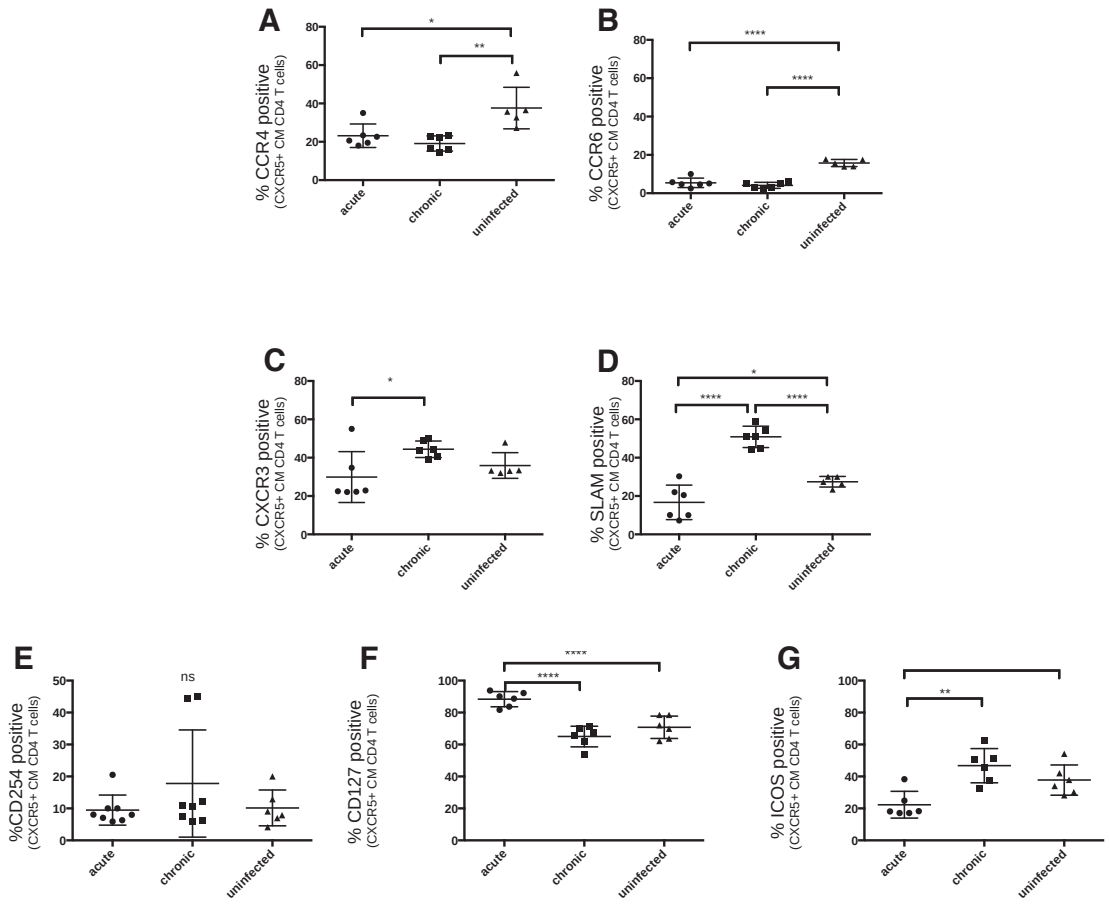


Fig 4.7 Expression patterns of cell surface receptors on lymph node CXCR5+ memory CD4 T cells in SIV infection. Percentage of CXCR5+ central memory CD4 T cells expressing (A) CCR4, (B) CCR6, (C) CXCR3, (D) SLAM, (E) CD25, (F) CD127, and (G) ICOS in acutely infected, chronically infected, and uninfected lymph nodes. Significant differences were determined using a one way ANOVA test.

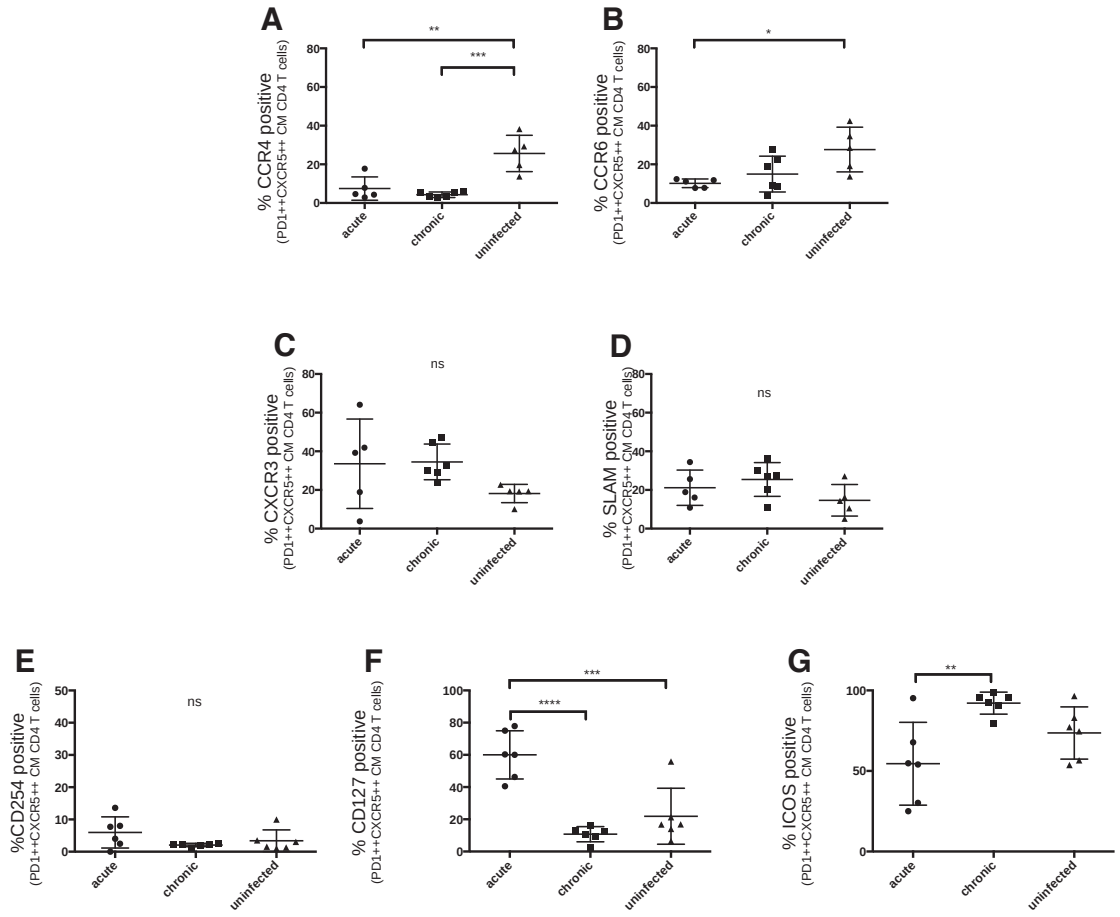


Fig 4.8 Expression patterns of cell surface receptors on lymph node TFH CD4 T cells in SIV infection. Percentage of T follicular helper (CXCR5++PD1++) CD4 T cells expressing (A) CCR4, (B) CCR6, (C) CXCR3, (D) SLAM, (E) CD25, (F) CD127, and (G) ICOS in acutely infected, chronically infected, and uninfected lymph nodes. Significant differences were determined using a one way ANOVA test.

4.2.2 Development of spliced RNA primers and probes

4.2.2.1 Design of primers probes & testing

HIV/SIV infection at the cellular level can be measured in several ways, depending on the type of sample and the research questions. In viral outgrowth assays, resting CD4 T cells from infected donors are activated and cultured alongside uninfected cells, and virus in cell supernatants is measured (typically via ELISA) after days to weeks of cell culture. Donor cells can be cultured in limiting dilution to estimate the frequency of infected cells. HIV/SIV constructs containing green fluorescent protein (GFP) can be used in cell culture infections to measure productive infection via flow cytometry. Quantitative PCR and reverse transcription PCR detect the presence of nucleic acids, and in the case of a retrovirus, can indicate whether transcription has occurred based on the presence of viral RNAs. We used PCR and reverse transcription PCR to quantify the presence of viral nucleic acids in T cell subsets, to understand whether infection was productive (i.e. not only had virus integrated into the cellar genome but that the virus was able to replicate and differentiate between early spliced RNAs and late unspliced RNA). Well-defined primers and probes to detect SIVmac251 Gag DNA

and RNA have been previously published by Lifson et al[326], and we generated absolute standards for that primer/probe set alongside the spliced RNA standards. To measure spliced RNA, we designed primers and probes that spanned the D4/A7 splice site in the SIVmac251 genome (Fig 4.9). To ensure maximum sensitivity to sequence heterogeneity within samples, we used a consensus alignment of 15 SIV full genome sequences including mac251, mac239 and smE660 to design 21-30 nucleotide primers and 21-33 nucleotide probes, and included degenerate bases. The probes span the splice site such that full-length RNAs and singly spliced RNAs (env, vif, tat) do not contain the target sequence. We verified the specificity of the primers by doing side-by-side qPCR with SIV DNA (containing only full length genomes) and qRT-PCR with SIV RNA (containing a mix of full length and spliced genomes), both extracted from in vitro infections of H9 cells with SIVmac251 (Fig 4.10). We verified infection in both cell lines and in primary cells using p24 staining, with 8% of H9 cells infected at 5 days post infection and saw no staining in the uninfected controls. The culture was split and DNA and RNA were extracted for qPCR, qRT-PCR, and synthesis of standards. After testing several combinations of primers and probes, we selected three forward primers, three reverse primers, and 3 probes that had the maximum sensitivity The tat/rev primers and probes did not produce any signal in the DNA qPCR, but detected spliced RNA in the qRT-PCR (Figure 4.11).

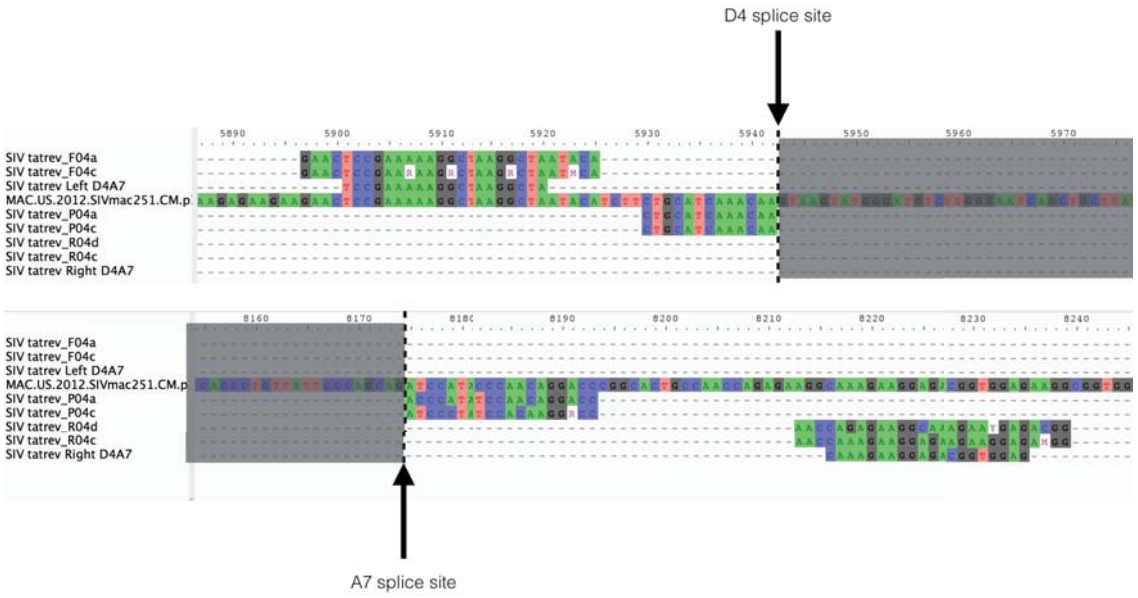


Fig 4.9 Primers and probes for spliced tat/rev RNA detection. Six primers (three forward, three reverse) and two probes were designed to uniquely amplify spliced SIV RNAs. Probes spanned the D4/A7 splice site, and all sequences were designed using a consensus alignment of SIV genomes.

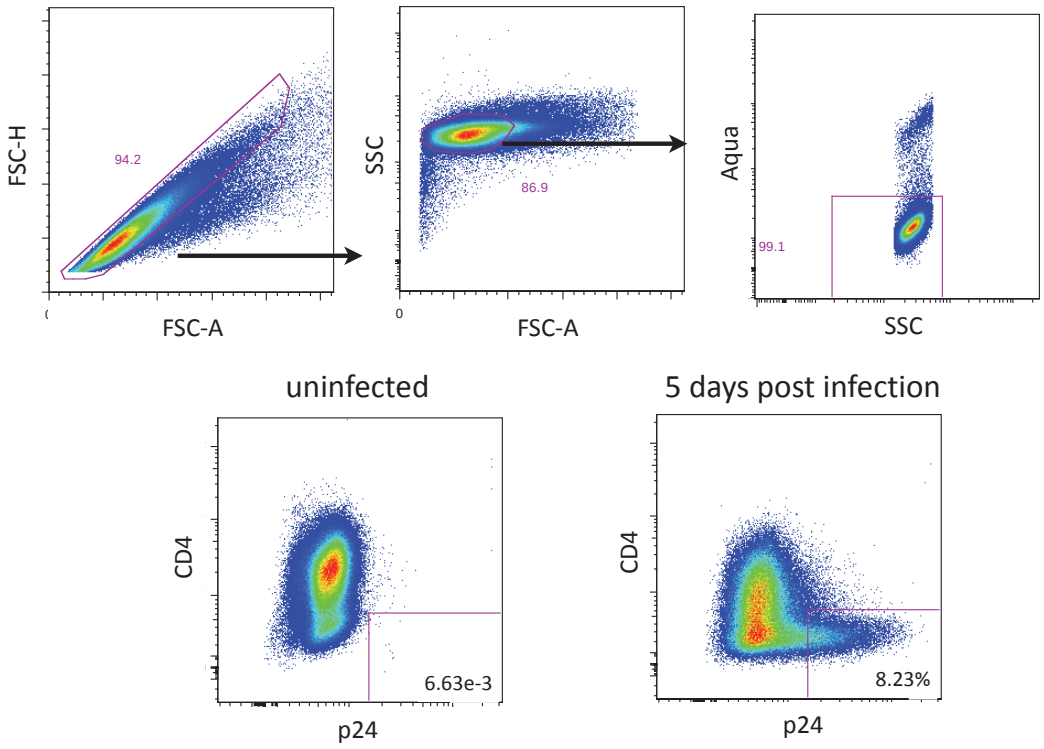


Fig. 4.10 *in vitro* infection of H9 cells. Cell line H9 was cultured and infected with SIVmac251 for 5 days to generate stocks of viral RNA and DNA for creation of SIV standards. Productive infection was verified using p24 staining on CD4 down-regulated cells.

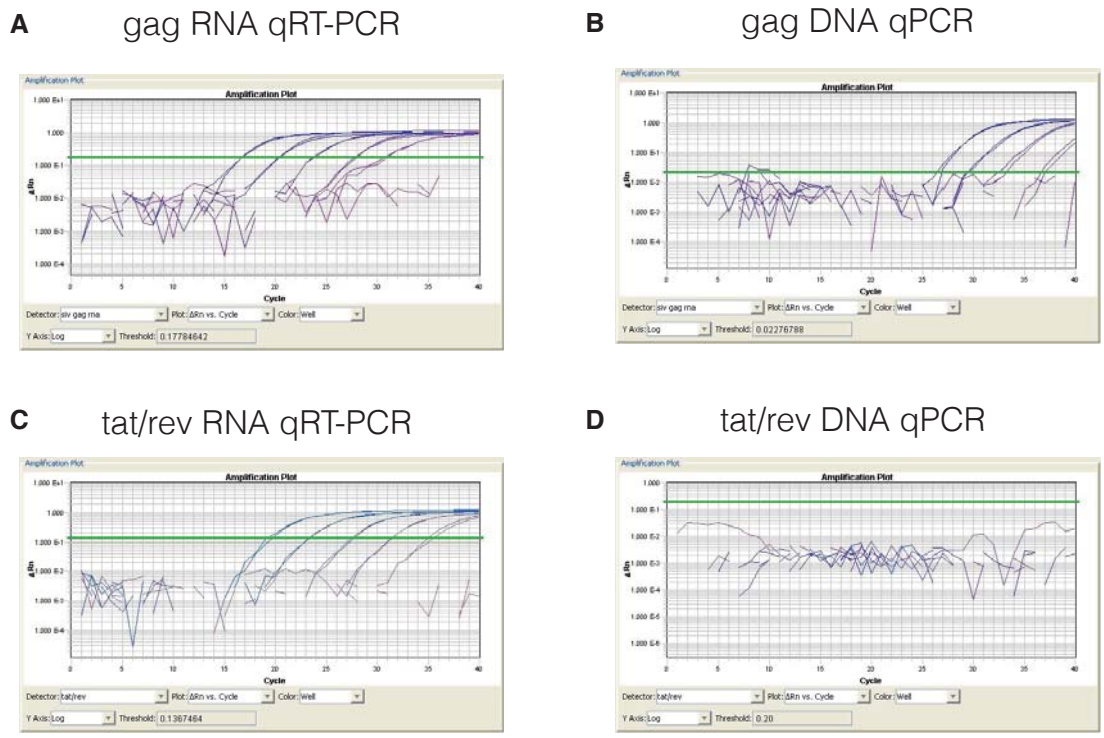


Fig 4.11 Tat/rev primers and probes uniquely amplify RNA. Serial dilutions of RNA (A, C) and DNA (B, D) from SIV-infected cells were used with gag (A, B) and tat/rev (C, D) primers and probes to verify linear amplification of the target RNA or DNA.

4.2.2.2 Synthesis of standards for absolute quantification of spliced and unspliced RNA

To quantify spliced SIV RNA, we used the tat/rev primers to amplify a 106 nucleotide region spanning the D4/A7 splice site and used the pGEM-T vector to transform DH5-a cells to produce plasmids containing the target sequence. Twenty recombinant colonies were picked and the presence of the DNA fragment was confirmed via qPCR. Samples were sent for Sanger sequencing to confirm that the correct, un-mutated sequence was present in the plasmids before purification for RNA synthesis using MEGAscript T7 Transcription enzymes. After RNA synthesis and purification, samples were re-checked for detection of SIV RNA using qPCR and qRT-PCR, and RNA was quantified (Fig 4.12) and diluted to yield standards ranging from 100 copies to $1x10^7$ copies. We performed the same protocol using published SIV gag primers to generate gag RNA standards.

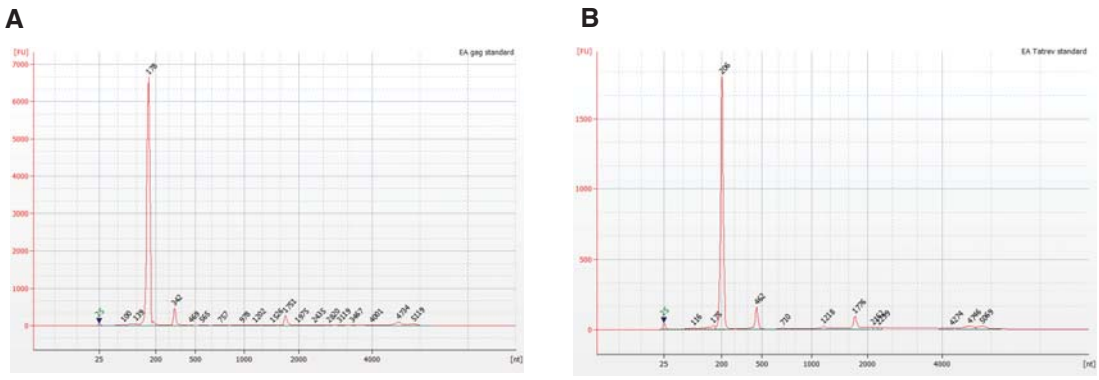


Fig 4.12 High sensitivity quantification of RNA standards for gag and tat/rev primers and probes. (A) SIV gag and (B) SIV tat/rev RNA standards were verified for size and quantity for serial dilution of standards for qRT-PCR. Bioanalyzer traces showed purity of product (>99%) of the appropriate length.

4.2.3 CD4 down-regulation as a marker of active infection

4.2.3.1 Sorting and quantification of SIV in CD4 dim cells

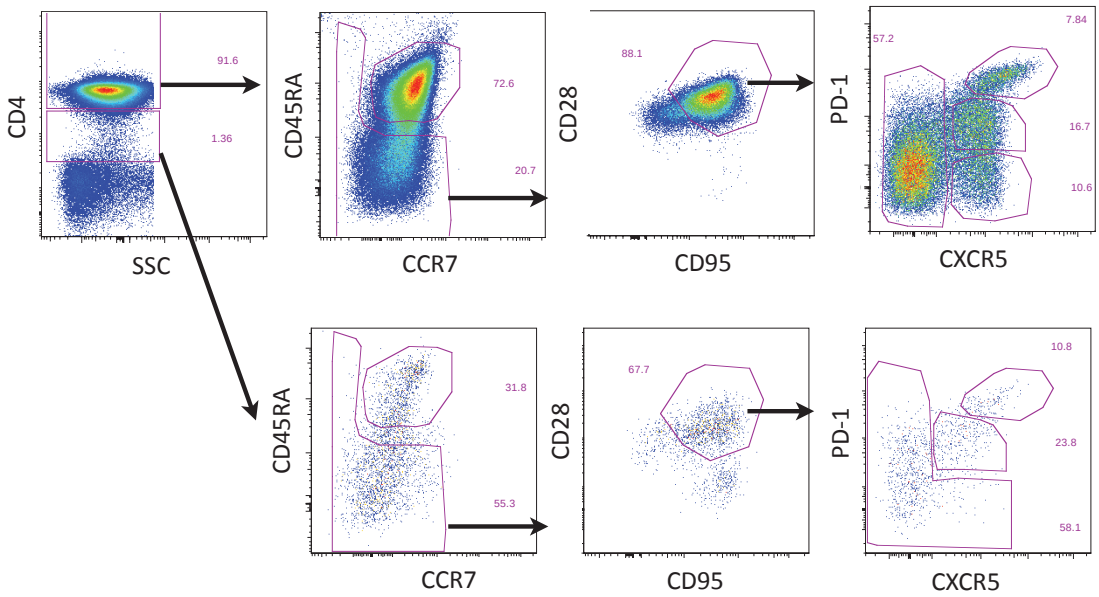


Fig 4.13 Representative gating schematic for sorting CD4 T cells subsets from acutely SIV-infected lymph node biopsies. Lymph node single cell suspensions were stained and sorted after gating on forward/side scatter, lymphocytes, live CD14-CD16-CD20-gdTCR-CD8- CD3+ T cells (as in Fig. 4.3) and subsequently on CD4 bright and CD4 dim expression. Four populations, based on expression of CXCR5 and PD-1, were sorted from memory (CD45RA-CD28-CD95+) CD4 high cells, while three were sorted from CD4 dim cells. Naïve cells (CD45RA+CCR7+CD28+CD95-) cells were sorted from CD4 high cells.

SIV protein nef is produced early in viral replication and down-regulates several surface markers including CD3, CD4, and MHC I. To confirm that CD4 dim CD3+ cells contain a high proportion of actively infected cells, we sorted naïve cells and memory T cells from CD4 high and dim populations, and quantified both spliced and unspliced SIV RNA from acutely infected (5-28 days post infection) lymph node biopsies. The gating scheme is similar to previous samples but with the addition of CD45RA and CCR7 to define central memory populations (Fig. 4.13). Due to the low numbers of CD4 dim cells, we sorted only three subsets from central memory to compare TFH (both PD1++CXCR5++ and PD1+CXCR5+) to non-TFH. Copy numbers of Gag RNA per cell were similar between all four CD4 high memory populations (median of 0.065 to 0.16), and were significantly higher in CXCR5++PD1++, CXCR5+PD1+, and CXCR5- cell than in naïve CD4 T cells (Fig 4.14). Copy numbers were slightly higher in PD1++CXCR5++ cells than the other populations (0.16 copies/cell vs. 0.078, 0.091, and 0.065 copies/cell). There were higher copy numbers of spliced tat/rev RNA in PD1++CXCR5++ (1.3 copies/cell) and PD1+CXCR5+ (1.7 copies/cell) cells than in non TFH (0.096 in CXCR5+ and 0.33 in CXCR5-), and virtually no spliced tat/rev RNA detected in naïve cells. Within CD4 dim cells, there was no difference between the copy numbers per cell of spliced or unspliced SIV RNA in CXCR5++PD1++, CXCR5+PD1+, and non TFH (gag RNA 0.25 copies- 0.72 copies, tat rev 1.1 copies-2.6

copies/cell) (Fig. 4.15). However, in comparing the copy numbers between high and dim cells, there were significantly higher copy numbers of both gag RNA and tat/rev RNA in CD4 dim CXCR5++PD1++ cells, and significantly higher copy number of gag RNA in CXCR5+PD1+ cells (Fig 4.16). While there was considerable sample-to-sample variability between the copies of gag and tat/rev RNA per cell, copy numbers were consistently higher in CD4 dim cells than in CD4 high cells. Thus, CD4 dim cells likely have more productively infected, i.e. producing high copy numbers of spliced and unspliced RNA during active infection, cells than any subset of CD4 bright cells.

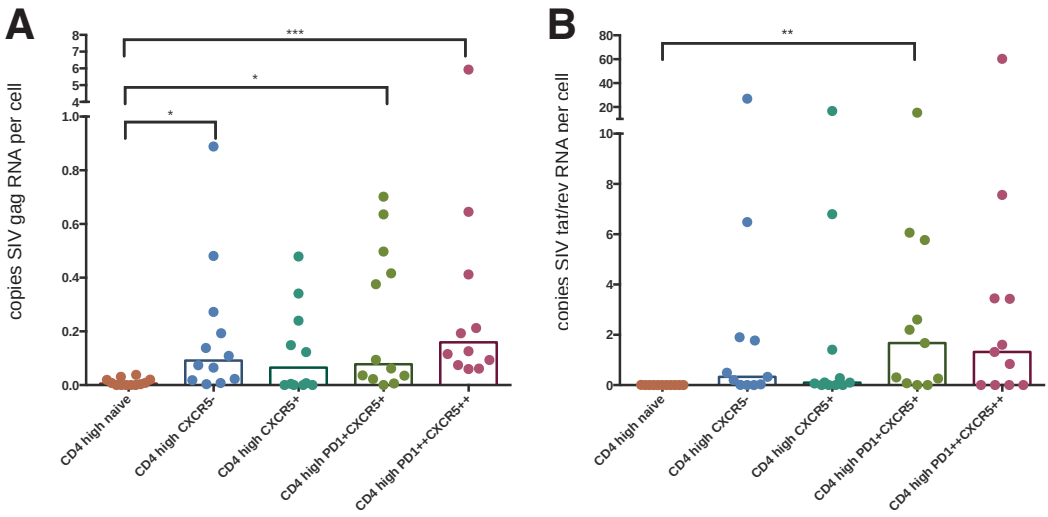


Fig 4.14 SIV RNA copies in CD4 bright T cell subsets. (A) unspliced gag RNA and (B) spliced tat/rev RNA copies per cell in naïve and CXCR5-, CXCR5+, PD1+CXCR5+, and PD1++CXCR5++ CD4 bright memory T cells. Significant differences were determined using the one-way ANOVA test.

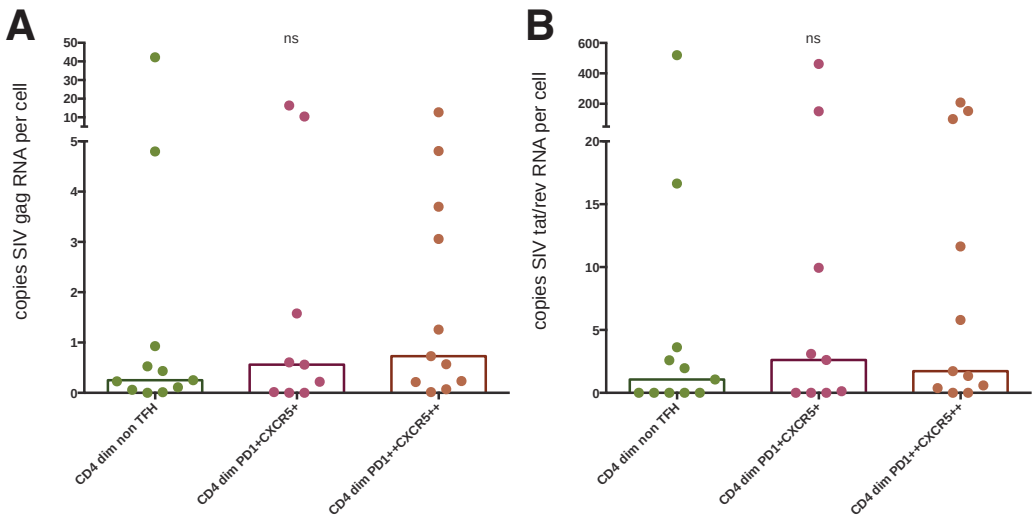


Fig 4.15 SIV RNA copies in CD4 dim T cell subsets. (A) unspliced gag RNA and (B) spliced tat/rev RNA copies per cell in non-TFH central memory, PD1+CXCR5+ central memory, and PD1++CXCR5++ central memory CD4 dim T cell subsets. No significant differences were found between population using one-way ANOVA.

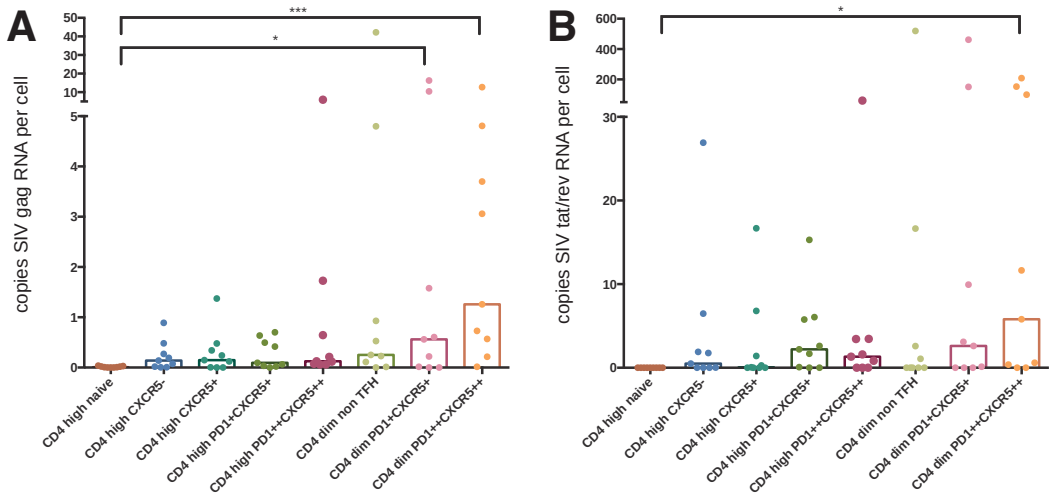


Fig 4.16 SIV RNA copies in CD4 bright and dim subsets. (A) unspliced gag RNA and (B) spliced tat/rev RNA copies per cell in CD4 bright (naïve, central memory CXCR5-, CXCR5+PD1-, CXCR5+PD1+, and CXCR5++PD1++) and CD4 dim (non TFH, CXCR5+PD1+, and CXCR5++PD1++) cells. Significant differences were determined using a one way ANOVA test.

4.2.4 Quantifying SIV in CD4 T cell subsets throughout infection

4.2.4.1 Measuring spliced and unspliced RNA in subsets

To track the proportion of infection in lymph node memory T cells, we used serial lymph node biopsies from SIVmac251 infected macaques and quantified SIV DNA and spliced and unspliced RNA at 13, 28, 76, and 176 days post infection. Gag DNA copies were highest per cell at 13 and 28 days post infection in all T cell populations, and were significantly lower by 76 days and 176 days post infection (Fig 4.17). CXCR5++PD1++ cells had the highest DNA copy numbers per cell at all time points, from a mean of 0.49 copies/cell at 13 days post infection and declining to 0.21 copies/cell at 28 days and 0.073 copies /cell at 176 days. DNA copy numbers were similar in CXCR5+PD1+, CXCR5+PD1-, and CXCR5- populations at 13 and 28 days post infection, ranging from 0.070 to 0.19 copies/cell. In chronic infection, SIV DNA copies per cell were relatively unchanged between CXCR5+PD1+, CXCR5+PD1-, and CXCR5- populations, with 0.042 copies/cell, 0.025 copies/cell, and 0.037 copies/cell respectively. Overall, CXCR5++PD1++ are infected at a higher frequency (as measured by copies of SIV DNA per cell) in early infection, and this difference is consistent throughout infection. While SIV DNA shows a history of infection, it does not indicate whether a cell is productively infected and whether the proviral genome is replication competent. Thus, we measured SIV RNA to better gauge where infection and replication were occurring throughout infection.

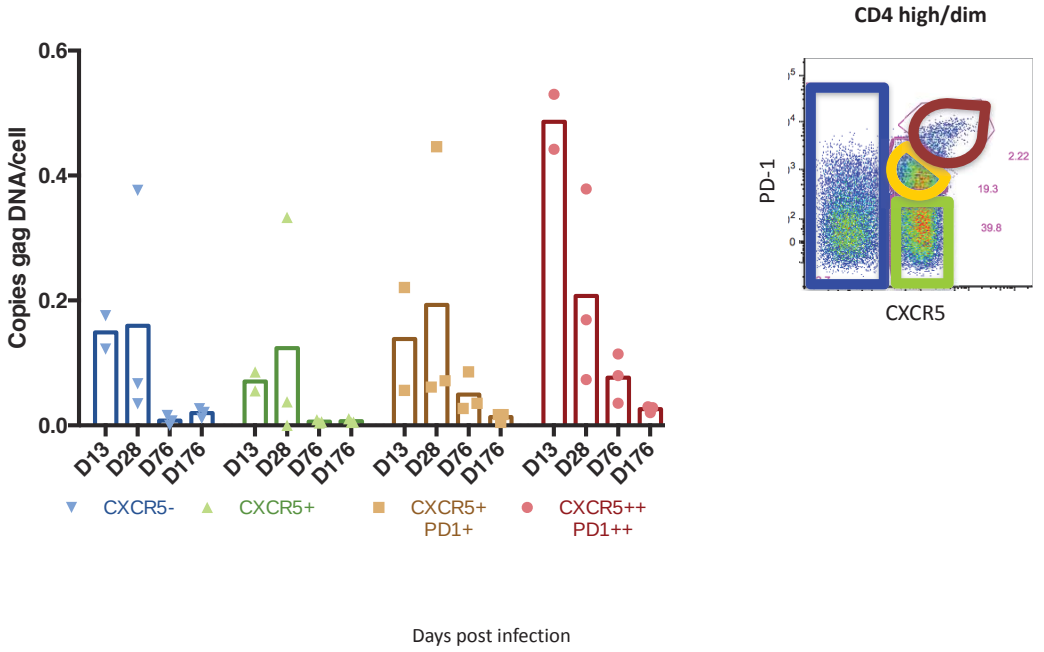


Fig 4.17 SIV gag DNA copies in CD4 bright and T cell subsets throughout infection. Copies of SIV gag DNA per cell at 13, 28, 76, and 176 days post infection in lymph node biopsies of SIV-infected macaques.

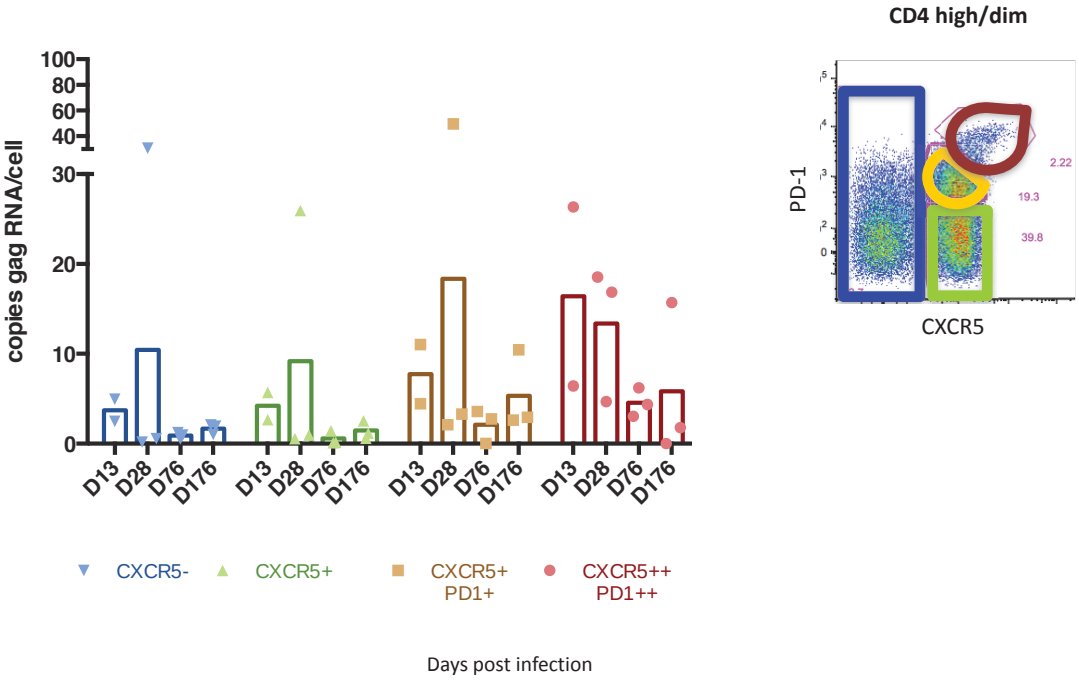


Fig 4.18 SIV unspliced gag RNA copies in CD4 bright and T cell subsets throughout infection. Left, copies of SIV gag RNA per cell at 13, 28, 76, and 176 days post infection in lymph node biopsies of SIV-infected macaques. Right, gating schematic for subsets of central memory CD4 T cells.

Gag RNA copies per cell had similar trends to DNA, with all T cell subsets from early time points (13 and 28 days) having higher copies per cell of unspliced RNA than later time points (means of 18 copies/cell to 3.7 copies/cell in acute samples vs. means of 5.8 copies/cell to 0.56 copies/cell in chronic) (Fig 4.18). CXCR5++PD1++ cells had the highest copy numbers of unspliced RNA per cell at 13 days, 76 days, and 176 days (16 copies/cell, 4.5 copies/cell, and 5.8 copies/cell), while CXCR5+PD1+ cells had the highest copies of gag RNA at 28 days

(18 copies/cell). CXCR5+PD1- cells and CXCR5- cells had lower copy numbers of gag RNA than CXCR5++PD1++ and CXCR5+PD1+ at all time points, but the difference was most dramatic at 76 days and 176 days, when they had on average fourfold fewer copies of SIV RNA per cell. Spliced Tat/rev RNA (Fig 4.19) had different dynamics, with variable copy numbers per cell across populations and over time. CXCR5++PD1++ cells and CXCR5+PD1+ cells had the highest copy numbers per cell of spliced RNA and were consistently above CXCR5- and CXCR5+PD1- cells. However, the net highest copy numbers per cell were measured at 76 days post infection in CXCR5++PD1++ cells. Due to the small numbers of cells sorted from each population (hundreds to thousands) and the low frequency of tat/rev producing cells in chronic infection, very low copy numbers per cell of tat/rev RNA were detected at 176 days post infection, and we were unable to measure tat/rev RNA in CXCR5++PD1++ cells. Overall, CXCR5++PD1++ cells and CXCR5+PD1+ cells contained higher copy numbers of spliced and unspliced RNA throughout infection, and the difference between RNA copy number per cell between these populations and CXCR5+PD1- and CXCR5- cells increased throughout infection.

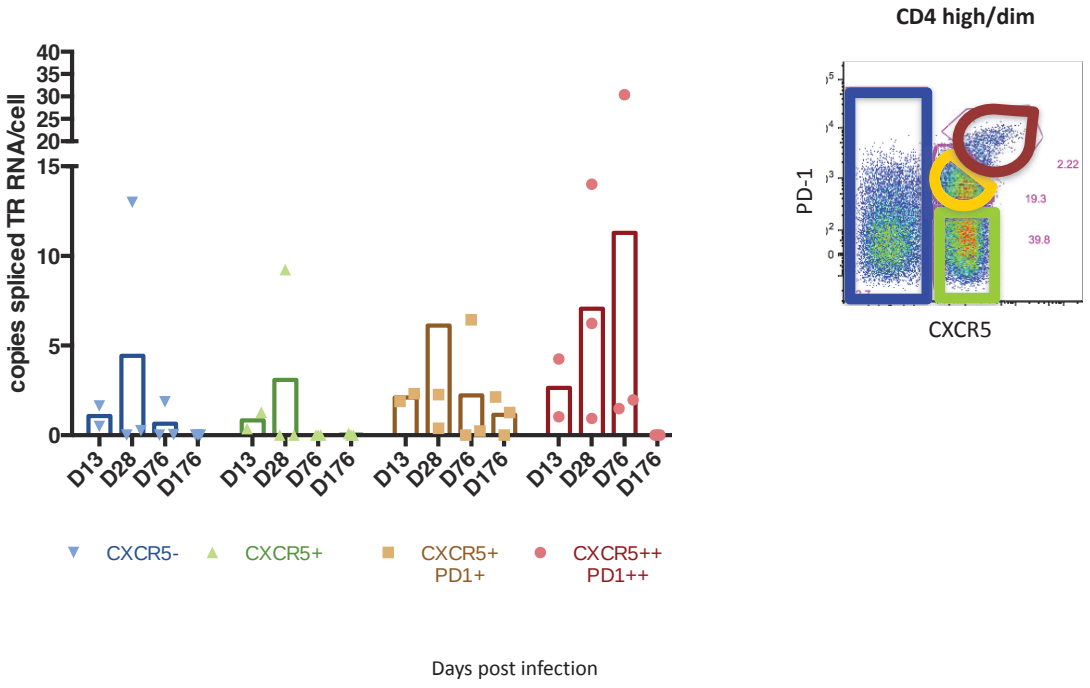


Fig 4.19 SIV spliced tat/rev RNA copies in CD4 bright and T cell subsets throughout infection. Left, copies of SIV tat/rev RNA per cell at 13, 28, 76, and 176 days post infection in lymph node biopsies of SIV-infected macaques. Right, gating schematic for subsets of central memory CD4 T cells.

Because of the low numbers of samples, we grouped acute (day 13/28) and chronic (day 76/176) samples to examine the compartmentalization of infection (Fig 4.20). In this analysis, gag DNA copies per cell was significantly higher in CXCR5++PD1++ cells than in CXCR5++ in both acute (mean of 0.32 copies/cell vs. 0.10 copies/cell) and chronic samples

(0.051 copies/cell vs. 0.0067 copies/cell). There was no statistical difference between gag RNA copies in subsets tin acute infection (although CXCR5++PD1++ and CXCR5+PD1+ cells had the highest copy number with means of 14.5 copies/cell and 14.1 copies/cell compared to 7.2 and 7.8 in CXCR5+PD1- and CXCR5-). In chronic infection, mean copies of gag RNA were significantly higher in CXCR5++PD1++ (3.2 copies/cell) than in CXCR5+PD1- (0.51 copies/cell) and CXCR5- (0.84 copies/cell), with CXCR5+PD1+ in between (1.5 copies/cell). Similarly in tat/rev RNA, there was no significant difference between copies per cell in acute infection (means of 2.2 copies/cell to 5.3 copies/cell), while in chronic infection CXCR5++PD1++ had the highest copy number of RNA per cell (mean of 7.9 copies/cell) and was significantly more than in CXCR5+ (0.18 copies/cell). Overall, productive infection was evenly distributed in acute infection (all subsets showed similar copy numbers of SIV RNA per cell), but in chronic infection, the bulk of infection was in CXCR5++PD1++ and CXCR5+PD1+ cells.

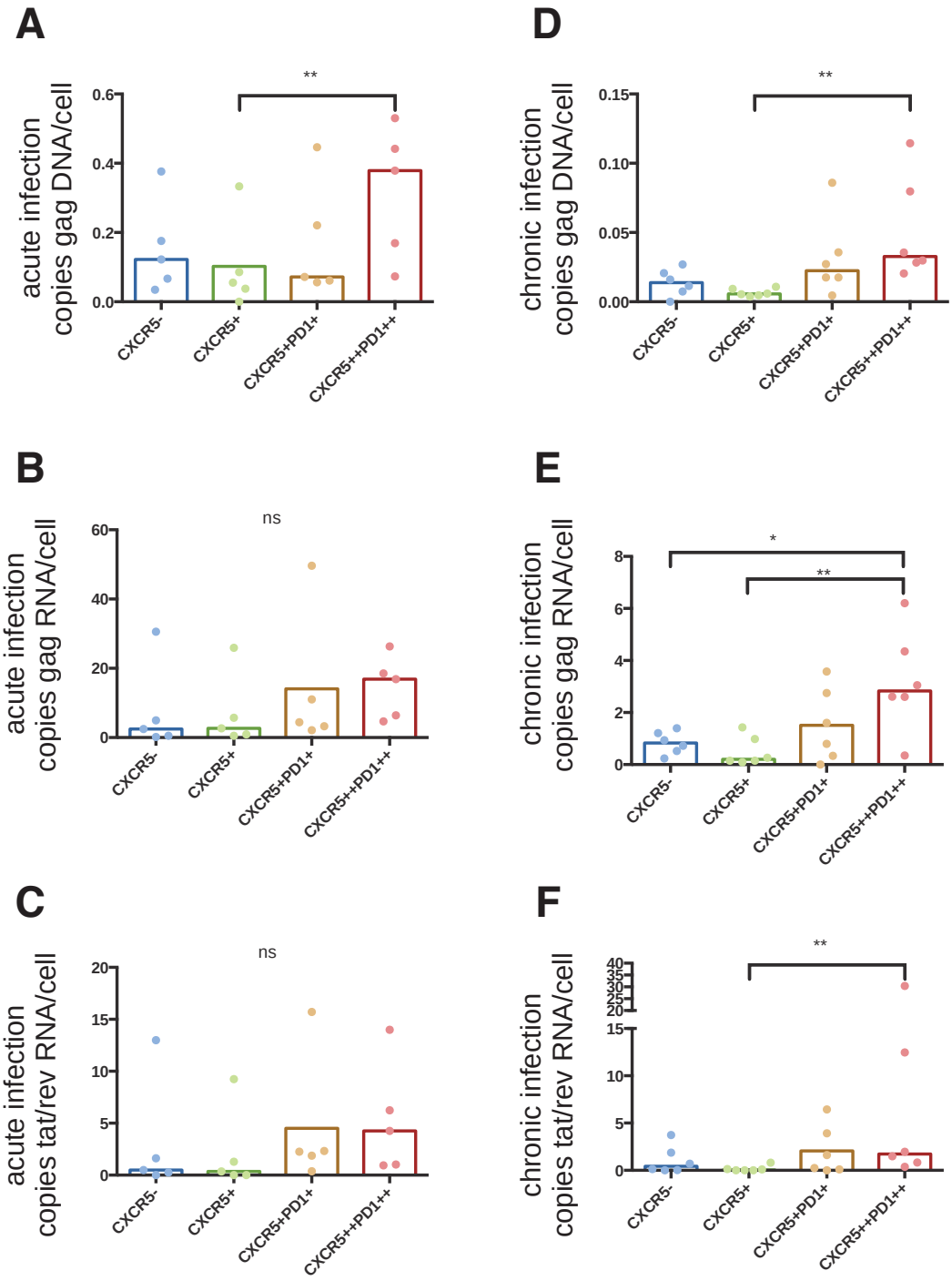


Fig 4.20 Distribution of SIV nucleic acids in acute and chronic infection. SIV RNA copies in CD4 bright and T cell subsets throughout infection. Copies of (A) gag DNA, (B) gag RNA and (C) tat/rev RNA in CXCR5- CXCR5+PD1-, CXCR5+PD1+, and CXCR5++PD1++ central memory CD4 T cells in acute infection. Copies of (D) gag DNA, (E) gag RNA and (F) tat/rev RNA in CXCR5- CXCR5+PD1-, CXCR5+PD1+, and CXCR5++PD1++ central memory CD4 T cells in chronic infection. Significant differences were determined using a one way ANOVA test.

4.2.4.2 Changes in TFH throughout infection

Previous studies have shown that TFH are increased as a fraction of all central memory T cell in lymph node in chronic HIV/SIV infection[170]. We confirmed this in our samples by

measuring the proportion of CXCR5++PD1++, CXCR5+PD1+, CXCR5+PD1-, and CXCR5- cells as a percentage of central memory (CD45RA-CCR7-CD28+CD95+) CD4 T cells (Fig. 4.21). CXCR5++PD1++ cells increased from a mean of 4.8% at 10/13 days post infection to 36% at 176 days post infection. CXCR5+PD1+ cells remained constant over the course of infection (18%-23% of central memory CD4 T cells). Both CXCR5+PD1- cells and CXCR5- cells decreased significantly over the course of infection, dropping from 19.4% to 2.8% and 51% to 32%, respectively. This shift in proportions of memory T cell subsets has an impact on the compartmentalization of virus replication and sites of infected cells.

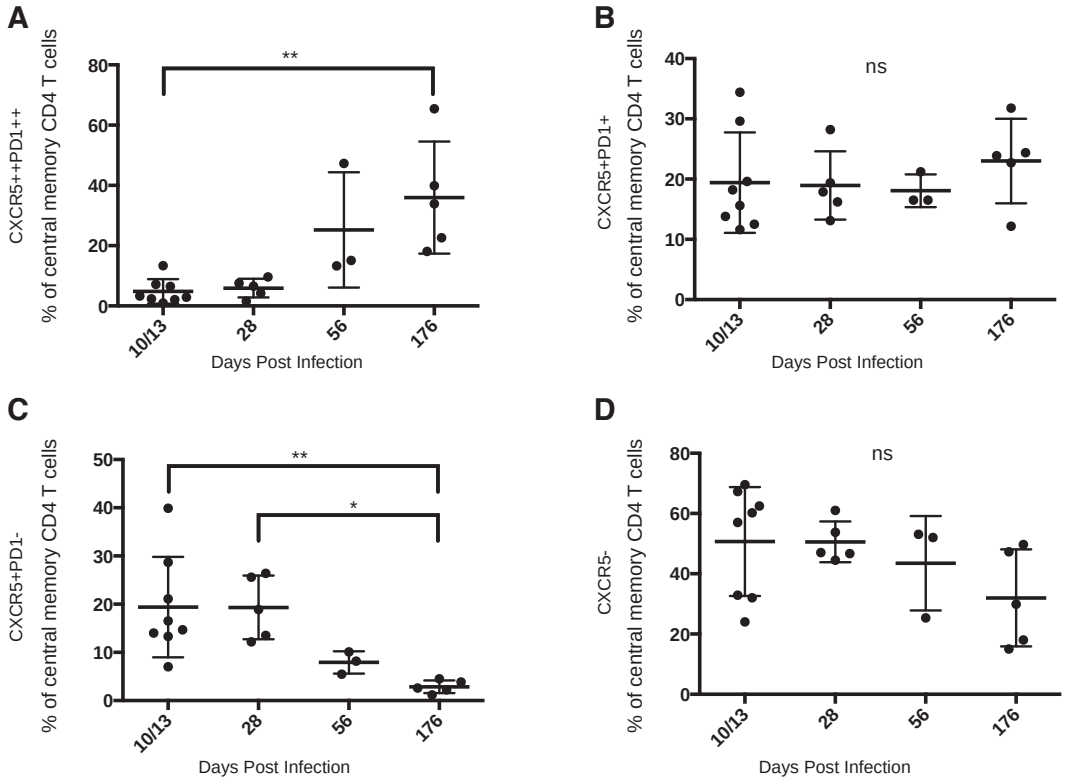


Fig 4.21 Frequencies of T cell subsets throughout infection. (A) CXCR5++PD1++, (B) CXCR5+PD1+, (C) CXCR5+PD1-, and (D) CXCR5- cells as a percentage of central memory CD4 T cells from early to late infection. Significant differences were determined using a one way ANOVA test.

While CXCR5++PD1++ and CXCR5+PD1+ cells had higher copy numbers of gag DNA and gag and tat/rev RNA per cell in all populations, the frequencies of CD4 T cell memory cells shift throughout the course of infection and impact the composition of the pool of infected cells. Using the viral nucleic acid copy numbers per cell and the proportion of each cell subset within the central memory, we can estimate the contribution of each subset to the total pool of SIV DNA and RNA (Fig. 4.22). In these calculations, we see that despite only representing a small percentage of all central memory T cells at peak infection, TFH (here defined as CXCR5++PD1++ and CXCR5+PD1+ cells) contain an outsized amount of gag DNA and gag and tat/rev RNA (40%, 49%, 46% respectively). At 28 days post infection when plasma

viremia had decreased to viral setpoint, these frequencies had diminished slightly (25%, 30%, and 29%), but later in infection, when TFH represent a higher fraction of all central memory cells, contribute most of viral DNA and RNA (73%, 84%, 99% at 176 days). In summary, CXCR5++PD1++ and CXCR5+PD1+ T follicular helper cells represent a significant source of SIV DNA and RNA in early infection, but by later infection are the major compartment of SIV infection and replication.

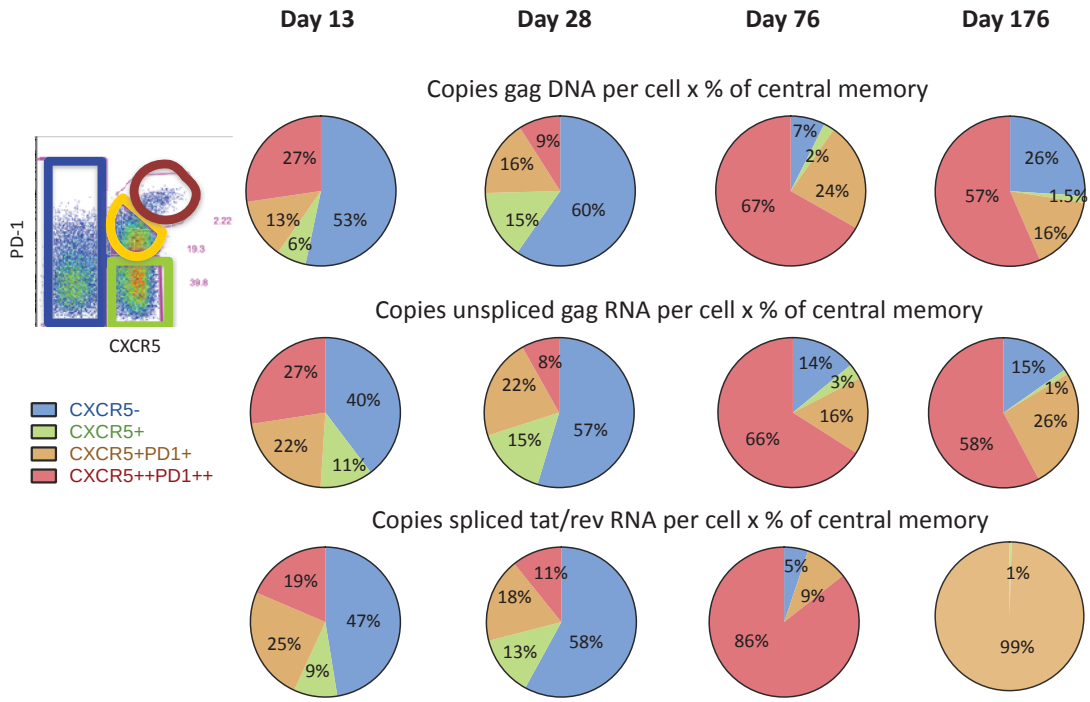


Fig 4.22 Contributions to SIV DNA and RNA from memory T cell subsets. Top, contributions to total gag DNA in central memory T cells by subset at 13, 28, 76, and 176 days post infection. Middle, contributions to total unspliced gag RNA in central memory T cells by subset at 13, 28, 76, and 176 days post infection. Bottom,, contributions to total spliced tat/rev RNA in central memory T cells by subset at 13, 28, 76, and 176 days post infection. Copies per cell of SIV DNA and RNA in four memory subsets were multiplied by the proportion of that subset in the total memory compartment to derive the contribution of each subset to the total pool of SIV.

4.2.5 RNA flow cytometry – detecting and phenotyping individually infected cells

4.2.5.1 Testing RNA probe staining

Quantifying viral nucleic acids yields detailed information of infection of bulk subsets of CD4 T cells, but does not directly yield the proportion of infection of different populations. Furthermore, populations must be stained and fixed in advance of DNA/RNA quantification, preventing the subsequent analysis of the features of infected cells within sorted populations. To gain a more detailed understanding of the cellular markers of infected cells, we used RNA probes in flow cytometry to identify individually infected cells. The use of RNA probes in flow cytometry is based upon techniques of *in situ* hybridization in fixed tissues, and has recently been developed as a commercially available kit [377]. We first tested the PrimeFlow kit (eBiosciences) against probes and reagents developed by ACD for use in RNAScope [378]

staining on formalin-fixed, paraffin-embedded tissue slices to determine which assay was more sensitive. Both assays are based on similar technologies and use branching amplified to RNA target probes with fluorescent detectors. We performed a side-by-side comparison of the two sets of probes, amplifies, and detectors to determine which assay had greater sensitivity (Fig. 4.23). While both ACD and eBiosciences reagents detected RNA-positive CD4 T cells from splenocytes at 13 days post infection, the eBiosciences reagents had better staining on central memory T cells and subsets. In the eBiosciences staining, 2.47% of central memory T cells were RNA probe-positive compared to 1.39% in the ACD protocol. The separation on the eBiosciences samples was better than on the ACD samples, with a separation of nearly two logs in the Biosciences samples and just over one log in the ACD samples. Both samples had equivalent background staining on cell populations not infected with SIV (CD8 T cells and CD3- cells) and on uninfected controls.

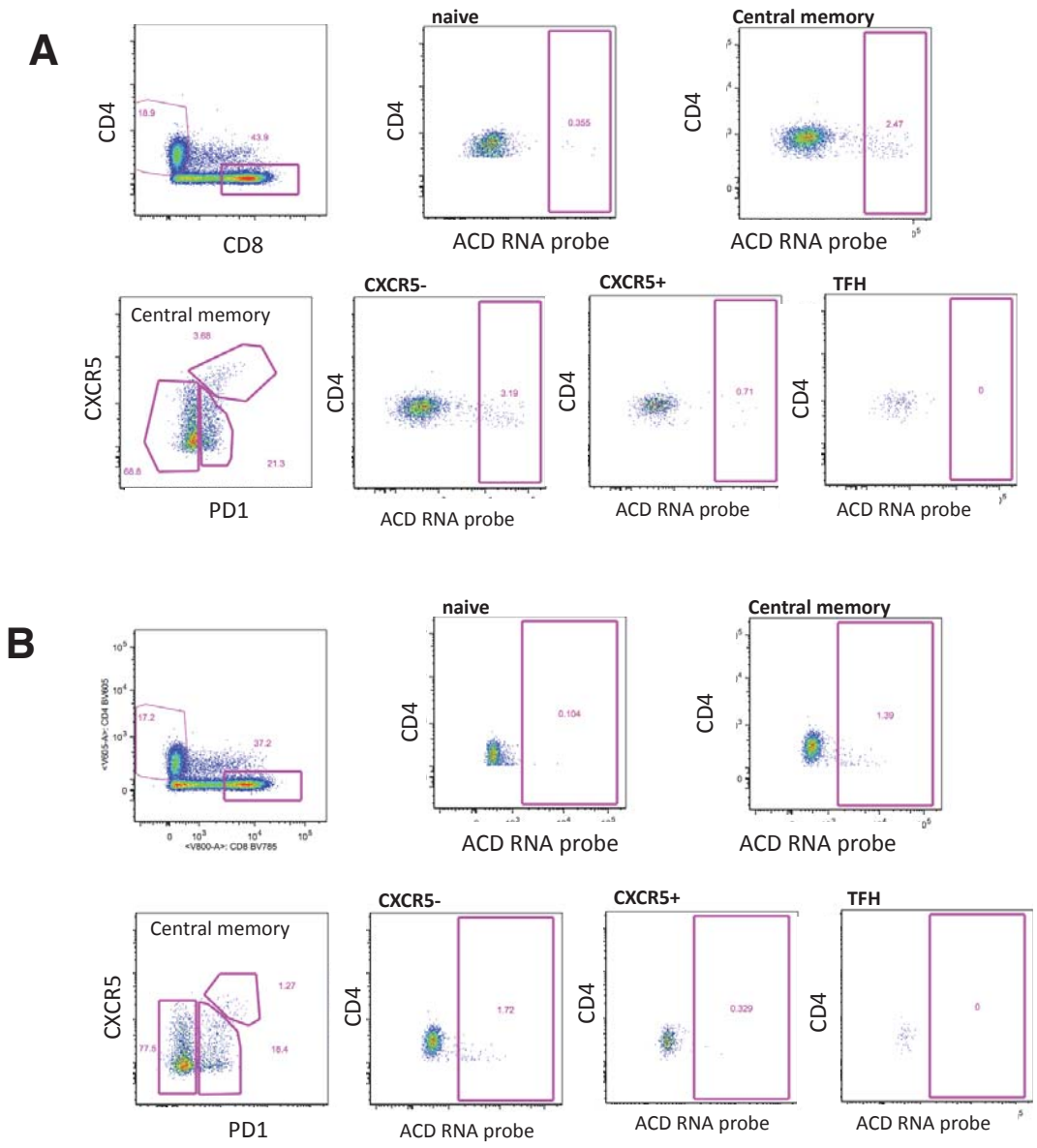


Fig. 4.23 Comparison of RNA flow cytometry staining using eBiosciences and ACD reagents. (A) eBiosciences regent staining and (B) ACD reagent staining of SIV infected splenocytes from 13 days post infection. In both figures, plots show staining of probe on naïve (top middle) and memory (top right) cells (from the CD4+CD8- compartment, top left), and from within the central memory compartment (CD95+CD28+), probe staining on CXCR5-, CXCR5+, and TFH (bottom row).

4.2.5.2 Specificity of RNAflow SIV probe staining

To confirm the specificity of the SIV RNA probes in infected cells, we used probes against beta-2 microglobulin (B2M) as a positive control and bacterial enzyme dapB as a negative control (Fig. 4.24). We observed no dapB staining on CD4 or CD8 cells, whereas 88-91% of CD4 and CD8 T cells were positive for B2M. Staining with SIV probes was specific to CD4 T cells (1.6% of CD4 T cells, isolated from splenocytes of an acutely infected macaque) and while there was background staining on CD8 T cells, it was significantly below staining on

CD4 cells. Nonetheless, we increased the number of wash steps in the protocol to reduce background staining.

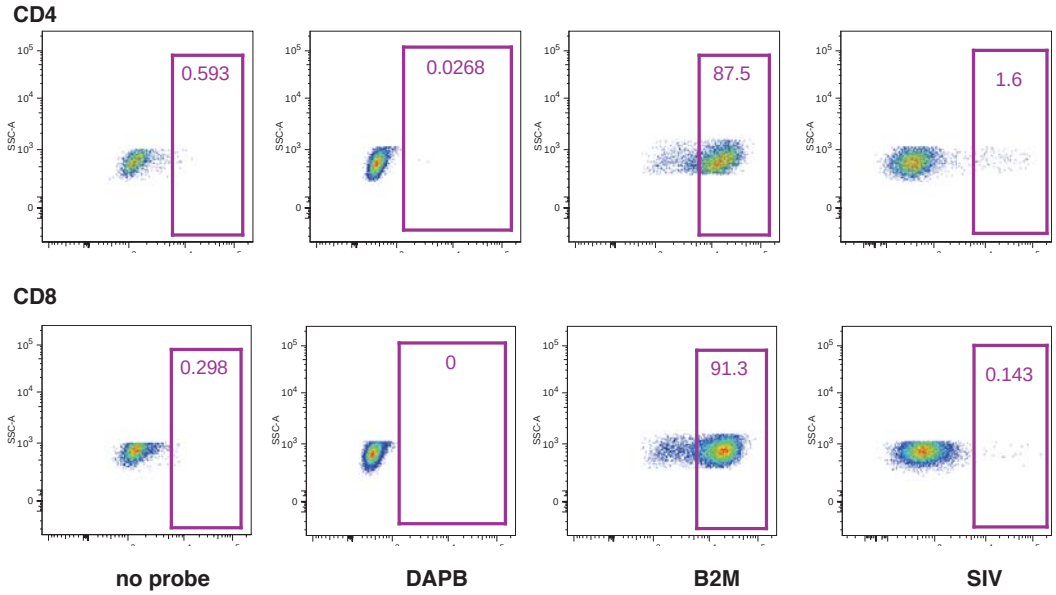


Fig 4.24 Staining of RNA flow probes and controls. Staining of DAPB (negative control), B2M (positive control), and SIV RNA probes alongside no probe control on (top row) CD4 T cells and (bottom row) CD8 T cells.

To further confirm the specificity of the RNA probes, we infected CEMx174 cells with SIVmac251 and measured the proportion of infected cells via intracellular p24 staining and RNA probe staining (Fig. 4.25). Samples at five days post infection showed clear p24 staining on CD4 down-regulated cells with an infection proportion of 8%, whereas staining at six days post infection with RNA probes had an infection frequency of 37.7%. The infection proportion is higher in RNA probe staining than with p24, as p24 is produced only at certain stages in the SIV lifecycle whereas SIV RNAs are present during initial infection and throughout the later stages of cellular infection, and the p24 probe is designed for HIV core antigen and shows only partial cross-reactivity with SIV proteins.

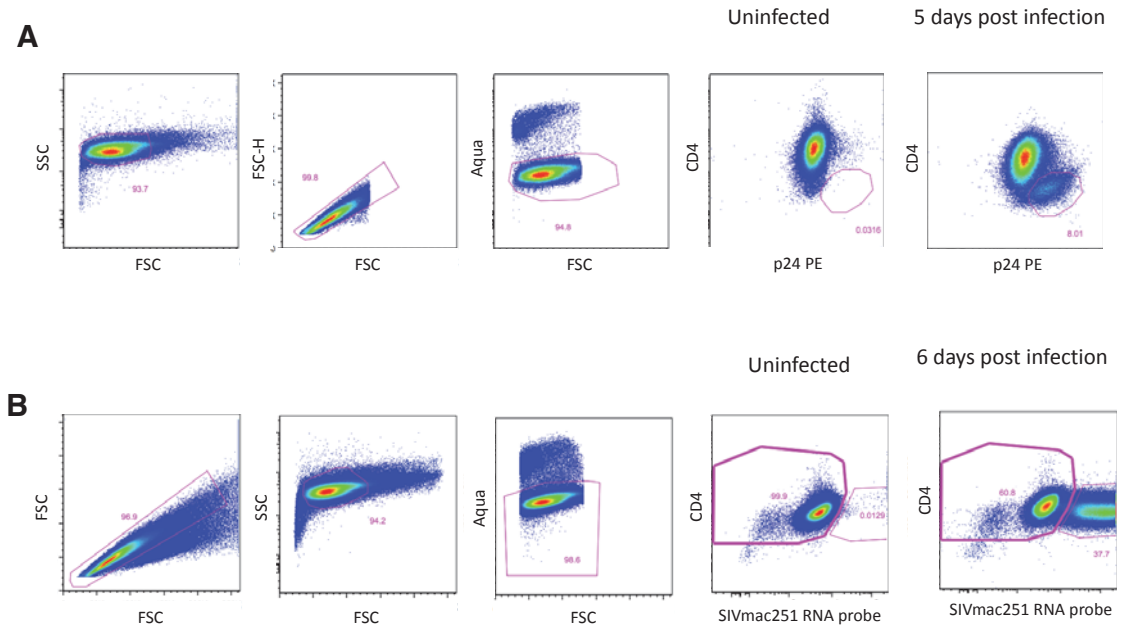


Fig 4.25 Infection and staining of cell lines with RNA flow probes. (A) Gating and staining of CEMx174 cells infected with SIVmac251 at zero and 5 days post infection. Cells were intracellularly stained with p24. (B) Gating and staining of CEMx174 cells infected with SIVmac251 at zero and six days post infection. Cells were stained with RNAflow SIV probes.

We developed a flow cytometry staining panel to evaluate the proportional infection of CD4 T cell subsets that was compatible with the long duration and high temperatures of the PrimeFlow protocol, which involves 18 wash steps, 6 hours of incubation at 40°C for 6 hours, and fourteen hours from the start of the protocol to acquisition of samples (Fig. 4.26). In addition, not all commercially available fluorophores are compatible with the PrimeFlow buffers and reagents, so a minimal panel was designed to accommodate the limitations of the kit while still differentiating CD4 T cell memory subsets and TFH in particular. To confirm the gating and staining after the PrimeFlow protocol, for each sample we stained a subset of cells with the surface panel only and fixed and acquired the samples without undergoing the incubations and washes. While there were some differences in the fluorescence between the two protocols, we could confidently distinguish the cellular populations in the full protocol samples. In all infected samples, we also ran a no-probe control to account for nonspecific binding of the amplifiers and label probes. This fluorescence –minus-one control was used to set the gating for probe-positive samples in all infected samples.

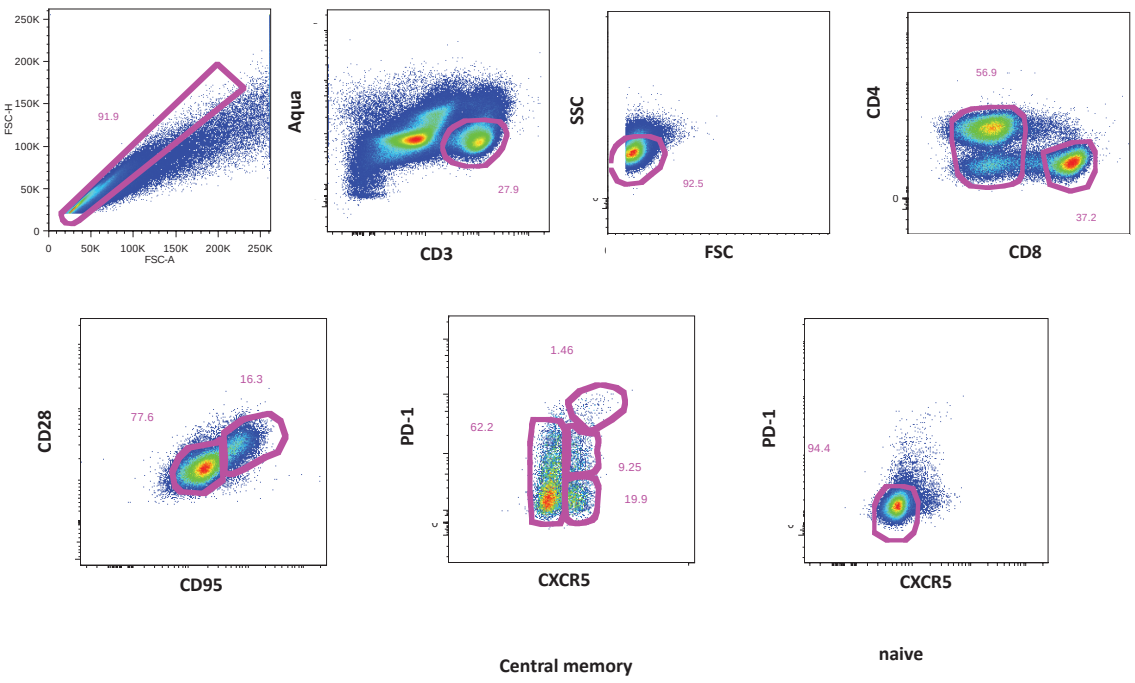


Fig 4.26 Representative gating schematic for RNA probe staining from SIV-infected lymph node biopsies. Lymph node single cell suspensions were stained and acquired on a flow cytometer after gating on forward/side scatter, lymphocytes, live -CD8- CD3+ T cells. Memory cells were defined as CD28+CD95+ and subsequently gated on expression of CXCR5 and PD-1. Gates were set based on staining on naïve and CD8 T cells. For all samples, a no-probe FMO control was stained alongside a surface-only sample to ensure accurate gating of cell populations

Probe staining on acutely infected and chronically infected lymph node samples showed that SIV probe staining is primary on CD4 T cells, with only minimal background on CD8 T cells (Fig. 4.26). When CD4 T cells were gated on CD28 and CD95 to differentiate naïve and central memory populations, the proportion of infection on central memory cells was over tenfold higher than in naïve cells. Acutely infected animals (5-13 days post infection) had significantly higher frequencies of infection than animals that had controlled viremia to viral setpoint (28/76 days post infection) or chronically infected animals (6 months-2 years post infection). Staining on uninfected animals showed only minimal background staining on CD4 and CD8 T cells and on central memory and naïve CD4 T cells, comparable to staining on CD8 T cells in infected animals. Acutely infected animals have a higher proportion of SIV-infected cells, and we wanted to confirm that we could distinguish chronically infected cells from background staining of RNA probes. Background staining in acutely and chronically infected samples varied from sample to sample (Fig. 4.27), but was reliably below the probe-positive samples at 176 days and 2 years post infection. Thus the RNA probes are specific to productive infection in CD4 T cells, and follows expected distributions of infection based on cellular phenotype.

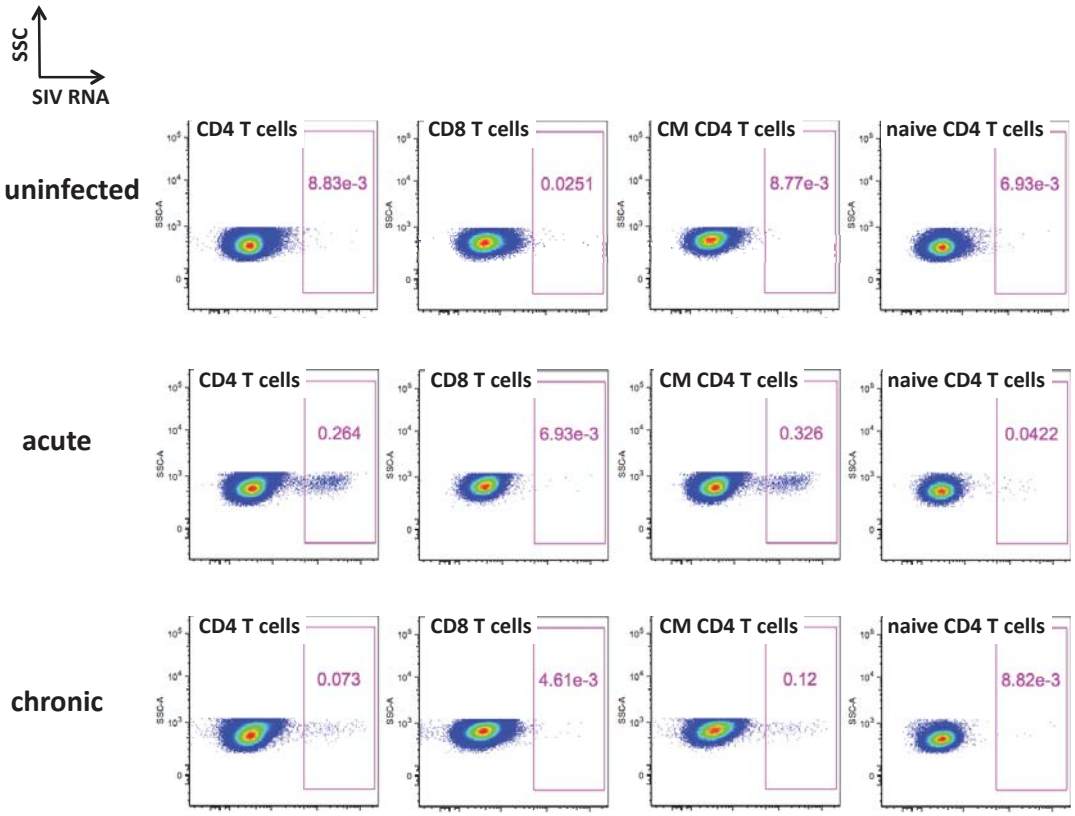


Fig 4.27 Staining of RNA flow probes on CD3 T cell populations. Staining of SIV probes on uninfected (top), acutely infected (middle), and chronically infected (bottom) lymph nodes CD4 T cells (far left column), CD8 T cells (left), central memory CD4 T cells (right), and naïve CD4 T cells (far right).

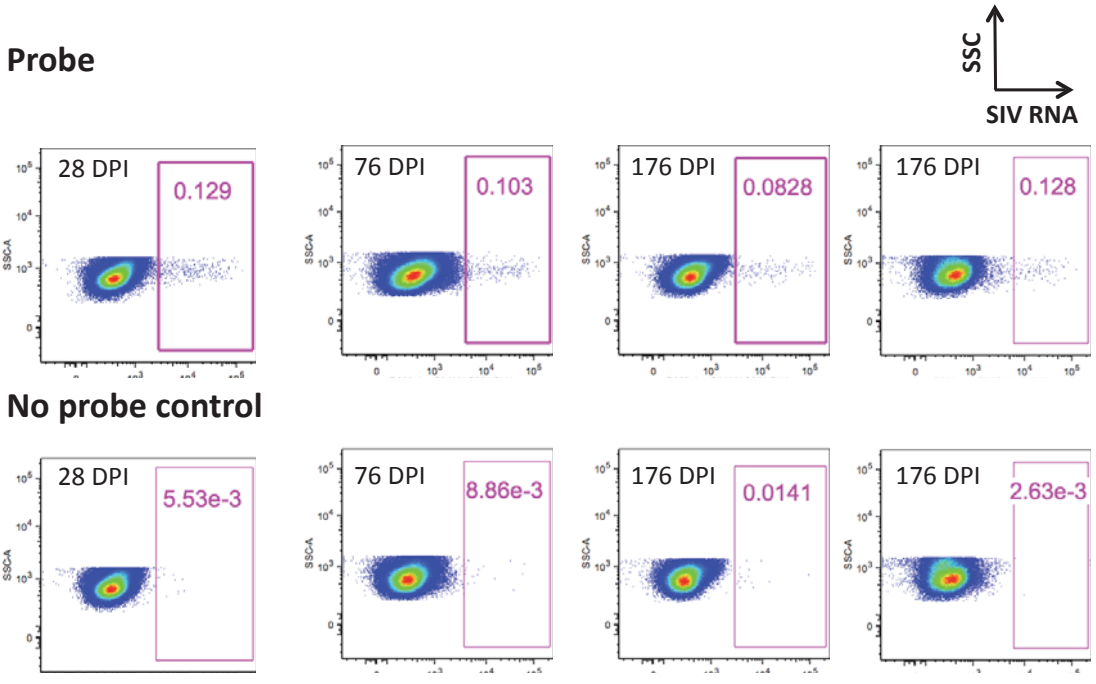


Fig 4.28 RNA probe staining on chronically infected lymph nodes. To test the background staining of the amplifiers used to detect the RNA probes, CD4 T cells from SIV-infected samples from 28 days (far left), 76 days (left), and 176 days (right and far right, two separate specimens) were stained with SIV RNA probes (top row) and compared to FMO controls (bottom row).

4.2.5.3 Staining is specific to memory T cell populations

We used CXCR5 and PD-1 to distinguish TFH in the RNA flow samples, and measured the proportion of RNA probe positive cells within four central memory populations. The highest proportions of infection were at 7-13 days post infection, when plasma virus levels were at or near peak. CXCR5++PD1++ cells had the highest proportion of infection (mean 1.5%), followed by CXCR5+PD1+ and CXCR5- (0.9% and 0.8%) and CXCR5+PD1- cells have the lowest proportion of infection (0.5%). By 1-3 months post infection when plasma virus was at setpoint, the proportion of RNA positive cells had declined significantly in CXCR5-, CXCR5+PD1-, and CXCR5+PD1+ cells to 0.32% to 0.56%. However, CXCR5++PD1++ cells still had a significantly higher proportion of infection at 1.67%. In chronic infection, the proportion of RNA probe positive cells declined in CXCR5-, CXCR5+PD1-, and CXCR5+PD1+ cells (ranging from 0.18% to 0.06%) but remained highest in CXCR5++PD1++ at 0.32%. Thus, TFH as defined by CXCR5 and PD1 staining are consistently infected at higher rates than other memory T cells, but this difference is most pronounced in later infection.

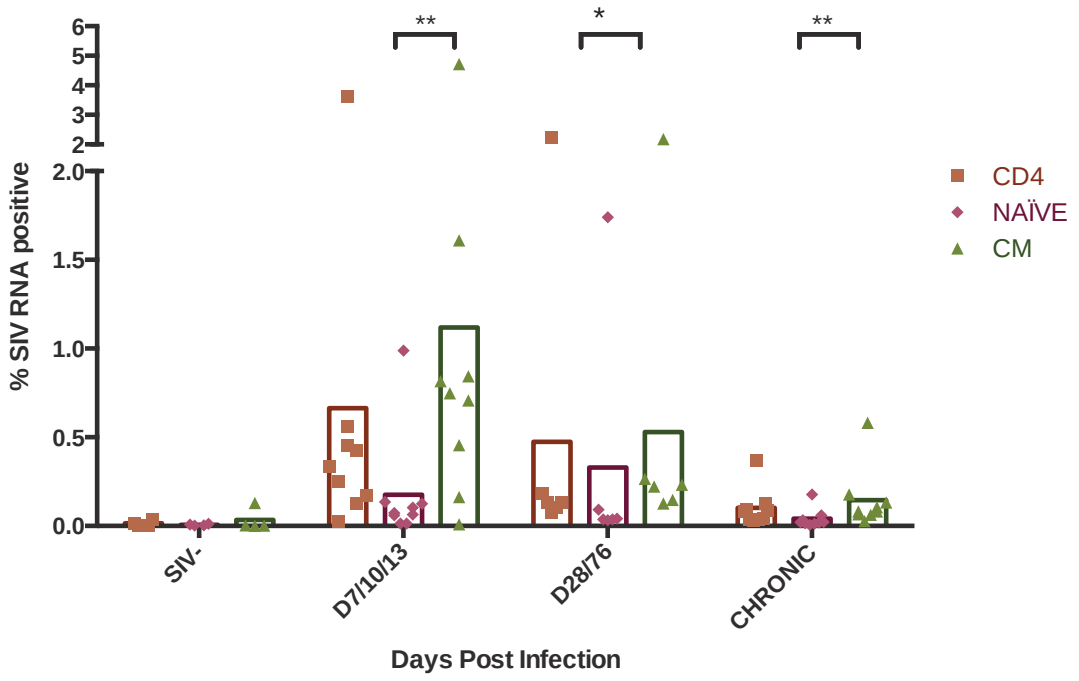


Fig 4.29 Frequencies of RNA probe staining on CD4 T cells. Percentage of RNA probe-positive cells in uninfected, 7/10/13 days post infection, 28/76 days post infection, and chronically infected lymph node biopsies. Within CD4 cells, samples were divided into naïve and central memory (CM) cells based on expression of CD95 and CD28. Significant differences were determined using a one way ANOVA test.

4.2.5.4 Results on acute samples, chronic

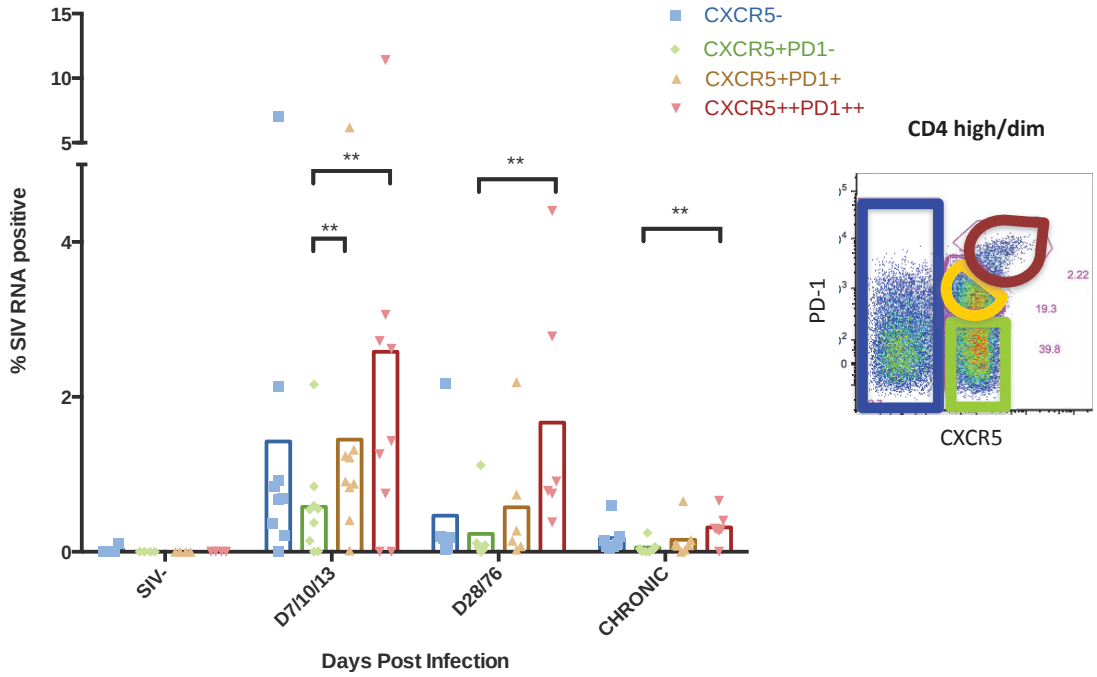


Fig 4.30 Frequencies of RNA probe staining on memory T cell subsets. Percentage of RNA probe-positive cells in CXCR5/PD1 subsets in uninfected, 7/10/13 days post infection, 28/76 days post infection, and chronically infected lymph node biopsies. Significant differences were determined using a one way ANOVA test.

While the increased proportion of SIV RNA-positive TFH compared to other T cells is significant on its own, when examined in the context of the relative frequencies of the central memory T cell populations is more notable. As described earlier in this thesis and in previously published work, the proportion of TFH as a proportion of memory T cells increases in chronic infection, with TFH representing up to 50% or more of central memory T cells in chronic infection. By multiplying the proportion of RNA probe positive cells in each population by that population's proportion in the total central memory pool, we can measure the contribution of each population to the total pool of virus-producing cells (Fig. 4.30). In acute samples, when TFH make up only a small percentage of memory T cells, in spite of having a higher proportion of infection only make a small contribution to infected cells (3% from CXCR5++PD1++ and 9% from CXCR5+PD1+). However, by 28 days post infection when plasma viremia had declined to viral setpoint and infection is largely under control, TFH contribute over 1/3 of all RNA probe positive cells (30% from CXCR5++PD1++ and 8% from CXCR5+PD1+). This trend continues into chronic infection, where TFH are responsible for nearly half of all RNA positive cells (37% from CXCR5++PD1++ and 11% from CXCR5+PD1+). This also shows that TFH are the major compartment of SIV replication and infection in chronic infection. This mirrors the qPCR/qRT-PCR data of SIV DNA/RNA copies per cell, indicating that T follicular helper cells contain the bulk of SIV

infection both in absolute quantities of viral nucleic acids and in proportion of SIV infected cells.

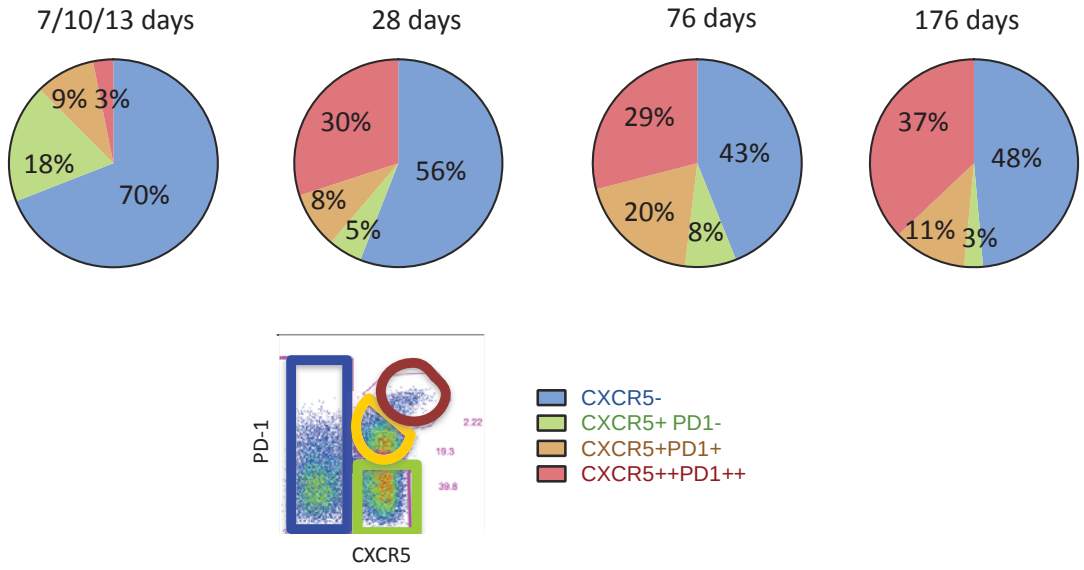


Fig 4.31 Contributions to SIV infected cells from subsets of central memory CD4 T cells. Percentage of RNA probe-positive cells multiplied by the fraction of the central memory yields the net contribution of CXCR5-, CXCR5+PD1-, CXCR5+PD1+ and CXCR5++PD1++ cells to the total pool of infected cells.

4.3 Conclusions

HIV/SIV rapidly infects and depletes CD4 T cells from the periphery and secondary lymphoid organs before being brought under partial control by the host immune system[36]. Even in the absence of clinical symptoms during the latent phase of infection, viral replication is on-going in lymph nodes and induces major changes in the phenotype and function of CD4 T cells [379,380]. While HIV/SIV infection requires at minimum expression of CD4 and a co-receptor (CCR5 or CXCR4), only a fraction of "eligible" CD4 T cells are infected[381]. A detailed understanding of the precise features and markers of the infected cells that harbour replication-competent virus is critical to the design of therapeutic antibodies (such as bispecific antibodies that can direct killing of infected cells) and cure strategies (demarcating the latent reservoir). We observed several changes in the expression of cell surface markers on CD4 T cells in lymph nodes during acute and chronic infection, reflecting increased immune activation and perturbations in T cell subsets. Acutely infected samples had high CD127 expression and low ICOS expression, whereas chronically infected samples had higher ICOS expression and higher SLAM expression than uninfected samples. TFH had distinct patterns of chemokine and activation marker expression when compared to CXCR5- and CXCR5+ CD4 T cells, with lower CCR4, ICOS, and CD25.

TFH are particularly important in HIV/SIV infection because of their role in providing B cell help to generate an effective antibody response. Accumulation of TFH, and their preferential infection in chronic infection, has been reported [170,172], and we examined the history of infection of TFH from early to chronic infection. We designed primers and probes to detect spliced SIV RNAs, to differentiate between cells that were infected by SIV and had an integrated proviral genome (and would therefore show infection via DNA qPCR), and cells that bypassed any cellular restriction and were capable of transcribing viral genes (as evidenced by the presence of spliced and unspliced RNA). Early in infection, most subsets of T cells were equally infected by DNA and RNA qPCR, likely reflecting the availability of target cells and the lack of an effective host immune response. In later infection, follicular T cells had higher copy numbers of both SIV DNA and RNA per cell. This indicates TFH are not only more infected (containing a proviral genome) than other populations, but also produce more viruses (indicated by transcription of viral RNAs) in chronic infection.

Assaying the proportion of infected cells by measuring nucleic acids from bulk sorted samples via FACS is limited in that populations of interest must be determined in advance of quantifying nucleic acids. Information on the number of individual cells infected is lost in the bulk extraction of DNA and RNA, and even clearly delineated subsets of T cells show variable expression of surface markers. We used novel SIV RNA probes in a flow cytometry assay to measure individually infected cells within lymph node populations, allowing us to calculate the proportion of infection within TFH and non-TFH throughout infection. We found that in acute infection, follicular T cells make up only a small percentage of SIV RNA producing cells, but by chronic infection contain the bulk of viral RNAs. The role of follicular T cells as the major compartment of SIV infection and replication is intensified by the relative increase in the proportion of TFH (as a fraction of central memory T cells) from acute to chronic infection. CXCR5+PD1+ and CXCR5++PD1++ CD4+ T cells are not only infected at a higher frequency than other T cells, but make up a greater fraction of the lymph node CD4+ T cells in chronic infection.

5 Intra-host Evolution of Simian Immunodeficiency Virus

5.1 Introduction and aims

As a retrovirus, HIV/SIV is present as both DNA and RNA within an infected cell. After the virus binds to the cell surface and gains entry to the cytoplasm, it undergoes capsid un-coating, reverse transcription of the RNA genome and insertion of the newly formed proviral DNA into the cellular genome. Most HIV/SIV infected cells contain a single proviral copy of viral DNA, indicating a single round of infection[36,382]. Recombination (also called superinfection) occurs when two genetically distinct viruses infect a single cell, and is estimated to occur at a rate of 1% per genome per generation[383]. Genomic integration does not guarantee expression, which requires cellular activation and the HIV/SIV DNA inserted into a favourably transcribed region for viral genes to be expressed. Cellular restriction factors, including the APOBEC family of mRNA editing enzymes, can also limit the replication of HIV/SIV by introducing lethal G-to-A mutations throughout the viral genome. Thus, the pool of DNA-infected cells contains only a fraction of cells that produce viral RNA and new, infectious virus.

Upon initiation of cART, viral RNA levels in plasma decline significantly, dropping by 99% within two weeks due to the elimination of free virus and productively infected cells [384]. The decay in cells containing HIV DNA is less dramatic, with an estimated 8 fold to 30 fold decline immediately following therapy and plateauing after 4 years of cART[385]. In spite of this, the number of cells with replication-competent HIV is low (estimated at 1 in a million PBCMs from patients on long-term ART). This gap in frequencies, between the total DNA-infected cells and the smaller subset that produce virus, is also reflected in the viral sequences derived from the two populations. DNA-infected cells include so-called “graveyard” sequences, and recent studies suggest that the integration site of proviral DNA [69,386,387] may drive persistence, clonal expansion, and RNA expression. In active untreated infection, turnover of different T cell subsets, as well as surveillance of CD8 T cells to different regions of lymph nodes, may affect which T cells harbor graveyard sequences and produce the most viral RNA.

To investigate the evolution of SIV in lymph nodes, we adapted a protocol from Gall et all[327] to amplify and sequence nearly full-length SIV genomes from plasma, cell-associated RNA, and proviral DNA. We sequenced SIV genomes from proviral DNA and cell-associated RNA in lymph node T cell subpopulations to examine the relationship between integrated genomes

and transcribed viral RNA. We also sequenced plasma virus from the same time points to examine the contributions of lymph node cells to plasma sequences throughout infection.

The aims of this chapter are to:

- Develop a protocol to sequence SIV cell-associated RNA and proviral DNA near-full-length genomes
- Evaluate the relationship between circulating plasma virus and cell-associated viral sequences in lymph node T cell subsets.
- Measure the compartmentalization of virus in lymph node T cell subsets, and the relative mutation rates and diversity of proviral DNA and cell-associated RNA
- Compare the intra-host evolution and compartmentalization of virus between individual animals.

5.2 Results

5.2.1 A near-full-length genome sequencing protocol for SIV DNA and RNA

While sequencing a single gene or region of the HIV/SIV genome can identify viral subtypes and provide insight into viral adaptations, full genomes offer a more complete picture of virus evolution. Given the short time frame of this study (twenty four weeks) and the relatively small number of viral mutations expected within a single host in that time period, we wanted to sequence near full length SIV genomes to be better equipped to detect viral intra-host evolution. We adapted the protocol used by Gall et al for HIV sequencing, using four overlapping amplicons spanning the entire coding region of the genome. We designed new primers targeting semi-conserved regions to amplify 2-3 kb overlapping regions of the SIV genome with help from the O'Connor laboratory at the University of Wisconsin, using SIVmac239 as a reference sequence. After a single round of RT-PCR (RNA samples) or PCR (DNA samples), we visualized amplicons on an agarose gel to verify amplification (Fig. 5.1) before pooling amplicons for Illumina HiSeq sequencing.

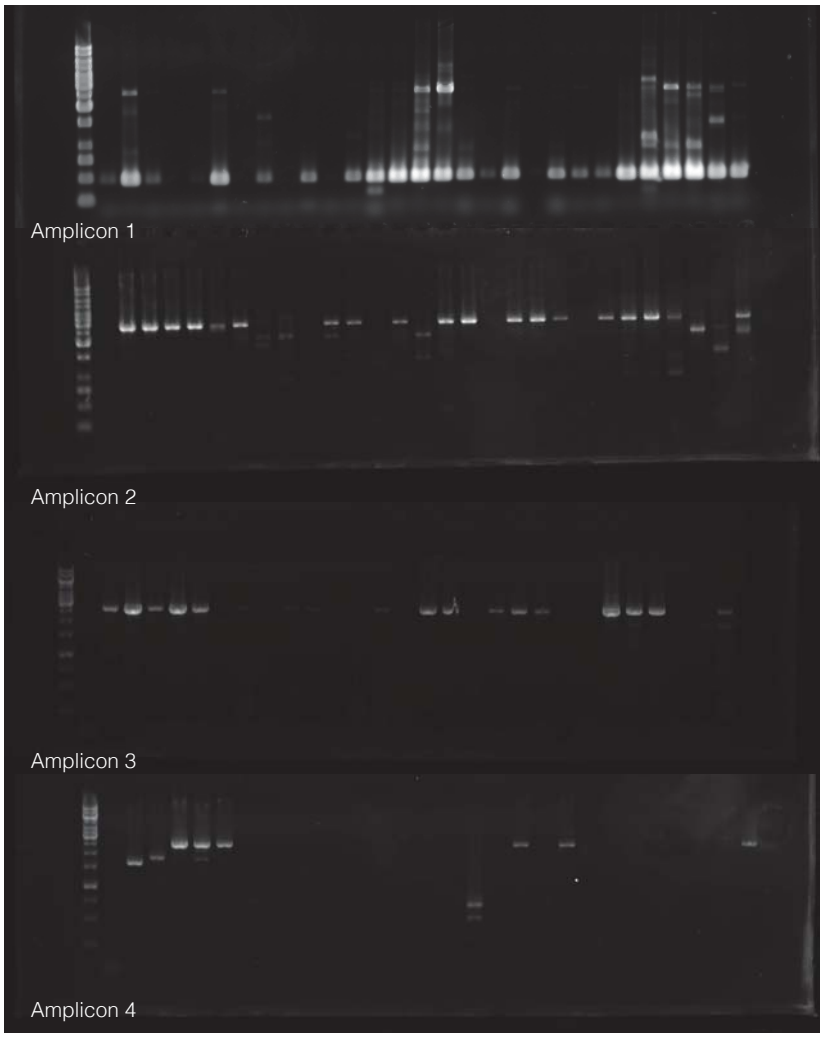


Figure 5.1 Verification of amplicons from SIV RNA (amplicon 1) and DNA (amplicons 2-4). Samples were run on an agarose gel with a 1kb ladder (left) to confirm amplification of the four PCR products (2-3 kb in length) from SIV RNA and DNA. Shown are (top to bottom) representative gels of amplicons 1, 2, 3, and 4. Amplicons 2 and 4 had the highest success rate (indicated by multiple lanes with bands in the 2-3kb range) while amplicon 1 had the highest failure rate.

5.2.1.1 Plasma virus sequencing

Plasma virus was sequenced from three animals (ZF61, ZF76, ZG13) at four time points, at 2, 4, 10, and 24 weeks post infection. Full-genome sequences, with all four amplicons sequenced and assembled into a single contiguous sequence (contig), were generated from six of the twelve samples (Table 5.1). Five of the samples had coverage of 75% of the genome, with three amplicons successfully sequenced and assembled, and the remaining sample had two amplicons sequenced and assembled. Amplicon 1 (covering the first 3kb of the genome) failed to assemble in five of the samples, while amplicon 4 failed to assemble in two of the samples. Amplicons 2 and 3 assembled in all samples.

Table 5.1 SIV PCR Amplicons sequenced from plasma RNA, proviral DNA, and cell-associated RNA

	Plasma RNA	Proviral DNA	Cell-associated RNA
Total Sequences	**12**	**62**	**65**
4 amplicons	6	0	30
3 amplicons	5	42	16
2 amplicons	1	15	8
1 amplicon	0	5	10
0 amplicons	0	0	1
Amplicon Failure			
Amplicon 1	5	62	27
Amplicon 2	0	4	5
Amplicon 3	0	3	19
Amplicon 4	2	18	15

5.2.2 Sequencing and assembly of SIV DNA and cell-associated RNA

To measure how lymph node SIV RNA sequences reflected circulating plasma virus and compare viral sequences present in different T cell subpopulations, we sorted six subsets of central memory CD4 T cells from iliac lymph nodes (at 2, 4, 10, and 24 weeks post infection) and mesenteric lymph nodes (at 24 weeks post infection). From CD3+CD8- CD4- cells, we sorted CXCR5- and CXCR5- cells, and from CD4 bright and dim central memory we sorted CXCR5++PD1++, CXCR5+PD1+, CXCR5+PD1-, and CXCR5- cells (for sample gating criteria, see chapter 4). We extracted DNA and RNA from the samples and split each nucleic acid sample four ways for amplification (Fig. 5.1) Sixty-two DNA samples that had sufficient material for sequencing library preparation and were sequenced on the Illumina HiSeq platform, although none had full genomes assembled (Table 5.2). However, forty-two had three of four amplicons sequenced and assembled, fifteen had two amplicons, and five had a single amplicon assemble. Amplicon 1 failed to assemble in all of the DNA samples sequenced, and amplicons 4 assembled in only eighteen of the sixty two DNA sequences. Amplicons 2 and 3 were assembled in nearly all the samples, failing only in 4 and 3 cases, respectively. Sixty-five cell-associated RNA sequences had sufficient material for sequencing, and thirty had full genomes assembled from the four PCR amplicons. Sixteen had three amplicons successfully amplified, eight had two amplicons, and ten had a single amplicons assemble. Amplicon 1 failed in twenty-seven of the samples, amplicons three failed in nineteen samples, amplicons 4 failed in fifteen samples, and amplicon 2 failed in five samples.

Of the six memory CD4 T cell subsets sorted for viral sequencing, CXCR5++PD1++ and CXCR5+PD1+ CD4 central memory T samples had the highest recovery of DNA amplicons, with 86% and 79% of samples containing 3 amplicons. CXCR5+PD1- and CXCR5- samples

had 67% and 50% of samples with three amplicons sequenced and assembled, while CD4 null samples had 43% and 67% of samples assembled (for CXCR5+ and CXCR5-, respectively). While amplicon 1 failed to assemble in all the DNA samples, amplicons 2 and 3 failed in an average of 18% of DNA samples, while amplicon 4 failed in 41% of samples. In the RNA samples, three of the four CD4 high populations had 3 or 4 amplicons assemble (79% of CXCR5++PD1++, 92% of CXCR5+PD1+, and 82% of CXCR5-). Null CXCR5+ samples also had good recovery of assembled amplicons, with 75% of samples with 3 or 4 amplicons. CD4 bright CXCR5+PD1- cells had only 33% of samples with 3 or 4 amplicons, and CD4 null CXCR5- cells had 40% of samples with 3 or 4 amplicons. Amplicon 1 failed to assemble in 25%-83% of samples, amplicon 2 failed to assemble in 8%-40% of samples, amplicon 3 failed in 18%-60%, and amplicon 4 failed in 9%-58% of samples. Overall, amplicons 2 and 3 were most likely to assemble, while amplicon 4 and particularly amplicon 1, did not assemble in most of the samples. The different CD4 T cell subsets had similar efficiencies of sequence recovery from DNA, while in RNA samples, CD4 bright CXCR5+PD1- and CD4 null CXCR5- samples did not assemble and recover as many full genomes as the other populations.

Table 5.2 Amplicons sequenced per sample and amplicon failure rate.

	CD4 Bright/Dim				CD4 Null	
Proviral DNA	**CXCR5++ PD1++**	**CXCR5+ PD1+**	**CXCR5+ PD1-**	**CXCR5-**	**CXCR5+**	**CXCR5-**
Total Sequences	14	14	12	12	7	3
4 amplicons	0	0	0	0	0	0
3 amplicons	12	11	8	6	3	2
2 amplicons	1	3	4	0	1	0
1 amplicon	1	0	0	6	3	1
0 amplicons		0	0	0	0	0
Amplicon Failure						
Amplicon 1	14	14	12	12	7	3
Amplicon 2	4	3	1	1	2	1
Amplicon 3	3	3	1	1	2	1
Amplicon 4	5	6	5	7	3	0

	CD4 Bright/Dim				CD4 Null	
Cell-associated RNA	**CXCR5++ PD1++**	**CXCR5+ PD1+**	**CXCR5+ PD1-**	**CXCR5-**	**CXCR5+**	**CXCR5-**
Total Sequences	14	12	12	11	8	5
4 amplicons	8	9	1	7	4	0
3 amplicons	3	2	3	2	2	2
2 amplicons	0	0	5	0	0	1
1 amplicon	3	0	3	2	2	2
0 amplicons		1	0	0	0	0
Amplicon Failure						
Amplicon 1	6	3	10	4	3	3
Amplicon 2	3	2	1	1	1	2
Amplicon 3	6	2	6	2	3	3
Amplicon 4	5	2	7	1	1	2

5.2.3 Hypermutation in assembled DNA contigs

For each sample, a consensus genome was created by merging overlapping contigs generated in IVA[333]. In some samples, multiple contigs covering the same genomic region were generated, indicating the presence of multiple distinct viral genomes within the sample. All contigs mapping to the SIV genome underwent hypermutation analysis using Hypermut [334], which detects APOBEC-driven G-to-A mutations in HIV and SIV genomes. Hypermutated sequences are replication-incompetent but if included in reconstructions of viral phylogenies will skew the trees by exaggerating the apparent rate of viral evolution. We used this program to detect and remove hypermutated contigs from the sequence assembly. Hypermut calculates the number of APOBEC G-to-A mutations (with a downstream RD (A/G, A/G/T) relative to control G-to-A mutations (with downstream YN or RC), and calculates a Fisher exact P value that evaluates the probability of APOBEC induced mutations relative to a predetermined reference sequence. The sixty-two DNA samples contained a total of one hundred forty eight contigs that mapped to the SIV reference genome. Seventy contigs had significantly more APOBEC-induced G-to-A mutations than random mutations ($p<0.00001$), indicating that a

significant proportion of the proviral genomes from those samples contained hypermutated regions so that they were reflected in the sequence assembly and were separately mapped into contigs. Hypermutated contigs were assembled in forty one of sixty two samples (66%), but did not differ by animal (69% of samples in ZF61, 64% in ZF76, and 66% in ZG13 contained hypermutated contigs) (Table 5.3). Early sequences (2 and 4 weeks) did not contain more hypermutated sequences (68%, 15/22) than 10 week and 24 week samples (65%, 26/40). The percentage of samples with hypermutated contigs varied widely by population, with 93% (13/14) of CD4 bright CXCR5++PD1++ samples containing hypermutated contigs, 71% of CD4 bright CXCR5+PD1+ samples, 75% of CD4 bright CXCR5+PD1-, and 58% (7/12) of CXCR5- samples. Fewer hypermutated contigs were detected in CD4 null samples (29% of CXCR5+ and 0% of CXCR-), however, there were fewer CD4 null samples overall that had sufficient material for amplification and analysis (7 samples and 3 samples, respectively). Hypermutated contigs were discarded from the assembly of consensus genomes for each sample, and subsequent Hypermut analysis of the resulting genomes confirmed that there was no evidence of APOBEC induced G-to-A mutations in the consensus proviral genomes.

Table 5.3 Hypermutated contigs in SIV DNA samples

ZF61	CD4 bright/dim				CD4 null	
	CXCR5++PD1++	CXCR5+PD1+	CXCR5+PD1-	CXCR5-	CXCR5+	CXCR5-
W4						
W10						
W24 ILN						
W24 MLN						

ZF76	CD4 bright/dim				CD4 null	
	CXCR5++PD1++	CXCR5+PD1+	CXCR5+PD1-	CXCR5-	CXCR5+	CXCR5-
W2						
W4						
W10						
W24 ILN						
W24 MLN						

ZG13	CD4 bright/dim				CD4 null	
	CXCR5++PD1++	CXCR5+PD1+	CXCR5+PD1-	CXCR5-	CXCR5+	CXCR5-
W2						
W4						
W10						
W24 ILN						
W24 MLN						

only non-hypermutated contigs
both hypermutated and non-hypermutated contigs
no contigs

5.2.4 Diversity of sequences within and between animals

HIV/SIV undergoes a complete replication cycle in approximately 48 hours, and mutated sequences arise rapidly within the viral swarm. To measure the evolutionary distance between sequences in different compartments and over time, we calculated p-distances using the proportion of nucleotide sites that differed between each pair of aligned sequences. While this method does not correct for multiple substitutions (e.g. A to T to A) or rate biases between different regions of the genome, it nonetheless offers a metric for the mean diversity within a group of sequences. A large mean distance indicates that many of the sequences in that group contain different mutations, while a small distance indicates that the sequences are similar.

There was no difference in the mean pairwise distance of plasma samples between animals, with an average distance between all plasma sequences of 0.00198 for ZF61, 0.00192 for ZF76, and 0.00180 for ZG13 (Fig. 5.2). However, the mean distance between all samples (plasma, proviral DNA and cell-associated RNA) was significantly lower in ZG13 (0.00195) than in ZF61 (0.00255, $p<0.00001$) and ZF76 (0.00248, $p<0.00001$). There was no significant

difference between the mean distances of ZF61 and ZF76 total sequences (p=0.38) or DNA sequences (p=0.33) but their mean distance between RNA sequences was slightly higher in ZF61 (0.00297 vs. 0.00258, p=0.036). ZG13 had significantly lower mean distances between its DNA (0.00193 vs. 0.00235 and 0.00250, both p<0.00001) and RNA (0.00196 vs. 0.00297 and 0.00258, both p<0.00001) sequences than in ZF61 or ZF76. This indicates that the sequences from animal ZG13 were generally closer to each other (i.e. less divergent) than the sequences from animals ZF76 or ZF61. The diversity in the RNA sequences in ZF61 was higher than the DNA.

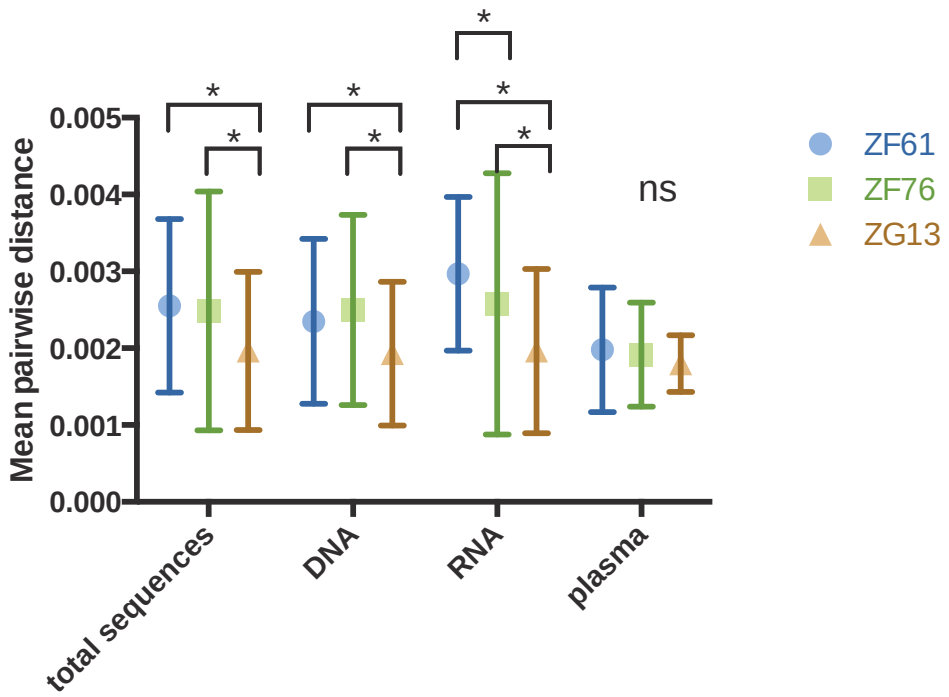


Fig. 5.2 Mean pairwise distance between sequences in individual animals. Mean pairwise distance was calculated for all sequences, proviral DNA only, cell-associated RNA only, and plasma from all time points for animals ZF61, ZF76, and ZG13. Significant differences were determined using a one way ANOVA test.

Next, we examined whether different CD4 T cell subsets contained different diversities of sequences (Fig 5.3). The mean distance between DNA sequences was not significantly different in CXCR5++PD1++, CXCR5+PD1+, and CXCR5+PD1- cells (means of 0.0033-0.0036), but was significantly lower in CXCR5- cells (0.0025, p<0.0003). However, in RNA from the same samples, CXCR5++PD1++ samples had the lowest mean distance (0.0027) and were significantly lower than the distances between CXCR5+PD1+, CXCR5+PD1-, and CXCR5- samples (p<0.01). CXCR5+PD1+ samples, with a mean distance of 0.0039, were significantly higher (p<0.006) than CXCR5+PD1- (0.0031) and CXCR5- (0.0034). The differences in the mean distance between the DNA and RNA samples varied, with CXCR5++PD1++ and CXCR5+PD1- samples having a greater genetic distance in DNA than in RNA (p<0.00001 and p<0.02), but CXCR5+PD1+ and CXCR5- samples with a greater genetic distance in RNA than in DNA (p<0.004 and p<0.0001). Thus, in DNA sequences,

CXCR5- cells were less diverse than other CD4 T cell subsets. For RNA sequences, CXCR5++PD1++ sequences were most similar, while CXCR5+PD1+ sequences differed the most.

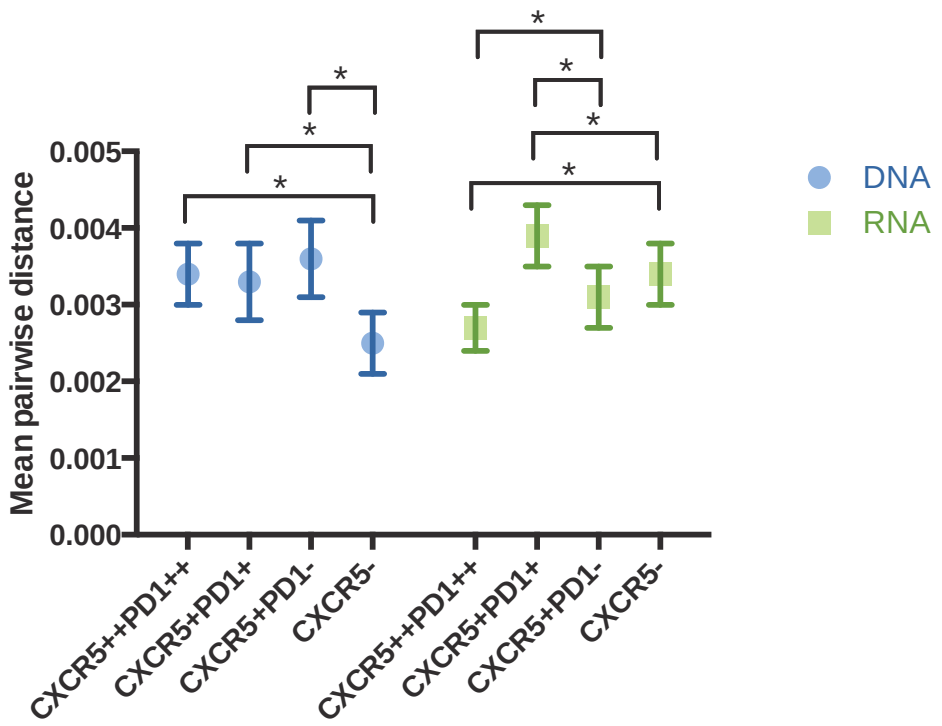


Fig. 5.3 Mean pairwise distance between T cell subpopulations. Mean pairwise distance was calculated between all DNA and RNA sequences isolated from populations of CXCR5++PD1++, CXCR5+PD1+, CXCR5+PD1- and CXCR5- CD4 T cells. Significant differences were determined using a one way ANOVA test.

Next, we examined the distances between TFH (CXCR5++PD1++ and CXCR5+PD1+) and non TFH (CXCR5+PD1- and CXCR5-) samples in early infection (2 and 4 weeks) and late infection (ten weeks and twenty-four weeks). There was no significant difference between the mean distances in DNA and RNA samples from early TFH and early non-TFH, or between late TFH and late non-TFH (Fig. 5.4). However, the mean distance between early and late samples was significant ($p<0.0001$) in both DNA and RNA TFH, with higher mean distance in the late samples (0.0038 in DNA and RNA) than in the early samples (0.0021 and 0.0023). There was also a significant difference ($p<0.00001$) between the mean distance in early vs. late samples in non-TFH, with an increase from 0.0017 and 0.0018 to 0.0038 and 0.0039 in DNA and RNA, respectively. Over the time frame in this study, later sequences show a wider range of mutations and diversity, while there is not a difference in the mean pairwise distances within TFH and non-TFH.

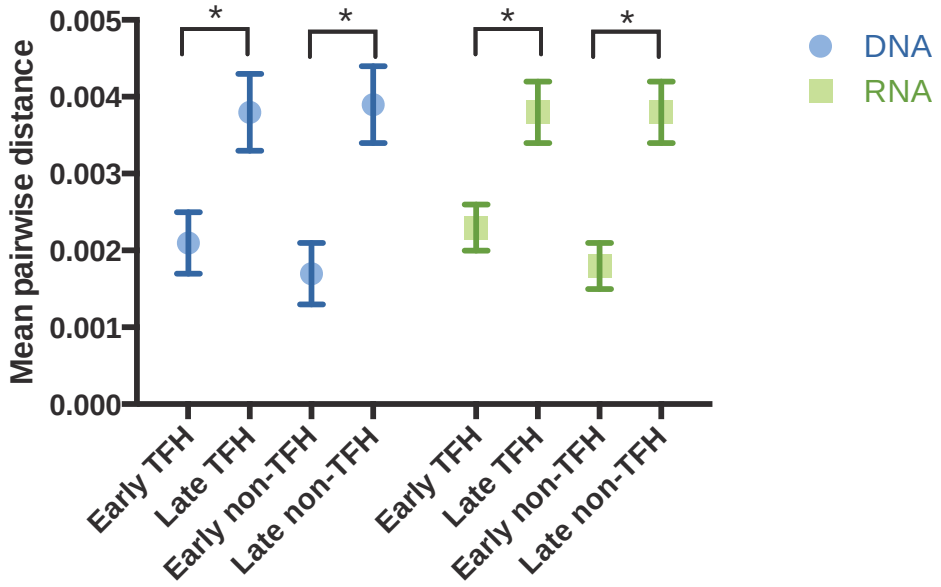


Fig. 5.4 Mean pairwise distance between early and late T cell subpopulations. Mean pairwise distance was calculated for all early (2 and 4 weeks post infection) and late (10 and 24 weeks post infection) DNA and RNA sequences from TFH (CXCR5++PD1++ and CXCR5+PD1+) and non-TFH (CXCR5+PD1- and CXCR5-) CD4 T cells. Significant differences were determined using a one way ANOVA test.

5.2.5 Phylogeny of SIV early infection

To explore the phylogenetic relationships between viruses sampled in this study, we constructed four phylogenetic trees: one with plasma sequences, one tree for each animal containing all the sequences from that animal, and one tree of sequences from all three animals. all samples from individual animals, and all viruses in the study. Sequences were aligned using MAFFT in Aliview and trees were constructed using FastTree with bootstrapping. All alignments included reference sequence KC52232, the consensus genome of the infectious stock used to infect the animals in this study[388]. All trees were rooted on the reference genome and are displayed with nodes in decreasing order.

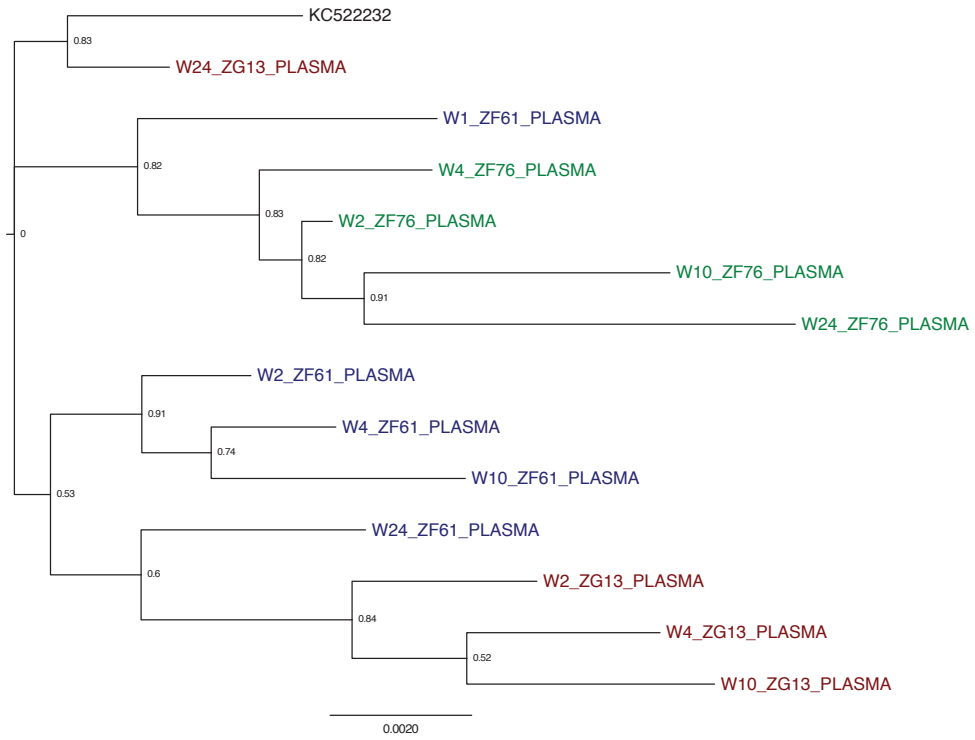


Fig. 5.5 Maximum likelihood phylogenetic tree of near-full-length genome plasma sequences. The maximum likelihood tree rooted on reference sequence KC52232. Bootstrap values on each node indicate the reliability of the split and were calculated using 1,000 resamplings.

5.2.5.1 SIV plasma phylogeny

We first constructed a phylogenetic tree of all the plasma consensus sequences, consisting of samples from infection weeks 2, 4, 10, and 24 in each of three animals (one animal, ZF61, had an additional sample from 1 week post infection) (Fig. 5.5). All four sequences from animal ZF76 (in green) clustered together, with the earlier sequences (weeks 2 and 4) close to the root and the week 10 and 24 sequences successively farther away. The four ZF76 sequences and the earliest ZF61 sequence (in blue) made up a larger cluster, distinct from the remaining ZF61 sequences and three of the four ZG13 sequences (in red). Three ZF61 sequences (from weeks 2, 4, and 10) formed a small cluster, with the week 24 sequences in a nearby branch. The week 2, 4, and 10 ZF61 sequences had increasing branch length with time of sampling. Animal ZG13 had three of four sequences (from 2, 4, and 10 weeks) in a cluster farthest away from the root, also with branch length increasing with time post infection. However, the latest ZG13 plasma sequence (week 24) was closest to the root but far from the earlier ZG13 sequences.

5.2.5.2 Interhost phylogeny

Next, we wanted to look at the relationship between all the sequences in this study. The first phylogenetic tree containing plasma, RNA, and DNA sequences from all animals is divided into three well-supported clades which largely cluster by animal (Fig. 5.6). The first, which

shows the least divergence from the reference sequence, contained most of the sequences from animal ZF61 (blue) and two sequences from ZF76 (green). The second major clade is almost entirely made up of sequences from animal ZF76, with one sequence from ZG13 (red). The third clade contains all but one of the ZG13 sequences, but also several ZF61 and ZF76 sequences. The first two clades, from ZF61 and ZF76, generally show a classic model of on-going viral escape from immune pressure. The virus in each of these two animals is close to the reference infection sequence at early time points, and later (4 weeks to 24 weeks) sequences are derived from the earlier circulating viruses. Some early viruses explore alternative sequence space but were ultimately unsuccessful, and most of the samples from ZF61 and ZF76 belong to major lineages. In contrast, sequences from animal ZG13 do not form a single lineage, but instead show more extensive sampling of the sequence space and evolution of several minor lineages over time. The sequences from ZG13 make up several sub-linages that temporally co-circulate within the animal. Overall, these data suggest two models of within-host viral evolution. In the first, one lineage dominates the sequences sampled with directed evolution over time, likely due to constant immune pressure. The other shows several smaller lineages arising throughout the sampling period, but without any single lineage dominating the samples over the course of the infection. The most significant branching in the tree was due to animal, not time point, nucleic acid, or T cell population.

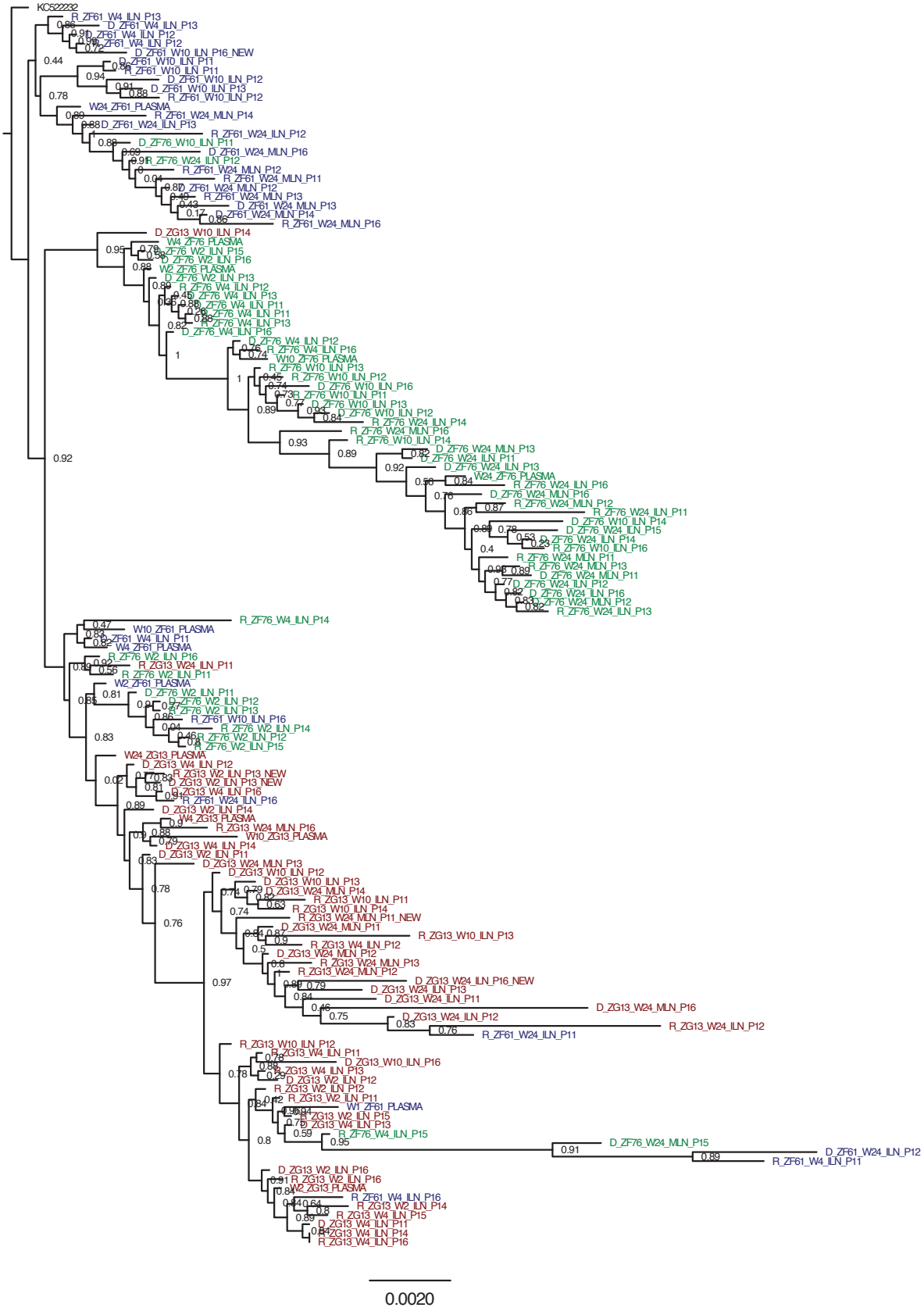


Fig. 5.6 Maximum likelihood phylogenetic tree of near-full-length genome sequences from all animals. An alignment of plasma, cell-associated RNA, and proviral DNA sequences from animal ZF76 were used to construct a maximum likelihood tree rooted on reference sequence KC52232. Bootstrap values on each node indicate the reliability of the split and were calculated using 1,000 resamplings. Sequences are named by nucleic acid (D-DNA, R-RNA), animal, time point (W2, W4, W10, W24), tissue (ILN-iliac, MLN-mesenteric lymph node), and sample population (P11- CXCR5-, P12- CXCR5++PD1++, P13 – CXCR5+PD1+, P14- CD4 null CXCR5+, P15- CD4 null CXCR5-, P16 – CXCR5+PD1-)

5.2.5.3 Intrahost phylogeny

Next, we looked at the evolution and phylogenetic relationship within each of the three animals. In ZF61, the sequences formed four distinct clusters, broadly clustering by time points (Fig. 5.7). The cluster closest to the root sequence contained most of the week 4 sequences, while the next-nearest cluster contained the early plasma sequences. The most distant cluster contained most of the week 24 sequences, although the week 1 plasma sequence was in the middle of this cluster. The week 10 sequences lay between the early and late sequences. At 24 weeks post infection, where sequences from two distinct lymph nodes (mesenteric and iliac) were obtained, the mesenteric sequences clustered together far from the root, while some of the iliac sequences clustered with week 10 sequences. The branching structure of the ZF61 tree suggests rapid exploration of the sequence space by the virus (although mostly unsampled in this study) with no single lineage predominant, until week 24 when the bottom clade becomes the most prevalent. RNA and DNA sequences were evenly dispersed throughout the tree and did not cluster together. Although we did not recover matched DNA and RNA sequences from all cell populations, in those that we did we were able to investigate how closely the RNA sequences reflected the DNA. In some cases, the DNA and RNA sequences were nearly identical branches on the same node, but in others they were separated by several branches and sequences.

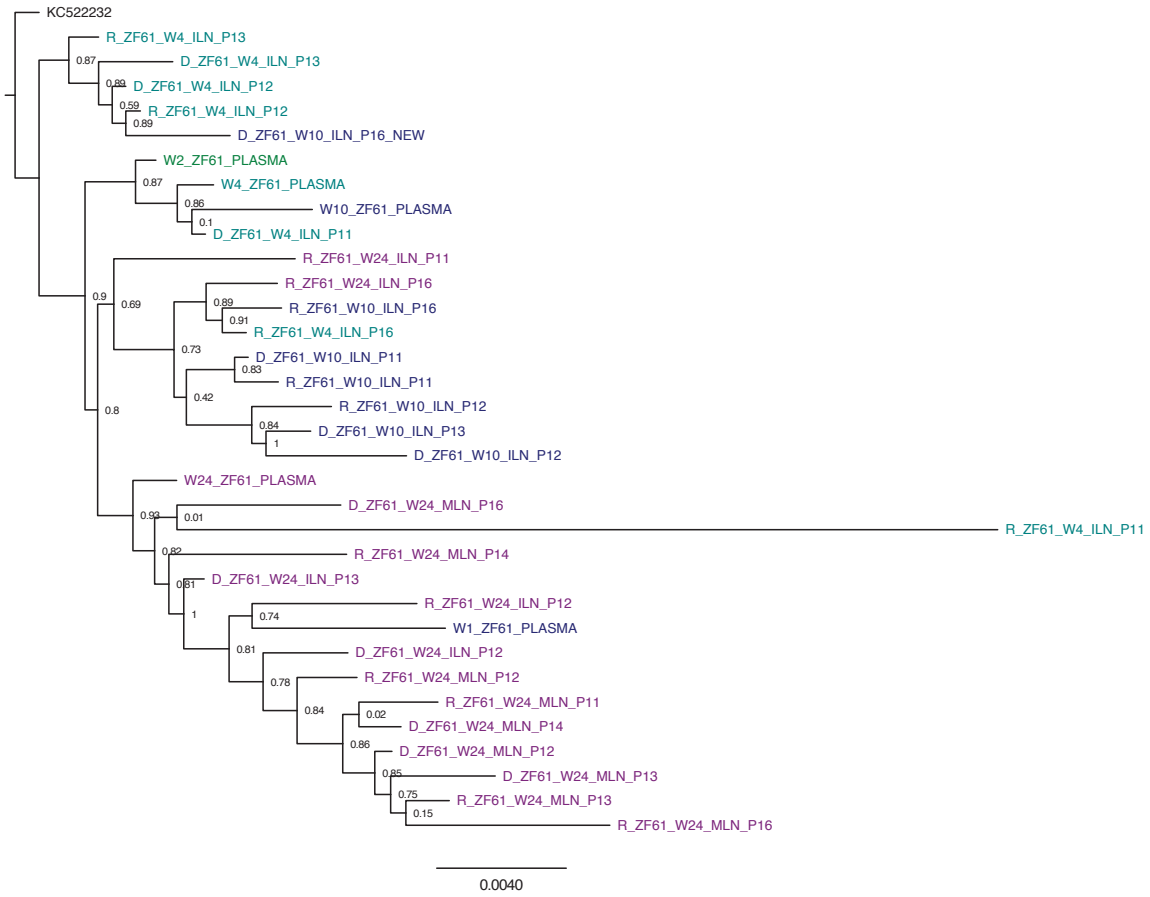


Fig. 5.7 Maximum likelihood phylogenetic tree of near-full-length genome sequences from animal ZF61. An alignment of plasma, cell-associated RNA, and proviral DNA sequences from animal ZF76 were used to construct a maximum likelihood tree rooted on reference sequence KC52232. Bootstrap values on each node indicate the reliability of the split and were calculated using 1,000 resamplings. Sequences are named by nucleic acid (D-DNA, R-RNA), animal, time point (W2, W4, W10, W24), tissue (ILN-iliac, MLN-mesenteric lymph node), and sample population (P11- CXCR5-, P12- CXCR5++PD1++, P13 – CXCR5+PD1+, P14- CD4 null CXCR5+, P15- CD4 null CXCR5-, P16 – CXCR5+PD1-)

ZF76 sequences formed two major clusters (Fig. 5.8). The first and smaller cluster contained most of the week 2 sequences and was close to the root, while the remaining week 2 sequences were part of the large branch that made up the majority of the tree. This large branch had a very strong time resolution, with the week 2 sequences closest to the root, followed by the week 4 sequences and week 10 sequences, with the week 24 sequences farthest away from the root. There were two sequences, from weeks 10 and 4, were on a separate branch close to the reference. DNA and RNA sequences were evenly distributed throughout the tree. The mesenteric and iliac sequences from week 24 were not clustered by lymph node, but were evenly intermingled. While there were small differences in the DNA and RNA sequences obtained from the same time point and cell subset, they were generally close to their counterparts on the tree, indicating that the proviral DNA did not contain a majority of substantially different genomes than the expressed RNA. Plasma sequences clustered with the DNA/RNA from the same time point. Like the ZF61 tree, the ZF76

samples showed evidence of a single lineage gradually mutating over time in response to host immune pressure, with early sequences closest to the root reference sequences, and later sequences on the most distant nodes.

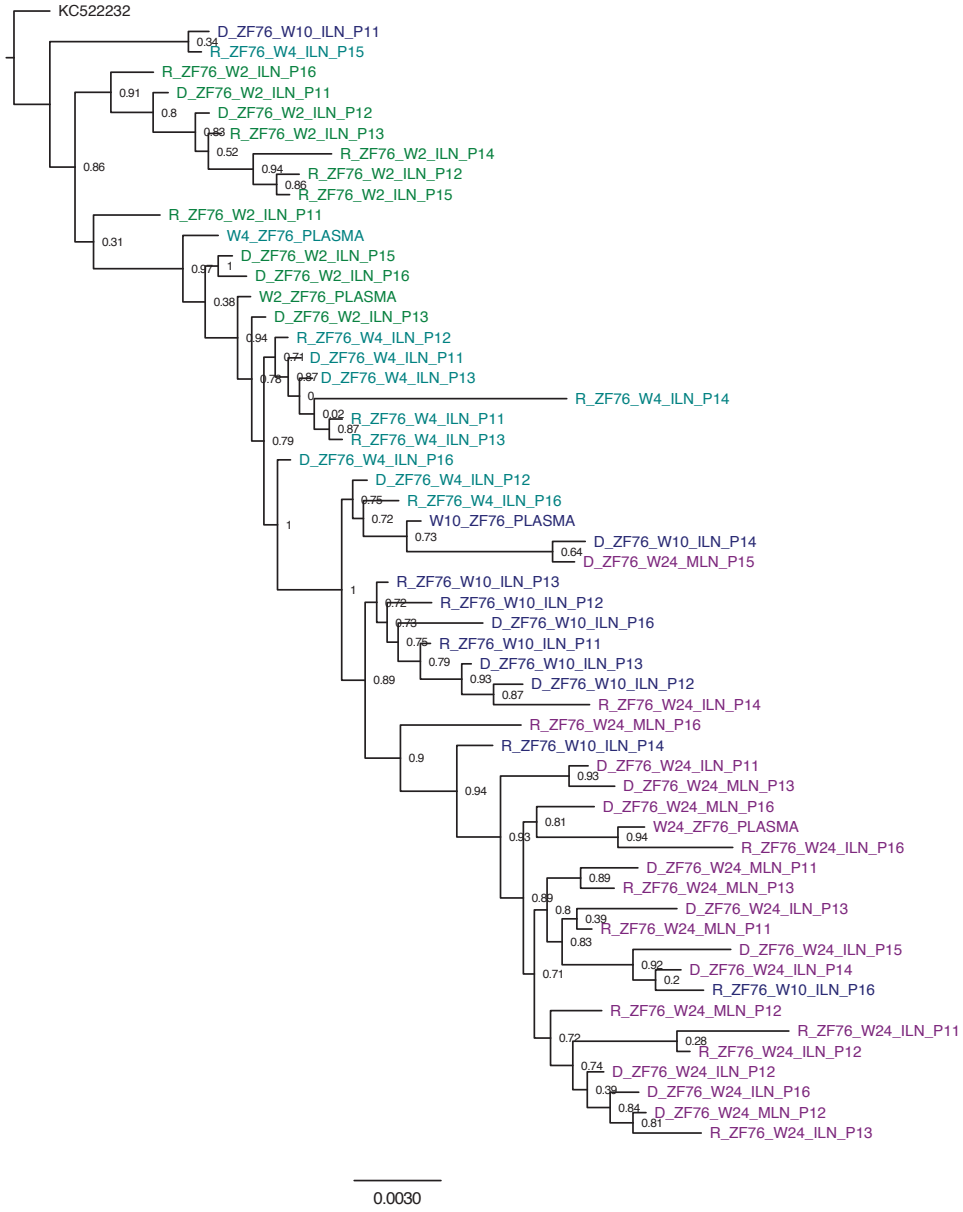


Fig. 5.8 Maximum likelihood phylogenetic tree of near-full-length genome sequences from animal ZF76. An alignment of plasma, cell-associated RNA, and proviral DNA sequences from animal ZF76 were used to construct a maximum likelihood tree rooted on reference sequence KC52232. Bootstrap values on each node indicate the reliability of the split and were calculated using 1,000 resamplings. Sequences are named by nucleic acid (D-DNA, R-RNA), animal, time point (W2, W4, W10, W24), tissue (ILN-iliac, MLN-mesenteric lymph node), and sample population (P11- CXCR5-, P12- CXCR5++PD1++, P13 – CXCR5+PD1+, P14- CD4 null CXCR5+, P15- CD4 null CXCR5-, P16 – CXCR5+PD1-)

Unlike the other two animals, ZG13 did not have a single distinct lineage with early sequences near the root and later sequences at the greatest distance (Fig. 5.9). Instead, ZG13 sequences formed several smaller clusters with sequences from multiple time points in each cluster. Most week 2 and week 4 sequences were in two clusters: a small cluster near the root, as well as a larger cluster that was farthest from the root. The week 4 plasma sequence was in

a separate cluster with week 10 and week 24 sequences. A separate cluster had both DNA and RNA sequences from weeks 4, 10, and 24, from multiple T cell populations. As in the other trees, DNA and RNA sequences were distributed throughout the tree, and many of the DNA and RNA sequences from the same subset and time period were clustered closely. Overall, several distinct lineages were present in the early sampling but not in later samples, while others emerged later and were not closely related to the early samples. Some of the later sequences more closely resemble the root sequence, and may be present at low frequencies in early samplings but were not represented in this tree since we used only majority consensus sequences.

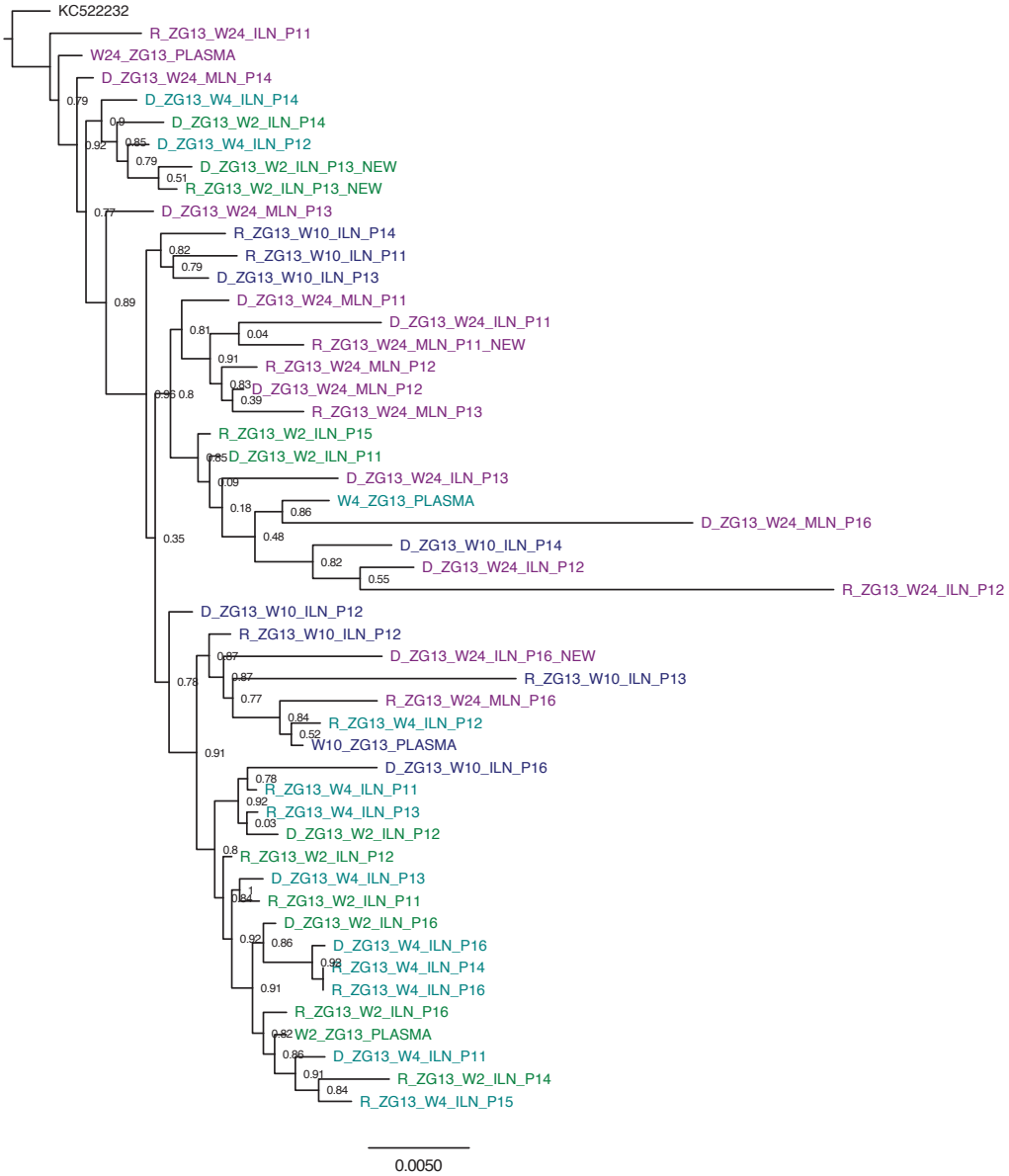


Fig. 5.9 Maximum likelihood phylogenetic tree of near-full-length genome sequences from animal ZG13. An alignment of plasma, cell-associated RNA, and proviral DNA sequences from animal ZF76 were used to construct a maximum likelihood tree rooted on reference sequence KC52232. Bootstrap values on each node indicate the reliability of the split and were calculated using 1,000 resamplings. Sequences are named by nucleic acid (D-DNA, R-RNA), animal, time point (W2, W4, W10, W24), tissue (ILN-iliac, MLN-mesenteric lymph node), and sample population (P11- CXCR5-, P12- CXCR5++PD1++, P13 – CXCR5+PD1+, P14- CD4 null CXCR5+, P15- CD4 null CXCR5-, P16 – CXCR5+PD1-)

5.3 Conclusions

Sequencing SIV genomes throughout infection offers a perspective on the evolution of the virus within a single host. As the host immune system begins to bring viremia under control, the virus adapts to the humoral and cell-mediated defences and escapes. Here, we infected three animals with the same infectious stock of SIVmac251, and tracked the evolution of virus in plasma and in lymph nodes. We adapted a protocol for sequencing near-full length genomes from HIV plasma RNA for our SIV cell-associated RNA and proviral DNA samples. The protocol for sequencing proviral DNA alongside cell-associated RNA was relatively robust, although dependent on sample input, with additional efficiencies to be gain from addition optimization of primer design. While we obtained at least partial genomes for all of the samples sent for sequencing (after screening for amplification using visualization on an agarose gel), there were variable efficiencies between the amplicons. The sensitivity of the assay was also a factor in the recovery of SIV sequences, as many of the samples had only a few thousand cells, of which only a fraction were expected to contain proviral DNA, and a subset of those containing RNA. An RNA producing cell can produce hundreds of thousands of viruses each with two copies of the RNA genome[389], whereas each infected cell contains an average of 1.5 copies of SIV DNA, which contributes to the lower recovery of DNA genomes. The presence of hypermutated sequences in the DNA samples indicates that a significant portion of the proviral genomes in lymph node CD4 T cells contained APOBEC-mutated genomes, with the highest frequencies of hypermutated sequences detected in CXCR5++PD1++ cells.

Plasma virus sequences were closely related to lymph node sequences from the same time point, but were not clustered with a specific T cell subpopulation. CXCR5- DNA samples generally had the lowest diversity (as measured by mean pairwise distance between sequences), while CXCR5+PD1+ RNA samples had the highest diversity. Both TFH and non-TFH sequences 2 and 4 weeks post infection, had lower diversity than later (10 and 24 week) samples.

A maximum likelihood phylogenetic tree with all the sequences showed that the genomes clustered largely by animal, and each animal explored a distinct sequence space. In two of the animals, nearly all the sequences in each animal were in a single lineage that grew from the founder sequence and accumulated mutations over the course of infection. In the third animal, several smaller lineages were present and there was not a clear linear evolution from early to late infection. We used consensus genomes for each sample, and do not in this study

investigate minority variants. In the two animals where a single lineage appears to persist over the course of infection, we may be missing low frequency viral mutations and haplotypes. In the third animal, where we see several linages appear and disappear, we may be missing the low-level persistence of some of those lineages.

6 Discussion

The immunological events of early HIV infection are critical to establishing the latent reservoir and determining the severity of disease progression. Most HIV-infected individuals fail to develop a broad and potent antibody response, and those who do develop broadly neutralizing antibodies typically do so only after years of infection. The development of a robust humoral response requires the aid of T follicular helper cells, a specialised subset of CD4 T cells that reside in the B cell follicle and promote somatic hypermutation and affinity maturation of antigen-experienced B cells. TFH are preferentially infected in chronic HIV infection, but whether they are infected in early infection and the role that this might play in the development of the early antibody response and in the establishment of the latent reservoir is not well defined. Lymph nodes and secondary lymphoid tissues are the primary site of HIV infection and replication, but are difficult to obtain in clinical studies, especially in the acute phase of infection.

We used the SIV-rhesus macaque model system to investigate three aspects of early HIV/SIV infection. One of the key advantages of the macaque model system is the ability to infect with a known virus stock and observe very early immunological events in compartments that are largely inaccessible in human clinical studies. By serially sampling lymph nodes, bone marrow, and blood from the onset of infection to peak plasma viremia and subsequently through viral setpoint and chronic infection, we can capture and compare different aspects of the major stages of infection. First, we sequenced the SIV envelope-specific immunoglobulin response to evaluate the development of the antibody repertoire from the onset of the adaptive immune response. Second, we performed phenotypic profiling and measured infection of TFH and other CD4 T cell subsets in lymph nodes. Finally, we sequenced SIV in plasma and in lymph node T cell subsets to measure differences in archived and expressed virus and track the evolution of the virus in response to host immune pressures.

We find that:

- The IgG SIV-specific immune response begins to develop characteristics of a broadly neutralizing response within six months of infection, with increased mean VH gene mutation, higher proportion of sequences with long CDR3 regions, and increased diversity of V and J gene usage.

- SIV-specific heavy chain sequences have distinct V and J gene usage from total memory B cells, and there are more similar sequences between animals in the SIV-specific compartment than in the total memory compartment.
- Several chemokine markers are differentially expressed on acutely, chronically, and uninfected lymph node CD4 T cells, including markers of trafficking, activation, and proliferation.
- In very early infection, central memory CD4 T cell subsets are equally infected, but as virus is controlled and plasma viremia reaches set-point, follicular T cells make up the majority of productively infected cells.
- Within-animal virus sequencing shows strong time-driven evolution, with greater diversity in later sequences, and intra-host evolution exhibits different patterns of evolution likely as a result of differential immune pressures.

The expansion and diversification of the SIV-specific antibody response we observe indicates that the humoral response to SIV infection explores a successively wider range of targets and epitopes as infection progresses. The early IgG response is narrow and non-neutralizing, as the first generations of antibodies have not undergone the multiple rounds of somatic hypermutation and affinity maturation that are present later in infection. The IgM responses, in contrast, do not show increased mutation or expansion over the course of infection. This suggests the IgM response is primarily made of up newly minted, un-mutated SIV-specific B cells, whereas the IgG response is likely a combination of newly class switched B cells as well as highly mutated B cells. In the period observed here, the pool of SIV-specific IgG memory B cells contains a mixture of less-mutated sequences and, as time progresses, more and more highly mutated sequences that have undergone multiple rounds of somatic hypermutation and affinity maturation before exiting the germinal center.

The non-SIV specific memory B cells represent an archive of all the responses to infections in the natural history of the individual animals. The early SIV-specific response is distinct from the total memory, but as infection progresses resembles the total memory response more in V and J gene usage and measures of diversity and clonality, suggesting that the later SIV-specific humoral response is maturing and bears some of the hallmarks of previous, successful humoral immunity. This is confirmed by the neutralization of the infection stock later in infection and the decrease in plasma viremia after several weeks of infection. However, the failure to completely control virus shows that the antibody response is indeed unable to keep equal pace with viral evolution and is instead playing catch-up to circulating virus.

While we did not detect a signature V/J gene response to SIV envelope probe in the animals studied here, it was clear that the early SIV-specific response utilizes a more narrow pool of V and J genes compared to the total memory response in each animal. This is likely reflective both of the range of epitopes presented on the SIV envelope alone (instead of envelope, gag, pol, and other SIV proteins), and the specific SIV envelope probe used to isolate SIV-specific B cells in this study. Probe technology is evolving rapidly, and using newer generations of trimers including SOSIP probes would select a broader selection of SIV-specific cells. Further studies, including functional analysis of cloned antibodies, would yield important insights into specific epitopes targeted by the antibody response at different time points. This analysis would also help elucidate which BCRs within the total SIV-specific B cell repertoire contribute to the neutralization of virus. This study is also limited by the small number of animals used, and the conclusions would be strengthened by the inclusion of a greater number of specimens.

While long CDR3 regions are associated with several HIV bnAbs because of their ability to access hidden pockets and residues on HIV envelope trimers, over the time period in this study we did not detect a significant shift in the median CDR3 length of SIV-specific heavy chains. However, it is unclear whether sequences with longer CDR3s need to make up a significant portion of the total antibody repertoire to provide neutralization or if they can have an impact on neutralization even as a small fraction of the pool of antibodies. Physiochemical features of CDR3 regions offer some insight into how the humoral immune response is tuned to this particular pathogen and set of epitopes, but like CDR3 regions, may require analysis of individual antibodies instead of being used as a bulk measure to gain additional insight into correlates of neutralization.

The assays used to measure immunoglobulin repertoires can affect the interpretation and scope of the results. While paired heavy and light chains offer a more complete portrait of the antibody response, there is evidence that in HIV/SIV infection, heavy chains provide most of the binding to antigen and can be informative on their own in understanding the evolution of the humoral response. We briefly investigate single-cell sequencing and compare it to the 5’ RACE bulk method. Single cell offers paired heavy and light chains and in the method used here, a direct conduit into performing antibody cloning and neutralization assays. We find that single cell assays using VH primers may miss some heavily mutated BCRs, although they are not prone to PCR amplification bias in counting multiple clones the way that non-multiplexed

bulk BCR sequencing is. The field of immunoglobulin repertoire sequencing is changing rapidly to reflect the challenges of sequencing this population of cells, and newer methods will continue to improve on the sequencing of disease-specific repertoires.

One of the key underlying assumptions of a B cell lineage immunogen HIV vaccine is that the same features of a protective immune response can be induced equally across individuals regardless of genetic background and stochastic events. Public immunoglobulin repertoires in HIV infection is a new area that requires a great deal more study, but our initial observations of similar sequences in the SIV-specific IgG heavy chains repertoires of different animals suggests that while there is great diversity within the individual host response to SIV, parts of the SIV-specific response is shared among animals. We see that a measureable part of the antigen-specific repertoire is shared between samples- significant given the limited sampling (a single lymph node or several mL of blood representing tens of thousands of cells out of the billions of total B cells). Although broadly neutralizing antibodies are rare and have proven difficult to isolate, the similarities between the SIV-specific repertoires between animals signals that they have much in common in adapting to SIV.

An effective B cell response to infection requires help from T follicular helper cells: thus, any finding that TFH are particularly important to sustaining HIV infection has implications for their role in providing the necessary help. While we do not directly investigate the consequences of preferential infection of TFH on the B cell response, the phenotypic changes in TFH in both acute and chronic infection compared to uninfected animals and their high levels of infection show that they differ substantially from TFH in normal, resolved infections. This analysis primarily uses surface staining of CXCR5+ and PD-1 to identify TFH, but does not distinguish T follicular regulatory cells from TFH. TFR represent a small but significant percentage of CXCR5+PD1+ CD4 T cells in healthy infection, and an unknown percentage in acute and chronic HIV infection. Thus, the populations of TFH studied in this thesis are “contaminated” to some degree with TFR that do not directly participate in TFH-B cell interactions.

While the combination of SIV DNA and RNA qPCR in bulk cell populations gives an estimate of the frequency of infection, the RNA probe technology yields more detailed information on infection at the single cell level. By identifying individually infected cells in combination with a multi-parameter flow cytometry panel, we can track changes in the actively infected cells throughout infection and confirm that CXCR5+PD1+ and

CXCR5++PD1++ CD4 T cells contain a higher percentage of SIV-infected cells and this difference is more pronounced in chronic infection. This is a subtle but important distinction, that the CD4+ T cells in the B cell follicle do not just produce high copy numbers of SIV nucleic acids but also contain the greatest percentage of SIV-producing cells.

We show here that in acute infection, follicular T cells are infected more or less equally by DNA and RNA qPCR – given the massive infection and high levels of plasma (and presumably lymphoid) viremia, it is logical that all susceptible CD4+ T cells would be equally infected. Shortly after peak viremia, follicular T cells have a higher proportion of infection than other central memory T cells, and as infection progresses contain more and more of virus-producing cells. That these cells are not only more infected (by measuring SIV proviral integration) but also contain more virus (measuring cell-associated RNA), in conjunction with the increase in the proportion of central memory CXCR5+PD1+ and CXCR5++PD1++ CD4+ T cells, raises further questions about why this population of cells makes up a significant proportion of HIV/SIV infection and replication.

The changes we observe in the phenotypic profiles of acutely and chronically infected lymph node CD4 T cells reflect the consequences of prolonged inflammation, most notably increased expression of ICOS and low CD127. We do not directly measure the ability of TFH to provide B cell help, but it is clear from this and other studies that their accumulation and altered phenotype have an impact on B cell responses in HIV/SIV infection. The possible mechanisms between preferential infection of TFH – their location in an immune-privileged area with limited penetration by CTL, the increased activation and accumulation of TFH in the chronic inflammation of HIV disease, or other yet to be determined factors – are an important area of further study. We are hoping to begin to address these questions using single cell RNA sequencing of TFH in acute and chronic infection and look for markers of longevity and/or turnover.

We also, for the first time that we know of, sequence both SIV proviral DNA and cell-associated RNA from lymph node T cell subpopulations to track virus compartmentalization and evolution within individual animals. Though this protocol is sensitive to sample input, we were able to reliably detect partial SIV genomes with input of as few as two to three thousand cells, of which only a fraction were SIV-infected. Although we expect variation in the viruses infecting the individual cells in a single subpopulation, because of the difficulties in reconstructing haplotypes from next-generation bulk sequencing data we use a single

consensus sequence for each sample in our distance and phylogenetic analyses. This represents the most frequent mutations within a given population, and yields a general impression of the viruses circulating within a population instead of a detailed accounting of the individual infected cells. Cross-priming during PCR amplification is an issue with bulk sequencing and further complicates the assembly of individual viral haplotypes within a sample, making consensus sequences represent the majority mutations, not necessarily a true circulating viral haplotype. While an imperfect measure, bulk sequencing of virus and consensus sequences for subsequent phylogenetic analysis does offer a snapshot of the most prominent circulating viruses. Clonal sequencing of individual cells would offer a more fine-grained portrait of the diversity within and between compartments, and we hope to apply this approach in future studies. As an intermediate approach, we aim to use the bulk sequencing data on heterogeneity and minority variants in addition to consensus sequences to examine the emergence of new haplotypes and fixed mutations in the total viral population.

In the absence of a controlling immune response or effective antiretroviral therapy, the functional proviral DNA sequences closely resemble the cell-associated RNA sequences. We do detect a significant amount of hypermutated sequences from the proviral DNA sequencing – although it is not quantitative and cannot tell us which percentage of proviral DNA sequences are hypermutated, it illustrates the bottleneck between integrating the nascent viral DNA into the cellular genome and later expression of viral RNAs. Removing these hypermutated sequences from the phylogenetic analysis was a critical step in assessing true viral evolution rates, as including them would assume falsely high mutation rates within the virus and skew interpretations of the phylogenetic trees.

One of the questions we aimed to answer was whether virus was compartmentalized within central memory T cell subsets, or if other factors, such as time and anatomical location, would have a greater effect of the evolutionary distance between viruses. In comparing the consensus infection stock sequence and the viruses circulating within the lymph nodes, the data suggest that in active infection, the greatest distinction between viral samples is correlated with time, not with T cell subset, or the type of nucleic acid sequenced. This implies that, in the absence of natural control of infection or cART and over the period studied here, viruses from early in infection do not continue to circulate in lymph nodes at high frequencies. Although T cells located in the B cell follicle have the highest levels of infection, the viruses that they contain are not substantially different from the viruses in cells outside the follicle at a given time point. No one population of cells was closest to plasma

virus, which is not surprising given than the plasma sequence represents a melange of viruses originating from throughout the blood and other lymphoid organs.

In examining the patterns of viral evolution, two of the animals studied here show a typical intra-host, directed evolutionary pattern with a strong backbone that is reminiscent of the phylogenies observed in influenza. There is a clear and linear relationship between successive clades of viruses that are organized in a time-dependent fashion, and there are no large gaps or jumps in virus evolution between time points, tissue types, or nucleic acid. This indicates that in these animals, the majority of viruses are evolving steadily away from immune pressures. In contrast, the third animal exhibits a different pattern of intra-host viral evolution that more closely resembles inter-host HIV phylogeny with a star-like pattern and no single pathway that dominates the viral evolution. In this animal, virus appears to test several branches and possible clades but has not settled into a single dominant clade. Virus evolution is not directed or restricted to a single cluster, and the sequences from various populations appear to use more of the total theoretical sequence space that the virus can sample to escape host immune pressures. These interpretations rely on accurate and sufficient sampling of the representative sequences within each host – while we sequence multiple central memory T cell subsets, we do not investigate effector memory or stem cell memory cells that harbour virus, albeit typically at lower levels than central memory CD4 T cells.

In the gross measures of plasma neutralization, quantity of IgG and IgM SIV-specific B cells, and levels of VH mutation or CDR3 length in the repertoire, there is not a single correlate that predicts the pattern of viral evolution. It is tempting to speculate that some combination of these or other factors (such as CTL) are the driving forces behind the two patterns of evolution, with stronger immune responses forcing an animal to sample more of the sequencing space and weaker responses reflected in a strong backbone, flu-like evolution. However, studies that follow the animals for longer periods of time or examine in greater detail the immunological pressures on the virus would offer additional insight into the factors driving the patterns of viral evolution. The progression to AIDS-like symptoms is more rapid in macaques than in humans, with animals typically developing severe immunodeficiency within two years of infection. Six months of SIV infection is sufficient to recapitulate the beginning of chronic infection in humans, but following animals for longer periods would further show the continual evolution of both the SIV-specific immunoglobulin repertoire and tissue-resident virus. We did not measure CTL in this study, although we do control for MHC alleles known to be associated with control (none of the animals in this study are

MamuA*01). While SIVmac251 infection is more difficult to neutralize than SIVsmE660, we did not see broad and potent neutralization in the animals in this study. Many macaques fail to ever produce bnAbs when infected with SIV, and the use of SHIV, viruses with SIV genetic backbones but HIV envelopes, would offer a useful comparison of antibody evolution.

This study uses samples from animals with uncontrolled viral infection to examine the preferential infection of T cells located in the B cell follicle and the co-evolution of virus and antibodies. Investigating the same phenomena, in macaques that either naturally control infection (such as Mamu A*01 animals) or in animals that are treated with cART, would yield insights into how reduced inflammation and immune control affect the evolution of antibodies and localization of infected cells. In particular, they can answer to what extent the intra-host viral evolutionary patterns we see here are reflected in the debate over whether there is on-going evolution of virus in lymph nodes during suppressive treatment.

This study of the development of the antigen-specific B cell repertoire and the preferential infection of follicular T cells helps inform our understanding of the immunological events of acute infection in secondary lymphoid organs. This is critical to building on our understanding of HIV immunology from studies in peripheral blood in HIV-infected individuals. Both B cell lineage immunogen vaccine strategies and studies of the latent reservoir rely on understanding of the perturbations in B cell-TFH interactions during HIV infection, and the consequences of preferential infection of follicular T cells. In this work, I expand our understanding of early B cell ontogenesis in chronic viral infection, and help to advance the cure strategies by identifying major compartments of HIV/SIV replication and viral evolution within those compartments.

7 References

1. Barre-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (80-.).* **220,** 868–871 (1983).
2. Gallo, R. *et al.* Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science (80-.).* **224,** 500–503 (1984).
3. Levy, J. A. HIV pathogenesis: 25 years of progress and persistent challenges. *Aids* **23,** 147–160 (2009).
4. Rambaut, A., Posada, D., Crandall, K. A. & Holmes, E. C. The causes and consequences of HIV evolution. *Nat. Rev. Genet.* **5,** 52–61 (2004).
5. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb Perspect Med* **1,** (2011).
6. Gao, F. *et al.* Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397,** 436–441 (1999).
7. Lemey, P. *et al.* Tracing the origin and history of the HIV-2 epidemic. *Proc. Natl. Acad. Sci.* **100,** 6588–6592 (2003).
8. Martin Stoltzfus, C. Chapter 1 Regulation of HIV-1 Alternative RNA Splicing and Its Role in Virus Replication. *Advances in Virus Research* **74,** 1–40 (2009).
9. Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strains. *Science (80-.).* **288,** 1789–1796 (2000).
10. Lynch, R. M., Shen, T., Gnanakaran, S. & Derdeyn, C. A. Appreciating HIV type 1 diversity: subtype differences in Env. *AIDS Res. Hum. Retroviruses* **25,** 237–248 (2009).
11. Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20,** W13–W23 (2006).
12. Gale, C. V, Myers, R., Tedder, R. S., Williams, I. G. & Kellam, P. Development of a novel human immunodeficiency virus type 1 subtyping tool, Subtype Analyzer (STAR): analysis of subtype distribution in London. *AIDS Res. Hum. Retroviruses* **20,** 457–64 (2004).
13. Faria, N. R. *et al.* HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science (80-.).* **346,** 56–61 (2014).
14. Ariën, K. K., Vanham, G. & Arts, E. J. Is HIV-1 evolving to a less virulent form in humans? *Nat. Rev. Microbiol.* **5,** 141–151 (2007).
15. Foley, B. *et al.* HIV Sequence Compendium 2010 Editors. *Analysis* (2010).
16. Maddon, P. J. *et al.* The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47,** 333–348 (1986).
17. Mcdougal, J. S., Kennedy, J. M., Sligh, S. P., Cort, A. & Mawle, J. K. A. Binding of HTLV-III/LAV to T4 T Cells by a Complex of the 110K Viral Protein and the T4 Molecule. *Science (80-.).* **231,** 382–385 (1986).
18. Choe, H. *et al.* The B-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85,** 1135–1148 (1996).
19. Deng, H. *et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381,** 661–666 (1996).
20. Dragic, T. *et al.* HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381,** 667–673 (1996).
21. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272,** 872–877 (1996).
22. Grivel, J. C. & Margolis, L. B. CCR5- and CXCR4-tropic HIV-1 are equally cytopathic for their T-cell targets in human lymphoid tissue. *Nat. Med.* **5,** 344–346 (1999).
23. Douek, D. C., Picker, L. J. & Koup, R. A. T Cell Dynamics in HIV-1 Infection. *Annu. Rev. Immunol.* **21,** 265–304 (2003).
24. Esté, J. a *et al.* Shift of clinical human immunodeficiency virus type 1 isolates from X4 to R5 and prevention of emergence of the syncytium-inducing phenotype by blockade of CXCR4. *J. Virol.* **73,** 5577–85 (1999).
25. Caputi, M. *RNA Processing*. (InTech, 2011). doi:10.5772/832
26. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7552–7 (2008).
27. Abrahams, M.-R. *et al.* Quantitating the multiplicity of infection with human

immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* **83,** 3556–67 (2009).
28. McMichael, A. J., Borrow, P., Tomaras, G. D., Goonetilleke, N. & Haynes, B. F. The immune response during acute HIV-1 infection: clues for vaccine development. *Nat Rev Immunol* **10,** 11–23 (2010).
29. Wiley, C. A., Schrier, R. D., Nelson, J. A., Lampert, P. W. & Oldstone, M. B. Cellular localization of human immunodeficiency virus infection within the brains of acquired immune deficiency syndrome patients. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 7089–93 (1986).
30. Wu, L. & Kewal Ramani, V. N. Dendritic-cell interactions with HIV: infection and viral dissemination. *Nat. Rev. Immunol.* **6,** 859–68 (2006).
31. Luciw, P. A., Pratt-Lowe, E., Shaw, K. E., Levy, J. A. & Cheng-Mayer, C. Persistent infection of rhesus macaques with T-cell-line-tropic and macrophage-tropic clones of simian/human immunodeficiency viruses (SHIV). *Proc. Natl. Acad. Sci. U. S. A.* **92,** 7490–4 (1995).
32. Pope, M. *et al.* Conjugates of dendritic cells and memory T lymphocytes from skin facilitate productive infection with HIV-1. *Cell* **78,** 389–398 (1994).
33. Brenchley, J. M. *et al.* CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J. Exp. Med.* **200,** 749–59 (2004).
34. Veazey, R. S. *et al.* Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. *Science (80-.).* **280,** 427–431 (1998).
35. Wang, L. *et al.* A novel mechanism of CD4 lymphocyte depletion involves effects of HIV on resting lymphocytes: induction of lymph node homing and apoptosis upon secondary signaling through homing receptors. *J. Immunol.* **162,** 268–76 (1999).
36. Mattapallil, J. J. *et al.* Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature* **434,** 1093–1097 (2005).
37. Engelman, A. & Cherepanov, P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat. Rev. Microbiol.* **10,** 279–90 (2012).
38. Permanyer, M., Ballana, E. & Este, J. A. Endocytosis of HIV: Anything goes. *Trends Microbiol.* **18,** 543–551 (2010).
39. Chan, D. C. & Kim, P. S. HIV entry and its inhibition. *Cell* **93,** 681–684 (1998).
40. Gomez, C. & Hope, T. J. The ins and outs of HIV replication. *Cell. Microbiol.* **7,** 621–626 (2005).
41. Gallay, P., Hope, T., Chin, D. & Trono, D. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 9825–30 (1997).
42. Bukrinsky, M. I. *et al.* A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* **365,** 666–669 (1993).
43. Haffar, O. K. *et al.* Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. *J. Mol. Biol.* **299,** 359–68 (2000).
44. Jordan, A., Defechereux, P. & Verdin, E. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J.* **20,** 1726–1738 (2001).
45. Schröder, a. R. W. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110,** 521–529 (2002).
46. Preston, B. D., Poiesz, B. J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase. *Science (80-.).* **242,** 1168–1171 (1988).
47. Roberts, J. D., Bebenek, K. & Kunkel, T. a. The accuracy of reverse transcriptase from HIV-1. *Science* **242,** 1171–1173 (1988).
48. Pollard, V. W. & Malim, M. H. The HIV-1 Rev Protein: Overview of the Retroviral Life Cycle. *Ann* **52,** 491–532 (1998).
49. Sodroski, J. *et al.* A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature* **321,** 412–417 (1986).
50. Malim, M. H. *et al.* HIV-1 structural gene expression requires binding of the rev trans-activator to its RNA target sequence. *Cell* **60,** 675–683 (1990).
51. Kirchhoff, F. Is the high virulence of HIV-1 an unfortunate coincidence of primate lentiviral evolution? *Nat. Rev. Microbiol.* **7,** 467–476 (2009).
52. Landi, A., Iannucci, V., Van Nuffel, A., Meuwissen, P. & Verhasselt, B. One Protein to Rule them All: Modulation of Cell Surface Receptors and Molecules by HIV Nef. *Curr. HIV Res.* **9,**

496–504 (2011).
53. Heaphy, S. *et al.* HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell* **60,** 685–693 (1990).
54. Fischer, U., Huber, J., Boelens, W. C., Mattajt, L. W. & Luhrmann, R. The HIV-1 Rev Activation Domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82,** 475–483 (1995).
55. Emerman, M. HIV-1 Regulatory/Accessory Genes: Keys to Unraveling Viral and Host Cell Biology. *Science (80-.).* **280,** 1880–1884 (1998).
56. Peterlin, B. M. & Trono, D. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nat. Rev. Immunol.* **3,** 97–107 (2003).
57. Schwartz, O., Maréchal, V., Le Gall, S., Lemonnier, F. & Heard, J. M. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat. Med.* **2,** 338–42 (1996).
58. Bubeník, J. MHC class I down-regulation: tumour escape from immune surveillance? *Int. J. Oncol.* **25,** 487–91 (2004).
59. Choremi-Papadopoulou, H. *et al.* Downregulation of CD28 surface antigen on CD4+ and CD8+ T lymphocytes during HIV-1 infection. *J Acquir Immune Defic Syndr* **7,** 245–253 (1994).
60. Münch, J. *et al.* T-cell receptor:CD3 down-regulation is a selected in vivo function of simian immunodeficiency virus Nef but is not sufficient for effective viral replication in rhesus macaques. *J. Virol.* **76,** 12360–4 (2002).
61. Mariani, R. *et al.* High frequency of defective nef alleles in a long-term survivor with nonprogressive human immunodeficiency virus type 1 infection. *J. Virol.* **70,** 7752–64 (1996).
62. Chowers, M. Y. *et al.* Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene. *J. Virol.* **68,** 2906–14 (1994).
63. Daniel, M. D., Kirchhoff, F., Czajak, S. C., Sehgal, P. K. & Desrosiers, R. C. Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene. *Science* **258,** 1938–1941 (1992).
64. Kirchhoff, F., Greenough, T. C., Brettler, D. B., Sullivan, J. L. & Desrosiers, R. C. Absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. *N. Engl. J. Med.* **332,** 228–232 (1995).
65. Malim, M. H. & Bieniasz, P. D. HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb. Perspect. Med.* **2,** a006940 (2012).
66. Doyle, T., Goujon, C. & Malim, M. H. HIV-1 and interferons: who's interfering with whom? *Nat. Rev. Microbiol.* **13,** 403–413 (2015).
67. Freed, E. O. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* **13,** 484–496 (2015).
68. Checkley, M. A., Luttge, B. G. & Freed, E. O. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J. Mol. Biol.* **410,** 582–608 (2011).
69. Brady, T. *et al.* HIV integration site distributions in resting and activated CD4+ T cells infected in culture. *AIDS* **23,** 1461–1471 (2010).
70. Coiras, M., Lopez-Huertas, M. R., Perez-Olmeda, M. & Alcami, J. Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. *Nat Rev Microbiol* **7,** 798–812 (2009).
71. Pomerantz, R. J. & Horn, D. L. Twenty years of therapy for HIV-1 infection. *Nat. Med.* **9,** 867–873 (2003).
72. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387,** 183–188 (1997).
73. Bullen, C. K., Laird, G. M., Durand, C. M., Siliciano, J. D. & Siliciano, R. F. New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. *Nat. Med.* **20,** 425–9 (2014).
74. Chun, T. W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 13193–7 (1997).
75. Finzi, D. *et al.* 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 (1999).
76. Siliciano, J. M. & Siliciano, R. F. The remarkable stability of the latent reservoir for HIV-1 in resting memory CD4+ T cells. *J. Infect. Dis.* **212,** 1345–1347 (2015).

77. Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15,** 893–900 (2009).
78. Lorenzo-Redondo, R. *et al.* Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* **530,** 51–56 (2016).
79. Persaud, D. *et al.* Continued production of drug-sensitive human immunodeficiency virus type 1 in children on combination antiretroviral therapy who have undetectable viral loads. *J. Virol.* **78,** 968–79 (2004).
80. Ruff, C. T. *et al.* 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,** 9481–92 (2002).
81. Cohen, M. S. Preventing sexual transmission of HIV. *Clin. Infect. Dis.* **45 Suppl 4,** S287-92 (2007).
82. Quinn, T. C. *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N. Engl. J. Med.* **342,** 921–9 (2000).
83. Ward, H. & Rönn, M. The contribution of STIs to the sexual transmission of HIV. *Curr. Opin. HIV AIDS* **5,** 305–310 (2010).
84. Auvert, B. *et al.* Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: The ANRS 1265 trial. *PLoS Med.* **2,** 1112–1122 (2005).
85. Curran, J. W. *et al.* Epidemiology of HIV infection and AIDS in the United States. *Science (80-.).* **239,** 610–616 (1988).
86. Cooper, E. R. *et al.* Combination antiretroviral strategies for the treatment of pregnant HIV-1-infected women and prevention of perinatal HIV-1 transmission. *J. Acquir. Immune Defic. Syndr.* **29,** 484–94 (2002).
87. Semba, R. D. *et al.* Maternal vitamin A deficiency and mother-to-child transmission of HIV-1. *Lancet* **343,** 1593–7 (1994).
88. Coovadia, H. M. *et al.* Mother-to-child transmission of HIV-1 infection during exclusive breastfeeding in the first 6 months of life: an intervention cohort study. *Lancet* **369,** 1107–1116 (2007).
89. Fox, J. & Fidler, S. Sexual transmission of HIV-1. *Antiviral Research* **85,** 276–285 (2010).
90. Mastro, T. D. & de Vincenzi, I. Probabilities of sexual HIV-1 transmission. *Aids* **10,** 75–82 (1996).
91. Fiebig, E. W. *et al.* Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS* **17,** 1871–1879 (2003).
92. Busch, M. P. & Satten, G. A. Time course of viremia and antibody seroconversion following human immunodeficiency virus exposure. in *American Journal of Medicine* **102,** 117–124 (1997).
93. Pasternak, A. O., Lukashov, V. V & Berkhout, B. Cell-associated HIV RNA: a dynamic biomarker of viral persistence. *Retrovirology* **10,** 41 (2013).
94. Veazey, R. S. *et al.* Identifying the target cell in primary simian immunodeficiency virus (SIV) infection: highly activated memory CD4(+) T cells are rapidly eliminated in early SIV infection in vivo. *J. Virol.* **74,** 57–64 (2000).
95. Veazey, R. S. *et al.* Dynamics of CCR5 expression by CD4(+) T cells in lymphoid tissues during simian immunodeficiency virus infection. *J. Virol.* **74,** 11001–7 (2000).
96. Grossman, Z., Meier-Schellersheim, M., Paul, W. E. & Picker, L. J. Pathogenesis of HIV infection: what the virus spares is as important as what it destroys. *Nat. Med.* **12,** 289–295 (2006).
97. Deeks, S. G., Tracy, R. & Douek, D. C. Systemic Effects of Inflammation on Health during Chronic HIV Infection. *Immunity* **39,** 633–45 (2013).
98. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496,** 469–76 (2013).
99. Wu, X. *et al.* Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333,** 1593–602 (2011).
100. Lempicki, R. a *et al.* Impact of HIV-1 infection and highly active antiretroviral therapy on the kinetics of CD4+ and CD8+ T cell turnover in HIV-infected patients. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13778–13783 (2000).
101. Douek, D. C. Immune activation, HIV persistence, and the cure. *Top. Antivir. Med.* **21,** 128–32 (2013).

102. Lifson, A. R. *et al.* Long-term human immunodeficiency virus infection in asymptomatic homosexual and bisexual men with normal CD4+ lymphocyte counts: immunologic and virologic characteristics. *J. Infect. Dis.* **163,** 959–65 (1991).
103. Deeks, S. G. & Walker, B. D. Human Immunodeficiency Virus Controllers: Mechanisms of Durable Virus Control in the Absence of Antiretroviral Therapy. *Immunity* **27,** 406–416 (2007).
104. Pereyra, F. *et al.* Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *J. Infect. Dis.* **197,** 563–571 (2008).
105. Migueles, S. A. *et al.* HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2709–14 (2000).
106. Walker, B. D. & Yu, X. G. Unravelling the mechanisms of durable control of HIV-1. *Nat. Rev. Immunol.* **13,** 487–498 (2013).
107. Okoye, A. A. & Picker, L. J. CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunol. Rev.* **254,** 54–64 (2013).
108. Sheppard, H. W. & Ascher, M. S. The Natural History and Pathogenesis of HIV Infection. *Annu. Rev. Microbiol.* **46,** 533–64 (1992).
109. Biancotto, A. *et al.* Abnormal activation and cytokine spectra in lymph nodes of persons chronically infected with HIV-1. **109,** 1–31 (2007).
110. Schacker, T. W. *et al.* Collagen deposition in HIV-1 infected lymphatic tissues and T cell homeostasis Rapid Publication. **110,** 1133–1139
111. Zeng, M. *et al.* Cumulative mechanisms of lymphoid tissue fibrosis and T cell depletion in HIV-1 and SIV infections. *J. Clin. Invest.* **121,** 998–1008 (2011).
112. Colineau, L. *et al.* HIV-Infected Spleens Present Altered Follicular Helper T Cell (Tfh) Subsets and Skewed B Cell Maturation. *PLoS One* **10,** e0140978 (2015).
113. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271,** 1582–1586 (1996).
114. Araújo, L. A. L. & Almeida, S. E. M. HIV-1 diversity in the envelope glycoproteins: Implications for viral entry inhibition. *Viruses* **5,** 595–604 (2013).
115. Williamson, S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. *Mol. Biol. Evol.* **20,** 1318–25 (2003).
116. Zhuang, J. *et al.* Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J. Virol.* **76,** 11273–82 (2002).
117. Brenchley, J. M., Price, D. A. & Douek, D. C. HIV disease: fallout from a mucosal catastrophe? *Nat Immunol* **7,** 235–239 (2006).
118. Gonzalez, V. D., Landay, A. L. & Sandberg, J. K. Innate immunity and chronic immune activation in HCV/HIV-1 co-infection. *Clinical Immunology* **135,** 12–25 (2010).
119. Sandler, N. G. *et al.* Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression. *Nature* **511,** 601–605 (2014).
120. Mandl, J. N. *et al.* Divergent TLR7 and TLR9 signaling and type I interferon production distinguish pathogenic and nonpathogenic AIDS virus infections. *Nat Med* **14,** 1077–1087 (2008).
121. Sandler, N. G. *et al.* Plasma levels of soluble CD14 independently predict mortality in HIV infection. *J. Infect. Dis.* **203,** 780–790 (2011).
122. Vargas-Inchaustegui, D. a, Xiao, P., Tuero, I., Patterson, L. J. & Robert-Guroff, M. NK and CD4+ T cell cooperative immune responses correlate with control of disease in a macaque simian immunodeficiency virus infection model. *J. Immunol.* **189,** 1878–85 (2012).
123. Gonzalez, V. D., Landay, A. L. & Sandberg, J. K. Innate immunity and chronic immune activation in HCV/HIV-1 co-infection. *Clin. Immunol.* **135,** 12–25 (2010).
124. Reeves, R. K. *et al.* Antigen-specific NK cell memory in rhesus macaques. *Nat. Immunol.* **16,** 927–932 (2015).
125. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–650 (2002).
126. Donahue, J. P., Vetter, M. L., Mukhtar, N. A. & D'Aquila, R. T. The HIV-1 Vif PPLP motif is necessary for human APOBEC3G binding and degradation. *Virology* **377,** 49–53 (2008).

127. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–430 (2008).
128. Stremlau, M. *et al.* The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427,** 848–853 (2004).
129. Yap, M. W., Nisole, S., Lynch, C. & Stoye, J. P. Trim5alpha protein restricts both HIV-1 and murine leukemia virus. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10786–91 (2004).
130. Tomaras, G. D. *et al.* Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J. Virol.* **82,** 12449–12463 (2008).
131. Gray, E. S. *et al.* Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. *J. Virol.* **81,** 6187–6196 (2007).
132. Stamatatos, L., Morris, L., Burton, D. R. & Mascola, J. R. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nat. Med.* **15,** 866–870 (2009).
133. Richman, D. D., Wrin, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 4144–9 (2003).
134. Crotty, S. T Follicular Helper Cell Differentiation, Function, and Roles in Disease. *Immunity* **41,** 529–542 (2014).
135. Ma, C. S., Deenick, E. K., Batten, M. & Tangye, S. G. The origins, function, and regulation of T follicular helper cells. *J. Exp. Med.* **209,** 1241–1253 (2012).
136. Miller, J. F. A. P., De Burgh, P. M. & Grant, G. A. Thymus and the Production of Antibody-plaque-forming Cells. *Nature* **208,** 1332–1334 (1965).
137. Breitfeld, D. *et al.* Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production. *J. Exp. Med.* **192,** 1545–52 (2000).
138. Schaerli, P. *et al.* CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function. *J. Exp. Med.* **192,** 1553–62 (2000).
139. Dobner, T., Wolf, I., Emrich, T. & Lipp, M. Differentiation specific expression of a novel G protein coupled receptor from Burkitt's lymphoma. *Eur. J. Immunol.* **22,** 2795–2799 (1992).
140. Legler, D. F. *et al.* B cell-attracting chemokine 1, a human CXC chemokine expressed in lymphoid tissues, selectively attracts B lymphocytes via BLR1/CXCR5. *J. Exp. Med.* **187,** 655–60 (1998).
141. Rock, K. L. & Shen, L. Cross-presentation: underlying mechanisms and role in immune surveillance. *Immunol Rev* **207,** 166–183 (2005).
142. Hardtke, S., Ohl, L. & Förster, R. Balanced expression of CXCR5 and CCR7 on follicular T helper cells determines their transient positioning to lymph node follicles and is essential for efficient B-cell help. *Blood* **106,** 1924–1931 (2005).
143. Linterman, M. A. *et al.* IL-21 acts directly on B cells to regulate Bcl-6 expression and germinal center responses. *J. Exp. Med.* **207,** 353–63 (2010).
144. Zotos, D. *et al.* IL-21 regulates germinal center B cell differentiation and proliferation through a B cell-intrinsic mechanism. *J. Exp. Med.* **207,** 365–378 (2010).
145. Lu, K. T. *et al.* Functional and Epigenetic Studies Reveal Multistep Differentiation and Plasticity of In Vitro-Generated and In Vivo-Derived Follicular T Helper Cells. *Immunity* **35,** 622–632 (2011).
146. Li, H. & Pauza, C. D. CD25 $^{+}$ Bcl6 ^{low} T follicular helper cells provide help to maturing B cells in germinal centers of human tonsil: Molecular immunology. *Eur. J. Immunol.* **45,** 298–308 (2015).
147. Rasheed, A. U., Rahn, H. P., Sallusto, F., Lipp, M. & Muller, G. Follicular B helper T cell activity is confined to CXCR5hiICOShi CD4 T cells and is independent of CD57 expression. *Eur. J. Immunol.* **36,** 1892–1903 (2006).
148. Chtanova, T. *et al.* T Follicular Helper Cells Express a Distinctive Transcriptional Profile, Reflecting Their Role as Non-Th1/Th2 Effector Cells That Provide Help for B Cells. *J. Immunol.* **173,** 68–78 (2004).
149. Onabajo, O. O., George, J., Lewis, M. G. & Mattapallil, J. J. Rhesus Macaque Lymph Node PD-1hiCD4+ T Cells Express High Levels of CXCR5 and IL-21 and Display a CCR7loICOS+Bcl6+ T-Follicular Helper (Tfh) Cell Phenotype. *PLoS One* **8,** 2–9 (2013).
150. Bryant, V. L. *et al.* Cytokine-mediated regulation of human B cell differentiation into Ig-secreting cells: predominant role of IL-21 produced by CXCR5+ T follicular helper cells. *J.*

Immunol. **179,** 8180–8190 (2007).
151. Chtanova, T. *et al.* T Follicular Helper Cells Express a Distinctive Transcriptional Profile, Reflecting Their Role as Non-Th1/Th2 Effector Cells That Provide Help for B Cells. *J. Immunol.* **173,** 68–78 (2004).
152. Johnston, R. J. *et al.* Bcl6 and Blimp-1 Are Reciprocal and Antagonistic Regulators of T Follicular Helper Cell Differentiation. *Cell Differ.* **325,** 1006–1010 (2010).
153. Kitano, M. *et al.* Bcl6 Protein Expression Shapes Pre-Germinal Center B Cell Dynamics and Follicular Helper T Cell Heterogeneity. *Immunity* **34,** 961–972 (2011).
154. Bauquet, A. T. *et al.* The costimulatory molecule ICOS regulates the expression of c-Maf and IL-21 in the development of follicular T helper cells and TH-17 cells. *Nat. Immunol.* **10,** 167–175 (2009).
155. Johnston, R. J., Choi, Y. S., Diamond, J. a., Yang, J. a. & Crotty, S. STAT5 is a potent negative regulator of TFH cell differentiation. *J. Exp. Med.* **209,** 243–250 (2012).
156. Ballesteros-Tato, A. *et al.* Interleukin-2 Inhibits Germinal Center Formation by Limiting T Follicular Helper Cell Differentiation. *Immunity* **36,** 847–856 (2012).
157. Tubo, N. J. *et al.* Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell* **153,** 785–796 (2013).
158. Choi, Y. S. *et al.* ICOS Receptor Instructs T Follicular Helper Cell versus Effector Cell Differentiation via Induction of the Transcriptional Repressor Bcl6. *Immunity* **34,** 932–946 (2011).
159. Kerfoot, S. M. *et al.* Germinal Center B Cell and T Follicular Helper Cell Development Initiates in the Interfollicular Zone. *Immunity* **34,** 947–960 (2011).
160. Yusuf, I. *et al.* Germinal center T follicular helper cell IL-4 production is dependent on signaling lymphocytic activation molecule receptor (CD150). *J. Immunol.* **185,** 190–202 (2010).
161. Linterman, M. A. *et al.* Follicular helper T cells are required for systemic autoimmunity. *J. Exp. Med.* **206,** 561–76 (2009).
162. Linterman, M. A. *et al.* Foxp3+ follicular regulatory T cells control the germinal center response. *Nat. Med.* **17,** 975–82 (2011).
163. Chung, Y. *et al.* Follicular regulatory T (Tfr) cells with dual Foxp3 and Bcl6 expression suppress germinal center reactions. *Nat. Med.* **17,** 983–988 (2011).
164. Sage, P. T., Francisco, L. M., Carman, C. V & Sharpe, A. H. The receptor PD-1 controls follicular regulatory T cells in the lymph nodes and blood. *Nat. Immunol.* **14,** 152–61 (2013).
165. Ding, Y. *et al.* Interleukin-21 promotes germinal center reaction by skewing the follicular regulatory T cell to follicular helper T cell balance in autoimmune BXD2 mice. *Arthritis Rheumatol.* **66,** 2601–2612 (2014).
166. Wang, C. J. *et al.* CTLA-4 controls follicular helper T-cell differentiation by regulating the strength of CD28 engagement. *Proc Natl Acad Sci U S A* **112,** 524–529 (2015).
167. Sage, P. T. & Sharpe, A. H. T follicular regulatory cells. *Immunol. Rev.* **271,** 246–259 (2016).
168. Nutt, S. L. & Tarlinton, D. M. Germinal center B and follicular helper T cells: siblings, cousins or just good friends? *Nat. Immunol.* **131,** 472–477 (2011).
169. Lindqvist, M. *et al.* Expansion of HIV-speci c T follicular helper cells in chronic HIV infection. **122,** (2012).
170. Petrovas, C. *et al.* CD4 T follicular helper cell dynamics during SIV infection. *J. Clin. Invest.* **122,** 3281–94 (2012).
171. Hong, J. J., Amancha, P. K., Rogers, K., Ansari, A. A. & Villinger, F. Spatial Alterations between CD4+ T Follicular Helper, B, and CD8+ T Cells during Simian Immunodeficiency Virus Infection: T/B Cell Homeostasis, Activation, and Potential Mechanism for Viral Escape. *J. Immunol.* **188,** 3247–3256 (2012).
172. Perreau, M. *et al.* Follicular helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and production. *J. Exp. Med.* **210,** 143–56 (2013).
173. Cubas, R. a *et al.* Inadequate T follicular cell help impairs B cell immunity during HIV infection. *Nat. Med.* **19,** 494–9 (2013).
174. Fukazawa, Y. *et al.* B cell follicle sanctuary permits persistent productive simian immunodeficiency virus infection in elite controllers. *Nat. Med.* **21,** 132–139 (2015).
175. Simpson, N. *et al.* Expansion of circulating T cells resembling follicular helper T cells is a fixed phenotype that identifies a subset of severe systemic lupus erythematosus. *Arthritis*

Rheum. **62,** 234–44 (2010).
176. Forcade, E. *et al.* Circulating T follicular helper cells with increased function during chronic graft-versus-host-disease. *Blood* (2016). doi:10.1182/blood-2015-12-688895
177. Boswell, K. L. *et al.* Loss of Circulating CD4 T Cells with B Cell Helper Function during Chronic HIV Infection. *PLoS Pathog.* **10,** e1003853 (2014).
178. Morita, R. *et al.* Human Blood CXCR5+CD4+ T Cells Are Counterparts of T Follicular Cells and Contain Specific Subsets that Differentially Support Antibody Secretion. *Immunity* **34,** 108–121 (2011).
179. Bentebibel, S. *et al.* Induction of ICOS+CXCR3+CXCR5+ TH cells correlates with antibody responses to influenza vaccination. *Sci. Transl. Med.* **5,** 176ra32 (2013).
180. Locci, M. *et al.* Human Circulating PD-1+CXCR3−CXCR5+ Memory Tfh Cells Are Highly Functional and Correlate with Broadly Neutralizing HIV Antibody Responses. *Immunity* **39,** 758–769 (2013).
181. Schultz, B. T. *et al.* Circulating HIV-Specific Interleukin-21+CD4+ T Cells Represent Peripheral Tfh Cells with Antigen-Dependent Helper Functions. *Immunity* **44,** 167–178 (2016).
182. Pallikkuth, S. *et al.* Peripheral T Follicular Helper Cells Are the Major HIV Reservoir Within Central Memory CD4 T Cells in Peripheral Blood from chronic HIV infected individuals on cART. *J. Virol.* **90,** JVI.02883-15 (2015).
183. Cubas, R. *et al.* Reversible Reprogramming of Circulating Memory T Follicular Helper Cell Function during Chronic HIV Infection. *J. Immunol.* (2015). doi:10.4049/jimmunol.1501524
184. Allen, C. D. C., Okada, T., Tang, H. L. & Cyster, J. G. Imaging of germinal center selection events during affinity maturation. *Science* **315,** 528–531 (2007).
185. Victora, G. D. *et al.* Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* **143,** 592–605 (2010).
186. Gitlin, A. D., Shulman, Z. & Nussenzweig, M. C. Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature* **509,** 637–40 (2014).
187. Shulman, Z. *et al.* Dynamic signaling by T follicular helper cells during germinal center B cell selection. *Science* **345,** 1058–1062 (2014).
188. Cyster, J. G. B cell follicles and antigen encounters of the third kind. *Nat. Immunol.* **11,** 989–996 (2010).
189. El Shikh, M. E. M. & Pitzalis, C. Follicular dendritic cells in health and disease. *Frontiers in Immunology* **3,** (2012).
190. Tarlinton, D. M. & Smith, K. G. C. Dissecting affinity maturation: A model explaining selection of antibody-forming cells and memory B cells in the germinal centre. *Immunology Today* **21,** 436–441 (2000).
191. Jackson, K. J. L., Kidd, M. J., Wang, Y. & Collins, A. M. The Shape of the Lymphocyte Receptor Repertoire: Lessons from the B Cell Receptor. *Front. Immunol.* **4,** 1–12 (2013).
192. Sundling, C., Phad, G., Douagi, I., Navis, M. & Karlsson Hedestam, G. B. Isolation of antibody V(D)J sequences from single cell sorted rhesus macaque B cells. *J. Immunol. Methods* **386,** 85–93 (2012).
193. Karlsson Hedestam, G. B. *et al.* The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat. Rev. Microbiol.* **6,** 143–155 (2008).
194. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32,** (2014).
195. Lefranc, M. P. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Research* **31,** 307–310 (2003).
196. Lefranc, M. P. in *Immunoinformatics* 1–18 (2008). doi:10.1007/978-0-387-72968-8_1
197. Brochet, X., Lefranc, M. P. & Giudicelli, V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* **36,** (2008).
198. Schatz, D. G. & Swanson, P. C. V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45,** 167–202 (2011).
199. Benedict, C. L., Gilfillan, S., Thai, T. H. & Kearney, J. F. Terminal deoxynucleotidyl transferase and repertoire development. *Immunol Rev* **175,** 150–157 (2000).
200. Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. in *Immunobiology 5* 892 (2001). doi:10.1111/j.1467-2494.1995.tb00120.x
201. Gay, D., Saunders, T., Camper, S. & Weigert, M. Receptor editing: an approach by

autoreactive B cells to escape tolerance. *J. Exp. Med.* **177,** 999–1008 (1993).
202. Tiegs, S. L., Russell, D. M. & Nemazee, D. Receptor editing in self-reactive bone marrow B cells. *J. Exp. Med.* **177,** 1009–20 (1993).
203. Pelanda, R. & Torres, R. M. Receptor editing for better or for worse. *Curr. Opin. Immunol.* **18,** 184–190 (2006).
204. Kouskoff, V. & Nemazee, D. Role of receptor editing and revision in shaping the B and T lymphocyte repertoire. *Life Sciences* **69,** 1105–1113 (2001).
205. Kabat, E. A., Wu, T. Te, Foeller, C., Perry, H. M. & Gottesman, K. S. *Sequences of Proteins of Immunological Interest*. (DIANE Publishing, 1992). at <https://books.google.com/books?hl=en&lr=&id=3jMvZYW2ZtwC&pgis=1>
206. Link, J., Ivanov, I., Ippolito, G. & Schroeder, H. in *The Antibodies* 43–67 (CRC Press, 2002). doi:doi:10.1201/9780203216514.ch3
207. Corbett, S. J., Tomlinson, I. M., Sonnhammer, E. L. ., Buck, D. & Winter, G. Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, 'minor' D segments or D-D recombination. *J. Mol. Biol.* **270,** 587–597 (1997).
208. Tuaillon, N. & Capra, J. D. Use of D gene segments with irregular spacers in terminal deoxynucleotidyltransferase (TdT)+/+ and TdT-/- mice carrying a human Ig heavy chain transgenic minilocus. *Proc Natl Acad Sci U S A* **95,** 1703–1708 (1998).
209. Gu, H., Kitamura, D. & Rajewsky, K. B cell development regulated by gene rearrangement: arrest of maturation by membrane-bound D mu protein and selection of DH element reading frames. *Cell* **65,** 47–54 (1991).
210. Benichou, J. *et al.* The restricted DH gene reading frame usage in the expressed human antibody repertoire is selected based upon its amino acid content. *J. Immunol.* **190,** 5567–77 (2013).
211. Paul, W. E. *Fundamental Immunology*. *Fundamental Immunology* (2015).
212. Ippolito, G. C. *et al.* Forced usage of positively charged amino acids in immunoglobulin CDR-H3 impairs B cell development and antibody production. *J. Exp. Med.* **203,** 1567–78 (2006).
213. Xu, J. L. & Davis, M. M. Diversity in the CDR3 Region of VH Is Sufficient for Most Antibody Specificities. *Immunity* **13,** 37–45 (2000).
214. Hardy, R. R. Chapter 7: B Lymphocyte Development and Biology. *Fundam. Immunol.* 1–66 (2010).
215. Max, E. E. Chapter 6 Immunoglobulins : Molecular Genetics Overview of Immunoglobulin V Gene Assembly. *Genetics* 1–101 (2015).
216. McHeyzer-Williams, L. J., Milpied, P. J., Okitsu, S. L. & McHeyzer-Williams, M. G. Class-switched memory B cells remodel BCRs within secondary germinal centers. *Nat. Immunol.* **16,** 296–305 (2015).
217. Morbach, H., Eichhorn, E. M., Liese, J. G. & Girschick, H. J. Reference values for B cell subpopulations from infancy to adulthood. *Clin. Exp. Immunol.* **162,** 271–9 (2010).
218. Kwong, P. D. & Mascola, J. R. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* **37,** 412–425 (2012).
219. Kwong, P. D., Mascola, J. R. & Nabel, G. J. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nat. Rev. Immunol.* **13,** 693–701 (2013).
220. Buchacher, a *et al.* Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Res. Hum. Retroviruses* **10,** 359–69 (1994).
221. Liao, H.-X. *et al.* High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies. *J. Virol. Methods* **158,** 171–9 (2009).
222. Robinson, W. H. Sequencing the functional antibody repertoire--diagnostic and therapeutic discovery. *Nat. Rev. Rheumatol.* **11,** 171–82 (2015).
223. Montefiori, D. C. Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol. Biol.* **485,** 395–405 (2009).
224. Montefiori, D. C. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protoc. Immunol.* **Chapter 12,** Unit 12.11 (2005).
225. Seaman, M. S. *et al.* Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J Virol* **84,** 1439–1452 (2010).
226. Li, M. *et al.* Genetic and neutralization properties of subtype C human immunodeficiency virus

type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J. Virol.* **80,** 11776–90 (2006).
227. deCamp, A. *et al.* Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* **88,** 2489–2507 (2014).
228. Kwong, P. D. & Mascola, J. R. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* **37,** 412–25 (2012).
229. Zhu, J. *et al.* Somatic Populations of PGT135-137 HIV-1-Neutralizing Antibodies Identified by 454 Pyrosequencing and Bioinformatics. *Front Microbiol* **3,** 315 (2012).
230. Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329,** 856–61 (2010).
231. Zhou, T. *et al.* Structural Basis for Broad and Potent. *Science (80-.).* **329,** 811 (2010).
232. Larimore, K., McCormick, M. W., Robins, H. S. & Greenberg, P. D. Shaping of human germline IgH repertoires revealed by deep sequencing. *J Immunol* **189,** 3221–3230 (2012).
233. Haynes, B. F., Kelsoe, G., Harrison, S. C. & Kepler, T. B. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat. Biotechnol.* **30,** 423–33 (2012).
234. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480,** 336–43 (2011).
235. Alam, S. M. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Nature* **39,** 1315–6 (2013).
236. Alam, S. M. *et al.* Role of HIV membrane in neutralization by two broadly neutralizing antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 20234–9 (2009).
237. Letvin, N. L. *et al.* Immune and Genetic Correlates of Vaccine Protection Against Mucosal Infection by SIV in Monkeys. *Sci. Transl. Med.* **3,** 81ra36 (2011).
238. Moore, J. S. *et al.* Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS* **19,** 381–389 (2005).
239. Ackerman, M. E. *et al.* Natural variation in Fc glycosylation of HIV-specific antibodies impacts antiviral activity. *J. Clin. Invest.* **123,** 2183–92 (2013).
240. Banerjee, K. *et al.* IgG subclass profiles in infected HIV type 1 controllers and chronic progressors and in uninfected recipients of Env vaccines. *AIDS Res. Hum. Retroviruses* **26,** 445–58 (2010).
241. Alter, G. & Moody, M. A. The Humoral Response to HIV-1: New Insights, Renewed Focus. *J. Infect. Dis.* **202,** S315–S322 (2010).
242. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32,** 158–68 (2014).
243. DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* **31,** 166–169 (2013).
244. Baum, P. D., Venturi, V. & Price, D. A. Wrestling with the repertoire: the promise and perils of next generation sequencing for antigen receptors. *Eur. J. Immunol.* **42,** 2834–9 (2012).
245. Zhu, J. *et al.* De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E4088-97 (2013).
246. Tan, Y. C. *et al.* High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clin. Immunol.* **151,** (2014).
247. Scheid, J. F. *et al.* Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. *Science (80-.).* **333,** 1633–1637 (2011).
248. Sanders, R. W. *et al.* Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J. Virol.* **76,** 8875–89 (2002).
249. Sanders, R. W. *et al.* Variable-loop-deleted variants of the human immunodeficiency virus type 1 envelope glycoprotein can be stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits. *J. Virol.* **74,** 5091–5100 (2000).
250. The rgp120 HIV Vaccine Study Group. Placebo-Controlled Phase 3 Trial of a Recombinant Glycoprotein 120 Vaccine to Prevent HIV-1 Infection. *J. Infect. Dis.* **191,** 654–665 (2005).
251. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393,** 705–11 (1998).
252. Doria-Rose, N. A. *et al.* New Member of the V1V2-Directed CAP256-VRC26 Lineage That Shows Increased Breadth and Exceptional Potency. *J. Virol.* **90,** 76–91 (2015).

253. Stowell, R. E., Smith, E. K., Espana, C. & Nelson, V. G. Outbreak of malignant lymphoma in rhesus monkeys. *Lab. Invest.* **25,** 476–479 (1971).
254. Murphey-Corb, M. *et al.* Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys. *Nature* **321,** 435–7 (1986).
255. Apetrei, C., Robertson, D. L. & Marx, P. a. 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 (2004).
256. Lifson, J. D. & Haigwood, N. L. Lessons in nonhuman primate models for AIDS vaccine research: from minefields to milestones. *Cold Spring Harb Perspect Med* **2,** a007310 (2012).
257. Morgan, C. *et al.* The use of nonhuman primate models in HIV vaccine development. *PLoS Med.* **5,** 1200–1204 (2008).
258. Amos, J. D. *et al.* Rapid Development of gp120-Focused Neutralizing B Cell Responses during Acute Simian Immunodeficiency Virus Infection of African Green Monkeys. *J. Virol.* **89,** 9485–9498 (2015).
259. Jain, S., Trivett, M. T., Ayala, V. I., Ohlen, C. & Ott, D. E. African Green Monkey TRIM5α Restriction in Simian Immunodeficiency Virus-Specific Rhesus Macaque Effector CD4 T Cells Enhances Their Survival and Antiviral Function. *J. Virol.* **89,** 4449–4456 (2015).
260. Jacquelin, B. *et al.* Nonpathogenic SIV infection of African green monkeys induces a strong but rapidly controlled type I IFN response. *J. Clin. Invest.* **119,** 3544–55 (2009).
261. Nath, B. M., Schumann, K. E. & Boyer, J. D. The chimpanzee and other non-human-primate models in HIV-1 vaccine research. *Trends Microbiol.* **8,** 426–431 (2000).
262. Loffredo, J. T. *et al.* Mamu-B*08-Positive Macaques Control Simian Immunodeficiency Virus Replication. *J. Virol.* **81,** 8827–8832 (2007).
263. Mothé, B. R. *et al.* Expression of the Major Histocompatibility Complex Class I Molecule Mamu-A * 01 is associated with control of Simian Immunodeficiency Virus SIV mac 239 replication. *J. Virol.* **77,** 2736 (2003).
264. Sauermann, U. *et al.* Homozygosity for a conserved Mhc class II DQ-DRB haplotype is associated with rapid disease progression in simian immunodeficiency virus-infected macaques: results from a prospective study. *J. Infect. Dis.* **182,** 716–724 (2000).
265. Yant, L. J. *et al.* The High-Frequency Major Histocompatibility Complex Class I Allele Mamu-B * 17 Is Associated with Control of Simian Immunodeficiency Virus SIVmac239 Replication. *Society* **80,** 5074–5077 (2006).
266. Zhang, Z. *et al.* Mamu-A * 01 Allele-Mediated Attenuation of Disease Progression in Simian-Human Immunodeficiency Virus Infection Mamu-A ٭ 01 Allele-Mediated Attenuation of Disease Progression in Simian-Human Immunodeficiency Virus Infection. **76,** 12845–12854 (2002).
267. Goulder, P. J. R. & Watkins, D. I. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat. Rev. Immunol.* **8,** 619–630 (2008).
268. Goulder, P. J. *et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat. Med.* **3,** 212–217 (1997).
269. Allen, T. M. *et al.* Characterization of the peptide binding motif of a rhesus MHC class I molecule (Mamu-A*01) that binds an immunodominant CTL epitope from simian immunodeficiency virus. *J. Immunol.* **160,** 6062–6071 (1998).
270. Hatziioannou, T. & Evans, D. T. Animal models for HIV/AIDS research. *Nat Rev Microbiol* **10,** 852–867 (2012).
271. Wilson, D. P. *et al.* Estimating the infectivity of CCR5-tropic simian immunodeficiency virus SIV(mac251) in the gut. *J. Virol.* **81,** 8025–8029 (2007).
272. Means, R. E., Greenough, T. & Desrosiers, R. C. Neutralization sensitivity of cell culture passaged simian immunodeficiency virus. *J. Virol.* **71,** 7895–7902 (1997).
273. Yeh, W. W. *et al.* Autologous neutralizing antibodies to the transmitted/founder viruses emerge late after simian immunodeficiency virus SIVmac251 infection of rhesus monkeys. *J. Virol.* **84,** 6018–32 (2010).
274. Sun, Y. *et al.* Antibody-Dependent Cell-Mediated Cytotoxicity in Simian Immunodeficiency Virus-Infected Rhesus Monkeys. *J. Virol.* **85,** JVI.00326-11- (2011).
275. Strickland, S. L. *et al.* Significant Genetic Heterogeneity of the SIVmac251 Viral Swarm Derived from Different Sources. *AIDS Res. Hum. Retroviruses* **27,** 1327–1332 (2011).
276. Rud, E. W. *et al.* Molecular and biological characterization of simian immunodeficiency virus

macaque strain 32H proviral clones containing nef size variants. *J. Gen. Virol.* **75,** 529–543 (1994).
277. Lopker, M. *et al.* Heterogeneity in neutralization sensitivities of viruses comprising the simian immunodeficiency virus SIVsmE660 isolate and vaccine challenge stock. *J. Virol.* **87,** 5477–92 (2013).
278. Wu, F. *et al.* Sequential Evolution and Escape from Neutralization of Simian Immunodeficiency Virus SIVsmE660 Clones in Rhesus Macaques. *J. Virol.* **86,** 8835–8847 (2012).
279. Sato, S. *et al.* Potent antibody-mediated neutralization and evolution of antigenic escape variants of simian immunodeficiency virus strain SIVmac239 in vivo. *J. Virol.* **82,** 9739–9752 (2008).
280. Johnson, W. E. *et al.* Assorted mutations in the envelope gene of simian immunodeficiency virus lead to loss of neutralization resistance against antibodies representing a broad spectrum of specificities. *J. Virol.* **77,** 9993–10003 (2003).
281. Cranage, M. P. *et al.* Macaques infected with live attenuated SIVmac are protected against superinfection via the rectal mucosa. *Virology* **229,** 143–154 (1997).
282. Johnson, R. P. *et al.* Induction of vigorous cytotoxic T-lymphocyte responses by live attenuated simian immunodeficiency virus. *J. Virol.* **71,** 7711–8 (1997).
283. Nixon, D. F. *et al.* Simian immunodeficiency virus-specific cytotoxic T lymphocytes and protection against challenge in rhesus macaques immunized with a live attenuated simian immunodeficiency virus vaccine. *Virology* **266,** 203–210 (2000).
284. Koff, W. C. *et al.* HIV vaccine design: insights from live attenuated SIV vaccines. *Nat. Immunol.* **7,** 19–23 (2006).
285. Reynolds, M. R. *et al.* Macaques Vaccinated with Simian Immunodeficiency Virus SIVmac239 nef Delay Acquisition and Control Replication after Repeated Low-Dose Heterologous SIV Challenge. *J. Virol.* **84,** 9190–9199 (2010).
286. Staprans, S. I. & Feinberg, M. B. The roles of nonhuman primates in the preclinical evaluation of candidate AIDS vaccines. *Expert Rev. Vaccines* **3,** S5-32 (2004).
287. Feinberg, M. B. & Moore, J. P. AIDS vaccine models: challenging challenge viruses. *Nat. Med.* **8,** 207–10 (2002).
288. Girard, M. P., Osmanov, S., Assossou, O. M. & Kieny, M.-P. Human immunodeficiency virus (HIV) immunopathogenesis and vaccine development: A review. *Vaccine* **29,** 6191–6218 (2011).
289. Gautam, R. *et al.* Pathogenicity and mucosal transmissibility of the R5-tropic simian/human immunodeficiency virus SHIV(AD8) in rhesus macaques: implications for use in vaccine studies. *J. Virol.* **86,** 8516–26 (2012).
290. Song, R. J. *et al.* Molecularly cloned SHIV-1157ipd3N4: a highly replication- competent, mucosally transmissible R5 simian-human immunodeficiency virus encoding HIV clade C Env. *J. Virol.* **80,** 8729–8738 (2006).
291. Ng, C. T. *et al.* Passive neutralizing antibody controls SHIV viremia and enhances B cell responses in infant macaques. *Nat. Med.* **16,** 1117–1119 (2010).
292. Hessell, A. J. *et al.* Early short-term treatment with neutralizing human monoclonal antibodies halts SHIV infection in infant macaques. *Nat. Med.* **22,** 1–9 (2016).
293. Barouch, D. H. *et al.* Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature* **503,** 224–228 (2013).
294. Reimann, K. A. *et al.* A chimeric simian/human immunodeficiency virus expressing a primary patient human immunodeficiency virus type 1 isolate env causes an AIDS-like disease after in vivo passage in rhesus monkeys. *J. Virol.* **70,** 6922–8 (1996).
295. Parker, R. A., Regan, M. M., Reimann, K. A., Parker, R. A. & Regan, M. M. Variability of Viral Load in Plasma of Rhesus Monkeys Inoculated with Simian Immunodeficiency Virus or Simian-Human Immunodeficiency Virus : Implications for Using Nonhuman Primate AIDS Models To Test Vaccines and Therapeutics Variability of Viral Load in. **75,** 11234–11238 (2001).
296. Yamamoto, T. *et al.* Quality and quantity of T FH cells are critical for broad antibody development in SHIV AD8 infection. *Sci. Transl. Med.* **7,** 1–12 (2015).
297. Nishimura, Y. *et al.* Generation of the pathogenic R5-tropic simian/human immunodeficiency virus SHIVAD8 by serial passaging in rhesus macaques. *J. Virol.* **84,** 4769–81 (2010).

298. Shingai, M. *et al.* Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia. *Nature* **503,** 277–80 (2013).
299. Jia, M., Lu, H., Markowitz, M., Cheng-Mayer, C. & Wu, X. Development of broadly neutralizing antibodies and their mapping by monomeric gp120 in HIV-1 infected humans and SHIV $_{SF162P3N}$ infected macaques. *J. Virol.* JVI.02898-15 (2016). doi:10.1128/JVI.02898-15
300. Fouts, T. *et al.* Crosslinked HIV-1 envelope-CD4 receptor complexes elicit broadly cross-reactive neutralizing antibodies in rhesus macaques. *Proc. Natl. Acad. Sci.* **99,** 11842–11847 (2002).
301. Namikawa, R., Kaneshima, H., Lieberman, M., Weissman, I. L. & McCune, J. M. Infection of the SCID-hu mouse by HIV-1. *Science* **242,** 1684–1686 (1988).
302. Goldman, J. P. *et al.* Enhanced human cell engraftment in mice deficient in RAG2 and the common cytokine receptor γ chain. *Br. J. Haematol.* **103,** 335–342 (1998).
303. Melkus, M. W. *et al.* Humanized mice mount specific adaptive and innate immune responses to EBV and TSST-1. *Nat. Med.* **12,** 1316–1322 (2006).
304. M Barry, S. Trial, Error, and Breakthrough: A Review of HIV Vaccine Development. *J. AIDS Clin. Res.* **5,** (2014).
305. Leslie, A. J. *et al.* HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med.* **10,** 282–9 (2004).
306. Boutwell, C. L., Rolland, M. M., Herbeck, J. T., Mullins, J. I. & Allen, T. M. Viral evolution and escape during acute HIV-1 infection. *J. Infect. Dis.* **202 Suppl,** (2010).
307. Duerr, A., Wasserheit, J. N. & Corey, L. HIV vaccines: new frontiers in vaccine development. *Clin. Infect. Dis.* **43,** 500–11 (2006).
308. Hütter, G. *et al.* Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N. Engl. J. Med.* **360,** 692–8 (2009).
309. Whitney, J. B. *et al.* Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512,** 74–7 (2014).
310. Berry, N. *et al.* Early potent protection against heterologous SIVsmE660 challenge following live attenuated SIV vaccination in mauritian cynomolgus macaques. *PLoS One* **6,** (2011).
311. Baba, T. W. *et al.* Live attenuated, multiply deleted simian immunodeficiency virus causes AIDS in infant and adult macaques. *Nat. Med.* **5,** 194–203 (1999).
312. Angel, J. B., Routy, J.-P., Graziani, G. M. & Tremblay, C. L. The Effect of Therapeutic HIV Vaccination With ALVAC-HIV With or Without Remune on the Size of the Viral Reservoir (A CTN 173 Substudy). *J. Acquir. Immune Defic. Syndr.* **70,** 122–8 (2015).
313. Moore, J. P. *et al.* Primary isolates of human immunodeficiency virus type 1 are relatively resistant to neutralization by monoclonal antibodies to gp120, and their neutralization is not predicted by studies with monomeric gp120. *J. Virol.* **69,** 101–9 (1995).
314. Mascola, J. R. *et al.* Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. The National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluati. *J. Infect. Dis.* **173,** 340–8 (1996).
315. Graham, B. S. *et al.* Phase 1 safety and immunogenicity evaluation of a multiclade HIV-1 DNA candidate vaccine. *J. Infect. Dis.* **194,** 1650–60 (2006).
316. Cooney, E. L. *et al.* Safety of and immunological response to a recombinant vaccinia virus vaccine expressing HIV envelope glycoprotein. *Lancet (London, England)* **337,** 567–72 (1991).
317. Hu, S. L., Kosowski, S. G. & Dalrymple, J. M. Expression of AIDS virus envelope gene in recombinant vaccinia viruses. *Nature* **320,** 537–40
318. Hu, S. L. *et al.* Effect of immunization with a vaccinia-HIV env recombinant on HIV infection of chimpanzees. *Nature* **328,** 721–3
319. Mothe, B. *et al.* Safety and immunogenicity of a modified vaccinia Ankara-based HIV-1 vaccine (MVA-B) in HIV-1-infected patients alone or in combination with a drug to reactivate latent HIV-1. *J. Antimicrob. Chemother.* **70,** 1833–42 (2015).
320. Bonsignori, M. *et al.* Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. *Cell* **165,** 449–463 (2016).
321. Hansen, S. G. *et al.* Immune clearance of highly pathogenic SIV infection. *Nature* **502,** 100–4 (2013).
322. Hansen, S. G. *et al.* Cytomegalovirus Vectors Violate CD8+ T Cell Epitope Recognition Paradigms. *Science (80-.).* **340,** 1237874–1237874 (2013).

323. Excler, J.-L., Ake, J., Robb, M. L., Kim, J. H. & Plotkin, S. A. Nonneutralizing functional antibodies: a new 'old' paradigm for HIV vaccines. *Clin. Vaccine Immunol.* **21,** 1023–36 (2014).
324. Douek, D. C. *et al.* Assessment of thymic output in adults after haematopoietic stemcell transplantation and prediction of T-cell reconstitution. *Lancet* **355,** 1875–1881 (2000).
325. Palmer, S. *et al.* New Real-Time Reverse Transcriptase-Initiated PCR Assay with Single-Copy Sensitivity for Human Immunodeficiency Virus Type 1 RNA in Plasma. *J. Clin. Microbiol.* **41,** 4531–4536 (2003).
326. Lifson, J. D. *et al.* Role of CD8(+) lymphocytes in control of simian immunodeficiency virus infection and resistance to rechallenge after transient early antiretroviral treatment. *J. Virol.* **75,** 10187–99 (2001).
327. Gall, A. *et al.* Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J. Clin. Microbiol.* **50,** 3838–44 (2012).
328. Mason, R. D. *et al.* Targeted Isolation of Antibodies Directed against Major Sites of SIV Env Vulnerability. *PLoS Pathog.* **12,** e1005537 (2016).
329. Tiller, T. *et al.* Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods* **329,** 112–124 (2008).
330. Watson, S. J. *et al.* Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120205 (2013).
331. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–10 (1990).
332. Bashford-Rogers, R. J. *et al.* Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome Res* (2013). doi:10.1101/gr.154815.113
333. Hunt, M. *et al.* IVA: Accurate de novo assembly of RNA virus genomes. *Bioinformatics* **31,** 2374–2376 (2015).
334. Rose, P. P. & Korber, B. T. Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. *Bioinformatics* **16,** 400–1 (2000).
335. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
336. Koichiro, T. *et al.* MEGA5: Molecular Evolutionary Genetics Analysis. *Mol. Biol. Evol.* **28,** 2731–2739 (2011).
337. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** (2010).
338. Zhu, J. *et al.* Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc Natl Acad Sci U S A* **110,** 6470–6475 (2013).
339. Lifson, J. D. *et al.* The extent of early viral replication is a critical determinant of the natural history of simian immunodeficiency virus infection. *J. Virol.* **71,** 9508–9514 (1997).
340. Kilgore, K. M. *et al.* Characterization and Implementation of a Diverse Simian Immunodeficiency Virus SIVsm Envelope Panel in the Assessment of Neutralizing Antibody Breadth Elicited in Rhesus Macaques by Multimodal Vaccines Expressing the SIVmac239 Envelope. *J Virol* **89,** 8130–8151 (2015).
341. Francica, J. R. *et al.* Analysis of immunoglobulin transcripts and hypermutation following SHIVAD8 infection and protein-plus-adjuvant immunization. *Nat. Commun.* **6,** 6565 (2015).
342. Galson, J. D. *et al.* BCR repertoire sequencing: different patterns of B cell activation after two Meningococcal vaccines. *Immunol. Cell Biol.* 1–11 (2015). doi:10.1038/icb.2015.57
343. Yu, L. & Guan, Y. Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. *Front. Immunol.* **5,** 250 (2014).
344. Sundling, C. *et al.* High-resolution definition of vaccine-elicited B cell responses against the HIV primary receptor binding site. *Sci. Transl. Med.* **4,** 142ra96 (2012).
345. Euler, Z. & Schuitemaker, H. Cross-reactive broadly neutralizing antibodies: timing is everything. *Front. Immunol.* **3,** 215 (2012).
346. Volpe, J. M. & Kepler, T. B. Large-scale analysis of human heavy chain V(D)J recombination patterns. *Immunome Res.* **4,** 1–10 (2008).
347. Sundling, C. *et al.* Single-Cell and Deep Sequencing of IgG-Switched Macaque B Cells Reveal a Diverse Ig Repertoire following Immunization. *J. Immunol.* **192,** 3637–44 (2014).

348. Phad, G. E. *et al.* Diverse Antibody Genetic and Recognition Properties Revealed following HIV-1 Envelope Glycoprotein Immunization. *J. Immunol.* 1500122 (2015). doi:10.4049/jimmunol.1500122
349. Moir, S. & Fauci, A. S. B cells in HIV infection and disease. *Nat. Rev. Immunol.* **9,** 235–45 (2009).
350. Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17,** 282–3 (2001).
351. Hoi, K. H. & Ippolito, G. C. Intrinsic bias and public rearrangements in the human immunoglobulin Vλ light chain repertoire. *Genes Immun.* **14,** 1–6 (2013).
352. Galson, J. D. *et al.* In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. *Front. Immunol.* **6,** 1–13 (2015).
353. Quigley, M. F. *et al.* Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19414–9 (2010).
354. Venturi, V., Price, D. A., Douek, D. C. & Davenport, M. P. The molecular basis for public T-cell responses? *Nat. Rev. Immunol.* **8,** 231–8 (2008).
355. Kedzierska, K. *et al.* Quantification of Repertoire Diversity of Influenza-Specific Epitopes with Predominant Public or Private TCR Usage. *J. Immunol.* **177,** 6705–6712 (2006).
356. Price, D. A. *et al.* Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. *J. Exp. Med.* **206,** 923–36 (2009).
357. Parameswaran, P. *et al.* Convergent antibody signatures in human dengue. *Cell Host Microbe* **13,** 691–700 (2013).
358. Wu, Y.-C. C. *et al.* High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* **116,** 1070–1078 (2010).
359. Kidera, A., Konish, Y., Oka, M., Ooi, T. & Scheraga, H. a. *Statistical Analysis of the Physical Properties of the 20 Naturally Occurring Amino Acids*. *Journal of Protein Chemistry* **4,** (1985).
360. Moore, D. S. Amino acid and peptide net charges: A simple calculational procedure. *Biochem. Educ.* **13,** 10–11 (1985).
361. Ikai, A. Thermostability and Aliphatic Index of Globular Proteins. *J. Biochem.* **88,** 1895–1898 (1980).
362. Wang, Y. *et al.* High-Resolution Longitudinal Study of HIV-1 Env Vaccine-Elicited B Cell Responses to the Virus Primary Receptor Binding Site Reveals Affinity Maturation and Clonal Persistence. *J. Immunol.* (2016). doi:10.4049/jimmunol.1502543
363. Schnittman, S. M. *et al.* Preferential infection of CD4+ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 6058–62 (1990).
364. Brenchley, J. M. *et al.* T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis. *J. Virol.* **78,** 1160–8 (2004).
365. Stieh, D. J. *et al.* Th17 Cells Are Preferentially Infected Very Early after Vaginal Transmission of SIV in Macaques. *Cell Host Microbe* **19,** 529–40 (2016).
366. Gosselin, A. *et al.* Peripheral blood CCR4+CCR6+ and CXCR3+CCR6+CD4+ T cells are highly permissive to HIV-1 infection. *J. Immunol.* **184,** 1604–16 (2010).
367. Yamamoto, T. *et al.* Quality and quantity of T FH cells are critical for broad antibody development in SHIV AD8 infection. **7,** 1–12 (2015).
368. Ma, A., Koka, R. & Burkett, P. Diverse functions of IL-2, IL-15, and IL-7 in lymphoid homeostasis. *Annu. Rev. Immunol.* **24,** 657–79 (2006).
369. Zhang, S.-Y. *et al.* Progressive CD127 down-regulation correlates with increased apoptosis of CD8 T cells during chronic HIV-1 infection. *Eur. J. Immunol.* **39,** 1425–34 (2009).
370. Hutloff, A. *et al.* ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* **397,** 263–6 (1999).
371. Jin, H.-T., Ahmed, R. & Okazaki, T. Role of PD-1 in regulating T-cell immunity. *Curr. Top. Microbiol. Immunol.* **350,** 17–37 (2011).
372. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. Appendix II. CD Antigens. (2001). at <http://www.ncbi.nlm.nih.gov/books/NBK10772/>
373. Yamazaki, T. *et al.* CCR6 regulates the migration of inflammatory and regulatory T cells. *J. Immunol.* **181,** 8391–8401 (2008).
374. Panzer, U. *et al.* Chemokine receptor CXCR3 mediates T cell recruitment and tissue injury in nephrotoxic nephritis in mice. *J. Am. Soc. Nephrol.* **18,** 2071–84 (2007).

375. Faustino, L. *et al.* Regulatory T cells migrate to airways via CCR4 and attenuate the severity of airway allergic inflammation. *J. Immunol.* **190,** 2614–21 (2013).
376. Dunham, R. M. *et al.* CD127 and CD25 expression defines CD4+ T cell subsets that are differentially depleted during HIV infection. *J. Immunol.* **180,** 5582–5592 (2008).
377. Porichis, F. *et al.* High-throughput detection of miRNAs and gene-specific mRNA at the single-cell level by flow cytometry. *Nat. Commun.* **5,** 5641 (2014).
378. Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14,** 22–9 (2012).
379. Pantaleo, G. *et al.* HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* **362,** 355–8 (1993).
380. Connors, M. *et al.* HIV infection induces changes in CD4+ T-cell phenotype and depletions within the CD4+ T-cell repertoire that are not immediately restored by antiviral or immune-based therapies. *Nat. Med.* **3,** 533–540 (1997).
381. Haase, A. T. Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues. *Annu. Rev. Immunol.* **17,** 625–56 (1999).
382. Josefsson, L. *et al.* Majority of CD4 + T cells from peripheral blood of HIV-1–infected individuals contain only one HIV DNA molecule. doi:10.1073/pnas.1107729108
383. Batorsky, R. *et al.* Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 5661–6 (2011).
384. Perelson, A. S. *et al.* Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387,** 188–91 (1997).
385. Besson, G. J. *et al.* HIV-1 DNA decay dynamics in blood during more than a decade of suppressive antiretroviral therapy. *Clin. Infect. Dis.* **59,** 1312–21 (2014).
386. Pruss, D., Bushman, F. D. & Wolffe, A. P. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5913–7 (1994).
387. Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C. & Bushman, F. D. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17,** 1186–94 (2007).
388. Del Prete, G. Q. *et al.* Comparative Characterization of Transfection- and Infection-Derived Simian Immunodeficiency Virus Challenge Stocks for In Vivo Nonhuman Primate Studies. *J. Virol.* **87,** 4584–4595 (2013).
389. Chen, H. Y., Di Mascio, M., Perelson, A. S., Ho, D. D. & Zhang, L. Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 19079–84 (2007).

8 Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
BCR	B Cell Receptor
bnAbs	Broadly neutralizing antibodies
CD	Cluster of Differentiation
CDR	Complementarity-determining region
CM	Central Memory
CTL	Cytotoxic T lymphocyte
DC	Dendritic Cell
ELISA	Enzyme-Linked Immunosorbent Assay
FACS	Fluorescence-activated cell sorting
FR	Framework Region
GC	Germinal Center
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HLA	Human Leukocyte Antigen
IFN	Interferon
IGH	Immunoglobulin heavy chain
IL	Interleukin
LN	Lymph node
LTNP	Long term nonprogressor
MHC	Major histocompatibility complex
mAb	Monoclonal antibody
NHP	Nonhuman primate
NK	Natural killer cell
NKT	Natural killer T cell
PBMC	Peripheral blood mononuclear cells
RACE	Rapid amplification of cDNA ends
RPMI	Roswell Park Memorial Institute media
SIV	Simian immunodeficiency virus
SHM	Somatic hypermutation
TFH	T follicular helper cells
TFR	T follicular regulatory cells
TCR	T cell receptor
VDJ	Variable, diversity, and joining regions