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.