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].

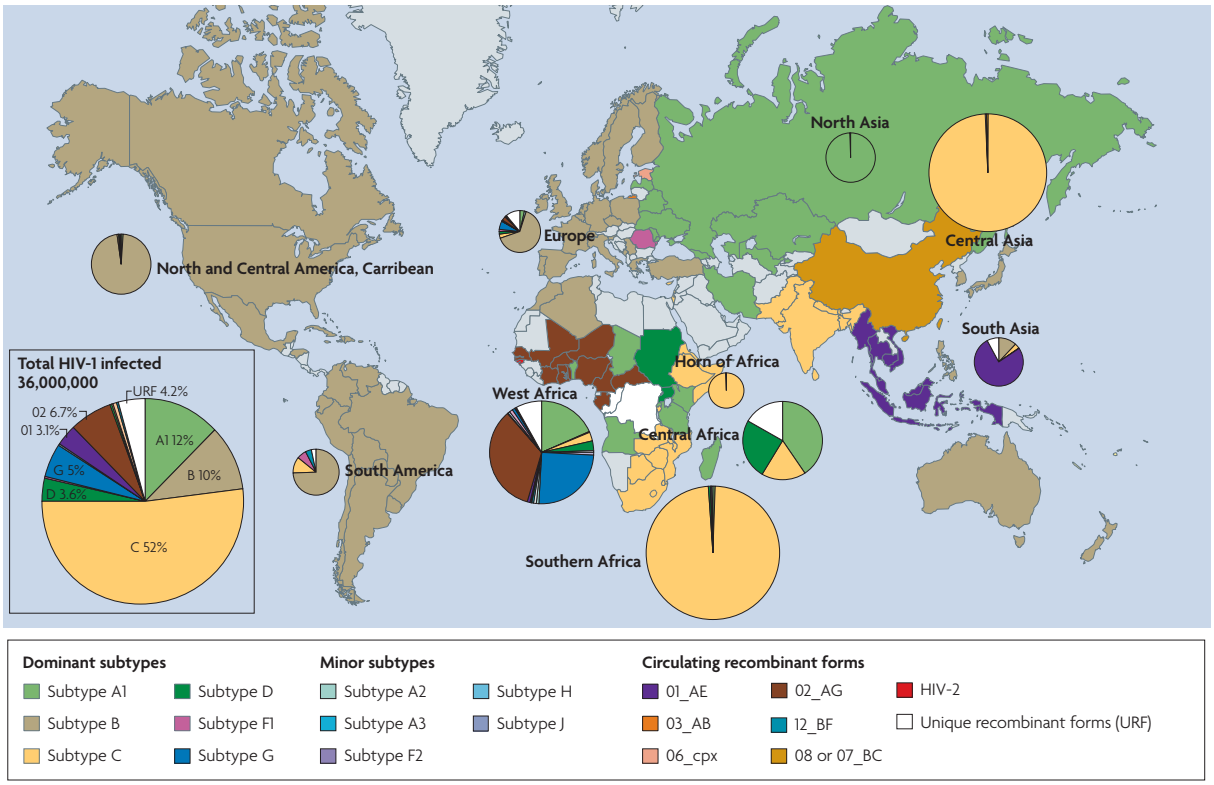

Figure 1.1. Global prevalence and distribution of HIV-1 subtypes. From Ariën, Vanham, and Arts 2007 Nature Reviews Microbiology[14]. Subtype C is the most common worldwide and is largely found in southern and eastern Africa and in the Indian subcontinent. The majority of infections in the Americas and 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

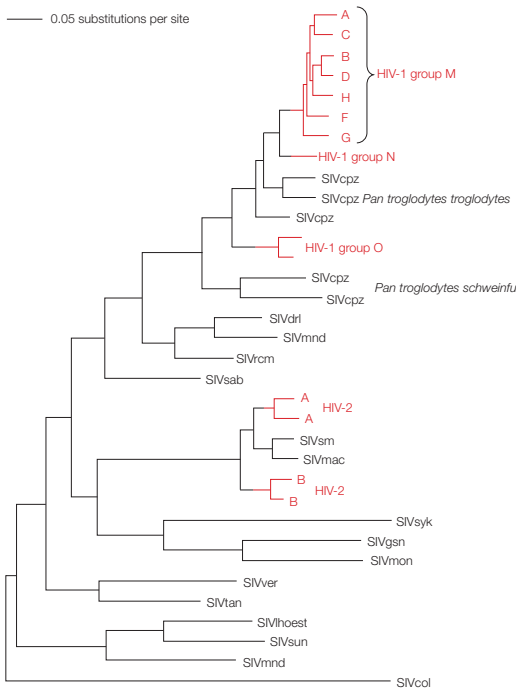

Fig 1.2. Phylogeny and ancestry of primate lentiviruses. From Pybus and Rambaut, 2009 Nature Reviews Genetics[4] The two HIV lineages (HIV-1 and HIV-2) fall between SIV branches and represent independent cross-species transmission events rather than a single transmission event. HIV-2 is closest to SIV 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

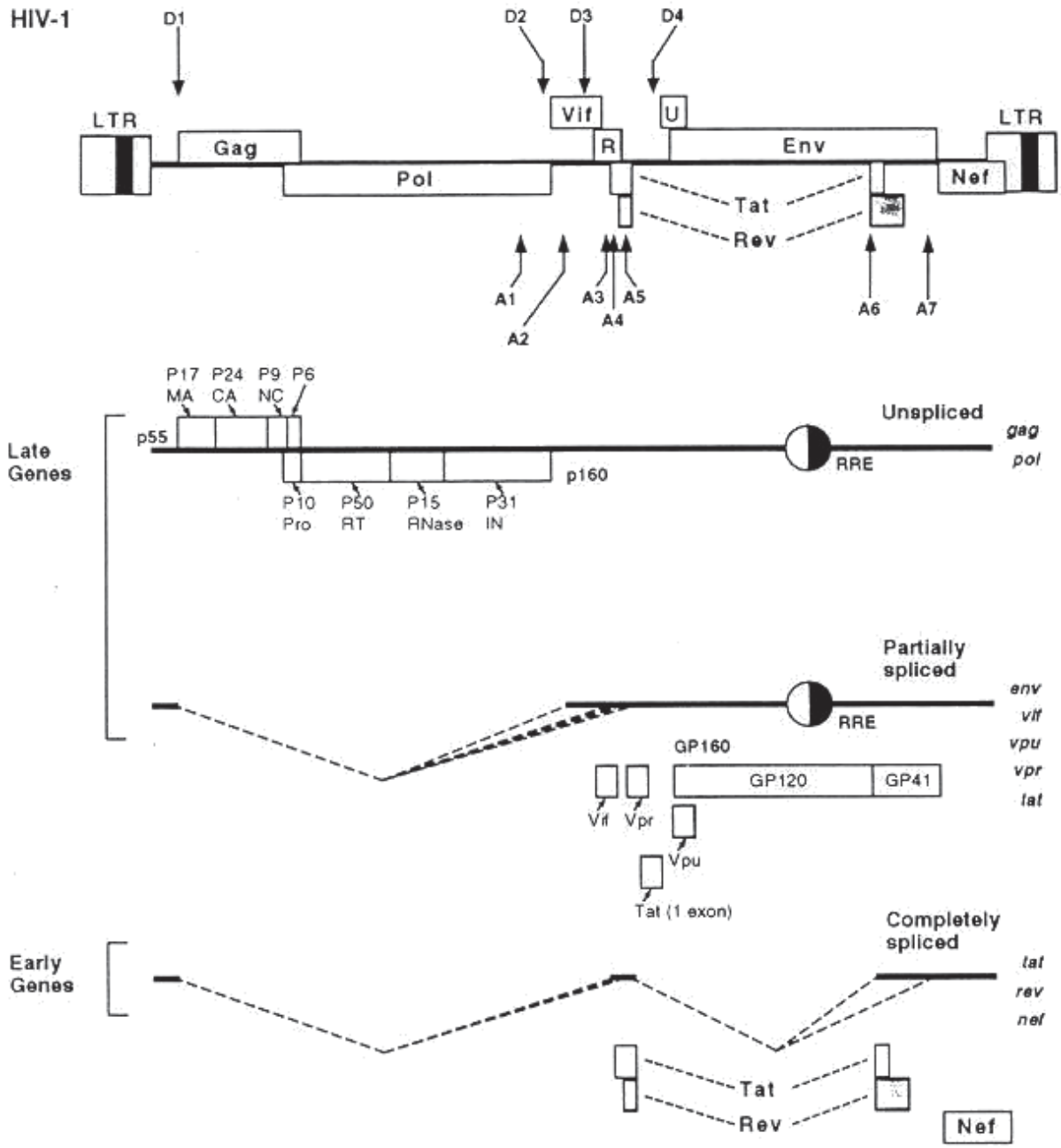

Fig 1.3 [25] **HIV/SIV genome organisation and splicing.** From http://www2.gsu.edu/~biotkf/bio475/475lecture8.htm . The HIV/SIV genome contains nine genes: three structural genes and six accessory genes. The 2kb RNAs are multiply spliced and are produced first in viral infection (Tat, rev, nef), followed 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].

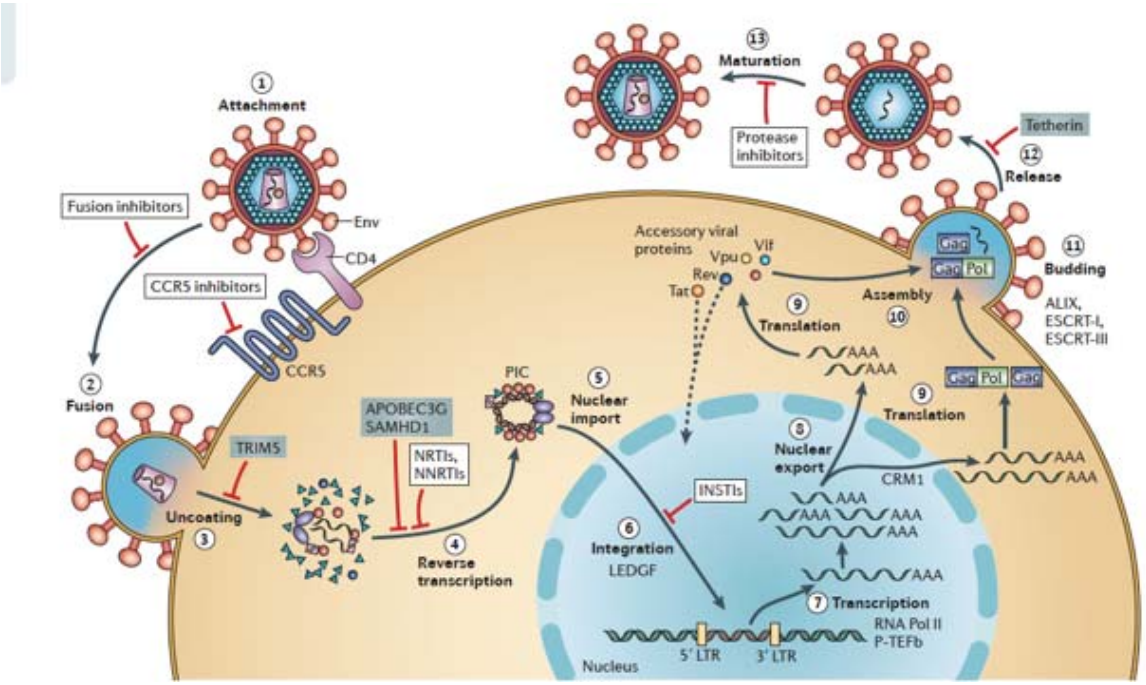

Fig 1.5 HIV/SIV Infection and lifecycle. From Engelmen and Cherepanov, 2012 Nature Reviews Microbiology[37]. HIV binds to the cell surface via CD4 and a co-receptor (CCR5 or CXCR4), and undergoes membrane fusion and release of the viral capsid into the cytoplasm. After capsid uncoating and reverse transcription 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].

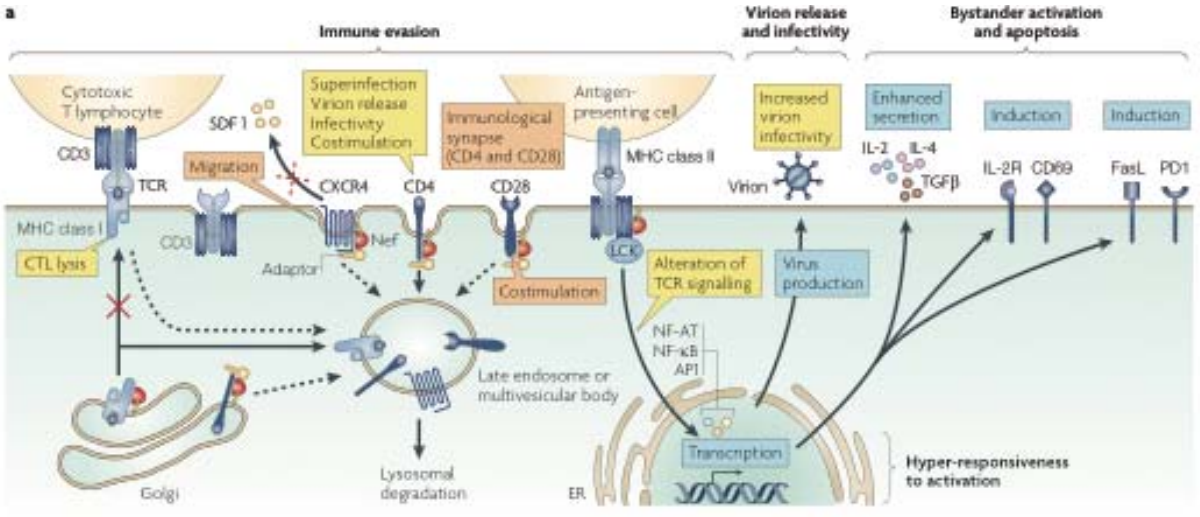

Fig 1.4 Immune evasion by nef. From Kirchoff, 2009 Nature Reviews Microbiology[51]. Nef downregulates CD4, CD28, MHC-1, MHC-II and in the case of SIV, CD3 from the cell surface. This allows the infected cell to avoid adaptive immune responses from CTL and antigen-presenting cells. Nef also promotes virus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

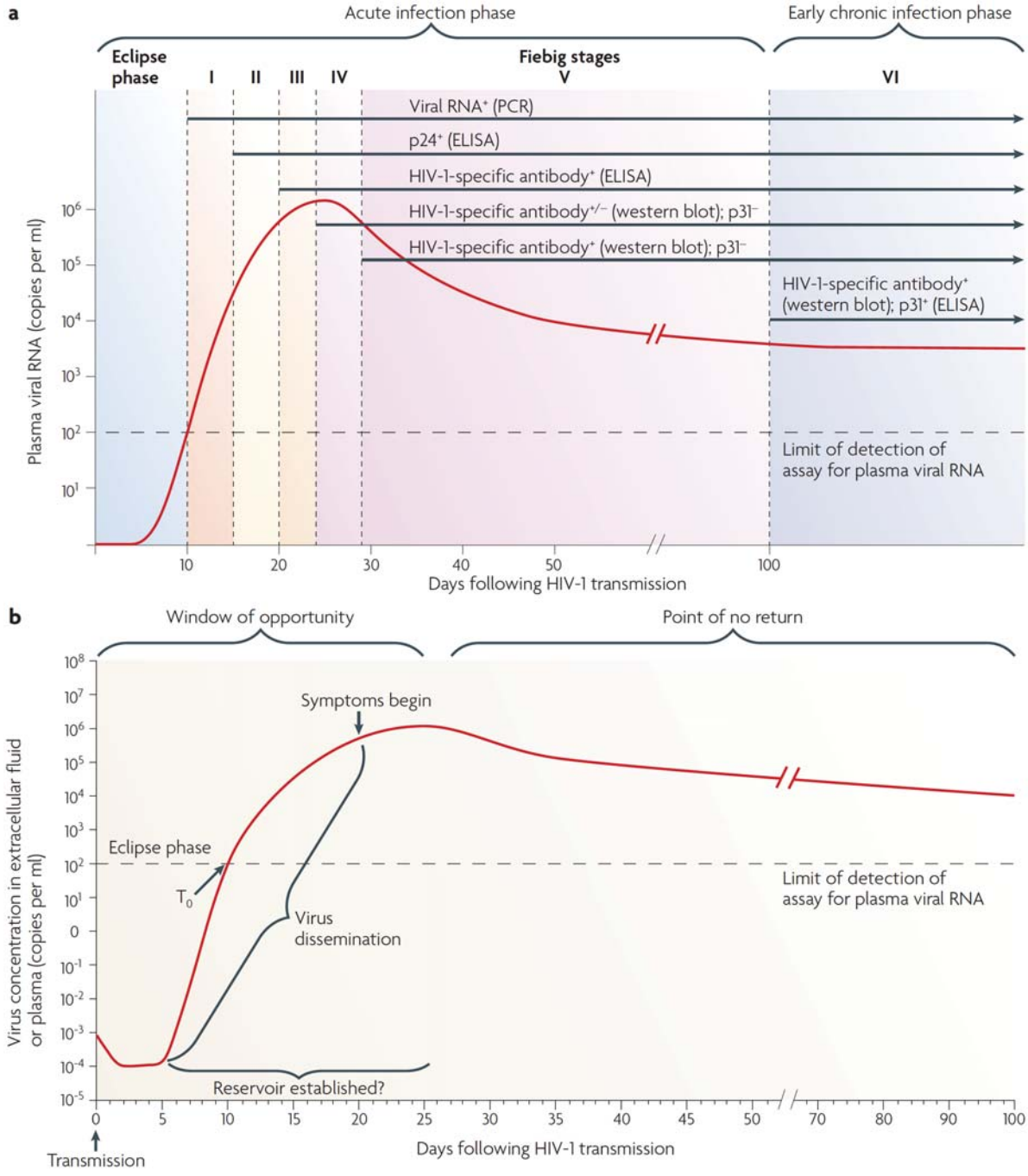

Fig 1.6. Stages of HIV infection. From McMichael, Borrow, Tomaras, Goonetilleke, and Haynes, 2010 Nature Reviews Immunology[28]**.** Fiebig stages of infection are determined by the type of test required to diagnose infection (PCR, ELISA, Western blot). The eclipse phase of infection refers to the period 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].

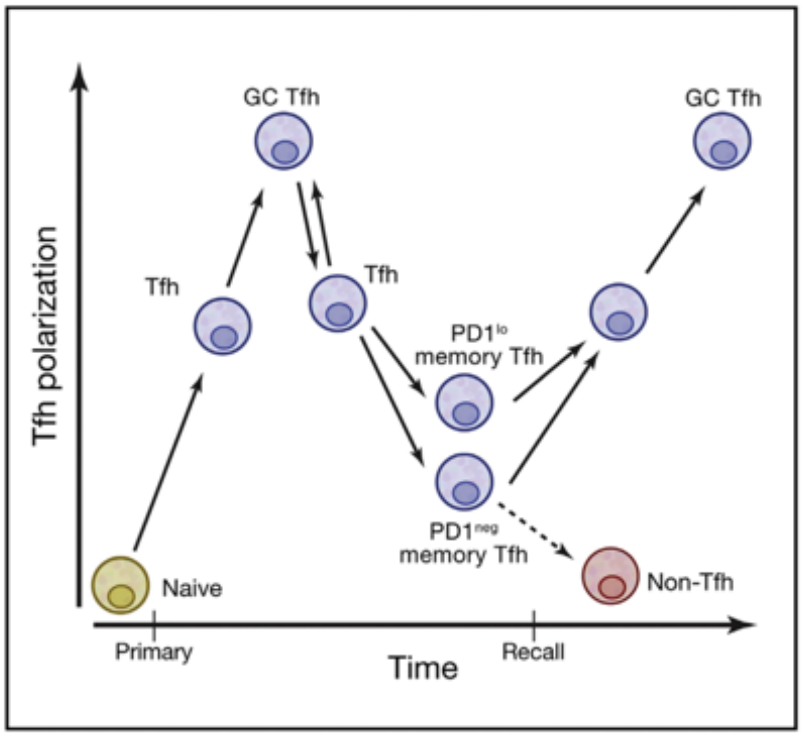

Fig 1.7 Differentiation and plasticity of T follicular helper cells in lymph nodes. From Crotty 2014, Immunity[134]. TFH develop from naïve T cells and gain TFH characteristics through interactions with dendritic cells and B cells, and can enter the germinal center upon up-regulation of CXCR5 and PD1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

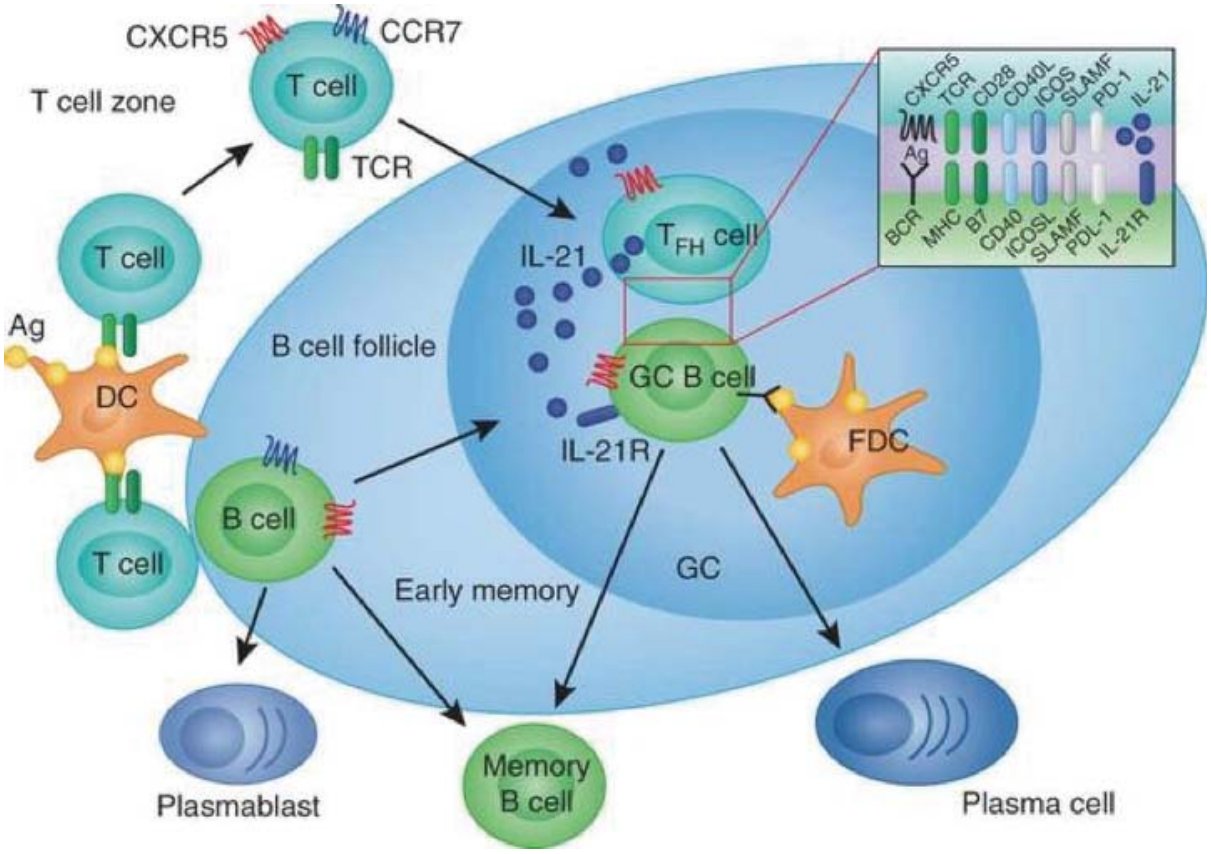

Fig 1.8. TFH localisation and migration. From Nutt and Tarlinton 2011, Nature Immunology[168]. T cells are primed on dendritic cells in the T cell zone, upregulate CXCR5 and downregulate CCR7 and migrate to the B cell follicle and into germinal centers. There, they interact with B cells and signal using 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].

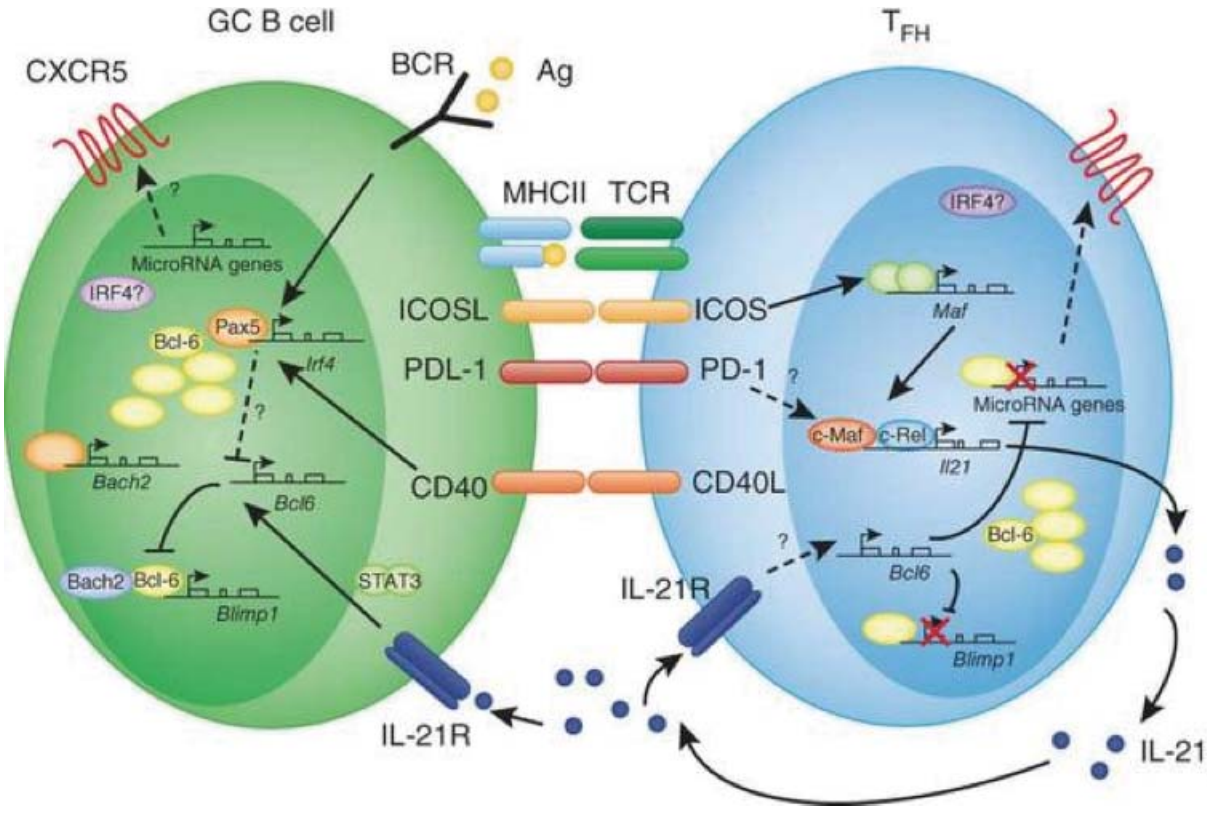

Fig 1.9. TFH and B cell interactions and transcriptional networks. From Nutt and Tarlinton 2011, Nature Immunology[168]

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].

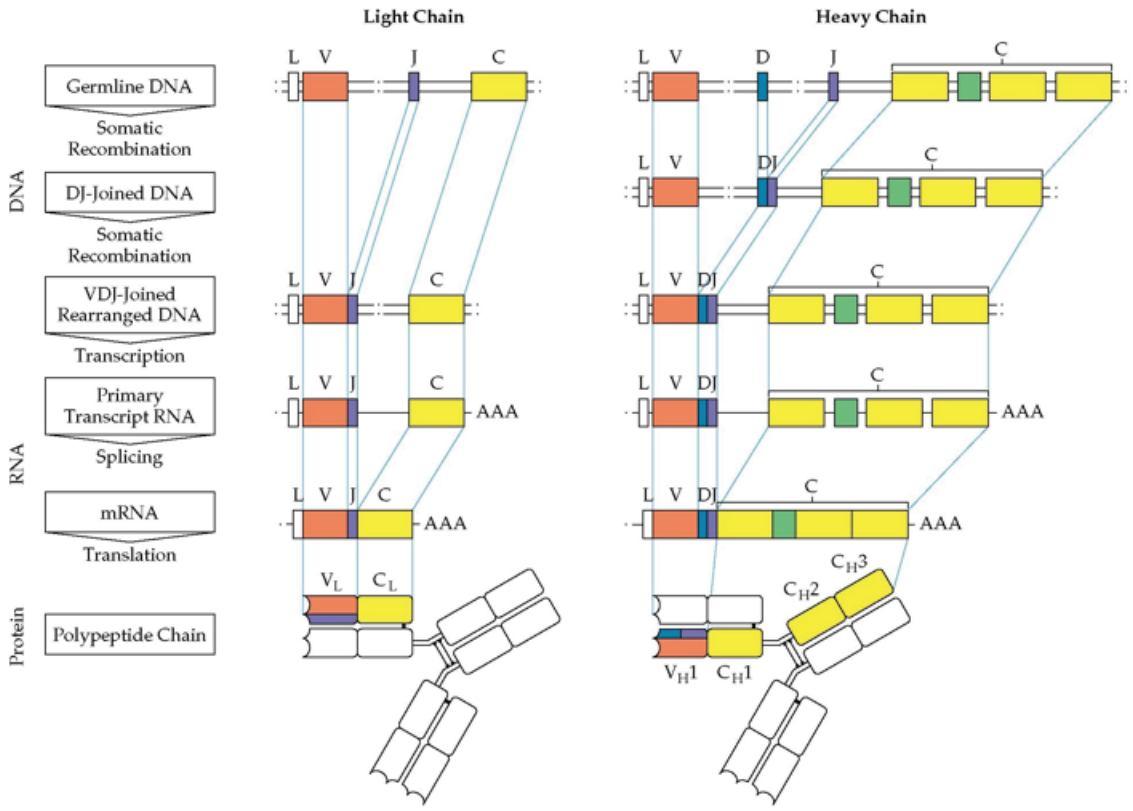

Fig 1.10. Germline rearrangement of heavy and light chains. Adapted from Schatz and Swanson 2011 Annual review of genetics[198]. At the germline level, chromosomal DNA is cut so that a single V, D, and J gene are joined to form a heavy chain exon (in light chains, a single D gene and J gene are joined 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]

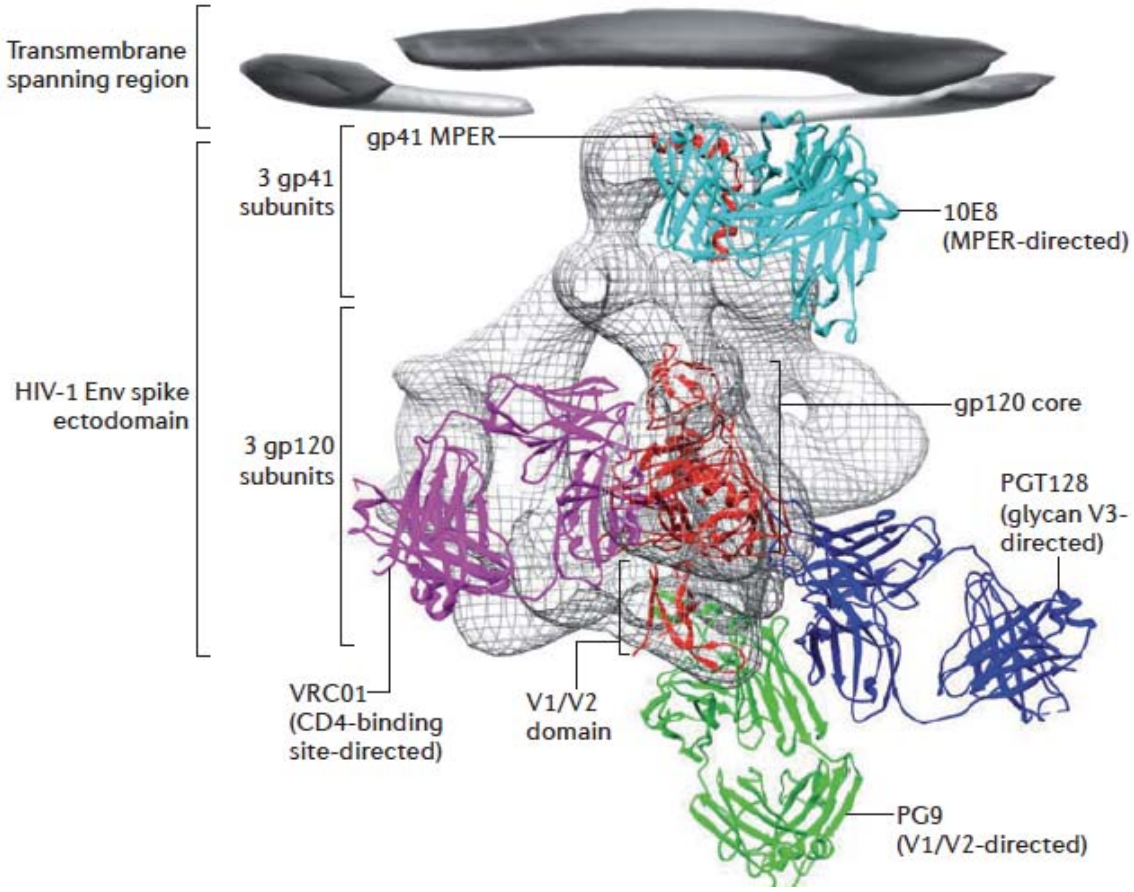

Fig 1.11. Sites of HIV broadly neutralizing antibody binding to envelope. From Kwong, Mascola, and Nabel 2013 Nature Reviews Immunology[219]. From HIV bnAbs are grouped based on the area where their class of antibodies bind, and include CD4 binding site, V1/V2 directed, glycan/V3 directed, membrane proximal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