Systems biology of natural SIV infections

Steven E. Bosinger\textsuperscript{a}, Béatrice Jacquelin\textsuperscript{b}, Arndt Benecke\textsuperscript{c,d}, Guido Silvestri\textsuperscript{a}, and Michaela Müller-Trutwin\textsuperscript{b,*}
\textsuperscript{a}Yerkes National Primate Research Center, and Emory Vaccine Center, Emory University, Atlanta, Georgia
\textsuperscript{b}Institut Pasteur, Unité Régulation des Infections Rétrovirales, Paris, France
\textsuperscript{c}Institut des Hautes Études Scientifiques, Bures sur Yvette, France
\textsuperscript{d}Vaccine Research Institute, INSERM U955, Institut Mondor de Recherche Biomédicale, Créteil, France

Abstract

\textbf{Purpose of review—} A key factor driving AIDS-associated immunopathogenesis is chronic immune activation. SIV infection of African natural host species leads to high viremia, but low immune activation and absence of disease. Considerable progress in our understanding of pathological immune activation have come from comparative studies of SIV infection in pathogenic Asian macaque species and natural hosts. The focus of this review is to highlight recent work on the natural host model using high throughput genomics.

\textbf{Recent findings—} Several groups have independently conducted microarray gene expression profiling comparing in vivo SIV infection in natural and non-natural hosts. A consistent finding between these studies is that both pathogenic SIV infection of macaques and nonpathogenic infections of natural hosts have strong induction of interferon-stimulated genes (ISGs) early on, but a key difference was that natural hosts downmodulated the interferon response rapidly after acute infection. The development of new genome-based resources for further study of the natural host model is discussed.

\textbf{Summary—} Initial efforts using high throughput biology to study SIV infection of natural hosts have effectively identified the ability of natural hosts to resolve interferon responses and immune activation. Further application of ‘omic’-based technologies coupled with integrative systems-based analysis should continue to yield progress.

\textbf{Keywords} \\
Simian Immunodeficiency Virus; Microarray; Interferon; ISG; interferon stimulated genes; Genome; Sooty Mangabey; Cercopithecus Atys; African green monkey; vervet; Chlorocebus

\textsuperscript{*}Corresponding author Unité de Régulation des Infections Rétrovirales Institut Pasteur 25, rue du Dr Roux 75015 Paris, France Tel.: 0033 (0)1 40613969 Fax : 0033 (0)1 45689957 michaela.muller-trutwin@pasteur.fr.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Introduction

HIV-induced immunodeficiency is characterized by progressive functional impairment, loss of CD4+ T cells, and excessive, generalized immune activation. Chronic immune activation is theorized to drive CD4+ T cell depletion and progression towards AIDS[1], and understanding the underlying pathophysiology remains one of the most important questions in HIV research. African nonhuman primates (NHPs) represent the only reservoir of simian immunodeficiency viruses (SIV) in the wild. These lentiviruses are the ancestors of the human immunodeficiency viruses (HIV-1 and HIV-2)[2-4]. For three African NHPs, sooty mangabeys (SMs), African green monkeys (AGMs) and mandrills, it has been demonstrated that SIV infection is nonpathogenic[5,6]. In contrast, Asian and New World NHPs are not infected with SIV in their natural habitat. Macaque species, endemic to Asia, can be experimentally infected with virus strains derived from SM, collectively termed SIVmac, and progress to AIDS with clinical hallmarks similar to human disease. An important distinction between nonpathogenic infection of natural reservoir species of SIV compared to pathogenic SIV infection of Asian macaques species is that natural hosts do not exhibit chronic immune activation despite virus replication levels in blood and gut as high as in HIV-1 infection in humans and SIVmac infection in macaques[5]. Comparative study of SIV infection in African natural host species and nonnatural, pathogenic macaques is an important tool for understanding the pathophysiology of AIDS[5]. Several hypotheses, not entirely mutually exclusive, have been proposed in the literature to explain the lack of chronic immune activation in natural hosts[5,6]

Amongst viral factors, a role for the differential activity of the NEF protein between SIVmac and SIVsm strains has been suggested[7]. NEF derived from SIVmac strains cannot downregulate CD3 on T cells as efficiently as viruses from natural hosts, which may protect infected cells from TCR-mediated activation[8]. Considering host factors, several important differences have been observed for CD4+ T cells between natural and nonnatural hosts. In SIV-infected natural hosts, CD4+ T cells do not show increased susceptibility to apoptosis and bystander cell death[9-13]. Infection of central memory CD4+ T cells (TCM) is limited in natural hosts compared to which might confer protection[14-17]. In the gut, TCM are dramatically reduced during acute infection in natural hosts, with partial reconstitution. Gut-resident Th17 cells, and the balance between Th17 and CD25+ Treg cells, are preserved in natural hosts[18,19], which may contribute to the lack of microbial translocation and chronic immune activation[14,20]. A comprehensive review comparing the immune system of natural and nonnatural hosts during SIV infection has recently been reported[21]. The remainder of this review will examine the impact of genomic technology on the study of natural host biology and highlight developments that will accelerate future research.
Lessons learned about SIV/natural host biology using high-throughput genomics: resolution of immune activation

In 2009 four separate studies were published using microarray technology to characterize genome-wide expression during in vivo SIV infection of SMs[22] and AGMs[19,23,24]. These studies were conducted in a comparative fashion, performing parallel SIV infection of natural hosts with non-natural host macaque species (Table 1). Despite using distinct platforms, the results from all four studies were in remarkable concordance. SIV infection induced expression of an expansive array of antiviral genes in all species during the acute infection phase (1 to ~30-60 days post-infection). This antiviral response was comprised of transcripts encoding restriction factors, chemokines, genes regulating adaptive immune responses, and interferon stimulated genes (ISGs). Collectively, these works demonstrated conclusively that natural host species exhibit widespread induction of innate immunity, including the interferon (IFN) system, in response to SIV. An important difference in the immune response between natural and pathogenic hosts was the duration: after the initial wave of ISG induction in SMs and AGMs, expression returned to baseline levels in the post-acute phase (>30 days). In contrast, rhesus macaques (RMs) maintained elevated expression of ISGs chronically (>180 days).

Previously, most work on natural hosts had used cross-sectional analyses of chronically infected animals, or limited sampling during acute infection and did not observe the initial burst of antiviral responses[12,25,26]. As a result, initial models proposed that innate pathways in the anti-SIV response of natural hosts were silenced or largely reduced. Subsequent studies took advantage of more frequent sampling during very early intervals of infection (3-30 days) and reported three key findings: (i) during acute infection, elevated levels of activation (DR) or proliferation (Ki67) markers on CD4+ T cells in SMs[27] and CD8+ T cells in AGMs[28] and SMs[29]; (ii) transient recruitment of pDCs to lymph nodes during acute SIVagm infection[30]; and (iii) recruitment of cells expressing the activation/exhaustion marker PD1 in lymph nodes after SIVsm infection[31]. Despite these observations, and the assertion by the authors that timely homeostatic mechanisms were at play, the concepts that natural hosts (i) displayed an acute-phase immune response to SIV, and (ii) subsequently resolved early innate responses, were not fully appreciated. In this regard, high-throughput genomics were key, as they showed that early immune activation after SIV infection in natural hosts was not limited to isolated pathways involving the CD8+ T cell response, but was indicative of system-wide activation.

The most striking finding in the microarray screens was the observation that SMs and AGMs exhibited widespread upregulation of ISGs during acute SIV infection, and that, for the majority of ISGs, expression returned to baseline during chronic infection[19,22-24]. In contrast, ISG expression remained elevated in chronic infection in RMs and PTMs (pig-tailed macaques). Furthermore, expression of IFNα protein in the lymph nodes of SIV-infected SMs and AGMs was readily apparent during acute infection[32]. Collectively, these studies indicate that SIV infection of natural hosts induces an intact IFN response during the acute phase of infection that resolves during the transition to chronic infection, while non-natural hosts maintain ISG expression indefinitely[33]. The role of Type I IFN in HIV infection has long been considered to be a double-edged sword; it promotes an antiviral state
in target cells that act to limit viral replication and spread, however it also induces nonspecific activation in lymphocytes, inhibits hematopoietic potential, skews Th differentiation and contributes to T cell apoptosis and activation[34,35]. Similar to SIV-infected macaques, persistent ISG expression has also been demonstrated in CD4+ and CD8+ T cells from patients chronically infected with HIV[36,37], thus, the ability of natural host species to limit the type I Interferon response may be a critical mechanism to avoid long term immune activation.

Recent unpublished work from our group tested the hypothesis that IFNα is capable of driving immune activation in vivo by administering recombinant RM IFNα2 (rmIFNα) to SIV+ SMs weekly for a period of 4 months. Microarray analysis showed induction of ISG expression in blood for the first 21 days, albeit at lower fold-changes compared to those seen in acute SIV infection; by day 42 ISG expression had largely disappeared, coinciding with the appearance of rmIFNα-specific antibodies in sera. However, the period of ISG upregulation/rmIFNα bioactivity yielded several interesting observations: (i) plasma viral loads decreased by approximately one-log (ii) peripheral blood CD8+ T lymphocytes significantly decreased (iii) peripheral CD4+ T cells declined (iv) Ki67+ expression in CD4+ cells and CD4+ memory subsets were elevated, although modestly. While short-lived, the observed transient increase in Ki67+ lymphocytes, even in the face of reduced viremia, suggests that IFNα may be able to contribute to immune activation in natural hosts if left unchecked.

On the horizon - new technologies, new cohorts, new genomes

Microarrays, as applied to the study of natural hosts, are hampered by a few technical limitations: (i) static content (ii) probes specific for RM and/or human sequences (iii) limited dynamic range[38]. RNA-seq technology employs next-generation sequencing platforms for transcriptomic profiling[39]. The study of SIV infection in natural hosts using RNA-seq offers insight beyond microarrays, particularly in identifying sequence variation between orthologous transcripts and splicing differences and will be particularly useful for investigating the role of noncoding RNAs, such as snRNAs, IncRNAs, and regulatory miRNAs. RNA-seq in NHPs has been limited, in part, due to a lack of reference genomes to efficiently identify sequencing reads. To improve the accuracy and efficiency of RNA-seq in primates, a co-operative effort led by the Katze laboratory at the University of Washington, between Illumina Inc. and the primate genetics community, has assembled pooled tissue RNA libraries from 14 NHP species, obtained high density sequencing data and is currently annotating the transcriptomes for reference usage (http://physiology.med.cornell.edu/faculty/mason/lab/styled/). The advent of next generation sequencing technology has substantially reduced the cost of de novo genome sequencing projects such that they can be borne by individual centers. Sequencing of the SM and AGM genomes is currently underway, with details in Table 2. The complete genome sequences of two natural host species will be an invaluable tool for comparative genetics aimed at identifying genes potentially under selective pressure by SIV.

A conceptual challenge in the study of natural hosts is that of translation: comparative differences between SIV-infected natural hosts and macaques do not necessarily represent...
HIV infection. Recently, however, a rare group of HIV-infected patients, termed Viremic Non-progressors (VNP), have been identified that exhibit clinical features similar to those observed in natural hosts: stable CD4+ T cell count, low levels of lymphocyte immune activation despite high plasma viremia[40]. Extensive genetic analysis of five VNPs was recently reported[41]. Interestingly, SMs shared overlapping features: genes upregulated by SIV in SMs (but not RMs) were enriched in VNPs compared to rapid progressors (RPs); further, ISGs identified as being downregulated by SMs were expressed at a higher level in RPs compared to VNPs. This latter finding is of special interest considering that plasma viral loads of VNPs were remarkably higher than RPs (median 5.4 vs 4.7 log10 copies/ml). Thus, a common response between SIV-infected SMs and VNPs is a lowered interferon response. The observation that VNPs and SMs have shared gene expression responses indicates that natural hosts are able to, at least in part, recapitulate immunologic features of lentivirus infection relevant to humans.

Harnessing the potential of systems biology to understand the natural host phenotype

The field of systems biology has been defined in different ways, but is consistently comprised of three activities: (i) acquisition of high-throughput data at multiple biological levels over time or between phenotypic classes; (ii) a computational component to integrate diverse data and make predictive models; (iii) an iterative cycle of local perturbations of the system that validate predictions and refine the model[42-45]. Microarray-based studies of transcriptional changes induced in SMs and AGMs by SIV infection have yielded considerable insight. However, a complete understanding of the molecular underpinnings that regulate the natural host phenotype will require a fully integrated analysis that incorporates the wealth of available clinical, immunological, genetic, and transcriptomic data. In the remaining section, we describe how systems-based approaches can be used to: (i) explore the restriction of SIV infection to ‘expendable’ target cells in natural hosts; (ii) investigate the resolution of immune activation, (iii) identify gene signatures predictive of disease progression.

Viral replication in SMs and AGMs occurs primarily in short-lived cells, i.e. CD4+ T cells, with estimates similar to HIV infection[46-48]. However, evidence indicates that natural hosts have evolved mechanisms to limit SIV entry into the CD4+ T cell pool: reduced levels of CCR5 on target cells[16]; loss-of-function mutations in CCR5 in SMs[49]; CD4null memory subsets capable of TH activity in AGMs[15] and reduced expression of CCR5 on CD4+ TCM in SMs[50]. Because the depletion of the CD4+ TCM pool predicts disease progression in SIV-infected macaques[51] reduction of entry receptors may be a mechanism by which natural hosts shift infection from the TCM pool into other ‘expendable’ targets. However, entry blockade is not the only weapon in the host arsenal to prevent lentiviral infection. Mammalian genomes encode genes for restriction factors, which function to make target cells nonpermissive to productive retroviral infection[52]. During SIV infection, natural and non-natural hosts express an overlapping, but different, array of restriction factors[22]. The prospect that restriction factors may be partially responsible for the observed differences in the pattern of infected cells between natural and nonnatural hosts is intriguing, however testing this hypothesis with conventional experimentation is challenging due to the vast number of potential restriction factors and interspecies diversity. Recently, a
A high-throughput method for screening the activity of restriction factors was described [53]. Restriction factor screening, when integrated with comparative genetics and transcriptomic data from discrete lymphocyte pools, should accelerate discovery of intrinsic antiviral effector molecules contributing to the natural host phenotype.

As described above, an important feature of the natural host phenotype is resolution of the interferon response [33], predominately regulated by pDCs (reviewed in [53]). A number of seminal studies have elegantly used systems-level analyses to model TLR and antiviral responses, such as negative-feedback mechanisms in LPS/TLR4 signaling [54], reviewed by [45]; signaling cascades downstream of TLRs in dendritic cells [55], and influenza infection [56], reviewed in [57]. An example of how similar approaches can be applied to comparatively study SIV interactions with natural hosts aimed at identifying differential regulatory pathways of IFN production is depicted in Figure 1.

The oncology field established proof-of-principle for using transcriptional profiles to distinguish biological classes (e.g. clinically distinct subtypes of leukemias) over a decade ago [58, 59]. Recently, similar molecular classification strategies have been adopted in the nascent field of ‘systems vaccinology’ [60-62] to identify gene signatures predictive of the immunogenicity of well-established vaccines [63-65]. An advantage of this approach is that predictive models are built from direct associations between transcript measurements and phenotypes of interest; disadvantages are: (i) robust prediction algorithms require large numbers of biological replicates separated into independent ‘training’ and ‘validation’ groups [66]; (ii) identified classifier genes, although predictive, may not necessarily regulate the biological process under study. For biological discovery in the natural host model, the quantity of animals required may reduce its feasibility compared to conventional gene expression profiling experiments. Instead, molecular discrimination techniques may have greater utility in identifying early marker gene signatures capable of predicting clinical progression. To this aim, the VNP phenotype will likely be an instructive model, and further establishment of this cohort is warranted.

Conclusion

High throughput genomics have allowed unprecedented insight into the immunological interplay between SIV and natural host species in vivo. Gene expression profiling studies from multiple groups have led to the clear proof that African natural hosts mount an intact Type I interferon response during acute infection. A key difference with pathogenic HIV/SIV infections is that they only show a transient response, whereas interferon expression persists in infected Asian macaques and humans indefinitely. Ongoing developments in the natural host field include the genome sequencing of SMs and AGMs, development of NHP reference genome resources for RNA-seq, and the establishment of a novel cohort of HIV infected patients, VNPs, in which the natural host phenotypes is recapitulated. Chronic immune activation remains an important cause of disease in HIV infection, even in patients with good virologic responses to antiretroviral therapy. The use of system-based analyses to uncover pathways by which natural hosts avoid disease will hopefully identify novel targets for ameliorating HIV-associated immune activation.
Acknowledgments

The authors would like to thank Dr. Donald Sodora and Dr. Zach Johnson for critical review of this manuscript. The authors apologize for the absence or abbreviated discussion of all relevant publications due to space limitations. This work was supported in part by the National Institutes of Health (R37-AI66998 to GS), Agence Nationale des Recherches sur le SIDA et les hépatites virales (ANRS, to MMT, AB), the Institut Pasteur (MMT), the Institut des Hautes Études Scientifiques, the Genopole Evry (AB) and the AREVA Foundation.

Abbreviations

AGM     African Green monkey
PTM     pig-tailed macaque
SIV     simian immunodeficiency virus
IFN     interferon
ISG     interferon stimulated genes
VNP     viremic non-progressor
RM      rhesus macaque
T_{CM}  central memory T cell
T_{EM}  effector memory T cell
SM      sooty mangabey

References


19. Favre D, Lederer S, Kanwar B, Ma ZM, Proll S, Kasakow Z, Mold J, Swainson L, Barbour JD, Baskin CR, et al. Critical loss of the balance between Th17 and T regulatory cell populations in pathogenic SIV infection. PLoS Pathog. 2009; 5:e1000295. [PubMed: 19214220] ** This study is one of two studies that were the first to characterize genome wide expression in natural hosts during in vivo SIV infection. The results are framed in the context of Th17 dynamics during acute SIV infection in RMs and AGMs.


23. Lederer S, Favre D, Walters KA, Proll S, Kanwar B, Kasakow Z, Baskin CR, Palermo R, McCune JM, Katze MG. **Transcriptional profiling in pathogenic and non-pathogenic SIV infections revealed significant distinctions in kinetics and tissue compartmentalization. PLoS Pathog. 2009; 5:e1000296. [PubMed: 19214219] ** This study one of two articles (Favre 2009) that the first to characterize genome wide expression during in vivo SIV infection of natural hosts. It demonstrated that ISGs are downregulated as early as 45 days post-infection in AGMs.


This article conducted an exhaustive characterization of transcriptomic changes in purified CD4+ cells during SIV infection of African green monkeys. The data provide the proof of an innate response during acute SIVagm infection.


47. Gord www.bosingerlab.com

50. Paiardini M, Cervasi B, Reyes-Aviles E, Micci L, Ortiz AM, Chahroudi A, Vinton C, Gordon SN, Bosinger SE, Francella N, et al. Low levels of SIV infection in sooty mangabey central memory CD T cells are associated with limited CCR5 expression. Nat Med. 2011; 17:830–836. [PubMed: 21706028] ** This study demonstrated that during activation and transition to central memory cells, CD4+ lymphocytes in sooty mangabeys have lower expression of CCR5, in contrast to rhesus macaques, and that the ratio of SIV infected Tcm/Tem CD4+ T cells in vivo was lower in sooty mangabeys than in rhesus macaques. Thus natural hosts may maintain a CD4+ T cell pool by restricting virus entry in Tcm cells.


Keypoints

- The immunological mechanisms of pathological immune activation in AIDS pathogenesis have recently been studied by using high-throughput gene expression profiling to compare pathogenic simian immunodeficiency virus (SIV) infection in RMs and nonpathogenic infection of sooty mangabeys and African green monkeys.
- Genome-wide expression profiling identified key difference in the innate immune response to SIV between pathogenic and natural hosts was strong induction of interferon stimulated genes (ISGs), which persisted indefinitely in pathogenic infection, but resolved to baseline rapidly in natural hosts.
- A novel cohort HIV-infected patients, termed Viremic Nonprogressors, have been described that share immunological features with natural host species, including preserved CD4+ T cell counts, low immune activation, and lowered ISG expression in CD4+ and CD8+ lymphocytes.
- Ongoing work to develop resources for the study of SIV infection in natural hosts include the sequencing of the genome of SMs and AGMs, and the development of reference transcriptomes for RNA-seq transcript analysis.
pDCs have been implicated as the predominant source of interferon production in lymph nodes of acutely infected SMs and AGMs [32]. (A) Transcriptional profiles of pDCs from natural and non-natural hosts activated over time with endogenous strains of SIV can be used to identify in a longitudinal analysis waves of genes expression and early activating ‘afferent’ genes (i.e. signaling molecules; transcription factors) from ‘late’ effector genes. The ability of pDCs to suppress interferon expression despite chronic viremia suggest an epigenetic mechanism, and deep sequencing can be used to profile epigenetic modifications in pDCs before and after SIV exposure. The genome sequence of AGMs and SMs can be leveraged against epigenetic and transcriptome data to probe for species specific variation in the regulatory regions of genes with differential epigenetic/transcriptional profiles. (B) Importantly, potential candidates for negative feedback of interferon signaling in natural host species may not be readily identified as differential on highthroughput screens. Detailed reconstruction of gene networks, boolean and enrichment analysis of genes with sequence variation/differential behavior between natural and pathogenic hosts, kinetic modeling of transcription and correlation with IFNα/β production can be integrated to predict potential candidates and map biologically meaningful divergence in innate signaling. (C) Prioritized candidates can be perturbed in vitro using gene-knockdown techniques and assayed for their impact on prolonged interferon production. Candidates can also be examined using comparative in vivo study of SIV infected NHPs.
Table 1

Large scale microarray analyses characterizing in vivo SIV infection of natural host species.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Natural Host Species &amp; Virus Strain</th>
<th>Pathogenic Species &amp; Virus Strain</th>
<th>Tissue Array Platform # chips</th>
<th>Unique Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lederer et al. PLoS Pathog 2009 [23]</td>
<td>AGM C. sabeus SIVagm.92018</td>
<td>PTM M. nemestrina SIVagm.92018</td>
<td>Blood LN Colon 90</td>
<td>Same strain of virus for both species; includes colon; monitors Th17 cells in gut</td>
</tr>
<tr>
<td>Jacquelin et al. J Clin Invest 2009</td>
<td>AGM C. sabeus SIVagm.92018</td>
<td>RM M. mulatta SIVmac251</td>
<td>PB CD4 LN CD4 201</td>
<td>Transcriptomes from isolated CD4+ T cells; extensive longitudinal sampling (&gt;1 yr)</td>
</tr>
<tr>
<td>Bosinger et al. J Clin Invest 2009 [22]</td>
<td>SM C. alysi SIVsmt</td>
<td>RM M. mulatta SIVsmt &amp; SIVmac239</td>
<td>Bshed LN 120</td>
<td>Used sooty mangabey; included infected RM cohorts of high and low pathogenicity; Same strain of virus for natural and pathogenic species</td>
</tr>
</tbody>
</table>
### Table 2

Genome sequencing projects of SIV natural host species.

<table>
<thead>
<tr>
<th>Natural Host</th>
<th>Species</th>
<th>Centers</th>
<th>Sequencing Platforms</th>
<th>Details</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM</td>
<td>C. Atys</td>
<td>Baylor Human Genome Sequencing Center, Yerkes National Primate Research Center</td>
<td>Pacific Biosciences, Illumina</td>
<td>80x average coverage; Reference transcriptomes of multiple tissues</td>
<td>n/a</td>
</tr>
</tbody>
</table>