ASSOCIATION OF INTERFERON REGULATORY FACTOR-1 POLYMORPHISMS WITH RESISTANCE TO INFECTION BY HIV-1 IN KENYAN FEMALE SEX WORKERS

BY

HEZHAO JI

A Thesis
Submitted to the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

Department of Medical Microbiology
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

Association of Interferon Regulatory Factor-1 Polymorphisms with Resistance to Infection by HIV-1 in Kenyan Female Sex Workers

by

Hezhao Ji

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree of

Doctor of Philosophy

Abstract

Resistance to Human Immunodeficiency Virus-1 (HIV-1) infection has been observed in multiple cohort studies including a female sex worker cohort established in Nairobi, Kenya. Both host genetic variations and HIV-1 specific immune responses have been described as important components correlating with resistance to HIV-1. Interferon regulatory factor 1 (IRF-1), a promoter to host immunity and also a transactivator of HIV-1, has a dual role to play in HIV/AIDS.

This study reconfirmed our previous finding that 179 allele at microsatellite (MS) region in IRF-1 correlated with HIV-1-resistant phenotype in this cohort. Upon this, by near-complete gene sequencing, we demonstrated that the IRF-1 gene in Kenyan cohort was highly polymorphic. Fifty-three single nucleotide polymorphisms (SNP) (26 newly identified), 2 insertions and 1 deletion mutation were identified. We identified 35 consistent discrepancies between IRF-1 GenBank sequences and our population-based sequencing data, suggesting that the current GenBank reference sequences for IRF-1 is incomplete. The sequence of IRF-1 gene and its upstream promoter region was re-established (GenBank: DQ789232).

Statistical analysis revealed that, together with 179 allele at the MS region, SNPs at 619 (A/C) and 6516 (G/T) were also significantly correlated with HIV-1 resistance. Despite their intronic locations, further functional investigations revealed that PBMCs from subjects with protective IRF-1 genotypes/haplotypes showed: (1) significantly depressed IRF-1 protein expression at both a basal expression level and in response to exogenous IFN-γ stimulation or infection by an artificially pseudotyped HIV-1 construct;

(2) significantly increased chances of skipping exon 2 and 3 which contain the IRF-1 start codon as well as encode the DNA binding domain of IRF-1 protein; (3) significantly decreased frequencies of skipping exon 7 and 8; (4) A clear trend of decreased efficiency of HIV-1 replication early upon infection by HIV-1, suggestive of that these cells are less competent in supporting HIV-1 replication.

This study suggests that different IRF-1 genotypes/haplotypes and consequently altered expression and function of IRF-1 likely compose a significant determinant for the heterogeneity of susceptibility to HIV-1 in our Kenyan subject cohort. This study provides insight to natural immunity to HIV-1 infection and suggests that effective anti-HIV-1 strategies should target not only host immunity, but also factors important in the establishment of HIV-1 infection.

Dedication

More than any other, I thank my wife, Xuefen Yang, and my two daughters, Jiakai and Jiaqi, for their love, understanding and great support for my study and work. Just because of me, Xuefen gave up her career as a gynecologist/Obstetrician in China and came to Canada with my daughter, Jiakai. I will never forget those days you worked in Winnipeg even as a sewing machine operator for 11 months! I would say half of my achievement should belong to you, my love. Without your support, I cannot even imagine how possible it could be for me to pursue a Ph.D degree with a second language in a foreign country!

We did it, together!

Acknowledgements

First of all, I thank my supervisor, Dr. Frank Plummer, for his guidance and great support for this project. I want to say that I have learned a lot from you regarding to not only dedicating to science but also becoming a scientist with deep insight in both my own HIV/AIDS research field and also a global vision of the infectious disease control. You also showed me how to raise and answer scientific questions with scientific thinking, which is the key to be a successful scientist. From you, I also witnessed how important a good personality is in becoming a successful scientist and scientific supervisor/director.

I give my special thank to Dr. Blake Ball, my daily supervisor. Without his previous hard work and important findings, this project wouldn't have been even established. I thank him for his generous help, dedicated daily guidance, scientific input, and great friendship, all of which contributed greatly to the accomplishment of this project.

I would also like to thank my advisory committee, Drs. Keith Fowke, Xi Yang and Aaron Marshall, and Drs. Xiaojian Yao and Ma Luo for their significant scientific input and constructive comments and suggestions invaluable for the accomplishment of this project and the success of myself. Special thank to Dr. Xiaojian Yao for his unique plasmid constructs which constitute a major tool for the functional assays in this project. Also I will never forget the great friends and colleagues I worked with during this 5 years' study in the University of Manitoba including Dr. Xiaojuan Mao, Jill Waruk, Shezhad Iqbal, Lyle Mckinnon, Paul Maclaren, Zhujun Ao, Ben Liang, Julius Oyugi,

Crystal Marlin, John Schelenberg, John Rutherford, Leslie Slaney, Sue Ramdahin, Carol Sigurdson, Lynn Gauthier and all other members in our HIV research group. It is really a wonderful and entertaining experience to work with all of you. Thank you all for your invaluable friendship and unselfish help! I really appreciate it! Thanks to Mary Cheang for all her assistance in statistical analysis. I also want to thank the department administrative support, Dr. Joanne Embree, Angela Nelson, Jude Zieske, Sharon Tardi, and Robin Scott, among others who kept me on track.

I also want to say thanks to all my colleagues working at Nairobi side in Majengo Clinic, Drs. Joshua Kimani, Charles Wachihi and Richard Lester, Jane Njoki, Jane Kamene, Elizabeth Bwibo and Edith Amatiwa. Although I didn't working with them directly, I know that I wouldn't be able to finish this project without their support from behind the scene.

I also want to send my great appreciation to all women who participated in ML research cohort for their cooperation, dedication and invaluable contribution to our HIV-1 resistance research, without which this project would never be accomplished.

I thank ICID/CIHR National Training Program and MHRC for offering me the multidisciplinary training and financial support during my study and NIH and CIHR for funding this project.

Table of Contents

List of Figures

List of Tables

Introduction

Overview

Since the first case of acquired immunodeficiency syndrome (AIDS) was reported in 1981 (Gottlieb et al., 1981) and human immunodeficiency virus type-1 (HIV-1) was identified as the causative organism for this lethal disease in 1983 to 1984 (Barre-Sinoussi et al., 1983; Popovic et al., 1984), HIV/AIDS has spread dramatically around the world without any sign of abatement. With 40.3 million people living with HIV/AIDS and 4.9 million newly infections alone in 2005 (UNAIDS/WHO, 2005), the AIDS pandemic has inarguably become one of the most dreadful disease the global community has had to confront over decades. With a dramatic decrease of life expectancy, high mortality rates in infant infection, severe labor loss and huge financial burden for the family and the affected society, HIV/AIDS has gone far beyond a health crisis to become a social, economical and political issue in both HIV affected regions and the world.

After more than two decades of unprecedented and intensive research efforts, over 200,000 publications related to HIV/AIDS have been generated. We have obtained a better understanding of HIV-1 retrovirology, HIV-1 pathogenesis, the clinical characteristics of HIV/AIDS, characteristics of the HIV epidemic regionally and globally and, perhaps most importantly for vaccine development efforts, a better understanding of host immunity against HIV-1. However, there are many more unanswered questions remaining mainly including exact pathogenesis of HIV-1 infection and effective anti-HIV-1 immune mechanisms capable of containing/sterilizing HIV-1, which compose the

hindrance of the development of efficient prophylactic and therapeutic strategies. The relentless spread of HIV-1 is still escalating especially in Africa, Eastern Europe, Central and East Asia (UNAIDS/WHO, 2005). The joint United Nations Program on HIV/AIDS has forecasted that from now until 2010 we will see 5 million new infections every year and actually a similar steady increasing rate has been seen for nearly 10 years so far. Although not curative, the development of antiretroviral therapy (ART) has dramatically decreased the morbidity and mortality due to HIV infection (Burger et al., 2004; Palella et al., 1998) (Walensky et al., 2006) and mother-to-child transmission of HIV-1 in developed countries (Soorapanth et al., 2006). But the reality is that 95% of HIV infections are occurring in developing countries instead where ART remains inaccessible to more than 93% of HIV-1 infected patients (Esparza, 2005). As has been demonstrated in many other infectious diseases, a highly efficient and inexpensive HIV-1 vaccine is believed to be the best long term solution for bringing the HIV/AIDS pandemic under control. Despite the enormous research efforts and funds that have been devoted to HIV-1 vaccine development, an effective vaccine for HIV-1 remains an unrealized goal even after 25 years' exploration (Esparza, 2005). Two of the main hindrances in developing efficient HIV-1 therapeutic and prophylactic strategies are an incomplete understanding of what the exact mechanisms of pathogenesis of HIV/AIDS and a lack of knowledge of the mechanisms are capable of protecting the host from HIV-1 infection or disease progression.

Although HIV-1 is highly pathogenic, it is noteworthy that not all individuals exposed to HIV-1 become infected and progress to AIDS. There is a growing body of evidence indicating the existence of natural resistance to HIV-1 infection (Fowke et al.,

1996; Koning et al., 2004; Kulkarni et al., 2003; Plummer et al., 1999; Shacklett, 2006; Shearer and Clerici, 1996). The presence of HIV-1 highly exposed but uninfected population not only is an evidence for the existence of natural resistance against HIV-1 infection, but also provides an invaluable model system for the exploration of protective mechanisms against HIV-1 infection. The observation of the natural resistance to smallpox in milkmaids had lead to the generation of vaccinia vaccine which contributed the most to the global eradication of smallpox. A significant breakthrough in HIV-1/AIDS research may be obtained by focusing on the study of populations showing altered susceptibility to HIV-1 infection. Understanding the correlates of natural protection against HIV-1 will provide invaluable information and help to develop effective strategies to contain the HIV/AIDS pandemic.

Studies on varied susceptibility to HIV-1 have mainly focused on the effects of the innate and adaptive host immune response and host genetics. In studies of host immunity, known aspects of potential antiviral immunity have been investigated extensively, including innate immunity, CD4 T helper responses, CD8 viral specific cytotoxic T cell responses, humoral immune responses and also mucosal immune responses. It is believed that the HIV-1 exposed uninfected subjects have effective innate and adaptive immune responses to HIV-1 despite of their uninfected status. But why these protective responses develop only in this small subset of population and how they are maintained remains elusive. Researchers have devoted enormous attention to genetic studies as well with the hypothesis that these individuals are predisposed to be less susceptible to HIV-1 due to their unique genetic compositions. Not surprisingly, the main targets for these studies have been genetic variations in natural HIV-1 receptors and co-receptors, which are

responsible for the viral attachment and entry into the target cells, as well as the extremely polymorphic human leukocyte antigens (HLA) which is responsible for the generation and regulation of efficient adaptive immune responses. Although numbers of publications have been generated from studies on HIV-1 exposed uninfected subjects and some interesting findings been disclosed, none of these findings provide a satisfying interpretation of altered susceptibility applicable to all different subject cohorts. This suggests that HIV-1 resistance is a complicated multifactorial phenomenon and that the potential determinant (s) for HIV-1 resistance will not likely be a single immune or genetic factor, but rather a comprehensive network involving both the host genetic background and host immunity and may be related to the infecting virus itself as well.

Discovery of HIV as the causative agent of AIDS

In 1981, a cluster of five otherwise healthy men were diagnosed with *Pneumocystis Carinii* Pneumonia (PCP) and reported to the Centers for Disease Control in Atlanta, USA (1982). By 1982, several new cases of PCP were reported as well as other conditions that were normally associated with severe immunodeficiency (Masur et al., 1981). This condition was later termed as acquired immune deficiency syndrome or AIDS and notably it was affecting gay men, intravenous drug users and individuals with frequent blood transfusions such as hemophiliacs. AIDS was later identified in other groups as well and the evidence from these cases strongly indicated an infectious causative agent might be responsible for this novel disease (Chamberland et al., 1984; Kreiss et al., 1985).

The first isolation of a possible causative agent for AIDS occurred in 1983. A human retrovirus similar to human T-cell leukemia virus (HTLV) was isolated from the lymph node of a man suffering from AIDS (Barre-Sinoussi et al., 1983). Concomitantly, a similar virus from another AIDS patient was isolated by a separate research group (Gallo et al., 1983). These groups studied their isolates separately and soon discovered that they were not in fact HTLV but a virus very similar to HTLV. Several names had been given to these new viruses including HTLV III, Lymphadenopathy Associated Virus (LAV) and AIDS Related Virus (ARV) depending on the research group involved. In 1986, the virus was recognized as novel and eventually named as Human Immunodeficiency Virus (HIV) and classified as a member of the lentivirus group (Coffin et al., 1986). In 1985, a second separate subtype was identified in Cameroon, Western Africa (Clavel et al., 1986) which was transmitted in much the same way as the initial isolates but was less virulent and had a longer disease progression and less severe disease course. This subtype was later named HIV type-2 with the original etiological agent of the North American and European AIDS being named HIV type-1.

Although the AIDS epidemic first recognized in the early eighties, the oldest verified HIV-1 infection case dates back to 1959 in Zaire. An HIV positive blood sample from an anonymous man was discovered in the blood archives of a Zairian (which was then called Congo) STD clinic in Kinshasa and analyzed in 1998 (Nahmias et al., 1986; Zhu et al., 1998). There is also a possible case dating back to 1934 that was suspected, but has not been verified due to the lack of tissue and/or blood samples. This was to some extent verified by a phylogenetic statistical analysis of the evolution of the retroviral genome of HIV conducted in 2000 using complex mathematical modeling (Korber et al., 2000). This

computational model correctly placed the genome of the 1959 case on the evolutionary tree and estimated that the most reliable time of the origin of the disease in human should be around 1931±15 (Korber et al., 2000). There is now convincing evidence suggesting that both HIV-1 and HIV-2 had passed to humans from chimpanzees who harbored a related simian immunodeficiency virus (SIV) (Gao et al., 1999). SIVs have been identified in no fewer than 36 different nonhuman primate species in sub-Saharan Africa; however, only two SIV strains, SIVcpz from chimpanzees (*Pan troglodytes troglodytes*) and SIVsmm from sooty mangabeys (*Cercocebus atys atys*), are known to have crossed the species barrier and believed to be the origins of HIV-1 and HIV-2 respectively based on genetic analysis (Hahn et al., 2000; Santiago et al., 2005). A recent report of detecting SIVcpz antibodies and nucleic acids in fecal samples from wild living *Pan troglodytes troglodytes* in southern Cameroon and further phylogenetic analyses on these SIVcpz strains and HIV-1 further affirmed that SIVcpz is the ancestor of HIV-1 which crossed species barrier and resulted in the pandemic of AIDS(Keele et al., 2006).

HIV-1 retrovirology

HIV-1 is a retrovirus and belongs to the lentivirus subfamily. It is an enveloped virus with double stranded RNA genome. The HIV-1 genome is made of 9749 base pairs (bp) and contains 3 main genes encoding HIV-1 structural proteins or enzymes necessary for HIV-1 replication, named **gag, pol** and **env,** 2 genes encoding regulatory proteins essential for HIV-1 replication are named **tat** and **rev**, and 4 genes encoding accessory proteins are named **vif, nef, vpr** and **vpu** (Levy, 1993).

The **gag** gene is approximately 2000 bp in length and encodes several internal structural proteins including the matrix protein P17, the capsid protein P24, the nuclear protein P9 and protein P6 whose function remains unclear. The **env** gene includes approximately 1800 bp and encodes the surface proteins gp120 and gp41 that protrude from the lipid envelope and mediate the viral attachment to target cells via binding to corresponding receptors on the surface of target cells. The **Pol** gene is about 2900 bp and encodes three viral enzymes including proteinase (**Pro**), reverse transcriptase (**RT**) and integrase (**IN**), which are all required for HIV-1 replication. Two regulatory proteins, the transactivator protein (**Tat**) and regulator of expression of virus protein (**Rev**) are encoded by **tat** and **rev** genes respectively. Another four accessory proteins, virus infectivity factor (**Vif**), misnamed negative regulator factor, but really an enhancing regulatory factor (**Nef**), virus protein R (**Vpr**), and virus protein U (**Vpu**) are encoded by the named genes.

HIV-1 exhibits an extraordinarily high genetic diversity, which derives mainly from high mutation and genetic recombination rates. HIV-1 variants are divided into three groups: M (major), N (non-M, non-O), and O (outlier). Group M, which is responsible for the global pandemic, is by far the largest and most common group and it is then subdivided into at least nine subtypes or clades based on phylogenetic analyses of full-length genomes: A, B, C, D, F, G, H, J, and K. Different clades are distributed unevenly in different geographic regions around the world. The A-clade is prevalent in Eastern Africa, Eastern Europe and central Asia. The B-clade is dominant in the Americas, Western Europe, Australia and Japan. Clades C is dominant in South and Eastern Africa and Asia and D is dominant in Africa (Geretti, 2006). Group O contains about thirty

subtypes found mainly in West African countries such as Cameroon and Gabon. In the September 1, 1998 issue of Nature Medicine, F. Simon announced the discovery of a variant of HIV-1 that fits neither the M nor O group (Simon et al., 1998). It seems to fall between the M-group and the simian immunodeficiency virus (SIV) and classified as N-group.

Recombinant strains originated from co-infection of different but related strains of HIV-1 have also been noted. When viruses of different subtypes co-infect the same cell and exchange their genetic materials, a new hybrid virus can be created (Burke, 1997). Many of these new strains do not survive for long, but those that are transmissible and able to infect more than one person are known as "circulating recombinant forms" or CRFs, of which at least 19 strains are currently identified(Casado et al., 2005; Robertson et al., 2000). The 8th Conference on Retroviruses and Opportunistic Infections in 2001 saw the publication indicating that as many as 14% of new infections are recombinant.

HIV-1 life cycle

HIV-1 follows a typical retrovirus life cycle as it infects target cells and reproduces. Without any intervention, a series of successive steps are followed in fulfilling the infection and replication of HIV-1. The initial step is attachment in which the virion adheres to the target cell and this is mediated through the interaction between viral envelope glycoproteins (gp120 and gp41) and receptor molecules (CD4, CCR5, CXCR4) expressed on the surface of target cells. The upregulation of these molecules makes activated T cells the main targets to HIV-1 infection. The second step is fusion and

penetration in which the viral envelope fuses with the host cell membrane and viral entry follows. Significant conformational changes occur in all involved molecules during this process. After entry into the target cell, the virion uncoats and viral genomic RNA core is released into cytoplasm. Single stranded viral RNA (ssRNA) is then reverse transcribed into single stranded DNA (ssDNA) by the viral enzyme RT. Host DNA synthesis system is mainly exploited to convert the viral ssDNA into dsDNA. With the aid of the viral RT enzyme which can function not only as RNA dependent DNA polymerase for synthesis of ssDNA out of ssRNA, but also acts as an RNase H to remove RNA primers during second strand cDNA synthesis. The dsDNA migrates into the nucleus of the cell and then randomly integrated into the host genome. This process is mediated mainly by the viral enzyme integrase. The integrated DNA form of the virus is also called a provirus. Once integrated into the host cell genome, the host cell is henceforth converted into the machinery for further viral production. The further transcription of HIV-1 will mainly be dependent on the complex interplay between viral and host transcription factors. The infected cells could remain dormant (non-replicating) which lead to a latent infection and this can be observed in infected resting CD4+ T cells. Otherwise it could immediately begin producing more viral RNA which occurs in pre-activated infected cells of which multiple host transcriptional factors are upregulated. The newly synthesized genomic viral RNA migrates out of the nucleus and gathers near the inner surface of the cell membrane. Meanwhile, viral mRNA undergoes protein synthesis again utilizing the host translational machinery. The long viral proteins are then cleaved into smaller functional pieces by viral protease. Viral assembly occurs as RNA packaging starts near the inner surface of cell membrane. Virion release is accomplished by either budding or cell lysis.

It is noteworthy that a high mutation rate and therefore sequence variation is common during HIV-1 replication due to the lacking of proof-reading during viral transcription with an error rate at 1/10Kb. This causes a high and unpredictable viral variability especially on the antigenicity of the envelope glycoproteins (i.e. gp120) which facilitates the immune evasion from HIV-1 epitope specific cytotoxic T cell and neutralizing antibody responses and composes the main hindrance of the generation of persistent and cross reactive antiviral immunity and the development of effective prophylactic strategies such as vaccine design.

Regulatory role of HIV-1 LTR in viral replication

Following the integration of the HIV-1 gene into the host genome as provirus DNA, HIV-1 replication requires a complex interaction between viral and host factors to initiate HIV-1 transcription and then translation in order to produce progeny virus. HIV-1 replication is mainly controlled at transcription level through intricate interaction between viral and cellular trans-regulatory proteins and regulatory sequences located in the long terminal repeat regions (LTR) at both ends of HIV-1 proviral genome (Cullen, 1991; Tang et al., 1999). HIV-1 LTR consists of a redundant (R) sequence and two unique (U) sequences, U5 and U3, found at the 5' and 3' ends, respectively. An important modification as a result of reverse transcription and integration is the duplication of U5 and U3 sequences, such that the provirus contains the sequence U3-R-U5 at both termini. Transcription initiates at the 5' U3/R junction and proceeds through to the polyadenylation (polyA) signal located in the 3' R repeat (Al-Harthi and Roebuck, 1998).

The transcriptional regulatory elements in HIV-1 LTR are located in both upstream and downstream of the transcription start site. The upstream elements are located within the U3 region, mainly containing NRE (negative regulatory element), NF-κB (nuclear factor κB) and Sp1 binding sites. The transcription response (TAR) element is located at the beginning of R region and binds to viral transactivator Tat, which is potent activator of HIV transcription and functions as a transcription factor to increase the number of full-length transcripts(Jones, 1993). The downstream elements are positioned within the R and U5 regions which contains binding regions for AP1 (activator protein 1), Sp1, NF-AT (nuclear factor of activated T cells) and an IRF (interferon regulatory factor) region (Al-Harthi and Roebuck, 1998; Tang et al., 1999). All these sites are strongly conserved and play a critical role in HIV-1 transcription and replication. The NRE negatively impacts transcription, as the loss of this region results in an increase in gene expression. The NRE has numerous candidate protein-binding sites including TCF-1, USF, NF-AT, ILF-1, GRE, AP-1, COUP-TF, RAR, NRT1/2, T cell factor B, myb, and GATA3. The NF-κB proteins, the major regulators of host gene activation, also activate and synergize with Tat in transcriptional activation of HIV-1 and enhance HIV-1 replication in infected cells. Both Sp1 and AP-1 also function as enhancers of HIV-1 transcription (Tang et al., 1999). The IRF region spans nt +200 to 217 (TTGAAAGCGAAAGGGAAACC) and is homologous to the interferon stimulated response element (ISRE) sequence present in the promoter of interferon stimulated genes (ISGs) which members of interferon regulatory factor family can interact with (Harada et al., 1989) . It has been shown that this sequence is a binding site for IRF proteins in particular IRF-1 and IRF-2. By binding to this

specific sequence, IRF-1 exerts important impact on HIV-1 replication as well (as discussed in following section) (Sgarbanti et al., 2002) .

HIV transmission

AIDS is not as contagious as many other common infectious diseases. HIV-1 is not an airborne virus and cannot be transmitted through casual contact, insect vectors, rarely by kissing or urine. Horizontal transmission is accomplished mainly through sexual contact, exchange of blood, blood products and other body fluids or tissues from infected individuals encompassing both medical procedures and intravenous drug use. Perinatal transmission refers to the transmission routes from HIV-1 infected mother to her child (Burns and Mofenson, 1999; Signorelli and Joseph, 1998).

Heterosexual transmission of HIV-1 is now the main route and it counts approximately 80% of current HIV-1 infections in adults (Quinn and Overbaugh, 2005). Sexual transmission occurs with the contact between sexual secretions of one partner with the rectal, genital or oral mucous membranes of another. The rate of heterosexual HIV-1 transmission per coital act follows a U-shaped curve, being highest during the post-seroconversion period, lower during latency, and increasing with advancing disease(Wawer et al., 2005). There is a direct relationship between viral load in peripheral blood and HIV-1 sexual transmission with a cutoff at approximately 1,500 copies of viral RNA per milliliter of blood (Quinn et al., 2000). Although an undetectable plasma viral load does not necessarily indicate a low viral load in the seminal liquid or genital secretions(Coombs et al., 2003), each 10-fold increment of seminal HIV RNA is

associated with an 81% increased rate of HIV transmission (Laga et al., 1991; Tovanabutra et al., 2002). Women are more susceptible to HIV-1 infection due to hormonal changes, vaginal microbial ecology and physiology, and a higher prevalence of sexually transmitted diseases (Lavreys et al., 2004a; Lavreys et al., 2004b; Sagar et al., 2004). People who are infected with HIV-1 can be re-infected by a heterologous HIV strain and this is defined as superinfection (Smith et al., 2005). Cofactors contributing to the increased susceptibility to HIV-1 sexual transmission include: concomitant sexually transmitted diseases (Fleming and Wasserheit, 1999; Plummer, 1998; Royce et al., 1997), genital tract trauma (Laga et al., 1993; Mehendale et al., 1995; Plummer et al., 1991; Simonsen et al., 1988), lack of circumcision in men (Moses et al., 1999), presence of bacterial vaginosis and altered vaginal flora (Taha et al., 1998), and contraceptive methods (Royce et al., 1997), although the exact roles of many of these cofactors in HIV-1 transmission remain unclear. Sexually transmitted diseases are believed to be the strongest contributing factors affecting the risk of HIV transmission and infection because they cause the disruption of the normal epithelial barrier by genital ulceration and/or microulceration; and by accumulation of pools of HIV-susceptible or HIV-infected cells (lymphocytes and macrophages) in semen and vaginal secretions. Epidemiological studies from sub-Saharan Africa, Europe and North America have suggested that there is approximately a four times greater risk of becoming HIV-infected in the presence of a genital ulcer such as caused by syphilis and/or chancroid. There is also a significant increased risk of HIV acquisition in the presence of STIs such as gonorrhea, chlamydial infection and trichomoniasis which cause local accumulations of lymphocytes and macrophages (Laga et al., 1991).

In sexual transmission model, it is believed that HIV-1 infection follows a limited and set pattern. The viral inoculum first encounter mucosal barriers, including the intact epithelial lining, mucus trapping and antiviral chemicals inactivation from mucosal secretions (Shattock and Moore, 2003). Only a small portion of inoculum survives the mucosal screening and manages to cross the mucosal barrier and become founder viruses (Pope and Haase, 2003). The screening process occurs within hours after the inoculation/exposure. Following this step is the primary establishment of HIV-1 infection in local susceptible target cells, including activated $CD4^+$ T cells, macrophages and DCs, converting them into "founder" infected cells and also the machinery to generate progeny virus, the virus becomes established at the point of entry. This process is thought to take hours to days. The contact of HIV-1 and "founder" target cells in sexual transmission might occur via one of three routes: a) transcytosis of virus through columnar cells in genital tract to immune cells underneath; b) binding of virus to DCs on the surface and transferred to lymphoid tissues; c) infection of susceptible macrophages and/or T cells directly on the surface. The virus and infected cells are then disseminated through lymph and later bloodstream to draining lymph nodes and subsequently to other lymphoid tissues such as gut-associated lymphoid tissue (GALT), where high density of target cells reside, the infection magnifies and systemic infection is then well established. This process covers the first several weeks of HIV-1 infection (Haase, 2005). The "fierce" battle between host immunity and virus lasts from commencement of the infection to all through the infection period. When the host immunity is strong, the host remains clinically healthy and $CD4^+$ cell counts remain steady or in slight decline but still fairly capable of maintaining their immune functions. This latency period normally spans years

after the initial stage of infection. Eventually, the host immune system succumbs to HIV-1 and slow but continued depletion of $CD4^+$ T cells occurs nearly in all infected patients and eventually progress to AIDS, characterized by CD4+ cell count drops to below 200 cells/ μl and the occurrence of frequent and fatal opportunistic infections or malignancies.

Transmission through blood, blood products and other infected body fluid exchange route is particularly important for hemophiliacs and recipients of blood transfusions and blood products and intravenous drug users (Donegan et al., 1990; Kaplan and Heimer, 1995). Sharing and reusing syringes contaminated with HIV-infected blood represents a major risk for infection with not only HIV-1, but also other blood-borne pathogens such as hepatitis B and hepatitis C viruses. Needle sharing is the cause of one third of all new HIV-infections and 50% of hepatitis C infections in Northern America, China and Eastern Europe. The risk of being infected with HIV-1 from a single prick with a needle that has been used on an HIV-1 infected person though is thought to be about 1 in 150 (Hung Fan, 2004). HIV-1 infection also rarely occurs after accidental exposure to HIV-1 by piercing the skin with contaminated instruments in medical procedures or ear-piercing, tattooing and acupuncture (Clerici et al., 1994b; Pinto et al., 1995).

Mother to child transmission of HIV-1 from infected mother to the fetus or infants occurs via placenta during pregnancy, during delivery or through breastfeeding (Burns and Mofenson, 1999). In the absence of treatment, the transmission rate between the mother to the child during pregnancy, labor and delivery is 25%. However, when the mother has access to antiretroviral therapy and gives birth by caesarean section, the chance of transmission is just about 1% (Coovadia, 2004; Sperling et al., 1996).

HIV-1 pathogenesis

Generally, HIV-1 infects cells that are arrested in the cell cycle either through cell activation or exposure to other factors resulting in "partial activation" and pushing cells out of G0 of the cell cycle. HIV-1 cannot infect naive quiescent CD4+ T cells or monocytes isolated from peripheral blood that are in the G0 stage, but it is able to infect cells in G0 that are terminally differentiated and post-mitotic (Yamashita and Emerman, 2006). HIV-specific CD4+ T cells are preferentially infected by HIV-1 *in vivo* (Douek et al., 2002). The pathogenesis of HIV-1 infection involves infection and replication within cells bearing the CD4 molecule, the natural ligand of HIV-1 gp120, and other co-receptors including CCR5 and CXCR4, which facilitate the interaction of CD4 and gp120. Unfortunately, these target cells compose the core of host immunity and are important in the induction and maintenance of host immune defense against pathogenic infections and immune homeostasis to keep the host healthy, such as $CD4^+$ T cells, macrophages and a small portion of $CD8^+$ T cells. C-type lectin receptors (CLRs) expressed on different subsets of dendritic cells (DC), including DC-sign (CD209), langerin (CD207) and mannose receptor (CD206), have recently been identified as HIV-1 co-receptors on DCs and thus place DC also on the list of HIV-1 target cells (Geijtenbeek et al., 2000; Turville et al., 2002). The ongoing replication of HIV-1 and consequent host immune response leads to the profound destruction and depletion of infected and even uninfected bystander CD4+ T cells by multiple mechanisms. These include cytotoxic T cell mediated cytotoxicity in HIV-1 infected cells, apoptosis mediated by Fas-mediated mechanisms in infected cells and bystander cells, natural killer (NK) cell mediated cytotoxicity and

cytopathogenicity induced by syncytium formation (Alimonti et al., 2003; Paranjape, 2005). Due to the vital role of $CD4^+$ T-cells in regulating and amplifying host immune response, depletion of these cells leads to profound immunosuppression resulting in deficits in both humoral and cellular immunity and subsequently the development of a series of symptoms at late stage of infection clinically recognized as AIDS. The virus debilitates the immune system leaving the host vulnerable to opportunistic infections and malignancies which are often lethal for these patients. Since the HIV-1 can not be eradicated once established in the host, the initial stages and antecedent events critical for the establishment of productive HIV infection have been extensively investigated. The primary establishment encompasses only the first hours to days after initial viral entry (Haase, 2005) . Mounting evidence suggests that virus at this stage is vulnerable and the infection could be prevented or eradicated at the point of entry (Haase, 2005). However, the window period for containing HIV infection by effective immune response or medical intervention at this stage ranges only from 1 day to 1 week (Pope and Haase, 2003).

The typical duration of HIV infection is approximately 5 to 10 years. WHO, the Centers for Disease Control and the Walter Reed Army Medical Center in the United States all have a staging system which outlines the course of HIV infection and the development of AIDS. The Walter Reed system has 6 stages while the CDC has 3 clinical categories which are then subdivided into 3 CD4+ T cell count (cells per microliter of blood) categories. The Walter Reed system begins at stage 0 which is the point at which infection first occurs. The highest viral titers occur during this primary acute stage and it is at this time that an infected individual is most likely to transmit the virus. Non-specific flu-like symptoms may occur during this stage. This stage lasts

approximately six weeks until a full immune recognition occurs and the viral load undergoes a dramatic decline. Stage 1 begins when the body starts to fight back. Antibodies (humoral response) and cell-mediated immunity begin to develop against the invading pathogens. Stage 2 then begins when the infected cells begin to gather in the lymph nodes. This stage lasts 3-5 years and the patients remain "healthy" physically and clinically. During this time, $CD4^+$ T cell counts remain normal though they do slightly decline. Stages 3 and 4 are linked. During this time the $CD4^+$ T cell decline goes from 700 cells/ μl to 400 cells/ μl. Stage 4 begins when the $CD4^+$ T cell count remains steadily under 400 cells/cubic mm. Stage 5 is when the patient has $CD4^+$ T cell counts below 400 cells/cubic mm but remain above 200. A patient is considered to be in stage 5 if they develop an opportunistic infection since this is a definitive sign that HIV-1 is winning the battle. In stage 6, the $CD4^+$ T cell count remains below 200 cells/ μl and numerous opportunistic infections may develop including PCP and Kaposi's sarcoma, a rare herpes virus-8 related cancer. Patients in stage 6 are considered to have AIDS and on average survive no more than 5 years (Redfield et al., 1986).

HIV/AIDS epidemic

Since the first case of AIDS was noted in 1981, HIV/AIDS has swept all populated continents globally and generated catastrophic impacts in the stricken areas. The numbers and facts of this epidemic are shocking: over 60 million cases of HIV infection and over 25 million deaths have been reported in total; 40.3 (36.7~45.3) million people around the world were living with HIV/AIDS in 2005; approximately 11 out of every 1,000 adults

(ages 15 to 49) are HIV-1 infected globally; 4.9 (4.3~6.6) million new HIV infections occurred in 2005 and this rate has sustained for about 10 years; 3.1 (2.8~3.6) million people lost their lives in 2005; 25 million children will be orphans by 2010 because of AIDS (UNAIDS/WHO, 2005).

Sub-Saharan Africa is the region of the world where the HIV epidemic has been the worst and where its impact is still increasing. Among the over 25 million deaths due to AIDS worldwide to date, three-quarter of them have occurred in Sub-Saharan Africa (Piot et al., 2001). In 2005 alone, 2.4 out of all 3.1 million deaths due to HIV/AIDS occurred in Sub-Saharan Africa (UNAIDS/WHO, 2005). This region has just over 10% of the world's population, but is home to 64% of all people living with HIV/AIDS, 25.8 (23.8~28.9) million people are currently infected by HIV-1 and 13.5 (12.5~15.1) million of them are women. Adult prevalence in Sub-Saharan Africa was 7.2 (6.6-8.0) % in 2005; In 2005, an estimated 3.2 (2.8 ~3.9) million people in the region became newly infected which counts 65% of all new infections worldwide (UNAIDS/WHO, 2005).

Kenya is one of the countries seriously stricken by HIV in Sub-Saharan Africa. According to the Kenya Demographic and Health Survey (KDHS) 2003 (Ministry of Health, 2005), 6.7% of Kenyan adults were infected with HIV. About 1.1 million adults between 15 and 49 are currently infected, with another 60,000 infected over the age of 50, approximately 100,000 children are currently living with HIV-1 in Kenya. Urban residents have a significantly higher risk of HIV infection (10%) than in rural residents (6%). HIV-1 prevalence in women age 15 ~ 49 is 8.7% while it is 4.6% for men. Remarkably, the prevalence in women aged 20~24 is 9% and 13% in women aged 25~29, which highlights young women as a main target population of HIV intervention strategies.

The human immune system against HIV-1

Host immunity is composed of two different but complementary branches in defending against invading pathogens: innate immunity and adaptive immunity. Innate immunity provides the front line of defense and consists of physical (intact epithelial linings), chemical (i.e. antiviral compounds in mucosal secretions, serum, saliva and lymph), biological (normal flora) and immunological barriers (immunocytes, such as macrophages and NK cells). All these contribute to the disruption of the initial establishment of HIV-1 infection, a key step in HIV pathogenesis. Only a tiny portion of viral inoculum can manage to reach susceptible target cells residing in the lamina propria and become founder viruses (Haase, 2005; Miller et al., 2005). Innate immune responses function by not only recognizing and repelling non-self pathogens directly but also facilitating the induction of adaptive immunity through antigen presenting cells (APC), which provides an even stronger, pathogen specific and long-lasting immune protection. Comprehensive studies in fully understanding the role of innate immunity in HIV/AIDS remain insufficient and urgently needed.

Adaptive immunity also has two arms: humoral immunity and cell-mediated immunity. Both rely upon the clonal expansion of lymphocytes capable of recognizing and responding to specific epitopes on the pathogens/antigens and generating pathogen/antigen specific effector cells and molecules to repel the invaders. The effector responses can be broadly categorized into humoral responses, mediated primarily by antibody producing B cells and antibodies they produce, and cellular response which is mediated by effector $CD8^+$ T cells. $CD4^+$ T helper cells (Th) are critical for the

generation and regulation of both humoral and cellular immune responses. $CD4^+$ helper T cells mainly contain two subsets: Th1 and Th2. After antigenic stimulation, Th1 cells produce IFN-γ and IL-2, and promote cell mediated immune responses, while Th2 cells produce IL-4, IL-5, IL-6 and IL-10 which influence the B cell development and augment humoral responses.

The infected subjects mount HIV-1 specific humoral and cell–mediated responses. HIV-1 specific IgM and IgG can be detected in all infected subjects as early as 1 to 6 weeks after the initial infection and a reasonably high titer of HIV-1 specific IgG can be detected all through the infection period (Tindall and Cooper, 1991). The HIV-1 envelope glycoproteins, gp120 and gp41, appear to be the major targets for host humoral immune response (Parren et al., 1997). Although high concentration of HIV-1 specific IgG antibody can be detected in sera from all HIV-1 infected patients which makes it an important diagnostic index for HIV-1 infection, the role of systemic antibodies in combating HIV-1 remains controversial (Paranjape, 2005). There is some suggestion that long term non-progressors host higher titer of HIV-1 neutralizing antibodies with broader antigen spectrum (Pilgrim et al., 1997; Wrin et al., 1994) and neutralizing antibodies exert significant selective pressure on autologous HIV-1(Rybarczyk et al., 2004). But neutralizing antibodies can be detected 8 week or more after infection which is after initial containment of HIV-1 replication in newly infected individuals, suggesting that these antibodies provide no protection at the early stages of infection (Pilgrim et al., 1997; Richman et al., 2003; Wei et al., 2003). Constant HIV-1 mutation generates viral mutants resistant to the existing neutralizing antibodies continuously and development of corresponding neutralizing antibodies usually fall behind, indicating that systemic

antibody response is at a disadvantage in competing with HIV (Wei et al., 2003; Wrin et al., 1994). The majority of patients will eventually progress to AIDS despite HIV-1 specific antibodies which can be detected persistently throughout the HIV-1 infection period (Paranjape, 2005). This suggests that systemic humoral response against HIV-1 has only a limited role to play in preventing progression to AIDS. On the contrary, secretary HIV-1 specific IgA antibody can be detected in mucosal secretions in both HIV-1 infected and interestingly in some exposed seronegative subjects and they have been shown to be able to neutralize HIV-1 and block the transcytosis of HIV-1 in the mucosal tissue (Kaul et al., 2001b). Identification of characteristics of protective epitopes recognizable by neutralizing mucosal IgA antibodies would be of great value in the development of prophylactic HIV-1 vaccines.

HIV-1 specific cell-mediated immunity is mediated mainly by $CD8^+$ cytotoxic T-lymphocytes (CTL) and it was first described in mid 1980s (Plata et al., 1987; Walker et al., 1986). These cells recognize and destroy viral infected cells bearing virus specific antigens on cell surface or are otherwise displaying non-self characteristics such as cancerous cells. Based on the role of CTL in other intracellular pathogenic infection, the CTL response is believed to be the main "sterilizing" response to HIV-1 infection. Differentiated CTL responses have been detected in long term non-progressors (LTNP) and also in HIV-1 exposed but uninfected individuals suggesting that they are indeed critical in fighting HIV-1 infection. HIV-1 disease progression has also been correlated with a loss of CTL activity (Brander and Walker, 1999) and CTL activity had been shown to be directly related to HIV-1 and SIV viral load in human and non-human primates (Ogg et al., 1998; Schmitz et al., 2005; Schmitz et al., 1999). $CD4^+$ T

lymphocytes control both arms of acquired immunity and can modulate the innate immune system as well through production of cytokines with differing regulatory effects on both innate and adaptive immunity system. Strong and efficient T helper responses had been associated with the maintenance of efficient CTL response (Brander and Walker, 1999; Rosenberg et al., 1997). Differential $CD4^+$ T cell responses have also been associated with differential susceptibility to HIV-1 infection (Alimonti et al., 2005). It has been previously proposed that the imbalance in the Th1 and Th2 responses significantly contributes to immune dysregulation associated with HIV-1 infection and disease progression. A switch of Th1- to Th2- like cytokine responses has been believed to be a critical step in the etiology of HIV-1 infection which was supported by kinetics of cytokine production. HIV-1 infected subjects clearly showed that, after stimulation, Th1 cytokines (IFN-γ and IL-2) production dominated the early infection stage while Th2 cytokines (IL-4 and IL-10) were overwhelming in late HIV-1 infection (Clerici and Shearer, 1993). The detection of significant HIV-1 specific Th1 responses in exposed uninfected subjects provides further support to this hypothesis (Alimonti et al., 2005; Clerici and Shearer, 1993).

Previous studies in our Nairobi cohort indicated that the nature of the cytokine response was remarkablely different in resistant and susceptible women. When compared to HIV-1 susceptible women, the resistant women demonstrated significantly higher frequency of exhibiting detectable IFN-γ in culture with HIV-1 (86% vs 18%, $P<0.001$), but lower frequency of detectable IL-4 response (14% vs 91%, $P<0.0002$). Resistant women exhibited more intense IFN-γ production but an over 20-fold lower of IL-4 production in response to HIV-1 stimulation. When these data are examined as the ratio

of virus-stimulated IL-4: IFN-γ production, resistant subjects exhibit patterns commonly termed 'type 1' dominance , whereas other HIV-infected subjects display 'type 2'-dominated responses (IL-4:IFN-γ ratios of 0.27 vs. 46, $P<0.0002$) (Trivedi et al., 2001). All these data highlight the critical role of Th1-Th2 balance in the pathogenesis to HIV-1 infection.

Although all arms of host immunity are capable of responding to HIV-1 in a variety of different manners, it is notable that host HIV-specific immune responses are in most instances ultimately unable to contain or eradicate the virus. The host immune system is at a serious disadvantage in combating infection by HIV-1. Three main factors contribute to the difficulty faced by host anti-HIV-1 immune responses: (1) HIV-1 infection results in a profound depletion of CD4 expressing cells and impairment of corresponding immune functions of these cells due to infection induced apoptosis, cell mediated cytotoxicity and direct cytopathogenicity ; (2) the high mutation rate in HIV-1 transcription due to error-prone polymerase coupled with high number of viral progeny produced daily lead to great viral variability which allows the rapid evolution of the virus capable of evading the host antiviral humoral and cellular immunity (Borrow et al., 1997; Goulder and Watkins, 2004; Price et al., 1997; Richman et al., 2003); (3) The counterbalancing effects of immunoactivation: it is required for efficient host immune responses, but it also generates more target cells for HIV-1 replication which in turn causes further destruction of the immune system.

Main challenges regarding HIV/AIDS

The HIV-1 epidemic and the devastating toll it claims on human life and society makes HIV-1 one of the most challenging pathogens to combat for researchers, clinicians, vaccinologists and pharmaceutical professionals. Although multiple antiretroviral drugs have been licensed since 1987 and morbidity and mortality owing to HIV-1 has dramatically declined in developed nations where they are available, none of current regimens is curative and able to eradicate HIV-1 in vivo (Fauci, 2003; Saksena and Potter, 2003). They merely inhibit viral replication, block acute symptoms of HIV-1 infection or slow down the HIV-1 disease progression, but they do not solve the core problem which is the eradication of the infection (Chun et al., 1998; Phillips et al., 2003; Saag, 2003). Additionally, these drugs can cause severe side effects, result in the emergence of drug resistance in many cases and can be extremely expensive. Meanwhile, anti-retroviral therapy remains inaccessible to the majority of HIV-1 infected patients with the fact that about 95% of HIV-1 infections occur in developing nations. The "3 by 5" initiative, launched by UNAIDS and WHO in 2003, targeted to provide three million people living with HIV/AIDS in low-and middle-income countries with life-prolonging antiretroviral treatment (ART) by the end of 2005. It significantly promoted the process to making universal access of HIV/AIDS prevention and treatment for all in need. The treatment gaps are narrowing gradually, but not at a pace required to effectively contain the epidemic. To gaining the upper-hand against the AIDS epidemic, the world requires a rapid and sustained expansion in effective HIV-1 therapy and especially HIV-1 prevention strategies. Prophylactic vaccines have proven to be the most effective control

strategy for many infectious diseases. Despite great expectations and considerable efforts on the development of HIV-1 vaccines, safe, effective and inexpensive HIV vaccine suitable for human use still remains an ideal.

In summary, the main challenges hindering the development of effective HIV-1 vaccines include:

1. High viral mutation rate due to the lack of proof-reading during HIV-1 replication: New mutations of HIV-1 can be generated all the time and evasion from established immune responses occurs continuously.
2. HIV-1 pathogenesis has not been fully understood.
3. Effective and protective anti-HIV-1 immunity against HIV-1 infection or HIV-1 disease progression has not been fully clarified.

There is, therefore, little doubt that preventing HIV-1 entry and infection of the host is the best way to contain the HIV-1 epidemic. Once HIV-1 infection is established, there is little evidence that sterilizing immunity ever results. So priorities in HIV-1 research should be defining the mechanisms of effective immunological mechanisms which are likely assist to prevent the entry of HIV-1 and/or disrupt the establishment of HIV-1 infection.

Altered susceptibility to HIV-1 infection has been identified in multiple cohort studies

Human HIV-1 infection is characterized by great variability in outcome. The majority of people with a sufficient exposure to HIV-1 become infected, undergo progressive decay of immune function and develop AIDS at a late stage. It is noteworthy that, similar to other infectious diseases, altered susceptibility has been observed in HIV-1/AIDS. HIV-1 exposed yet uninfected "non-permissive" individuals had been identified among commercial sex workers (CSW) (Fowke et al., 1996; Plummer et al., 1999; Rowland-Jones et al., 1995), discordant couples (Beretta et al., 1996; Bernard et al., 1999; Clerici et al., 1992; Langlade-Demoyen et al., 1994; Mazzoli et al., 1997; Stranford et al., 1999), uninfected children of infected mothers (De Maria et al., 1994; Rowland-Jones et al., 1993), and also accidentally exposed health care workers (Bernard et al., 1999; Clerici et al., 1994b; Pinto et al., 1995). Many terms have been coined in describing individuals who resist HIV-1 infection despite of multiple high-risk exposures, such as exposed seronegative (ESN), exposed uninfected (EU), exposed seronegative persons (ESN), highly exposed persistently negative (HEPS), high-risk seronegative (HRSN) (Koning et al., 2005) or HIV-1-resistant (HIV-R) directly(Kulkarni et al., 2003). One of the best characterized HIV-R groups is identified in our female sex worker cohort established in Nairobi, Kenya since 1985, of which some individuals can be epidemiologically defined as resistant to infection by HIV-1 (Fowke et al., 1996) .

It is inarguable that the main hindrance in the development of an effective prophylactic and therapeutic strategies against HIV/AIDS is an incomplete understanding

of HIV-1 pathogenesis and protective anti-HIV-1 immunity. Groups such as the HIV-1 resistant women from Nairobi not only suggest the existence of effective anti-HIV-1 mechanisms, but also provide an important model system for researchers to elucidate these mechanisms.

HIV resistance is a complicated multifactorial process. Although viral factors, such as low dose exposure, certain HIV-1 subtypes and viral attenuations may contribute to decreased risk of transmission after exposure (Hogan and Hammer, 2001), they cannot adequately explain resistance to HIV-1 infection especially for the HIV-1 resistance in CSWs who might exposed to multiple subtypes/strains of HIV-1 with different virulence at very high frequency. Behavioral and environmental factors are also presumed to contributing to the altered susceptibility to HIV-1, but no convincing data exists to prove it so far in the Nairobi cohort as these factors shown to be similar in resistant women and those who seroconverted. On the other hand, the growing body of information suggests that this relative resistance to HIV-1 is natural and inherent and host immunological and genetic factors play critical roles in the development of it.

Immunological correlates in resistance to HIV-1

The best studied innate factors associating with HIV-1 resistance are chemokines which are secreted by immune cells (i.e. T, NK and Mφ) and play important roles in both host innate and adaptive immunity due to their chemoattractant properties. In 1996, two chemokine receptors, CCR5 and CXCR4, were identified as primary co-receptors for macrophage-tropic and T-cell tropic HIV-1 strains respectively, comprehensive studies

have been conducted on the role of chemokines, especially the natural ligands of these two receptors, in HIV pathogenesis (Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Feng et al., 1996). Physiologically, CCR5, which is present on the surface of various cell types (i.e. T cell, DC amd Mφ), binds CC(β)-chemokines, including regulation on activation, normal T cell–expressed and secreted (RANTES), macrophage inflammatory protein (MIP)–1α, and MIP-1β, and then induces signal transduction for immune activation and mediate host tissue inflammatory responses (Luster, 1998). These β-chemokines have been shown to suppress the entry or fusion of macrophage-tropic HIV (R5) strains in vitro (Cocchi et al., 1995; Paxton et al., 1998), presumably by competing for the site of the co-receptor for virus binding. These proteins were assumed to be the major soluble components secreted mainly by $CD8^+$ cells that inhibit the transmission of R5 HIV strains (Paxton et al., 1996; Zagury et al., 1998). Some studies have demonstrated that HIV-1 exposed yet uninfected individuals had significantly elevated serum levels RANTES, MIP–1α, and MIP-1β (Shieh et al., 2001); other studies have shown that CD4+ cells from these subjects produced significantly higher level of RANTES, MIP-1α, MIP-1β upon PHA stimulation and also HIV-1 antigenic peptides (Furci et al., 1997; Paxton et al., 1996). It is noteworthy that the high levels of aforementioned chemokines were not identified in all exposed uninfected subjects (Fowke et al., 1998). Meanwhile, the majority of these studies have assessed systemic responses and the role of compartmentalized mucosal immunity has not been fully investigated yet. Results from our group indicated that HIV-1-resistant CSWs had a 10-fold increase in RANTES expression in genital secretion when compared with HIV-1-uninfected CSWs (Iqbal et al., 2005).

HIV-1 specific adaptive immune responses have also been identified in HIV-1 highly exposed persistently seronegative subjects although they are not infected. Cytotoxic T cell response is the typical immune response to HIV-1 infection (Kaleem et al., 1987) and HIV-1 specific CD8+ cellular immune response against multiple HIV-1 epitopes, including nef, gag, pol and env peptides, have been detected in a significant proportion of HIV-1 highly exposed persistently seronegative individuals including HIV-1 resistant women from Nairobi (Akridge et al., 1999; Bernard et al., 1999; Bienzle et al., 2000; Cheynier et al., 1992; Fowke et al., 2000; Goh et al., 1999; Kaul et al., 2001a; Rowland-Jones et al., 1995; Rowland-Jones et al., 1998; Rowland-Jones et al., 1993; Skurnick et al., 2002; Sriwanthana et al., 2001). HIV-1 exposed uninfected subjects also have systemic HIV-1-specific T-helper responses and systemic and mucosal cytotoxic T Lymphocyte responses (Alimonti et al., 2005; Beretta et al., 1996; Clerici et al., 1992; Clerici and Shearer, 1994; Fowke et al., 2000; Kaul et al., 1999; Mazzoli et al., 1997; Rowland-Jones et al., 1995), HIV-1 specific mucosal IgA in the genital tract (Alimonti et al., 2005; Beyrer et al., 1999; Kaul et al., 1999; Mazzoli et al., 1997; Rowland-Jones et al., 1998), and a global hyporesponsive interleukin-4 (IL-4) response to HIV-1 and other antigens (Trivedi et al., 2001). These data strongly suggest that these individuals may exhibit a cellular (type-1) bias in immune responses to HIV-1. Low level of CD4+ T cell activation and lower T cell responsiveness has also been shown to be associated with low susceptibility to HIV-1 infection (Koning et al., 2005) . Results from our cohort also suggested increased cervical CD4+ and CD8+ T cell counts in HIV-1 resistant subjects, which is not reflected in the systemic lymphocyte compartment (Iqbal et al., 2005).

Genetic correlates in resistance to HIV-1

It is widely accepted that the host genetic background can affect susceptibility to certain pathogens. Various genetic variations have been identified to play a role in the altered susceptibility to infection by HIV. Polymorphisms shown to affect HIV-1 susceptibility have been described in HIV-1 co-receptor and their natural ligands genes including CCR5 (Bienzle et al., 2000; Dean et al., 1996; Huang et al., 1996; Lockett et al., 1999; Rugeles et al., 2002; Shrestha et al., 2006; Zimmerman et al., 1997), DC-SIGN(Liu et al., 2004), CCR2-64I (Louisirirotchanakul et al., 2002), SDF-1 (Soriano et al., 2002; Winkler et al., 1998)) and in HLA genes (HLA-A11(Sriwanthana et al., 2001), A2/6802 supertype(MacDonald et al., 2000), DQ4 (Rohowsky-Kochan et al., 1998), DR5 (Lockett et al., 2001), DRB1*01 (MacDonald et al., 2000), B53 (Rohowsky-Kochan et al., 1998), B18 (Beyrer et al., 1999)). Most of these genetic polymorphisms have also been investigated in multiple HIV-R cohort studies.

In 1996, it was reported that CD4+ T lymphocytes and macrophages from two HIV-1 exposed uninfected subjects were highly resistant to infection by macrophage-tropic HIV-1 and this resistance was due to the inability of viral entry into the target cell(Connor et al., 1996; Paxton et al., 1996). The mechanism of this phenomenon was then revealed that these two individuals were both homologous for a 32 bp deletion in the gene for CCR5 (CCR5Δ32) which abolishes the cell surface expression of CCR5, a key co-receptor for R5 and dual tropic viruses (Liu et al., 1996). Heterozygotes expressing low levels of CCR5 have been shown to have delayed disease progression, although there are no demonstrable changes in susceptibility to infection (Eugen-Olsen et al., 1997;

Mulherin et al., 2003; Samson et al., 1996; Stewart et al., 1997). When examined for its role in mother to child transmission, the deletion showed no effect on susceptibility regardless of homozygous vs. heterozygous state but did delay progression of illness in children with the mutation (Ioannidis et al., 2003; Philpott et al., 1999). It is noteworthy that the protection of chemokine receptor gene polymorphisms is not absolute. Cohorts of CSW resistant subjects are mostly from Africa and Asia, where the frequency of CCR5Δ32 homozygosity is negligible. Whereas CCR5 mutations are not lethal, CXCR4 mutations are rare because this molecule is vital to cells which express it. Mutations have been noted in its ligand, stromal-derived factor-1 (SDF-1). SDF-1, an α-chemokine that binds to CXCR4, regulates the trafficking of CXCR4+ haemato/lymphopoietic cells, their homing/retention in major haemato/ lymphopoietic organs and accumulation of CXCR4+ immune cells in tissues affected by inflammation (Kucia et al., 2004). A "G→A" single nucleotide polymorphism (SNP) had been identified at nucleotide 801 in SDF-1 3′ untranslated region (UTR) (Winkler et al., 1998). This change had been shown to be associated with reduced susceptibility to HIV-1 infection in adults although no effect in mother to child transmission(Soriano et al., 2002; Sriwanthana et al., 2001; Winkler et al., 1998). Due to its physical location in untranslated region, the manner in which this polymorphism affect HIV-1 susceptibility is unknown although it is speculated that the mutation increases translation of the mRNA and causes an excess of the ligand that outcompetes HIV-1 for CXCR4 (Soriano et al., 2002). The chemokine receptor CCR2 had also been identified as an alternative HIV-1 co-receptor and a variant called CCR2-64I was shown to be associated with delayed HIV disease progression in homozygous subjects (Ioannidis et al., 1998; Kostrikis et al., 1998). Study in HIV-1 discordant couples

showed that CCR2-64I homozygosity was overrepresented in exposed uninfected subjects(Louisirirotchanakul et al., 2002), but this is not the case in commercial sex workers(Sriwanthana et al., 2001). Dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN) is able to bind the HIV-1 surface protein, gp120, with high affinity and in a CD4-independent manner(Curtis et al., 1992; Geijtenbeek et al., 2000). DC-SIGN is abundantly expressed primarily on DCs, including those derived from monocytes and those located beneath the genital surface and HIV-1 may use DCs as carriers to gain entry into lymph nodes and, subsequently, to infect $CD4^+$ T cells. The neck region of DC-SIGN between the C-terminal domain and the transmembrane domain is formed by 6~8 repeats of 69 bp that encode repeating units of 23 amino acids. Allele 7 (with 7 repeats in this region) is regarded as the wild type, whereas alleles 6 and 8 are rare and regarded as mutations. The rare 7/6 genotype is significantly over-represented in HIV-1 exposed uninfected subjects and has been suggested to be associated with reduced susceptibility to HIV-1(Liu et al., 2004). A "C→G" mutation was identified at position -28 in RANTES upstream promoter region and RANTES -28G was shown to be associated with increased transcription and expression of the RANTES gene and delayed disease progression to AIDS in HIV-1-infected individuals(Liu et al., 1999). When this polymorphism was investigated in a CSW cohort from Thailand, a slightly higher frequency RANTES -28G was noted in HEPS subjects when compared to controls, but these differences were not statistically significant (Sriwanthana et al., 2001).

Due to the vital role of HLA in the generation and regulation of host immune responses, HLA polymorphism is the best studied genetic factor in HIV-1 exposed

uninfected cohort studies. HLA-DR5, DQ4, B53 had been identified to be significantly overrepresented in exposed uninfected subjects in cohort studies conduct in discordant couples(Lockett et al., 2001; Rohowsky-Kochan et al., 1998). HLA-A11, HLA-A2/6802 supertype (consist of alleles A*0202, A*0205, A*0214 and A*6802), HLA-B18, and HLA-DRB1*01 have been shown to be associated with HIV-1 "resistant phenotype" in commercial sex worker cohort studies(Beyrer et al., 1999; MacDonald et al., 2000; Sriwanthana et al., 2001). The available data suggests that the association between HLA genotypes and HIV-1 resistance may, to some extent, be mediated through protective HLA alleles selectively presenting peptide epitopes from highly conserved regions of HIV-1 and subsequently triggering protective immune responses (Rowland-Jones et al., 1998). It should be noted that most of these associations were seen only in certain population (s) suggesting that at least some associations between HLA and HIV-1 resistance may be due to diversity of populations and diversity of infecting viruses and thus not applicable to all populations.

Although a variety of factors have been postulated to confer protection against HIV infection, definitive proof that these factors are actually responsible for natural resistance is lacking. In virtually all cases, a possible role in altered susceptibility to infection by HIV-1 has been inferred from the occurrence of these factors in differentially susceptible subjects than in appropriate control subjects. Association is not causation and all these factors may contribute to HIV resistance observed in our Nairobi cohort in some way, but none of them can be conclusively associated with HIV resistance. More direct functional evidence is needed to demonstrate that the factor is really protecting the cells/host from getting infected. HIV resistance is apparently a multifactorial phenomenon and a

spectrum of factors might be involved. Numerous of studies on these HIV-R individuals are ongoing especially on novel mechanisms that enable them to repel the virus and clear it before it has a chance to become established. Knowledge of these mechanisms may ultimately assist in developing strategies for efficient HIV-1 prevention and AIDS therapy.

Studies specific to our Kenyan female sex worker cohort revealed that HLA genotypes and phenotypes, but not chemokine receptor polymorphisms correlated with HIV-1-resistance (Dunand et al., 1997; Fowke et al., 1998; MacDonald et al., 2000; MacDonald et al., 2001). HLA associations appear to be population-dependent, likely due to complex interactions between HLA subtypes and clade variability of HIV-1 and do not fully explain resistance in this cohort; additional factors are likely involved. As resistance to HIV-1 infection is clearly immune mediated and a Th1-biased immune response is present in the HIV-1 resistant subjects (Alimonti et al., 2005; Kaul et al., 1999; Rowland-Jones et al., 1998; Trivedi et al., 2001), regulation of host Th1-Th2 immune responses is likely critical in determining resistance.

Collectively, HIV-R is a multifactorial process and mediated by multiple mechanisms. The potential determinant(s) will not be solo immune or genetic factor(s), but rather a comprehensive network concerning both genetic background and host immunity and may also related to invading virus itself. Factors affecting HIV-1 transcription and replication may also be involved in this process due to their ability in influencing the establishment of HIV-1 infection. Potential genetic polymorphism(s) influencing both host immunity and HIV-1 itself might exist and play a dominant role in mediating the natural resistance to HIV-1 infection.

IL-4 gene cluster screening revealed IRF-1 as a potential candidate for further investigation

The human IL-4 gene cluster consists of genes critical to both Th1 and Th2 responses. Predominant Th1 or Th2 responses had been shown to determine susceptibility to a number of infections including Leishmainiasis, chlamydial infection in mouse models (Lucey et al., 1996; Scott, 1998; Yang et al., 1996), and had been linked to the Th2 gene cluster on murine chromosome 11 (analogous to the IL-4 gene cluster on human chromosome 5), and possibly the interferon regulatory factor -1 (IRF-1) gene (Gorham et al., 1996). HIV resistant women in our cohort have an immunologic response to HIV-1 suggestive of a dominant HIV-1 specific Th1 response. With the hypothesis that significant genetic variation(s) in this region might exert significant impact on varied HIV-1 susceptibility, experiments were conducted in which six microsatellite (MS) markers spanning the IL-4 gene cluster (D5S666, D5S1984, IL-4 MS, IRF-1 MS, D5S2115 and D5S399) were characterized in our cohort to determine if they were associated with resistance to HIV-1. Significant differences in the allele distributions for the D5S1984 and the IRF-1 MS markers between the HIV-1 resistant and susceptible groups were identified. Although the allele distribution in the D5S1984 marker differed significantly between the two groups, no specific allele was associated with the resistance phenotype. In contrast, the IRF-1 MS marker clearly had one allele (allele 179 with 12 "GT" dinucleotide repeats in intron 7 of IRF-1) significantly associated with the HIV-1 resistance phenotype although exerted a weak protective effect with HIV-1 resistance phenotype($P=0.0051$, $OR=0.53$, 95% CI: 0.32~0.87) (Ball, 2001). These findings suggest

that IRF-1 is an important candidate gene for further investigation.

IRF-1 plays important role in host antiviral immunity

Interferon regulatory factors (IRFs) consist of a family of transcriptional regulators involved in multiple physiological processes (Mamane et al., 1999; Nguyen et al., 1997; Taniguchi et al., 2001). So far, nine members have been identified and named IRF-1 to IRF-9 (IRF-8 was previously named as IRF-consensus-sequence-binding protein or ICSBP and IRF-9 was previously known as IFN-stimulated gene factor 3γ or ISGF3γ). These transcription regulators play important physiological roles in the host (Lohoff and Mak, 2005) . All members of IRF family share homology in their first 125 amino acids encompassing the DNA-binding domain (DBD) of IRFs characterized by five tryptophan repeats (WWWWW) spaced by 10~18 amino acids(Schaper et al., 1998; Uegaki et al., 1995). They can bind to similar DNA sequences termed IRF-element (IRF-E) with consensus sequence G(A)AAAg/ct/cGAAAg/ct/c which is present within the promoters of IFN-β and other IFN stimulated genes (ISGs) (Tanaka et al., 1993). This sequence almost overlaps the interferon-stimulated response element (ISRE) (a/gNGAAANNGAAACT) targeted by IFN signaling (Kroger et al., 2002; Marsili et al., 2003). IRF-1 is the first identified member of this family and it was first described as a transcriptional regulatory protein that mediates the transcription of interferon β genes, which was also the reason IRF was named (Fujita et al., 1988; Miyamoto et al., 1988). More recently, it has been recognized as transcription activator for a number of interferon inducible genes (ISGs) via interaction of the IRF binding domain located in the promoter

region of target genes and regulator of their corresponding functions (Kroger et al., 2002). The IRF-1 targeted genes encompass genes actively involved in antiviral responses such as IFN-α/β, 2′, 5′-OAS and protein kinase R (PKR), antibacterial responses such as iNOS and $gp91^{phox}$, cell apoptosis such as caspase 1 and 7, and genes critical for other innate and adaptive immune responses such as IL-12, IL-15, and TAP-1/LMP-2. In addition to antiviral and antibacterial responses, these genes are also involved in the regulation of host immunity, cell differentiation, cytokine signaling and hematopoietic cell development (Kroger et al., 2002; Mamane et al., 1999).

The IRF-1 gene is located in IL-4 gene cluster on human chromosome 5q31 and spans 7721 nucleotides right next to its 495 bp promoter sequence (**Figure 1**). Most functional data on IRF-1′s role in host adaptive and innate immunity was derived from IRF-1 knockout mice studies. IRF-1 knockout mice have a number of significant immunological alterations including: (1) A defect in the development of thymic $CD8^+$ cells and a reduced number of mature $CD8^+$ cells (Matsuyama et al., 1993; Penninger and Mak, 1998); (2) Profound phenotypic changes in $CD4^+$ T cells in spite of "normal" maturation process (Matsuyama et al., 1993), a greater number of memory/effector CD4+ T cells and less naïve $CD4^+$ cells observed, (3) Their $CD4^+$ cells showed decreased production of IL-2 and IFN-γ (Th1 cytokine) and increased IL-3, 4, 5, and 6 (Th2 cytokine) production after exogenous stimulation which suggested that their $CD4^+$ cells preferablely underwent a Th2 differentiation(Mamane et al., 1999), the defect of Th1 differentiation is most likely due to a impaired IL-12 production by macrophages in IRF-$1^{-/-}$ mice and a IRF-1 binding site had been found in the promoter region of IL-12 p40 subunit (Ma et al., 1996; Taki et al., 1997); (4) Significant decrease in both the numbers

Figure 1 Depiction of IL-4 gene cluster and genomic structure of IRF-1 gene

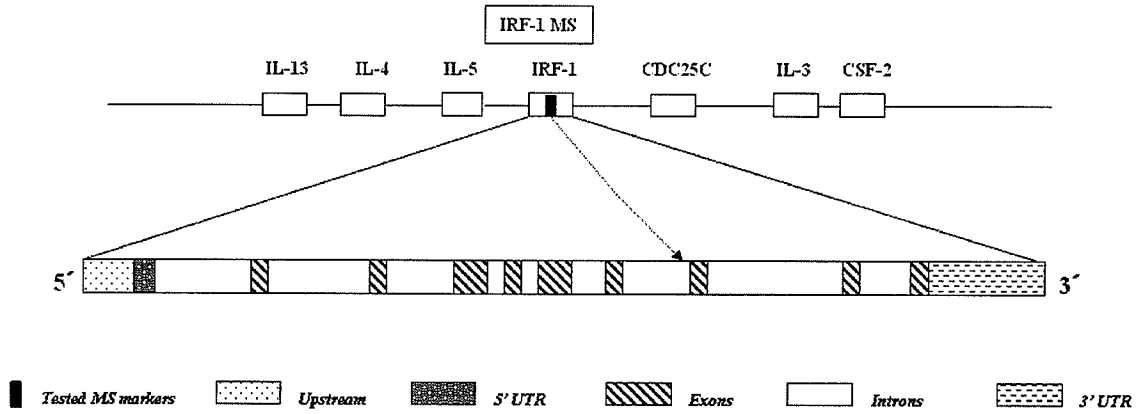


Figure 1 Depiction of IL-4 gene cluster and genomic structure of IRF-1 gene.

The top half of this diagram depicts the composition of human IL-4 gene cluster located between 5q23 and 5q31 (not to scale). Genes encoding Interleukin (IL)-13, IL-4, IL-5, Interferon Regulatory Factor 1 (IRF-1), CDC 25C, IL-3, and Colony Stimulation Factor (CSF) 2 are located within this region. The black bar indicates the microsatellite (MS) region in IRF-1. The lower section shows the genomic structure of the human IRF-1 gene, which encompasses 7748 nt and contains one 5′untranslated region (UTR), 9 exons, 9 introns, one 3′ UTR and a 495 nt immediate upstream promoter region.

and functions of NK cells potentially due to the lack of IL-15 which is transcriptionally regulated by IRF-1 (Ogasawara et al., 1998; Ohteki et al., 1998; Taki et al., 1997); (5) Decreased MHC I expression likely as consequence of reduced expression of transporter associated with antigen processing-1 (TAP-1) and the low molecular weight protein-2 (LMP-2) which are crucial for class I antigen expression and their gene expressions are IRF-1 dependent (Hobart et al., 1997; Penninger and Mak, 1998; White et al., 1996); (6) A decreased polymeric immunoglobulin receptor (PIGR) expression, which is responsible for the transportation of polymeric IgA across epithelial barrier (Blanch et al., 1999). It is apparent that IRF-1 plays a critical role in the development and regulation of host immune responses, especially cellular immunity and mucosal immunity which are key components in antiviral immunity. This aligns well with the current data derived from our immunologic studies which strongly suggests that resistance to HIV-1 infection is immune mediated and that HIV-1 resistant individuals may exhibit a Th1-biased immune response to HIV-1 with stronger HIV-1 specific cellular, but not humoral responses, as well as what appears to be a compartmentalized mucosal cellular and antibody response (Kaul et al., 1999; Rowland-Jones et al., 1998). We hypothesized that IRF-1 plays a critical role in generating successful protective immune response to HIV-1.

IRF-1 regulates the transcription of HIV-1 LTR

Interestingly, IRF-1 has also been suggested to bind to an IRF sequence element in the HIV-1 5'-LTR spanning nt +200 to 217 (TTGAAAGCGAAAGGGAAACC) and to initiate HIV-1 LTR transcription independently of Tat, a key viral transactivator of HIV-

1 replication. IRF-1 may act at initial stages of viral replication prior to Tat. It also synergizes with Tat in amplifying HIV-1 replication at later stages (Marsili et al., 2003; Sgarbanti et al., 2002). HIV-1 is able to induce IRF-1 response early upon infection and prior to the expression of Tat in both Jurkat cell line and primary CD4+ T lymphocytes (Sgarbanti et al., 2002). IRF-1 binding to the target sequence on HIV-LTR results in the activation of HIV-LTR transcription. By promoting HIV-LTR transcription and HIV replication, IRF-1 can also activate the transcription of Tat and, in turn, amplify LTR-directed gene expression. IRF-1 also binds to Tat and cooperates with suboptimal doses of Tat to activate HIV-1 transcription (Sgarbanti et al., 2002). These findings suggested a key role of IRF-1 in the early phase of viral replication and/or during viral activation from latency, when viral transactivators are absent or present at very low levels (Sgarbanti et al., 2002). Therefore, IRF-1 may be critical in the establishment of HIV-1 infection and a potential determinant of the outcome of initial HIV-1 infection. In addition to its role in HIV-LTR transcription initiation during primary infection, the previous studies also suggested a potential role of IRF-1 during viral reactivation from latency. Latently infected resting T cells are a main and stable reservoir of HIV-1 in infected subjects. Reactivation of latent provirus occurs when these cells are activated and this depends greatly on the host transcription factors (Greene, 1990; Siliciano, 1999). IRF-1 can be upregulated upon T cell stimulation and might thus contribute to viral reactivation independent of the presence of Tat (Nelson et al., 1996). In a similar vein, proinflammatory cytokines, such as IFN-γ and TNF-α, which could induce cell activation and stimulate HIV-1 replication also result in strong IRF-1 responses(Marsili et al., 2003).

These data suggest that IRF-1 may play a significant role in HIV-1 pathogenesis. Taken together with previous data from our group which demonstrated that a particular MS allele ("179") in the IRF-1 gene was associated with HIV-1 resistance phenotype (Ball, 2001), it is clear that further studies of IRF-1 genetics in this population are warranted.

Hypothesis and Objectives

Here we hypothesized that genetic polymorphisms in the IRF-1 gene are potential determinants of resistance to HIV-1 infection. Further to that, polymorphisms in IRF-1 will have significant functional consequences which contribute to variation in susceptibility to HIV-1 infection.

To investigate this, the following objectives were established:

1. To identify all IRF-1 gene polymorphisms in the subject cohort via complete sequencing of the IRF-1 gene and its immediate upstream promoter region;
2. To identify all key polymorphisms showing strong correlation with HIV-1 resistance by comprehensive association analysis;
3. To explore the underlying mechanisms of the association between IRF-1 gene polymorphisms and HIV-1 resistance at the protein level, the mRNA level and in the context of in vitro HIV-1 infection models.

Subjects, Materials and Methods

Subjects:

All subjects were from a well-described female sex worker cohort established in Nairobi, Kenya, in 1985. All studies involving these subjects have been approved by both the University of Manitoba and the University of Nairobi ethics review panels.

All enrollees were tested for HIV-1 by serological test upon enrollment and during every visit in regular twice-a-year resurvey. Confirmative HIV-1 testing was conducted using a second serological EIA test. All subjects' case history data including HIV-1 status, $CD4^+$ and $CD8^+$ cell counts were archived in our database for reference. Peripheral blood mononuclear cells (PBMCs) or whole blood samples and plasma samples were obtained and cryopreserved in liquid nitrogen before shipment back to Winnipeg, Canada. All biological samples tested for these subjects were from our sample reservoir at the University of Manitoba in Winnipeg, Canada.

All subjects involved in this study were divided into three different categories according to the established criteria: (1) HIV-1-resistants: subjects who were HIV-1 negative at enrollment and remained HIV-1 negative for over 3 years while continuing active sex work; (2) HIV-1-infected: subjects who were HIV-1 positive at enrollment; (3) Seroconvertors: those who were identified as HIV-1 negative at enrollment and seroconverted during follow-up. The later two groups were together considered as HIV-1- susceptibles.

Materials:

Commercial kits used:

1. Big Dye® 3.1 terminator V3.1 cycle sequencing kit (Applied Biosystems, Part number: 4336917)
2. ECL Advance TM Western Blotting Detection Kit (Amersham Bioscience, Catalog number: RPN2135)
3. Expand High Fidelity PCR System (Roche, Catalog number: 11732650001)
4. Luciferase Assay System (Promega, Catalog number: E1500)
5. Luciferase Assay System (Promega, Catalog number: E1500)
6. QAIamp DNA Mini Kit (QIAGEN, Catalog number 51304): for DNA isolation
7. QIAGEN Plasmid Mega Kit for plasmid purification: (QIAGEN, Catalog number: 12181)
8. QIAGEN RNeasy Mini Kit (QIAGEN, Catalog number: 74104): for total RNA extraction
9. QIAprep 96 Turbo Miniprep kit (Qiagen, Catalog number:27191)
10. SuperScript ^{TM} First-strand Synthesis System for RT-PCR (Invitrogen, Catalog number: 11904-018)
11. Taq polymerase PCR system (Invitrogen, Catalog number: 10342-020)

Commercial reagent sources:

1. 30% Acrylamide/Bis (Bio-Rad, Catalog number:161-0154)
2. Ammonium Persulfate (Bio-Rad, Catalog:161-0700)
3. Antibiotic-antimycotic solution (100×) (Multicell, Catalog number: 450-115-EL)
4. Betaine (5M solution) (Sigma, Catalog number: B-0300)
5. Blotto non-fat dry milk (Santa-Cruz, Catalog number: sc-2325)
6. Custom primers (mostly from Invitrogen)
7. Dimethyl Sulphoxide (DMSO) Hybri-max® (100ml) (Sigma, Catalog number : D2650)
8. DMEM medium (Sigma, Catalog number: D5671)
9. DNA ladder, 100 bp (Invitrogen, Catalog number: 15628-019)
10. DNase I, RNase free (Qiagen, Catalog number: 1010395)
11. ELISA Substrate Tablets (Sigma, Catalog number: 104)
12. Ethidium Bromide (10 mg/ml) (Sigma, Catalog number D2650)
13. Fetal bovine serum (Hyclone, Catalog number: SH30396.03)
14. Fluorescent labeled forward primer (4000 pmol total): 6FAM- (for MS typing)
15. GeneScan™ -120 Size Standard (Applied Biosystems, Part number: 4322362)
16. Glycine (Bio-Rad, Catalog number:161-0718)
17. Goat anti-Actin polyAb (Santa Cruz, Catalog number: sc-1616,)
18. Hi-Di ™ Formamide (Applied Biosystems, Part number: 4311320)
19. LymphoprepTM (AXIS-SHIELD Poc AS, Norway Ref: 1114547)
20. P24 ELISA coating antibody: NIH, Catalog number1513

21. Peroxidase-conjugated AffiniPure Goat anti-Rabbit IgG(H+L) (Jackson ImmunoResearch, Catalog number: 111-035-003)
22. Peroxidase-conjugated AffiniPure Rabbit anti-Goat IgG(H+L) (Jackson ImmunoResearch, Catalog number: 305-035-003)
23. POP-6^{TM} performance Optimized Polymer 6 (Applied Biosystems, Part number: 402837)
24. Prestained SDS-PAGE protein standards, Broad Range (BIO-RAD, Catalog number: 161-0318)
25. Rabbit anti human IRF-1 (Santa Cruz, Catalog number: sc-497,)
26. Rabbit polyclonal anti P24 (NIH, Catalog number: 384)
27. RPMI 1640 medium (Hyclone, Catalog number: SH30027.01)
28. Taq DNA polymerase recombinant (Invitrogen, Catalog number: 10342-020)
29. Temed (Bio-Rad, Catalog number: 161-0801)

Self-prepared Solutions:

1. 2 ×HeBS (Hepes Buffer Saline)

	100ml	300ml	500ml
NaCl (g):	1.6	4.8	8.0
KCl (g)	0.074	0.222	0.37
$Na_2HPO_4.2H_2O$ (g):	0.027	0.081	0.135
(OR Na_2HPO_4 (g)	***0.021***	***0.063***	***0.11)***
Dextrose (Glucose) (g)	0.2	0.6	1.0

Hepes (g)	1.0	3.0	5.0
d.d.H_2O (ml)	90	270	450

--Adjust pH to 7.05 using 0.5N NaOH

--Adjust the volume to 100ml, 300ml or 500 ml with H_2O.

--Filter with 0.22 μm micron

--Aliquot to 10ml/tube, keep at -20 °C.

2. Trypsin – EDTA

Trypsin (Sigma 1:250): 1.25g

0.5M EDTA stock: 1 ml

(OR EDTA-4Na: 0.19g)

-----Dissolve in 500 ml PBS (Without Ca^{++} and Mg^{++}),

-----Adjust pH to 7.0 to 7.2.

-----Sterilize by filtration using bottle-top filter.

3. TE buffer (pH8.0) For 1 liter

10 ml Tris-HCl (1M)

10 mM Tris-Cl (pH 8.0) 2 ml EDTA (0.5 M)

1mM EDTA (pH8.0) 988 ml d.d H_2O

4. 1 M Hepes:

47.66 g Hepes + 150 ml d.d.H_2O: dissolve thoroughly and then add d.d. H_2O to 200ml. Keep at 4°C.

5. 10 × TBE buffer (for 1 liter):

Tris Base (MW 121.1): 108 g

Boric acid (MW 61.83): 55 g

EDTA (Disodium, dehydrate, MW: 372.2): 7.44 g

---- Distilled H_2O to 1 liter.

6. 6× PCR Gel Loading buffer:

   0.25 bromophenol blue

   0.25 Xylene Cyanol

   30% Glycerol in H_2O

   ----dissolve well, 0.2um syringe filter to filter, aliquot and store in -20°C.

7. 2 × master mix for PCR

   Recipe for 2× master mixture used in this project:

Concentration of Stock Buffer	Quantity	Final Conc. in 2× mixture
1) 833 mM (or 1M) Tris-HCl pH9.0	144 µl	120mM Tris-HCl pH9.0
2) 50 mM $MgCl_2$	60 µl	3 mM $MgCl_2$
3)1M $(NH_4)_2$ SO_4	30 µl	30 mM $(NH_4)_2$ SO_4
4) 1mM dNTP Mix	200 µl	200 uM dNTP Mix
5) 1% Gelatin	200 µl	0.2% Gelatin
6) D.D.W (distilled deironized)	276 µl	
Total volume	910 µl	

Alternatively, we can use another 2× mix if higher demand is required

Concentration of Stock Buffer	Quantity	Final Conc. in 2× mixture
1) 833 mM (or 1M)Tris-HCL pH9.0	144 µl	120mM Tris-HCL pH9.0
2) 50 mM $MgCl_2$	60 µl	3 mM $MgCl_2$
3)1M $(NH_4)_2$ SO_4	30 µl	30 mM $(NH_4)_2$ SO_4
4) 1mM dNTP Mix	200 µl	200 uM dNTP Mix

5) DMSO	100 µl	10% DMSO
6) 1% Gelatin	10 µl	0.01% Gelatin
7) D.D.W (double autoclaved)	366 µl	
Total volume	910 µl	

8. 10% SDS

---100 g SDS

--- 900 ml water

--- Adjust the volume to 1L and filter with 0.2 um syringe filter

9. 6 ×Western Blot sampling buffer:

0.5 M Tris (pH 6.8) : 7.0 ml

Glycerol: 3.8 g (drop to weigh)

SDS 1.0 g

DTT 0.93 g

Brown phenolblue 1 %

---- Make sure everything is dissolved well, keep at -20°C.

10. Making separation gel (10ml for one 1.5mm or two 0.75mm mini-gels)

	6 %	7.5%	10 %	12 %	15 %
d.d.H_2O	5.3 ml	4.8 ml	4.0 ml	3.3 ml	2.3 ml
1.5M Tris (pH8.8)	2.5 ml	2.5 ml	2.5 ml	2.5 ml	2.5 ml
30% Acrylamide	2.0 ml	2.5 ml	3.3 ml	4.0 ml	5.0 ml
10% SDS	100 µl	100 µl	100 µl	100 µl	100 µl
10% APS	100 µl	100 µl	100 µl	100 µl	100 µl
Temed	8 µl	8 µl	6 µl	5 µl	4 µl

11. Make stack gel (5 ml for two 0.75mm gels)

	10 ml (for 4 gels)	5 ml (for 2 gels)
d.d.H_2O	6.2 ml	3.1 ml
0.5 M Tris (pH 6.8)	2.5 ml	1.25 ml
30% Acrylamide	1.33 ml	0.65 ml
10% SDS	100 µl	50 µl
10% APS	100 µl	50 µl
Temed	10 µl	5 µl

12. 5 × Running buffer:

	600ml	1 L	4 L
Tris base:	9 g	15 g	60 g
Glycine:	43.2 g	72 g	288 g
SDS	3 g	5 g	20 g

----- Dissolve in d.d.H_2O and bring to the final volume.

13. 1 × PBS (for 1 Liter, Final pH7.4):

NaCl----------8 g

KCl---------- 0.2 g

Na_2HPO_4----------1.44 g

KH_2PO_4----------0.24 g

----- Dissolve in d.d.H_2O and bring to the final volume.

Plasmids used (generously offered by Dr. Xiaojian Yao, University of Manitoba)

1. pNL-Bru/E-/luc+
2. SVCMV-VSV-G

Cell lines used (generously offered by Dr. Xiaojian Yao, University of Manitoba)

1. 293 T cell line

Main consumables used:

1. 15 ml, 50 ml tubes
2. Cell culture plates and flasks
3. Centricon column (100,000MW). Catalog number 91008, 15 ml capacity)
4. 0.5 ml, 1.5 ml eppendorf tubes
5. Immobilon ^{TM} _P Transfer membrane for Western Blotting (Millipore, Catalog number: P-15552)
6. Membrane for RT assay (Fisherbrand Glass Fiber Filter Circles (G4) Catalog number 09-804-24C)
7. MicroAmp Optical 96-well reaction plate (Applied Biosystems, Part number : N801-0560)
8. MultiScreen ^{TM} PCR Plate (Millipore, Catalog number: MANU03050)
9. PCR tubes and plates

10. Petri dishes (cell culture use)
11. Pipettes and tips
12. Snap-cap tubes (4ml and 15ml)
13. Transfer pipettes

Equipment used:

1. ABI Prism® 310 Genetic Analyzer
2. ABI Prism® 3100 Genetic Analyzer (Hitachi, Japan)
3. Centrifuges
4. EG&G Berthold Microplate Luminometer (Model: LB 96V)
5. Eppendorf Master Cycler *Gradient*: (Eppendorf, NO: 5331 12211)
6. Fluorchem™ 8800 Imaging System (Alpha Innotech Corporation, U.S.)
7. Gel Doc 2000 (Bio-Rad Laboratories, Segrate, Italy)
8. Gel Pump (SAVANT Instruments Inc. Farmingdale, USA, Model : GP110)
9. Incubators
10. Microplate Spectrophotometer (Molecular Devices, Sunnyvale, USA)
11. Millipore filtering set (Catalog number XX270255) for radioactives (for RT)
12. Millpore's Amicon Microcon-PCR centrifugal Filter Devices
13. MultiScreen® Filtration System (Millipore, Catalog number:MAVM0960R)
14. Peltier Thermal Cycler (MJ Research, Model: PTC-200)
15. Power-pack (Bio-Rad, Model number: PowerPac300)
16. Scintillation counter (Liquid Scintillation System, Beckman, LS6000TA)
17. Trans-Blot® SD Semi-dry Transfer Cell (Bio-Rad, Model: Trans-Blot® SD Cell)

Methods

Complete IRF-1 gene sequencing

To identify all gene polymorphisms in IRF-1 gene, 507 subjects were randomly selected from enrollees in our Pumwani female sex worker cohort in Nairobi, Kenya. Genomic sequencing of IRF-1 gene and its upstream promoter region was conducted comprehensively. Twenty-eight healthy local donors were recruited from staff members in the University of Manitoba as control population, which consists of individuals from a variety of ethic groups including Asians, Europeans and Caucasians.

Extraction of genomic DNA

Peripheral blood samples were obtained from the described individuals. PBMCs were extracted following regular Ficoll-Hypaque isolation protocol. DNA was extracted either from PBMCs or from whole blood directly using Qiagen DNA extraction kit (Qiagen, Inc., Mississauga, ON, Canada) or Puregene DNA Extraction Kit (Gentra Systems, Minneapolis, USA) following manufacturer's instructions. All DNA samples were kept at -20°C for further experiments.

PCR amplification of IRF-1 fragments

According to the current GenBank references, the IRF-1 gene is 7,721 bp in size and has a 495 bp immediate upstream promoter region. To facilitate further sequencing, the

entire IRF-1 gene and its immediate upstream promoter region were divided into 17 over-lapping fragments and PCR amplified using specifically designed primers (**Figure 2, Table 1**). All primers were designed based on reference sequences from GenBank for IRF-1 (accession numbers: L05072) or its immediate upstream promoter region (accession numbers: X53095). PCR amplification was conducted as follows: the 50 µl PCR reaction mixture contained: 60 mM Tris-HCl (pH 9.0), 1.5 mM $MgCl_2$, 15 mM $(NH_4)_2SO_4$, 100 µM dNTP mixture, 25 pmol of each primer; 1.25 U Taq DNA Polymerase (Invitrogen Life Technologies, Burlington, ON, Canada) and 100-200 ng of DNA template. PCR was performed on a GeneAmp PCR-System 9600 thermocycler (Applied Biosystems, Foster City, CA, USA) using the following setting: (i) an initial denaturation step of 3 min at 94°C; (ii) 35 cycles of 30 seconds at 94°C, 30 seconds at the indicated annealing temperature (Tm) (**Table 1**), and 2 min elongation at 72°C followed by a final extension step of 10 min at 72°C. Five microliters of PCR products were visually screened by Ethidium Bromide staining and agarose gel electrophoresis.

PCR product purification

PCR amplicons were purified using Millpore's Amicon Microcon-PCR Centrifugal filter devices (Millipore Bedford, MA, USA) prior to sequencing and harvested with distilled deionized H_2O. In brief, the PCR products from above step were transferred to the multiscreen plate (Millipore, Catalog number: MANU03050). The multiscreen plate was then mounted onto provided manifold platform. Filtering of PCR products was performed by using Gel pump with suction at 24″ Hg for 5 min. One hundred microliters

Figure 2 Distribution of sequenced fragments in IRF-1 gene and its promoter region

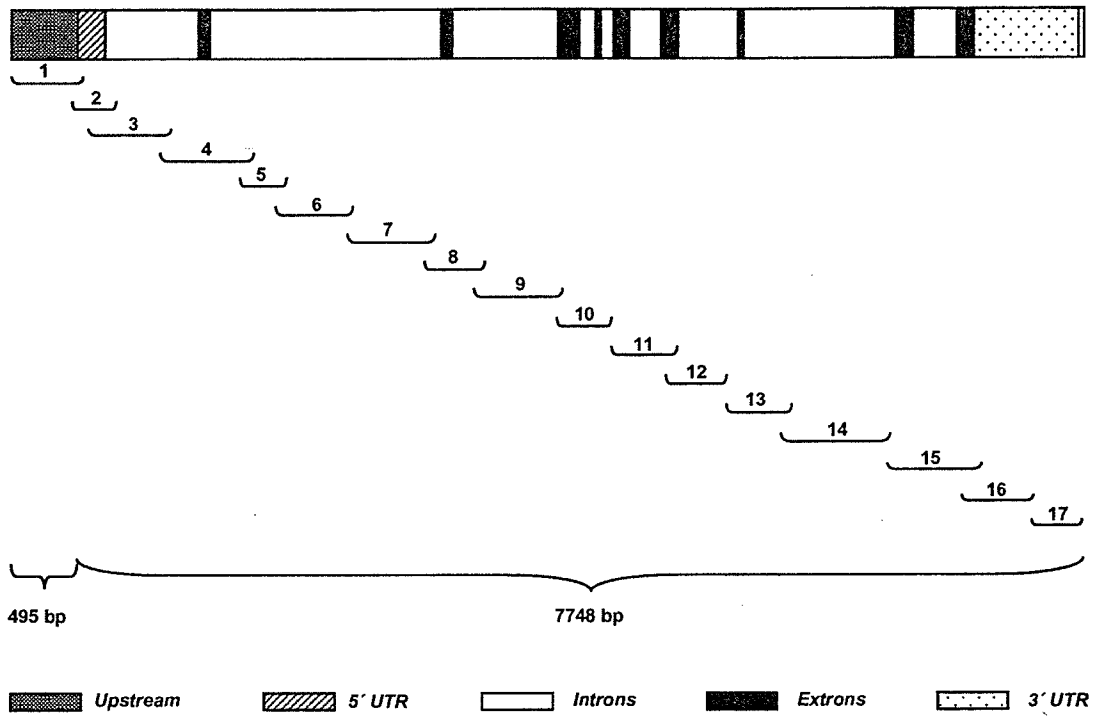


Figure 2 Distribution of sequenced fragments in IRF-1 gene and its promoter region.

The topmost panel shows a diagram of the genomic structure of human IRF-1gene and its immediate upstream promoter region. IRF-1 gene encompasses 7748 nt and contains 9 exons, 9 introns, one untranslated region (UTR) at both 5′ and 3′ ends. Its upstream promoter region is 495 nt in length. To facilitate sequencing experiment, the entire region was divided into 17 overlapping fragments numerated as 1 to 17 as shown in this diagram.

Table 1 Oligonucleotide Primers used for first round IRF-1 sequencing

Locus Amplified*	Amplification Primer sequences		Tm(°C)	Sequencing Primer sequences			
	Forward	Reverse		Forward	Tm(°C)	Reverse	Tm (°C)
-489~-33	gagga gccag gctgc cagtc	cactt agcgg gattc cccag	54.8	ggagc caggc tgcca gt	53.0	cactt agcgg gattc ccca	55.0
-85~250	cgcgc ggctg tccgg gggcg	gtgcc ccggc ctcaa aggcg	65.6	aggge tgggg aatcc cgc	53.0	gtgcc ccggc ctcaa ag	53.0
13~662	tcctt agtcg aggca agac	cagta agcag cccti gcc	54.8	tagtc gaggc aagac gtg	53.0	agccc ttgcc accag caca	55.0
611~1262	cccag gcgac aggtt ttg	ctcca gtctg gaatc tgttg a	50.0	caggc gacag gtttt gggct	53.0	ctgga atctg ttgaa cagta cct	53.0
1171~1550	tagca gctgg ctgag gtgag	tcatg caagt gaggc ccagg	63.1	tagca gctgg ctgag gtg	53.0	tcatg caagt gaggc cca	51.0
1422~2086	cttct ggagg gacca aagct	ctgag tcgag ccgtt caagt	57.0	gacca aagct tcaga tgcag	53.0	tcaag ttcaa atctt ggctt c	51.0
2008~2678	cagag gcact gagag gtaga	ctgag tctga gaccc aggc	59.0	gaggc actga gaggt agatc	47.0	gcctg cccag ccaga cag	47.0
2611-3046	gggaa aaaag gggaa gtcag	ctaag atctt ggctt ttccc	55.8	gaaaa aaggg gaagt cag	47.0	agatc ttggc tttte cct	47.0
2989~3642	ctgga gaaga gagtc gatca tc	gcttt gtatc ggcct agagg	59.0	cgatc atctc agcaa attct c	53.0	tagag ggcag gaga aaaag agg	53.0
3591-4040	catgg ccctg acgat cctct	cacct tagca tttct tccac	52.0	tggcc ctgac gatcc tct	53.0	cttag cattt cttcc actgc	53.0
4043~4545	ctggg cctaa gctgc tttc	acagt caagc agcag gcca	61.0	tgggc ctaag ctgct ttct	53.0	agtca agcag caggc cag	53.0
4482~4899	tgtca cccat gccct ccac	gtgga cctat ctgcc ataga g	57.0	tgccc tccac ctctg aag	53.0	cctat ctgcc ataga gattc ca	53.0
4951~5390	aaggg caggg ggatc tggtg	tgggg catgt tgggg cagag	64.7	aggge agggg gatct ggt	55.0	gcatg ttggg gcaga ggat	55.0
5323~6247	cagct ctgtg agtca ggcac	ctggg gacag aagag caaga	57.0	gtcag gcact atcag cgatg	55.0	agcaa gaggc tatca actcc	55.0
6211~6916	gaagg gagtt gatag cctct	cttgg cagtg gggtc aca	55.0	ttgat agcct cttgc tcttc	53.0	cttgg ctgtt gaggg gc	53.0
6641~7200	cctca tacaa ccttc ttcac	tctct tagtg cagag ccagc	54.8	catac aacct tcttc acatc	51.0	tctct tagtg cagag cca	49.0
7155-7705	cagaa aagca taaca ccaat	cattc ctaca agtta aattt c	44.0	aagca taaca ccaat cccag	53.0	tccta caagt taaat ttctc tt	51.0

Table 1 Oligonucleotide Primers used for first round IRF-1 sequencing.

* Nucleotide numbers indicating locus amplified are numerated according to GenBank sequence of human IRF-1 gene (accession number L05072), the immediate upstream promoter sequence (495 nucleotides) was numerated by -1~-495 (accession number: X53095); Tm=Annealing temperature.

of TE buffer for washing was then added into all sample wells and the plate was filtered for another 5 min using same pump setting. The multiscreen plate was blotted once with paper towel and a proper volume of distilled and deionized H_2O was added into each well to elute the bound purified DNA. Purified DNA samples were then be ready for sequencing PCR or could be transferred into 0.5 ml centrifuge tubes and stored at –20°C. For some samples, Millipore centrifugal columns (Millipore Bedford, MA, USA, Millipore, Catalog number: MAVM0960R) were used for the purification step. Briefly, the labeled columns were placed in a collection tube and 455 µl TE buffer (PH 8.0) and 45 µl PCR reaction to each column were then added onto each column. The column was then centrifuged at 1000 g (3,300 rpm) for 15 min. Proper volume of distilled and deionized H_2O was used to recover purified DNA from the membrane.

Sequencing PCR and sequence resolution

Sequencing PCR was conducted using ABI Prism BigDye Terminator Version 3.0 and 3.1 Cycle Sequencing System (Applied Biosystems, Foster City, CA, USA). Each reaction contained: 2 µl of Big Dye (Version 3 or 3.1, prepared by 1:1 dilution of original solution with provided dilution buffer), 1.5 µl of 10 µM primer (forward or reverse) (**Table 1**) and 2 µl of purified PCR amplicon templates from above steps. Sequencing cycling reactions were programmed as follow: (i) an initial denaturation step of 3 min at 96°C; (ii) 80 reaction cycles with each cycle consisting of 30 seconds at 96°C, 30 seconds at the specified Tm (**Table 1**), and 4 min at 60°C.

The amplified samples were precipitated using a NaOAc - ethanol method. To each cycle sequenced reaction, 1 µl of 3M NaOAc (pH 5.2) and 20 µl 95% ethanol were added and mixed by vortexing. The samples were left to precipitate for 2~24 hrs in the dark at room temperature. Samples were then centrifuged at 12,000 rpm for 30 min to pellet the DNA. The DNA pellets were then washed with 150 µl of 70% ethanol once and dried using 90°C heat-block for 2 minutes. The DNA samples were then ready for sequence resolution.

For sequence resolution, the DNA pellets were reconstituted with 20 µl high grade formamide (Applied Biosystems, Part number: 4311320) and resolved using Applied Biosystems (ABI) Prism 3100 Genetic Analyzer (Hitachi, Japan). Polymorphisms in a randomly selected 16 individuals were confirmed using Expand High Fidelity Plus PCR system (Roche, Catalog number: 11732650001) to eliminate the possibility of errors introduced due to imperfect proofreading in PCR amplification with Taq DNA polymerase.

Sequence analysis and polymorphism identification

Sequencing data was analyzed using Sequencher software (Version 4.0.5, Gene Codes Corporation, U.S.) and gene variations were identified by alignment with published GenBank reference sequences (accession numbers: L05072 or X53095). All polymorphisms were confirmed by sequencing in both directions. Allele frequencies in the sample population were then calculated and genotypes were determined accordingly. All single nucleotide polymorphisms (SNPs) were tested for Hardy-Weinberg

equilibrium. Pair-wise linkage disequilibrium (LD) analysis between SNPs was conducted using PyPop software (Lancaster et al., 2003).

IRF-1 full length sequence determination

Based on our population-based IRF-1 sequence analysis, we noted that there were 35 consistent discrepancies between GenBank reference sequences for IRF-1 and its promoter region and our sequencing data. This observation was confirmed in all tested Kenyan subjects and 28 local donors representing multiple ethnic groups, including Asian, Caucasian and European, suggesting that these discrepancies were real and that the GenBank information on human IRF-1 was not accurate and representative for all populations (Ji et al., 2004). Following the above sequencing protocol, full-length sequencing for IRF-1 and its upstream promoter sequence were conducted on selected subjects who were homozygous for wild-type alleles (consist with current GenBank sequences) at all loci with identified genetic variations. Several new primers were designed and applied to cover the entire IRF-1 gene and its upstream region in this round of sequencing (**Table 2, Figure 3**). The number of fragments amplified and sequenced in this round of sequencing decreased to 16 by combining the first two segments together. The obtained sequences were aligned with current IRF-1 Genbank sequences (L05072 and X53095). Potential splicing pattern changes due to these discrepancies were analyzed using online splicing prediction software (NetGene2--http://www.cbs.dtu.dk/ services/NetGene2/).

Table 2 Oligonucleotide Primers used for full-length IRF-1 sequencing

Locus Amplified**	Amplification Primer sequences		Tm (°C)	Sequencing Primer sequences			
	Forward	Reverse		Forward	Tm (°C)	Reverse	Tm (°C)
-495~-250	aagct tgagg agcca ggctg	gtgcc ccggc ctcaa aggcg	66.0	ggagc caggc tgcca gt	53.0	cactt agcgg gattc ccca	55.0
				agggc tgggg aatcc cgc	53.0	gtgcc ccggc ctcaa ag	53.0
13~662	tcctt agtcg aggca agac	cagta agcag ccctt gcc	54.8	tagtc gaggc aagac gtg	53.0	agccc ttgcc accag caca	55.0
611~1262	cccag gcgac aggtt ttg	ctcca gtctg gaatc tgttg a	50.0	caggc gacag gtttt gggct	53.0	ctgga atctg ttgaa cagta cct	53.0
1171~1550	tagca gctgg ctgag gtgag	tcatg caagt gaggc ccagg	63.1	tagca gctgg ctgag gtg	53.0	tcatg caagt gaggc cca	51.0
1422~2086	cttct ggagg gacca aagct	ctgag tcgag ccgtt caagt	57.0	gacca aagct tcaga tgcag	53.0	tcaag ttcaa atctt ggctt c	51.0
2008~2678	cagag gcact gagag gtaga	ctgag tctga gaccc aggc	59.0	gaggc actga gaggt agatc	47.0	gcctg cccag ccaga cag	47.0
2611-3046	gggaa aaaag gggaa gtcag	ctaag atctt ggctt ttccc	55.8	gaaaa aaggg gaagt cag	47.0	agatc ttggc ttttc cct	47.0
2989~3642	ctgga gaaga gagtc gatca tc	gcttt gtatc ggcct agagg	59.0	cgatc atctc agcaa attct c	53.0	tagag ggcag gaga aaaag agg	53.0
3591-4140	catgg ccctg acgat cctct	tggtc atcag gcaga gtgga	55.0	tggcc ctgac gatcc tct	53.0	cttag cattt cttcc actgc	53.0
4043~4545	ctggg cctaa gctgc tttc	acagt caagc agcag gcca	61.0	tgggc ctaag ctgct ttct	53.0	agtca agcag caggc cag	53.0
4482~5020	tgtca cccat gccct ccac	agagg ttggc tggca gtcag	57.0	tgccc tccac ctctg aag	53.0	cctat ctgcc ataga gattc ca	53.0
4951~5390	aaggg caggg ggatc tggtg	tgggg catgt tgggg cagag	64.7	agggc agggg gatct ggt	55.0	gcatg ttggg gcaga ggat	55.0
5323~6247	cagct ctgtg agtca ggcac	ctggg gacag aagag caaga	57.0	gtcag gcact atcag cgatg	55.0	agcaa gaggc tatca actcc	55.0
6143~6916	caagg gtgtg atact ggct	cttgg cagtg gggtc aca	55.0	ttgat agcct ctgc tcttc	53.0	cttgg ctgtt gaggg gc	53.0
6644~7200	cctca tacaa ccttc ttcac	tctct tagtg cagag ccagc	52.0	catac aacct tcttc acatc	51.0	tctct tagtg cagag cca	49.0
7155-7721	cagaa aagca taaca ccaat	ccata aaaag taaaa tcatt cct	44.0	aagca taaca ccaat cccag	53.0	tccta caagt taaat ttctc tt	51.0

Table 2 Oligonucleotide Primers used for full-length IRF-1 sequencing.

* Nucleotide numbers indicating locus amplified are numerated according to GenBank sequence of human IRF-1 gene (accession number L05072), the immediate upstream promoter sequence (495 nucleotides) was numerated by -1~-495 (accession number: X53095); Tm=Annealing temperature.

Figure 3 Distribution of fragments for complete IRF-1 gene sequencing

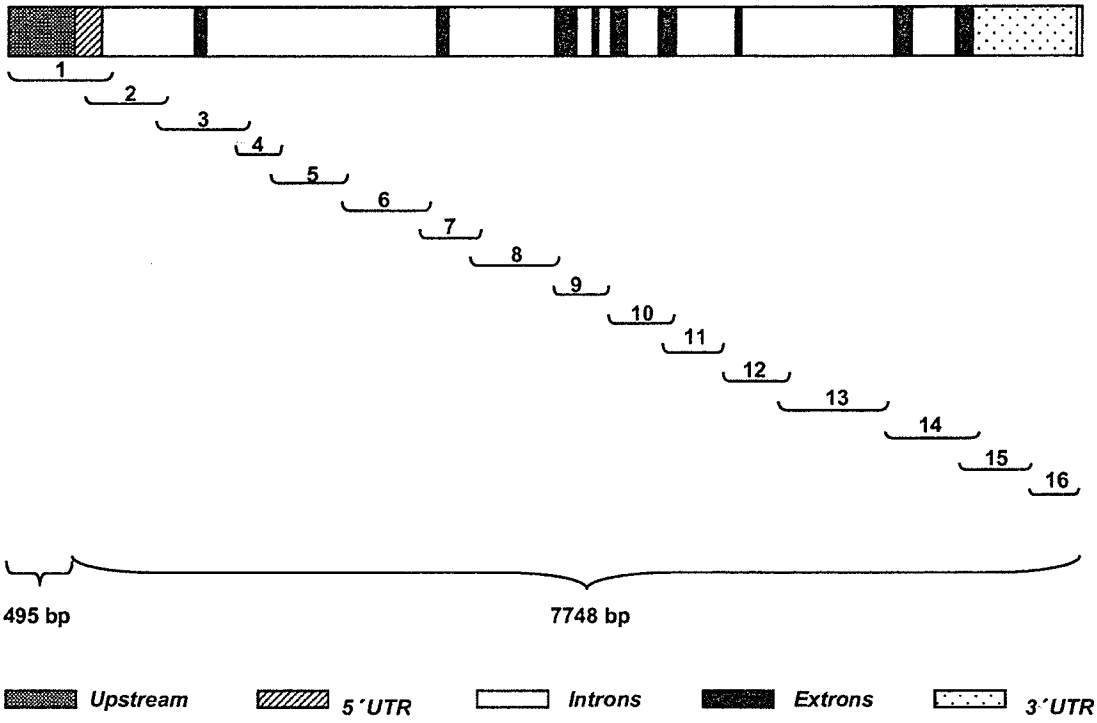


Figure 3 Distribution of fragments for complete IRF-1 gene sequencing.

The topmost panel shows a diagram of the genomic structure of human IRF-1gene and its immediate upstream promoter region. IRF-1 gene encompasses 7748 nt and contains 9 exons, 9 introns, one untranslated region (UTR) at both 5′ and 3′ ends. Its upstream promoter region is 495 nt in length. To facilitate sequencing experiment, the entire region was divided into 16 overlapping fragments numerated as 1 to 16 as shown in this diagram.

Microsatellite (MS) Typing

The IRF-1 MS marker is located in intron 7 of IRF-1 gene and contains a variable number of "GT" di-nucleotide repeats. MS typing was performed as previously described (Hughes, 1993). Briefly, primer oligonucleotides (5′-3′, annealing temperatures in parenthesis) designed on GenBank reference (accession number: L05072) specific for the IRF-1 MS marker were used: (f) ATGGCAGATAGGTCCACCGG and (r) TCATCCTCATCTGTTG TAGC (55°C). The primer oligonucleotides were obtained from PE-Applied Biosystems (Chesire, UK) with reverse primer end-labeled with FAM fluorescent marker. The 15 µl PCR reaction mixture contained 10 mM Tris- HCL (pH 8.3), 25 mM KCl, 1mM $MgCl_2$, 15 mM $(NH_4)_2SO_4$, 200 µM dNTP mixture, 1.5 pmol of each primer; 0.4 U of Amplitaq Gold Polymerase (Pekin-Elmer, Norwalk CT) and 50 ng of DNA template. PCR was performed on a thermocycler as follows: (i) an initial denaturation step of 14 min at 94°C; (ii) 35 cycles of 15 s at 94°C, 30 s at the specified annealing temperatures, and 30 s elongation at 72°C. The samples were then resolved on a ABI Prism 310 Genetic Analyzer (Applied BioSystems) and genotypes determined using GeneScan analysis software.

Association analysis on IRF-1 polymorphisms and resistance to HIV-1 infection

Differences in MS allele distributions and IRF-1 allele frequencies between the HIV-1 resistant and susceptible subjects were analyzed using Chi-square test. Log linear (multivariate) analysis was conducted to evaluate the independence of significant correlations of different IRF-1 polymorphisms and HIV-1 resistance. As SNPs at 619,

6516 and genotypes at IRF-1MS region were all found to be significantly associated with HIV-1 resistance phenotype in the first round of screening and all of them were in significant linkage disequilibrium, haplotype analysis were also performed to detected the correlation of different IRF-1 haplotypes and resistance to HIV-1 infection also using Chi-Square analysis. To assess the protective effects of IRF-1 gene variations in the HIV-1 free survival, Kaplan-Meier and Cox Regression survival analysis were conducted to compare HIV-1-free time prior to seroconversion in subjects who became infected during follow up. To assess the protective effects of IRF-1 gene variations in HIV-1 disease progression, Kaplan-Meier survival analysis was also conducted to assess the time differences of $CD4^+$ T cell decline in HIV-1 positive subjects with different IRF-1 genotypes/haplotypes. $CD4^+$ count decline was measured as the time between the first visit with $CD4^+$ T cell count above 400 cells/μl and the visit at which a count of below 400 cells/μl or 200 cells/μl was reached. To test the protective alleles frequency changes over time, those allele frequencies in subjects enrolled before and after 1993 were compared by Chi-Square test.

Functional examinations of key IRF-1 genetic variations

Although the three gene polymorphisms showing significant association with HIV-1 resistance phenotype in our subject cohort all reside in the non-coding region of IRF-1 gene, we hypothesized that functional impacts must be accompanying these key genetic variations. The intronic localization of the key IRF-1 polymorphisms correlated with HIV-1 resistance composes an obstacle to direct examination of their functions. We conducted functional tests by directly measuring IRF-1 basal level expression, IRF-1 response to exogenous stimulation, as well as IRF-1 mRNA expression and also the efficacy of HIV-1 transcription in PBMCs with different IRF-1 genotypes/haplotypes.

Basal level IRF-1 protein expression determination via Western blot

With the exception of early embryonic cells, IRF-1 is expressed at a low basal level in all cell types (Kroger et al., 2002). To initiate our functional tests, PBMCs from HIV-1 uninfected individuals with different IRF-1 genotypes concerning IRF-1 MS, 619A/C and 6516G/T were isolated following Ficoll-Hypaque isolation. For basal level IRF-1 expression assessment, 5×10^5 PBMCs from each subject were washed with PBS twice and then lysed with 1× Western Blot sample buffer (10 µl/10^5 cells). All cell lysate samples were boiled for 3 min and equal amount samples were then loaded into 10% SDS-polyacrylamide gel (10 µl /lane). The gel electrophoresis was performed at 160 volts for 60 min. The gel-separated proteins were then transferred onto a nitrocellulose membrane using a semi-dry transfer apparatus (BioRad) for 30 min at 15 volts. The

efficacy of transfer was visually checked by brief ponceau staining. The membrane was blocked with 5% nonfat milk in PBS with 0.05% Tween-20 (PBS-T) for 60 min at room temperature. Following a brief wash with PBS-T, the membrane was incubated overnight at 4 °C with a rabbit anti-human IRF-1 polyclonal antibody (1:500 in 2% nonfat milk in PBS-T) (Santa Cruz Biotechnology, Catalog number: sc-497). After incubation, the membrane was washed 6 times for 5 min each with PBS-T to remove unbound antibodies. The membrane was then incubated with peroxidase-conjugated AffiniPure goat anti-rabbit IgG (H+L) (1:16,000 in 2% nonfat milk in PBS-T) (Jackson ImmunoResearch, Catalog number: 111-035-003) for 1 h at room temperature. The membrane was washed again 6 times for 5 min each with PBS-T. Immunoreactive bands were detected by ECL Advance™ Western Blotting Detection Kit (Amersham Bioscience, Catalog number: RPN2135). As an internal control, the membrane was then washed and re-probed for actin, a house-keeping protein, with a goat anti-actin polyclonal antibody (1:1,000) (Santa Cruz Biotechnology, Catalog number: sc-1616) and peroxidase-conjugated AffiniPure rabbit anti-goat IgG (H+L) (1:20,000) (Jackson ImmunoResearch, Catalog number: 305-035-003). The procedures were the same as that for IRF-1 detection. Semi-quantitative analysis was conducted via spot densitometry quantification using Fluorchem™ 8800 Imaging System (Alpha Innotech Corporation, U.S.). The IRF-1/Actin ratio was used as comparator for intra-group comparisons using nonparametric Mann Whitney test (two-tailed).

Detection of IRF-1 responsiveness to exogenous IFN-γ stimulation

IRF-1 expression can be up-regulated by multiple agents including viral infection (Kroger et al., 2002). To test the IRF-1 responsiveness to exogenous stimulation, IFN-γ, the strongest and most commonly used stimulus for IRF-1 expression, was used in this experiment. PBMC samples from subjects with different IRF-1 genotypes were prepared as above and cultured with or without 1 ng/ml of IFN-γ for 18 hrs at 37°C with 5% CO_2. The cells were then harvested, washed with PBS and lysed with 1 × Western Blot sample buffer. Western blots were conducted for IRF-1 expression detection with actin as loading control again as described. The fold-increase in IRF-1 expression after IFN-γ stimulation over the basal level in the non-stimulated control cells was determined by semi-quantitative analysis via spot densitometry. This value was then utilized as the index for statistical comparison among different genotypic subgroups. Nonparametric two-tailed Mann-Whitney test was employed for statistical intra-group comparisons.

IRF-1 mRNA expression determination

To investigate the potential functional consequences of intronic polymorphisms in IRF-1 expression and function, we also examined the IRF-1 mRNA expression in selected HIV-1 uninfected subjects with distinct IRF-1 genotypes. Altered IRF-1 splicing pattern and skipping of different exons in IRF-1 mRNA had been previously reported and been associated with pathogenesis of chronic and acute myeloid leukemia (Green et al., 1999; Harada et al., 1994a; Tzoanopoulos et al., 2002). Both IRF-1 mRNA expression levels and exon skipping patterns in different subject groups were analyzed and compared in this set of experiments.

First-Strand cDNA Synthesis

Total RNA was extracted from unstimulated PBMCs with different IRF-1 haplotypes (concerning IRF-1 MS, 619A/C and 6516G/T) using Qiagen RNeasy Mini Kit (Qiagen, Catalogue number: 74104) following the manufacturer's instructions and the extracted RNA was quantified by spectrometry. Five hundred nanograms of RNA extract was used to perform first-strand cDNA synthesis with oligo (dT) primer and SuperScript ^{TM} First-strand Synthesis System (Invitrogen, Catalogue number: 11904-018) for reverse transcription PCR following instructions from manufacturer. The final cDNA products were treated with RNase H (Invitrogen, catalog number: 18021-014) at 37°C for 20 min before proceeding to PCR amplification of the target gene.

IRF-1 specific reverse transcription PCR (RT-PCR)

IRF-1 specific primers were designed to amplify a IRF-1 cDNA segment encompassing nt 67 to nt 1072 in the IRF-1 mRNA sequence (forward primer: TAAGAACCAGGCAACC; reverse primer: TGTAGACTCAGCCCAATA; annealing temperature: 57°C) (Tzoanopoulos et al., 2002). The PCR protocol described previously was used here for IRF-1 specific RT-PCR reactions. As an internal control, human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was amplified on an aliquot of the identical amount cDNA samples for all subjects using GAPDH specific primers (forward primer: TGATGACATCAAGAAGGTGGTGAAG; reverse primer: TCCTTGGAGGCCAT GTGGGCCAT; Tm: 55°C; 240bp in size)(Shimonovitz et al.,

1994). The size of the RT-PCR products was assessed by Ethidium Bromide staining and agarose gel electrophoresis.

TA cloning and sequencing analysis

Purified PCR products (5 to 10 ng) from above RT-PCR step were then TA-cloned using a TOPO TA cloning kit (Invitrogen, Carlsbad, Calif.), according to manufacturer's instructions. Thirty-two colonies of transformed *Escherichia coli* were picked for each reaction, cultured at 37 °C for 24 hours while shaking at 300 rpm, and used to prepare plasmid DNA using QIAprep 96 Turbo Miniprep kit (Qiagen, Catalog number:27191) by following instructions from the manufacturer.

To confirm the presence of inserts, restriction endonuclease digestion of cloned plasmids was conducted. The EcoRI restriction sites in the plasmid were utilized to release the cloned insert fragment by incubating the plasmid DNA and EcoRI at 37 °C for 1 hr (**Figure 4**). The products after EcoRI restriction digestion were screened for inserts by 1% agarose gel electrophoresis after ethidium bromide staining. The nucleotide sequences of insert DNAs were directly determined with the purified plasmid. Sequencing PCR was conducted using ABI Prism BigDye Terminator Version 3.0 and 3.1 Cycle Sequencing System. The total volume of each reaction was 10 µl containing 2 µl of Big Dye (Version 3 or 3.1, 1:1 dilution with provided dilution buffer), 10 ng of primer (T7 or T3 primer provided by manufacturer) and 450 ng of plasmid DNA. Sequencing cycling reactions were programmed as follow: (i) an initial denaturation step of 3 min at 96°C; (ii) 80 reaction cycles with each cycle consisting of 30 seconds at 96°C,

Figure 4 TA cloning plasmid demonstration and TOPO plasmid structure

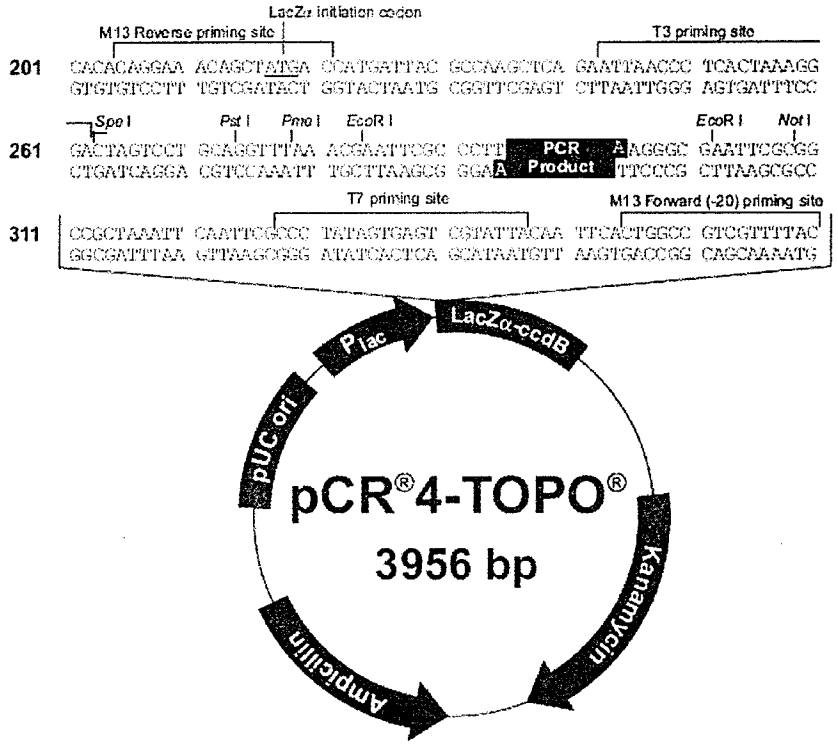


Figure 4 TA cloning demonstration and TOPO plasmid structure.

Cited from: TOPO Cloning kit manual (Invitrogen life technology). The general construction of TOPO cloning plasmid is shown in the bottom section of the diagram. Details of LacZα gene fusion site, EcoR I enzymatic digestion sites, T3 and T7 primer loci are all depicted in the top section.

30 seconds at the 50°C and 4 min at 60°C. All samples were sequenced using T7 and T3 primers (supplied in the TOPO TA cloning kit) for a cross-check. The amplified samples were precipitated using a NaOAc - ethanol method. To each cycle sequenced reaction, 1 µl of 3M NaOAc (pH 5.2) and 20 µl 95% ethanol was added and mixed well by vortexing. The samples were left to precipitate for 3 hrs in the dark at room temperature. Samples were then centrifuged at 12,000 rpm for 30 min to pellet the DNA content. The DNA pellets were then washed with 150 µl of 70% ethanol once and dried using 90°C heat-block for 2 minutes. The DNA samples were then ready for sequence resolution.

For sequence resolution, the DNA pellets were reconstituted with 20 µl high grade formamide (Applied Biosystems, Foster City, CA, USA) and resolved using Applied Biosystems (ABI) Prism 3100 Genetic Analyzer (Hitachi, Japan). Sequences obtained (1006 bp) were aligned and compared to GenBank reference sequences for human IRF-1 mRNA (NM_002198) using Sequencher software (Version 4.0.5, Gene Codes Corporation, U.S.). The exon distribution in IRF-1 mRNA was designated based on GenBank submission L05072. The exon expression profiles were determined by checking the presence of corresponding exon sequences in the TA cloned PCR products using Sequencher software (Version 4.0.5, Gene Codes Corporation, U.S.). Since the original start codon of IRF-1 mRNA is located at the beginning of previously described exon 2 region (L05072), all sequences with exon 2 skipped lose the original translation starting site and their open reading frame shifted downstream accordingly. Similar reading frame shift occurs while other exon skipping is present. The potential open reading frame shifts for sequences with skipped exon(s) were analysized using NCBI online Open Reading Frame Finder analysis tool

(http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The obtained IRF-1 nucleotide and predicted amino acid sequences were aligned with IRF-1 reference sequences using Mega 3.1 software (Kumar et al., 2004).

HIV-1 infection experiments

HIV-1 infection experiments are usually conducted using activated PBMCs or $CD4^+$ T cell pre-treated either by PHA or other stimuli. Such stimuli may affect IRF-1 expression in target cells and subsequently confound experiments aimed at testing the role of IRF-1 in HIV-1 transcription. Therefore, we established an alternative method to test the unstimulated PBMCs which more closely mimics the *in vivo* situation where HIV-1 infection naturally occurs.

A one-cycle pseudovirus construct which expressed vesicular stomatitis virus envelope G protein (VSV-G) and had a HIV-1 pNL4.3 backbone was used to infect PBMCs with different IRF-1 genotypes for functional assays in this set of experiments. The main advantages for using this pseudotyped HIV-1 include: (1) One cycle infection: it can only infect the cell once and no infectious progeny virus will be generated after initial infection; (2) Infecting all nucleated cell types with no requirement of pre-activation due to the possession of VSV-G envelope protein; (3) Sensitivity: the luciferase indicator in the virus construction facilitates the detection of HIV-1 LTR transcription at the earliest stages in a highly sensitive fashion.

HIV-1 pseudovirus construction

Two plasmids were employed in construction of the pseudovirus: pNL-Bru/E-/luc+ and SVCMV-VSV-G. For construction of pNL-Bru/E-/luc+ plasmid, a ApaI/SalI fragment of HIV-1 NLlucΔBglΔRI provirus (Ao et al., 2005) was replaced by a corresponding fragment from HxBru(Yao et al., 1995). This plasmid has a pNL4.3 HIV-1 strain backbone but the Env gene was knocked out and the Nef gene which is adjacent to the LTR region was replaced by firefly indicator luciferase gene (**Figure 5**). The second plasmid SVCMV-VSV-G encodes the vesicular stomatitis virus G protein. The two plasmids used were both constructed and generously provided by Dr. Xiaojian Yao (University of Manitoba, Canada). The VSV-G pseudotyped HIV-1 viral particles were produced by co-transfection of 293T cells with the pNL-Bru/E-/luc+ and the VSV-G expressor within biosafety level 3 containment. As the resulting progeny pseudoviruses lack an envelope as it is not encoded by the HIV-1 proviral constructs, functional viral replication and envelope-mediated syncytium formation is prevented, leading to a single viral replication cycle. Infection with VSV-G-pseudotyped HIV-1 viruses allows for a simultaneous infection of almost all cell types in the population (Bartz et al., 1996).

Transfection of 293T cells was conducted as follows in biosafety level 3 containment. Plasmid DNAs were prepared by combining 8 μg of pNL-Bru/E-/luc+ plasmid, 4.2 μg of SVCMV-VSV-G plasmid, 50 μl of 2.5M $CaCl_2$ and 450μl of distilled deionized H_2O in a 1.5 ml eppendorf tube in order and mixed well by pipetting. This mixture was then overlaid onto 500 μl 2× HeBS buffer in a separate tube. The bilayers were allowed to sit at room temperature for 20 min before being used to transfect 293T cells in later steps.

Figure 5 Construction of pNL-Bru/E$^-$/luc$^+$ plasmid

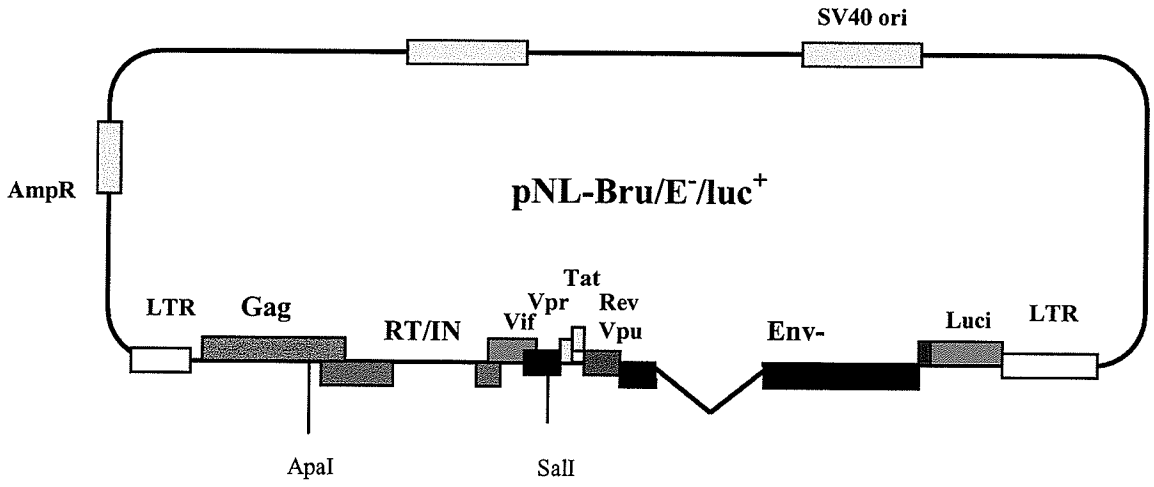


Figure 5 Construction of pNL-Bru/E$^-$/luc$^+$ plasmid.

The depicted pNL-Bru/E-/luc+ plasmid is one of the two plasmids utilized for the generation of pseudotyped HIV-1. For construction of pNL-Bru/E-/luc+ plasmid, an ApaI/SalI fragment of HIV-1 NLlucΔBglΔRI provirus was replaced by a corresponding fragments from HxBru. This plasmid has a pNL4.3 HIV-1 strain backbone but part of the Env gene was removed and the Nef gene which is right next to the LTR region was replaced by an indicator firefly luciferase gene.

DMEM medium containing 10% fetal calf serum (D10) was used for culturing 293T cells. Four million target 293T cells were plated in 100 × 15 mm Petri dishes the day before transfection so that cell monolayers were 60-80% confluent prior to transfection. Two tubes of the bi-layer plasmid/HeBS buffer preparations were mixed well by gentle pipetting and added into the culture medium. The cell culture medium and the added plasmid preparations were mixed well by gently rocking the dishes. All dishes were then put back into the incubator and incubated for another 8~12 hrs at which point the culture medium was replaced by an equal volume of fresh RPMI 1640 containing 20% FCS (R20) without disturbing the monolayer and the cells were cultured for another 36~40 hrs. The cell culture supernatant containing the pseudovirus particles was then collected into 50 ml tubes. The harvested supernatants were spun at 3000 rpm for 15 min to pellet the cell debris and filtered with 0.45μm syringe filter. The pseudovirus constructs were concentrated using a Centricon column (Millipore, 100,000MW, 15 ml capacity, Catalogue number: 91008, USA) by centrifugation for 30 min at 3000 rpm. The virus concentrate was then collected and kept at -70°C for further use. A reverse transcriptase assay as described below was utilized to determine viral titer in the concentrated pseudovirus stock.

Reverse Transcriptase (RT) activity determination

The RT assay was performed using reaction cocktails by adding the reagents listed in **Table 3** in order. In brief, Tris Hydrochloride, $MgCl_2$, KCL, EGTA, Triton X-100, Ethylene glycol stock solution and H_2O as indicated were added into a 1.5 ml centrifuge

Table 3 Reagents for reverse transcriptase assay cocktail

	Stock	Volume	Final conc
1) Tris Hydrochloride (pH 7.9)	2 M	2.5 µl	50 mM
2) $MgCl_2$	0.2 M	2.5 µl	5 mM
3) KCl	2 M	7.5 µl	150 mM
4) EGTA	20 mM	2.5 µl	0.5 mM
5) Triton X-100	2 %	2.5 µl	0.05 %
6) Ethylene glycol	80 %	2.5 µl	2 %
7) H_2O		17 µl	
8) DTT (dithiotheritol)	0.2 M	2.5 µl	5 mM
9) GSH (reduced gluthatione)	12 mM	2.5 µl	0.3 mM
10) Poly-A-oligo (dT)	10 U/ml	4 µl	50 µg/ml
11) ^3H dTTP	2.5 µci/µl	4 µl	20 µci
12) HIV-1 culture fluid		50 µl	
Total volume		100 µl	

tube. These reaction tubes were then brought into the Level 3 containment laboratory and 50 µl of 1:10, 1:100, 1:200 diluted pseudovirus concentrate were added to the reaction tubes. Fifty milliliter of fresh R20 medium was added to the negative control tube. Inactivation of the pseudovirus was conducted by incubating the above mixture at room temperature for 20 min. All reaction tubes were then taken out of level 3 confinement. The indicated volumes of DTT, GSH, Poly-A-oligo (dT) and ^{3}H dTTP were then added into all reactions and mixed well by gentle vortexing. RT activity was determined by incubating at 33°C for 22 hrs and stopped by adding 1 ml of refrigerated 10% trichloroacetic acid (TCA). The reaction tubes were then placed on ice for 2 hrs for precipitation.

The RT reactions were then harvested using the following procedures. The filter membranes (Fisher Scientific, Glass Fiber Filter Circles (G4) Catalogue number: 09-804-24C) were pre-wet with 10% TCA before being mounted onto the Millipore filtering set (Millipore, Catalogue number: XX270255). The vacuum pump was then turned on and the reaction mixtures from above steps were then added into the corresponding wells. The reaction tubes were rinsed with 5% cold TCA twice and then the vacuum pump was turned off when the liquid has passed through in all wells. The membranes were then transferred into labeled scintillation vials and dried for 2 hrs at 37°C or 1 hr at 70°C. Five milliliters scintillation liquid was then added to each vial and the scintillation count was read using a Beckman scintillation counter (Liquid Scintillation System, Beckman, LS6000TA). The Counts per Minute (cpm) readouts were taken for the determination of virus concentration. A viral titer of five cpm/cell would be used in all further infection experiments.

Pseudovirus infection of PBMCs

Frozen PBMC samples obtained from 13 HIV-1 negative subjects with distinct IRF-1 haplotypes were employed for this set of experiments. The cells were thawed at 37°C, washed with RPMI1640 medium with 20% FCS (R20) once and incubated at 37°C overnight in R20 for complete recovery. Cell viability was measured using Trypan blue staining. Only samples with viability higher than 90% were utilized in further infection experiments. Viable cells were resuspended to 1×10^7 cells/ml with R20 and pseudovirus stock was added into the cell suspension to a final viral concentration at 5 cpm/cell. The cell-virus mixture was then cultured at 37°C for 18 hrs with frequent shaking to infect the host cells. The cells were then pelleted by centrifugation and washed twice with fresh R20 to remove the extracellular free viral particles. The cells were then resuspended with R20, aliquoted into 48-well plate at 1×10^6 cells/500 µl /well and cultured at 37°C with 5% CO_2. The cell samples were harvested at 0, 24, 48, 96, 144 hrs after infection. At each time point, the cells were transferred into 1.5 ml eppendorf tubes and washed with PBS twice and pelleted after the final wash. The cells pellets were then dislodged by gentle vortex and fifty microliter of 1 × luciferase assay lysis buffer was then added to each cell sample to lyse the cells completely. The cell lysates were then stored under -70°C for further luciferase activity assay and Western Blot analysis.

Luciferase activity determination

Luciferase activity was determined by utilizing a Microplate Luminometer (EG&G

Berthold Microplate Luminometer LB 96V) which was programmed for a 5 seconds delay and a 5 seconds measurement time prior to the luciferase activity measurement. Ten microliter of the cell lysate obtained as described above was added into Nunc-Immuno^{TM} 96-well plate (Nalge Nunc International, Rochester) and then 50 µl of luciferase assay reagent was added to the same well. The plate was then mounted onto the luminometer and luminance signal measurement was immediately determined. The read-out was calculated and plotted by calculating the fold increase of the luciferase activity at specific time points (24h, 48h,96h,144h post infection) compared to the baseline at time 0.

Determination of IRF-1 response after pseudotyped HIV infection

A 10 µl aliquot of cell lysate from the luciferase assay from 3 subjects representing 3 distinct IRF-1 genotypes (1 subject with "179179 619AA6516GG" haplotype, 1 with "179^{-}179^{+}+619AC6516GT" and 1 with "179^{-}179^{-}619CC6516TT") was taken for determination of IRF-1 response after infection with the pseudotyped HIV-1. This aliquot was mixed with same volume of 2× Western Blot sample buffer. All cell lysate samples were then boiled for 3 min and then loaded into 10% SDS-polyacrylamide gel. The gel electrophoresis was performed at 160 volts for 60 min. The gel-separated proteins were then transferred onto a nitrocellulose membrane using a semi-dry transfer apparatus (BioRad) for 30 min at 15 volts. The efficacy of transfer was visually checked by brief ponceau staining. The membrane was blocked with 5% nonfat milk in PBS with 0.05% Tween-20 (PBS-T) for 60 min at room temperature. Following a brief wash with PBS-T, the membrane was incubated overnight at 4 °C with a rabbit anti-human IRF-1 polyclonal

antibody (1:500 in 2% nonfat milk in PBS-T) (Santa Cruz Biotechnology, Catalog number: sc-497). After incubation, the membrane was washed 6 times for 5 min each with PBS-T to remove unbound antibodies. The membrane was then incubated with peroxidase-conjugated AffiniPure goat anti-rabbit IgG (H+L) (1:16,000 in 2% nonfat milk in PBS-T) (Jackson ImmunoResearch, Catalog number: 111-035-003) for 1 h at room temperature. The membrane was washed again 6 times for 5 min each with PBS-T. Immunoreactive bands were detected by ECL Advance™ Western Blotting Detection Kit (Amersham Bioscience, Catalog number: RPN2135). As internal control, the membrane was then washed and re-probed for actin, a house-keeping protein, with a goat anti-actin polyclonal antibody (1:1,000) (Santa Cruz Biotechnology, Catalog number: sc-1616) and peroxidase-conjugated AffiniPure rabbit anti-goat IgG (H+L) (1:20,000) (Jackson ImmunoResearch, Catalog number: 305-035-003). The procedures were the same as that for IRF-1 detection. Semi-quantitative analysis was conducted via spot densitometry quantification using Fluorchem™ 8800 Imaging System (Alpha Innotech Corporation, U.S.). The fold increase of IRF-1/actin ratio upon time 0 control in spot densitometry was taken as index for intra-group comparisons. These Western Blot results were then compared with the luciferase assay readouts to investigate potential correlation between IRF-1 protein level and HIV-1 replicative ability.

Results

Section One: IRF-1 gene sequencing and polymorphism identification

IRF-1 gene is highly polymorphic in Kenyan subjects

In previous studies, genetic variations in the IRF-1 MS marker were correlated with the resistance to HIV-1 infection phenotype after screening of host IL-4 gene cluster on human chromosome 5 (Ball, 2001). The potential role of IRF-1 as a candidate gene to explain, in part, resistance to infection by HIV-1 was supported by previous data demonstrating that IRF-1 plays a key role in host innate and adaptive immunity (Kroger et al., 2002; Mamane et al., 1999), and in mediating HIV-1 transcription and replication (Battistini et al., 2002; Sgarbanti et al., 2002). The IRF-1 MS marker is located in the 7^{th} intronic region of IRF-1 gene and encodes no known functional changes. Hence we assumed that it might act as a biomarker for other functional gene variation (s) in the IRF-1 gene that mediates the association between IRF-1 and HIV-1 resistance. The IRF-1 gene is 7721 bp in length and has a 495 bp immediate upstream promoter region. It was feasible to conduct population-based complete gene sequencing on a number of subjects to see if the IRF-1 MS is in linkage disequilibrium with a functional polymorphism and so that we can conduct further analysis on them.

In total, 507 subjects from our Kenyan female sex worker cohort were randomly selected and their IRF-1 genes were sequenced. We discovered that our tested Kenyan population displayed extensive diversity throughout the IRF-1 locus. All primers and

PCR reactions worked as expected and representative PCR amplified fragments of the expected sizes are shown in **Figure 6**. Fifty-three SNPs were identified in the IRF-1 gene and its promoter region. The genotype and allelic frequencies of all SNPs were calculated and are depicted in **Table 4**. All 53 SNPs had been submitted to Single Nucleotide Polymorphism Database (dbSNP) of NCBI's Entrez system and the corresponding SNP identification numbers are listed in **Table 5**. Twenty-six of these SNPs were novel and had not been previously reported. Twenty-seven of them had been previously described either in publications or by the International Human Genome Sequencing Consortium through the NCBI-based SNP database (Database1, ; Database2, ; Donn et al., 2001; Nakao et al., 2001; Noguchi et al., 2000; Saito et al., 2001; Saito et al., 2002). However, a number of previously reported SNPs were not observed in the subjects from this study, including 5551T/G, 4950A/G, 5558T/C, 5636G/A, 7662C/- and 6355G/A, demonstrating a high degree of IRF-1 genetic polymorphism in different ethnic populations. Beside the 26 novel SNPs, two novel insertion mutations were identified in the second intron of IRF-1 ("CA" insertion between nt 2592 and 2593 in L05072 or nt 131900079 and 131900078 in chromosome 5 (TG→T*CA*G")) and in the 3′ UTR (an "A or G" insertion between nt 7607 and 7608 or nt 131895046 ~ 131895044 in chromosome 5 (AG→A*A/G*G)) with genotype frequencies at 16.33% and 93.33% respectively. A 16 nt deletion was identified in intron 7 (**Table 6, Figure 7**) in 28.30% of subjects tested. This mutation was completely linked with the "T" allele of SNP at 4816. Interestingly, this deletion was located close to the IRF-1 microsatellite region with multiple "GT" di-nucleotide repeats (4907→28 in L05072), a putative site for the formation of Z-DNA (Cha et al., 1992).

Figure 6 PCR amplified segments of IRF-1 gene

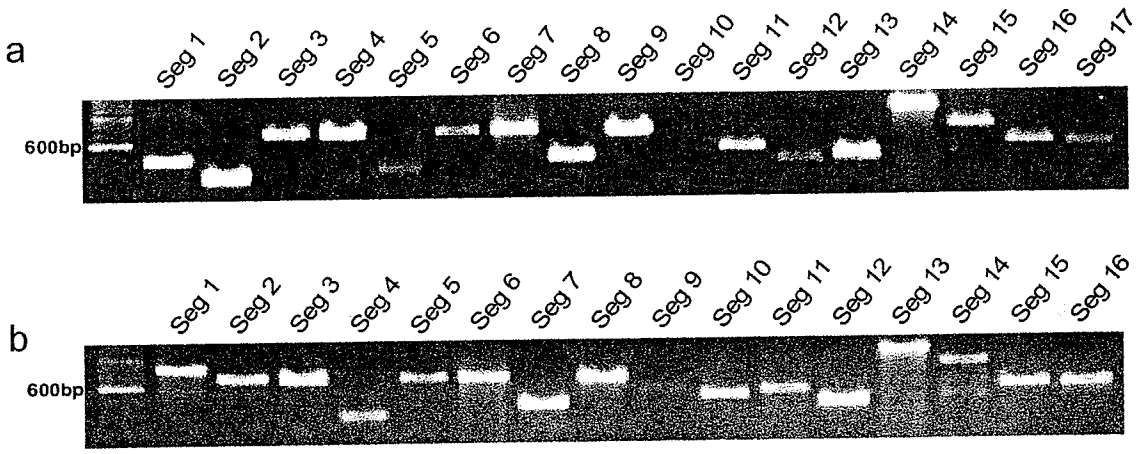


Figure 6 PCR amplified segments of IRF-1 gene.

The top panel (a) shows all 17 PCR amplified IRF-1 gene segments using primers listed in table 1. For full length IRF-1 sequence determination, several new primers were re-designed and segment 1 and 2 were merged in the first round of PCR. The bottom panel (b) of the figure shows all 16 PCR amplified segments of IRF-1 gene during IRF-1 full length determination using primers listed in Table 2. Bands for segment 10 in (a) and segment 9 in (b) are not as strong as the other bands but these didn't affect the final sequencing results.

Table 4 Polymorphisms in the IRF-1 gene and its promoter region detected in a Kenyan population

Loci*	Genotype	%*	Allele	%*
-415 (n=187)#	AA	18.7	A	44.4
	AC	51.3	C	55.6
	CC	29.9		
-386 (n=188)	CC	96.8	C	98.4
	CT	3.2	T	1.6
-281 (n=188)	GG	99.5	G	99.7
	GA	0.5	A	0.3
-90 (n=187)	GG	98.9	G	99.5
	GA	1.1	A	0.5
142 (n=289)	CC	91.7	C	95.8
	CT	8.3	T	4.2
379 (n=269)	CC	85.1	C	92.6
	CT	14.9	T	7.4
858 (n=48)	CC	97.9	C	99.0
	CT	2.1	T	1.0
2529 (n=56)	AA	94.6	A	97.3
	AG	5.4	G	2.7
3267 (n=48)	TT	33.3	T	59.4
	TC	52.1	C	40.6
	CC	14.6		
4318 (n=93)	AA	24.7	A	51.6
	AG	53.8	G	48.4
	GG	21.5		
4420 (n=90)	CC	98.9	C	99.4
	CT	1.1	T	0.6
4659 (n=130)	GG	29.2	G	51.9
	GA	45.4	A	48.1
	AA	25.4		
4816 (n=120)	GG	34.2	G	52.9
	GT	37.5	T	47.1
	TT	28.3		
5617 (n=44)	TT	31.8	T	59.1
	TC	54.6	C	40.9
	CC	13.6		
6467 (n=376)	AA	21.8	A	46.9
	AG	50.3	G	53.1
	GG	27.9		
6639 (n=171)	AA	32.7	A	54.4
	AG	43.3	G	45.6
	GG	24.0		
7238 (n=90)	CC	97.8	C	98.9
	CT	2.2	T	1.1
7447 (n=90)	CC	98.9	C	99.4
	CA	1.1	A	0.6

Loci	Genotype	%	Allele	%
-410 (n=187)	GG	18.2	G	43.9
	GA	51.3	A	56.1
	AA	30.5		
-300 (n=188)	GG	26.1	G	54.0
	GA	55.8	A	46.0
	AA	18.1		
-280 (n=188)	GG	99.5	G	99.7
	GC	0.5	C	0.3
-65 (n=188)	GG	98.9	G	99.5
	GA	1.1	A	0.5
154 (n=290)	CC	34.1	C	58.6
	CT	49.0	T	41.4
	TT	16.9		
619 (n=507)	AA	8.5	A	29.7
	AC	42.4	C	70.3
	CC	49.1		
1970 (n=48)	TT	35.4	T	60.4
	TG	50.0	G	39.6
	GG	14.6		
2936 (n=83)	TT	34.9	T	61.4
	TC	53.0	C	38.6
	CC	12.1		
4220 (n=94)	TT	95.7	T	97.9
	TC	4.3	C	2.1
4379 (n=93)	TT	24.7	T	51.6
	TC	53.8	C	48.4
	CC	21.5		
4583 (n=129)	AA	28.7	A	51.6
	AC	45.7	C	48.4
	CC	25.6		
4692 (n=130)	CC	99.2	C	99.6
	CT	0.8	T	0.4
5203 (n=93)	CC	24.7	C	51.6
	CT	53.8	T	48.4
	TT	21.5		
6163 (n=47)	GG	34.0	G	59.6
	GT	51.1	T	40.4
	TT	14.9		
6516 (n=483)	GG	9.1	G	29.6
	TG	41.0	T	70.4
	TT	49.9		
6936 (n=36)	CC	94.4	C	97.2
	CT	5.6	T	2.8
7303 (n=90)	TT	98.9	T	99.4
	TC	1.1	C	0.6
7489 (n=90)	CC	97.8	C	98.9
	CT	2.2	T	1.1

Loci	Genotype	%	Allele	%
-388 (n=188)	TC	4.3	T	2.1
	CC	95.7	C	97.9
-297 (n=186)	GG	98.9	G	99.5
	GA	1.1	A	0.5
-203 (n=188)	GG	99.5	G	99.7
	GA	0.5	A	0.3
53 (n=271)	CC	91.9	C	95.9
	CG	8.1	G	4.1
197 (n=291)	GG	98.6	G	99.3
	GA	1.4	A	0.7
831 (n=48)	GG	91.7	G	95.8
	GC	8.3	C	4.2
2392 (n=66)	AA	47.0	A	67.4
	AT	40.9	T	32.6
	TT	12.1		
3168 (n=48)	TT	97.9	T	99.0
	TC	2.1	C	1.0
4227 (n=95)	AA	25.3	A	52.1
	AG	53.7	G	47.9
	GG	21.1		
4396 (n=93)	AA	24.7	A	51.6
	AG	53.8	G	48.4
	GG	21.5		
4656 (n=130)	GG	87.7	G	93.8
	GA	12.3	A	6.2
4709 (n=130)	GG	28.5	G	51.5
	GT	46.1	T	48.5
	TT	25.4		
5261 (n=93)	GG	24.7	G	51.6
	GA	53.8	A	48.4
	AA	21.5		
6460 (n=377)	GG	21.8	G	46.8
	GA	50.1	A	53.2
	AA	28.1		
6637 (n=171)	CC	22.2	C	42.7
	CT	40.9	T	57.3
	TT	36.8		
7175 (n=36)	CC	94.4	C	97.2
	CT	5.6	T	2.8
7311 (n=92)	GG	23.9	G	51.6
	GA	55.4	A	48.4
	AA	20.7		

Table 4 Polymorphisms in the IRF-1 gene and its promoter region detected in a Kenyan population.

* Nucleotide numbers indicating site of polymorphism were numerated according to the GenBank reference sequence for human IRF-1 (accession number L05072), the immediate upstream promoter sequence (495 nucleotides) was numerated by -1~-495 (accession number: X53095); # Numbers in brackets indicate the number of individuals tested; ^ Genotype or allele frequencies were calculated directly based upon number sequenced for each polymorphism .

Table 5 SNPs in IRF-1 gene and its promoter region submitted to dbSNP

NCBI Assay ID	Submitter SNP ID	Allele	Sample size	refSNP ID
ss28462041	HJsequ2936	T/C	166	rs13170412
ss28502689	HJsequ-203	G/A	376	rs17848387
ss28502690	HJsequ-90	G/A	374	rs17848388
ss28502691	HJsequ-65	G/A	376	rs17848389
ss28502692	HJsequ53	C/G	302	rs17848390
ss28502693	HJsequ142	C/T	312	rs17848391
ss28502694	HJsequ154	C/T	312	rs17848392
ss28502695	HJsequ197	G/A	306	rs17848393
ss28502696	HJsequ379	C/T	210	rs17848394
ss28502697	HJsequ619	A/C	208	rs17848395
ss28502698	HJsequ831	G/C	96	rs17848396
ss28502699	HJsequ-415	A/C	374	rs17848397
ss28502700	HJsequ858	C/T	96	rs17848398
ss28502701	HJsequ1970	T/G	96	rs17848399
ss28502702	HJsequ2392	A/T	132	rs17848400
ss28502703	HJsequ2529	A/G	112	rs17848401
ss28502704	HJsequ3168	T/C	96	rs17848402
ss28502705	HJsequ3267	T/C	96	rs17848403
ss28502706	HJsequ4220	T/C	186	rs17848404
ss28502707	HJsequ4227	A/G	186	rs17848405
ss28502708	HJsequ4318	A/G	186	rs17848406
ss28502709	HJsequ4379	T/C	186	rs17848407
ss28502710	HJsequ-410	G/A	374	rs17848408
ss28502711	HJsequ4396	A/G	186	rs17848409
ss28502712	HJsequ4420	C/T	180	rs17848410
ss28502713	HJsequ4583	A/C	258	rs17848411
ss28502714	HJsequ4656	G/A	260	rs17848412
ss28502715	HJsequ4659	G/A	260	rs17848413
ss28502716	HJsequ4692	C/T	260	rs17848414
ss28502717	HJsequ4709	G/T	260	rs17848415
ss28502718	HJsequ4816	G/T	240	rs17848416
ss28502719	HJsequ5203	C/T	186	rs17848417
ss28502720	HJsequ5261	G/A	186	rs17848418
ss28502721	HJsequ-388	T/C	376	rs17848419
ss28502722	HJsequ5617	T/C	88	rs17848420
ss28502723	HJsequ6163	G/T	94	rs17848421
ss28502724	HJsequ6460	G/A	554	rs17848422
ss28502725	HJsequ6467	A/G	554	rs17848423
ss28502726	HJsequ6516	G/T	526	rs17848424
ss28502727	HJsequ6637	C/T	92	rs17848425
ss28502728	HJsequ6639	A/G	92	rs17848426
ss28502729	HJsequ6936	C/T	72	rs17848427
ss28502730	HJsequ7175	C/T	72	rs17848428
ss28502731	HJsequ7238	C/T	180	rs17848429
ss28502732	HJsequ-386	T/C	376	rs17848430
ss28502733	HJsequ7303	T/C	180	rs17848431
ss28502734	HJsequ7311	G/A	184	rs17848432
ss28502735	HJsequ7447	C/A	180	rs17848433
ss28502736	HJsequ7489	C/T	180	rs17848434
ss28502737	HJsequ-300	G/A	376	rs17848435
ss28502738	HJsequ-297	G/A	372	rs17848436
ss28502739	HJsequ-281	G/A	376	rs17848437
ss28502740	HJsequ-280	G/C	376	rs17848438

Table 5 SNPs in IRF-1 gene and its promoter region submitted to dbSNP.

The listed were all 53 SNPs identified in the Kenyan subject population and submitted to data base for SNPs (dbSNP) of NCBI's Entrez system with submission handle name: UMPLUMMERLAB. Related dbSNP information for these SNPs is depicted.

Table 6 Deletion mutation identified in intron 7 of IRF-1 gene

	$4787^{\$}\rightarrow$ $\leftarrow 4819^{\$}$
$L05072^{\#}$:	AGGGTGAGGAAGGAAGTAGGGTAGGGGTG G GAA
With del :	AGGG**********GTAGGGGTG T GAA

Table 6 Deletion mutation identified in intron 7 of IRF-1 gene.

A 16- nucleotide deletion mutation was identified at 4791 to 4806 in intron 7 of IRF-1 gene. The detailed indication of deleted nucleotides is shown in this table. $^{\$}$ Nucleotides were numerated according to GenBank submission L05072; $^{\#}$ GenBank reference sequence for human IRF-1 from submission L05072.

Figure 7 A 16- nucleotide deletion mutation identified in intron 7 of IRF-1

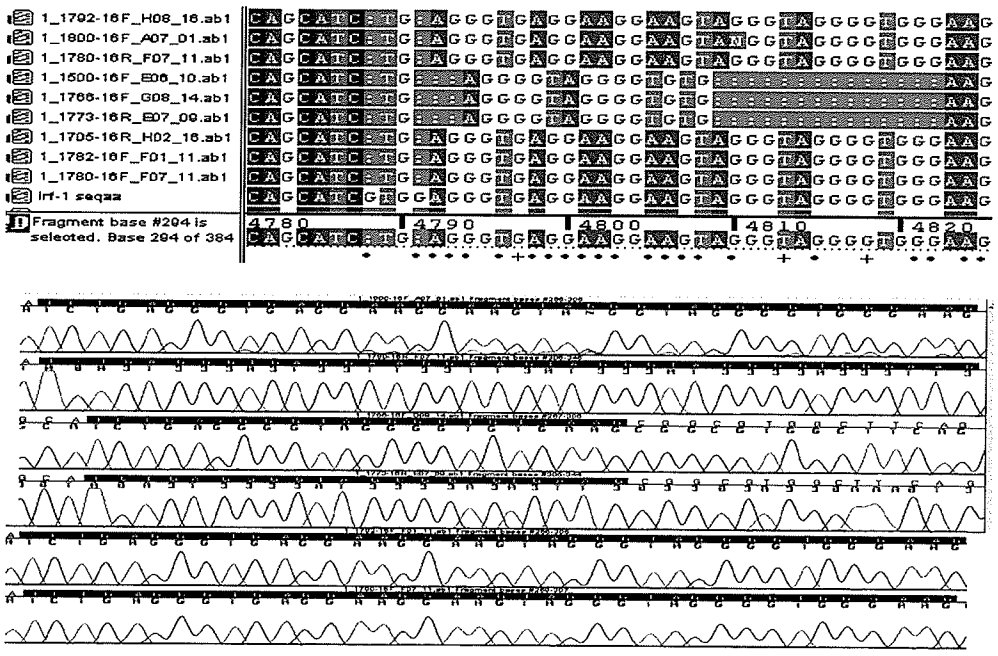


Figure 7 A 16- nucleotide deletion mutation identified in intron 7 of IRF-1

A 16- nucleotide deletion mutation was identified at 4791 to 4806 in intron 7 of IRF-1 gene. Alignment of obtained IRF-1 sequences from several tested subjects with GenBank reference sequence (L05072) using Sequencher software is shown in the top panel of this figure. The corresponding chromatogram demonstration is shown in the bottom section.

IRF-1 polymorphisms mainly reside in the intronic and promoter regions

Of all the 53 SNPs identified in IRF-1 gene and its promoter region, 42 resided in the IRF-1 gene itself. Except for two silent mutations at 4396 A/G and 4420 C/T in exon 7, all the other 40 SNPs, the 2 insertions and the deletion mutation were located in non-coding regions of IRF-1 gene. Among these 40 SNPs, 4 of them were located in exon 1 (5' untranslated region (UTR)) (53 C/G, 142 C/T, 154 C/T, 197 G/A), and 7 in the 3' UTR (6936 C/T, 7175 C/T, 7238 C/T, 7303 T/C, 7311 G/A, 7447 C/A, 7489 C/T) and the others were distributed in introns I, II, III, VI, VII, VIII and IV respectively (**Figure 8**). Another previously reported silent mutation 6355G/A (Noguchi et al., 2000) was not found in any subjects. The remaining 11 SNPs were located in the immediate upstream IRF-1 promoter region (-415 A/C, - 410 G/A, -388 T/C, -386 C/T, -300 G/A, -297 G/A, -281 G/A, -280 G/C, -203 G/A, -90 G/A, and -65 G/A). Among them, SNPs at -386, -203 and -65 were located within three Sp1 binding sites respectively while none of these SNPs were related with IFN-γ-activated sequence (GAS) and NF-κB recognition (**Figure 9**). The putative recognition sites for known transcriptional factors are referred to: Harada H, et al. (1994) Mol Cell Biol, 14(2):1500-1509).

Linkage disequilibrium (LD) shown to be common in IRF-1 gene

Linkage disequilibrium (LD) between SNPs was observed throughout the IRF-1 gene. Pair-wise LD analysis revealed that LD was extensive in the IRF-1 gene and it was observed to some degree to be present between most SNPs. As shown in **Table 7**, complete linkage was observed among SNPs at positions 4227, 4318, 4379, 4396, 5203,

Figure 8 IRF-1 genomic structure and distribution of polymorphic sites

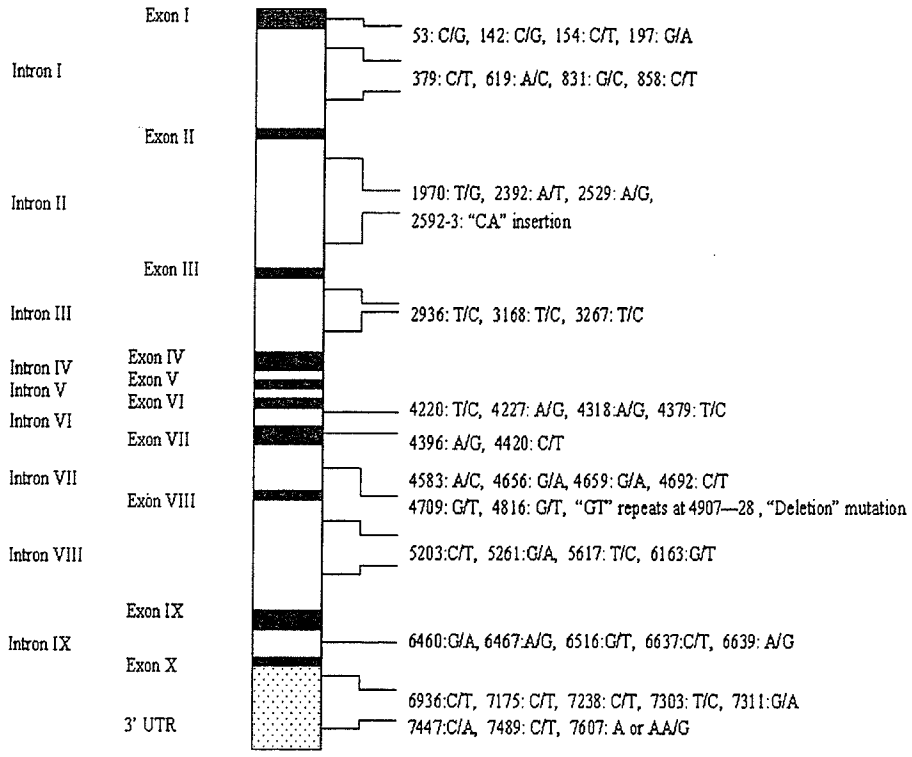


Figure 8 IRF-1 genomic structure and distribution of polymorphic sites.

Nucleotide numbers indicating loci were assigned according to GenBank sequence for IRF-1 (accession number L05072). The human IRF-1 gene and its immediate upstream promoter sequences had been mapped to chromosome 5q31.1. IRF-1 gene is composed of 10 exons (exon 1 is untranslated terminal region, 5' UTR), 9 introns and one 3' UTR. Forty SNPs were distributed in the 5' UTR, introns 1, 2, 3, 6, 7, 8, 9, and in the 3' UTR. Two silent mutations (4396 A/G and 4420C/T) occurred in exon 7. Two insertion mutations were identified in intron 2 and 3' UTR respectively. A 16 nt deletion mutation was discovered just before the "GT" dinucleotide repeat (IRF-1 microsatellite marker) in intron 7.

Figure 9 IRF-1 promoter sequence and distribution of polymorphic sites

Figure 9 IRF-1 promoter sequence and distribution of polymorphic sites.

The immediate upstream promoter region of the IRF-1 gene is 495 bp in length. The putative recognition sites for known transcription factors are indicated; GAS: IFN-γ-activated sequence; NF-kB: nuclear factor-kappa B; Sp1: transcription factor specificity protein1; CAAT box: a CCAAT sequence being recognized by some transcriptional factors, i.e. nuclear factor-Y (NF-Y). Eleven SNPs marked with * dispersed in the IRF-1 promoter region. SNPs at -386, -203 and -65 were located within 3 Sp1 binding sites respectively. No SNPs were found regions related with GAS and NF-κB recognition. This figure was adapted from: Harada H, et al. (1994) Mol Cell Biol, 14:1500-1509. * Nucleotide numbers indicating loci were assigned according to GenBank sequence for IRF-1 promoter (accession number X53095) and correspond to nt 131903176~131902682 of human chromosome 5.

Table 7 Complete linkage between 4227 and other SNPs within the IRF-1 gene

4227[#]	4318	4379	4396	5203	5261	6460	6467	7311
A	A	T	A	C	G	G	A	G
A/G	A/G	T/C	A/G	C/T	G/A	G/A	A/G	G/A
G	G	C	G	T	A	A	G	A

Table 7 Complete linkage between 4227 and other SNPs within the IRF-1 gene.

Complete linkage between SNPs at 4227, 4318, 4379, 4396, 5203, 5261, 6460, 6467 and 7311 was discovered in IRF-1 gene. The allele compositions for each locus are listed in this table. This linkage was observed in all tested subjects including our local healthy controls. [#] Nucleotide numbers indicating loci were according to GenBank sequence for IRF-1 (accession number: L05072).

5261, 6460, 6467 and 7311 (identified hereafter as "4227 cluster"). Pair-wise LD analysis results for all identified SNPs with minor allele frequencies higher than 0.1 is presented in **Table 8** and all these SNPs were found to be in Hardy-Weinberg equilibrium. The 4816 SNP was also associated with the 16 bp deletion mutation in intron 7, suggesting this region might be undergoing some selective pressure. LD existed even between SNPs separated by long intervals and distributed in different functional regions of IRF-1 gene. Partial linkage disequilibrium was also observed between the "GT" di-nucleotide repeats in IRF-1 MS region in intron 7, the "4227 cluster" and other identified SNPs such as 619A/C and 6516G/T (data not shown). This suggested that multivariate or haplotype analysis might be necessary for full evaluation of the effect of IRF-1 gene variations in their functional context or in correlation with disease conditions. We were unable to confirm a previous report that nearly complete LD existed between SNPs at –300 and 4396 (Noguchi et al., 2000).

Consistent discrepancies identified between GenBank references and our population-based sequence dada

It is noteworthy that thirty-five consistent discrepancies were identified between our population-based sequence data and the current GenBank sequences for IRF-1 and its upstream promoter. Consistent discrepancies were defined here as the same nucleotide variation found in all tested subjects but different from the GenBank submissions. These discrepancies are depicted in **Table 9**. The presence of these altered nucleotides has been confirmed in a control population of 28 healthy local donors with non-African genetic backgrounds, including Asians, Europeans and North Americans, suggesting these

Table 8 Pairwise LD estimates for all SNPs with minor allele frequency higher than 0.1

	-410	-300	154	619	1970	2392	2936	3267	4227	4583	4659	4709	4816	5617	6163	6516	6637	6639
-415	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
-410		***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
-300			***	**	***	***	***	***	***	***	***	***	***	***	***	***	***	***
154				***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
619					***	***	*	***	***	***	***	***	***	***	***	***	***	***
1970						***	***	***	***	***	***	***	***	***	***	***	***	***
2392							***	***	***	***	***	***	***	***	***	***	***	***
2936								***	***	***	***	***	***	***	***	***	***	***
3267									***	***	***	***	***	***	***	***	***	***
4227										***	***	***	***	***	***	***	***	***
4583											***	***	***	***	***	***	***	***
4659												***	***	***	***	***	***	***
4709													***	***	***	***	***	***
4816														***	***	***	***	***
5203														***	***	***	***	***
5617															***	***	***	***
6163																***	***	***
6460																***	***	***
6516																	***	***
6637																		***

Table 8 Pair-wise LD estimates for all SNPs with minor allele frequency higher than 0.1.

Pair-wise LD analysis were conducted for all identified SNPs with minor allele frequencies higher than 0.1 by using PyPop software. The results were shown based on P value (* $P<0.05$, ** $P<0.01$, *** $P<0.001$). Nucleotide numbers indicating loci were numerated according to GenBank sequence for IRF-1 (accession number: L05072). The "4227 cluster" was represented by the 4227 SNP itself.

Table 9 Consistent discrepancies between GenBank sequences and our detected sequences

Position[Δ]	L05072	Variation	Position	L05072	Variation	Position	L05072	Variation
97-100	CGCG*	CGGCGG*#	1751	C	CC*	4746	T	—*
111	C*	A#	1867	T	TT*	4782-6	CGTGG	C—T—G*
153-6	TCCG	TC/TCGCGCG*#	2044	G	A*	5000	G	GG*
242-3	CC	CGC*	2077-9	CTC	TCT*	5160	G	—*
335-9	TGTCG	T—T—G*	2544-5	CC	CCC*	5253	C	G*
416-7	CG	GC*	3212-3	TT	TTT*	5260-1	TG	TG/AGGAAG*
603-4	CG	GC*	3418-9	CC	CCC*	5293-4	GC	GGCC*
609-10	GC	CG*	3565	G	GG*	5552	G	GG*
655	G	GG*	4534	C	CC*	5685	G	GG*
688	C	CC*	4552	C	CC*	6081	C	CCC*
833	C	G*	4626	C	CCC*	6183-4	GG	GGG*
1217-9	GGG	GGGG*	4735	T	CT*			

Table 9 Consistent discrepancies between GenBank sequences and our detected sequences.

Δ All numbers indicating positions are according to GenBank sequence for IRF-1 (L05072); # indicates that the shown nucleotides are consistent with reference sequence for IRF-1 promoter region (X53095); * indicates that these discrepancies could be detected in all tested samples.

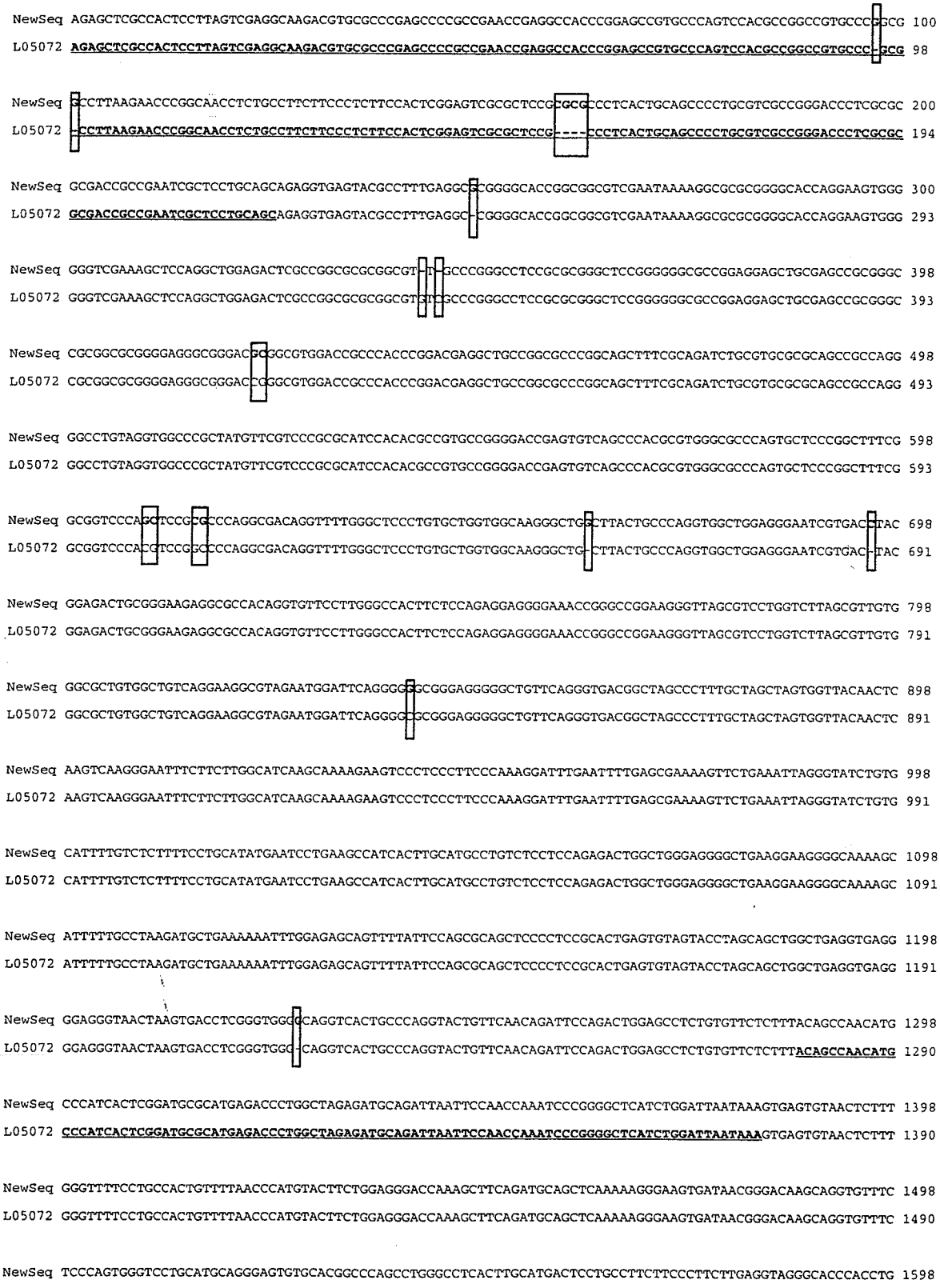
discrepancies were consistent among all ethnic groups tested and the current IRF-1 reference sequences might not be accurate and representative for all populations. We also noted differences between the two GenBank sequences (X53095 and L05072) in the overlapping region (nt 496 to 669 of X53095 and nt 1 to 173 of L05072), further suggesting that these discrepancies be due to the inaccuracy of the available IRF-1 sequence data.

Complete IRF-1 and its promoter sequences were re-established

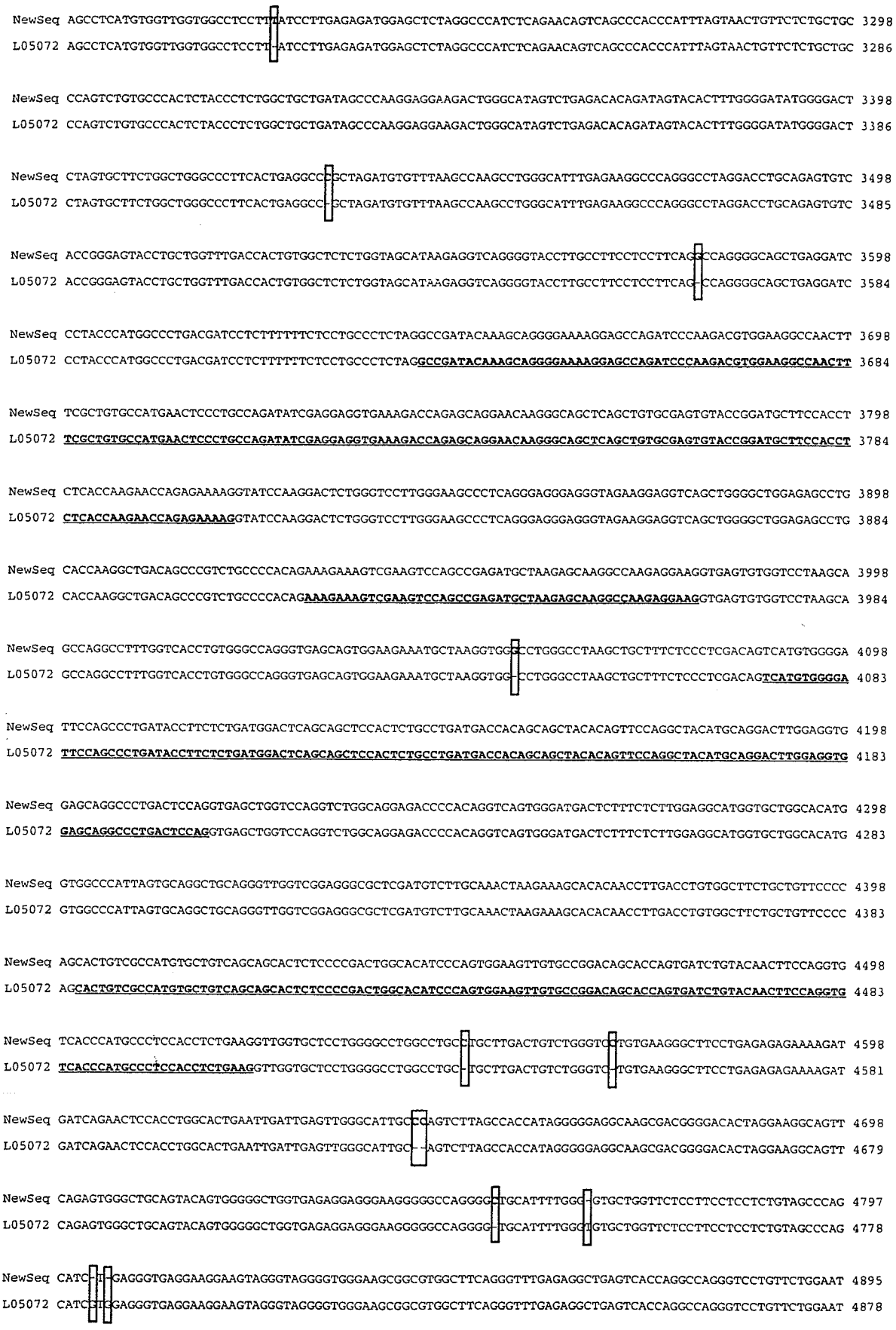
The accuracy of a reference sequence for any human gene is critical for further genetic and functional studies. The current GenBank sequence for human IRF-1 (accession number: L05072) was derived from a human placenta DNA library and reported in 1992 (Cha et al., 1992). According to our population-based sequencing data, this reference sequence is not representative for all populations/subjects and an update to these sequences was required. We conducted complete IRF-1 gene sequencing on one selected individual who was homozygous for alleles that coincide with the current GenBank sequences at all polymorphic loci (Ji et al., 2004). This selection criterion aimed to avoid the polymorphic sites which might bias our new full length IRF-1 sequence determination.

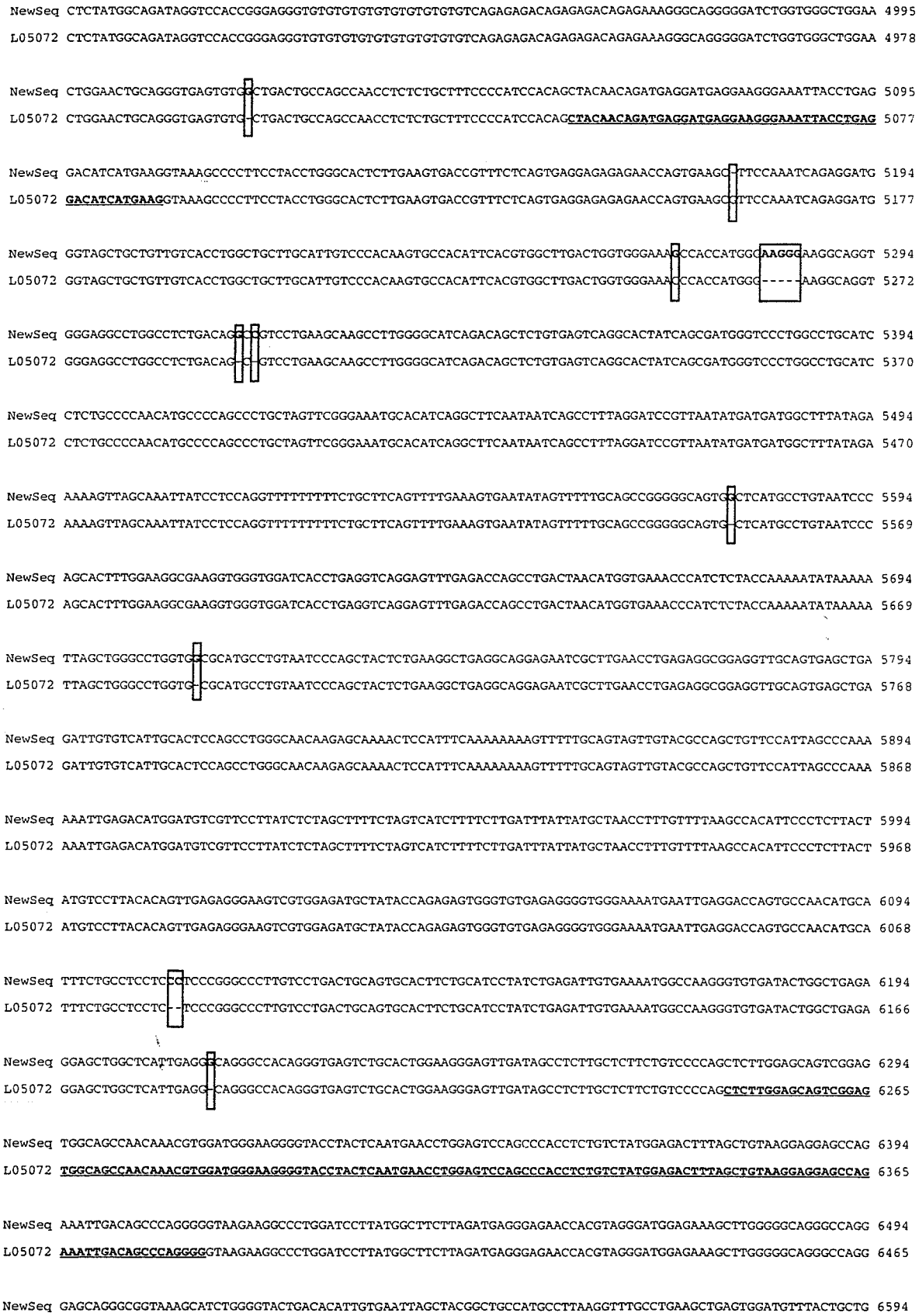
The identified full length genomic IRF-1 nucleotide sequence is shown in **Figure 10**. In comparison to submission L05072, our sequence contains 35 additional nucleotides while 8 nucleotides were absent leading to the size of the identified IRF-1 gene being 7,748 nucleotides in length instead of 7721 as described in L05072. There were twelve additional nucleotide sequence changes as well. Among these differences, 7 nucleotide

Figure 10 Complete IRF-1 gene sequence in alignment with current GenBank reference sequence



```
L05072 TCCCAGTGGGTCCTGCATGCAGGGAGTGTGCACGGCCCAGCCTGGGCCTCACTTGCATGACTCCTGCCTTCTTCCCTTCTTGAGGTAGGGCACCCACCTG 1590

NewSeq AAGGCACTTCCAGTTTCCAGCAGCAAGACTTTCCAGCATCTGCAGAGCTGGAGTTCTGCTCTCCTCTAAGCGAGACCCTTACAAACATACACAGCACTCT 1698
L05072 AAGGCACTTCCAGTTTCCAGCAGCAAGACTTTCCAGCATCTGCAGAGCTGGAGTTCTGCTCTCCTCTAAGCGAGACCCTTACAAACATACACAGCACTCT 1690

NewSeq GCAGGGCTCCAATCGAACAAATAGAAGACTGAGAAGTGGATGCTGCTGGGCAGAAACGTGCCTGGCTTAGCAGAGGACAAACGAGTTAATCTTGCACCAG 1798
L05072 GCAGGGCTCCAATCGAACAAATAGAAGACTGAGAAGTGGATGCTGCTGGGCAGAAACGTGC-TGGCTTAGCAGAGGACAAACGAGTTAATCTTGCACCAG 1789

NewSeq TCACTCTGGCCCAAGAAGCCTATAGCTGGTGCACTTGGGGCAACATAGACCCTATAGACTTAGTAGCAATGATAGTATTCATAATAATAGCTAATGCTTA 1898
L05072 TCACTCTGGCCCAAGAAGCCTATAGCTGGTGCACTTGGGGCAACATAGACCCTATAGACTTAGTAGCAATGATAGTAT-CATAATAATAGCTAATGCTTA 1888

NewSeq CTGAACACTCCCTGTGTGCCTGGCACCTGCTAAGTATGTTATTTACATTGTGTCATTTAATCCTCGCAGTAGTCCTGTGGGTTAGATCTTACTAATGTCA 1998
L05072 CTGAACACTCCCTGTGTGCCTGGCACCTGCTAAGTATGTTATTTACATTGTGTCATTTAATCCTCGCAGTAGTCCTGTGGGTTAGATCTTACTAATGTCA 1988

NewSeq TCATTTTCAGATAAGTAAACAGAGGCACTGAGAGGTAGATCATAAGATCACACAAAAAGTGATGAAGCCAAGATTTGAACTTGAACGGTCTGACTCAGAA 2098
L05072 TCATTTTCAGATAAGTAAACAGAGGCACTGAGAGGTAGATCATAAGATCACACAA-AAGTGATGAAGCCAAGATTTGAACTTGAACGGCTCGACTCAGAA 2088

NewSeq ATCTTTACTGTTAACCATAAGTGATATAATAACAGTAAGACCTTAGACTTCATATTTGTCACTGTGTCCCTACACATCCTCTGGTTTTTAATCCTCAAAA 2198
L05072 ATCTTTACTGTTAACCATAAGTGATATAATAACAGTAAGACCTTAGACTTCATATTTGTCACTGTGTCCCTACACATCCTCTGGTTTTTAATCCTCAAAA 2188

NewSeq TTTTGTTGGATATGTTTTCTCATTTCCGAGAAGAGAAAACTGAGGGGCAAAGAGATACAGTGACAATGCCAGGGTTACACAGTGTTCACCATCCAAGTCT 2298
L05072 TTTTGTTGGATATGTTTTCTCATTTCCGAGAAGAGAAAACTGAGGGGCAAAGAGATACAGTGACAATGCCAGGGTTACACAGTGTTCACCATCCAAGTCT 2288

NewSeq AGCCCAGAGCTCCCTCAGTGGTATGACCAGGACCCCCTGTGTAAGAGCCCATGCTCCCAGGTGTCCTGAGGAGTCCTTTCTAATGGAAGAAGTTCTTACT 2398
L05072 AGCCCAGAGCTCCCTCAGTGGTATGACCAGGACCCCCTGTGTAAGAGCCCATGCTCCCAGGTGTCCTGAGGAGTCCTTTCTAATGGAAGAAGTTCTTACT 2388

NewSeq TCCATGTGGGTGCTTACAAGCCAGAGAGAAACATCCCAGAGCTTCAAAACCAGGGCTTTGGGGGAGGGTGCCCTGTGTGGGTCCTAGCACATGTGTAACA 2498
L05072 TCCATGTGGGTGCTTACAAGCCAGAGAGAAACATCCCAGAGCTTCAAAACCAGGGCTTTGGGGGAGGGTGCCCTGTGTGGGTCCTAGCACATGTGTAACA 2488

NewSeq GGCAGAGGGAGGTCTTTGTGAGCTAATAATGCTGCAGCTCATCCAAACTAGGTGTCCCTCCTGAGAGATCCAGAGTGGTCTGTTTAAGCCAGCCTCAAGA 2598
L05072 GGCAGAGGGAGGTCTTTGTGAGCTAATAATGCTGCAGCTCATCCAAACTAGGTGTCC-TCCTGAGAGATCCAGAGTGGTCTGTTTAAGCCAGCCTCAAGA 2587

NewSeq TGGGTGTCCAAGCCAGATGTCAGGGGAAAAAAGGGGAAGTCAGCCTTTTCTCAGACCTGTCTGGCTGGGCAGGCCTGGGTCTCAGACTCAGCCCCAAAGT 2698
L05072 TGGGTGTCCAAGCCAGATGTCAGGGGAAAAAAGGGGAAGTCAGCCTTTTCTCAGACCTGTCTGGCTGGGCAGGCCTGGGTCTCAGACTCAGCCCCAAAGT 2687

NewSeq CTGTGGTCTCTGACCTGACACAGCCTTATGTGTATGTGTGTATTGTTCAGGAGGAGATGATCTTCCAGATCCCATGGAAGCATGCTGCCAAGCATGGCTG 2798
L05072 CTGTGGTCTCTGACCTGACACAGCCTTATGTGTATGTGTGTATTGTTCAGGAGGAGATGATCTTCCAGATCCCATGGAAGCATGCTGCCAAGCATGGCTG 2787

NewSeq GGACATCAACAAGGATGCCTGTTTGTTCCGGAGCTGGGCCATTCACACAGGTGTGTGCCTGGGACTCAGGCCTAGGAAGCCCAGGGTAGAGACAAGAGGA 2898
L05072 GGACATCAACAAGGATGCCTGTTTGTTCCGGAGCTGGGCCATTCACACAGGTGTGTGCCTGGGACTCAGGCCTAGGAAGCCCAGGGTAGAGACAAGAGGA 2887

NewSeq GGCACTCACGTTAACACAGAGGCTCTTCACTGGGGTCCCTGAGCTCCCTGAGACAACATGCAGAATTACTGGGAAGAGGGGCTGGTGGCAGACTTGTGTT 2998
L05072 GGCACTCACGTTAACACAGAGGCTCTTCACTGGGGTCCCTGAGCTCCCTGAGACAACATGCAGAATTACTGGGAAGAGGGGCTGGTGGCAGACTTGTGTT 2987

NewSeq TCTGGAGAAGAGAGTCGATCATCTCAGCAAATTCTCAAAGGGAAAAGCCAAGATCTTAGAAAGTGTGTGGCTTCAGGGGGTTTGTGGCTAGATGAAAGTT 3098
L05072 TCTGGAGAAGAGAGTCGATCATCTCAGCAAATTCTCAAAGGGAAAAGCCAAGATCTTAGAAAGTGTGTGGCTTCAGGGGGTTTGTGGCTAGATGAAAGTT 3087

NewSeq CTCCCTGGCAAAAGCATCTGTGAAAAGCAGCTGTAAGCCAGGGCACTGAAAGAGACCCAGGTCTGCCTTTTTCTTCGTGTTGACCAAGGCCCTTGGTCCA 3198
L05072 CTCCCTGGCAAAAGCATCTGTGAAAAGCAGCTGTAAGCCAGGGCACTGAAAGAGACCCAGGTCTGCCTTTTTCTTCGTGTTGACCAAGGCCCTTGGTCCA 3187
```


NewSeq AGCCTCATGTGGTTGGTGGCCTCCTTCATCCTTGAGAGATGGAGCTCTAGGCCCATCTCAGAACAGTCAGCCCACCCATTTAGTAACTGTTCTCTGCTGC 3298
L05072 AGCCTCATGTGGTTGGTGGCCTCCTT-ATCCTTGAGAGATGGAGCTCTAGGCCCATCTCAGAACAGTCAGCCCACCCATTTAGTAACTGTTCTCTGCTGC 3286
NewSeq CCAGTCTGTGCCCACTCTACCCTCTGGCTGCTGATAGCCCAAGGAGGAAGACTGGGCATAGTCTGAGACACAGATAGTACACTTTGGGGATATGGGGACT 3398
L05072 CCAGTCTGTGCCCACTCTACCCTCTGGCTGCTGATAGCCCAAGGAGGAAGACTGGGCATAGTCTGAGACACAGATAGTACACTTTGGGGATATGGGGACT 3386
NewSeq CTAGTGCTTCTGGCTGGGCCCTTCACTGAGGCCCGCTAGATGTGTTTAAGCCAAGCCTGGGCATTTGAGAAGGCCCAGGGCCTAGGACCTGCAGAGTGTC 3498
L05072 CTAGTGCTTCTGGCTGGGCCCTTCACTGAGGCC-GCTAGATGTGTTTAAGCCAAGCCTGGGCATTTGAGAAGGCCCAGGGCCTAGGACCTGCAGAGTGTC 3485
NewSeq ACCGGGAGTACCTGCTGGTTTGACCACTGTGGCTCTCTGGTAGCATAAGAGGTCAGGGGTACCTTGCCTTCCTCCTTCAGCCCAGGGGCAGCTGAGGATC 3598
L05072 ACCGGGAGTACCTGCTGGTTTGACCACTGTGGCTCTCTGGTAGCATAAGAGGTCAGGGGTACCTTGCCTTCCTCCTTCAG-CCAGGGGCAGCTGAGGATC 3584
NewSeq CCTACCCATGGCCCTGACGATCCTCTTTTTTCTCCTGCCCTCTAGGCCGATACAAAGCAGGGGAAAAGGAGCCAGATCCCAAGACGTGGAAGGCCAACTT 3698
L05072 CCTACCCATGGCCCTGACGATCCTCTTTTTTCTCCTGCCCTCTAGGCCGATACAAAGCAGGGGAAAAGGAGCCAGATCCCAAGACGTGGAAGGCCAACTT 3684
NewSeq TCGCTGTGCCATGAACTCCCTGCCAGATATCGAGGAGGTGAAAGACCAGAGCAGGAACAAGGGCAGCTCAGCTGTGCGAGTGTACCGGATGCTTCCACCT 3798
L05072 TCGCTGTGCCATGAACTCCCTGCCAGATATCGAGGAGGTGAAAGACCAGAGCAGGAACAAGGGCAGCTCAGCTGTGCGAGTGTACCGGATGCTTCCACCT 3784
NewSeq CTCACCAAGAACCAGAGAAAAGGTATCCAAGGACTCTGGGTCCTTGGGAAGCCCTCAGGGAGGGAGGGTAGAAGGAGGTCAGCTGGGGCTGGAGAGCCTG 3898
L05072 CTCACCAAGAACCAGAGAAAAGGTATCCAAGGACTCTGGGTCCTTGGGAAGCCCTCAGGGAGGGAGGGTAGAAGGAGGTCAGCTGGGGCTGGAGAGCCTG 3884
NewSeq CACCAAGGCTGACAGCCCGTCTGCCCCACAGAAAGAAAGTCGAAGTCCAGCCGAGATGCTAAGAGCAAGGCCAAGAGGAAGGTGAGTGTGGTCCTAAGCA 3998
L05072 CACCAAGGCTGACAGCCCGTCTGCCCCACAGAAAGAAAGTCGAAGTCCAGCCGAGATGCTAAGAGCAAGGCCAAGAGGAAGGTGAGTGTGGTCCTAAGCA 3984
NewSeq GCCAGGCCTTTGGTCACCTGTGGGCCAGGGTGAGCAGTGGAAGAAATGCTAAGGTGGGCCTGGGCCTAAGCTGCTTTCTCCCTCGACAGTCATGTGGGGA 4098
L05072 GCCAGGCCTTTGGTCACCTGTGGGCCAGGGTGAGCAGTGGAAGAAATGCTAAGGTGG-CCTGGGCCTAAGCTGCTTTCTCCCTCGACAGTCATGTGGGGA 4083
NewSeq TTCCAGCCCTGATACCTTCTCTGATGGACTCAGCAGCTCCACTCTGCCTGATGACCACAGCAGCTACACAGTTCCAGGCTACATGCAGGACTTGGAGGTG 4198
L05072 TTCCAGCCCTGATACCTTCTCTGATGGACTCAGCAGCTCCACTCTGCCTGATGACCACAGCAGCTACACAGTTCCAGGCTACATGCAGGACTTGGAGGTG 4183
NewSeq GAGCAGGCCCTGACTCCAGGTGAGCTGGTCCAGGTCTGGCAGGAGACCCCACAGGTCAGTGGGATGACTCTTTCTCTTGGAGGCATGGTGCTGGCACATG 4298
L05072 GAGCAGGCCCTGACTCCAGGTGAGCTGGTCCAGGTCTGGCAGGAGACCCCACAGGTCAGTGGGATGACTCTTTCTCTTGGAGGCATGGTGCTGGCACATG 4283
NewSeq GTGGCCCATTAGTGCAGGCTGCAGGGTTGGTCGGAGGGCGCTCGATGTCTTGCAAACTAAGAAAGCACACAACCTTGACCTGTGGCTTCTGCTGTTCCCC 4398
L05072 GTGGCCCATTAGTGCAGGCTGCAGGGTTGGTCGGAGGGCGCTCGATGTCTTGCAAACTAAGAAAGCACACAACCTTGACCTGTGGCTTCTGCTGTTCCCC 4383
NewSeq AGCACTGTCGCCATGTGCTGTCAGCAGCACTCTCCCCGACTGGCACATCCCAGTGGAAGTTGTGCCGGACAGCACCAGTGATCTGTACAACTTCCAGGTG 4498
L05072 AGCACTGTCGCCATGTGCTGTCAGCAGCACTCTCCCCGACTGGCACATCCCAGTGGAAGTTGTGCCGGACAGCACCAGTGATCTGTACAACTTCCAGGTG 4483
NewSeq TCACCCATGCCCTCCACCTCTGAAGGTTGGTGCTCCTGGGGCCTGGCCTGCCTGCTTGACTGTCTGGGTCGTGTGAAGGGCTTCCTGAGAGAGAAAAGAT 4598
L05072 TCACCCATGCCCTCCACCTCTGAAGGTTGGTGCTCCTGGGGCCTGGCCTGC-TGCTTGACTGTCTGGGTC-TGTGAAGGGCTTCCTGAGAGAGAAAAGAT 4581
NewSeq GATCAGAACTCCACCTGGCACTGAATTGATTGAGTTGGGCATTGCCCAGTCTTAGCCACCATAGGGGGAGGCAAGCGACGGGGACACTAGGAAGGCAGTT 4698
L05072 GATCAGAACTCCACCTGGCACTGAATTGATTGAGTTGGGCATTGC--AGTCTTAGCCACCATAGGGGGAGGCAAGCGACGGGGACACTAGGAAGGCAGTT 4679
NewSeq CAGAGTGGGCTGCAGTACAGTGGGGGCTGGTGAGAGGAGGGAAGGGGGCCAGGGGCTGCATTTTGGG-GTGCTGGTTCTCCTTCCTCCTCTGTAGCCCAG 4797
L05072 CAGAGTGGGCTGCAGTACAGTGGGGGCTGGTGAGAGGAGGGAAGGGGGCCAGGGG-TGCATTTTGGG-GTGCTGGTTCTCCTTCCTCCTCTGTAGCCCAG 4778
NewSeq CATC-T-GAGGGTGAGGAAGGAAGTAGGGTAGGGGTGGGAAGCGGCGTGGCTTCAGGGTTTGAGAGGCTGAGTCACCAGGCCAGGGTCCTGTTCTGGAAT 4895
L05072 CATCGTGAGGGTGAGGAAGGAAGTAGGGTAGGGGTGGGAAGCGGCGTGGCTTCAGGGTTTGAGAGGCTGAGTCACCAGGCCAGGGTCCTGTTCTGGAAT 4878


NewSeq CTCTATGGCAGATAGGTCCACCGGGAGGGTGTGTGTGTGTGTGTGTGTGTCAGAGAGACAGAGAGACAGAGAAAGGGCAGGGGGATCTGGTGGGCTGGAA 4995
L05072 CTCTATGGCAGATAGGTCCACCGGGAGGGTGTGTGTGTGTGTGTGTGTGTCAGAGAGACAGAGAGACAGAGAAAGGGCAGGGGGATCTGGTGGGCTGGAA 4978
NewSeq CTGGAACTGCAGGGTGAGTGTGGCTGACTGCCAGCCAACCTCTCTGCTTTCCCCATCCACAGCTACAACAGATGAGGATGAGGAAGGGAAATTACCTGAG 5095
L05072 CTGGAACTGCAGGGTGAGTGTG-CTGACTGCCAGCCAACCTCTCTGCTTTCCCCATCCACAGCTACAACAGATGAGGATGAGGAAGGGAAATTACCTGAG 5077
NewSeq GACATCATGAAGGTAAAGCCCCTTCCTACCTGGGCACTCTTGAAGTGACCGTTTCTCAGTGAGGAGAGAGAACCAGTGAAGCGTTCCAAATCAGAGGATG 5194
L05072 GACATCATGAAGGTAAAGCCCCTTCCTACCTGGGCACTCTTGAAGTGACCGTTTCTCAGTGAGGAGAGAGAACCAGTGAAGCGTTCCAAATCAGAGGATG 5177
NewSeq GGTAGCTGCTGTTGTCACCTGGCTGCTTGCATTGTCCCACAAGTGCCACATTCACGTGGCTTGACTGGTGGGAAAGCCACCATGGCAAGGGAAGGCAGGT 5294
L05072 GGTAGCTGCTGTTGTCACCTGGCTGCTTGCATTGTCCCACAAGTGCCACATTCACGTGGCTTGACTGGTGGGAAAGCCACCATGGC-----AAGGCAGGT 5272
NewSeq GGGAGGCCTGGCCTCTGACAGGCCGTCCTGAAGCAAGCCTTGGGGCATCAGACAGCTCTGTGAGTCAGGCACTATCAGCGATGGGTCCCTGGCCTGCATC 5394
L05072 GGGAGGCCTGGCCTCTGACAG-C-GTCCTGAAGCAAGCCTTGGGGCATCAGACAGCTCTGTGAGTCAGGCACTATCAGCGATGGGTCCCTGGCCTGCATC 5370
NewSeq CTCTGCCCCAACATGCCCCAGCCCTGCTAGTTCGGGAAATGCACATCAGGCTTCAATAATCAGCCTTTAGGATCCGTTAATATGATGATGGCTTTATAGA 5494
L05072 CTCTGCCCCAACATGCCCCAGCCCTGCTAGTTCGGGAAATGCACATCAGGCTTCAATAATCAGCCTTTAGGATCCGTTAATATGATGATGGCTTTATAGA 5470
NewSeq AAAAGTTAGCAAATTATCCTCCAGGTTTTTTTTTCTGCTTCAGTTTTGAAAGTGAATATAGTTTTTGCAGCCGGGGGCAGTGGCTCATGCCTGTAATCCC 5594
L05072 AAAAGTTAGCAAATTATCCTCCAGGTTTTTTTTTCTGCTTCAGTTTTGAAAGTGAATATAGTTTTTGCAGCCGGGGGCAGTG-CTCATGCCTGTAATCCC 5569
NewSeq AGCACTTTGGAAGGCGAAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGACTAACATGGTGAAACCCATCTCTACCAAAAATATAAAAA 5694
L05072 AGCACTTTGGAAGGCGAAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGACTAACATGGTGAAACCCATCTCTACCAAAAATATAAAAA 5669
NewSeq TTAGCTGGGCCTGGTGGCGCATGCCTGTAATCCCAGCTACTCTGAAGGCTGAGGCAGGAGAATCGCTTGAACCTGAGAGGCGGAGGTTGCAGTGAGCTGA 5794
L05072 TTAGCTGGGCCTGGTG-CGCATGCCTGTAATCCCAGCTACTCTGAAGGCTGAGGCAGGAGAATCGCTTGAACCTGAGAGGCGGAGGTTGCAGTGAGCTGA 5768
NewSeq GATTGTGTCATTGCACTCCAGCCTGGGCAACAAGAGCAAAACTCCATTTCAAAAAAAAGTTTTTGCAGTAGTTGTACGCCAGCTGTTCCATTAGCCCAAA 5894
L05072 GATTGTGTCATTGCACTCCAGCCTGGGCAACAAGAGCAAAACTCCATTTCAAAAAAAAGTTTTTGCAGTAGTTGTACGCCAGCTGTTCCATTAGCCCAAA 5868
NewSeq AAATTGAGACATGGATGTCGTTCCTTATCTCTAGCTTTTCTAGTCATCTTTTCTTGATTTATTATGCTAACCTTTGTTTTAAGCCACATTCCCTCTTACT 5994
L05072 AAATTGAGACATGGATGTCGTTCCTTATCTCTAGCTTTTCTAGTCATCTTTTCTTGATTTATTATGCTAACCTTTGTTTTAAGCCACATTCCCTCTTACT 5968
NewSeq ATGTCCTTACACAGTTGAGAGGGAAGTCGTGGAGATGCTATACCAGAGAGTGGGTGTGAGAGGGGTGGGAAAATGAATTGAGGACCAGTGCCAACATGCA 6094
L05072 ATGTCCTTACACAGTTGAGAGGGAAGTCGTGGAGATGCTATACCAGAGAGTGGGTGTGAGAGGGGTGGGAAAATGAATTGAGGACCAGTGCCAACATGCA 6068
NewSeq TTTCTGCCTCCTCCGTCCCGGGCCCTTGTCCTGACTGCAGTGCACTTCTGCATCCTATCTGAGATTGTGAAAATGGCCAAGGGTGTGATACTGGCTGAGA 6194
L05072 TTTCTGCCTCCTC--TCCCGGGCCCTTGTCCTGACTGCAGTGCACTTCTGCATCCTATCTGAGATTGTGAAAATGGCCAAGGGTGTGATACTGGCTGAGA 6166
NewSeq GGAGCTGGCTCATTGAGGGCAGGGCCACAGGGTGAGTCTGCACTGGAAGGGAGTTGATAGCCTCTTGCTCTTCTGTCCCCAGCTCTTGGAGCAGTCGGAG 6294
L05072 GGAGCTGGCTCATTGAGG-CAGGGCCACAGGGTGAGTCTGCACTGGAAGGGAGTTGATAGCCTCTTGCTCTTCTGTCCCCAGCTCTTGGAGCAGTCGGAG 6265
NewSeq TGGCAGCCAACAAACGTGGATGGGAAGGGGTACCTACTCAATGAACCTGGAGTCCAGCCCACCTCTGTCTATGGAGACTTTAGCTGTAAGGAGGAGCCAG 6394
L05072 TGGCAGCCAACAAACGTGGATGGGAAGGGGTACCTACTCAATGAACCTGGAGTCCAGCCCACCTCTGTCTATGGAGACTTTAGCTGTAAGGAGGAGCCAG 6365
NewSeq AAATTGACAGCCCAGGGGGTAAGAAGGCCCTGGATCCTTATGGCTTCTTAGATGAGGGAGAACCACGTAGGGATGGAGAAAGCTTGGGGGCAGGGCCAGG 6494
L05072 AAATTGACAGCCCAGGGGGTAAGAAGGCCCTGGATCCTTATGGCTTCTTAGATGAGGGAGAACCACGTAGGGATGGAGAAAGCTTGGGGGCAGGGCCAGG 6465
NewSeq GAGCAGGGCGGTAAAGCATCTGGGGTACTGACACATTGTGAATTAGCTACGGCTGCCATGCCTTAAGGTTTGCCTGAAGCTGAGTGGATGTTTACTGCTG 6594

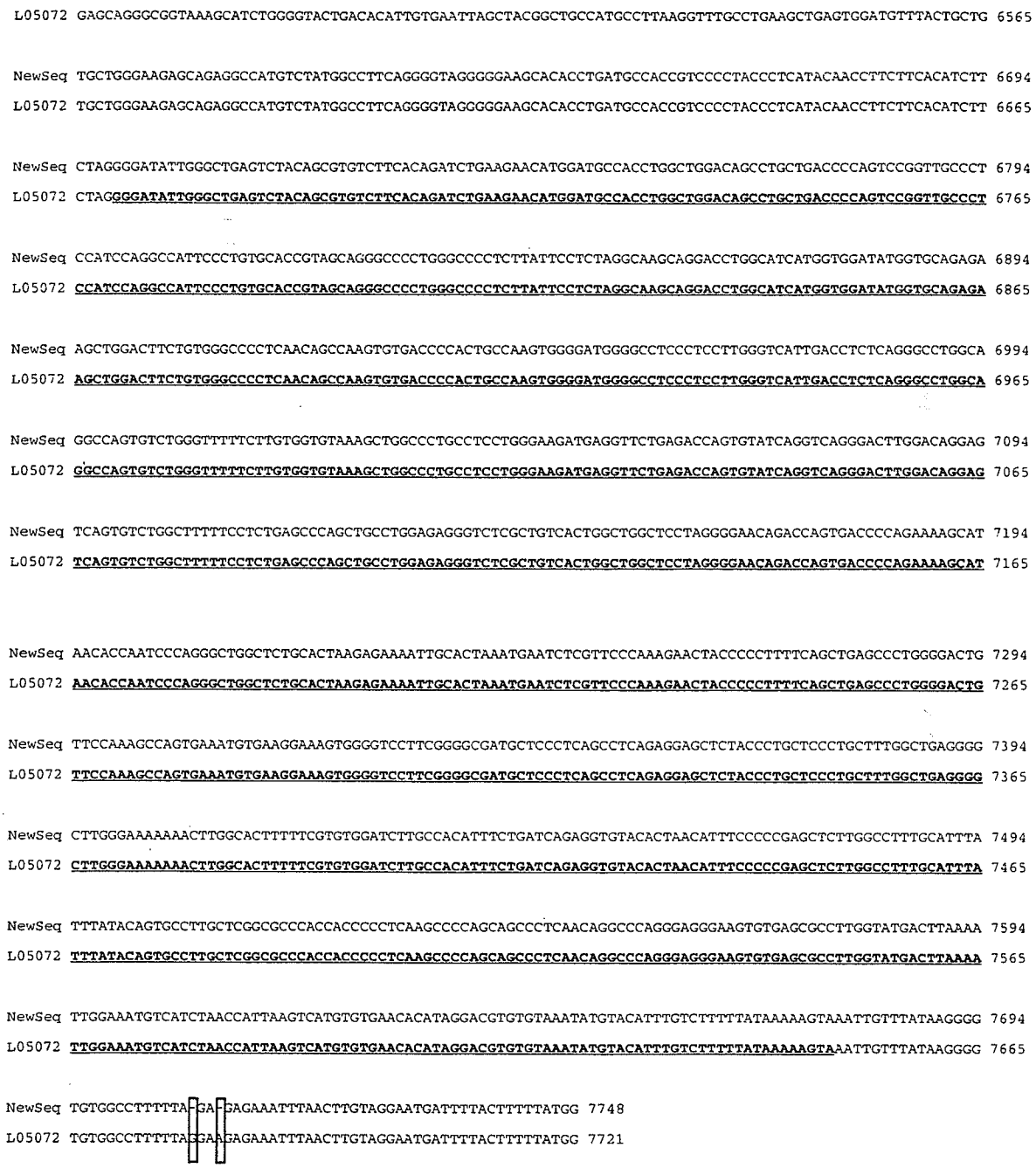

L05072 GAGCAGGGCGGTAAAGCATCTGGGGTACTGACACATTGTGAATTAGCTACGGCTGCCATGCCTTAAGGTTTGCCTGAAGCTGAGTGGATGTTTACTGCTG 6565
NewSeq TGCTGGGAAGAGCAGAGGCCATGTCTATGGCCTTCAGGGGTAGGGGGAAGCACACCTGATGCCACCGTCCCCTACCCTCATACAACCTTCTTCACATCTT 6694
L05072 TGCTGGGAAGAGCAGAGGCCATGTCTATGGCCTTCAGGGGTAGGGGGAAGCACACCTGATGCCACCGTCCCCTACCCTCATACAACCTTCTTCACATCTT 6665
NewSeq CTAGGGGATATTGGGCTGAGTCTACAGCGTGTCTTCACAGATCTGAAGAACATGGATGCCACCTGGCTGGACAGCCTGCTGACCCCAGTCCGGTTGCCCT 6794
L05072 CTAGGGGATATTGGGCTGAGTCTACAGCGTGTCTTCACAGATCTGAAGAACATGGATGCCACCTGGCTGGACAGCCTGCTGACCCCAGTCCGGTTGCCCT 6765
NewSeq CCATCCAGGCCATTCCCTGTGCACCGTAGCAGGGCCCCTGGGCCCCTCTTATTCCTCTAGGCAAGCAGGACCTGGCATCATGGTGGATATGGTGCAGAGA 6894
L05072 CCATCCAGGCCATTCCCTGTGCACCGTAGCAGGGCCCCTGGGCCCCTCTTATTCCTCTAGGCAAGCAGGACCTGGCATCATGGTGGATATGGTGCAGAGA 6865
NewSeq AGCTGGACTTCTGTGGGCCCCTCAACAGCCAAGTGTGACCCCACTGCCAAGTGGGGATGGGGCCTCCCTCCTTGGGTCATTGACCTCTCAGGGCCTGGCA 6994
L05072 AGCTGGACTTCTGTGGGCCCCTCAACAGCCAAGTGTGACCCCACTGCCAAGTGGGGATGGGGCCTCCCTCCTTGGGTCATTGACCTCTCAGGGCCTGGCA 6965
NewSeq GGCCAGTGTCTGGGTTTTTCTTGTGGTGTAAAGCTGGCCCTGCCTCCTGGGAAGATGAGGTTCTGAGACCAGTGTATCAGGTCAGGGACTTGGACAGGAG 7094
L05072 GGCCAGTGTCTGGGTTTTTCTTGTGGTGTAAAGCTGGCCCTGCCTCCTGGGAAGATGAGGTTCTGAGACCAGTGTATCAGGTCAGGGACTTGGACAGGAG 7065
NewSeq TCAGTGTCTGGCTTTTTCCTCTGAGCCCAGCTGCCTGGAGAGGGTCTCGCTGTCACTGGCTGGCTCCTAGGGGAACAGACCAGTGACCCCAGAAAAGCAT 7194
L05072 TCAGTGTCTGGCTTTTTCCTCTGAGCCCAGCTGCCTGGAGAGGGTCTCGCTGTCACTGGCTGGCTCCTAGGGGAACAGACCAGTGACCCCAGAAAAGCAT 7165
NewSeq AACACCAATCCCAGGGCTGGCTCTGCACTAAGAGAAAATTGCACTAAATGAATCTCGTTCCCAAAGAACTACCCCCTTTTCAGCTGAGCCCTGGGGACTG 7294
L05072 AACACCAATCCCAGGGCTGGCTCTGCACTAAGAGAAAATTGCACTAAATGAATCTCGTTCCCAAAGAACTACCCCCTTTTCAGCTGAGCCCTGGGGACTG 7265
NewSeq TTCCAAAGCCAGTGAAATGTGAAGGAAAGTGGGGTCCTTCGGGGCGATGCTCCCTCAGCCTCAGAGGAGCTCTACCCTGCTCCCTGCTTTGGCTGAGGGG 7394
L05072 TTCCAAAGCCAGTGAAATGTGAAGGAAAGTGGGGTCCTTCGGGGCGATGCTCCCTCAGCCTCAGAGGAGCTCTACCCTGCTCCCTGCTTTGGCTGAGGGG 7365
NewSeq CTTGGGAAAAAAACTTGGCACTTTTTCGTGTGGATCTTGCCACATTTCTGATCAGAGGTGTACACTAACATTTCCCCCGAGCTCTTGGCCTTTGCATTTA 7494
L05072 CTTGGGAAAAAAACTTGGCACTTTTTCGTGTGGATCTTGCCACATTTCTGATCAGAGGTGTACACTAACATTTCCCCCGAGCTCTTGGCCTTTGCATTTA 7465
NewSeq TTTATACAGTGCCTTGCTCGGCGCCCACCACCCCCTCAAGCCCCAGCAGCCCTCAACAGGCCCAGGGAGGGAAGTGTGAGCGCCTTGGTATGACTTAAAA 7594
L05072 TTTATACAGTGCCTTGCTCGGCGCCCACCACCCCCTCAAGCCCCAGCAGCCCTCAACAGGCCCAGGGAGGGAAGTGTGAGCGCCTTGGTATGACTTAAAA 7565
NewSeq TTGGAAATGTCATCTAACCATTAAGTCATGTGTGAACACATAGGACGTGTGTAAATATGTACATTTGTCTTTTTATAAAAAGTAAATTGTTTATAAGGGG 7694
L05072 TTGGAAATGTCATCTAACCATTAAGTCATGTGTGAACACATAGGACGTGTGTAAATATGTACATTTGTCTTTTTATAAAAAGTAAATTGTTTATAAGGGG 7665
NewSeq TGTGGCCTTTTTA-GA-GAGAAATTTAACTTGTAGGAATGATTTTACTTTTTATGG 7748
L05072 TGTGGCCTTTTTAGGAAGAGAAATTTAACTTGTAGGAATGATTTTACTTTTTATGG 7721

Figure 10 Complete IRF-1 gene sequence in alignment with current GenBank reference sequence.

The alignment of obtained population-based complete IRF-1 sequence with current GenBank reference sequence (L05072) is shown in this figure. Compared to submission L05072, our population-based IRF-1 sequence contains: 35 nucleotides addition, 8 nucleotides removal and another 12 nucleotides replacements. These lead to a re-established IRF-1 gene sequence with 7748 nucleotides instead of 7721 nucleotides (in L05072) in length. None of these discrepancies was found to be located on the predicted potential alternative splicing sites after analysis with online splicing prediction software (NetGene2--http://www.cbs.dtu.dk/services/NetGene2/). "**NewSeq**" stands for our reporting IRF-1 genome sequence and L05072 refers to the current GenBank reference sequence. The framed nucleotides indicate the differences between the two compared sequences. The underlined regions indicate the proposed exon regions in L05072.

additions occurred in exon 1 (the untranslated terminal region in IRF-1 mRNA) while the other differences were located in different intronic regions. None of these discrepancies were located in predicted potential alternative splicing sites after analysis with online splicing prediction software (NetGene2--http://www.cbs.dtu.dk/services/NetGene2/). Another single nucleotide discrepancy in the IRF-1 promoter was also identified between our sequence data and submission X53095 at nucleotide 608 (A→C) (**Figure 11**). It is noteworthy that this nucleotide difference in IRF-1 promoter region is located in the overlapping region of L05072 and X53095 and it is identical to the sequence in L05072. The combined newly established genomic sequence of IRF-1 and upstream region has been submitted to GenBank (accession number is: DQ789232). This sequence is more representative of general population including African, European, Asian and Caucasian.

Section Two: Association of IRF-1 polymorphisms with HIV resistance

IRF-1 179 allele associated with HIV-1 resistance and decreased likelihood of seroconversion

In previous studies, a specific microsatellite allele in IRF-1 gene significantly correlated with resistance to HIV-1 infection in our study cohort (Ball, 2001). Using the most current epidemiologic data, we re-analyzed the effect of this polymorphism on altered susceptibility to HIV-1 infection. It was confirmed that the 179 allele (with 12 "GT" repeats in intron 7) in the IRF-1 MS marker was strongly associated with the HIV-1 resistance phenotype. The allele frequencies for 179 allele was 47.1% (81/172) and 35.8% (426 /1190) in HIV-1-resistants and HIV-1-susceptibles. The difference is

Figure 11 Complete IRF-1 upstream promoter sequence

```
  1 aagcttgagg agccaggctg ccagtcggga gattcggccc agtgttccca ctggagaggg cggcaagtgc ccgggcgatc
 81 acctcgcctg cgttcgggag atatacctcc gcccccgccc cgccaggagg gtgaaaagat ggccccagga gccagccggc
161 tgggacaagg cggagtgaga ggacaggctg gggccggggg cgctgggctg tcccgggcag ccctcctccg ggcaagccgg
241 agcaggggtg gattgggagc gctcggggcg ggcccgcggt ggccccgggg cggtggcgcc cggccggaga gggtggggcg
321 gagcagccgc cctgtacttc cccttcgccg ctagctctac aacagcctga tttccccgaa atgacggcac gcagccggcc
401 aatgggcgcc cgcgcggctg tccgggggcg gggccggcca gggctgggga atcccgctaa gtgtttggat tgctcggtgg
481 cgccgctgcc ctggcagagc tcgccactcc ttagtcgagg caagacgtgc gcccgagccc cgccgaaccg aggccacccg
561 gagccgtgcc cagtccacgc cggccgtgcc cggcggcctt aagaaccCgg caacctctgc cttcttccct cttccactcg
641 gagtcgcgct ccgcgcgccc tcactgcag
```

Figure 11 Complete IRF-1 upstream promoter sequence.

The re-established human IRF-1 promoter sequence is shown in this figure. The underscored capitalized nucleotide is the only nucleotide discrepancy between the reporting sequence and current GenBank reference sequence (X53095) in which "A" instead of "C" is present at this position. It is notable that this nucleotide difference is located in the overlapping region of L05072 and X53095 and it is identical with L05072.

statistically significant with p value at 0.0054 (**Table 10, Figure 12**). Genotype analysis showed similar result that 55.8% of HIV-1 resistant subjects had at least one copy of 179 allele (genotype $179^{+}179^{+}$ or $179^{+}179^{-}$) in comparison to 42.0% (250/595) in HIV-1 susceptible subjects.

To confirm the protective effect of this allele, we conducted a survival analysis in the same cohort to determine if the IRF-1 179 allele provided protection in prospective analysis. Kaplan-Meier and Cox Regression survival analysis demonstrated that the 179 allele was associated with longer HIV-1 free survival. HIV-1 negative subjects with at least one copy of the 179 allele had significantly longer HIV-1-free time, and therefore a decreased rate of seroconversion (p =0.039; Hazard Ratio: 0.708; 95% CI: 0.509~0.985; Log Rank: 4.24) (**Figure 13a**), demonstrating that the 179 allele not only associated with HIV-1 resistance, but also a reduced likelihood of seroconversion. These results confirmed previous findings and solidified the bases for our further studies.

Two SNPs in IRF-1 correlate with resistance to HIV-1 infection and decreased likelihood of seroconversion

As the IRF-1 MS showing positive association with resistance to HIV-1 lies within an intronic region and has no clear functional role, we conducted comparative sequence analysis of the IRF-1 gene and its promoter from a subset of 507 individuals to determine if the MS marker was in linkage disequilibrium (LD) with potential functional mutations. As described, the IRF-1 gene demonstrated extensive polymorphism, containing 53 SNPs, 2 insertions, and 1 deletion, none of which encoded amino acid changes (Ji et al., 2004).

Table 10 Distributions of specific IRF-1 alleles and genotypes in HIV-1-resistant and HIV-1-susceptible subjects

Variation sites	Allele	Allele frequencies				*p* values Odd Ratios (95% CI)
		Affected		Resistant		
		n	%	*n*	%	
IRF-1 MS	179	426	35.8	81	47.1	p =0.0054
	Non-179	764	64.2	91	52.9	OR=1.60(1.14~2.23)
619	A	143	28.8	76	43.2	p =0.00073
	C	353	71.2	100	56.8	OR=1.88 (1.29~2.72)
6516	G	152	29.8	69	39.2	p =0.030
	T	358	70.2	107	60.8	OR=1.52 (1.05~2.20)

Variation sites	Genotype	Genotype frequencies				*p* values In Chi-Square (2×3)
		Susceptible		Resistant		
		n	%	n	%	
IRF-1 MS	$179^{+}179^{+}$	176	29.6	33	38.4	p=0.0529
	$179^{+}179^{-}$	74	12.4	15	17.4	
	$179^{-}179^{-}$	345	58.0	38	44.2	
619	619AA	24	9.8	13	14.9	p =0.0007
	619AC	96	39.0	50	57.5	
	619CC	126	51.2	24	27.6	
6516	6516GG	25	9.8	14	16.1	p =0.087
	6516GT	102	40.0	40	46.0	
	6516TT	128	50.2	33	37.9	

Table 10 Distributions of specific IRF-1 alleles and genotypes in HIV-1-resistant and HIV-1-susceptible subjects.

Distribution of different alleles and genotypes for all three key IRF-1 polymorphisms in HIV-1 resistant and susceptible subjects were listed. Statistic analysis results comparing the differences between two subject groups by Chi-square test are depicted as well. Only results from polymorphisms significantly associated with resistance to HIV-1 infection are listed here.

The top panel shows the allele frequencies of three IRF-1 polymorphisms (at 619, IRF-1 MS and 6516) correlated with HIV-1 resistance. The frequency data shows the distribution of the specific allele on either chromosome (2n), so the subject numbers of each group equaled to total/2. Data showed that 179 allele at 619A, IRF-1 MS and 6516 G were significantly over-represented in HIV-1 resistant subjects compared to the alternative alleles at these loci. The *p* values, Odd ratios and 95% confidence intervals from Chi-Square tests (2×2) are listed as well.

The bottom panel shows the genotypes frequencies of the three IRF-1 polymorphisms (at 619, IRF-1 MS and 6516) correlated with HIV-1 resistance in different subject groups. Frequency data indicates the distribution of the specific genotypes in different subject groups. The data show that genotypes containing 619A, 179 allele at IRF-1 MS and 6516 G were over-represented in HIV-1 resistant subjects compared to the genotypes without these alleles. The *p* values from Chi-Square tests (2×3) are listed as well.

Figure 12 Correlation between IRF-1 gene variations and resistance to HIV-1 infection

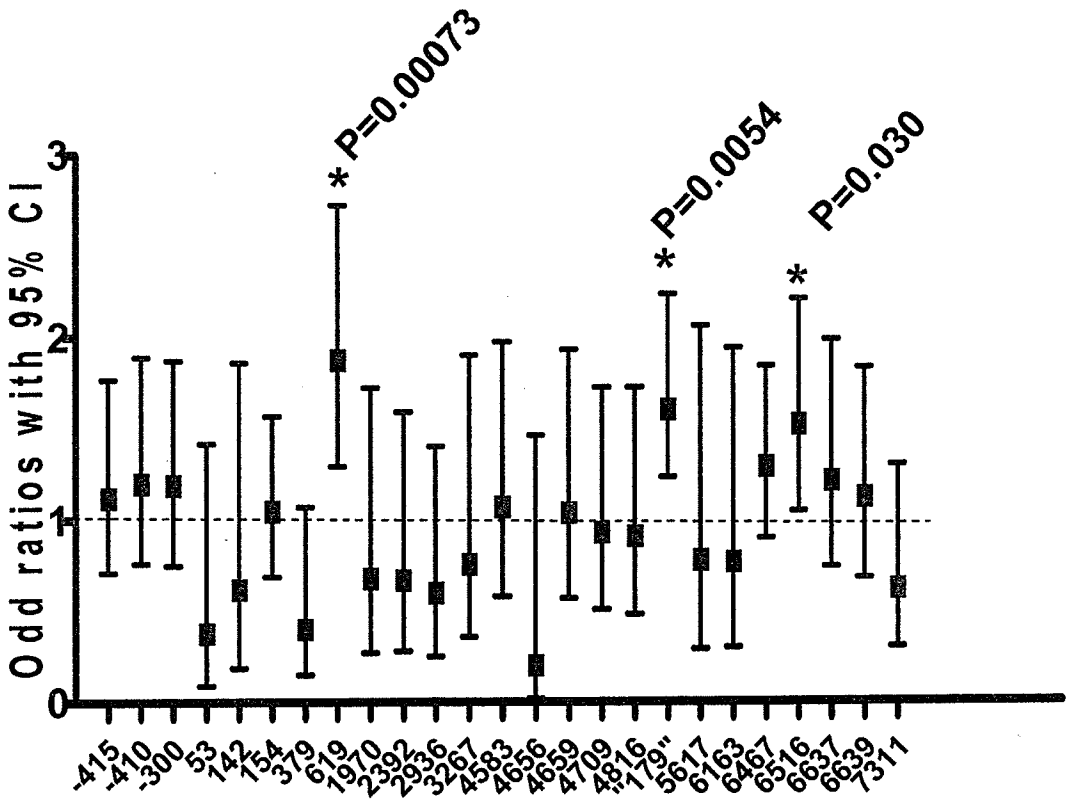


Figure 12 Correlation between IRF-1 gene variations and resistance to HIV-1 infection.

The association between IRF-1 gene variations and resistance to HIV-1infection was analyzed by comparing the allele distributions in HIV-1 susceptible and resistant groups using Chi-square test. Only results from polymorphisms with minor allele frequencies higher than 4 % are depicted. The odd ratios with the 95% confidence intervals (CI) and *p* values for each polymorphism showing significant associations with HIV-1 resistance are plotted. SNPs at 619 and 6516 and "179" allele at IRF-1 microsatellite region were shown to be significantly correlated with HIV-1 resistance.

Figure 13 Kaplan-Meier survival analysis comparing individuals with different IRF-1 genotypes

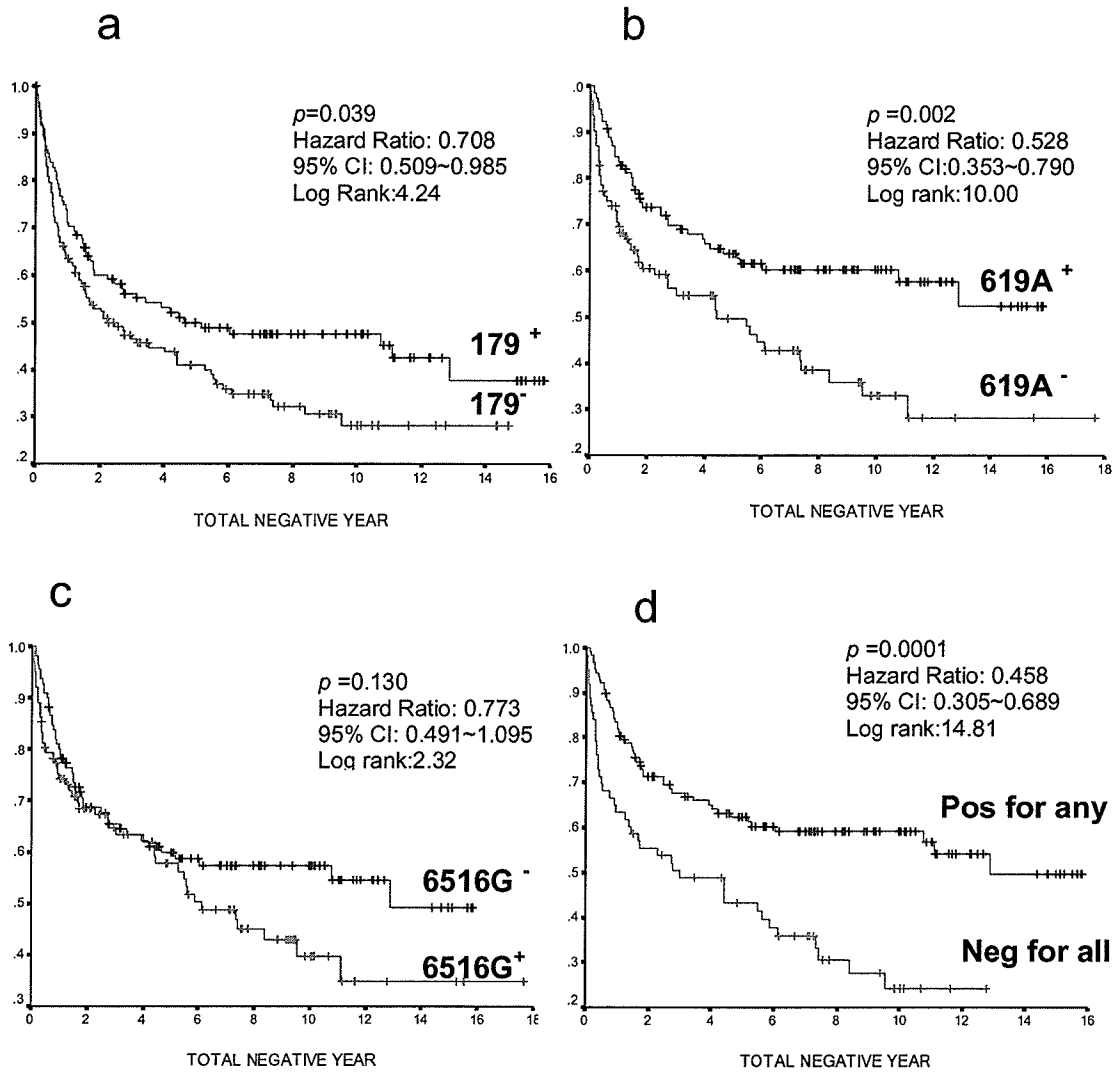


Figure 13 Kaplan-Meier survival analysis comparing individuals with different IRF-1 genotypes.

All initially HIV-1 uninfected subjects who were followed over time were included in this analysis. The HIV-1- free time was counted and plotted with the corresponding IRF-1 genotypes. Statistic analysis was conducted by Kaplan-Meier and Cox Regression survival analysis. *p* values, hazard ratios, 95% confidence interval (CI) and log rank values from these analyses are listed. **(a)** Those individuals who had at least 1 copy of 179 allele at IRF-1 microsatellite site were significantly less likely to seroconvert. **(b)** 619A allele also showed a significant protective effect again HIV-1 infection in the subject group. **(c)** 6516G allele showed a strong trend towards a significant association with decreased susceptibility to HIV-1 infection with significance at 0.130. **(d)** Additive protection is observed when all three protective alleles are combined.

We analyzed these additional polymorphisms to determine if they correlated with HIV-1 resistance phenotype. Of all the 53 SNPs investigated, only 2 SNPs showed significant associations with HIV-1 resistance. SNPs at positions 619 (A/C) (rs17848395) and 6516 (G/T) (rs17848424) were differentially represented in HIV-1 resistant subjects compared to HIV-1 susceptible women (**Figure 12, Table 10**). The 619A and 6516G alleles were over-represented in HIV-1 resistant subjects, while the alternative alleles (619C and 6516T) showed significantly higher frequencies in HIV-1 susceptible subjects (p=0.00073 and 0.030 respectively). The allele frequencies for 619A and 6516G were 43.2% (76/176) and 39.2% (69/176) in HIV-1-resistants and 28.8% (143/496) and 29.8% (152/496) in HIV-1-susceptibles. Genotype analysis showed similar result: 74.4% (63/87) and 62.1% (54/87) of HIV-1 resistant subjects had at least one copy of 619A (619AA or 619AC) or 6516G allele (6516GG or 6516GT) in comparison to 48.8% (120/246) and 49.8% (127/255) respectively in HIV-1 susceptible subjects. The differences between HIV-1 resistant and susceptible groups were statistically significant with p values at 0.0002 for 619A and marginally significant for 6516G with p values at 0.06.

We also noted that, in this population, the protective alleles occurred at a much lower frequencies (619A: 29.7%; 179 at IRF-1 MS: 36.8%; 6516G: 29.6%) than the alternative alleles (619C: 70.3%; non-179 alleles at IRF-1 MS: 63.2%; 6516T: 70.4%). When we compared protective IRF-1 allele frequencies between Kenyan subjects and our local donors with a diversity of non-African genetic backgrounds, we noticed that frequencies of 619A and 6516G were significantly lower in Kenyan subjects (29.7% vs. 48% for 619A and 29.6% vs. 50% for 6516G). One could hypothesize that the lower frequencies of protective IRF-1 alleles in general population might contribute to the severe HIV-1

endemic in Africa to some extent.

We conducted Kaplan-Meier and Cox Regression survival analysis in an independent prospective study to determine the effects of these alleles on seroconversion rates in HIV-1 uninfected subjects. Similar to what we found in analysis of the 179 allele at IRF-1 MS, individuals with at least one copy of the 619A allele had significantly lower seroconversion rates than those lacking this allele (p =0.002; Hazard Ratio: 0.528; 95% CI: 0.353~0.790; Log Rank: 10.00) (**Figure 13b**). Although not statistically significant, a similar trend was observed for the 6516G allele (p =0.130; Hazard Ratio: 0.773; 95% CI: 0.491~1.095; Log Rank: 4.24) (**Figure 13c**).

So far, 619A, the 179 allele at IRF-1 MS and 6516G have been identified to be significantly associated with resistance to HIV-1 infection. They all showed significant associations with both HIV-1 resistance phenotype and longer HIV-1 free survival in prospective studies. We conducted a separate set of analyses in evaluating the protective effects of possessing one or more of these alleles in resistance to HIV-1. When we assessed the effects of having at least one copy of protective alleles, either 619A, 179 in MS, or the 6516G allele, we noted that 78.8% of HIV-1 resistant individuals carried at least one copy of these protective IRF-1 alleles compared to 55.6% in HIV-1 infected subjects (p =0.0003, OR=2.97, 95% CI: 1.61~5.53). When we considered the effect of these three protective alleles together in prospective survival analysis, we found that subjects having at least one copy of these alleles had a significantly longer HIV-1 disease free survival and reduced likelihood of seroconversion compared to individuals lacking these alleles (p =0.0001; Hazard Ratio: 0.458; 95% CI: 0.305~0.689: Log Rank: 14.81) (**Figure 13d**). These results further confirmed our findings from analysis on these alleles

individually as to the importance of these alleles in resistance/susceptibility to infection by HIV-1.

Haplotypes containing more than one protective alleles associated with resistance to HIV-1 infection

Linkage disequilibrium had been shown to be common in IRF-1 in our studied subject cohort and the three polymorphisms correlated with resistance to HIV infection were in significant LD as well. The protective alleles tend to show up together as a haplotype. To assess if these alleles provide additive protective effects we conducted haplotype analysis where possible, in individuals homozygous for two or more protective alleles. Comparing haplotype frequencies containing one or more alleles associated with HIV-1 resistance, not surprisingly we found that all combinations of protective alleles (619A+179, 179+6516G, 619A+6516G and 619A+179+6516G) significantly associated with protection from HIV-1 infection (619A +179: p=0.0114, OR=1.80(1.14~2.86); 179+6516G: P=0.0005, OR= 1.91 (1.21~3.04); 619A+6516G: p=0.0006, OR=2.71 (1.49~4.89); 619A+179 + 6516G: p =0.0008, OR=2.41 (1.23~4.70)) (**Table 11**). Thus, compared to individual markers, the presence of additional protective alleles appears to provide additive protection against HIV-1 infection. Numbers of subjects with discrete haplotypes were insufficient to conduct survival analysis.

Table 11 Distributions of specific IRF-1 haplotypes consist of variations at IRF-1 MS, 619 and 6516 in HIV-1-resistant and HIV-1-susceptible subjects

Haplotypes	Haplotype frequencies				*p* values Odd Ratios (95% CI)
	Affected*		Resistant*		
	n	%	*n*	%	
619A +179	83	18.8	40	29.4	*p* =0.0114
	359	81.2	96	70.6	OR=1.80(1.14~2.86)
179+6516G	83	19.2	42	31.3	*p* =0.0005
	349	80.8	92	68.7	OR=1.91 (1.21~3.04)
619A+6516G	41	13.3	27	29.3	*p* =0.0006
	267	86.7	65	70.7	OR=2.71 (1.49~4.89)
619A+179 + 6516G	32	10.3	19	21.6	*p* =0.0008
	280	89.7	69	78.4	OR=2.41 (1.23~4.70)

Table 11 Distributions of specific IRF-1 haplotypes consist of variations at IRF-1 MS, 619 and 6516 in HIV-1-resistant and HIV-1-susceptible subjects.

Distributions of different IRF-1 haplotypes consisting of polymorphisms at IRF-1 MS, 619 and 6516 in HIV-1 resistant and susceptible subjects were depicted. Statistic analysis results comparing the differences between two subject groups by Chi-square test were listed as well. (*) The top numbers indicated the frequencies of listed haplotypes and the bottom numbers showed the alternative haplotypes in the corresponding subject groups. Data showed that haplotypes containing 179 allele at IRF-1 MS, 619A or 6516 G were significantly over-represented in HIV-1 resistant subjects compared to the alternative haplotypes at these loci.

Polymorphism at 619 is the predominant and independent indicator of HIV-1 resistance

Since all three polymorphisms in IRF-1 (IRF-1 MS, 619A/C, 6516G/T) showed protective effects in resistance to HIV-1 infection were in linkage disequilibrium, the multivariate analysis was required to identify the likely dominant polymorphism providing the most significant protection. Log Linear multivariate analysis revealed that 619A/C was the only independent and significant indicator for HIV-1 resistance (p=0.0022), while the other two (179 at IRF-1 MS and 6516G/T) were associated with HIV-1 resistance likely due to their high correlation with 619A/C.

IRF-1 alleles that are associated with resistance to infection by HIV-1 do not correlate with HIV disease progression

It is rational to assume that, whatever the protective biologic effects the polymorphisms in IRF-1 may provide, the alleles that appear to protect against HIV-1 infection might also play a role in protecting against disease progression in HIV-1 infected subjects. For instance, homozygotes for CCR-Δ32 mutation showed resistance to HIV-1 infection and heterozygotes for this allele exhibited significantly slower disease progression (Dean et al., 1996). To determine if these polymorphisms protect against disease progression, subjects with different IRF-1 genotypes at the three loci were compared. Kaplan-Meier and Cox regression survival analysis was performed comparing the durations between the first HIV-1 positive day in the cohort and first day their $CD4^+$

cell count decreased to <400 cells/μl or 200 cells/μl, which were utilized as milestones reflecting disease progression to intermediate or late stage infection. The results of this analysis are shown in **Figure 14**. Specifically, neither 619 A nor 6516 G showed any protective effects on the HIV disease progression as determined by the time for $CD4^+$ cell count to drop to below either 400 cells/μl or 200 cells/μl. The only marginal significant difference was observed when comparing the time of CD4 count drop to 200 cells/μl in subjects with different genotypes at IRF-1 MS although no significant difference was observed when comparing the time of CD4 count drop to 400 cells/μl . Heterozygous subjects having one copy of 179 allele have marginally significantly shorter time to progress to CD4 count below 200 cells / μl (p=0.0426). These results indicated that the seemingly protective IRF-1 genotypes only provide protection from infection at the onset of the infection, but do not affect the course of disease progression.

The frequencies of protective IRF-1 alleles decreased over time

This is the first report revealing gene polymorphisms in IRF-1, an immune regulatory factor, correlate with resistance to HIV-1 infection. This suggests that some specific IRF-1 alleles are providing protection to the host from being infected by HIV-1. It is noteworthy that the epidemic of HIV/AIDS showed no sign of waning despite the fact that people have identified multiple protective mechanisms mediating resistance to HIV-1 infection including genetic polymorphisms. Theoretically, the accumulation of protective alleles in the population should be observed and this positive evolution will help to

Figure 14 Kaplan-Meier Survival analysis comparing disease progression for individuals with different IRF-1 genotypes

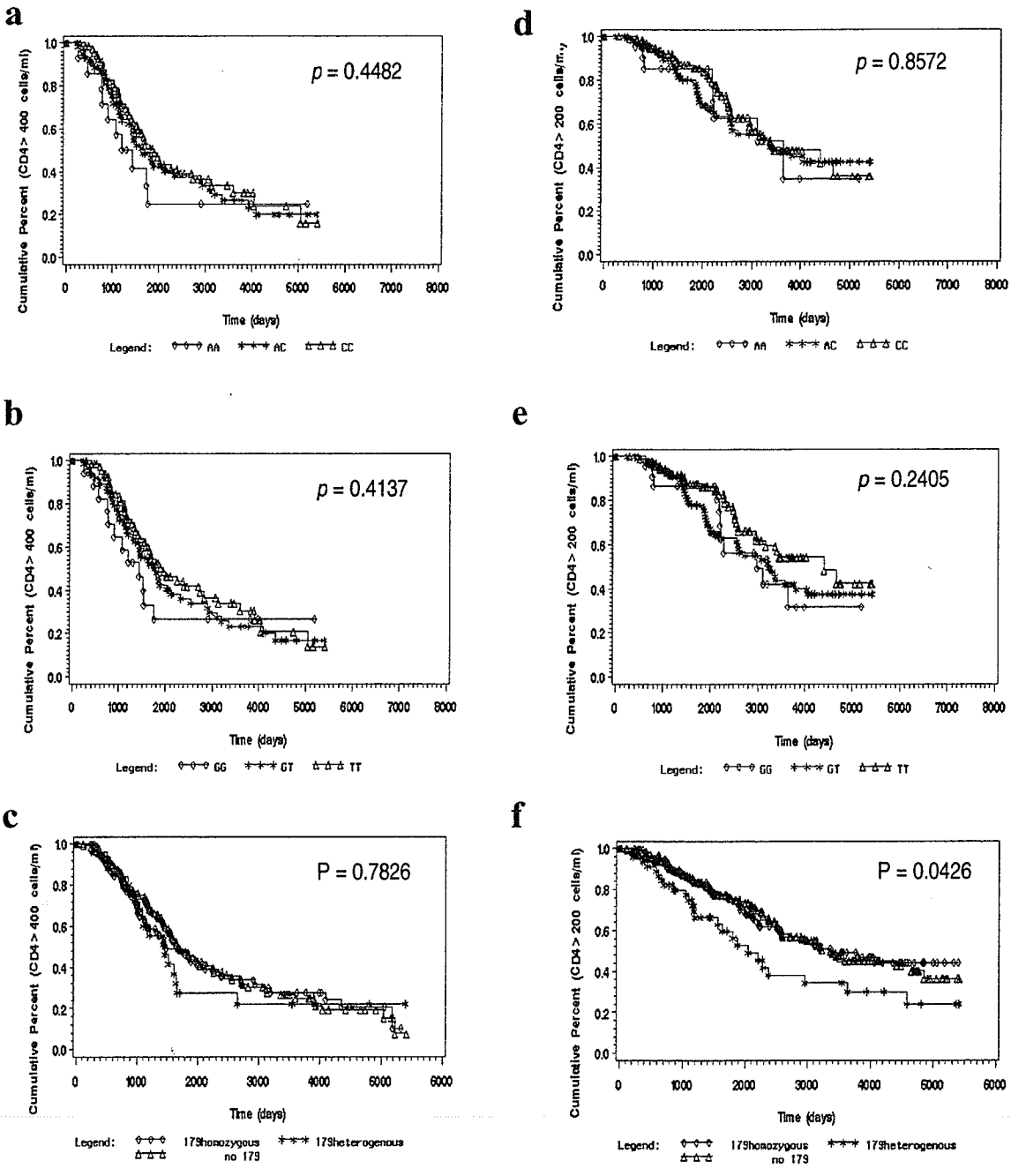


Figure 14 Kaplan-Meier Survival analysis comparing disease progression for individuals with different IRF-1 genotypes.

Only HIV-1 infected subjects were included in this analysis. The durations of the first HIV-1 positive day in the cohort to the first day of CD4 count declined to below 400 cells/μl (**a, b** and **c**) or 200 cells/ μl (**d, e** and **f**) were compared between subjects with different IRF-1 genotypes showing significant correlations with HIV-1 resistance. *p* values after Log-Rank statistic analysis are plotted as well. Except for genotypes at IRF-1 MS, no significant differences were observed in HIV disease progression when comparing subjects with and without detected protective IRF-1 alleles (**a** and **d** for 616A/C, **b** and **e** for 6516G/T and **c** and **f** for 179 allele). The only marginally significant difference is observed when comparing the time of CD4 count drop to 200 cells / μl in subjects with different genotypes at IRF-1 MS (**f**). Heterozygous subjects having one copy of 179 allele appear to progress faster to CD4 count below 200 cells / μl.

contain the HIV/AIDS pandemic naturally over time. Thus we performed an epidemiological analysis to investigate how the frequencies of protective IRF-1 alleles changed over time. According to the date of enrollment, all subjects were divided into two groups: those who enrolled before 1993 and those with entry in 1993 or later. As shown in **Table 12**, the frequencies of all protective IRF-1 alleles significantly decreased over time when comparing the two study groups. These results are contrary to our original hypothesis and unpublished data from our group indicating the protective HLA allele frequencies increased when doing a similar comparison in this cohort (Unpublished data from Drs. Frank Plummer and Ma Luo). When we compared these allele frequencies between Kenyan subjects and our local donors with a diversity of non-African genetic backgrounds, we noticed that the protective allele frequencies of 619A and 6516G were significantly lower in Kenyan subjects (29.7% vs. 48% for 619A and 29.6% vs. 50% for 6516G). The lower frequencies of protective IRF-1 alleles in general population might contribute to the severe HIV-1 endemic in Africa to some extent.

Table 12 IRF-1 allele frequencies before and after 1993

Variations		Before 1993	After 1993	
619 A/C	619A	158 (33.6%)	139 (26%)	Chi-Square test (2X2)
	619C	312 (66.4%)	389 (74%)	p =0.0145
				OR=1.417 (1.07-1.88)
6516G/T	6516G	161 (38.1%)	124 (25.6%)	Chi-Square test (2X2)
	6516T	311 (61.9%)	364 (74.4%)	p =0.004
				OR=1.52 (1.14-2.03)
IRF-1 MS	179	455 (38.2%)	82 (30.6%)	Chi-Square test (2X2)
	Non-179	737 (61.8%)	186 (69.4%)	p =0.024
				OR=1.40 (1.04-1.88)

Table 12 IRF-1 allele frequencies before and after 1993.

The allele frequencies in subjects enrolled before or after 1993 were compared for all three IRF-1 gene variations showing correlations with resistance to HIV-1 infection. The data shown in the table were allele numbers (percentage of all presented alleles) in the two subject groups. Chi-Square test results for comparison were depicted as well. In general, the frequencies of protective alleles, 619A, 179 at IRF-1 MS and 6516G, were all significantly decreased in subjects enrolled after 1993 when compared to those before 1993.

Section Three: Functional analysis of identified key IRF-1 polymorphisms

With the observation that IRF-1 gene polymorphisms were significantly associated with natural resistance to HIV-1 infection in our female sex worker subjects, we hypothesized that these polymorphisms exert important functional impacts on the expression and function of IRF-1 and these functional alterations could be mediating the observed association. Extensive functional analysis was conducted on the roles of the three key IRF-1 polymorphisms at both protein and mRNA levels. Functional examination of the impacts of these IRF-1 gene variations on HIV-1 transcription at the initial stage of infection were conducted as well.

IRF-1 genotypes correlated with resistance to HIV-1 are associated with reduced basal level IRF-1 expression

With the exception of early embryonic cells, low basal levels of IRF-1 mRNA are expressed in all cell types (Kroger et al., 2002). To investigate if polymorphisms associating with HIV-1 resistance have functional consequences, we first tested basal IRF-1 protein expression in peripheral blood mononuclear cells (PBMCs) from HIV-1 uninfected, IRF-1 genotyped Kenyan subjects by semi-quantitative Western Blot. The ratios of IRF-1/actin based upon spot densitometry analysis readouts from subjects with different IRF-1 genotypes were calculated and utilized as index for intra-group comparisons. The results showed that PBMCs from subjects homozygous for the 619A allele (genotype 619AA) had significantly reduced IRF-1 protein expression compared to

PBMCs from subjects with one copy of this allele (genotype 619AC) or lacking this allele (genotype 619CC) (p=0.0011 and 0.0010, respectively). No significant differences were detected in PBMCs from subjects with 619AC and 619CC genotypes (**Figure 15a**). We obtained similar results in experiments comparing IRF-1 expression between subjects with different genotypes at the 6516 locus. Those homozygous for the 6516 G alleles (genotype 6516GG) also showed reduced basal IRF expression compared to heterozygous individuals (genotype 6516GT) or individuals homozygous for the alternative allele (genotype 6516TT) (p=0.0028 and 0.0015 respectively) but no significant difference observed between cells with 6516GT and 6516TT genotypes (**Figure 15b**). A similar trend was observed when comparing PBMCs with different IRF-1 MS genotypes, although significant differences were only observed between individuals homozygous for 179 allele and individuals lacking the 179 allele entirely (p =0.0093) (**Figure 15c**). These data suggested that the identified polymorphisms in IRF-1 gene result in functional differentiation in basal level IRF-1 expression.

IRF-1 genotypes correlated with resistance to HIV-1 are associated with reduced responsiveness to exogenous IFN-γ stimulation

IRF-1 and IFN responses represent the typical initial host immune responses to viral infection and usually lead to a potent antiviral state in the host and bystander cells (Marsili et al., 2003). The ability of IRF-1 to respond to exogenous stimulation represents an important character of the IRF-1 function in containing viral infection. A number of

Figure 15 IRF-1 genotypes correlated with resistance to HIV-1 infection are associated with reduced basal IRF-1 expression and responsiveness to exogenous IRF-γ stimulation

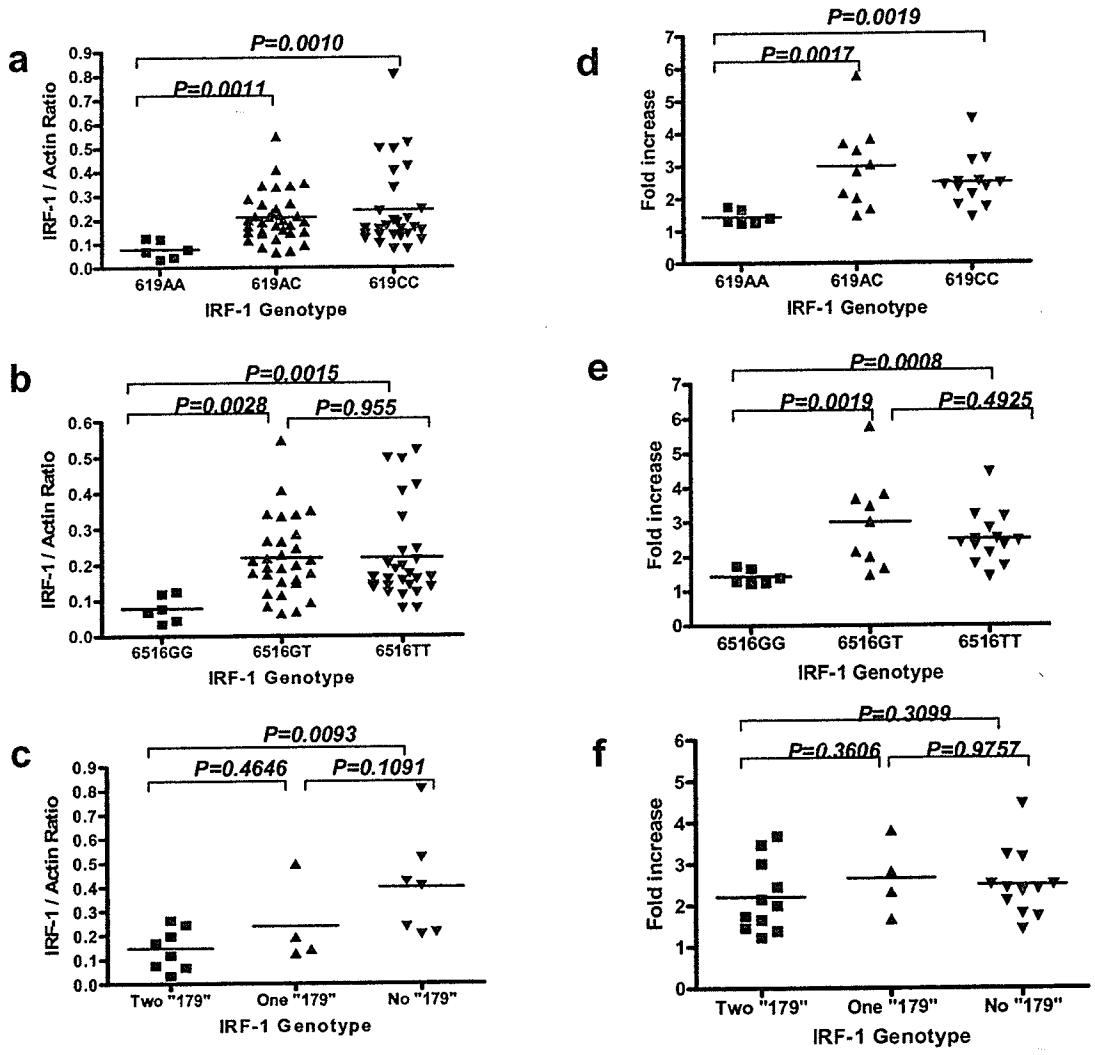


Figure 15 IRF-1 genotypes correlated with resistance to HIV-1 infection are associated with reduced basal IRF-1 expression and responsiveness to exogenous IRF- γ stimulation. Basal level IRF-1 expression in unstimulated peripheral blood mononuclear cells (PBMCs) from HIV-1 negative individuals with different IRF-1 genotypes were analyzed by Western Blot (**a**, **b** and **c**). Semi-quantitative comparison was done by comparing the IRF-1/actin ratio in spot densitometry read-outs and analyzed by non-parametric Mann-Whitney test (2 tailed). **a)**: Cells with 619AA genotype had significantly lower IRF-1 basal expression level when compared to those with either 619AC or 619CC genotype although no significant difference observed between the later two groups; **b)**: Similar results also obtained when comparing IRF-1 expression in PBMCs from individual with genotypes of 6516GG, 6516GT and 6516TT; **c)**: This trend also exists when doing same analysis on IRF-1 MS although significant difference was only observed only between 179179 homozygotes and individuals without this 179 allele at all.

IRF-1 responsiveness to exogenous IFN-γ stimulation was analyzed. Semi-quantitative comparison was done by comparing the spot densitometry read-outs and the fold increase of IRF-1 expression after IFN-γ stimulation as an index for intra-group comparison by non-parametric Mann-Whitney test (2 tailed) (**d**, **e**, and **f**). Similar to the results from the basal level comparison, PBMCs from homozygotes with "protective" alleles at 619 and 6519 showed significantly lower IRF-1 response indicated by lower fold-increase to the basal level expression when compared to heterozygotes or homozygotes with the alternative alleles (**d** and **e**); No significant differences observed when doing the same analysis on IRF-1 MS markers (**f**).

agents are known to induce IRF-1 responses including infection by HIV-1 (Kroger et al., 2002; Sgarbanti et al., 2002). Using IFN-γ, the strongest and most commonly used IRF-1 stimulus, we determined if key IRF-1 polymorphisms were associated with differential responsiveness to exogenous stimulation. PBMCs from individuals with different IRF-1 genotypes were cultured for 18 hrs in presence or absence of IFN-γ and then analyzed by semi-quantitative Western Blot. The fold-increases of IRF-1 expression upon unstimulated basal level expression after IFN-γ stimulation was calculated and compared among subjects with different IRF-1 genotypes. Similar to the previous results, PBMCs from individuals homozygous for alleles associated with HIV-1 resistance (genotype 619AA or 6516GG) showed significantly lower fold-increase in IRF-1 expression compared to individuals heterozygous or homozygous for the alternate alleles (p =0.0017 and 0.0019 for genotype 619AA vs. genotype 619AC and 619CC respectively; p =0.0019 and 0.0008 for genotype 6516GG vs. genotype 6516GT and 6516TT respectively). No differences were observed between individuals heterozygous or homozygous for the non-associated alleles (**Figure 15d** and **15e**), nor amongst the different subject groups with distinct genotypes at IRF-1 MS region (**Figure 15f**).

IRF-1 mRNA splicing patterns are distinct in cells with different IRF-1 genotypes

Multiple mechanisms can account for the association of intronic polymorphisms and differential protein expression (Herbert, 1996; Mattick, 1994; Moore, 1996). A possible consequence of intronic polymorphism is change in the mRNA splicing pattern due to alternative splicing. Alternative splicing, including the skipping of IRF-1 exon 2, 3, 7, 8

and 9 has been previously reported to be increased in leukemia patients and skipping of exon 2 and 3 has been proposed to be a potential mechanism of IRF-1 inactivation (Green et al., 1999; Harada et al., 1994a; Tzoanopoulos et al., 2002). To test if specific IRF-1 genotypes and intronic polymorphisms are associated with altered splicing patterns, we conducted IRF-1 specific reverse transcription PCR (RT-PCR) using RNA extracts from cells with different IRF-1 haplotypes. Nineteen HIV-1 negative subjects with three different IRF-1 haplotypes: 7 subjects with haplotypes "179179 + 619AA + 6516GG", 6 subjects with "$179^{+}179^{-}$ + 619AC + 6516GT and another 6 with haplotype "$179^{-}179^{-}$ + 619CC + 6516TT". These haplotypes covered the best, the medium and the worst IRF-1 allele compositions based on their correlations with HIV-1 resistance. GAPDH was also amplified and utilized as an internal control. The IRF-1 specific RT-PCR products were visualized by ethidium bromide staining (**Figure 16**).

The presence of a long tail of smaller RT-PCR IRF-1 amplicons in all samples tested hinted the likelihood of multiple bands with different lengths in these RT-PCR products instead of only a single PCR amplicon (**Figure 16**). One explanation for these minor bands would be alternative splicing resulting in mRNA with various lengths. To determine the detailed nucleotide sequences and thus mRNA transcript formats of these bands for further analysis, the RT-PCR amplicons were cloned into TOPO cloning plasmid vector and subsequently transformed and amplified in *Escherichia coli*. Thirty-two clones for each subject were randomly picked for subculture and the plasmid DNA was then extracted. The presence of inserts was confirmed by restriction endonuclease digestion and agarose electrophoresis. The EcoRI restriction sites in the plasmid were utilized to release the cloned insert fragments. Representative photographs of the

Figure 16 IRF-1 RT-PCR products after agarose gel electrophoresis

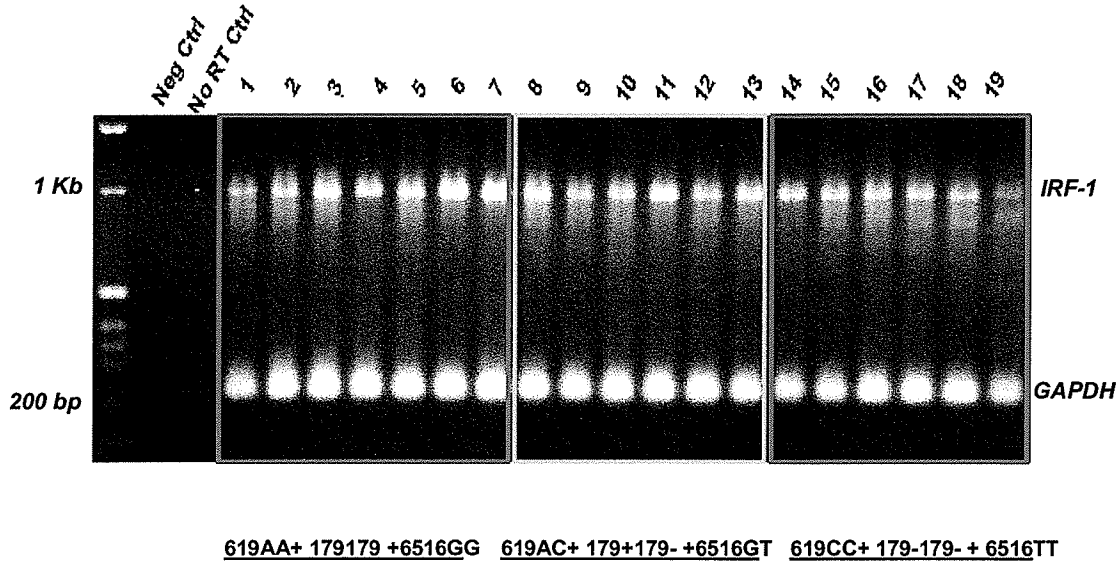


Figure 16 IRF-1 RT-PCR products after agarose gel electrophoresis.

Nineteen HIV-1 negative subjects with three different IRF-1 haplotypes were included in this experiment. Subjects 1 to 7 were homozygous for all three "protective" alleles (619AA+179179+6516GG), subjects 8 to 13 were heterozygous (619AC + $179^{+}179^{-}$ + 6516GT) and subject 14 to 19 were homozygous for the alternative alleles (619CC + $179^{-}179^{-}$ + 6516TT) at all three loci. Total RNA was extracted from all subjects and cDNA was synthesized subsequently. IRF-1 specific RT-PCR was performed on all cDNA samples to amplify a 1006 nt segments of IRF-1 cDNA. GAPDH was also amplified and taken as internal control. This figure shows the all IRF-1 RT-PCR products after agarose gel electrophesis with Ethidium Bromide staining. An obscure long tail follows all IRF-1 bands hinting that the IRF-1 RT-PCR products likely contain multiple bands with different lengths.

enzymic digested products from two subjects with two divergent IRF-1 haplotypes (619AA + 179179 + 6516GG vs 619CC + 179-179- + 6516TT) are shown in **Figure 17**. These results confirmed our assumption that the RT-PCR products contain multiple mRNA products with different lengths.

It is noteworthy that none of the IRF-1 intronic polymorphism associated with HIV resistance is located at predicted splicing sites (data not shown). To determine the nucleotide sequences of non-full-length IRF-1 amplicons' inserts, the purified plasmids were sequenced and aligned with published full length IRF-1 mRNA sequence (NM_002198). Exon skipping patterns were determined by analyzing the exon compositions of the cloned RT-PCR amplicons in comparison to reference mRNA sequence. The results confirmed that exon skipping was fairly common in IRF-1 mRNA and different exon skipping patterns were detected in all subjects. Exon skipping occurred mainly to exons 2, 3, 7, 8 and 9, rarely to exon 5. The representative mRNA expression patterns from the two subjects shown in **Figure 17** are depicted in **Figure 18**. Remarkably, skipped exons were mostly adjacent to the IRF-1 polymorphisms correlated with HIV-1 resistance phenotype. Exon 2 encompasses nt 1296 to nt 1382 in the IRF-1 gene and it is adjacent to intron 1 where polymorphism 619A/C resides; exon 7 (nt 4401 to 4523) and exon 8 (nt 5058 to 5107) are neighboring intron 7 where the IRF-1 MS marker is located ; and exon 9 (nt 6277 to 6412) is adjacent to 6516G/T which is in intron 9 of the IRF-1 gene. (**Figure 19**)

Figure 17 Agarose gel electrophoresis of TA cloned IRF-1 RT-PCR products

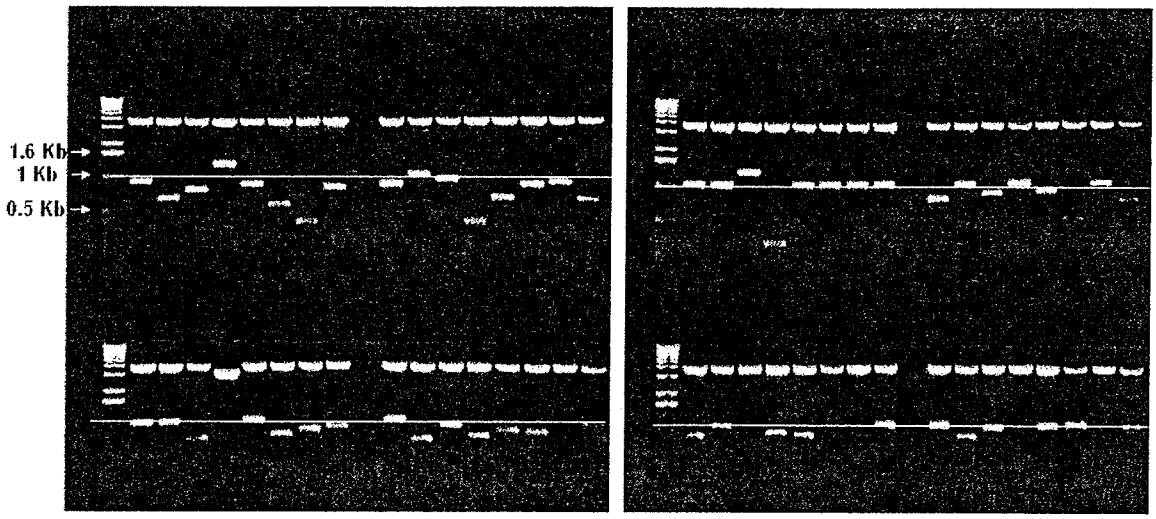


Figure 17 Agarose gel electrophoresis of TA cloned IRF-1 RT-PCR products

This figure shows two representative agarose gels photographs with ethidium bromide staining. Panel A shows the result from a patient with haplotype "179179/ 619AA/6516GG" and significant deviation of band sizes was observed. Panel B shows the result from a patient with haplotype "179179/ 619CC/6516TT" and fewer deviations of band sizes were detected in these subjects. The white lines indicate the expected size of full length IRF-1 RT-PCR product.

Figure 18 RepresentiveIRF-1 mRNA exon skipping patterns

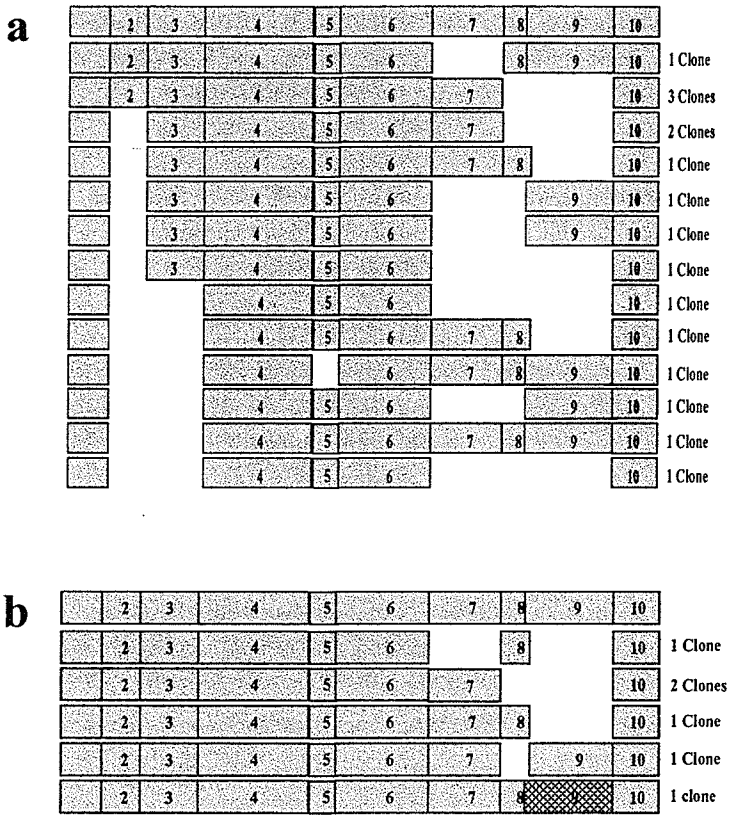


Figure 18 RepresentiveIRF-1 mRNA exon skipping patterns

This figure shows two representative exon skipping patterns in the same two subjects with distinct IRF-1 haplotypes as shown in Figure 17. Numbers of clones with different exon skipping pattern were indicated as well. (**a**) shows result from the patient with haplotype "179179/619AA/6516GG", which is significantly associated with resistance to HIV-1 infection. Frequent skipping of exons 2,3,7,8 and 9 was observed, a rare exon 5 skipping is also observed; (**b**) shows result from the patient with haplotype "179˙179˙ /619CC/6516TT", the combination of alternative alleles at these three loci. Skipping of exons 7, 8 and 9 was observed. The hatched exon 9 indicated that exon 9 sequence is significantly changed due to alternative splicing in this specific clone. Exon 2 and 3 skippings were rarely detected in subjects with this IRF-1 haplotype (not shown here). Significant difference was detected when comparing the frequencies of exon 2, 3, 7 and 8 skippings between subjects with these two IRF-1 haplotypes in Chi-Square tests (discussed in text).

Figure 19 Distribution of IRF-1 polymorphisms correlated with HIV-1 resistance and frequently skipped exons in IRF-1 mRNA

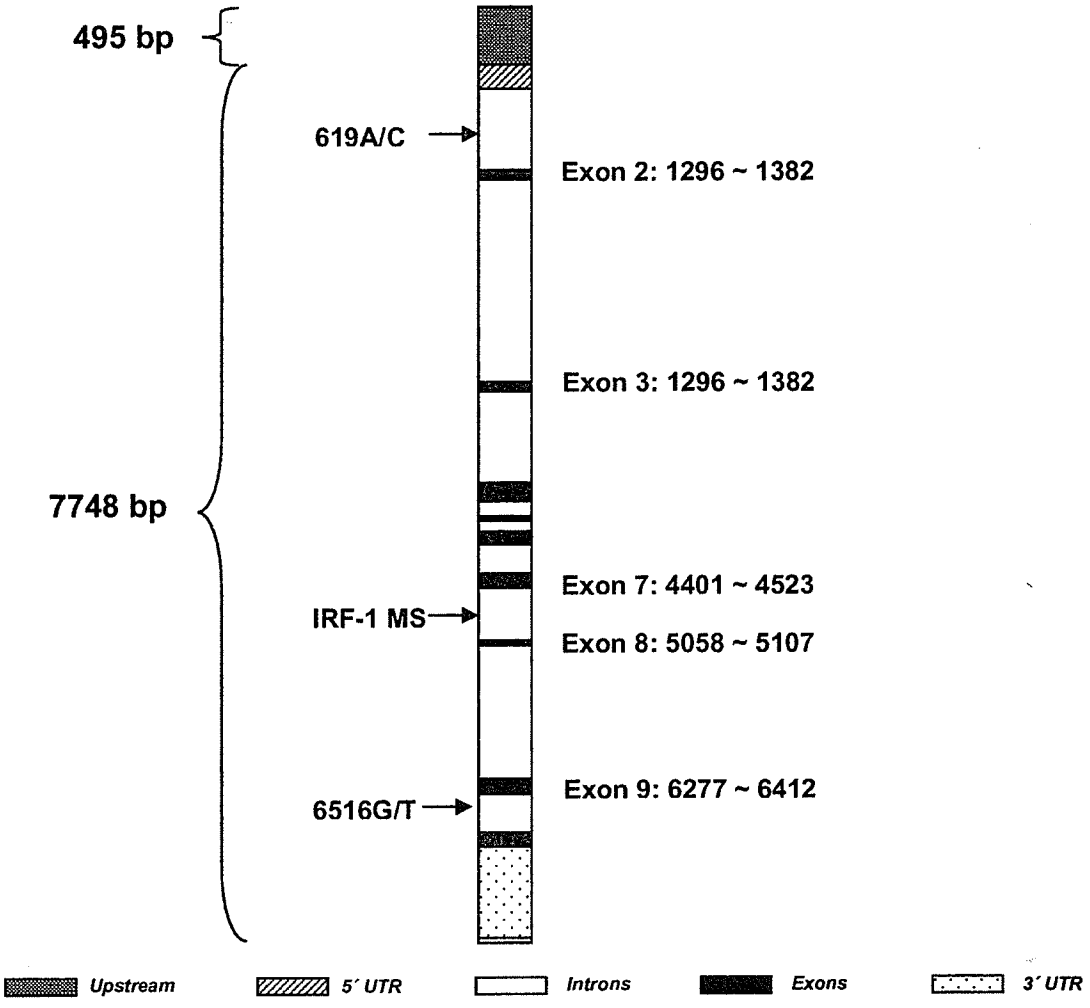


Figure 19 Distribution of IRF-1 polymorphisms correlated with HIV-1 resistance and frequently skipped exons in IRF-1 mRNA.

This diagram shows the distributions of three IRF-1 polymorphisms correlated with HIV-1 resistance and the frequently skipped exons in IRF-1 gene. IRF-1 gene is 7,748 bp in length and has a 495 bp immediate promoter region. IRF-1 gene has 10 exons, 10 introns and 2 untranslated terminal region at both ends. Exon skipping occurred mainly to exons 2, 3, 7, 8 and 9, rarely to exon 5. As indicated in this diagram, frequently skipped exons including exon 2, 7, 8 and 9 are closely related with the three key IRF-1 polymorphisms. Exon 2 encompasses nt 1296 to nt 1382 in the IRF-1 gene and it is adjacent to intron 1 where polymorphism 619A/C resides; exon 7 (nt 4401 to 4523) and exon 8 (nt 5058 to 5107) are neighboring intron 7 where the IRF-1 MS marker is located; and exon 9 (nt 6277 to 6412) is adjacent to 6516G/T which is in intron 9 of the IRF-1 gene.

The frequencies of skipping of certain exons were then statistically compared between different subject groups by Chi-Square test. Analysis of IRF-1 frequencies of exon skipping in the tested subjects revealed that individuals with different IRF-1 haplotypes had distinct skipping patterns. As described in **Table 13**, compared to the other two subjects groups (subjects with haplotype 619AC + $179^{+}179^{-}$ + 6516GT or 619CC + $179^{-}179^{-}$ + 6516TT), subjects with the protective IRF-1 haplotype (619AA+179179+6516GG) showed significantly higher frequencies of exon 2 skipping (p =0.0002, OR=2.72, 95%CI: 1.58~4.69) and exon 3 skipping (p =0.00005, OR=3.85, 95%CI: 1.90~7.86), but lower frequencies of Exon 7 (P=0.012, OR=0.53, 95%CI: 0.31~0.88) and exon 8 skipping (p =0.015, OR=0.51, 95%CI: 0.30~0.89). No significant difference was observed for exon 9 or exon 5 skippings in different subject groups.

It is interesting to note that the start codon for IRF-1 mRNA translation resides in exon 2 (exon 1 in IRF-1 mRNA is not encoding (Cha et al., 1992)). Open reading Frame analysis of the mRNA sequences missing exon 2 revealed that the "ATG" start codon shift to nt 7 ~ 9 of downstream exon 3. When exon 3 or both exon 2 and 3 are skipped, the start codon shifted further downstream to nt 66~68 in exon 4. None of these start codon shifts result in coding sequence changes downstream of the new start sites in comparison to GenBank reference sequence (NM_002198). However, skippings of exon 2, 3 or both in IRF-1 mRNA all result in translated proteins lacking the N-terminal DNA binding domain (the "WWWWW" repeat motif in aa 1 ~121 encoded by exons 2, 3 and 4 together (Schaper et al., 1998)) (**Figure 20**). Exon 7 skipping results in a truncated protein lacking the corresponding segment coded by exon 7 (**Figure 20**). Significant open reading frame shifts occurs when exon 8, 7 and 8, or exon 9 skipped. When Exon 8

Table 13 Frequencies of different exon skippings in subjects with different IRF-1 haplotypes

		Skipped exons			
		Exon 2	Exon 3	Exon 7	Exon 8
Exons skipping Frequencies	Group A (n=192)	37	26	25	21
	Group B (n=192)	13	2	46	40
	Group C (n=192)	18	13	39	34
p values In Chi-Square (2×2)	A vs B	0.0004	0.000007	0.009	0.02
	A vs C	0.009	0.04	0.075	0.08
	A vs (B+C)	0.0002	0.00005	0.012	0.015
	B vs C	0.45	0.008	0.46	0.52

Table 13 Frequencies of different exon skippings in subjects with different IRF-1 haplotypes.

Eighteen subjects are divided into 3 groups according to their IRF-1 haplotypes concerning polymorphisms at 619, IRF-1 MS and 6516. Group A are homozygotes for alleles associated with HIV-1 resistance at all three loci (619AA+179179+6516GG), group B are heterozygotes and have only one cope of the protective allele for each locus (619AC+179^{+}179^{-}+6516GT) and group C are subjects entirely lack of resistant allele (619CC+179^{-}179^{-}+6516TT). Each group consists of 6 for subjects. IRF-1 specific RT-PCR amplicons for all subjects were TA cloned and 32 clones for each subjects were randomly selected for sequencing and exon skipping determination (so 192 clones in total were examined for each group). Numbers of clones with certain exon skipped in each group were counted and plotted. Chi-Square test (2×2) has been utilized for statistic analysis in comparing the frequencies of certain exon skipping in different subject groups and the *p* value are plotted as well. In summary, the result shows that subjects in group A have significantly higher frequencies of skipping exon 2 and 3, but lower frequencies of skipping exon 7 or exon 8 when compared to either of the other two groups individually or in combination. Significant difference is also detected in exon 3 skipping between group B and C, although it is likely due to the rareness of exon skipping in group B.

Figure 20 Alignment of amino acid sequences encoded by IRF-1 mRNA with exon 2, 3, 2 and 3, or 7 skipped

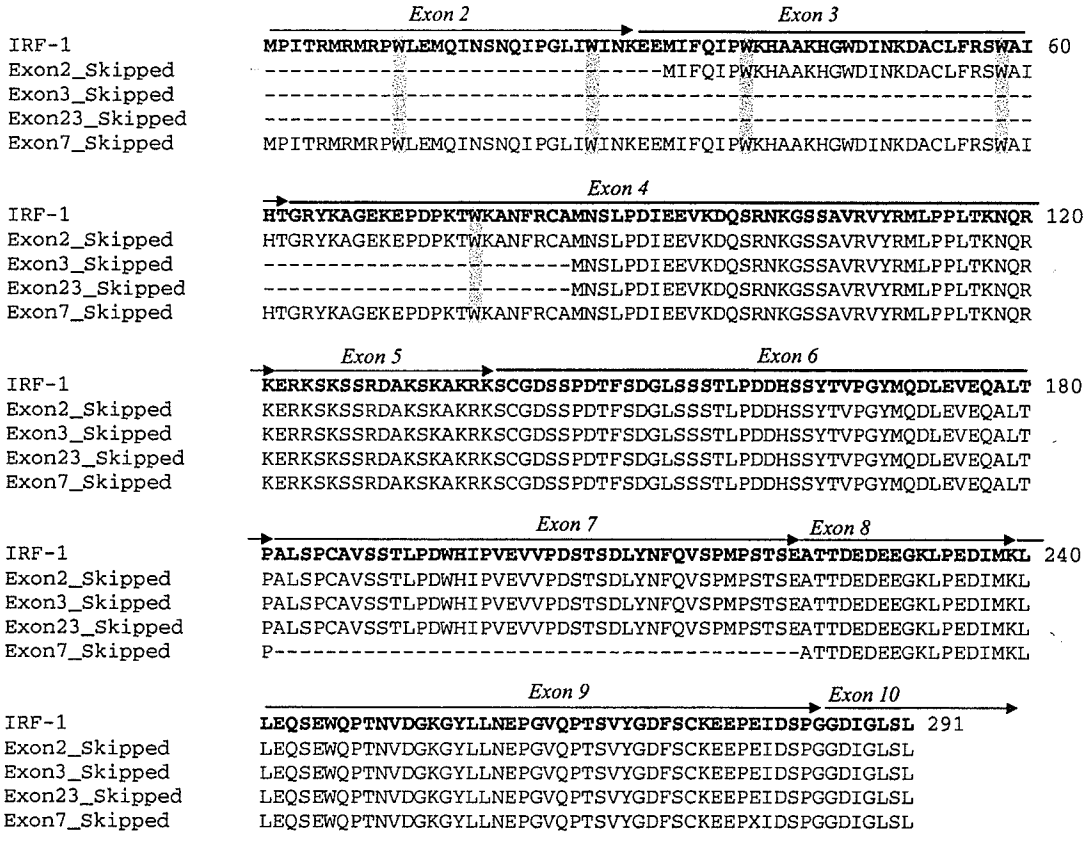


Figure 20 Alignment of amino acid sequences encoded by IRF-1 mRNA with exon 2, 3, 2 and 3, or 7 skipped

The alignment of reference IRF-1 amino acid sequence and sequences coded by IRF-1 mRNA with exon 2, 3, 2 and 3, or 7 skipped is depicted. The IRF-1 protein consists of 325 amino acids encoded by exon 2 to 10 of IRF-1 mRNA inclusively (GenBank NM-002198). The IRF-1 specific RT-PCR amplified a segment spanning nt 67 to 1072 of IRF-1 mRNA and covering coding sequence for aa 1 to 291. Corresponding sequence fragments coded by different exons are indicated. The residues forming the N-terminal "WWWWW" DNA binding motif located in the exon 2, 3 and 4 coded regions are highlighted. The amino acid sequences coded by mRNAs with exon 2, 3, 2 and 3, or 7 skipped were obtained by analyzing the corresponding nucleotide sequences using NCBI online Open Reading Frame Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). When exon 2 is skipped, the "ATG" start codon shift to the nt 7 ~ 9 of downstream exon 3. When exon 3 or both exon 2 and 3 are skipped, the start codon shifted further down to nt 66~68 in exon 4. None of these start codon shifts result in coding sequence changes downstream of new start sites in comparison to GenBank reference sequence. The figure shows that skipping exon 2, 3 or both all result in the loss of "WWWWW" motif in the coded protein products. Exon 3 skipping leads to same truncated IRF-1 products as the one with both exon 2 and 3 skipped. Exon 7 skipping results in a truncated protein exactly lacking the corresponding segment coded by exon 7.

*** " IRF-1" stands for the standard IRF-1 sequences numerated according to GenBank NM_002198.

is skipped, the open reading frame shifts occurs right after amino acid (aa) 222, ending up with a truncated 242 aa IRF-1 protein comparing with a 325 aa in full length IRF-1. When both exon 7 and 8 are skipped, the open reading frame is changed after aa 181 and the resulting truncated product has 201 aa. When Exon 9 is skipped, the open reading frame is changed after aa 239 and the coded truncated product has 243 aa in length. (**Figure 21**). These data suggest that intronic polymorphisms in IRF-1 genomic sequence result in significant splicing changes in IRF-1 mRNA and subsequent protein translation.

The efficiency of HIV-1 pseudovirus replication is distinct in PBMCs with different IRF-1 genotypes

Because multiple stimuli, including mitogens and some cytokines, can induce the up-regulation of IRF-1 expression and confound the detection of any effect of IRF-1 on HIV-1 infection, traditional HIV-1 culture techniques using PHA or IL-2 stimulated T cells would not be suitable for testing the impact of altered IRF-1 expression on HIV-1 replication. Additionally, HIV-1 can normally infect only activated CD4 bearing cells but not resting cells. These factors necessitate an alternative protocol to be established to study the effect of different IRF-1 genotypes on HIV-1 replication. A method that would allow us to infect unstimulated cells efficiently, require no pre-treatment that would affect IRF-1 expression and that mimics the natural process of HIV-1 infection is essential. Such a method would also need to be sensitive so that we could detect HIV-1 transcription early after infection, as IRF-1 function is believed to function at very early stage of infection and this may play a vital role in the establishment of HIV-1 infection.

Figure 21 IRF-1 open reading frame shifts after exon 8, 7 and 8, or exon 9 skipped

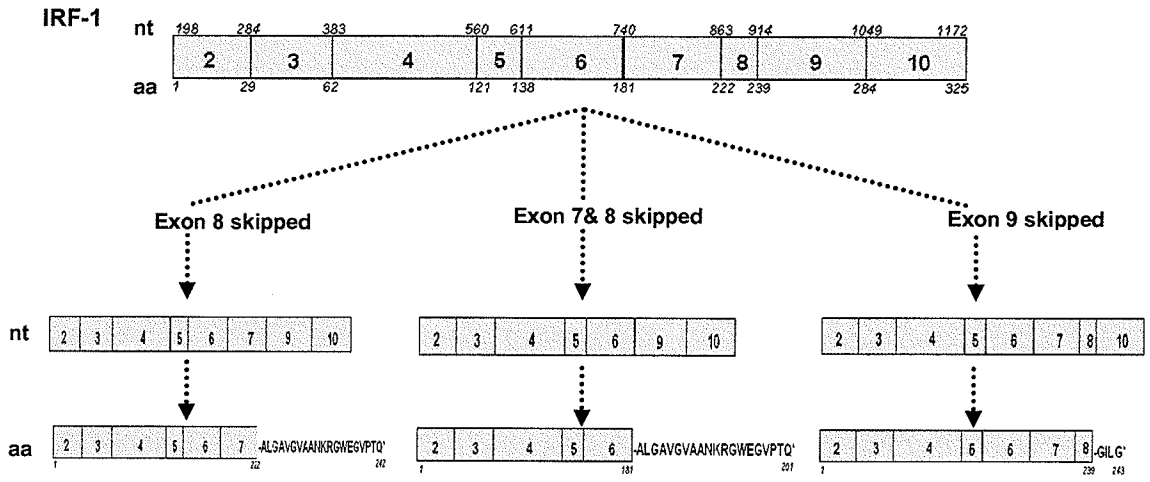


Figure 21 IRF-1 open reading frame shifts after exon 8, 7 and 8, or exon 9 skipped

The IRF-1 protein consists of 325 amino acids encoded by exon 2 to 10 of IRF-1 mRNA inclusively. The amino acid sequences with exon 8, 7 and 8, or 9 skipped were predicted by analyzing the corresponding nucleotide sequences using NCBI online Open Reading Frame Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The top panel shows the reference IRF-1 mRNA coding nucleotide and amino acid sequence frames numerated according to GenBank submission NM_002198. The bottom panel shows the predicted truncated IRF-1 products resulted from the open reading frame shifts due to mRNA exon skippings. When Exon 8 is skipped (**Left**), the open reading frame shifts occurs right after aa 222, ending up with a truncated 242 aa comparing with a 325 aa in full length IRF-1. When both exon 7 and 8 are skipped (**Middle**), the open reading frame is changed after aa 181 and the resulting truncated product has 201 aa. When Exon 9 is skipped (**Right**), the open reading frame is changed after aa 239 and the coded truncated product has 243 aa in length.

*** *nt: nucleotide ; aa: amino acid*

Dr. Yao constructed and gifted to us a HIV-1 pseudovirus which had a HIV-1 pNL4.3 backbone but expressed a VSV envelope protein (VSV-G) enabling its entry to all nucleated cells without a requirement for pre-activation. The Nef gene, which is located adjacent to the HIV-1 LTR region, of the parental HIV-1 strain was replaced by a firefly luciferase gene so that we could evaluate the efficiency of HIV-1 LTR transcription by measuring luciferase activity after the cells were infected with this pseudovirus (Ao et al., 2004; Yao et al., 1998; Yao et al., 1995). PBMC samples obtained from 13 HIV-1 negative subjects with distinct IRF-1 haplotypes were employed for this set of experiments (4 with "619AA+179179+6516GG" haplotype, 4 with 619AC+179^{-}179^{+} + 6516GT" and 5 with "619CC+179^{-}179^{-} +6516TT"). Since non-activated PBMCs cannot be maintained for a long time in cell culture and the role of IRF-1 in HIV-1 transcription appears to be critical at the very early stage of infection before overwhelming activity of viral Tat comes into play, we analyzed results at 0 hr, 24 hrs, 48 hrs, 96 hrs and 144 hrs post infection which is much earlier than traditional HIV-1 replication studies.

After infection with HIV-1 pseudovirus, the efficiency of HIV-1 transcription was shown to be distinct in PBMCs with different IRF-1 genotypes. As shown in **Figure 22,** the differences of HIV-1 transcription in different groups could be detected starting between 48 and 96 hrs post infection and became pronounced by 144 hrs. PBMCs from subjects homozygous for protective alleles at all three key loci (619AA+179179 +6516GG) had lower viral transcription throughout the experiment period, while subject heterozygous for all these three variations (619AC+ 179^{+}179^{-}+6516GT) had much higher level of transcription (Luciferase activity fold increase means: 373.3 vs 1740). PBMCs homozygous for the alternative alleles (619CC+ 179^{-}179^{-} +6516TT) had

Figure 22 Efficiency of HIV-1 transcription in PBMCs with different IRF-1 genotypes

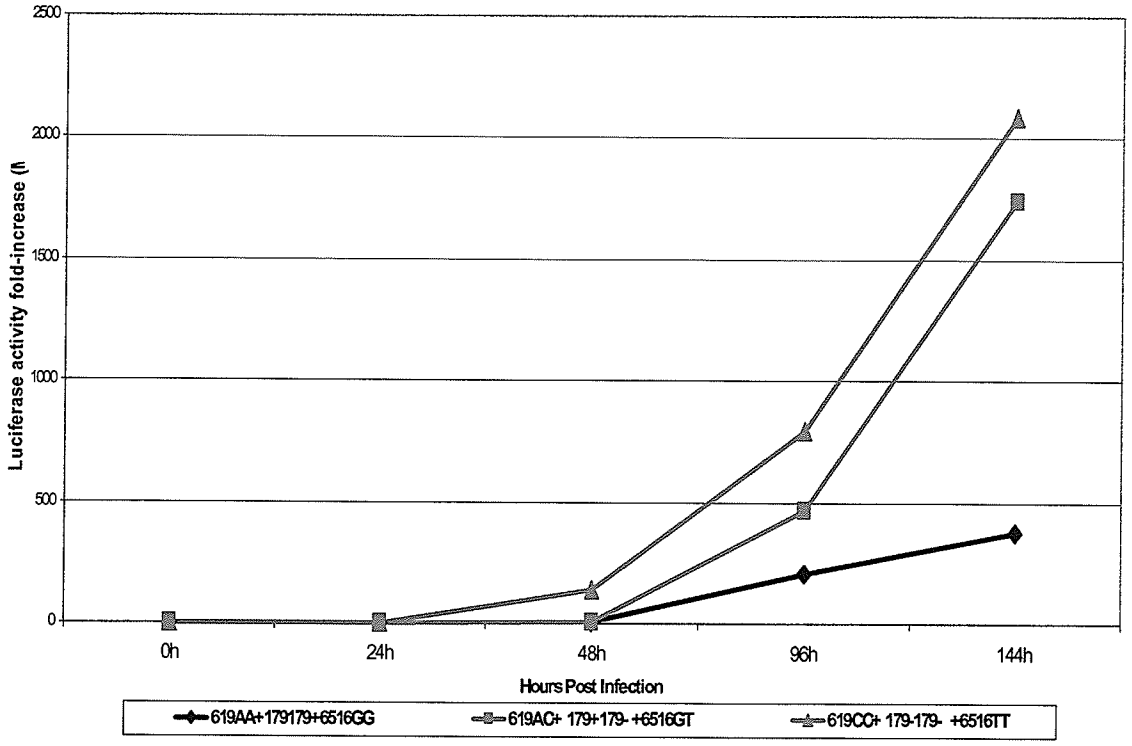


Figure 22 Efficiency of HIV-1 transcription in PBMCs with different IRF-1 genotypes.

PBMCs from subjects with different IRF-1 genotypes were infected with VSV-G pseudotyped HIV-1 virus and the viral transcription efficacy was determined by measuring the luciferase activity at certain time points post infection. The fold increases of luciferase activity (Mean) upon negative control (0 hr) were plotted here. This figure showed that differences in viral transcription after infection could be observed starting between 48 and 96 hrs post infection and becoming pronounced by 144 hrs. PBMCs with the best IRF-1 haplotype who were homozygous for protective alleles at all three loci (619AA+179179+6516GG) consistently had lower viral transcription (the group in blue line), while subject heterozygous for all these three variations (619AC+179^{+}179^{-} + 6516GT) had much higher level of transcription (the group in purple line). The luciferase activity fold increase means at 144 hrs post infection for these two groups are 373.3 vs 1740 respectively. PBMCs homozygous for the alternative alleles (619CC+ 179^{-}179^{-} +6516TT) had even higher viral transcription (the group in red line) and the mean luciferase activity fold increase at 144 hrs reaches 2080, which is much higher than that in PBMCs with the best haplotype and comparable to the heterozygotes. Nonparametric statistical analysis result comparing subjects with the most protective IRF-1 genotype (619AA+179179+6516GG) and those with alternative genotypes highly suggested that PBMCs with good IRF-1 genotype correlates with less HIV-1 LTR transcription at early HIV-1 infection (P=0.09).

even higher viral transcription and the mean luciferase activity fold reaches 2080 at 144 hrs, which was much higher than that in PBMCs with the best genotype and comparable to that in the heterozygotes. Nonparametric statistical analysis result in comparing subjects with the most protective IRF-1 genotype (619AA+179179+6516GG) and those with alternative genotypes highly suggested that PBMCs with good IRF-1 genotype correlates with less HIV-1 LTR transcription at early HIV-1 infection (P=0.09).

Varied IRF-1 responses to HIV-1 infection detected in PBMCs with different IRF-1 genotypes

It has been previously reported that HIV-1 infection can induce an IRF-1 response and IRF-1 transactivated HIV-1 LTR directed gene expression in a dose dependent manner (Sgarbanti et al., 2002). Hence we examined the IRF-1 response in PBMCs from HIV-1 negative subjects with different IRF-1 genotypes after infection with the HIV-1 pseudovirus. The result showed that HIV-1 infection induced IRF-1 response in all tested samples and it could be detected as early as 24 hrs after infection. Further analysis of IRF-1 expression in HIV-1 infected PBMCs revealed that cells from subjects with different IRF-1 genotypes showed distinct kinetics/patterns of IRF-1 responses. Representative results from three subjects with distinct IRF-1 genotypes were shown in **Figure 23.** All cells showed similar pattern of IRF-1 elevation within the first 24 to 48 hrs, but the upregulation of IRF-1 in cells with the protective IRF-1 genotype (179179+619AA+6516GG) peaked at 48 hrs post infection and remained at a moderate lower level throughout the experiment period, while it rose to much higher levels in cells

Figure 23 Varied IRF-1 responses to infection by HIV-1

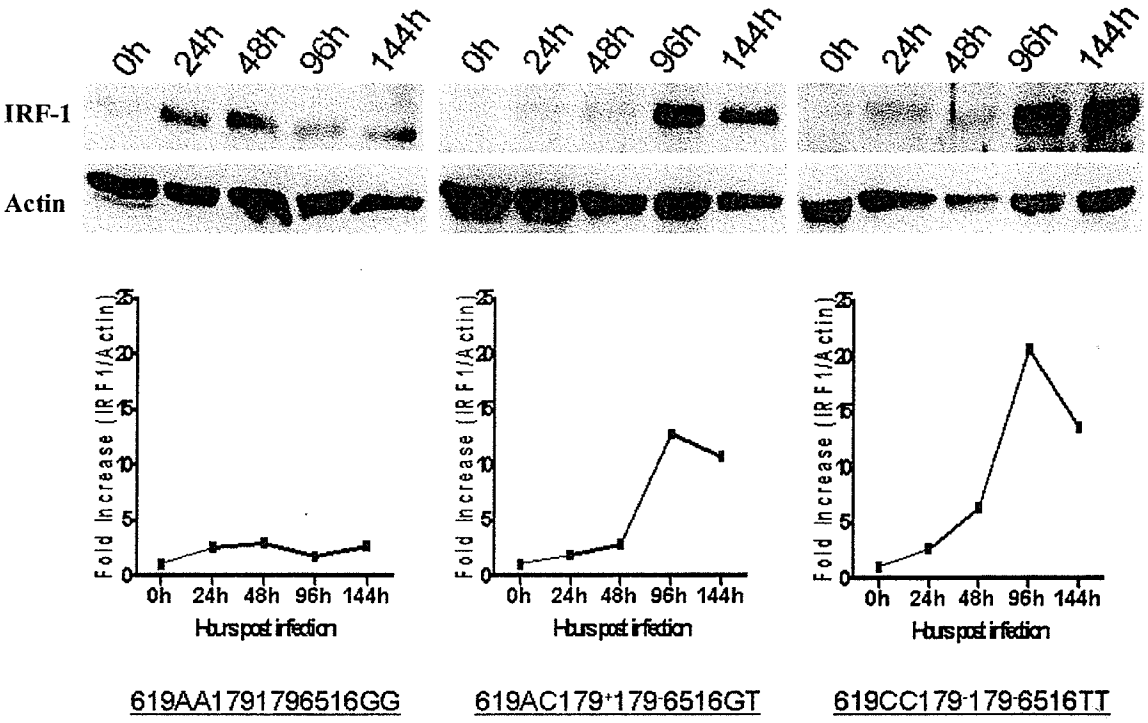


Figure 23 Varied IRF-1 responses to infection by HIV-1

PBMCs from 3 subjects with different IRF-1 genotypes from VSV-G pseudotyped HIV-1 infection experiment were included here and the IRF-1 response to HIV-1 infection was determined by measuring the elevation of IRF-1 protein expression at certain time points post infection. The fold increase of IRF-1/Actin ratio in spot densitometry was used as index for comparison. The top panel shows the Western Blot results for IRF-1 and action detection. The bottom panel shows the corresponding IRF-1/Actin ratio fold increase curves. The bottom line indicates the IRF-1 haplotypes of PBMCs tested. As shown in this figure, HIV-1 infection induced IRF-1 response in all tested samples and it could be detected as early as 24 hrs after infection. But PBMCs with different IRF-1 genotypes showed distinct patterns of IRF-1 responses. As shown in this figure, all cells showed similar pattern of IRF-1 elevation within the first 24 to 48 hrs, but the upregulation of IRF-1 peaked at 48 hrs post infection and remained at a moderate lower level in cells with the protective IRF-1 genotype (179179+619AA+6516GG) while it rose to much higher levels in cells with the other two alternative haplotypes (619AC + $179^{+}179^{-}$+ 6516TT and 619CC +$179^{-}179^{-}$ + 6516TT), both of which have an IRF-1 response peaked at 96 hrs post infection.

with the other two alternative haplotypes (619AC + 179+179-+ 6516TT and 619CC +179-179- + 6516TT), both of which have an IRF-1 response peaked at 96 hrs post infection. These results correspond precisely to the luciferase assay results in which all PBMCs with the best IRF-1 genotype had a continuous low level HIV-1 LTR transcription while the other two groups showed continuous increase of LTR transcription throughout the experimental period (**Figure 22**). Reduced IRF-1 responsiveness to HIV-1 infection further confirmed our finding from IRF-γ stimulation experiments indicating that PBMCs with protective IRF-1 genotypes exhibit lowered IRF-1 responsiveness to exogenous stimulation. These results showed that not only did PBMCs from subjects with protective IRF-1 genotypes support decreased pseudo-HIV replication, they also exhibited depressed IRF-1 responses to HIV infection.

Discussion

Immunological and genetic correlates of HIV-1 resistance in the Kenyan female sex worker cohort

The phenomenon of natural resistance to HIV-1 infection provides an excellent model to explore mechanisms that may be protecting the host from HIV-1 infection. One of the best characterized groups of highly exposed, uninfected individuals is the well described commercial sex worker cohort from Pumwani district of Nairobi, Kenya, of which some individuals can be epidemiologically defined as resistant to infection by HIV-1 (Fowke et al., 1996; Plummer et al., 1999). Following our strict criteria (HIV-1 negative for ≥ 3yrs while actively continuing sex work and remaining in a good health), we defined a portion of subjects in this cohort (approximately 5% in total) who appeared to be naturally resistant to HIV-1. HIV-1 prevalence in adults in Nairobi is 9.9% (11.9% in women and 7.8% in men) (Ministry of Health, 2005) and it was conservatively estimated that these female sex workers had at least 60 high-risk exposures/year to various HIV-1 strains/clades through sex with multiple partners (Fowke et al., 1996). Observational studies showed that approximately 65% of the enrollees were infected when they were recruited into the cohort while the majority of HIV-1 uninfected women seroconverted soon after follow-up. When evaluating the risk of seroconversion over time in this cohort, it was noted that the survival curve leveled off after around 3 yrs (Fowke et al., 1996). Those who were not infected within the first 3 yrs have a 10-fold reduction in risk of infection at later time points. Thus, we can be fairly confident that subjects who could

remain HIV-1 uninfected for at least 3 yrs are relatively resistant to HIV-1 infection. This is the rationale for the selection criteria for HIV-1 resistance subjects. Resistance to HIV-1 in these sex workers raises important research questions and provides an invaluable resource in trying to understand the underlying mechanisms protecting these subjects from contracting an infection despite of hundreds of exposures to HIV-1.

A series of studies had been conducted in this cohort. Results from these studies showed that the resistance to HIV-1 in these subjects was not due to insufficient exposure, differences in sexual behavior, condom use, resistance to other sexually transmitted infections, suggesting a biological basis for this resistance (Plummer et al., 1999) (Fowke et al., 1996). This resistance is not associated with plasma MIP-1α, MIP-1β, or RANTES level, nor with the secretion of these chemokines by PBMCs or altered CCR5 or CXCR4 expression(Fowke et al., 1998). It is also not related to plasma neutralizing IgG antibodies, allo-antibodies to HLA class I, or suppressive factors in plasma (Plummer et al., 1999). Cells from HIV-1 resistant subjects exhibited similar susceptibility to HIV-1 infection in traditional *in vitro* studies suggesting that innate cellular resistance to HIV-1 infection is not the explanation either (Fowke et al., 1998; Plummer et al., 1999) . These subjects do not have any known chemokine-receptor polymorphisms shown to be important for host susceptibility to HIV-1, such as CCRΔ32(Plummer et al., 1999). Further studies revealed that systemic HIV-1-specific T-helper responses (Alimonti et al., 2005; Fowke et al., 2000) as well as systemic and mucosal cytotoxic T lymphocyte responses were potentially involved in this process (Fowke et al., 2000; Rowland-Jones et al., 1998) although CTL responses were not prospectively associated with protection in a multivariate analysis that included HIV-1 exposure and duration of sex work (Kaul et al.,

2004). However, there is some suggestion that the "type" of immune response may be important as HIV resistant subjects showed significantly lower levels of CD4+ specific immune activation and apoptosis compared with those in the HIV infected controls (Alimonti et al., 2005). HIV-resistant individuals showed stronger proliferative rather than cytokine responses to HIV-1 P24 antigens, suggesting that altered types of immune responses may be present (Alimonti et al., 2005). A global hyporesponsive interleukin-4 (IL-4) response to HIV-1 and other antigens were also shown to be associated with this resistance (Trivedi et al., 2001). HIV-1-resistant individuals also had increased cervical $CD4^+$ and $CD8^+$ T cell counts in comparison with the HIV-1-uninfected CSWs but these increases were not reflected in the systemic lymphocyte compartment (Iqbal et al., 2005). These data suggest that these individuals may exhibit a cellular (type-1) bias in cellular immune responses to HIV-1.Although no HIV-1 specific IgG can be detected in either plasma or genital secretions in HIV-resistant subjects, HIV-1 specific neutralizing mucosal IgA in the genital tract could be detected in most of them (Devito et al., 2002; Devito et al., 2000; Kaul et al., 2001b). HIV-1-resistant women also have a 10-fold increase in RANTES expression in vaginal secretion, compared with the HIV-1 – uninfected CSWs (Iqbal et al., 2005), suggesting that innate immunity also plays a role in HIV-1 resistance. All together, data from studies in this specific cohort clearly indicate that the HIV-1 resistant subjects have detectable innate and adaptive immune response to HIV-1 in absence of HIV-1 infection, which suggests that they had been exposed to but not infected by HIV-1 likely due to their distinct immune responses to HIV-1 exposure and/or other unknown factors/mechanisms. HIV resistance seems to be a multifactorial phenomenon with a variety of immune factors involved although the underlying

interaction among them has not been fully understood.

It is well-known that in the majority of infectious diseases only a portion of population exposed to the particular pathogen become eventually infected, indicating the existence of natural and intrinsic resistance to infectious pathogens. Genetic variation is an evolutionary strategy utilized by all organisms and likely explains a good proportion of altered susceptibility. In humans, the best available strategies to identify the role of genetic factors in certain phenotypes are genetic epidemiologic studies. Many genetic variations have been associated with altered susceptibility to infections including HIV-1 and other immune-related diseases (Bidwell et al., 1999; Bidwell et al., 2001; Haukim et al., 2002; Hill, 1998; Hill, 2001; Julg and Goebel, 2005).

A familial clustering of the HIV-1 resistant phenotype was identified in this cohort (Plummer et al., 1999), further suggesting that a unique genetic composition might be a potential determinant of the altered susceptibility to HIV-1 in this population. Previous genetic studies have shown that HLA-A2/6802 supertype and DRB1*01 are associated with the HIV-1 resistant phenotype while HLA-A2301 is strongly correlated with increased susceptibility to infection further confirming a role for genetic composition in determining the host susceptibility to HIV-1 infection in this population (MacDonald et al., 2000) . The association between HLA class I and II genotypes and HIV-1 resistance would be explained if they selectively present peptide epitopes from highly conserved regions of HIV-1(Rowland-Jones et al., 1998). Remarkably, we noticed that the clustering of HIV-1 resistance in families in Pumwani cohort was independent of MHC associations and known chemokine receptor polymorphisms, suggesting that additional genetic factors are involved. The other genetic factor perhaps works in concert with the

protective MHC alleles and perhaps is involved in the generation of type 1 cellular immune responses (Plummer et al., 1999).

Combining the currently available data from this cohort, we conclude that resistance to HIV-1 in this cohort is related to an altered host immune response to HIV-1, but not solely dependent on it. Both published and unpublished data derived from this cohort suggest that low immune activation is universal in the HIV-1 resistant individuals and that a Th1 biased immune response contribute greatly to the resistance to HIV-1. Resistance is clearly intrinsic and inherent and HLA polymorphism may be involved in this process but is not the sole genetic factor involved. The detection of HIV-1 specific immune response suggests that their immune systems must have encountered HIV-1 but eliminated the infection before it had a chance to be well-established. The disruption of HIV-1 infection may occur either before (via secretary IgA and RANTES in the mucosal secretion) or after the viral entry (via CTL response and/or other uncharacterized mechanisms). Factor(s) involved in both unique genetic composition and host immunity may be the main player(s) dominating the HIV-1 resistance in this small portion of population.

IRF-1, a double-edged sword in HIV/AIDS

IRF-1 is located in the human IL-4 gene cluster which contains genes with important immune regulatory functions especially in the regulation of Th1/2 immune responses, such as interleukin 3 (IL-3), IL-4, IL-5 and IL-13. Polymorphisms in this region have been associated with differential susceptibility to a number of infections and pathologic

conditions including schistosomiasis and allergy/atopy (Bidwell et al., 1999), pathogenesis of which appears to be dependent on the regulation of the types of T-cell response. If a robust cell mediated immune response was important for protection against HIV-1 infection, it is reasonable to hypothesize that genetic mechanism for resistance to HIV-1 might be identified in the IL-4 gene cluster. Previous findings about the association between IRF-1 MS polymorphism and HIV-1 resistance gave us an important clue that immunoregulatory gene(s) could also be involved in the determination of host susceptibility to HIV-1 and that IRF-1 is likely an important factor. The potential role of IRF-1 as a candidate gene to explain, in part, resistance to infection by HIV-1 was supported by previous data demonstrating that IRF-1 plays a key role in host innate and adaptive immunity (Kroger et al., 2002; Mamane et al., 1999). By binding to an interferon response sequence element (ISRE) in the promoter regions of the target genes, IRF-1 can regulate the expression of multiples genes with important functions in host immunity, especially in innate immunity and the generation of efficient Th1 response (Kroger et al., 2002). Remarkably, IRF-1 can also regulate HIV-1 transcription and replication (Battistini et al., 2002; Sgarbanti et al., 2002). HIV-1 infection can induce IRF-1 response in infected CD4 cells and the increased IRF-1 mRNA level can be detected at as early as 3 hrs after infection and the increase of IRF-1 protein expression can be detected within 24 hrs. This is much faster than the key HIV-1 viral transcriptional regulatory protein, Tat, which is detectable at 24 hrs for mRNA and 48 hrs for protein respectively (Marsili et al., 2003; Sgarbanti et al., 2002). Thus, IRF-1 responses to HIV-1 infection may occur prior to Tat at the earliest stages of infection and may be the primary initiator of HIV-1 LTR transcription. At later stages, IRF-1 is thought to also

bind to Tat and synergize with Tat in activation of HIV-1 transcription. In the presence of low doses of Tat, IRF-1 increases tat-mediated HIV-1 transcription by direct physical interaction (Sgarbanti et al., 2002).

These data suggested that IRF-1 is a factor involved in both host antiviral immunity and HIV-1 replication. It promotes the host IFN responses and also induces Th1-like immune responses in combating viral infections. Meanwhile, as an important host transcription regulator, it appears to be actively involved in the intricate interplay between host and virus at HIV-1 transcription regulation level, mainly by promoting HIV-1 replication through HIV-1 LTR activation. All these suggest that IRF-1 has a contradictive dual role to play in HIV/AIDS. Hence we hypothesized that IRF-1 was a potential determinant of the natural resistance to HIV-1 infection.

IRF-1 polymorphisms in Kenyan subject cohort

Although the IRF-1 MS marker has been correlated with altered susceptibility to HIV-1, it is noteworthy that IRF-1 MS marker lies within an intronic region and has no clear functional role. Hence we assumed that the MS marker might be in linkage disequilibrium with other functional IRF-1 mutation(s) responsible for this correlation. The human IRF-1 gene and its immediate upstream promoter sequence had been assigned to 5q31.1 (Harada et al., 1994b; Itoh et al., 1991). The Genbank reference sequence for IRF-1 (L05072) were established in 1992 based on a placenta cloning from an Asian patient and was only 7721 bp in length (Cha et al., 1992). The IRF-1 promoter sequence was described in 1994 (X53095) with a length of 669 bp in which 174 bp were

overlapping with the nt 1~174 in IRF-1 gene (L05072) (Harada et al., 1994b). To identify all existing polymorphisms in IRF-1, we conducted near-full length sequencing analysis of the IRF-1 gene and its promoter from a subset of 507 individuals from our cohort.

In our population-based sequencing study, we discovered that Kenyan population displayed extensive diversity throughout the IRF-1 locus. Besides the IRF-1 MS marker, 53 SNPs, 2 insertions and one 16 nucleotide deletion were identified in IRF-1 and its immediate promoter region in this subject cohort, although some previously reported polymorphisms were not detected at all. These findings indicated a high degree of polymorphism in IRF-1 gene in the Kenyan population. When we examined the IRF-1 polymorphisms in our Manitoban local donors of non-African genetic background, we discovered that most of these polymorphisms could be detected in other ethnic groups as well although the allele frequencies varied.

In addition to these polymorphisms, we identified 35 consistent discrepancies between IRF-1 sequence obtained from our subjects and the GenBank reference sequences. These discrepancies were confirmed in our local donor population, indicating that they were not caused by an artifact of poor-proof reading in PCR reactions but rather inaccuracy of the current GenBank references. The accuracy of the reference sequence is critical for further downstream genetic and functional studies. We conducted complete full length IRF-1 sequencing from a single subject in our study. The new established IRF-1 sequence contained multiple differences in the nucleotide composition and was 7748 nucleotides in length instead of 7721 in L05072. Compared to X53095, the new IRF-1 promoter sequence had one nucleotide change in it. These two sequences were apparently more accurate and representative of the general populations. The combination of these

two sequences has been submitted to GenBank (accession number: DQ789232) and will serve as a more representative IRF-1 reference for future studies.

It is notable that, except for two silent mutations in exon 7, all of the other IRF-1 mutations occurred in the non-coding regions, either in the intronic region or in the promoter region. Although it is not surprising, it was contrary to our original assumption that functional coding changes might reside in the IRF-1 gene which were in linkage disequilibrium with the IRF-1 MS marker and responsible for the positive association between IRF-1 MS polymorphism and resistance to HIV-1 infection. However, there is a growing body of evidence showing that non-coding sequences also have functional consequences. We believed that the functional IRF-1 polymorphisms might be found in these non-coding polymorphisms. Polymorphisms in the IRF-1 promoter region may affect the transcriptional activation of the gene and its subsequent expression (Saito et al., 2001). Genetic variations in non-coding regions of cytokine genes have been correlated with multiple disease conditions (Bidwell et al., 1999; Bidwell et al., 2001; Haukim et al., 2002). All these suggested that we should further explore the effects of these existing promoter and intronic IRF-1 polymorphisms on IRF-1 expression and function.

Correlation of IRF-1 polymorphisms and HIV-1 resistance

Using the latest epidemiological data from our cohort of sex workers, we were able to confirm our previous findings. A single MS allele (IRF-1 179) showed significant association with HIV-1 resistance (P=0.0054) and also provided a protective effect in HIV-1-free survival analysis evaluating the risk of seroconversion over time (P=0.0396).

Although our comparative sequencing study of the entire IRF-1 gene in 87 HIV-1 resistant and over 300 susceptible women failed to identify any polymorphisms with obvious functional consequences, two additional SNPs (619A/C and 6516G/T) were found to be associated with resistance to HIV-1 infection (P=0.00073 and 0.030 respectively). The 619A and 6519G alleles at these two loci were significantly associated with HIV-1 resistance phenotype and reduced likelihood of infection and longer HIV-1 free survival while the alternative alleles (619C and 6516T) were associated with increased susceptibility to HIV-1. This observation coincides with the phenomenon of family clustering of HIV-1 resistant individuals as observed in our subject cohort.

Since linkage disequilibrium was found to be quite common in the IRF-1 gene and all protective alleles (179 at IRF-1 MS, 619A and 6516G) tend to co-segregate, we also conducted haplotype analysis on the association between different IRF-1 haplotypes and resistance to HIV-1 infection. The combinations of protective alleles at different loci generated an even stronger correlation with HIV-1 resistance. Multivariate analysis revealed that 619A/C was the only independent and significant indicator of HIV-1 resistance while the associations between the other two IRF-1 polymorphisms and HIV-1 resistance were potentially due to their correlation with 619A/C. In spite of this, the stronger association between combinations of protective alleles in IRF-1 with HIV-1 resistance suggests that these alleles synergize and provide additive protection against HIV infection.

The associations between IRF-1 genotypes and resistance to HIV-1 were revealed to be all independent of the HLA genotypes, such as HLA-DRB1*01 and HLA A2/6802 supertype (MacDonald et al., 2000), previously shown to correlate with HIV-1 resistance

in this cohort (Data not shown). This indicates that protective IRF-1 genotypes mediate inherent resistance to HIV-1 via a distinct mechanism in which a particular HLA genotype is not essential. Hence IRF-1 polymorphism may partly explain the phenomenon of HLA independent family clustering of HIV resistance.

IRF-1 polymorphism was not related with HIV-1 disease progression

Since polymorphisms in IRF-1 seem to be playing a role in protecting the host from being infected by HIV-1, it would be rational to assume that these protective alleles might also important in limiting disease progression in HIV-1 infected patients. We conducted Kaplan-Meier survival analysis comparing the durations between the first positive day in the cohort and first day of CD4+ cell count decrease to <400 cells/μl or 200 cells/μl in subjects with different IRF-1 genotypes. CD4+ cell counts is routinely taken as index for HIV disease progression and CD4+ cell count drop to below 400 cells/μl is redeemed as the later stage of latent HIV-1 infection (stage 5) and CD4+ cell count below 200 cells/μl is an indicator of disease progression to clinical AIDS (stage 6). This survival analysis failed to show any protective effect and none of the alleles protective against infection correlated with HIV-1 disease progression except that, with no clear explanation, subjects heterozygous for IRF-1 MS (with one 179 allele) appeared to progress faster to CD4+ counts below 200 cells/μl. These data suggest that the biologic effects due to the altered IRF-1 genotypes function in early events of HIV-1 replication within the host, perhaps by blocking viral entry or disrupting the early viral transcription, key steps in the establishment of HIV-1 infection, but not in limiting HIV infection in the

presence of an already established infection. IRF-1 is a transcription regulator and resides only in the cell nucleus. This means that IRF-1 cannot function directly on intact HIV-1 and is not likely involved in blocking the HIV-1 entry into the target cells, although HIV-1 inhibition may be mediated by other IRF-1 regulated downstream factors. On the other hand, IRF-1 is a potential key initiator of HIV-1 transcription. By binding to the ISRE-like sequence in the HIV-1 5'-LTR (spanning nt +200~+217), IRF-1 can initiate HIV-1 transcription in a dose-dependent fashion independently of and prior to Tat, a key transcription activator of HIV-1. These findings suggest a key role for IRF-1 in the early stage of HIV-1 replication when viral transactivators are absent or present at very low levels(Sgarbanti et al., 2002). It may be essential to the establishment of an HIV-1 infection. This supports the hypothesis that the association between IRF-1 polymorphisms and HIV-1 resistance might be mediated by the disruption of HIV-1 establishment at the early stage. The lack of association with disease progression may be explained by the overwhelming role of Tat in HIV-1 replication after HIV-1 infection is well established, when Tat is sufficiently present and effect of IRF-1 is relatively negligible.

Theoretically, in a population exposed to a lethal infectious agent, genetic factors involved in resistance to infection should accumulate over time. HIV-1 is undoubtedly putting selection pressure on the African population. Thus it was of interest to conduct an epidemiological survey to investigate how the frequencies of protective IRF-1 alleles changed over time. Surprisingly, the protective IRF-1 allele (179 at IRF-1 MS, 619A and 6516G) frequencies all significantly decreased over time when we compared the frequencies in subjects enrolled in our cohort before or after 1993. These results are

contradictory to our unpublished findings in an independent study conducted in the same cohort, which shows that protective HLA allele frequencies significantly increased over time (Unpublished data from Drs. Ma Luo and Frank Plummer). This discrepant finding helps to confirm the independence of IRF-1 and HLA genotypes in the context of HIV/AIDS. But it does not fit our hypothesis as we expected the protective IRF-1 allele frequencies increase as well. The fact that these protective allele frequencies decreased suggests that other mechanisms besides HIV-1 may be at work. Multiple mechanisms may be responsible for this "abnormal" phenomenon which include that there might be other unknown selection pressure(s) on IRF-1 gene evolution within this population or geographic area. Significant selection pressures on IRF-1 gene polymorphism might be exerted on this population by other infectious diseases that are highly prevalent in Sub-Saharan Africa, such as tuberculosis and malaria. When we compared the allele frequencies between Kenyan subjects and our local donors with a diversity of non-African genetic backgrounds, we noticed that the protective IRF-1 allele frequencies at all these three loci were significantly lower while the alternative "susceptible" alleles were much higher in the general Kenyan population. It is pure speculation that deserves more study but the increase of "susceptible" IRF-1 alleles may contribute to the severe and persistent HIV-1 endemic in Africa to some extent.

Low IRF-1 expression and responsiveness may contribute to resistance to HIV-1 infection

Although the IRF-1 polymorphisms correlating with HIV-1 resistance were located in

the non-coding regions, our current data clearly suggests that they are important in the determination of host susceptibility to HIV-1 infection. The obvious question then is how do these polymorphisms affect IRF-1 transcription? The most direct way to answer this question is to measure levels of IRF-1 protein directly. IRF-1 levels in PBMCs from subjects with protective IRF-1 genotypes were significantly lower both at baseline and in response to exogenous IFN-γ stimulation. We noted in the genetic analyses, alleles that associated with resistance acted in a dominant manner (i.e. one or more copies of the 179, 619A, and 6516G alleles provided a protective effect), while the differences in protein expression were most significant between the extreme genotypes (i.e. 619AA vs. 619CC). This is likely due to the complex nature of IRF-1 regulation, and the likelihood that this process is a complex event and thus only the most divergent genotypes exhibit detectable functional differences. Differential IRF-1 responses were also confirmed in PBMCs with different IRF-1 haplotypes after HIV-1 infection. That the protective genotypes associated with a lower IRF-1 protein expression was somewhat counter-intuitive, as one would think that this would result in lower interferon responses and perhaps a blunted cellular response. Yet, HIV-1 resistant women have cell mediated immune responses to HIV-1. While these responses are associated with HIV-1 resistance, it is entirely possible that these are a result of exposure to HIV-1, and not in fact, mediating protection. Cellular responses in HIV-1 resistant women tend to be weaker than those found in HIV infected patients (Kaul et al., 2001c), and did not clearly associate with protection in a prospective study (Kaul et al., 2004). Alternately, these responses may indeed be protective, and their induction is not solely dependent on IRF-1 expression or a low IRF-1 response is sufficient enough for the development of these immune responses.

When we examined the IRF-1 response to infection by the pseudotyped HIV-1, we were able to further confirm that HIV-1 infection induced elevated IRF-1 protein expression which could be detected as early as 24 hrs post infection. Although all PBMCs with different IRF-1 genotypes showed a similar level of IRF-1 upregulation prior to 48 hrs, distinct patterns were observed after 48 hrs post infection. PBMCs from subjects with protective IRF-1 haplotype (619AA+179179+6516GG) retained a low level of IRF-1 response while cells with the other two alternative haplotypes (619AC + $179^{+}179^{-}$ + 6516GT or 619CC + 179-179- + 6516TT) showed dramatic IRF-1 upregulation peaked at 96 hrs post infection.

The function of IRF-1 in HIV-1 infection appears to be a double-edged sword: it is a pivotal regulator for cell-mediated immune responses, but also appears to be critical in the establishment of HIV-1 replication. The outcome of a successful infection is likely dependent upon the counterbalance between these two competing effects. We hypothesize that HIV-1 resistant women, with a genetic predisposition to lowered IRF-1 expression, may be able to limit the initiation of HIV-1 replication, which may lead to the disruption of HIV-1 infection. Effective cell-mediated and mucosal antibody immune responses, regulated by the IRF-1 gene, would thus be able to develop in the widened window period to eliminate replicating virus. The subsequent immune memory to HIV-1 would then be able to further protect the host from later HIV-1 infection after re-exposure.

One recent report suggested that HIV-1 exposed uninfected people had higher percentages of naïve (CD45RO+CD27+) CD4 and CD8 T cells and lower percentages of activated (HLA-DR+CD38+CD70+) CD4 and proliferating (Ki67) CD4 and CD8 T cells and whole blood cultures from these subjects showed lower lymphoproliferative response

than healthy controls. The authors suggested that low levels of immune activation and low T cell responsiveness may contribute to low susceptibility (Koning et al., 2005). Other studies conducted in the Pumwani cohort suggest that the HIV-resistant population has a general profile of decreased T cell activation and a suppressed immune response to HIV-1 as well as other non-specific antigens (personal communication, Blake Ball and Paul McLaren). We believe that this low immune activation is important to HIV resistance and protective from infection because cell over-activation leads to: 1) more co-receptors expressed on the cell surface facilitating the binding of HIV-1; 2) more activation induced cell death leading to accelerated depletion of CD4 repertoire including memory T cells ; 3) activation of HIV-1 transcription activators, such as NF-κB, which could synergize with Tat in amplifying HIV-1 replication ; 4) activate other CD4 expressing cells (Macrophage and DCs) and make them susceptible to HIV-1. Due to the key role of IRF-1 in immunoregulation, the likely consequence of lowered IRF-1 expression might contribute to the generation and maintenance of lower immunoactivation.

IRF-1 genotypes may be associated with altered HIV-1 transcription at early stage of infection

How then, may lowered IRF-1 expression and IRF-1 responsiveness result in resistance to infection by HIV-1? Reconciliation may lie with the potential role of IRF-1 in modulating HIV-1 replication. It is reasonable to assume that lowered baseline IRF-1 expression and IRF-1 responsiveness to HIV-1 infection may suppress / postpone the

onset of HIV-1 transcription and reduce viral replication at the early stage. This could lead to a prolonged window period and reduced viral replication allowing host immune responses to develop. Indirect supporting evidence comes from non-human primate studies in which subinfectious doses of SIV intravaginal challenge lead to SIV-specific T cell responses and protection from subsequent intrarectal challenge (Clerici et al., 1994a). Non-symptomatic exposure to the less virulent SIV/HIV chimera elicit a transient viremia that evokes humoral and cellular immune responses to HIV and SIV antigens virus and protects the host from challenging with more virulent strains (Miller et al., 1997). Early control, or reduction of HIV replication prior to uncontrolled viral replication and subsequent dissemination throughout the body has been postulated as a key phase in blocking HIV infection (Haase, 2005). The window period of maximum vulnerability for the virus in which host immunity or other interventions could prevent or control the infection normally ranges from 1 day to 1 week (Haase, 2005; Miller et al., 2005). There is no doubt that longer window period with repressed viral replication allows the host some precious time to develop strategies in disrupting viral infection.

To test our hypothesis that distinct IRF-1 genotypes and altered subsequent IRF-1 expression play a vital role in the initiation of HIV-1 transcription, we conducted HIV-1 infection experiments in PBMCs from subjects with different IRF-1 genotypes using a one cycle infection system with VSV-G pseudotyped HIV-1. This one-cycle infection virus infects all cell types due to the possession of VSV-G envelope protein and provides a highly sensitive way in the detection of HIV-1 LTR transcription especially at the earliest stages. After infection with this pseudovirus, drastic difference in the efficiency of HIV-1 transcription was observed in PBMCs from subjects with different IRF-1

genotypes early after infection. We could detect the differences of HIV-1 transcription in different groups from after 48 hrs post infection, supporting the previous findings that HIV-1 induced accumulation of IRF-1 mRNA expression in primary CD4+ T cells peaked between 24 and 48 hrs post infection (Sgarbanti et al., 2002). At 144 hrs after infection, we could note a clear trend that PBMCs with the most protective IRF-1 haplotype (619AA+179179+6516GG) consistently had lower viral transcription, while subjects with haplotypes "619AC + $179^{+}179^{-}$ + 6516GT" or "619CC + 179-179- + 6516TT" had much higher level of transcription. One subject with "619CC+ 179-179- + 6516TT" haplotype showed a unique curve of transcription with significant earlier onset of transcription and also the second highest transcription at 144h of all tested subjects. Further investigation revealed that this patient seroconverted soon after the PBMC sampling date which suggested a correlation between earlier initiation of HIV-1 transcription and increased susceptibility to HIV-1. Regardless of the importance of this anecdotal evidence, these results are highly suggestive that protective IRF-1 genotypes are associated with a lower/delayed onset of HIV-1 transcription. The kinetics of IRF-1 response to pseudotyped HIV-1 infection coincided with the observation in HIV-1 LTR transcription detected by luciferase assay. This suggests that altered IRF-1 response is likely determining the magnitude of HIV-1 transcription at the initial stage of infection and mediating the association between IRF-1 genotypes and resistance phenotype to HIV-1 infection. Since it is difficult to sustain unstimulated and infected cells longer than 1 week, we couldn't compare the differences at later time points. But we speculate that the differences will be minimized after Tat, the HIV-1 viral transactivator, dominates the HIV-1 transcription.

Intronic IRF-1 polymorphisms resulted in altered IRF-1 splicing patterns

The obvious next step will be to try to determine what the underlying mechanisms are to link the intronic polymorphisms, lower IRF-1 protein expression and the reduced rate of HIV-1 transcription? Although IRF-1 protein expression in subjects with the protective genotypes was significantly altered, it is notable that the polymorphisms associated with HIV-1 resistance were located in non-coding regions. How do these mutations alter the expression of IRF-1? Polymorphisms in the IRF-1 promoter region may affect the transcriptional activation of the gene and its subsequent expression (Saito et al., 2001). We identified 11 promoter polymorphisms in target population, 2 of which were located in putative Sp1 binding sites. None of these polymorphisms showed significant correlation with HIV-1 resistance, including the SNP at -300 (G/A) which was shown to be correlated with altered cellular immunity to HCV infection (Saito et al., 2002). The three key variations showing positive association with HIV-1 resistance were all located in the intronic regions (619A/C in intron 1, IRF-1 MS in intron 7 and 6516 G/T in intron 9). How do they function?

Introns and other non-coding regions in the genome have often been regarded as "junk DNA" with little or no apparent function. But a growing body of evidence is showing that non-coding regions have significant function in gene expression, and the "junk" is not considered as "junk" anymore (Herbert, 1996; Mattick, 1994; Moore, 1996). Genetic variations in non-coding regions of cytokine genes have been correlated with multiple disease conditions (Bidwell et al., 1999; Bidwell et al., 2001; Haukim et al.,

2002). A SNP at 7311 in the IRF-1 3' UTR was shown to be correlated with occurrence of juvenile idiopathic arthritis (Donn et al., 2001). Also, allele distributions of IRF-1 MS ("GT" repeats in intron 7) had been shown to be associated with the onset time of childhood atopic asthma (Nakao et al., 2001). So it is entirely possible that these intronic polymorphisms do have functional consequences on IRF-1 protein expression. Introns are capable of affecting gene function by influencing transcription, mRNA splicing, translation, mRNA stability, or promoter activity (Chabot, 1996; Herbert, 1996; Mattick, 1994; Moore, 1996). Intronic variations in IRF-1 gene showed strong correlation with HIV-1 resistance phenotype and were clearly associated with reduced protein levels, suggesting they might be the functional polymorphisms acting through these mechanisms.

IRF-1 mRNA contains 10 exons in which exon 1 and part of exon 10 comprise the untranslated terminal regions (**Figure 24**). The start codon is located at the beginning of exon 2. Except for exon 1 which is untranslated terminal sequence, IRF-1 mRNA has 9 coding exons encoding corresponding functional domains of IRF-1 protein (Cha et al., 1992; Schaper et al., 1998) (**Figure 24**). Skipping of any of these coding exons will result in the loss of function mediated by corresponding domains. Skipping of IRF-1 exon 2, 3, 7, 8 and 9 has been previously reported to be increased in leukemia patients and skipping of exon 2 and 3 have been proposed to be a potential mechanism of IRF-1 inactivation (Green et al., 1999; Harada et al., 1994a; Tzoanopoulos et al., 2002). The 619 A/C variation resides in intron 1 which is adjacent to exon 2 which hosts the transcription start codon and encodes the functional domain of IRF-1 responsible for DNA binding (Schaper et al., 1998). The IRF-1 MS region is in intron 7 of the IRF-1 gene and is in a putative site for Z-DNA formation which might affect DNA structure, transcription and

Figure 24 Functional domains of IRF-1 and locations of key polymorphisms

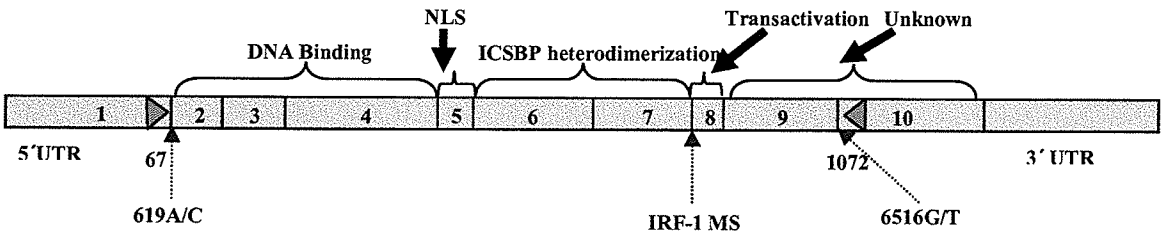


Figure 24 Functional domains of IRF-1 and location of key polymorphisms.

The IRF-1 mRNA contains two untranslated terminal regions (UTR) at both ends and 9 encoding exon sequences between them. Exons 2 to 4 encode the DNA binding domain of IRF-1 protein; Exon 5 encodes nuclear signaling domain; Exons 6 and 7 encodes the region mediating ICSBP heterodimerization; Exon 8 encodes the transaction domain. No known IRF-1 functions had been associated with exons 9 and 10. The arrows indicate the locations of designated RT-PCR primers and they amplify a segment of IRF-1 mRNA with 1006 nt in length encompassing nt 67 to nt 1072. The relative loci of intronic IRF-1 gene variations showing significant associations with HIV-1 resistance were indicated as well. It is noteworthy that these polymorphisms are closely related to the exons to which frequent exon skipping occurs.

translation (Cha et al., 1992; Li et al., 2002). The 6516G/T variation is located in intron 9 and it is next to exon 9 and 10 which encode domains of IRF-1 with unknown function(s) (Schaper et al., 1998). Together with the observation that IRF-1 specific RT-PCR products have a concomitant obscure tail, we believed that IRF-1 exon skipping might exist in our subjects as well.

Despite the fact that none of the IRF-1 intronic polymorphisms correlated with HIV resistance are located at predicted splicing sites (data not shown), our data in IRF-1 mRNA TA cloning and sequencing derived from PBMCs with different IRF-1 genotypes/haplotypes confirmed our expectation that exon skipping is present in our subjects and is a possible mechanism of altered IRF-1 expression. Our results indicated that exon skipping occurred frequently during IRF-1 mRNA splicing. Exon skipping was observed to be frequently occurring to exons 2, 3, 7, 8 and 9, rarely to exon 5. This coincides with previous observation in leukemia patient, suggesting these exons are functionally critical for IRF-1(Tzoanopoulos et al., 2002). When comparing the frequencies of clones with these exons skipped, we observed that subjects homozygous for the protective alleles at 619, IRF-1 MS and 6516 exhibited significantly more exon 2 and 3 skippings but less skipping for exon 7 and 8 than subjects heterozygous or homozygous for alternative allele at all these three loci. No significance was detected regarding to exon 9 or 5 skipping in different subjects.

This increased exon 2 and 3 skipping in protective IRF-1 genotype coincided with the data showing that 619A/C (resides in intronic region next to exon 2) was the dominant indicator for resistance to HIV-1 infection in the three IRF-1 polymorphisms correlated with HIV-1 resistance. The IRF-1 mRNA sequences lacking the exon 2 or 3 region lost

their original start codons and the typical five tryptophans repeats at N-terminal in coded sequences which composed the DNA binding domain of IRF-1 protein (Schaper et al., 1998). Exon 2 and 3 skippings may also result in the non-sense transcripts, potentially halting the further mRNA translation, and leading to the overall decrease of IRF-1 protein expression in subjects with protective IRF-1 genotypes. The loss of DNA binding motif will consequently impair the interaction of IRF-1 and HIV-1 LTR which is dependent on the binding of IRF-1 DNA binding domain and ISRE homologous sequence in HIV-1 LTR. IRF-1 may be an important primary initiator and amplifier of HIV-1 LTR transcription, it is then reasonable to hypothesize that a reduced or delayed onset of HIV-1 LTR transcription and decreased magnitude of HIV-1 transcription might occur when these subjects encounter HIV-1. HIV-1 appears to be persist and ineradicable once it is well established in the host. The early stage of HIV-1 infection leaves only hours to one week window period during which HIV-1 replication is tenuous and immunological or medical intervention could potentially eradicate the infection (Haase, 2005). If a reduced/delayed initiation of HIV-1 LTR transcription occurs, this window period might be widened to some extent and immune response will have more chance to sterilize the invaders and disruption of HIV-1 infection at an early stage. This might be one potential mechanism mediating the association between IRF-1 and HIV-1 resistance. Both decreased IRF-1 expression and increased exon 2 and 3 skippings might all be mediating the lower initiation of HIV-1 transcription at early stage of HIV-1 infection.

The decreased frequencies of exon 7 and exon 8 skipping might be related with the regulatory role IRF-1 plays in host immunity. IRF-1 exon 7 encodes part of the ICSBP heterodimerization domain which is involved in the interplay between IRF-1 and IRF-8

(previously known as ICSBP). IRF-8 and IRF-2 are two natural antagonists of IRF-1 under normal physiological conditions(Nguyen et al., 1995; Weisz et al., 1994; Weisz et al., 1992). IRF-8 can form a heterodimer with IRF-1 and repress function of IRF-1 through yet undecided mechanisms (Sharf et al., 1997). Therefore, exon 7 skipping might impair the negative regulation of IRF-1 function via IRF-8. Exon 8 encodes the transactivation domain which is the key regulator of IRF-1's ability to transactivate target genes and loss of this domain will deprive IRF-1 of transactivation ability which will consequently affect the functions of IRF-1 regulated genes (Lin et al., 1994; Schaper et al., 1998). Meanwhile, exon 8 skipping results in significant open reading frame change and mRNA with exon 8 skipped encodes a truncated IRF-1 protein missing exon 8 region and also exon 9 and 10 regions, whose functions remain unknown. How exon 7 and 8 skipping function in the subjects with specific IRF-1 genotype remains to be completely elucidated. The answer might lie in the altered regulation of antiviral host immunity: exon 7 and 8 skippings, especially exon 8, may significantly impair the normal immunoregulatory functions of IRF-1 motivation that normal IRF-1 does.

IRF-1 – A target in a survival game between HIV-1 and the host

Based on the currently available data on the interplay between HIV-1 and IRF-1, we are convinced that this is an intricate and complex process. IRF-1 apparently has a dual role to play in HIV/AIDS. It is a promoter for host antiviral immunity but also is an enhancer of HIV-1 transcription. How does a host antiviral factor also become an initiator and amplifier of HIV-1 transcription? We and others speculate that HIV-1 may be

utilizing this as a novel immune evasion strategy (Marsili et al., 2003).

To enhance their survival, mammalian hosts have developed strategies to fight the invading pathogens mainly via innate and adaptive immune responses. Viral infection normally results in the induction of cellular genes encoding proteins (cytokines and/or chemokines) that modulate host responses and/or have direct antiviral effects such as IFNs. These factors play important role in determination of the outcome of infection and greatly affect the viral pathogenesis as well. IFN and IRF responses represent the typical initial host immune responses to viral infection and they usually lead to a potent antiviral state in the host and bystander cells. The kinetics of IRF-1 induction by HIV-1 show a similar time course to that induced by other viruses such as Newcastle disease virus, where IRF-1 expression precedes IFN type I production (Marsili et al., 2003).

Thus induction of IRF-1 should be a negative event from the perspective of the pathogenic invaders. However, for their own survival purposes, pathogens have acquired some successful mechanisms of evading the immune defence as a consequence of significant selective pressure from the host. As a typical "Hit and Stay" virus, HIV-1 has developed multiple strategies for evading host immune attack, including constant mutational changes due to high replication and high mutational rate, the inaccessibility of epitopes recognizable by neutralizing antibodies and CTLs, latent infection, a large provirus reservoir, down modulation of the HLA class I expression on the surface of infected cells and profound depletion of activated and HIV-1 specific CD4 cells (Hilleman, 2004; Piguet and Trono, 2001). The role of IRF-1 in HIV-1 transcription suggests that HIV-1 may have effectively evolved another survival strategy in converting the IRF-1 response, an initial host antiviral reaction, to its own advantage (Harada et al.,

1989). The utilization of IRF-1 as a HIV-LTR transcription factor may represent an immune evasion strategy utilized by HIV-1 to bypass the host defences. Supportive evidence has shown that another early reactive protein, also a member of IRF family, named IRF-2. IRF-2 usually acts as a repressor of IRF-1 and IFN response induced by viral infection. It has been shown and IRF-2 can also be upregulated by HIV-1 viral protein Nef and has also a positive effect in favor of HIV-1 replication (Mamane et al., 1999; Marsili et al., 2003; Simmons et al., 2001).

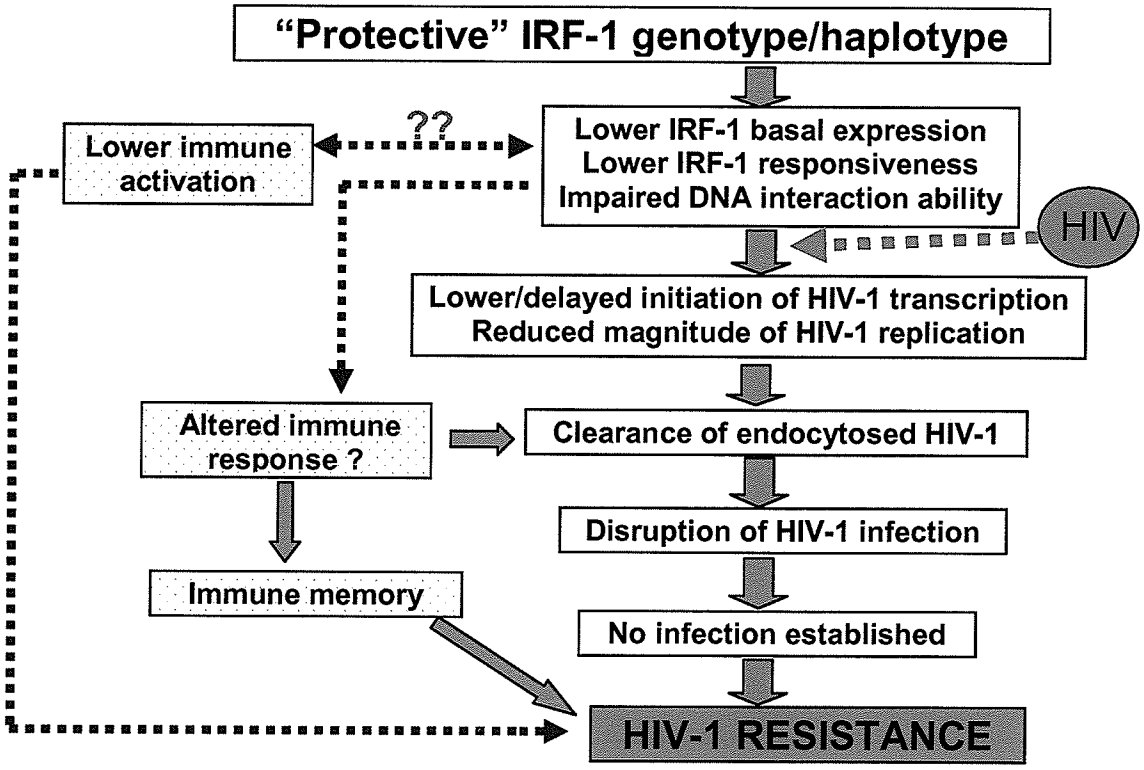
HIV-1 resistance – a consequence of successful decipher of HIV-1 immune evasion

The obvious question is how our findings fit in this complex interplay between HIV-1 and IRF-1, and how polymorphisms in IRF-1 contribute to the development of natural resistance to HIV-1 infection. HIV-1 infection induces IRF-1 upregulation early upon infection and this IRF-1 response likely benefits HIV-1 in the initiation of its replication and establishment of a productive infection. We discovered that certain IRF-1 genotypes showed a significant association with natural resistance to HIV-1 infection. These protective IRF-1 genotypes lead to lower IRF-1 basal level expression and lower IRF-1 responsiveness to exogenous stimulation including HIV-1 infection. These polymorphisms also correlate with altered IRF-1 mRNA splicing pattern that may lead to an increased likelihood of losing the DNA binding ability which mediates the interaction of IRF-1 and HIV-LTR. Both of these could contribute to the observed low level of HIV-1 transcription at the initial stage of infection as determined by our pseudovirus infection assay, potentially leading to the development of HIV-1 resistance in subjects hosting

these IRF-1 genotypes. Taking the exploitation of IRF-1 in HIV-1 transcription as a successful immune evasion strategy developed by HIV-1, these protective IRF-1 genotypes and functional changes in IRF-1 expression may also compose a protective mechanism evolved in the host in breaking the HIV-1 immune evasion from IRF-1 and IFN responses.

We raise a novel hypothesis which emphasizes the importance of limiting initial HIV-1 replication in developing resistance to HIV infection, and provide a new interpretation of this phenomenon. We hypothesize that protective IRF-1 genotypes/haplotypes predispose reduced IRF-1 basal level expression and decreased IRF-1 responsiveness to exogenous stimulation, including HIV-1 infection, and impaired IRF-1 ability to interact with HIV-1 LTR. As a consequence, HIV-1 transcription initiation is delayed and the magnitude of HIV-1 replication at early stage is reduced; this widens the window period for the host innate and adaptive immune responses to develop and disrupt the establishment of a productive HIV-1 infection. The induced HIV-1 immune memory would, in turn, enable the maintenance of long term protection (i.e. long time HIV-1 resistance). As one possible consequence of decreased IRF-1 expression and altered mRNA splicing patterns, reasonably lower level immunoactivation as detected in HIV-resistant subjects(Alimonti et al., 2005) may be an alternative mechanism underlying the identified correlation between IRF-1 polymorphisms and HIV-1 resistance (**Figure 25**).

Figure 25 A possible scenario in HIV-1 resistance



Significance

In conclusion, we have identified an association between polymorphisms in IRF-1 gene and resistance to infection by HIV-1. This is the first report indicating genetic variations in a transcriptional regulatory gene may affect susceptibility to HIV-1. We made significant contribution in identifying how and why these specific IRF-1 gene variations correlate with resistance to HIV-1 as well. Our data provide interesting hints on the mechanisms underlying this association including that: protective IRF-1 genotypes associate with lowered IRF-1 protein expression and reduced IRF-1 responsiveness upon exogenous stimulation, suggesting that lower IRF-1 protein levels may be one mechanism underlying this association; protective IRF-1 genotypes also associate with increased likelihood of exon 2 and 3 skippings and loss of DNA binding domain, which might also contribute to this correlation; reduced exon 7 and 8 skipping in subjects with protective IRF-1 genotypes suggested that altered immune regulation could also contribute to this altered host susceptibility to HIV-1 infection. This is also the first report indicating that factor regulating HIV-1 replication especially at initial stage is crucial for HIV-1 resistance. Lower LTR transcription in early infection in PBMCs from individuals with protective IRF-1 genotypes suggests that this may be a mechanism whereby these genotypes protect against HIV infection. These findings provide insight to natural anti-HIV-1 immunity and should impact studies on cell mediated immune responses to HIV-1 and inform the designs of future prophylactic and therapeutic strategies against HIV-1. These data also suggest that effective anti-HIV-1 strategies should target not only host immunity, but also factors important in the establishment of HIV-1 infection.

Future directions

The correlation between IRF-1 genetic polymorphisms and resistance to HIV-1 infection has been well established. Lowered IRF-1 basal expression and IRF-1 responsiveness and altered mRNA splicing associated with these polymorphisms have been suggested to be the potential underlying mechanisms mediating this correlation. A reduced initiation of HIV-1 transcription has been demonstrated in PBMCs from subjects with protective IRF-1 genotypes. To expand this study and conclusively prove these findings, the further functional studies might include:

1. IRF-1 knockdown experiments using siRNA in susceptible cell lines: This will conclusively prove that IRF-1 expression level and IRF-1 response to HIV-1 infection are essential for efficient HIV-1 initiation and establishment of productive HIV-1 infection.
2. Establishing an alternative experimental sytem using pseudotyped HIV-1 virus with a GFP repoter and using it to confirm the leuciferase assay results at single-cell level.
3. Testing the impacts of critical intronic IRF-1 polymorphisms on mRNA splicing and IRF-1 expression using artificial system, i.e. minigene system. This will help to clarify the exact role of these polymorphisms in IRF-1 mRNA splicing and further IRF-1 translation.
4. Real-time PCR to test the impact of intronic polymorphisms of IRF-1 on the expression level of its mRNA expression in a more quantitative manner. This will

help to both confirm the current western blot data and clarify the impact of these polymorphisms on IRF-1 transcription.

5. Testing the impacts of truncated IRF-1 proteins on HIV-1 LTR transcription to study the potential function impacts of IRF-1 exon skipping on HIV/AIDS.
6. Examining the efficiency and productivity of different HIV-1 viral strains in cells with different IRF-1 genotypes/haplotypes. This will help to evaluate the impact of IRF-1 polymorphisms in susceptibility to HIV-1 in general population.
7. Exploring the functional impacts of IRF-1 polymorphisms on expressions of IRF-1 regulated genes, i.e by microarray assay. It will help to fully evaluate the downstream functional impacts of these polymorphisms in IRF-1 related signal transduction and host immune alterations.
8. Exploring the potential ways to manipulate the IRF-1 expression in host cells and these could be exploited for the development potential prophylactic strategies, such as microbicide designs.

REFERENCES:

(1982) Update on acquired immune deficiency syndrome (AIDS)--United States. *MMWR Morb Mortal Wkly Rep*, **31**, 507-508, 513-504.

Akridge, R., Hladik, F., Markee, J., Alef, C., Kelley, H., Collier, A. and McElrath, M.J. (1999) Cellular immunity and target cell susceptibility in persons with repeated HIV-1 exposure. *Immunol Lett*, **66**, 15-19.

Al-Harthi, L. and Roebuck, K.A. (1998) Human immunodeficiency virus type-1 transcription: role of the 5'-untranslated leader region (review). *Int J Mol Med*, **1**, 875-881.

Alimonti, J.B., Ball, T.B. and Fowke, K.R. (2003) Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. *J Gen Virol*, **84**, 1649-1661.

Alimonti, J.B., Koesters, S.A., Kimani, J., Matu, L., Wachihi, C., Plummer, F.A. and Fowke, K.R. (2005) CD4+ T cell responses in HIV-exposed seronegative women are qualitatively distinct from those in HIV-infected women. *J Infect Dis*, **191**, 20-24.

Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M. and Berger, E.A. (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science*, **272**, 1955-1958.

Ao, Z., Fowke, K.R., Cohen, E.A. and Yao, X. (2005) Contribution of the C-terminal tri-lysine regions of human immunodeficiency virus type 1 integrase for efficient reverse transcription and viral DNA nuclear import. *Retrovirology*, **2**, 62.

Ao, Z., Yao, X. and Cohen, E.A. (2004) Assessment of the role of the central DNA flap in human immunodeficiency virus type 1 replication by using a single-cycle replication system. *J Virol*, **78**, 3170-3177.

Ball, T.B. (2001) A genetic basis for resistance to infection by HIV-1. *Department of Medical Microbiology*. University of Manitoba, Winnipeg.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.

Bartz, S.R., Rogel, M.E. and Emerman, M. (1996) Human immunodeficiency virus type 1 cell cycle control: Vpr is cytostatic and mediates G2 accumulation by a mechanism which differs from DNA damage checkpoint control. *J Virol*, **70**, 2324-2331.

Battistini, A., Marsili, G., Sgarbanti, M., Ensoli, B. and Hiscott, J. (2002) IRF regulation of HIV-1 long terminal repeat activity. *J Interferon Cytokine Res*, **22**, 27-37.

Beretta, A., Furci, L., Burastero, S., Cosma, A., Dinelli, M.E., Lopalco, L., DeSantis, C., Tambussi, G., Carrow, E., Sabbatani, S., Clerici, M., Lazzarin, A. and Siccardi, A.G. (1996) HIV-1-specific immunity in persistently seronegative individuals at high risk for HIV infection. *Immunol Lett*, **51**, 39-43.

Bernard, N.F., Yannakis, C.M., Lee, J.S. and Tsoukas, C.M. (1999) Human immunodeficiency virus (HIV)-specific cytotoxic T lymphocyte activity in HIV-exposed seronegative persons. *J Infect Dis*, **179**, 538-547.

Beyrer, C., Artenstein, A.W., Rugpao, S., Stephens, H., VanCott, T.C., Robb, M.L., Rinkaew, M., Birx, D.L., Khamboonruang, C., Zimmerman, P.A., Nelson, K.E. and Natpratan, C. (1999) Epidemiologic and biologic characterization of a cohort of human immunodeficiency virus type 1 highly exposed, persistently seronegative female sex workers in northern Thailand. Chiang Mai HEPS Working Group. *J Infect Dis*, **179**, 59-67.

Bidwell, J., Keen, L., Gallagher, G., Kimberly, R., Huizinga, T., McDermott, M.F., Oksenberg, J., McNicholl, J., Pociot, F., Hardt, C. and D'Alfonso, S. (1999) Cytokine gene polymorphism in human disease: on-line databases. *Genes Immun*, **1**, 3-19.

Bidwell, J., Keen, L., Gallagher, G., Kimberly, R., Huizinga, T., McDermott, M.F., Oksenberg, J., McNicholl, J., Pociot, F., Hardt, C. and D'Alfonso, S. (2001) Cytokine gene polymorphism in human disease: on-line databases, supplement 1. *Genes Immun*, **2**, 61-70.

Bienzle, D., MacDonald, K.S., Smaill, F.M., Kovacs, C., Baqi, M., Courssaris, B., Luscher, M.A., Walmsley, S.L. and Rosenthal, K.L. (2000) Factors contributing to the lack of human immunodeficiency virus type 1 (HIV-1) transmission in HIV-1-discordant partners. *J Infect Dis*, **182**, 123-132.

Blanch, V.J., Piskurich, J.F. and Kaetzel, C.S. (1999) Cutting edge: coordinate regulation of IFN regulatory factor-1 and the polymeric Ig receptor by proinflammatory cytokines. *J Immunol*, **162**, 1232-1235.

Borrow, P., Lewicki, H., Wei, X., Horwitz, M.S., Peffer, N., Meyers, H., Nelson, J.A., Gairin, J.E., Hahn, B.H., Oldstone, M.B. and Shaw, G.M. (1997) Antiviral

pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med*, **3**, 205-211.

Brander, C. and Walker, B.D. (1999) T lymphocyte responses in HIV-1 infection: implications for vaccine development. *Curr Opin Immunol*, **11**, 451-459.

Burger, D.M., van der Ven, A.J. and Koopmans, P.P. (2004) Highly active antiretroviral therapy for HIV infection: lessons for the future. *Neth J Med*, **62**, 407-408.

Burke, D.S. (1997) Recombination in HIV: an important viral evolutionary strategy. *Emerg Infect Dis*, **3**, 253-259.

Burns, D.N. and Mofenson, L.M. (1999) Paediatric HIV-1 infection. *Lancet*, **354 Suppl 2**, SII1-6.

Casado, G., Thomson, M.M., Sierra, M. and Najera, R. (2005) Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba. *J Acquir Immune Defic Syndr*, **40**, 532-537.

Cha, Y., Sims, S.H., Romine, M.F., Kaufmann, M. and Deisseroth, A.B. (1992) Human interferon regulatory factor 1: intron-exon organization. *DNA Cell Biol*, **11**, 605-611.

Chabot, B. (1996) Directing alternative splicing: cast and scenarios. *Trends Genet*, **12**, 472-478.

Chamberland, M.E., Castro, K.G., Haverkos, H.W., Miller, B.I., Thomas, P.A., Reiss, R., Walker, J., Spira, T.J., Jaffe, H.W. and Curran, J.W. (1984) Acquired immunodeficiency syndrome in the United States: an analysis of cases outside high-incidence groups. *Ann Intern Med*, **101**, 617-623.

Cheynier, R., Langlade-Demoyen, P., Marescot, M.R., Blanche, S., Blondin, G., Wain-Hobson, S., Griscelli, C., Vilmer, E. and Plata, F. (1992) Cytotoxic T lymphocyte responses in the peripheral blood of children born to human immunodeficiency virus-1-infected mothers. *Eur J Immunol*, **22**, 2211-2217.

Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P.D., Wu, L., Mackay, C.R., LaRosa, G., Newman, W., Gerard, N., Gerard, C. and Sodroski, J. (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, **85**, 1135-1148.

Chun, T.W., Engel, D., Berrey, M.M., Shea, T., Corey, L. and Fauci, A.S. (1998) Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc Natl Acad Sci U S A*, **95**, 8869-8873.

Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C. and et al. (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science*, **233**, 343-346.

Clerici, M., Clark, E.A., Polacino, P., Axberg, I., Kuller, L., Casey, N.I., Morton, W.R., Shearer, G.M. and Benveniste, R.E. (1994a) T-cell proliferation to subinfectious SIV correlates with lack of infection after challenge of macaques. *Aids*, **8**, 1391-1395.

Clerici, M., Giorgi, J.V., Chou, C.C., Gudeman, V.K., Zack, J.A., Gupta, P., Ho, H.N., Nishanian, P.G., Berzofsky, J.A. and Shearer, G.M. (1992) Cell-mediated immune response to human immunodeficiency virus (HIV) type 1 in seronegative

homosexual men with recent sexual exposure to HIV-1. *J Infect Dis*, **165**, 1012-1019.

Clerici, M., Levin, J.M., Kessler, H.A., Harris, A., Berzofsky, J.A., Landay, A.L. and Shearer, G.M. (1994b) HIV-specific T-helper activity in seronegative health care workers exposed to contaminated blood. *Jama*, **271**, 42-46.

Clerici, M. and Shearer, G.M. (1993) A TH1-->TH2 switch is a critical step in the etiology of HIV infection. *Immunol Today*, **14**, 107-111.

Clerici, M. and Shearer, G.M. (1994) The Th1-Th2 hypothesis of HIV infection: new insights. *Immunol Today*, **15**, 575-581.

Cocchi, F., DeVico, A.L., Garzino-Demo, A., Arya, S.K., Gallo, R.C. and Lusso, P. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science*, **270**, 1811-1815.

Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P. and et al. (1986) Human immunodeficiency viruses. *Science*, **232**, 697.

Connor, R.I., Paxton, W.A., Sheridan, K.E. and Koup, R.A. (1996) Macrophages and CD4+ T lymphocytes from two multiply exposed, uninfected individuals resist infection with primary non-syncytium-inducing isolates of human immunodeficiency virus type 1. *J Virol*, **70**, 8758-8764.

Coombs, R.W., Reichelderfer, P.S. and Landay, A.L. (2003) Recent observations on HIV type-1 infection in the genital tract of men and women. *Aids*, **17**, 455-480.

Coovadia, H. (2004) Antiretroviral agents--how best to protect infants from HIV and save their mothers from AIDS. *N Engl J Med*, **351**, 289-292.

Cullen, B.R. (1991) Regulation of HIV-1 gene expression. *Faseb J*, **5**, 2361-2368.

Curtis, B.M., Scharnowske, S. and Watson, A.J. (1992) Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc Natl Acad Sci U S A*, **89**, 8356-8360.

Database1. Entrez SNP database. http://www.ncbi.nlm.nih.gov/entrez.

Database2. Functionally Annotated Mutation Database. http://mutdb.org/AnnoSNP.

De Maria, A., Cirillo, C. and Moretta, L. (1994) Occurrence of human immunodeficiency virus type 1 (HIV-1)-specific cytolytic T cell activity in apparently uninfected children born to HIV-1-infected mothers. *J Infect Dis*, **170**, 1296-1299.

Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Smith, M.W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R. and O'Brien, S.J. (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science*, **273**, 1856-1862.

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M., Davis, C.B., Peiper, S.C., Schall, T.J., Littman, D.R. and Landau, N.R. (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, **381**, 661-666.

Devito, C., Hinkula, J., Kaul, R., Kimani, J., Kiama, P., Lopalco, L., Barass, C., Piconi, S., Trabattoni, D., Bwayo, J.J., Plummer, F., Clerici, M. and Broliden, K. (2002)

Cross-clade HIV-1-specific neutralizing IgA in mucosal and systemic compartments of HIV-1-exposed, persistently seronegative subjects. *J Acquir Immune Defic Syndr*, **30**, 413-420.

Devito, C., Hinkula, J., Kaul, R., Lopalco, L., Bwayo, J.J., Plummer, F., Clerici, M. and Broliden, K. (2000) Mucosal and plasma IgA from HIV-exposed seronegative individuals neutralize a primary HIV-1 isolate. *Aids*, **14**, 1917-1920.

Donegan, E., Stuart, M., Niland, J.C., Sacks, H.S., Azen, S.P., Dietrich, S.L., Faucett, C., Fletcher, M.A., Kleinman, S.H., Operskalski, E.A. and et al. (1990) Infection with human immunodeficiency virus type 1 (HIV-1) among recipients of antibody-positive blood donations. *Ann Intern Med*, **113**, 733-739.

Donn, R.P., Barrett, J.H., Farhan, A., Stopford, A., Pepper, L., Shelley, E., Davies, N., Ollier, W.E. and Thomson, W. (2001) Cytokine gene polymorphisms and susceptibility to juvenile idiopathic arthritis. British Paediatric Rheumatology Study Group. *Arthritis Rheum*, **44**, 802-810.

Douek, D.C., Brenchley, J.M., Betts, M.R., Ambrozak, D.R., Hill, B.J., Okamoto, Y., Casazza, J.P., Kuruppu, J., Kunstman, K., Wolinsky, S., Grossman, Z., Dybul, M., Oxenius, A., Price, D.A., Connors, M. and Koup, R.A. (2002) HIV preferentially infects HIV-specific CD4+ T cells. *Nature*, **417**, 95-98.

Dunand, V.A., Ng, C.M., Wade, J.A., Bwayo, J., Plummer, F.A. and MacDonald, K.S. (1997) HLA-DR 52- and 51-associated DRB1 alleles in Kenya, east Africa. *Tissue Antigens*, **49**, 397-402.

Esparza, J. (2005) The global HIV vaccine enterprise. *Int Microbiol*, **8**, 93-101.

Eugen-Olsen, J., Iversen, A.K., Garred, P., Koppelhus, U., Pedersen, C., Benfield, T.L., Sorensen, A.M., Katzenstein, T., Dickmeiss, E., Gerstoft, J., Skinhoj, P., Svejgaard, A., Nielsen, J.O. and Hofmann, B. (1997) Heterozygosity for a deletion in the CKR-5 gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. *Aids*, **11**, 305-310.

Fauci, A.S. (2003) HIV and AIDS: 20 years of science. *Nat Med*, **9**, 839-843.

Feng, Y., Broder, C.C., Kennedy, P.E. and Berger, E.A. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*, **272**, 872-877.

Fleming, D.T. and Wasserheit, J.N. (1999) From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sex Transm Infect*, **75**, 3-17.

Fowke, K.R., Dong, T., Rowland-Jones, S.L., Oyugi, J., Rutherford, W.J., Kimani, J., Krausa, P., Bwayo, J., Simonsen, J.N., Shearer, G.M. and Plummer, F.A. (1998) HIV type 1 resistance in Kenyan sex workers is not associated with altered cellular susceptibility to HIV type 1 infection or enhanced beta-chemokine production. *AIDS Res Hum Retroviruses*, **14**, 1521-1530.

Fowke, K.R., Kaul, R., Rosenthal, K.L., Oyugi, J., Kimani, J., Rutherford, W.J., Nagelkerke, N.J., Ball, T.B., Bwayo, J.J., Simonsen, J.N., Shearer, G.M. and Plummer, F.A. (2000) HIV-1-specific cellular immune responses among HIV-1-resistant sex workers. *Immunol Cell Biol*, **78**, 586-595.

Fowke, K.R., Nagelkerke, N.J., Kimani, J., Simonsen, J.N., Anzala, A.O., Bwayo, J.J., MacDonald, K.S., Ngugi, E.N. and Plummer, F.A. (1996) Resistance to HIV-1

infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet*, **348**, 1347-1351.

Fujita, T., Sakakibara, J., Sudo, Y., Miyamoto, M., Kimura, Y. and Taniguchi, T. (1988) Evidence for a nuclear factor(s), IRF-1, mediating induction and silencing properties to human IFN-beta gene regulatory elements. *Embo J*, **7**, 3397-3405.

Furci, L., Scarlatti, G., Burastero, S., Tambussi, G., Colognesi, C., Quillent, C., Longhi, R., Loverro, P., Borgonovo, B., Gaffi, D., Carrow, E., Malnati, M., Lusso, P., Siccardi, A.G., Lazzarin, A. and Beretta, A. (1997) Antigen-driven C-C chemokine-mediated HIV-1 suppression by CD4(+) T cells from exposed uninfected individuals expressing the wild-type CCR-5 allele. *J Exp Med*, **186**, 455-460.

Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., Leibowitch, J. and Popovic, M. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*, **220**, 865-867.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. and Hahn, B.H. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*, **397**, 436-441.

Geijtenbeek, T.B., Kwon, D.S., Torensma, R., van Vliet, S.J., van Duijnhoven, G.C., Middel, J., Cornelissen, I.L., Nottet, H.S., KewalRamani, V.N., Littman, D.R., Figdor, C.G. and van Kooyk, Y. (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell*, **100**, 587-597.

Geretti, A.M. (2006) HIV-1 subtypes: epidemiology and significance for HIV management. *Curr Opin Infect Dis*, **19**, 1-7.

Goh, W.C., Markee, J., Akridge, R.E., Meldorf, M., Musey, L., Karchmer, T., Krone, M., Collier, A., Corey, L., Emerman, M. and McElrath, M.J. (1999) Protection against human immunodeficiency virus type 1 infection in persons with repeated exposure: evidence for T cell immunity in the absence of inherited CCR5 coreceptor defects. *J Infect Dis*, **179**, 548-557.

Gorham, J.D., Guler, M.L., Steen, R.G., Mackey, A.J., Daly, M.J., Frederick, K., Dietrich, W.F. and Murphy, K.M. (1996) Genetic mapping of a murine locus controlling development of T helper 1/T helper 2 type responses. *Proc Natl Acad Sci U S A*, **93**, 12467-12472.

Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A. and Saxon, A. (1981) Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med*, **305**, 1425-1431.

Goulder, P.J. and Watkins, D.I. (2004) HIV and SIV CTL escape: implications for vaccine design. *Nat Rev Immunol*, **4**, 630-640.

Green, W.B., Slovak, M.L., Chen, I.M., Pallavicini, M., Hecht, J.L. and Willman, C.L. (1999) Lack of IRF-1 expression in acute promyelocytic leukemia and in a subset of acute myeloid leukemias with del(5)(q31). *Leukemia*, **13**, 1960-1971.

Greene, W.C. (1990) Regulation of HIV-1 gene expression. *Annu Rev Immunol*, **8**, 453-475.

Haase, A.T. (2005) Perils at mucosal front lines for HIV and SIV and their hosts. *Nat Rev Immunol*, **5**, 783-792.

Hahn, B.H., Shaw, G.M., De Cock, K.M. and Sharp, P.M. (2000) AIDS as a zoonosis: scientific and public health implications. *Science*, **287**, 607-614.

Harada, H., Fujita, T., Miyamoto, M., Kimura, Y., Maruyama, M., Furia, A., Miyata, T. and Taniguchi, T. (1989) Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes. *Cell*, **58**, 729-739.

Harada, H., Kondo, T., Ogawa, S., Tamura, T., Kitagawa, M., Tanaka, N., Lamphier, M.S., Hirai, H. and Taniguchi, T. (1994a) Accelerated exon skipping of IRF-1 mRNA in human myelodysplasia/leukemia; a possible mechanism of tumor suppressor inactivation. *Oncogene*, **9**, 3313-3320.

Harada, H., Takahashi, E., Itoh, S., Harada, K., Hori, T.A. and Taniguchi, T. (1994b) Structure and regulation of the human interferon regulatory factor 1 (IRF-1) and IRF-2 genes: implications for a gene network in the interferon system. *Mol Cell Biol*, **14**, 1500-1509.

Haukim, N., Bidwell, J.L., Smith, A.J., Keen, L.J., Gallagher, G., Kimberly, R., Huizinga, T., McDermott, M.F., Oksenberg, J., McNicholl, J., Pociot, F., Hardt, C. and D'Alfonso, S. (2002) Cytokine gene polymorphism in human disease: on-line databases, supplement 2. *Genes Immun*, **3**, 313-330.

Herbert, A. (1996) RNA editing, introns and evolution. *Trends Genet*, **12**, 6-9.

Hill, A.V. (1998) The immunogenetics of human infectious diseases. *Annu Rev Immunol*, **16**, 593-617.

Hill, A.V. (2001) The genomics and genetics of human infectious disease susceptibility. *Annu Rev Genomics Hum Genet*, **2**, 373-400.

Hilleman, M.R. (2004) Strategies and mechanisms for host and pathogen survival in acute and persistent viral infections. *Proc Natl Acad Sci U S A*, **101 Suppl 2**, 14560-14566.

Hobart, M., Ramassar, V., Goes, N., Urmson, J. and Halloran, P.F. (1997) IFN regulatory factor-1 plays a central role in the regulation of the expression of class I and II MHC genes in vivo. *J Immunol*, **158**, 4260-4269.

Hogan, C.M. and Hammer, S.M. (2001) Host determinants in HIV infection and disease. Part 1: cellular and humoral immune responses. *Ann Intern Med*, **134**, 761-776.

Huang, Y., Paxton, W.A., Wolinsky, S.M., Neumann, A.U., Zhang, L., He, T., Kang, S., Ceradini, D., Jin, Z., Yazdanbakhsh, K., Kunstman, K., Erickson, D., Dragon, E., Landau, N.R., Phair, J., Ho, D.D. and Koup, R.A. (1996) The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med*, **2**, 1240-1243.

Hughes, A.E. (1993) Optimization of microsatellite analysis for genetic mapping. *Genomics*, **15**, 433-434.

Hung Fan, R.F.C., Luis P. Villarreal. (2004) *AIDS: Science and Society*. Jones and Bartlett Publishers, Boston, MA.

Ioannidis, J.P., Contopoulos-Ioannidis, D.G., Rosenberg, P.S., Goedert, J.J., De Rossi, A., Espanol, T., Frenkel, L., Mayaux, M.J., Newell, M.L., Pahwa, S.G., Rousseau, C., Scarlatti, G., Sei, S., Sen, L. and O'Brien, T.R. (2003) Effects of CCR5-delta32

and CCR2-64I alleles on disease progression of perinatally HIV-1-infected children: an international meta-analysis. *Aids*, **17**, 1631-1638.

Ioannidis, J.P., O'Brien, T.R., Rosenberg, P.S., Contopoulos-Ioannidis, D.G. and Goedert, J.J. (1998) Genetic effects on HIV disease progression. *Nat Med*, **4**, 536.

Iqbal, S.M., Ball, T.B., Kimani, J., Kiama, P., Thottingal, P., Embree, J.E., Fowke, K.R. and Plummer, F.A. (2005) Elevated T cell counts and RANTES expression in the genital mucosa of HIV-1-resistant Kenyan commercial sex workers. *J Infect Dis*, **192**, 728-738.

Itoh, S., Harada, H., Nakamura, Y., White, R. and Taniguchi, T. (1991) Assignment of the human interferon regulatory factor-1 (IRF1) gene to chromosome 5q23-q31. *Genomics*, **10**, 1097-1099.

Ji, H., Ball, T.B., Kimani, J. and Plummer, F.A. (2004) Novel interferon regulatory factor-1 polymorphisms in a Kenyan population revealed by complete gene sequencing. *J Hum Genet*, **49**, 528-535.

Jones, K.A. (1993) Tat and the HIV-1 promoter. *Curr Opin Cell Biol*, **5**, 461-468.

Julg, B. and Goebel, F.D. (2005) Susceptibility to HIV/AIDS: an individual characteristic we can measure? *Infection*, **33**, 160-162.

Kaleem, K., Chertok, F. and Erhan, S. (1987) Novel materials from protein-polymer grafts. *Nature*, **325**, 328-329.

Kaplan, E.H. and Heimer, R. (1995) HIV incidence among New Haven needle exchange participants: updated estimates from syringe tracking and testing data. *J Acquir Immune Defic Syndr Hum Retrovirol*, **10**, 175-176.

Kaul, R., Dong, T., Plummer, F.A., Kimani, J., Rostron, T., Kiama, P., Njagi, E., Irungu, E., Farah, B., Oyugi, J., Chakraborty, R., MacDonald, K.S., Bwayo, J.J., McMichael, A. and Rowland-Jones, S.L. (2001a) CD8(+) lymphocytes respond to different HIV epitopes in seronegative and infected subjects. *J Clin Invest*, **107**, 1303-1310.

Kaul, R., Plummer, F., Clerici, M., Bomsel, M., Lopalco, L. and Broliden, K. (2001b) Mucosal IgA in exposed, uninfected subjects: evidence for a role in protection against HIV infection. *Aids*, **15**, 431-432.

Kaul, R., Rowland-Jones, S.L., Kimani, J., Fowke, K., Dong, T., Kiama, P., Rutherford, J., Njagi, E., Mwangi, F., Rostron, T., Onyango, J., Oyugi, J., MacDonald, K.S., Bwayo, J.J. and Plummer, F.A. (2001c) New insights into HIV-1 specific cytotoxic T-lymphocyte responses in exposed, persistently seronegative Kenyan sex workers. *Immunol Lett*, **79**, 3-13.

Kaul, R., Rutherford, J., Rowland-Jones, S.L., Kimani, J., Onyango, J.I., Fowke, K., MacDonald, K., Bwayo, J.J., McMichael, A.J. and Plummer, F.A. (2004) HIV-1 Env-specific cytotoxic T-lymphocyte responses in exposed, uninfected Kenyan sex workers: a prospective analysis. *Aids*, **18**, 2087-2089.

Kaul, R., Trabattoni, D., Bwayo, J.J., Arienti, D., Zagliani, A., Mwangi, F.M., Kariuki, C., Ngugi, E.N., MacDonald, K.S., Ball, T.B., Clerici, M. and Plummer, F.A. (1999) HIV-1-specific mucosal IgA in a cohort of HIV-1-resistant Kenyan sex workers. *Aids*, **13**, 23-29.

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Mpoudi Ngole,

E., Bienvenue, Y., Delaporte, E., Brookfield, J.F., Sharp, P.M., Shaw, G.M., Peeters, M. and Hahn, B.H. (2006) Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1. *Science*.

Koning, F.A., Jansen, C.A., Dekker, J., Kaslow, R.A., Dukers, N., van Baarle, D., Prins, M. and Schuitemaker, H. (2004) Correlates of resistance to HIV-1 infection in homosexual men with high-risk sexual behaviour. *Aids*, **18**, 1117-1126.

Koning, F.A., Otto, S.A., Hazenberg, M.D., Dekker, L., Prins, M., Miedema, F. and Schuitemaker, H. (2005) Low-level CD4+ T cell activation is associated with low susceptibility to HIV-1 infection. *J Immunol*, **175**, 6117-6122.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S. and Bhattacharya, T. (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science*, **288**, 1789-1796.

Kostrikis, L.G., Huang, Y., Moore, J.P., Wolinsky, S.M., Zhang, L., Guo, Y., Deutsch, L., Phair, J., Neumann, A.U. and Ho, D.D. (1998) A chemokine receptor CCR2 allele delays HIV-1 disease progression and is associated with a CCR5 promoter mutation. *Nat Med*, **4**, 350-353.

Kreiss, J.K., Kitchen, L.W., Prince, H.E., Kasper, C.K. and Essex, M. (1985) Antibody to human T-lymphotropic virus type III in wives of hemophiliacs. Evidence for heterosexual transmission. *Ann Intern Med*, **102**, 623-626.

Kroger, A., Koster, M., Schroeder, K., Hauser, H. and Mueller, P.P. (2002) Activities of IRF-1. *J Interferon Cytokine Res*, **22**, 5-14.

Kucia, M., Jankowski, K., Reca, R., Wysoczynski, M., Bandura, L., Allendorf, D.J., Zhang, J., Ratajczak, J. and Ratajczak, M.Z. (2004) CXCR4-SDF-1 signalling, locomotion, chemotaxis and adhesion. *J Mol Histol*, **35**, 233-245.

Kulkarni, P.S., Butera, S.T. and Duerr, A.C. (2003) Resistance to HIV-1 infection: lessons learned from studies of highly exposed persistently seronegative (HEPS) individuals. *AIDS Rev*, **5**, 87-103.

Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*, **5**, 150-163.

Laga, M., Manoka, A., Kivuvu, M., Malele, B., Tuliza, M., Nzila, N., Goeman, J., Behets, F., Batter, V., Alary, M. and et al. (1993) Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: results from a cohort study. *Aids*, **7**, 95-102.

Laga, M., Nzila, N. and Goeman, J. (1991) The interrelationship of sexually transmitted diseases and HIV infection: implications for the control of both epidemics in Africa. *Aids*, **5 Suppl 1**, S55-63.

Lancaster, A., Nelson, M.P., Meyer, D., Thomson, G. and Single, R.M. (2003) PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data. *Pac Symp Biocomput*, 514-525.

Langlade-Demoyen, P., Ngo-Giang-Huong, N., Ferchal, F. and Oksenhendler, E. (1994) Human immunodeficiency virus (HIV) nef-specific cytotoxic T lymphocytes in noninfected heterosexual contact of HIV-infected patients. *J Clin Invest*, **93**, 1293-1297.

Lavreys, L., Baeten, J.M., Martin, H.L., Jr., Overbaugh, J., Mandaliya, K., Ndinya-Achola, J. and Kreiss, J.K. (2004a) Hormonal contraception and risk of HIV-1 acquisition: results of a 10-year prospective study. *Aids*, **18**, 695-697.

Lavreys, L., Chohan, V., Overbaugh, J., Hassan, W., McClelland, R.S., Kreiss, J., Mandaliya, K., Ndinya-Achola, J. and Baeten, J.M. (2004b) Hormonal contraception and risk of cervical infections among HIV-1-seropositive Kenyan women. *Aids*, **18**, 2179-2184.

Levy, J.A. (1993) Pathogenesis of human immunodeficiency virus infection. *Microbiol Rev*, **57**, 183-289.

Li, Y.C., Korol, A.B., Fahima, T., Beiles, A. and Nevo, E. (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol*, **11**, 2453-2465.

Lin, R., Mustafa, A., Nguyen, H., Gewert, D. and Hiscott, J. (1994) Mutational analysis of interferon (IFN) regulatory factors 1 and 2. Effects on the induction of IFN-beta gene expression. *J Biol Chem*, **269**, 17542-17549.

Liu, H., Chao, D., Nakayama, E.E., Taguchi, H., Goto, M., Xin, X., Takamatsu, J.K., Saito, H., Ishikawa, Y., Akaza, T., Juji, T., Takebe, Y., Ohishi, T., Fukutake, K., Maruyama, Y., Yashiki, S., Sonoda, S., Nakamura, T., Nagai, Y., Iwamoto, A. and Shioda, T. (1999) Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. *Proc Natl Acad Sci U S A*, **96**, 4581-4585.

Liu, H., Hwangbo, Y., Holte, S., Lee, J., Wang, C., Kaupp, N., Zhu, H., Celum, C., Corey, L., McElrath, M.J. and Zhu, T. (2004) Analysis of genetic polymorphisms in CCR5, CCR2, stromal cell-derived factor-1, RANTES, and dendritic cell-specific

intercellular adhesion molecule-3-grabbing nonintegrin in seronegative individuals repeatedly exposed to HIV-1. *J Infect Dis*, **190**, 1055-1058.

Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A. and Landau, N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, **86**, 367-377.

Lockett, S.F., Alonso, A., Wyld, R., Martin, M.P., Robertson, J.R., Gore, S.M., Leen, C.L., Brettle, R.P., Yirrell, D.L., Carrington, M. and Brown, A.J. (1999) Effect of chemokine receptor mutations on heterosexual human immunodeficiency virus transmission. *J Infect Dis*, **180**, 614-621.

Lockett, S.F., Robertson, J.R., Brettle, R.P., Yap, P.L., Middleton, D. and Leigh Brown, A.J. (2001) Mismatched human leukocyte antigen alleles protect against heterosexual HIV transmission. *J Acquir Immune Defic Syndr*, **27**, 277-280.

Lohoff, M. and Mak, T.W. (2005) Roles of interferon-regulatory factors in T-helper-cell differentiation. *Nat Rev Immunol*, **5**, 125-135.

Louisirirotchanakul, S., Liu, H., Roongpisuthipong, A., Nakayama, E.E., Takebe, Y., Shioda, T. and Wasi, C. (2002) Genetic analysis of HIV-1 discordant couples in Thailand: association of CCR2 64I homozygosity with HIV-1-negative status. *J Acquir Immune Defic Syndr*, **29**, 314-315.

Lucey, D.R., Clerici, M. and Shearer, G.M. (1996) Type 1 and type 2 cytokine dysregulation in human infectious, neoplastic, and inflammatory diseases. *Clin Microbiol Rev*, **9**, 532-562.

Luster, A.D. (1998) Chemokines--chemotactic cytokines that mediate inflammation. *N Engl J Med*, **338**, 436-445.

Ma, X., Chow, J.M., Gri, G., Carra, G., Gerosa, F., Wolf, S.F., Dzialo, R. and Trinchieri, G. (1996) The interleukin 12 p40 gene promoter is primed by interferon gamma in monocytic cells. *J Exp Med*, **183**, 147-157.

MacDonald, K.S., Fowke, K.R., Kimani, J., Dunand, V.A., Nagelkerke, N.J., Ball, T.B., Oyugi, J., Njagi, E., Gaur, L.K., Brunham, R.C., Wade, J., Luscher, M.A., Krausa, P., Rowland-Jones, S., Ngugi, E., Bwayo, J.J. and Plummer, F.A. (2000) Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis*, **181**, 1581-1589.

MacDonald, K.S., Matukas, L., Embree, J.E., Fowke, K., Kimani, J., Nagelkerke, N.J., Oyugi, J., Kiama, P., Kaul, R., Luscher, M.A., Rowland-Jones, S., Ndinya-Achola, J., Ngugi, E., Bwayo, J.J. and Plummer, F.A. (2001) Human leucocyte antigen supertypes and immune susceptibility to HIV-1, implications for vaccine design. *Immunol Lett*, **79**, 151-157.

Mamane, Y., Heylbroeck, C., Genin, P., Algarte, M., Servant, M.J., LePage, C., DeLuca, C., Kwon, H., Lin, R. and Hiscott, J. (1999) Interferon regulatory factors: the next generation. *Gene*, **237**, 1-14.

Marsili, G., Borsetti, A., Sgarbanti, M., Remoli, A.L., Ridolfi, B., Stellacci, E., Ensoli, B. and Battistini, A. (2003) On the role of interferon regulatory factors in HIV-1 replication. *Ann N Y Acad Sci*, **1010**, 29-42.

Masur, H., Michelis, M.A., Greene, J.B., Onorato, I., Stouwe, R.A., Holzman, R.S., Wormser, G., Brettman, L., Lange, M., Murray, H.W. and Cunningham-Rundles,

S. (1981) An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N Engl J Med*, **305**, 1431-1438.

Matsuyama, T., Kimura, T., Kitagawa, M., Pfeffer, K., Kawakami, T., Watanabe, N., Kundig, T.M., Amakawa, R., Kishihara, K., Wakeham, A. and et al. (1993) Targeted disruption of IRF-1 or IRF-2 results in abnormal type I IFN gene induction and aberrant lymphocyte development. *Cell*, **75**, 83-97.

Mattick, J.S. (1994) Introns: evolution and function. *Curr Opin Genet Dev*, **4**, 823-831.

Mazzoli, S., Trabattoni, D., Lo Caputo, S., Piconi, S., Ble, C., Meacci, F., Ruzzante, S., Salvi, A., Semplici, F., Longhi, R., Fusi, M.L., Tofani, N., Biasin, M., Villa, M.L., Mazzotta, F. and Clerici, M. (1997) HIV-specific mucosal and cellular immunity in HIV-seronegative partners of HIV-seropositive individuals. *Nat Med*, **3**, 1250-1257.

Mehendale, S.M., Rodrigues, J.J., Brookmeyer, R.S., Gangakhedkar, R.R., Divekar, A.D., Gokhale, M.R., Risbud, A.R., Paranjape, R.S., Shepherd, M.E., Rompalo, A.E. and et al. (1995) Incidence and predictors of human immunodeficiency virus type 1 seroconversion in patients attending sexually transmitted disease clinics in India. *J Infect Dis*, **172**, 1486-1491.

Miller, C.J., Li, Q., Abel, K., Kim, E.Y., Ma, Z.M., Wietgrefe, S., La Franco-Scheuch, L., Compton, L., Duan, L., Shore, M.D., Zupancic, M., Busch, M., Carlis, J., Wolinsky, S. and Haase, A.T. (2005) Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. *J Virol*, **79**, 9217-9227.

Miller, C.J., McChesney, M.B., Lu, X., Dailey, P.J., Chutkowski, C., Lu, D., Brosio, P., Roberts, B. and Lu, Y. (1997) Rhesus macaques previously infected with simian/human immunodeficiency virus are protected from vaginal challenge with pathogenic SIVmac239. *J Virol*, **71**, 1911-1921.

Ministry of Health, R.o.K. (2005) *AIDS in Kenya: trends, interventions and impact.*

Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., Sudo, Y., Miyata, T. and Taniguchi, T. (1988) Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN-beta gene regulatory elements. *Cell*, **54**, 903-913.

Moore, M.J. (1996) Gene expression. When the junk isn't junk. *Nature*, **379**, 402-403.

Moses, S., Nagelkerke, N.J. and Blanchard, J. (1999) Analysis of the scientific literature on male circumcision and risk for HIV infection. *Int J STD AIDS*, **10**, 626-628.

Mulherin, S.A., O'Brien, T.R., Ioannidis, J.P., Goedert, J.J., Buchbinder, S.P., Coutinho, R.A., Jamieson, B.D., Meyer, L., Michael, N.L., Pantaleo, G., Rizzardi, G.P., Schuitemaker, H., Sheppard, H.W., Theodorou, I.D., Vlahov, D. and Rosenberg, P.S. (2003) Effects of CCR5-Delta32 and CCR2-64I alleles on HIV-1 disease progression: the protection varies with duration of infection. *Aids*, **17**, 377-387.

Nahmias, A.J., Weiss, J., Yao, X., Lee, F., Kodsi, R., Schanfield, M., Matthews, T., Bolognesi, D., Durack, D., Motulsky, A. and et al. (1986) Evidence for human infection with an HTLV III/LAV-like virus in Central Africa, 1959. *Lancet*, **1**, 1279-1280.

Nakao, F., Ihara, K., Kusuhara, K., Sasaki, Y., Kinukawa, N., Takabayashi, A., Nishima, S. and Hara, T. (2001) Association of IFN-gamma and IFN regulatory factor 1

polymorphisms with childhood atopic asthma. *J Allergy Clin Immunol*, **107**, 499-504.

Nelson, N., Kanno, Y., Hong, C., Contursi, C., Fujita, T., Fowlkes, B.J., O'Connell, E., Hu-Li, J., Paul, W.E., Jankovic, D., Sher, A.F., Coligan, J.E., Thornton, A., Appella, E., Yang, Y. and Ozato, K. (1996) Expression of IFN regulatory factor family proteins in lymphocytes. Induction of Stat-1 and IFN consensus sequence binding protein expression by T cell activation. *J Immunol*, **156**, 3711-3720.

Nguyen, H., Hiscott, J. and Pitha, P.M. (1997) The growing family of interferon regulatory factors. *Cytokine Growth Factor Rev*, **8**, 293-312.

Nguyen, H., Mustafa, A., Hiscott, J. and Lin, R. (1995) Transcription factor IRF-2 exerts its oncogenic phenotype through the DNA binding/transcription repression domain. *Oncogene*, **11**, 537-544.

Noguchi, E., Shibasaki, M., Arinami, T., Yamakawa-Kobayashi, K., Yokouchi, Y., Takeda, K., Matsui, A. and Hamaguchi, H. (2000) Mutation screening of interferon regulatory factor 1 gene (IRF-1) as a candidate gene for atopy/asthma. *Clin Exp Allergy*, **30**, 1562-1567.

Ogasawara, K., Hida, S., Azimi, N., Tagaya, Y., Sato, T., Yokochi-Fukuda, T., Waldmann, T.A., Taniguchi, T. and Taki, S. (1998) Requirement for IRF-1 in the microenvironment supporting development of natural killer cells. *Nature*, **391**, 700-703.

Ogg, G.S., Jin, X., Bonhoeffer, S., Dunbar, P.R., Nowak, M.A., Monard, S., Segal, J.P., Cao, Y., Rowland-Jones, S.L., Cerundolo, V., Hurley, A., Markowitz, M., Ho,

D.D., Nixon, D.F. and McMichael, A.J. (1998) Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science*, **279**, 2103-2106.

Ohteki, T., Yoshida, H., Matsuyama, T., Duncan, G.S., Mak, T.W. and Ohashi, P.S. (1998) The transcription factor interferon regulatory factor 1 (IRF-1) is important during the maturation of natural killer 1.1+ T cell receptor-alpha/beta+ (NK1+ T) cells, natural killer cells, and intestinal intraepithelial T cells. *J Exp Med*, **187**, 967-972.

Palella, F.J., Jr., Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J., Satten, G.A., Aschman, D.J. and Holmberg, S.D. (1998) Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med*, **338**, 853-860.

Paranjape, R.S. (2005) Immunopathogenesis of HIV infection. *Indian J Med Res*, **121**, 240-255.

Parren, P.W., Gauduin, M.C., Koup, R.A., Poignard, P., Fisicaro, P., Burton, D.R. and Sattentau, Q.J. (1997) Relevance of the antibody response against human immunodeficiency virus type 1 envelope to vaccine design. *Immunol Lett*, **57**, 105-112.

Paxton, W.A., Liu, R., Kang, S., Wu, L., Gingeras, T.R., Landau, N.R., Mackay, C.R. and Koup, R.A. (1998) Reduced HIV-1 infectability of CD4+ lymphocytes from exposed-uninfected individuals: association with low expression of CCR5 and high production of beta-chemokines. *Virology*, **244**, 66-73.

Paxton, W.A., Martin, S.R., Tse, D., O'Brien, T.R., Skurnick, J., VanDevanter, N.L., Padian, N., Braun, J.F., Kotler, D.P., Wolinsky, S.M. and Koup, R.A. (1996)

Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure. *Nat Med*, **2**, 412-417.

Penninger, J.M. and Mak, T.W. (1998) Thymocyte selection in Vav and IRF-1 gene-deficient mice. *Immunol Rev*, **165**, 149-166.

Phillips, A.N., Lepri, A.C., Lampe, F., Johnson, M. and Sabin, C.A. (2003) When should antiretroviral therapy be started for HIV infection? Interpreting the evidence from observational studies. *Aids*, **17**, 1863-1869.

Philpott, S., Burger, H., Charbonneau, T., Grimson, R., Vermund, S.H., Visosky, A., Nachman, S., Kovacs, A., Tropper, P., Frey, H. and Weiser, B. (1999) CCR5 genotype and resistance to vertical transmission of HIV-1. *J Acquir Immune Defic Syndr*, **21**, 189-193.

Piguet, V. and Trono, D. (2001) Living in oblivion: HIV immune evasion. *Semin Immunol*, **13**, 51-57.

Pilgrim, A.K., Pantaleo, G., Cohen, O.J., Fink, L.M., Zhou, J.Y., Zhou, J.T., Bolognesi, D.P., Fauci, A.S. and Montefiori, D.C. (1997) Neutralizing antibody responses to human immunodeficiency virus type 1 in primary infection and long-term-nonprogressive infection. *J Infect Dis*, **176**, 924-932.

Pinto, L.A., Sullivan, J., Berzofsky, J.A., Clerici, M., Kessler, H.A., Landay, A.L. and Shearer, G.M. (1995) ENV-specific cytotoxic T lymphocyte responses in HIV seronegative health care workers occupationally exposed to HIV-contaminated body fluids. *J Clin Invest*, **96**, 867-876.

Piot, P., Bartos, M., Ghys, P.D., Walker, N. and Schwartlander, B. (2001) The global impact of HIV/AIDS. *Nature*, **410**, 968-973.

Plata, F., Autran, B., Martins, L.P., Wain-Hobson, S., Raphael, M., Mayaud, C., Denis, M., Guillon, J.M. and Debre, P. (1987) AIDS virus-specific cytotoxic T lymphocytes in lung disorders. *Nature*, **328**, 348-351.

Plummer, F.A. (1998) Heterosexual transmission of human immunodeficiency virus type 1 (HIV): interactions of conventional sexually transmitted diseases, hormonal contraception and HIV-1. *AIDS Res Hum Retroviruses*, **14 Suppl 1**, S5-10.

Plummer, F.A., Ball, T.B., Kimani, J. and Fowke, K.R. (1999) Resistance to HIV-1 infection among highly exposed sex workers in Nairobi: what mediates protection and why does it develop? *Immunol Lett*, **66**, 27-34.

Plummer, F.A., Simonsen, J.N., Cameron, D.W., Ndinya-Achola, J.O., Kreiss, J.K., Gakinya, M.N., Waiyaki, P., Cheang, M., Piot, P., Ronald, A.R. and et al. (1991) Cofactors in male-female sexual transmission of human immunodeficiency virus type 1. *J Infect Dis*, **163**, 233-239.

Pope, M. and Haase, A.T. (2003) Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nat Med*, **9**, 847-852.

Popovic, M., Sarngadharan, M.G., Read, E. and Gallo, R.C. (1984) Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*, **224**, 497-500.

Price, D.A., Goulder, P.J., Klenerman, P., Sewell, A.K., Easterbrook, P.J., Troop, M., Bangham, C.R. and Phillips, R.E. (1997) Positive selection of HIV-1 cytotoxic T

lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A*, **94**, 1890-1895.

Quinn, T.C. and Overbaugh, J. (2005) HIV/AIDS in women: an expanding epidemic. *Science*, **308**, 1582-1583.

Quinn, T.C., Wawer, M.J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., Meehan, M.O., Lutalo, T. and Gray, R.H. (2000) Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med*, **342**, 921-929.

Redfield, R.R., Wright, D.C. and Tramont, E.C. (1986) The Walter Reed staging classification for HTLV-III/LAV infection. *N Engl J Med*, **314**, 131-132.

Richman, D.D., Wrin, T., Little, S.J. and Petropoulos, C.J. (2003) Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci U S A*, **100**, 4144-4149.

Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S. and Korber, B. (2000) HIV-1 nomenclature proposal. *Science*, **288**, 55-56.

Rohowsky-Kochan, C., Skurnick, J., Molinaro, D. and Louria, D. (1998) HLA antigens associated with susceptibility/resistance to HIV-1 infection. *Hum Immunol*, **59**, 802-815.

Rosenberg, E.S., Billingsley, J.M., Caliendo, A.M., Boswell, S.L., Sax, P.E., Kalams, S.A. and Walker, B.D. (1997) Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. *Science*, **278**, 1447-1450.

Rowland-Jones, S., Sutton, J., Ariyoshi, K., Dong, T., Gotch, F., McAdam, S., Whitby, D., Sabally, S., Gallimore, A., Corrah, T. and et al. (1995) HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women. *Nat Med*, **1**, 59-64.

Rowland-Jones, S.L., Dong, T., Fowke, K.R., Kimani, J., Krausa, P., Newell, H., Blanchard, T., Ariyoshi, K., Oyugi, J., Ngugi, E., Bwayo, J., MacDonald, K.S., McMichael, A.J. and Plummer, F.A. (1998) Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. *J Clin Invest*, **102**, 1758-1765.

Rowland-Jones, S.L., Nixon, D.F., Aldhous, M.C., Gotch, F., Ariyoshi, K., Hallam, N., Kroll, J.S., Froebel, K. and McMichael, A. (1993) HIV-specific cytotoxic T-cell activity in an HIV-exposed but uninfected infant. *Lancet*, **341**, 860-861.

Royce, R.A., Sena, A., Cates, W., Jr. and Cohen, M.S. (1997) Sexual transmission of HIV. *N Engl J Med*, **336**, 1072-1078.

Rugeles, M.T., Solano, F., Diaz, F.J., Bedoya, V.I. and Patino, P.J. (2002) Molecular characterization of the CCR 5 gene in seronegative individuals exposed to human immunodeficiency virus (HIV). *J Clin Virol*, **23**, 161-169.

Rybarczyk, B.J., Montefiori, D., Johnson, P.R., West, A., Johnston, R.E. and Swanstrom, R. (2004) Correlation between env V1/V2 region diversification and neutralizing antibodies during primary infection by simian immunodeficiency virus sm in rhesus macaques. *J Virol*, **78**, 3561-3571.

Saag, M.S. (2003) Antiretroviral therapy: select lessons from recent studies. *Top HIV Med*, **11**, 164-168.

Sagar, M., Lavreys, L., Baeten, J.M., Richardson, B.A., Mandaliya, K., Ndinya-Achola, J.O., Kreiss, J.K. and Overbaugh, J. (2004) Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. *Aids*, **18**, 615-619.

Saito, H., Tada, S., Ebinuma, H., Wakabayashi, K., Takagi, T., Saito, Y., Nakamoto, N., Kurita, S. and Ishii, H. (2001) Interferon regulatory factor 1 promoter polymorphism and response to type 1 interferon. *J Cell Biochem Suppl*, **Suppl 36**, 191-200.

Saito, H., Tada, S., Wakabayashi, K., Nakamoto, N., Takahashi, M., Nakamura, M., Ebinuma, H. and Ishii, H. (2002) The detection of IRF-1 promoter polymorphisms and their possible contribution to T helper 1 response in chronic hepatitis C. *J Interferon Cytokine Res*, **22**, 693-700.

Saksena, N.K. and Potter, S.J. (2003) Reservoirs of HIV-1 in vivo: implications for antiretroviral therapy. *AIDS Rev*, **5**, 3-18.

Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R.J., Collman, R.G., Doms, R.W., Vassart, G. and Parmentier, M. (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene [see comments]. *Nature*, **382**, 722-725.

Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noe, R., Peeters, M., Brookfield, J.F., Shaw, G.M., Sharp, P.M. and Hahn, B.H. (2005) Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J Virol*, **79**, 12515-12527.

Schaper, F., Kirchhoff, S., Posern, G., Koster, M., Oumard, A., Sharf, R., Levi, B.Z. and Hauser, H. (1998) Functional domains of interferon regulatory factor I (IRF-1). *Biochem J*, **335 (Pt 1)**, 147-157.

Schmitz, J.E., Johnson, R.P., McClure, H.M., Manson, K.H., Wyand, M.S., Kuroda, M.J., Lifton, M.A., Khunkhun, R.S., McEvers, K.J., Gillis, J., Piatak, M., Lifson, J.D., Grosschupff, G., Racz, P., Tenner-Racz, K., Rieber, E.P., Kuus-Reichel, K., Gelman, R.S., Letvin, N.L., Montefiori, D.C., Ruprecht, R.M., Desrosiers, R.C. and Reimann, K.A. (2005) Effect of CD8+ lymphocyte depletion on virus containment after simian immunodeficiency virus SIVmac251 challenge of live attenuated SIVmac239delta3-vaccinated rhesus macaques. *J Virol*, **79**, 8131-8141.

Schmitz, J.E., Kuroda, M.J., Santra, S., Sasseville, V.G., Simon, M.A., Lifton, M.A., Racz, P., Tenner-Racz, K., Dalesandro, M., Scallon, B.J., Ghrayeb, J., Forman, M.A., Montefiori, D.C., Rieber, E.P., Letvin, N.L. and Reimann, K.A. (1999) Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science*, **283**, 857-860.

Scott, P. (1998) Differentiation, regulation, and death of T helper cell subsets during infection with Leishmania major. *Immunol Res*, **17**, 229-238.

Sgarbanti, M., Borsetti, A., Moscufo, N., Bellocchi, M.C., Ridolfi, B., Nappi, F., Marsili, G., Marziali, G., Coccia, E.M., Ensoli, B. and Battistini, A. (2002) Modulation of human immunodeficiency virus 1 replication by interferon regulatory factors. *J Exp Med*, **195**, 1359-1370.

Shacklett, B.L. (2006) Understanding the "lucky few": the conundrum of HIV-exposed, seronegative individuals. *Curr HIV/AIDS Rep*, **3**, 26-31.

Sharf, R., Meraro, D., Azriel, A., Thornton, A.M., Ozato, K., Petricoin, E.F., Larner, A.C., Schaper, F., Hauser, H. and Levi, B.Z. (1997) Phosphorylation events modulate the ability of interferon consensus sequence binding protein to interact with interferon regulatory factors and to bind DNA. *J Biol Chem*, **272**, 9785-9792.

Shattock, R.J. and Moore, J.P. (2003) Inhibiting sexual transmission of HIV-1 infection. *Nat Rev Microbiol*, **1**, 25-34.

Shearer, G.M. and Clerici, M. (1996) Protective immunity against HIV infection: has nature done the experiment for us? *Immunol Today*, **17**, 21-24.

Shieh, B., Yan, Y.P., Ko, N.Y., Liau, Y.E., Liu, Y.C., Lin, H.H., Chen, P.P. and Li, C. (2001) Detection of elevated serum beta-chemokine levels in seronegative Chinese individuals exposed to human immunodeficiency virus type 1. *Clin Infect Dis*, **33**, 273-279.

Shimonovitz, S., Hurwitz, A., Dushnik, M., Anteby, E., Geva-Eldar, T. and Yagel, S. (1994) Developmental regulation of the expression of 72 and 92 kd type IV collagenases in human trophoblasts: a possible mechanism for control of trophoblast invasion. *Am J Obstet Gynecol*, **171**, 832-838.

Shrestha, S., Strathdee, S.A., Galai, N., Oleksyk, T., Fallin, M.D., Mehta, S., Schaid, D., Vlahov, D., O'Brien, S.J. and Smith, M.W. (2006) Behavioral risk exposure and host genetics of susceptibility to HIV-1 infection. *J Infect Dis*, **193**, 16-26.

Signorelli, D. and Joseph, R.J., 2nd. (1998) Epidemiology of HIV and AIDS. A retrospective look. *Clin Podiatr Med Surg*, **15**, 179-187.

Siliciano, R.F. (1999) Latency and reservoirs for HIV-1. *Aids*, **13 Suppl A**, S49-58.

Simmons, A., Aluvihare, V. and McMichael, A. (2001) Nef triggers a transcriptional program in T cells imitating single-signal T cell activation and inducing HIV virulence mediators. *Immunity*, **14**, 763-777.

Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F. and Brun-Vezinet, F. (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med*, **4**, 1032-1037.

Simonsen, J.N., Cameron, D.W., Gakinya, M.N., Ndinya-Achola, J.O., D'Costa, L.J., Karasira, P., Cheang, M., Ronald, A.R., Piot, P. and Plummer, F.A. (1988) Human immunodeficiency virus infection among men with sexually transmitted diseases. Experience from a center in Africa. *N Engl J Med*, **319**, 274-278.

Skurnick, J.H., Palumbo, P., DeVico, A., Shacklett, B.L., Valentine, F.T., Merges, M., Kamin-Lewis, R., Mestecky, J., Denny, T., Lewis, G.K., Lloyd, J., Praschunus, R., Baker, A., Nixon, D.F., Stranford, S., Gallo, R., Vermund, S.H. and Louria, D.B. (2002) Correlates of nontransmission in US women at high risk of human immunodeficiency virus type 1 infection through sexual exposure. *J Infect Dis*, **185**, 428-438.

Smith, D.M., Richman, D.D. and Little, S.J. (2005) HIV superinfection. *J Infect Dis*, **192**, 438-444.

Soorapanth, S., Sansom, S., Bulterys, M., Besser, M., Theron, G. and Fowler, M.G. (2006) Cost-effectiveness of HIV rescreening during late pregnancy to prevent mother-to-child HIV transmission in South Africa and other resource-limited settings. *J Acquir Immune Defic Syndr*, **42**, 213-221.

Soriano, A., Martinez, C., Garcia, F., Plana, M., Palou, E., Lejeune, M., Arostegui, J.I., De Lazzari, E., Rodriguez, C., Barrasa, A., Lorenzo, J.I., Alcami, J., del Romero, J., Miro, J.M., Gatell, J.M. and Gallart, T. (2002) Plasma stromal cell-derived factor (SDF)-1 levels, SDF1-3'A genotype, and expression of CXCR4 on T lymphocytes: their impact on resistance to human immunodeficiency virus type 1 infection and its progression. *J Infect Dis*, **186**, 922-931.

Sperling, R.S., Shapiro, D.E., Coombs, R.W., Todd, J.A., Herman, S.A., McSherry, G.D., O'Sullivan, M.J., Van Dyke, R.B., Jimenez, E., Rouzioux, C., Flynn, P.M. and Sullivan, J.L. (1996) Maternal viral load, zidovudine treatment, and the risk of transmission of human immunodeficiency virus type 1 from mother to infant. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. *N Engl J Med*, **335**, 1621-1629.

Sriwanthana, B., Hodge, T., Mastro, T.D., Dezzutti, C.S., Bond, K., Stephens, H.A., Kostrikis, L.G., Limpakarnjanarat, K., Young, N.L., Qari, S.H., Lal, R.B., Chandanayingyong, D. and McNicholl, J.M. (2001) HIV-specific cytotoxic T lymphocytes, HLA-A11, and chemokine-related factors may act synergistically to

determine HIV resistance in CCR5 delta32-negative female sex workers in Chiang Rai, northern Thailand. *AIDS Res Hum Retroviruses*, **17**, 719-734.

Stewart, G.J., Ashton, L.J., Biti, R.A., Ffrench, R.A., Bennetts, B.H., Newcombe, N.R., Benson, E.M., Carr, A., Cooper, D.A. and Kaldor, J.M. (1997) Increased frequency of CCR-5 delta 32 heterozygotes among long-term non- progressors with HIV-1 infection. The Australian Long-Term Non- Progressor Study Group. *Aids*, **11**, 1833-1838.

Stranford, S.A., Skurnick, J., Louria, D., Osmond, D., Chang, S.Y., Sninsky, J., Ferrari, G., Weinhold, K., Lindquist, C. and Levy, J.A. (1999) Lack of infection in HIV-exposed individuals is associated with a strong CD8(+) cell noncytotoxic anti-HIV response. *Proc Natl Acad Sci U S A*, **96**, 1030-1035.

Taha, T.E., Hoover, D.R., Dallabetta, G.A., Kumwenda, N.I., Mtimavalye, L.A., Yang, L.P., Liomba, G.N., Broadhead, R.L., Chiphangwi, J.D. and Miotti, P.G. (1998) Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *Aids*, **12**, 1699-1706.

Taki, S., Sato, T., Ogasawara, K., Fukuda, T., Sato, M., Hida, S., Suzuki, G., Mitsuyama, M., Shin, E.H., Kojima, S., Taniguchi, T. and Asano, Y. (1997) Multistage regulation of Th1-type immune responses by the transcription factor IRF-1. *Immunity*, **6**, 673-679.

Tanaka, N., Kawakami, T. and Taniguchi, T. (1993) Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. *Mol Cell Biol*, **13**, 4531-4538.

Tang, H., Kuhen, K.L. and Wong-Staal, F. (1999) Lentivirus replication and regulation. *Annu Rev Genet*, **33**, 133-170.

Taniguchi, T., Ogasawara, K., Takaoka, A. and Tanaka, N. (2001) IRF family of transcription factors as regulators of host defense. *Annu Rev Immunol*, **19**, 623-655.

Tindall, B. and Cooper, D.A. (1991) Primary HIV infection: host responses and intervention strategies. *Aids*, **5**, 1-14.

Tovanabutra, S., Robison, V., Wongtrakul, J., Sennum, S., Suriyanon, V., Kingkeow, D., Kawichai, S., Tanan, P., Duerr, A. and Nelson, K.E. (2002) Male viral load and heterosexual transmission of HIV-1 subtype E in northern Thailand. *J Acquir Immune Defic Syndr*, **29**, 275-283.

Trivedi, H.N., Plummer, F.A., Anzala, A.O., Njagi, E., Bwayo, J.J., Ngugi, E.N., Embree, J.E. and Hayglass, K.T. (2001) Resistance to HIV-1 infection among African sex workers is associated with global hyporesponsiveness in interleukin 4 production. *Faseb J*, **15**, 1795-1797.

Turville, S.G., Cameron, P.U., Handley, A., Lin, G., Pohlmann, S., Doms, R.W. and Cunningham, A.L. (2002) Diversity of receptors binding HIV on dendritic cell subsets. *Nat Immunol*, **3**, 975-983.

Tzoanopoulos, D., Speletas, M., Arvanitidis, K., Veiopoulou, C., Kyriaki, S., Thyphronitis, G., Sideras, P., Kartalis, G. and Ritis, K. (2002) Low expression of interferon regulatory factor-1 and identification of novel exons skipping in patients with chronic myeloid leukaemia. *Br J Haematol*, **119**, 46-53.

Uegaki, K., Shirakawa, M., Harada, H., Taniguchi, T. and Kyogoku, Y. (1995) Secondary structure and folding topology of the DNA binding domain of interferon regulatory factor 2, as revealed by NMR spectroscopy. *FEBS Lett*, **359**, 184-188.

UNAIDS/WHO. (2005) Global summary of the AIDS epidemic : December 2005.

Walensky, R.P., Paltiel, A.D., Losina, E., Mercincavage, L.M., Schackman, B.R., Sax, P.E., Weinstein, M.C. and Freedberg, K.A. (2006) The survival benefits of AIDS treatment in the United States. *J Infect Dis*, **194**, 11-19.

Walker, C.M., Moody, D.J., Stites, D.P. and Levy, J.A. (1986) CD8+ lymphocytes can control HIV infection in vitro by suppressing virus replication. *Science*, **234**, 1563-1566.

Wawer, M.J., Gray, R.H., Sewankambo, N.K., Serwadda, D., Li, X., Laeyendecker, O., Kiwanuka, N., Kigozi, G., Kiddugavu, M., Lutalo, T., Nalugoda, F., Wabwire-Mangen, F., Meehan, M.P. and Quinn, T.C. (2005) Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis*, **191**, 1403-1409.

Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., Komarova, N.L., Nowak, M.A., Hahn, B.H., Kwong, P.D. and Shaw, G.M. (2003) Antibody neutralization and escape by HIV-1. *Nature*, **422**, 307-312.

Weisz, A., Kirchhoff, S. and Levi, B.Z. (1994) IFN consensus sequence binding protein (ICSBP) is a conditional repressor of IFN inducible promoters. *Int Immunol*, **6**, 1125-1131.

Weisz, A., Marx, P., Sharf, R., Appella, E., Driggers, P.H., Ozato, K. and Levi, B.Z. (1992) Human interferon consensus sequence binding protein is a negative regulator of enhancer elements common to interferon-inducible genes. *J Biol Chem*, **267**, 25589-25596.

White, L.C., Wright, K.L., Felix, N.J., Ruffner, H., Reis, L.F., Pine, R. and Ting, J.P. (1996) Regulation of LMP2 and TAP1 genes by IRF-1 explains the paucity of CD8+ T cells in IRF-1-/- mice. *Immunity*, **5**, 365-376.

Winkler, C., Modi, W., Smith, M.W., Nelson, G.W., Wu, X., Carrington, M., Dean, M., Honjo, T., Tashiro, K., Yabe, D., Buchbinder, S., Vittinghoff, E., Goedert, J.J., O'Brien, T.R., Jacobson, L.P., Detels, R., Donfield, S., Willoughby, A., Gomperts, E., Vlahov, D., Phair, J. and O'Brien, S.J. (1998) Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). *Science*, **279**, 389-393.

Wrin, T., Crawford, L., Sawyer, L., Weber, P., Sheppard, H.W. and Hanson, C.V. (1994) Neutralizing antibody responses to autologous and heterologous isolates of human immunodeficiency virus. *J Acquir Immune Defic Syndr*, **7**, 211-219.

Yamashita, M. and Emerman, M. (2006) Retroviral infection of non-dividing cells: old and new perspectives. *Virology*, **344**, 88-93.

Yang, X., HayGlass, K.T. and Brunham, R.C. (1996) Genetically determined differences in IL-10 and IFN-gamma responses correlate with clearance of Chlamydia trachomatis mouse pneumonitis infection. *J Immunol*, **156**, 4338-4344.

Yao, X.J., Mouland, A.J., Subbramanian, R.A., Forget, J., Rougeau, N., Bergeron, D. and Cohen, E.A. (1998) Vpr stimulates viral expression and induces cell killing in human immunodeficiency virus type 1-infected dividing Jurkat T cells. *J Virol*, **72**, 4686-4693.

Yao, X.J., Subbramanian, R.A., Rougeau, N., Boisvert, F., Bergeron, D. and Cohen, E.A. (1995) Mutagenic analysis of human immunodeficiency virus type 1 Vpr: role of a predicted N-terminal alpha-helical structure in Vpr nuclear localization and virion incorporation. *J Virol*, **69**, 7032-7044.

Zagury, D., Lachgar, A., Chams, V., Fall, L.S., Bernard, J., Zagury, J.F., Bizzini, B., Gringeri, A., Santagostino, E., Rappaport, J., Feldman, M., O'Brien, S.J., Burny, A. and Gallo, R.C. (1998) C-C chemokines, pivotal in protection against HIV type 1 infection. *Proc Natl Acad Sci U S A*, **95**, 3857-3861.

Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M. and Ho, D.D. (1998) An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature*, **391**, 594-597.

Zimmerman, P.A., Buckler-White, A., Alkhatib, G., Spalding, T., Kubofcik, J., Combadiere, C., Weissman, D., Cohen, O., Rubbert, A., Lam, G., Vaccarezza, M., Kennedy, P.E., Kumaraswami, V., Giorgi, J.V., Detels, R., Hunter, J., Chopek, M., Berger, E.A., Fauci, A.S., Nutman, T.B. and Murphy, P.M. (1997) Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Mol Med*, **3**, 23-36.

Appendix 1

Abbreviations

AIDS: Acquired Immunodeficiency Syndrome

APC: Antigen Presenting Cells

APS: Ammonium Persulfate

CCR: Chemokine Receptor

CD4: Cluster Differentiation molecule 4

CD8: Cluster Differentiation molecule 8

CDC: Center for Disease Control, the United States of America

cDNA: Complementary Deoxyribonucleic Acid

CI: Confidence Interval

cpm: Counts per Minute

CSF: Colony Stimulation Factor

CSW: Commercial Sex Worker

CTL: Cytotoxic T Lymphocyte

CXCR: CXC Chemokine Receptor

D10: DMEM medium with 10% fetal calf serum

dbSNP: Database for Single Nucleotide Polymorphisms

DC: Dendritic Cells

DC-SIGN: DC Specific Intercellular-adhesion-molecule-3 Grabbing Nonintegrin

DEA: Diethanolamine

DNA: Deoxyribonucleic Acid

dNTP: Deoxyribonucleotides

dsDNA: Double stranded DNA

dsRNA: Double stranded RNA

DTT: Dithiotheritol

EDTA: Ethylenediamine Tetraacetic Acid

EGTA: Ethylene Glycol bis(2-aminoethyl ether)-N,N,N'N'-Tetraacetic Acid

ELISA: Enzyme-Linked Immunosorbent Assay

Env: Envelope

ESN: Exposed Seronegative

ESP: Exposed Seronegative Person

EtBr: Ethidium Bromide

EtOH: Ethanol

EU: Exposed Uninfected

FCS: Fetal Calf Serum

Gag: Group Specific Antigen

GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase

GAS: IFN-γ-activated Sequence

gp: Glycoprotein

HeBS: Hepes Buffer Saline

HEPS: Highly Exposed Persistently Seronegative

HIV: Human Immunodeficiency Virus

HIV-R: HIV-1 resistance

HLA: Human Leukocyte Antigen

HRP: Horseradish Peroxidase

HRSN: High-Risk Seronegative

HTLV: Human T-cell Leukemia Virus

IFN: Interferon

Ig: Immunoglobulin

IL: Interleukin

IN: Integrase

IRF: Interferon Regulator Factor

ISRE: Interferon-Stimulated Response Element

Kb: Kilobase

LD: Linkage Disequilibrium

LMP-2: Low Molecular weight Protein-2

LTNP: Long Term Non-Progressor

LTR: Long Terminal Region

Luc : Luciferase

mg: Miligram

MHC: Major Histocompatibility Complex

MIP: Macrophage Inflammatory Protein

ml : Mililiter

mM : Milimole

mRNA : Messenger RNA

MS: Microsatellite

MW: Molecular Weight

Nef : misnamed Negative Regulator Factor

NF-κB: Nuclear Factor κB

Ng: Nanogram

NK: Natural Killer cell

NLS: Nuclear Location signaling

nt: Nucleotide

OR: Odd Ratio

P24: HIV-1 viral protein with molecular weight at 24 kilo Dolton

PBMC : Peripheral Blood Mononuclear Cells

PBS : Phosphate Buffer Saline

PBS-T : Phosphate Buffer Saline with Tween-20

PCP: Pneumocistis Carinii Pneumonia

PCR: Polymerase Chain Reaction

PIGR: Polymeric Immunoglobulin Receptor

Pol: Polymerase

Pro: Proteinase

R10: RPMI1640 with 10% fetal calf serum

R20: RPMI1640 with 20% fetal calf serum

RANTES: Regulated upon Activation, Normal T-cell Expressed and Secreted

Rev: Regulator of Expression of Virus protein

RNA: Ribonucleic Acid

rpm: Revolutions Per Minute

RT: Reverse Transcriptase

RT-PCR : Reverse Transcription-Polymerase Chain Reaction

SAAP: Streptavidin-Alkphosphatase

SDF-1: Stromal cell-derived Factor 1

SDS: Sodium Dodecyl Sulfate

SDS-PAGE: Sodium Dodecyl (lauryl) Sulfate-Polyacrylamide Gel Electrophoresis

SIV: Simian Immunodeficiency Virus

SNP: Single Nucleotide Polymorphism

ssDNA: Single Stranded DNA

ssRNA: Single Stranded RNA

TAP-1: Transporter Associated with Antigen Presentation

Taq: Taq DNA polymerase

Tat: Transactivator protein

TBE: Tris Base EDTA Buffer

TCA: Trichloroacetic acid

TE: Tris-HCl EDTA

Th1: T Helper type-1

Th2: T Helper type-2

TLR: Toll-like Receptor

Tm: Annealing temperature

TNF: Tumor Necrosis Factor

Tris-HCL: Tris Hydrochloride

UTR: Untranslated Region

Vif: Virus Infectivity Factor

Vpr: Viral Protein R

Vpu: Viral Protein U

VSV-G: Vesicular Stomatitis Virus Glycoprotein

WB: Western Blot

β-2-ME : β-2-Mercaptoethanol

μg: Microgram

μl: Microliter

Appendix 2

Paper/manuscripts arise from this project:

1. **Ji H.**, Ball T.B., Kimani J. & Plummer F.A. Novel interferon regulatory factor-1 polymorphisms in a Kenyan population revealed by complete gene sequencing. J Hum Genet 49, 528-35 (2004).

2. **Ji H.**, Ball T.B., Liang B, Kimani J. & Plummer F.A. Human Interferon Regulatory Factor-1 gene and its promoter sequences Revealed by Population-based Complete Gene Sequencing (2006, submitted to **DNA Sequence**)

3. **Ball T.B.*, Ji H***, Kimani J., McLaren P., Marlin C., Hill, A.V., and Plummer F.A. Polymorphisms in interferon regulatory factor 1 associate with resistance to infection by Human Immunodeficiency Virus-1 (*Co-first authorship) (2006, Submitted to **AIDS**)

4. **Ji H**, Ball T.B., Ao Z., Kimani J, Yao X., and Plummer F.A. Varied initiation of HIV-1 transcription: a potential mechanism mediating association between IRF-1 polymorphisms and HIV-1 resistance (**In preparation**).

5. **Ji H**, Ball T.B., Mao X., Liang B., Kimani J., and Plummer F.A. Altered exon skipping patterns: another potential mechanism mediating association between IRF-1 polymorphisms and resistance to HIV-1 (**In preparation**).