Evaluation of Viral Genes and Host Immune Factors in HIV-1 infected patients

Submitted by:

Taskeen Aman

(NUST-00000203711)

Supervisor:

Dr. Aneela Javed

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

2019

Evaluation of Viral Genes and Host Immune Factors in HIV-1 infected patients

A Thesis submitted in partial fulfillment of requirement for the degree of

Master of Sciences

In

Healthcare Biotechnology

By

Taskeen Aman

(NUST-00000203711)

Supervised by:

Dr. Aneela Javed

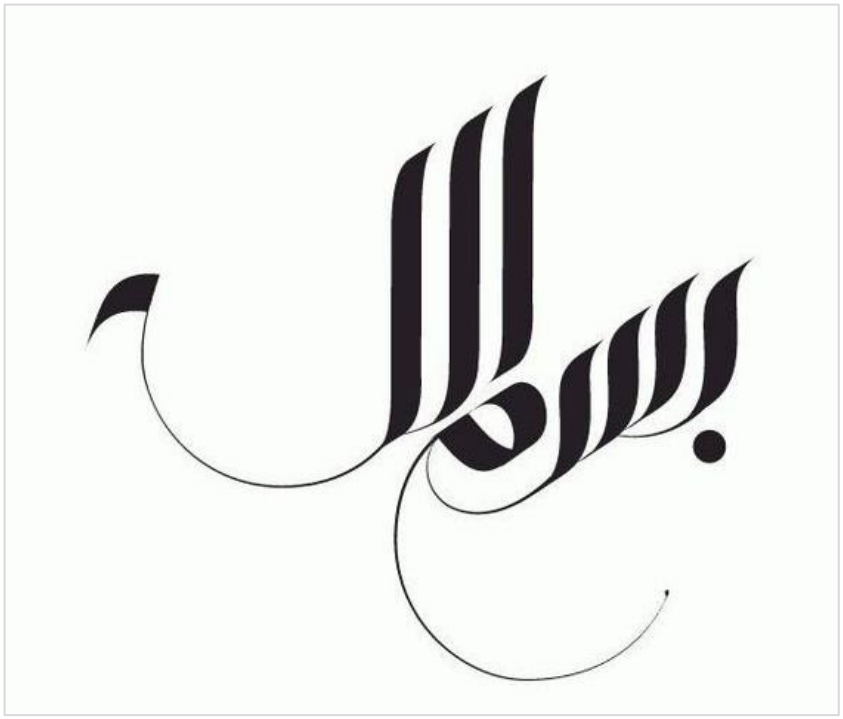
Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Islamabad-Pakistan

2019

TH-FORM


بسم الله

Dedicated to

Ammi & Baba ♡

ACKNOWLEDGMENTS

إِيَّاكَ نَعْبُدُ وإِيَّاكَ نَسْتَعِينُ

Praise be to Allah Almighty who is the most beneficent, the most merciful! Verily, Allah is the ultimate source of all knowledge, the One who has picked me up after every fall, the One who has granted me so much more than I deserve, and the only One who has instilled in me the spirit and passion to seek knowledge and pursue my dreams despite all the odds.

I would like to thank my mentor (from the past 6 years), the best teacher, and an expert supervisor – Dr. Aneela Javed. I owe all my success to her for she has shaped me into a professional and taught me the skills I will need in future. I hope I can make her very proud one day! Next, on an equally special note, I would like to thank Dr. Salik J. Kakar (my co-supervisor). He has been my pillar of support and a moral compass throughout the tough years of Masters. I have found in him an excellent guide, an amazing critic and a great confidante. The ease and comfort that I have been provided by my supervisors during Masters Research was 'unparalleled' and I consider myself lucky to work under their supervision. I would also like to thank Principal ASAB – Dr. Hussnain A. Janjua, HOD HCB – Dr. Touqeer Ahmed, esteemed faculty, staff members and lab fellows (especially Walifa Waqar) for all the support.

Last, but the most important, I owe all of my success and growth to my parents, friends (especially Shiza Navaid) and family. Their unconditional love and immense sacrifices put me where I am today, and I can't be grateful enough. I would conclude by pledging a solemn oath to serve my nation in the best of my abilities, with honesty, integrity and justice always in favor of my country – Pakistan!

Taskeen Aman.

Table of Contents

LIST OF ACRONYMS X

LIST OF FIGURES XII

LIST OF TABLES XIV

ABSTRACT XV

INTRODUCTION 16

HIV and AIDS 1

Viral and Host Factors influencing HIV Pathogenesis 3

AIMS AND OBJECTIVES 5

LITERATURE REVIEW 6

HIV Epidemiology 7

Origin & Aetiology of HIV 10

Viral Classification 11

HIV Subtyping 12

HIV Structure and Genome 13

HIV Gene Functions 15

Host Factors that influence Disease progression 16

Innate and Adaptive Immunity Factors 17

HIV Dependency Factors 19

Intrinsic anti-HIV restriction entities 21

Role of Host protein FAM26F in HIV Disease Progression 23

MATERIALS & METHODS 26

Formulation of Study Design 27

Sampling and Ethical Consent 28

Separation of Plasma and PBMCs from whole blood 29

Assessment of Viral Factors 29

RNA Extraction 29

Viral Load Quantification 31

cDNA Synthesis 31

Primers Designing and Optimization 33

Amplification of HIV genes via Nested Polymerase Chain Reaction 35

Gel Electrophoresis 36

PCR Product Purification for Sequencing 37

Assessment of Host Factors 38

Nucleic Acid Extraction 38

DNA Purification 40

RNA Purification 40

Nucleic Acid Quantification 41

cDNA Synthesis 41

FAM26F and Thioredoxin Primer Optimization - Amplification via Conventional Polymerase Chain Reaction 42

Primer Optimization and Primer Efficiency Analysis 46

Amplification via Real Time PCR (FAM/TXN/GAPDH/B-Actin) 48

Gel Electrophoresis 48

RESULTS 49

Patient Demographic Details 50

Separation of Blood Plasma and Peripheral Blood Mononuclear Cells 53

Viral Load Analysis 53

Assessment of Viral Factors 55

Amplification of HIV genes 55

Sequencing of Amplified Viral Genes 57

Assessment of Host Factors 58

Expression Analysis of Host genes 58

Analysis of FAM26F and Thioredoxin (TXN) gene expression in HIV positive patients with respect to Controls 64

DISCUSSION 69

FUTURE PROSPECTS 73

REFERENCES 75

LIST OF ACRONYMS

HLA	Human Leukocyte Antigen
SNP	Single Nucleotide Polymorphism
HIV-1	Human Immunodeficiency Virus Type 1
SIV	Simian Immunodeficiency Virus
FAM26F	Family with sequence similarity 26, member F
AIDS	Acquired Immunodeficiency Syndrome
PCR	Polymerase Chain Reaction
qPCR	Quantitative PCR (Real time PCR)
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
CD4	Cluster with Differentiation 4
CD4 cells	CD4+ T lymphocytes
IDUs	Injection Drug Users
HSWs	Hijra (Transgender) Sex Workers
MSMs	Men having sex with men
FSWs	Female sex workers
EXP	Expatriates/ Immigrants
SP	Spouse
MSP	Multiple Sexual Partners
ART	Antiretroviral Therapy
CCR5	C-C Chemokine Receptor 5

CXCR4	C-X-C Chemokine Receptor type 4
PIMS	Pakistan Institute of Medical Sciences
CRFs	Circulating Recombinant Forms
KIRs	Killer Cell Immunoglobulin like receptors
TXN	Thioredoxin gene
Trx	Thioredoxin protein

LIST OF FIGURES

Figure No.	Title of figures	Page No.
1	Prevalence of HIV-1 Worldwide; 2017 statistics	07
2	Global Progress toward 90-90-90 HIV-1 related targets.	08
3	Prevalence and Concentration groups of HIV-1 in Pakistan	09
4	Origins of HIV-1 and AIDS viruses	10
5	HIV Genome Structure	15
6	Host Factors involved in HIV Pathogenesis	17
7	Transmission Routes of HIV-1	18
8	Host Factors involved in HIV-1 Pathogenesis	20
9	Interactive Partner Proteins of FAM26F.	24
10	Structure of FAM26F protein	24
11	Schematic flow diagram of the methodology of study.	27
12	Conventional PCR Cycles and Setup.	36
13	RNA Extraction from PBMCs - protocol.	39
14	Conventional PCR Cycles and Setup (for FAM26F amplification).	44
15	Conventional PCR Cycles and Setup (For Thioredoxin amplification).	46
16	Age Distribution in HIV Positive patients.	50
17	Gender Distribution in HIV positive patients	51
18	Viral Transmission distribution in HIV positive patients	51
19	CD4+ count distribution in HIV positive patients involved in this study	52
20	Collection of HIV positive blood from patients and initial processing using Ficoll Hypaque	53

21	Visual Representation of amplified viral genes (gag and pol) in cDNA samples of patients (S1 and S2) extracted from plasma and in positive control (P) using NESTED PCR.	56
22	Visual Representation of amplified cDNA samples of patients extracted from plasma using NESTED PCR.	57
23	Representative Chromatogram of amplified pol gene generated via Sanger Sequencing (using forward primer).	58
24	Representative Figure: Shade Identities and Similarities in Pairwise and Multiple Alignment Window	58
25	Visual representation of verification of amplified cDNA products using Beta Actin primers	60
26	Optimization and Primer Efficiency Analysis of FAM26F gene.	61
27	Optimization and Primer Efficiency Analysis of Thioredoxin gene.	61
28	Representative amplification plot of qPCR (Real time PCR).	62
29	Representative Melt Curves of FAM26F, Thioredoxin and Beta Actin	63
30	Copies of FAM26F gene in healthy controls vs. HIV positive patients.	64
31	Copies of Thioredoxin gene in healthy controls vs. HIV positive patients.	65
32	Fold Change of FAM26F gene in healthy controls vs. HIV positive patients.	66
33	Fold Change of Thioredoxin gene in healthy controls vs. HIV positive patients	66
34	Correlation of copy numbers of FAM26F and Thioredoxin genes in healthy controls vs. HIV positive patients	67
35	Correlation of fold change difference of FAM26F and Thioredoxin genes in healthy controls vs. HIV positive patients.	68

LIST OF TABLES

Table No.	Title of tables	Page No.
1	Role of SNPs in APOBEC3 gene role in HIV-1 Pathogenesis.	22
2	cDNA preparation from RNA extracted from plasma (I).	31
3	cDNA preparation from RNA extracted from plasma (II).	32
4	Conventional PCR setup.	34
5	Conventional PCR Recipe.	35
6	PCR Product Purification Recipe.	37
7	cDNA synthesis from RNA extracted from PBMCs (I).	41
8	cDNA synthesis from RNA extracted from PBMCs (II).	42
9	FAM26F primer sequences.	43
10	Conventional PCR Recipe (for FAM26F amplification).	43
11	Thioredoxin Primer Sequences.	45
12	Conventional PCR Recipe (For Thioredoxin amplification).	45
13	Sequences of FAM26F and Thioredoxin Primers.	46
14	Real-time PCR (qPCR recipe) for FAM26F and Thioredoxin amplification.	47
15	Viral loads of obtained blood samples of HIV positive patients.	55
16	Representation of RNA concentrations of patient samples	59
17	Representation of RNA concentrations of control samples	59

ABSTRACT

HIV-1 infection has rapidly spread worldwide infecting more than 37.9 million people to date, making HIV-1 one of the leading causes of mortality in infectious diseases. Pakistan, too, has faced a shift in HIV status from 'low prevalence, high risk' to 'concentrated' epidemic. Negligent healthcare management and research in Pakistan has added to the plethora of disease due to insufficient modes of diagnosis and awareness about disease, improper management of HIV positive patients and unsatisfactory genotyping and screening of population-specific virus. Population specific HIV-1 remains unexplored in Pakistan leading to ineffective drug regimen, evolution of more virulent and drug resistant strains and ultimately a rise in HIV-1 status in the country. Up-to-date, in Pakistan only partial sequence characterization of HIV-1 has been performed leaving a huge question mark on the actual number of circulating subtypes, recombinants and modes of their transmission. There is a severe need to amplify, sequence and analyze whole population-specific HIV-1 in Pakistan. This study, has therefore, primarily focused on amplification and sequencing of HIV-1 genome within infected Pakistani patients. In addition to viral factors in disease pathogenesis, host factors also play a crucial role. Several host proteins either protect or render host more susceptible to HIV-1 infection. In quest of host factors that shield host from HIV-1 infection, differential expression of a relatively recently identified host protein FAM26F (family with sequence similarity 26, member F) and Thioredoxin have been checked. Differential role of these host factors in other infections has previously been established. This study has focused on the role of these host factors in HIV Pathogenesis. As a result of this study, viral genes (gag and pol) were successfully amplified. Selective host immune factors i.e. FAM26F and Thioredoxin genes were amplified and were found to express differentially in HIV positive patients in comparison to controls.

INTRODUCTION

HIV and AIDS

Human Immunodeficiency Virus (HIV) is a retrovirus belonging to lentivirus genus that causes the infamous acquired immunodeficiency syndrome (AIDS) (Blattner, Gallo and Temin 1988). There are intrinsically two types of HIV that ae responsible for causing infections in humans, i.e. HIV-1 and HIV-2. They mainly differ in their virulence and global prevalence. Majority of the reported AIDS cases worldwide have been attributed to HIV- type 1. Infection with HIV normally occurs by several modes i.e. transfer of infected blood, vaginal fluid, breast milk and other fluids within the body (Royce, et al. 1997) (Burgener, McGowan and Klatt 2015). HIV resides as either free virus particles or as a virus within infected cells within body fluids. Major routes of transmission of HIV are unprotected sexual intercourse, contaminated needles and transmission from an infected mother to baby during or after birth (through breast milk) (Read, et al. 1999).

HIV-1 has four distinct lineages through which it transferred to humans from the counterpart primates. The distinct HIV-1 lineages are group M, N, O and P. Of these groups, the most significantly occurring and persisting lineage worldwide has been the M group of HIV-1, basically accountable for the immense majority of human infections (Burke 1997). The M group can further can be classified into 9 subtypes such as A to D, E to H, J and K subtypes along with sub-subtypes (A1, A2, F1 and F2), circulating recombinant forms (CRFs) and unique recombinant forms (URFs); others such as O group has O subtype, N group has N subtype. These HIV-1 subtypes are crucial in respect of determining virus' etiology, epidemiology, laboratory diagnosis, clinical symptoms, drug screening and vaccine development techniques (Hemelaar, Elangovan, et al. 2019).

The prime target of HIV in body is the human immune system and its cells i.e. helper T cells (specifically CD4+). A body infected with HIV contains lower levels of CD4+ T cells in comparison to healthy controls. After the decline of these cells below a critical threshold, cell-mediated immunity is lost, and body develops progressive susceptibility to opportunistic infections up until body loses its potential to fight off pathogens and develops AIDS (acquired immunodeficiency syndrome) (Blattner, Gallo and Temin 1988). The most common infections developed at this stage include chronic cryptosporida diarrhea, cytomegalovirus-induced eye infection, pneumocystis pneumonia, toxoplasmosis, and tuberculosis as well as infections with members of the *Mycobacterium avium* complex (DiPerri, et al. 1989) (Fenner, et al. 1991) (Fisseh, Petros and WoldeMichael 1998) (Pereira-Chioccola, Vidal and Su 2009). In addition, the development of different types of cancer, such as invasive cervical cancer, Kaposi sarcoma, or lymphoma, is frequently observed. The disease is, at present, incurable.

HIV/ AIDS is a global phenomenon. The HIV/AIDS pandemic continues exacting an enormous toll globally, claiming more than 0.9 million lives in the year 2018 and a total of 35 million lives since AIDS was first recognized more than three decades ago (World Health Organization 2018). The statistics are daunting especially after knowing that the disease (owing to its mechanism of transmission) is a taboo in several countries around the globe and most cases are underrepresented. According to WHO statistics of 2018, 36.9 million individuals were living with HIV/AIDS by the end of 2017 while 1.8 million people became newly infected individuals in 2017 globally (Avert 2018). In Pakistan alone, there

were approximately 0.15 million people living with HIV infection by the end of 2017 (The World Bank 2012).

Viral and Host Factors influencing HIV Pathogenesis

The massive genetic variability of HIV-1, resulting primarily from the error prone nature of its reverse transcriptase (possessing 361025 mutations per nucleotide per replication cycle) and high rate of mutation and the rapid turnover of HIV-1 in vivo (1,036,109 particles per day) have been the prime causes of emergence of selective (to host) yet advantageous genetic alterations in the viral population over time (Vijay, et al. 2008). These changes, over the course of time, have made HIV-1 victorious into resisting the evolving host defense mechanisms. Viral recombination is another evolutionary source that has contributed to the genetic diversification of the virus and has made HIV more virulent, drug resistant, containing altered cell tropism over the period of its emergence. These viral factors possess huge prospects of compromising the efficacy of antiretroviral therapy and naturally present huge challenges to scientists and researchers for better vaccine design (Burke 1997).

Alongside viral factors, host factors such as innate and adaptive host factors, HIV dependency factors and intrinsic anti-virus restriction entities are also responsible for an altered (increased or decreased) HIV progression rate in humans (McLaren and Carrington 2015). Several polymorphisms in key receptor molecules (such as CCR5 and CXCR4) alter host's response to HIV infection. The one and only mutation in human CCR5 protein that has been documented up till date to completely protect individuals of HIV infection is a 32 base pair deletion that

leaves CCR5 receptor protein truncated and ultimately does not allow the virus to enter. On similar grounds, other mutations and polymorphisms leave individuals and certain populations more prone to HIV infection (DARC null genotype in African population) (An and Winkler 2010). FAM26F, on similar grounds, is a relatively recently discovered protein involved in various inflammatory pathways. The membrane protein has a major conserved domain of Ca_hom_mod and is speculated to play a role in inflammation, pathogenesis, and infection in several diseases that influence immune system (U. Malik, et al. 2017). Upregulation of FAM26F was reported in SIV infected rhesus models, and other infections. Role of FAM26F in HCV viral clearance was reported (Malik and Javed 2016).

Thioredoxin belongs to a class of small redox proteins (Nakamura, et al. 1996). Like FAM26F, Thioredoxin protein (Trx) is also a calcium modulator and has an immunological role of calcium regulation (Kohárýová M 2008). Upregulation of Trx has been reported in HIV pathogenesis and several cancers (Benhar, et al. 2016).

AIMS AND OBJECTIVES

The objectives of this study are therefore to;

- Amplify and Sequence HIV genes (*Pol* and *Gag*) of the virus specific to Pakistani population
- Check the expression of host factors FAM26F (family with sequence similarity 26, member F) and Thioredoxin (Trx) in HIV infected patients.

LITERATURE REVIEW

HIV Epidemiology

HIV continues to be the global public health concern. According to global 2017 statistics, 36.9 million people have been affected by HIV among whom 1.8 million are children (<15 years old). As the virus remains in dormant form in most cases, 25% of the globally affected HIV population remains unaware of their condition (Avert 2018). HIV has affected 77.3 million individuals since the beginning of this epidemic, and has taken lives of an approximate 35.4 million individuals globally. Majority of individuals living with HIV are located in low income and middle income countries, nonetheless the developed countries are also not completely free of the virus. 66% of the globally affected HIV patients happen to reside in sub-Saharan African countries (World Health Organization 2018). Therefore, a direct linkage of prevalent HIV status with socioeconomic status of countries is validated and needs to be worked upon.

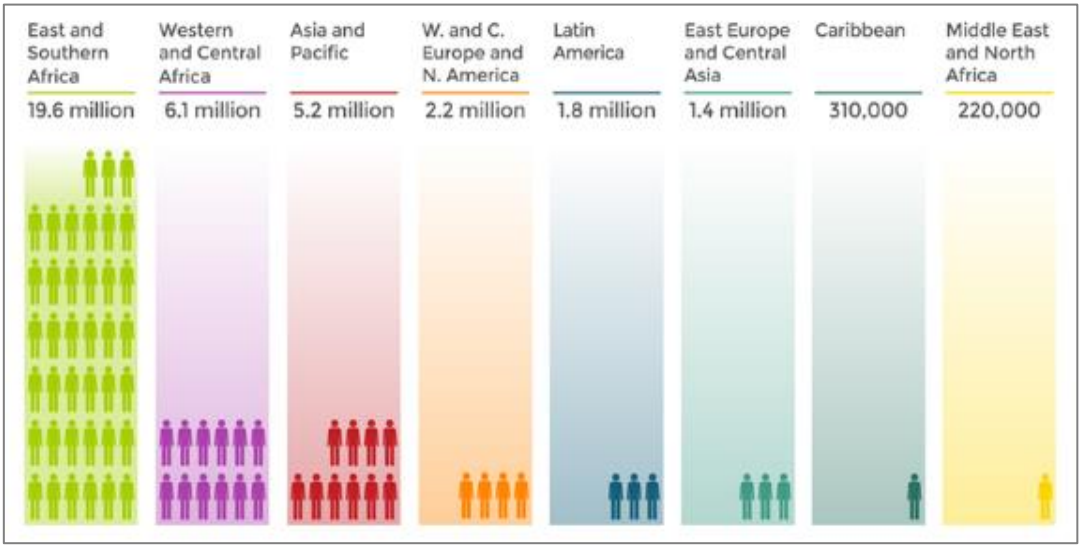


Figure 1: Prevalence of HIV-1 Worldwide; 2017 statistics *(Avert 2018)*

The annual number of new HIV infections globally has remained somewhat static over the past years. In 2017, the annual number of new HIV infections was 1.8 million, which is the same as it was in 2016. In the past seven years however, the global number of new HIV infections has declined consistently by 18%, with around 2.2 million new infections in 2010 and 1.8 million newly infected in 2017. This trend brings the world one step closer to the 90-90-90 target set by UNAIDS. This ambitious treatment target to help the World by eradicating the virus works if by 2020 90% of all HIV positive individuals are diagnosed, the 90% of those diagnosed are provided antiretroviral therapy (ART), and among those 90% of treated individuals successfully achieve viral suppression. The real-time progress toward this goal however has been marginally satisfactory as until 2017, 75% individuals knew HIV status, with further 79% and 81% getting treatment and successfully suppressing the virus respectively (Hemelaar, Elangovan, et al. 2019).

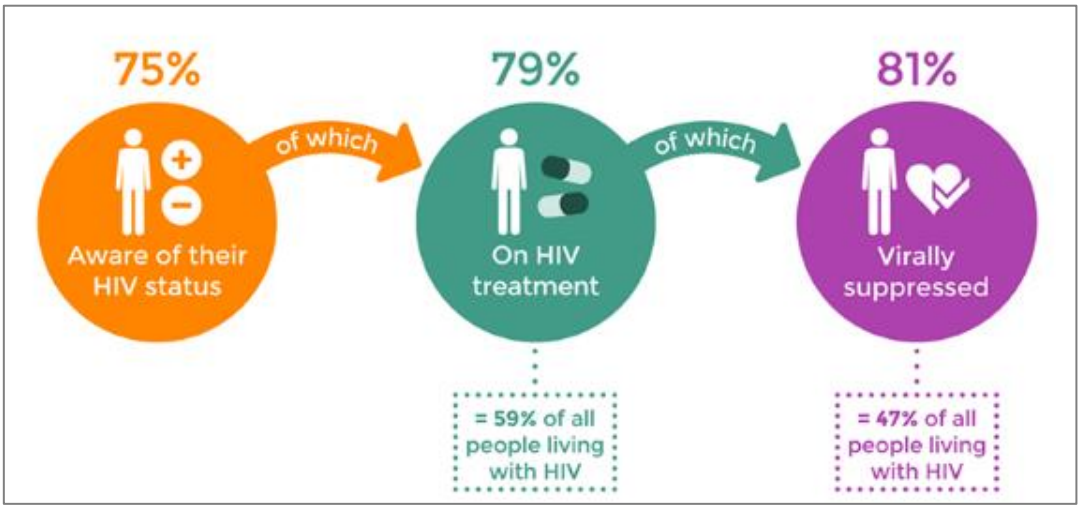
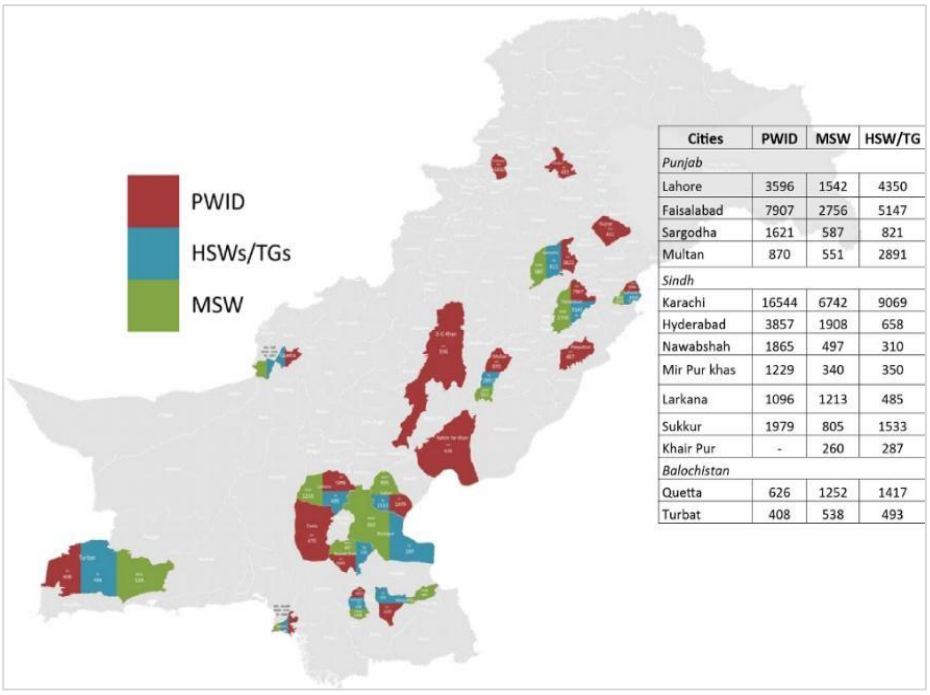


Figure 2: Global Progress toward 90-90-90 HIV-1 related targets. *(Avert 2018)*

Pakistan, being one of the developing economies as stated by United Nations, too faces rising HIV and AIDS concern as the country has faced a shift in HIV status

from ‘low-prevalence, high risk’ to concentrated epidemic. Around 0.1% of the adult Pakistani population is HIV positive and the trend seems to be rising since 1990s. The HIV and AIDS epidemic in Pakistan is particularly concentrated among Injection drug users (IDUs) and sex workers. Unlike in other parts of the world, the other key population groups are insignificant in Pakistan as compared to the IDUs and sex workers (The World Bank 2012). There are four major key population HIV-risk groups in Pakistan. These include Injection Drug Users (IDUs) followed by Hijra Sex Workers (HSWs), Men-who have-Sex with Men (MSMs) and Female Sex Workers (FSWs).



Cities	PWID	MSW	HSW/TG
Punjab			
Lahore	3596	1542	4350
Faisalabad	7907	2756	5147
Sargodha	1621	587	821
Multan	870	551	2891
Sindh			
Karachi	16544	6742	9069
Hyderabad	3857	1908	658
Nawabshah	1865	497	310
Mir Pur khas	1229	340	350
Larkana	1096	1213	485
Sukkur	1979	805	1533
Khair Pur	-	260	287
Balochistan			
Quetta	626	1252	1417
Turbat	408	538	493

Figure 3: Prevalence and Concentration groups of HIV-1 in Pakistan *(CCM Pakistan 2018).*

Origin & Aetiology of HIV

One of the earliest reported HIV/AIDS episode was discovered in a preserved blood sample of a man from Belgian Congo. The virus' origin has been explained by several theories, but amongst all the hunter theory seems most legitimate. Like HIV that is a lentivirus and attacks the immune system of the host, SIV (Simian Immunodeficiency Virus) also attacks the immune system of its host (monkeys and chimpanzees (Hirsch, et al. 1995). In 1999, researchers found a linkage between HIV and SIV that convinced them to do further research. A strain of SIV (SIVcpz) in a chimpanzee was found almost completely identical to HIV in humans. The transfer of this strain of virus was later explained via the hunter theory proposing that hunters in Africa who attempted to hunt down chimpanzees and monkeys as a meat source acquired the virus and transfected to fellow humans.

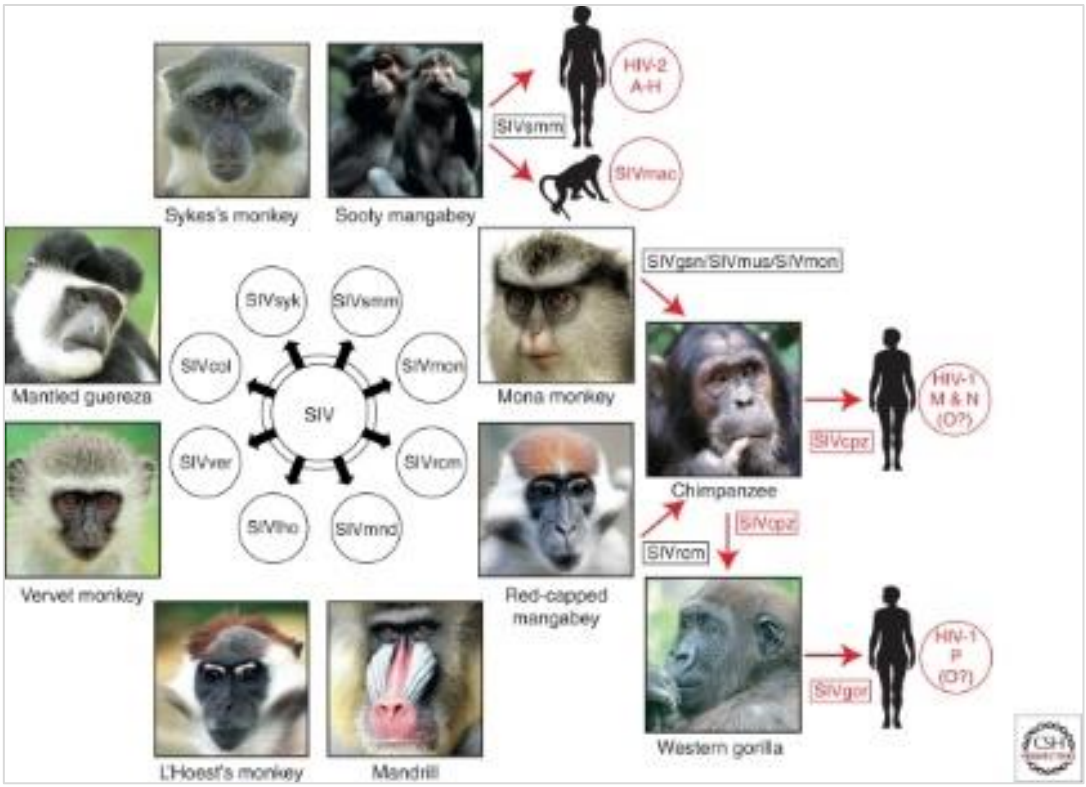


Figure 4: Origins of HIV-1 and AIDS viruses *(Sharp P.M 2011)**.***

Scientists have been able to formulate a family tree ancestry of HIV transmission and concluded that the first episode of transmission of SIV into HIV in humans occurred in 1920s in or around the region of Congo. In further course of events the sex trade and high population of migrants aggravated the virus and furthered the disease. HIV is further categorized into two types; HIV-1 and HIV-2 (Fauci 2003). However, the virus responsible for the global HIV havoc is HIV-1. HIV-1 isn't just one virus, instead it comprises four distinct lineages, that are termed groups M, N, O, and P, each of which resulted from the independent cross-species transmission phenomena.

Group M of HIV-1 was first discovered in 1950s and basically represents the pandemic form of HIV-1 globally; this group has infected millions worldwide and has been found in every country worldwide (Clutter, et al. 2016). Group O was discovered in 1990 and like other groups it is much less prevalent than group M. Group O represents <1% of HIV-1 infections worldwide. Group N was later identified in 1998, and is currently less prevalent than both groups M and O.

Viral Classification

Human Immunodeficiency viruses (HIV) are two species of lentivirus genus and belong to the family retroviridae. There are two types of HIV characterized; HIV-1 and HIV-2. HIV-1 is globally dominant, more infective and comparatively more virulent than HIV-2. Owing to its nature of infection, HIV-1 was originally named lymphadenopathy associated virus (LAV) and human T-lymphotropic virus 3

(HTLV-III). HIV-2 on the other hand is largely confined to Western Africa owing to its poor capacity for transmission. Based on its high importance and prevalence, HIV-1 is further classified into four groups; M, N, O and P. HIV-1 Group M predominates and is responsible for the AIDS pandemic around the globe. It is categorized into nine subtypes; A to D, F to H, J and K. Recombinant HIV are in turn derived from a combination of various subtypes and are termed circulating recombinant forms (CRFs) (Sinoussi 1996).

Currently, CRFs are reported all around the globe and approximately 80 CRFs have been completely established. CRF01-AE is the earliest identified recombinant subtype that was initially discovered in Thailand and has since been considered to be a major pandemic strain in Southeast Asian region. Globally, the region containing the fastest-growing HIV is the Sub-Saharan Africa. The region also contains the most diverse forms of epidemics. The epidemic recombinant CRF18 cpx identified in Cuba was originally associated with 38 HIV strains, most of which were based in Africa. The epidemic recombinant type of viruses formed because of recombinant gene mutations in the virus inherently change virus's biological characters, including its immune response and its sensitivity to drugs.

HIV Subtyping

Virus variation and recombination has resulted in the production and emergence of several HIV subtypes around the globe. The most common type of HIV-1 spread around the world is Group M, which is further classified into nine subtypes (Subtype A to D, F to H, J and K). HIV-1 Subtype C led infections globally account for 50% of HIV-infections in the world, and the viruses belonging to this subtype

are the most prevalent HIV strains in the world. HIV-1 Subtype B viruses are mainly spread in USA, Europe and Australia. However, the HIV-1 Subtype A viruses are mainly located in Eastern Africa, Eastern Europe and Central Asia. Globally, Africa contains the most HIV subtype distribution and almost all HIV subtypes reside in this continent, mainly because the virus originated in this region.

The HIV epidemic strains in different regions of World's continents are not the same. For instance, the epidemic strain in Central Asia is Subtype A, while in East Asia is C subtype. Southeast Asia is CRF01-AE recombinant subtype. Alternatively, patterns of geographical distribution of these subtypes are continually changing with changes in modes of transmission and immigration. This could be explained as in Eastern Europe the virus of subtype B was once the predominant strain, but a survey study conducted in 2003 found that HIV-1 subtype A1 has become the predominant strain in the area. Pakistan too has faced a shift in subtypes, and subtype C is more prevalent.

HIV Structure and Genome

The mature Human Immunodeficiency Virus is a round particle that measures 100 nm in diameter. The envelope of virus comprises of a lipid bilayer with 72 knobs containing trimers of envelope proteins giving it a complex structure. In mature viral particles, viral proteins SU and TM are also present on the envelope. The trimers of gp120 surface protein (SU) are anchored to the membrane by the trimers of transmembrane protein gp41 (TM). These viral proteins cover the outer capsid proteins originally made up of the matrix proteins. Viral particles free of gp120 as

a result of shedding or any other phenomenon are no longer infective. The conical capsid of HIV is assembled from the capsid protein p24.

Other than the outer membranes, the virus mainly possesses two identical genomic RNA molecules coupled with viral enzymes such as RT/RNase H and IN located within the capsid. Also present within a mature HIV particle are the oligopeptides, which arise as a result of the proteolytic processing of precursor proteins (p160 and p55). The mature HIV particle is formed when the combined Gag and Gag/Pol precursor proteins are cleaved into individual proteins after termination of budding of HIV particles from infected cell.

HIV genome comprises of two identical RNA molecules enclosed within the capsid. The integrated form of HIV-1, also called HIV proviral DNA, is the viral RNA reverse transcribed into DNA and integrated into the host genome. The proviral DNA comprises of long terminal repeats present at both ends. The reading frame comprises of several genes in 5' to 3' direction i.e. *gag* gene, followed by *pol* reading frame coding for enzymes protease (PR, p12), reverse transcriptase (RT, p51), RNase H (p53) and integrase (IN, p32). Next to *pol* gene is the *env* reading frame that codes for two envelope glycoproteins gp120 (surface protein, SU) and g41 (transmembrane protein, TM).

In addition to the structural proteins, HIV genome also codes for several regulatory proteins: Tat (transactivator protein), Nef (negative regulating factor), Rev (RNA splicing regulator), Vif (Viral infectivity factor), Vpr (Virus protein r), and Vpu (Virus protein unique). These regulatory proteins have an impact on viral replication, virus budding and pathogenesis.

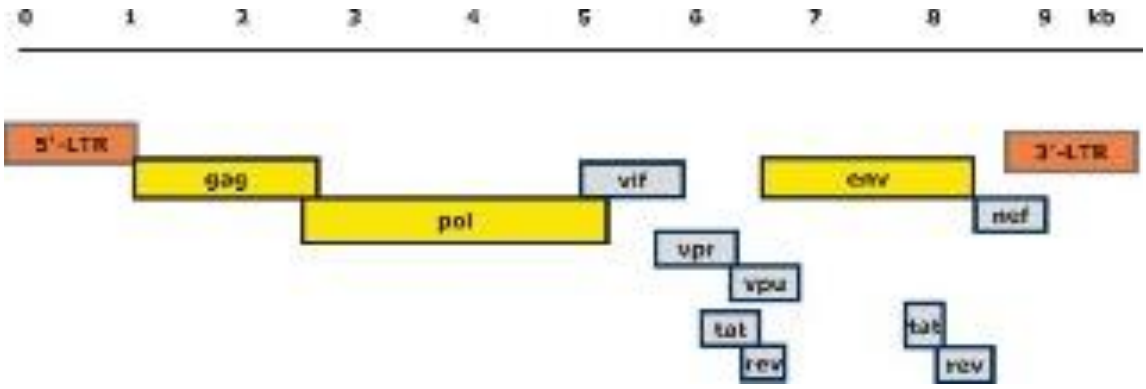


Figure 5: HIV Genome Structure. *HIV genome comprises of three major genes; gag, pol and env along with several regulatory/ accessory genes.*

HIV Gene Functions

HIV-1 contains three major structural genes and several other regulatory and accessory genes in its genome. The three major structural proteins, being *gag*, *pol* and *env*, play crucial role in the virus' assembly, virulence and dissemination. Gag gene gives rise to a 55-kilodalton (kD) gag protein precursor that is also called p55. During translation into protein, the gene's N-terminus is myristoylated and as a result it is associated to the cytoplasmic aspect of cell membrane. This association triggers the recruitment of RNA molecules and ultimately causes budding of the new virion. During its maturation however, p55 is cleaved into the four designated smaller proteins MA (matrix [p17]), CA (capsid [p24]), NC (nucleocapsid [p9]) and into p6. These proteins are separately responsible for virus particle stabilization, assembly, packaging and interactions of the protein.

Pol gene is responsible for coding crucial enzymes responsible for virus' replication: protease (Pro), reverse transcriptase (RT) and integrase (IN). Protease

is responsible for cleavage of gag and gag-pol polyprotein precursors during virion maturation. Reverse transcriptase is responsible for converting viral RNA molecules to DNA for it to get integrated into the host genome and influence translation. Because the polymerase does not contain any proof-reading activity, as a result of replication of viral particles is error prone and introduces point mutations into each new copy of the viral genome. Integrase, as the name suggests, is responsible for integrating the proviral DNA into host genome and ultimately mediating the efficient expression of viral genes.

Env gene encodes the envelope proteins PrGp160, and precursors of envelope proteins; surface glycoprotein (SU) and transmembrane protein (TM) that are cleaved by cellular proteases. The role of env gene is to successfully attach virus to the target cell by anchorage of gp120 to CD4 receptor on host cell and ultimate fusion of both entities. The Env-mediated high affinity interaction between these two receptor molecules is normally not operative in virus particles containing heterologous Env glycoproteins.

Host Factors that influence Disease progression

Host factors that influence HIV disease progression have been identified using a combination of genome wide association studies (GWAS) and specific candidate gene analysis (CGA). The genes responsible for influencing HIV disease progression are normally categorized into three classes;

- Innate and adaptive immunity
- HIV dependency factors (HDFs)

- Intrinsic anti-HIV restriction entities

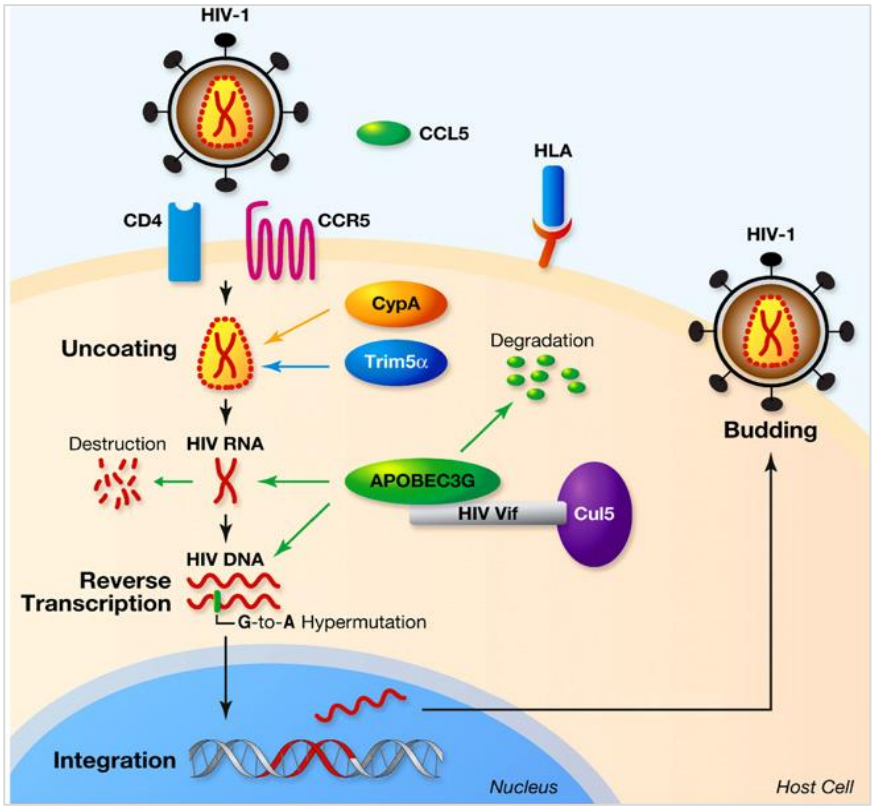


Figure 6: Host Factors involved in HIV Pathogenesis *(An and Winkler 2010).*

Innate and Adaptive Immunity Factors

The HLA system comprised of HLA and KIR mainly constitute innate and adaptive immunity factors influencing disease course. HLA class I alleles have in fact been the first set of host genetic factors that were identified to influence AIDS. Through differential HIV epitope binding, HLA alleles influence the effectiveness of immune response and henceforth affect HIV progression. Over the course of years, it has been established that HLA homozygosity in its 1, 2 or 3 class I loci has been a positive influencer in HIV disease progression. On the other hand, homozygosity in HLA-A, -B and -C leads to shorter AIDS-free survival period in humans.

Presence of another allele in HLA – HLA-B*35 Px allows for only a limited repertoire of HIV epitope to be recognized on HLA entity and is thus positively and dominantly affiliated with more rapid progression to AIDS. On the contrary, HLA alleles – HLA-B*27 and HLA-B*57 have been consistently linked to the delayed HIV disease progression in humans.

HLA system has an innate mechanism of representation of epitopes to killer CD8+ cytotoxic T lymphocytes via HLA class 1 receptor molecules. Failure to present epitope results in killing of HLA containing cells and therefore HIV intelligently suppresses this mechanism by inducing a mutation in the epitope and ultimately most often escapes this control. Innate host factors involved in disease progression, for instance HLA, help or prevent HIV to escape this control passively.

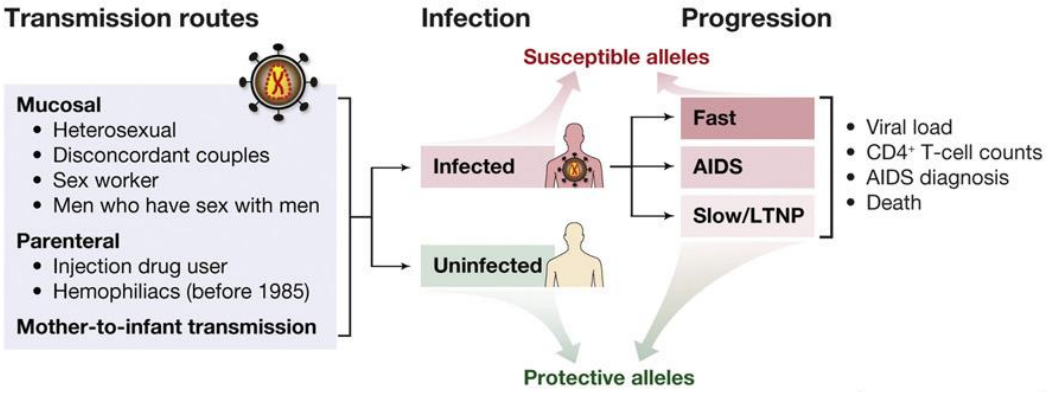


Figure 7: Transmission Routes of HIV-1 *(An and Winkler 2010).*

Killer Immunoglobulin-like receptors (KIRs) are interactive partners of HLA and represent the intersection between acquired and innate immunity. Suppression of HLA has been HIV's defence mechanism that allowed it to escape CTL killing. Natural killer (NK) cells have countered this defence mechanism by counter-checking cells and destroying those that fail to present appropriate set of HLA

receptors with the help of KIRs present upon these NK cells. But KIR mutations have also led us into classifying them as a host factor in influencing disease progression. According to studies, presence of an activating KIR3DS1 receptor allele along with the absence of its partner ligand HLA-Bw4 leads to an increased susceptibility to AIDS progression while in presence of the ligand of any of HLA-B alleles it confers a strong protective effect against HIV disease progression.

HIV Dependency Factors

Several deep bioinformatics studies and large scale small interfering RNA screening studies have revealed that more than 250 HIV dependency factors exist and out of those 36 host factors affecting HIV disease progression have been analysed and thoroughly confirmed to play a major role. Among these are the CD4, CXCR4, NMT1, Rab9p40, and components of the NF-kB and CREB transactivation pathways. Figure below shows the complex yet extravagant role of host proteins/genes in affecting HIV disease progression;

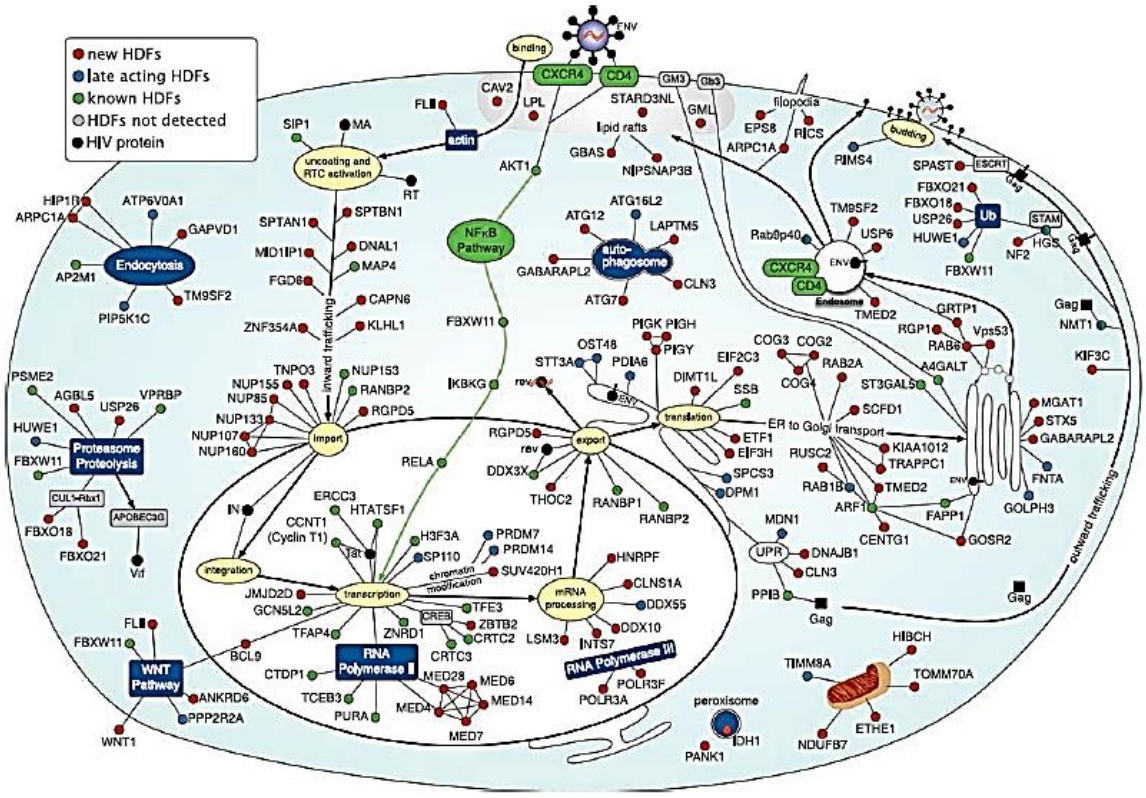


Figure 8: Host Factors involved in HIV-1 Pathogenesis *(Brass, et al. 2008).*

Since their discovery in 1996, chemokine receptors affecting HIV pathogenesis have been thoroughly investigated and it has been found that these receptors play a key role in HIV cell entry. HIV-1 enters human T-helper cells by binding to CD4 (major entry receptor) and one of the two co-receptor molecules i.e. CCR5 or CXCR4, based on either cell type. Chemokines CCL5 (RANTES), CCL4 (Mip1β) and CCL3 (Mip1α) are competitive inhibitory molecules inhibiting the entry of HIV as they are intrinsically ligands of CCR5 co-receptors. It has been reported and thoroughly investigated that a base pair deletion in CCR5 molecule (CCR5 Δ32, homozygous) results in a truncated CCR5 protein expressed on cell surface and almost inevitably protects individuals from conferring HIV infection. Heterozygosity in CCR5 Δ32 does not protect individuals completely from

infection, but indeed delays the disease progression. Up to date, CCR5 Δ32 is the only recognized mutation that completely inhibits viral infection.

CCL5, a CCR5 ligand, competitively inhibits viral entry. However, downregulation of CCL5 by regulatory SNPs is not only associated to HIV acquisition but also improved progression to AIDS in both European and Asian populations. On the contrary, increased CCL5 levels indicate lesser viral entry. This has been proven in a study where 10 fold increased levels of CCL5 were found in genital tracts of HIV uninfected commercial sex workers in comparison to uninfected or unexposed controls. On similar grounds, copy number variation is also predicted to play a huge role in HIV disease burden. Another ligand for CCR5, CCRL3L1, is encoded by 1 to 7 duplicated genes. Lower copy number of CCRL3L1 has been found to lead to improved HIV acquisition and AIDS progression.

Intrinsic anti-HIV restriction entities

Retroviruses in general and HIV in particular are intrinsically restricted within cells by the key mediators belonging to APOBEC3 family of proteins. Belonging to a class of cytidine deaminase enzyme, this protein family works either by incorporating enzymes within virion and functioning by editing the newly reverse transcribed viral DNA by deaminating dC to dU. This results in lethal G-to-A hypermutation for the virion and prevents viral spread. The APOBEC3 family also works by inhibiting viral spread at viral transcription or integration levels. HIV has coped this degradation by APOBEC3 family proteins by employing HIV-1 encoded viral infectivity factor (vif) and causing proteosomal degradation of

APOBEC3 by ubiquitination. In this quest, vif employs host proteins Cullin5, elongins B and C, and Rbx1 for completely degrading APOBEC3. Several intronic SNPs (shown in table below) influence APOBEC3 role in HIV progression;

Gene	Anti-HIV-1 activity	Resistance to vif
APOBEC3A	None	No
APOBEC3B	Strong	Complete
APOBEC3C	Weak	No
APOBEC3D/E	Weak	No
APOBEC3F	Strong	Partial
APOBEC3G	Strong	No
APOBEC3H	Variable	Variable

Table 1: Role of SNPs in APOBEC3 gene role in HIV-1 Pathogenesis.

Cullin5 and APOBEC3 therefore have a collective effect on HIV disease progression *in vivo*. Impact of APOBEC3 expression levels on HIV disease progression has been well documented and its clear range from high to low levels has been observed in non-progressors > LTNPs > progressors. This also suggested that a vif-resistant APOBEC3 molecule could restrict HIV infection.

Another compelling candidate gene, TRIM5α, has also been thoroughly investigated to check which of the several naturally occurring polymorphisms in this gene can influence HIV disease course. An *in vitro* study indicated that cells transduced with TRIM5 missense polymorphisms such as R136Q improved HIV restriction from the cell, however other polymorphisms such as H43Y, V112F,

G110E improved HIV replication directly/indirectly, while some such as G249D and H419Y had no effect on HIV progression. Another aider of the enemy residing within cells is Cyclophilin A (CypA). This protein improves viral infectivity by facilitating with proper uncoating of viral capsid. Regulatory SNPs in CypA are associated with faster progression to AIDS in infected individuals.

Role of Host protein FAM26F in HIV Disease Progression

Little is known about FAM26F (family with sequence similarity 26, member F) and its immune regulatory mechanism. Previously named as IRF-3–dependent NK-activating molecule (INAM), FAM26F is a rather recently known protein that has been observed to play a crucial role in various immune responses within human body. *In silico* analyses have revealed the molecular details about FAM26F gene i.e. it is located on human chromosome 6 with reference genomic contig NC_000006.12, mapped on to chromosomal position 6q22.1. The gene encodes 1141 base pair mRNA molecule containing 3 exons that further translate a 315 amino acid long stable protein molecule. It has a molecular weight of 34.258 kilo Dalton and is predicted to be a membrane protein consisting of 3-5 transmembrane helices as well as an immunoglobulin-like fold. The folds and domains point toward the interaction capability of FAM26F, they represent that FAM26F interacts with Ig-like domains via the beta sheets also present in structures of various antibodies and T cells. The interaction partners of FAM26F have however been reported to be;

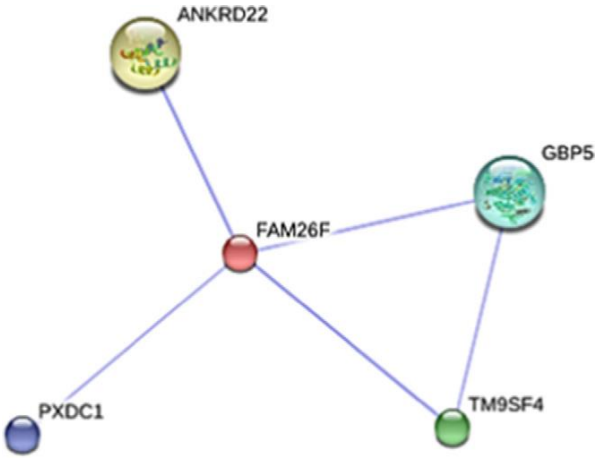


Figure 9: Interactive Partner Proteins of FAM26F. *FAM26F interactive protein partners were evaluated using BLASTn.*

FAM26F has been recognized as a cation channel that has a key role in trafficking smaller molecules and other macromolecules within cell, outside the cell and in between different cells. *In silico* analysis has shown that FAM26F contains only a single well conserved domain i.e. Ca_hom_mod (that covers 248 residues of the 315 aa protein). Ca_hom_mod belongs to a transmembrane family of proteins and is intrinsically involved in controlling cytosolic calcium via pore-forming ion channels as shown below;

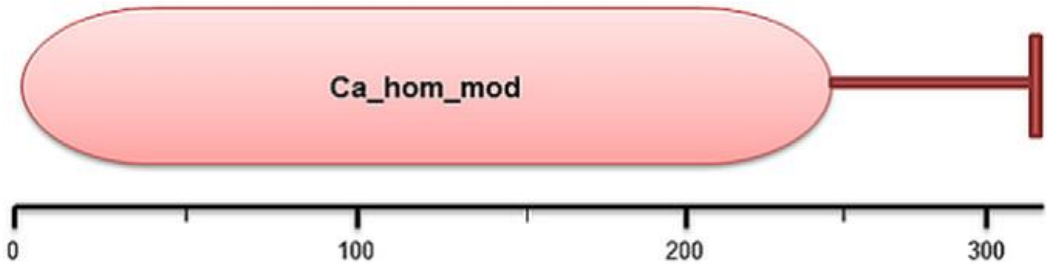


Figure 10: Structure of FAM26F protein. *Structure shows a Ca_hom_mod conserved domain within FAM26F structure that points toward its potential role in calcium modulation.*

Studies have reported role of FAM26F in immune mechanisms. FAM26F has been shown to be expressed in CD8+, CD4+ and CD20+ cells in different infections including Rheumatoid arthritis (RA) and Tuberculosis. Expression of FAM26F in NK cells, dendritic cells and macrophages has also been reported in mice models upon infection. Zhang and co-workers have reported that FAM26F is responsive to several key cytokines including IFN-ɣ, IFN-β and IL-10. In several *in vitro* studies, it has also been revealed that the RNA expression of FAM26F is only upregulated after IFN-ɣ stimulation of peripheral blood mononuclear cells. The promoter analysis of FAM26F molecule has revealed that the molecule possesses an interferon-stimulated response element (ISRE) along with STAT binding sites, concluding and affirming that FAM26F is an IFN-ɣ responsive gene. Although studies have shed light upon its expression levels and possible role in several key pathways influencing inflammation, pathogenesis and infection, the exact modulatory pathways and function of FAM26F is yet to be revealed.

Role of FAM26F in HIV pathogenesis has been a potential question mark. It has been postulated that like in many other infections, altered expression of FAM26F could also be involved in HIV pathogenesis that may potentially serve as a diagnostic marker for HIV infection.

MATERIALS & METHODS

Formulation of Study Design

This study was designed by project's principal investigator and was conducted in collaboration with NACP (National AIDS Control Program) and its referral lab at Pakistan Institute of Medical Sciences (PIMS), Islamabad as per memorandum of understanding (MOU) signed between ASAB NUST, NACP and PIMS.

The methodology of this project is depicted in flow chart shown below;

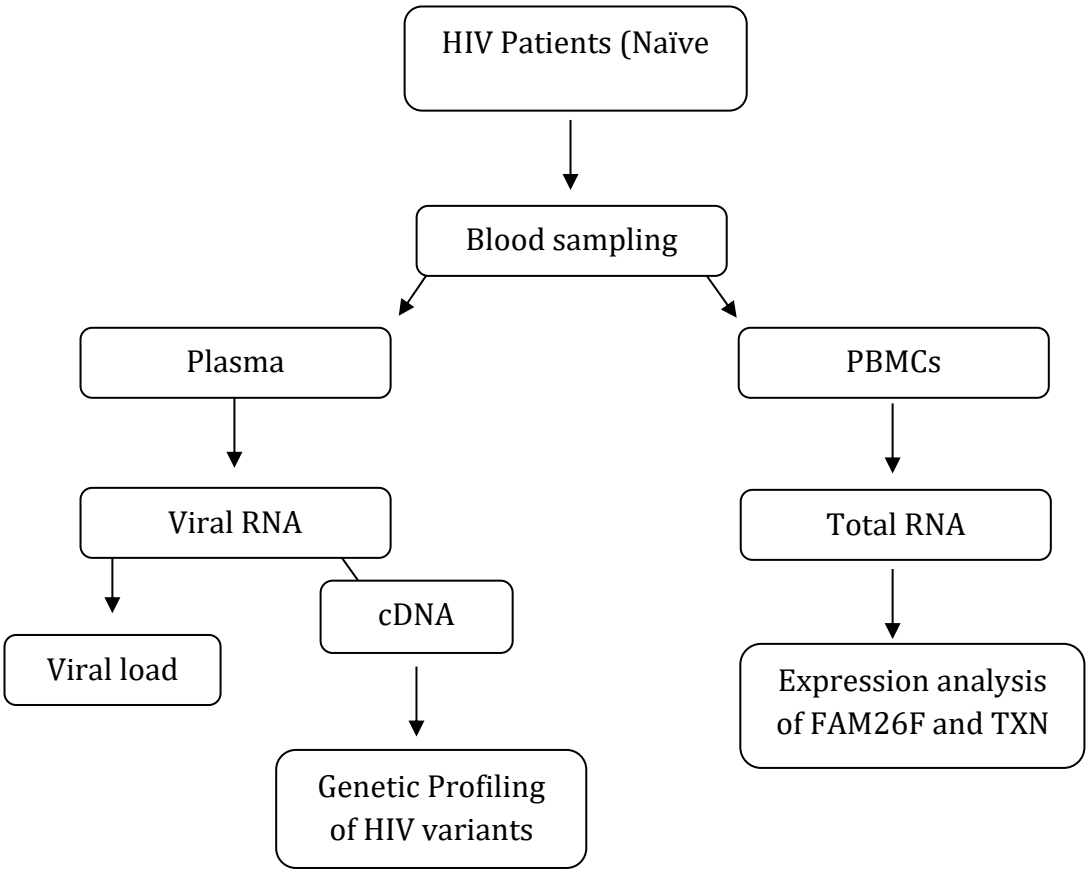


Figure 11: Schematic flow diagram of the methodology of study.

Sampling and Ethical Consent

All Patients recruited in this study were informed about the study design and objectives of this study. All patients provided a written consent to willingly participate in this study and were informed about the potential outcomes of this project. The study commenced after the written approval from the ethical committee of Pakistan Institute of Medical Sciences (PIMS), NACP and from institutional review board (IRB) of Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences & Technology (NUST) Islamabad, Pakistan after methodical assessment of all research protocols involved in this project. Ethical Approval was also sought from Rawalpindi Medical College (RMC) and affiliated hospitals to collect HIV positive blood samples.

Patients involved in this project involved HIV positive patients from Pakistan Institute of Medical Sciences, Islamabad and Hayatabad Medical Complex, Peshawar. HIV patients in Federal Territory (Islamabad and Rawalpindi) and Peshawar were targeted in the course of this study. The study had a peculiar exclusion criteria. According to the criterion, only those HIV positive patients who were treatment naïve or treatment failure were made part of this project. Subjects who corresponded to therapy were specifically excluded. Informed consent was obtained. Patients were also inquired about the systemic effects and possible related co-morbidities prior to or during the course of HIV infection. This was obtained because related co-morbidities could possibly alter the results of this project. In addition patient age, gender, viral load, and relevant clinical factors were also taken into account.

Separation of Plasma and PBMCs from whole blood

Evidently, HIV is a bio-hazardous agent therefore sampling and subsequent optimization procedures were conducted under strict biosafety Level 2 lab conditions which required undertaking of standard precautionary measures at sampling and processing sites. 5 ml blood was collected in EDTA tubes at the time of collection. Blood was then transferred to ficol hypaque tubes. The tubes were centrifuged at 1800-2000 rpm for 25 mins in order to separate blood plasma and PBMCs from RBCs. RBCs settle down owing to density gradient centrifugation and blood plasma is separated in a separate vial, whereas PBMCs are separated as well.

Next, the vial containing PBMCs was centrifuged at 1000 rpm for 10 mins in order to form a pallet. Subsequent washing steps with ice cold ethanol and final dissolution of pallet into PBS+FBS mixture was carried out. Both plasma and PBMCs were stored in cryovials at -80°C for either long term or short term storage.

Assessment of Viral Factors

RNA Extraction

Plasma taken from whole blood of HIV positive patients was further processed in order to isolate free viral particle nucleic acid and further proceed it.

As HIV is an RNA virus, RNA extraction was carried out. The method employed in this study was QIAamp Viral RNA Mini Kit protocol. The protocol normally allows user to purify viral RNA either via spin or vacuum method. Optimization was carried out using both methods and results were negligibly variable. Spin

protocol was further carried out in order to avoid possible contamination. Sample was thawed only once as repeated freeze-thawing of plasma samples leads to reduced viral titres. Carrier RNA dilutions were prepared followed by mixing of Buffer AVL with Buffer AVE-Carrier RNA mixture. 140 µl plasma was added to 560 µl Buffer AVL containing mixture (Buffer AVE + Carrier RNA) and pulse vortexed. After an incubation time of 10 minutes that allowed buffer mixture to completely lyse cells, the mixture was centrifuged briefly. 560 µl ethanol was added to the mixture, pulse vortexed and centrifuged shortly.

Next, 630 µl of mixture was taken into a tube containing spin column and centrifuged at max speed for 1 min. Filtrate was discarded and the step was repeated until all mixture passed through the column. Washing steps began when 500 µl of buffer AW1 was added to the column and spun at max speed for 1 min. After discarding filtrate, 500 µl of buffer AW2 was added to the column placed in new tube and centrifuged at full speed for 3 mins. To continue with elution, the column was placed in a clean 1.5 ml micro-centrifuge tube and old collection tube containing filtrate was discarded. 60 µl buffer AVE was added on to the column membrane carefully, keeping notice it doesn't dispense onto the rim but over the membrane. After 5 minutes of incubation at room temperature, the tube was spun at 10,000 ×g for two minutes. During troubleshooting, it was determined that plasma was divided into aliquots and during elution first aliquot was eluted with buffer and the remaining were eluted with filtrates.

Viral Load Quantification

During the course of troubleshooting and primer optimization, sample viral loads were quantified using artus HI virus-1 RG RT-PCR kit. Viral load quantification was performed at National Institute of Health (NIH, Islamabad).

cDNA Synthesis

Optimization steps involved cDNA synthesis using RevertAid Reverse Transcriptase, but further research concluded that viral RNA copy number in human plasma is significantly low as compared to that within cells. Henceforth, owing to better processing speed and improved resistance to a variety of inhibitors that interfere with cDNA synthesis, Superscript III Reverse Transcriptase was used and was proven fruitful.

Superscript III Reverse Transcriptase cDNA synthesis procedure involves primer annealing to template RNA, assembly of reaction mixture and addition of reaction mixture to annealed RNA as major steps. It primarily starts with combining the following components in a reaction tube:

Component	Volume
50 µM random hexamers, 50 µM Oligo(Dt)20 primers, or 2 µM gene specific primers	1 µL
10 mM dNTP mix (10 mM each)	1 µL
Template RNA (10 pg – 5 µg total RNA)	up to 11 µL

Table 2: cDNA preparation from RNA extracted from plasma (I).

After mixing, the reaction tube was briefly centrifuged. The machine was pre-heated to 65°C and following centrifugation, the mixture was heated at 65°C for 5 minutes and then incubated on ice right away for 2 minutes. During heating of RNA mixture, RT reaction mix was prepared by combining the following components in another reaction tube:

Component	Volume
5× SSIII buffer	4 µL
DTT	1 µL
RNase Out RNase inhibitor	1 µL
Superscript III Reverse Transcriptase (200 U / µL)	1 µL

Table 3: cDNA preparation from RNA extracted from plasma (II).

The RT reaction mix tube was capped, mixed and briefly centrifuged. The next step was addition of RT reaction mix to annealed RNA mix. Incubation steps followed that comprised of incubating the combined reaction mix at 50°C for 15 minutes followed by an inactivation phase of 10 mins at 80°C. As the expected product size was more than 1 kb, an optional step of RNA removal was crucial. To remove RNA, 1 µL of *E.coli* RNase H was added to the mixture and incubated at 37°C for 20 minutes.

In addition to preparing cDNA of template RNA, another reaction tube containing NRTC (non-reverse transcriptase control) was also prepared in order to check the efficacy of reaction and to rule out the possibility of DNA contamination in other reagents.

Primers Designing and Optimization

Primers were designed for *gag, pol* and *env* regions of the HIV genome in order to sequence the entire viral genome. Primers were designed manually with the help of online tools and other software including MegaX and Primer3 online portal (http://bioinfo.ut.ee/primer3-0.4.0/). For each HIV genomic region (among *gag, pol* and *env*) tetra-primers were designed (two forward primers and two reverse primers). This was done keeping downstream processing and application in mind. HIV genome extracted and reverse transcribed through aforementioned protocol requires amplification of genome via Nested PCR and not conventional method. Nested PCR employs two sets of forward and reverse primers.

The first step of primer designing therefore involved retrieval of HIV genome sequences specific to Pakistani population from Los Alamos HIV database (available at https://www.hiv.lanl.gov/). Among the retrieved sequences, a sequence of around 20-22 nucleotide base pairs was chosen and its binding proximity to HXB2 referral HIV sequence was analyzed using the online available QuickAlign Tool (available at https://www.hiv.lanl.gov/content/sequence/QUICK_ALIGNv2/). After selection of the most closely aligned sequence, the selected primers' consensus was checked to match to the conserved regions of aligned sequences with the help of MegaX software.

The primers that showed most consensus to the conserved regions were further scrutinized by evaluation of their GC content, self-complementarities and any hairpin loop structures within primers using the online OligoCalc tool (available at http://biotools.nubic.northwestern.edu/OligoCalc.html). Primer Tm (melting

temperatures) were also calculated manually using the standard formula and later reaffirmed using OligoCal tool. Finally, BLAST was performed to check primer specificity with the respective gene sites they ought to amplify with the help of HIV BLAST tool on the HIV database – Los Alamos (also available at https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/). QuickAlign online tool was also used to check for potential mutations or non-binding regions on Subtype A1 of HIV-1.

After primer designing, primer optimization was the next big step for continuation with Polymerase Chain Reaction. Primers were checked to amplify required product at a range of temperatures on gradient PCR ranging from 49°C to 55°C and mostly depended upon primer Tm. PCR extension temperature was set at 68°C to 72°C depending upon the enzyme utilized. 50°C annealing temperature and 68°C extension temperature ideally worked for the primers utilized, however the extension time varied by the combination of primers used.

Number of Stage	Cycle Number	Temperature	Duration
Stage 1	×1	94°C	2 minutes
Stage 2	×30	94°C 50°C 68°C	30 seconds 30 seconds variable
Stage 3	×1	68°C	10 minutes
Stage 4	×1	4°C	∞

Table 4: Conventional PCR setup.

Amplification of HIV genes via Nested Polymerase Chain Reaction

Nested Polymerase Chain Reaction was run to ensure the specificity of the product length and type. Nested PCR normally involves two sets of both forward and reverse primers and therefore outer set and inner set of primers were formed and used to get the desired product. In order to continue with PCR, each round of both rounds involved a combination of the following items in a clean reaction tube:

Reagents	Volume Added
Buffer	5 µL
MgSO4 (50 mM)	2 µL
dNTP (10 mM)	1 µL
Forward Primer (10 µM)	1 µL
Reverse Primer (10 µM)	1 µL
Taq Polymerase	0.4 µL
Sterile Water	34.6 µL
DNA (template)	5 µL

Table 5: Conventional PCR Recipe.

All reagents except DNA were made into a master mix depending upon the number of reactions to prepare. DNA was finally added to each reaction tube. The whole process was performed keeping mixture on ice. Normally for one reaction mix, three reactions were prepared i.e. Experimental, Positive Control and Negative Control. Positive control involved HXB2 plasmid in place of DNA added to it and NTC involved sterile water. After preparation of reaction mixture, the sample and controls were lightly mixed and thoroughly spun or mildly centrifuged so all the

components would settle down. The PCR was set according to the following conditions;

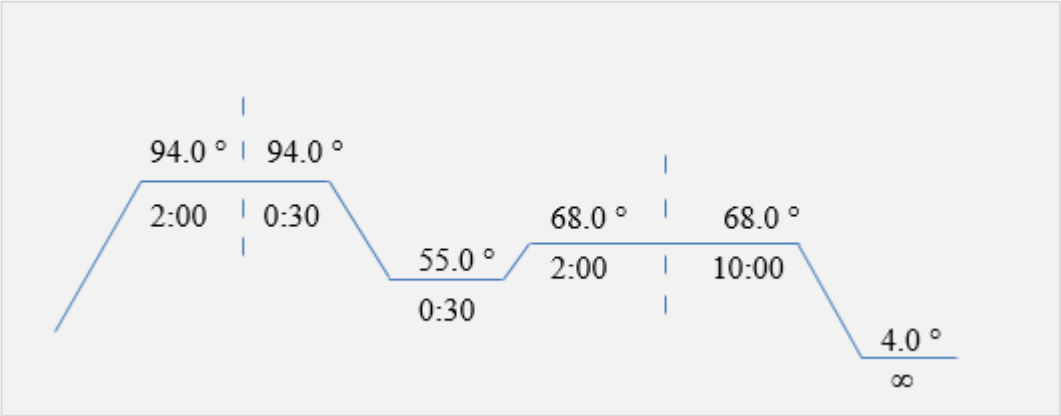


Figure 12: Conventional PCR Cycles and Setup.

Second round of PCR was run with internal primers and extension time was set accordingly. The PCR product was either directly proceeded to gel electrophoresis or stored at -20°C.

Gel Electrophoresis

Gel electrophoresis was carried out for the samples processed. TAE buffer (pH 8.0) was ideally used. 100 ml TAE buffer was taken in a flask and 1.25 g agarose was weighed and added to it. Buffer was made to boil until the solution was clear. 2.5 µL ethidium bromide was added to the solution after it cooled down a bit. After the gel cooled down, it was poured in gel caster and a setting time of 30 minutes was given for the gel to set. 5 µL of sample was added to a reaction tube, 2 µL of 6X dye was added to the sample. Respective ladder (1kb) was run along with the sample/s. Gel electrophoresis conditions were 120V, 500Amp, 45 minutes. Gel was visualized later using Dolphin Doc UV Gel Visualizer.

PCR Product Purification for Sequencing

The PCR products amplified using NESTED PCR were subsequently purified using GENEJET PCR Purification kit. The protocol comprises of 5 to 10 minutes, but has significance especially for pursuing downstream applications like sequencing, ligation, blotting or in-situ hybridization. As the next step of this project was sequencing of the amplified cDNA samples, purification of these products of PCR reaction mixture e.g. dNTPs, proteins, primers, unincorporated labelled nucleotides etc. was crucial. The silica based technology of the spin columns employed in this protocol allowed for apt separation of proteins from DNA and binding of DNA to columns only to be eluted in the final step. The components involved were;

Components	Volume/Amount Added
Binding Buffer	1:1 (of PCR products)
Wash Buffer	700 µL
Elution Buffer (10mM Tris-HCl, pH 8.5)	50 µL
GeneJET Purification Columns	variable
3M Sodium Acetate	variable
Isopropanol	variable
Amplified DNA product	variable

Table 6: PCR Product Purification Recipe.

The first step of PCR product purification involved addition of 1:1 volume of binding buffer to PCR products and mix thoroughly. Solution colour change to yellow was described as an indication of optimal pH. If the colour was violet or

orange, 10 µL of 3M sodium acetate (pH 5.2) was added to the solution and mixed till the colour changed to yellow. Next, as our desired DNA fragment getting amplified was ≥ 500 bp, 1:2 volume of isopropanol was added to the mixture. The mixture was mixed thoroughly. Next, all contents were transferred to GeneJET Purification columns placed in collection tubes and columns were centrifuged at max speed for 30 to 60 seconds. Flow-through was discarded. Next, 700 µL of wash buffer was added to the column, and centrifuged at max speed for 30 to 60 seconds. Flow through was discarded. As it was crucial to get the column completely rid of residual wash buffer, the empty column was subsequently centrifuged for an additional minute and it was made sure to remove all buffer from column before proceeding to next step. The final step of elution involved transferring the columns to clean 1.5 ml centrifuge tubes and addition of 50 µL elution buffer on top of the membrane. Caution was taken to not put buffer over the rim of column and only at the top of membrane. The columns were then centrifuged at max speed for 1 minute and the flow-through was proceeded to sequencing.

Assessment of Host Factors

Nucleic Acid Extraction

Peripheral Blood Mononuclear Cells (PBMCs) isolated using ficoll hypaque gradient centrifugation procedure were further processed to isolate nucleic acid (DNA and RNA) from them. RNA was further utilized for FAM experimentation while DNA was stored at -20°C. Nucleic acid was extracted from All Prep DNA/RNA kit that simultaneously purifies genomic DNA and total RNA from cellular or tissue samples. The protocol follows as;

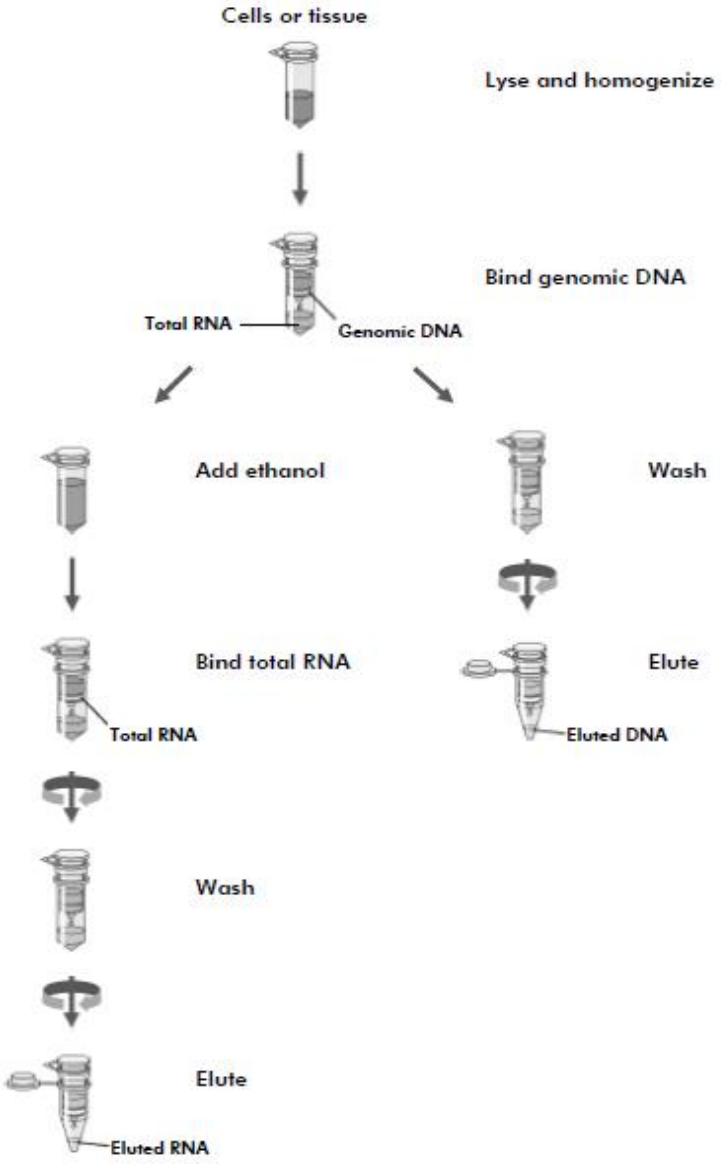


Figure 13: RNA Extraction from PBMCs - protocol.

Before proceeding with All Prep DNA/ RNA kit method of nucleic acid extraction, 10 µL β-mercaptoethanol was added to 1000 µL of Buffer RLT. The procedure started after harvesting cells (forming cell pallet). To the harvested cells, 600 µL of Buffer RLT containing β-mercaptoethanol was added and homogenized. The tube was flicked to loosen the pallet and vortexed thoroughly. The lysate was then transferred to DNA spin column placed in a 2 ml collection tube provided with the

kit. After the transfer, the tube was centrifuged at ≥ 8,000 ×g (≥ 10,000 rpm) for 30 seconds. The column was then placed in a new 2 ml collection tube and stored for later purification steps of either DNA or RNA.

DNA Purification

To the DNA Column, 500 µL AW1 buffer was added, and centrifuged at ≥ 8,000 ×g for 15 seconds. Flow through was discarded and column was kept. Next, 500 µL of buffer AW2 was added to column, and centrifuged at maximum speed for 2 minutes. Longer centrifugation time was ensured. After centrifugation, collection tube was emptied the column was centrifuged again at maximum speed. The column was placed in a new 1.5 ml collection tube. 100 µL buffer EB was added directly to spin column membrane and lid was closed and incubated for 1 minute. The tube was centrifuged at ≥ 8,000 ×g for 1 minute to elute DNA. Step was repeated to elute DNA further using the same collection tube.

RNA Purification

To the flow-through obtained in the initial processing step, 600 µL of 70% ethanol was added and mixed well by pipetting. Next, 700 µL of sample was transferred to RNeasy column already placed in 2 ml collection tube and centrifuged at ≥ 8,000 ×g for 15 seconds. Flow through was discarded but collection tube was reused. 700 µL of buffer RW1 was added to RNEasy column, and later centrifuged at ≥ 8,000 for 15 seconds. Flow through was discarded and collection tube was re-used. Next, 500 µL of buffer RPE was added to the column and centrifuged at ≥ 8,000 ×g for 2 minutes. The column was placed in 2 ml collection tube, and the old tube and filtrate were discarded. The new column containing column was spun at full speed for 1 minute. Last, the column was placed in a 1.5ml reaction tube. 30-50 µL RNase

free water was added directly to the membrane. The lid was closed and tube was centrifuged at ≥ 8,000 ×g for 1 minute to elute RNA.

Nucleic Acid Quantification

Nucleic Acid (DNA and RNA) were quantified after extraction using Colibri Microvolume Spectrometer. As RNA was eluted in RNAse free water and DNA in elution buffer, therefore RNAse free water and Elution Buffer were used to blank the machine. 1 µL of respective samples were then loaded onto the machine to measure the amount. Readings were obtained in ng/ µL and were taken twice or thrice to obtain an average at the end of procedure.

cDNA Synthesis

In order to formulate cDNA from template RNA, Thermoscientific RevertAid First Strand cDNA Synthesis Kit, that has the capacity to form first strand of cDNA up to 13kb, was used. Prior to proceeding with cDNA synthesis, template RNA was treated to removal genomic DNA from sample. In order to do so, an RNase-free tube was taken and the following components were added:

Component	Volume
RNA	1 µg
10× Reaction buffer with MgCl2	1 µL
DNase I, RNase-free	1 µL (1U)
Water, nuclease-free	up to 10 µL

Table 7: cDNA synthesis from RNA extracted from PBMCs (I).

The components were then incubated at 37°C for 10 minutes. Next, 1 µL of 50 mM EDTA was then added to the mixture and incubated at 65°C for 10 minutes in order to obtain DNA-free RNA for next steps of cDNA Synthesis.

In order to proceed with cDNA synthesis, the following reagents were added in a reaction tube placed on ice;

Component	Volume
Template RNA	11 µL
Primer	1 µL
5× Reaction Buffer	4 µL
RNase Inhibitor (20U/ µL)	1 µL
10 mM dNTP mix	2 µL
Revertaid M-MuL V RT (200U/ µL)	1 µL

Table 8: cDNA synthesis from RNA extracted from PBMCs (II).

After brief mixing of the aforementioned components added in a reaction tube, the tube was briefly centrifuged. The reaction tube was then incubated at 42°C for 60 minutes, and finally inactivated by heating at 70°C for 5 minutes. The prepared cDNA was either proceeded directly to PCR processing or stored at -20°C.

FAM26F and Thioredoxin Primer Optimization - Amplification via Conventional Polymerase Chain Reaction

In order to initiate with FAM26F and TXN primer optimization, amplification via conventional PCR was performed.

Amplification of FAM26F using Plasmid

First, plasmid containing FAM26F gene was used as a template and FAM26F primers were used that contain the following sequences;

Primers	Sequences
FAM-F	5'-CACCCGATGCCTATCTCCAG-3'
FAM-R	5'-TTTGCTGCCACTCTTTCATGC-3'

Table 9: FAM26F primer sequences.

The PCR process involved a combination of the following items in a clean reaction tube:

Reagents	Volume Added
Buffer	3 µL
MgSO4 (50 mM)	2 µL
dNTP (10 mM)	1 µL
Forward Primer (10 µM)	1 µL
Reverse Primer (10 µM)	1 µL
Taq Polymerase	0.4 µL
Sterile Water	9.6 µL
DNA (template)	2 µL

Table 10: Conventional PCR Recipe (for FAM26F amplification).

All reagents except DNA were prepared into a master mix depending upon the number of reactions to prepare. DNA/ Plasmid was finally added to each reaction tube. The whole process was performed keeping mixture on ice. For this step, two

reaction tubes were prepared; one contained plasmid and the other contained negative control that contained nuclease free water (NTC). Results were thereafter observed and analysed. Further processing initiated after positive results were obtained. After preparation of reaction mixture, the sample and controls were lightly mixed and thoroughly spun or mildly centrifuged so all the components would settle down. The PCR was set according to the following conditions;

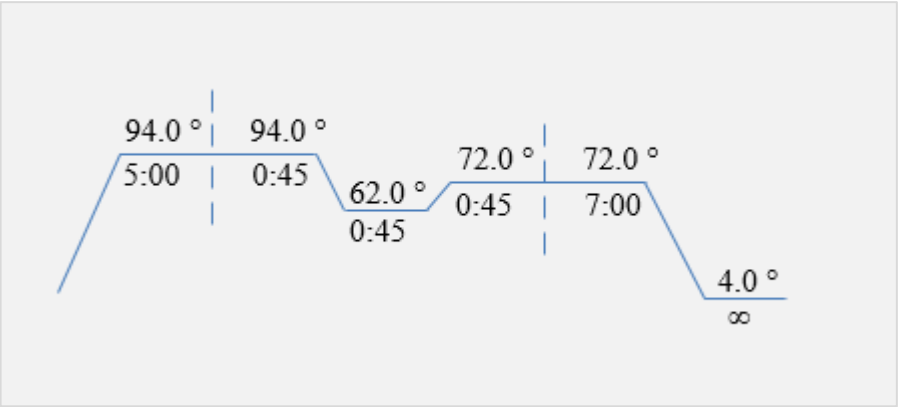


Figure 14: Conventional PCR Cycles and Setup (for FAM26F amplification).

The PCR product was either directly proceeded to gel electrophoresis or stored at -20°C.

Amplification of FAM26F & Thioredoxin using HIV positive patient cDNA samples

After proceeding with FAM primer optimization on plasmid, the next step was to amplify the HIV positive patient sample cDNA with FAM and TXN primers in order to check if the optimal annealing temperature worked for samples. The primer sequences used for TXN were:

Primers	Sequences
TXN-F	5’-CTCGTTTGGTGCTTTGGATCC-3’
TXN-R	5’-CAAGTTTATCACCTGCAGCGTC-3’

Table 11: Thioredoxin Primer Sequences.

Normally for one reaction, three reactions mixes were prepared i.e. Experimental, Positive Control and Negative Control. Positive control involved GAPDH or Beta Actin in place of template DNA added to it and NTC involved nuclease free water. The reagents added were:

Reagents	Volume Added
Buffer	3 µL
MgSO4 (50 mM)	2 µL
dNTP (10 mM)	1 µL
Forward Primer (10 µM)	1 µL
Reverse Primer (10 µM)	1 µL
Taq Polymerase	0.4 µL
Sterile Water	9.6 µL
DNA (template)	2 µL

***Table 12: Conventional PCR Recipe** (For Thioredoxin amplification).*

After preparation of reaction mixture, the sample and controls were lightly mixed and thoroughly spun or mildly centrifuged so all the components would settle down. The PCR was set according to the following conditions;

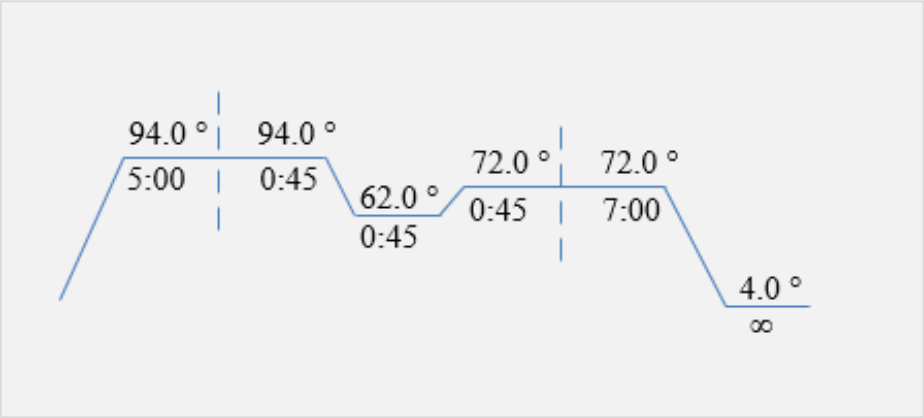


Figure 15: Conventional PCR Cycles and Setup *(For Thioredoxin amplification).*

The PCR product was either directly proceeded to gel electrophoresis or stored at -20°C.

Primer Optimization and Primer Efficiency Analysis

In order to proceed with FAM26F and Thioredoxin (TXN) primer efficiency analysis, primers specific to both genes were designed and optimized to check primer efficiency. The primers used had the following sequences;

Primers	Sequences
FAM-F	5'-CACCCGATGCCTATCTCCAG-3'
FAM-R	5'-TTTGCTGCCACTCTTTCATGC-3'
TXN-F	5'-CTCGTTTGGTGCTTTGGATCC-3'
TXN-R	5'-CAAGTTTATCACCTGCAGCGTC-3'

Table 13: Sequences of FAM26F and Thioredoxin Primers.

The primer efficiency was analysed using real-time polymerase chain reaction. Each cDNA sample was serially diluted at either 1:20 or 1:30 up to five serial

dilutions. Real time PCR reagents were combined in reaction tubes according to following measurements;

Reagents	Volume
Template cDNA	2 µL
Forward Primer	0.5 µL
Reverse Primer	0.5 µL
SYBR Green	4 µL
Nuclease Free Water	12 µL

Table 14: Real-time PCR (qPCR recipe) for FAM26F and Thioredoxin amplification.

As the goal was to check primer efficiency, therefore the reaction was performed in separate reaction tubes using successive dilutions of template cDNA. Duplicates of five successive reaction mixtures containing serial dilutions were made in the following order;

- 1st Reaction: 1: 30
- 2nd Reaction: 1: 900
- 3rd Reaction: 1: 27,000
- 4th Reaction: 1: 810,000
- 5th Reaction: 1: 24,300,000

The reaction was run on Real Time PCR machine and annealing temperature kept for both FAM26F and Thioredoxin (TXN) was 62°C. The reaction was also run

with FAM26F containing plasmid. The results were subsequently analysed, and the experiment was repeated until optimal primer efficiency was obtained.

Amplification via Real Time PCR (FAM/TXN/GAPDH/B-Actin)

After optimizing the annealing temperatures of primers and checking primer efficiency, real time PCR (with a similar profile) was performed to detect, characterize and quantify cDNA obtained from PBMCs of HIV positive treatment naïve and treatment failure patients were amplified. Each reaction was run in triplicates where experimental reaction mix involved amplification with FAM and TXN primers while GAPDH or β-actin (housekeeping genes) primers were used as positive control. 30 samples were in turn amplified and thereafter analysed.

Gel Electrophoresis

Gel electrophoresis was performed for the processed cDNA samples. TAE buffer (pH 8.0) was used. 100 ml TAE buffer was taken in a flask and 2.5 g agarose was weighed and added to it. Buffer was made to boil until the solution was clear. 2.5 µL ethidium bromide was added to the solution after it cooled down. After the gel cooled down, it was poured in gel caster and a setting time of 30 minutes was given for the gel to set. 6 µL of sample was added to a reaction tube, 2 µL of 6X dye was added to each sample. Respective ladder (50 bp) was run along with the sample/s. Gel electrophoresis conditions were 80V, 500Amp, 60 minutes. Gel was visualized later using Dolphin Doc UV Gel Visualizer.

RESULTS

Patient Demographic Details

Blood samples of HIV positive patients were obtained at Pakistan Institute of Medical Sciences (PIMS). The sole exclusion criterion was on-treatment patients. Patients were classified based on a number of demographic and biological factors. As a result, patients ranged in ages from 18 to 65. Age groups were formulated according to study plan and the age distribution was represented as:

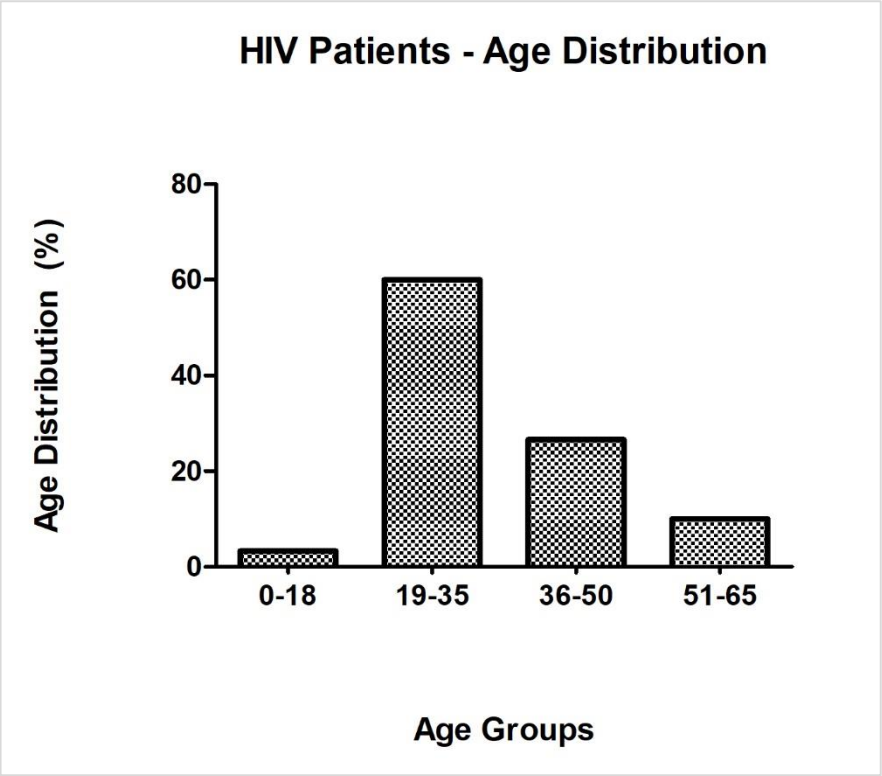


Figure 16: Age Distribution in HIV Positive patients. *The figure shows that the patients who participated in this study were divided into four age groups; 0-18 (children and adolescents), [19-35, 36-50] (adults) and 51-60 (senior adults). Most number of HIV positive patients were however observed in the first subgroup of adults i.e. 19-35 age group.*

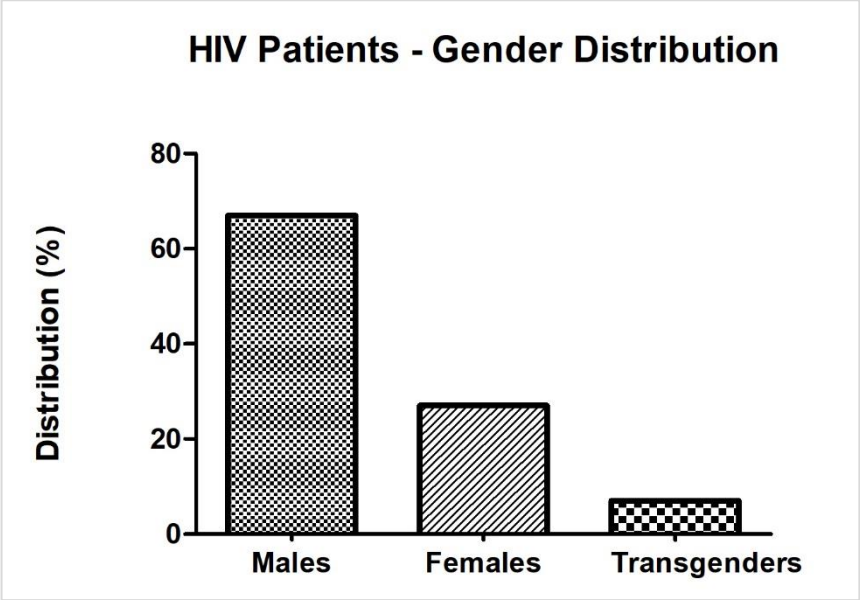


Figure 17: Gender Distribution in HIV positive patients. *The figure shows that Males, Females and Transgenders participated in this study. Males dominated being ~70% of total HIV positive patients in this study.*

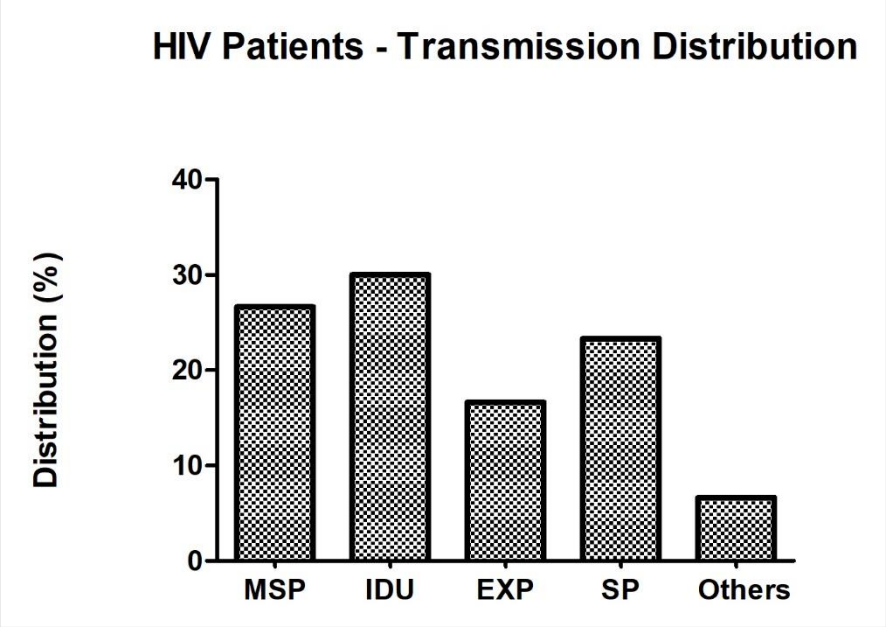


Figure 18: Viral Transmission distribution in HIV positive patients. *The figure shows that HIV positive patients in this study acquired virus from multiple modes; (I): Multiple sexual partners, (II): Injection Drug Users, (III): Expatriates/ Immigrants, (IV): Spouse and (V) Others such as blood transfusion/disease etc. Highest transmission route observed was Injection Drug Use. All female HIV positive patients acquired virus from spouse.*

As all patients were HIV positive patients, CD4 count of all patients was also obtained and is represented as:

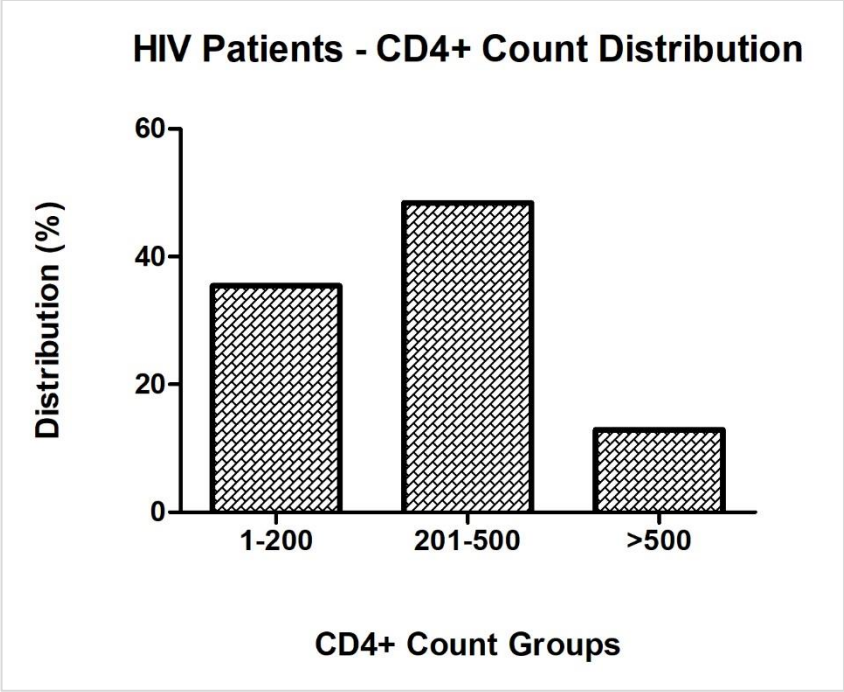


Figure 19: CD4+ count distribution in HIV positive patients involved in this study. *The figure shows that CD4 counts of most individuals were <500 and therefore according to UNAIDS classification of T-cell test, a significant number of patients involved in study had AIDS (CD4 count= <200) or were at the verge of acquiring AIDS.*

Separation of Blood Plasma and Peripheral Blood Mononuclear Cells

5-6 mL blood was withdrawn from every patient. Blood was collected in EDTA tubes and treated in Ficol Hypaque tubes afterwards. Based on principle of density gradient centrifugation, Ficol Hypaque media separated blood into its components (i.e. blood plasma, PBMCs and RBCs) as shown in figures below:

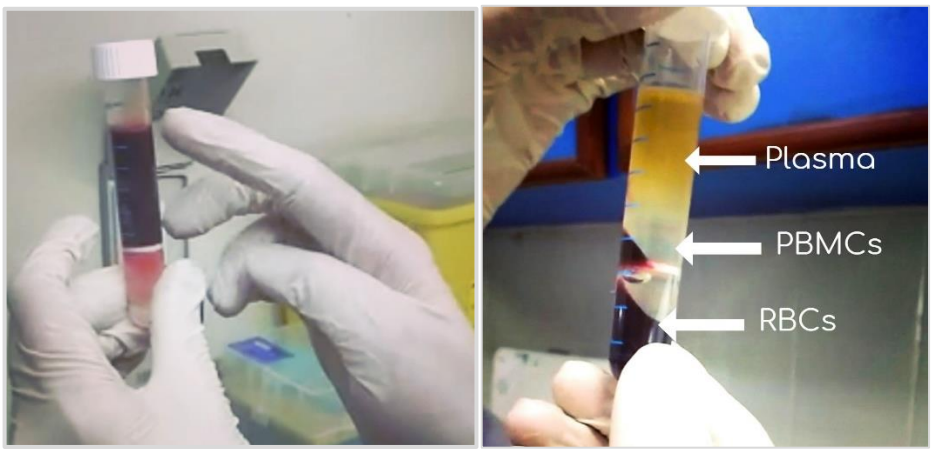


Figure 20: Collection of HIV positive blood from patients and initial processing using Ficoll Hypaque. *The figure represents how blood is separated into distinct plasma, PBMCs and RBCs after density gradient centrifugation using ficol hypaque vials.*

Viral Load Analysis

Serial No. of Samples	Viral Load (copies/ml)
01	**80,412**

02	**23,369,026**
03	**355,196**
04	**443,607**
05	**4,429,972**
06	**652,969**
07	**95,886**
08	**4,036,704**
09	**516,527**
10	**64,178**
11	**248,646**
12	**363,399**
13	**5,998,439**
14	**4,252,229**
15	**2,257,067**
16	**161,468**
17	**1,775,885**
18	**812,618**
19	**399,489**
20	**26,799,516**
21	**140,527**

22	**612,871**
23	**1,447,944**
24	**266,109**
25	**23,369,026**

Table 15: Viral loads of obtained blood samples of HIV positive patients. *Viral loads in both treatment failure and naïve patients are mentioned.*

Assessment of Viral Factors

Amplification of HIV genes

RNA was extracted from both plasma and PBMCs using RNA Mini Kit as well as All Prep RNA and DNA extraction kit respectively. RNA obtained from plasma was lesser in concentration as compared to that from PBMCs therefore RNA obtained from plasma was subsequently processed to prepare cDNA using Superscript III RT cDNA synthesis kit. cDNA obtained from plasma was amplified and verified. PCR was run in two rounds (Nested PCR), once with external pair of primers and once with an internal pair. This was done to improve specificity of the PCR amplicon. The primers pairs used to amplify gag were:

Therefore, the PCR amplicons verified using Gel electrophoresis had band sizes of 4,251 bp and 2969 bp for gag and pol respectively. The primers were initially tested on one patient sample and one control sample. The results are shown below;

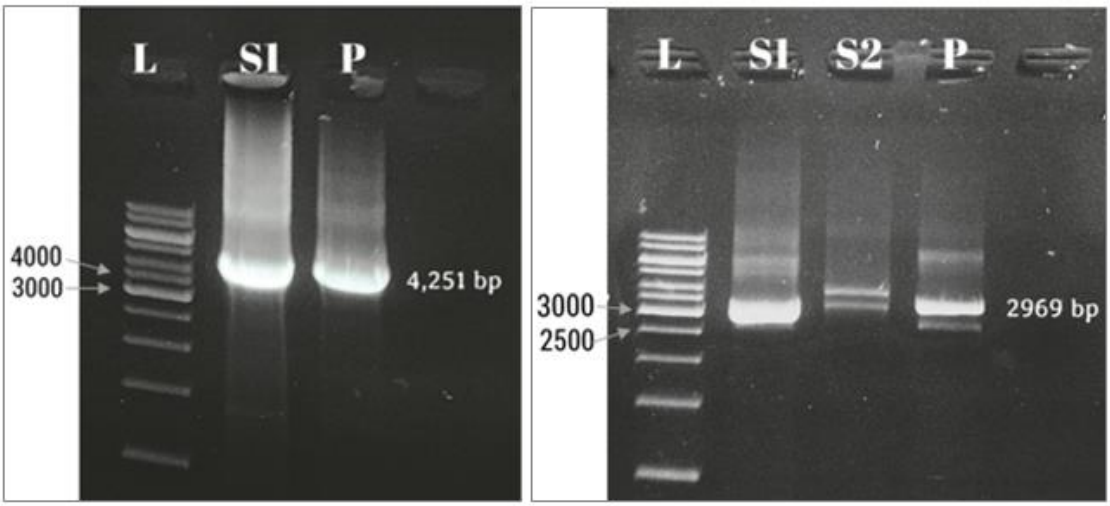


Figure 21: Visual Representation of amplified viral genes (gag and pol) in cDNA samples of patients (S1 and S2) extracted from plasma and in positive control (P) using NESTED PCR. *Samples S1 and S2 show products in patient sample group in which gag and pol gene were amplified using two sets of primers individually. The final amplicon size for gag was 4251 bp and was 2969 bp for pol. The gel also shows bands at about that location with respect to 1 kb ladders (L). Positive Control (P) contained an artificially constructed plasmid containing HIV genome amplified using same sets of primers for both viral genes.*

After verification, primers were used to amplify patient samples. Two rounds of PCR similarly followed and the genes were successfully amplified as shown in the representative figure;

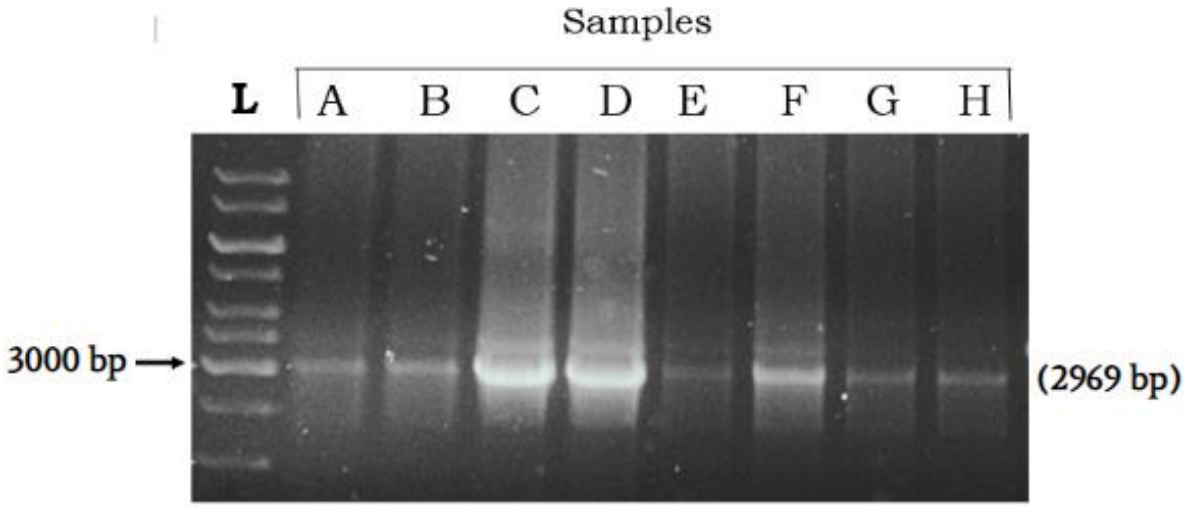


Figure 22: Visual Representation of amplified cDNA samples of patients extracted from plasma using NESTED PCR. *Samples A to H show products in patient sample group in which pol gene was amplified using two sets of primers. The final amplicon size was 2969 bp and the gel also shows bands at about that location with respect to 1 kb ladder.*

Sequencing of Amplified Viral Genes

Viral gene (Pol) was sequenced via Sanger Sequencing in order to verfiy the amplification of right product. The representative chromatogram of amplified *pol* gene and sequenced using forward primer is shown below:

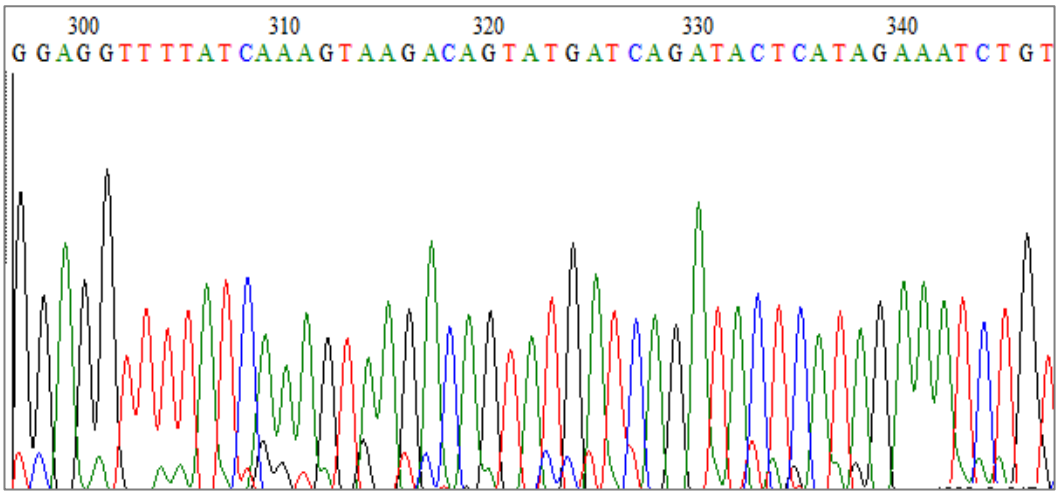


Figure 23: Representative Chromatogram of amplified pol gene generated via Sanger Sequencing (using forward primer).

The sequence was also analyzed for any sequence similarities and difference using pairwise and multiple alignment as shown below:

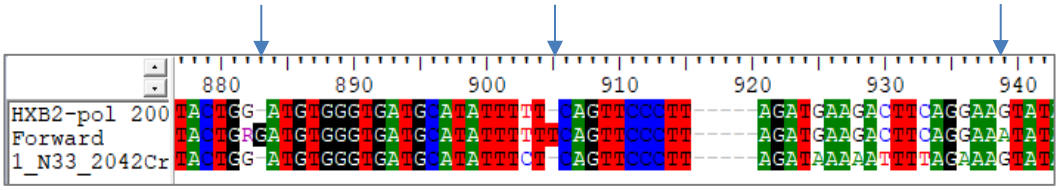


Figure 24: Representative Figure: Shade Identities and Similarities in Pairwise and Multiple Alignment Window; *Lane 1: HXB2 (positive control); Lane 2: Sequence under study; Lane 3: HIV Sequence from Pakistan.*

Assessment of Host Factors

Expression Analysis of Host genes

Peripheral Blood Mononuclear Cells (PBMCs) were processed after separation from whole blood. Cellular RNA and DNA were extracted from PBMCs. DNA

was stored at -80°C for future processing. Control and patient samples were taken for this purpose. Obtained DNA and RNA samples were quantified using Colibri Microvolume Spectrometer. Representative results obtained for both patient and control RNA samples are shown in figures:

Sample No.	RNA Concentration (ng/µL)	OD (260/280)
P1	30.49	2.05
P2	11.25	1.87
P3	2.75	1.95

Table 16: Representation of RNA concentrations of patient samples. *Samples contained acceptable RNA concentrations with optimal O.D. ready to be reverse transcribed and to be processed further.*

Sample No.	RNA Concentration (ng/µL)	OD (260/280)
C1	38.85	2.08
C2	85.30	2.07
C3	155.56	2.24

Table 17: Representation of RNA concentrations of control samples. *Samples contained improved RNA concentrations with optimal O.D. ready to be reverse transcribed and to be processed further.*

RNA was reverse transcribed into cDNA and subsequently amplified or verified to check the cDNA product using beta actin primers via conventional PCR. As a result of this, beta actin was amplified and checked using gel electrophoresis.

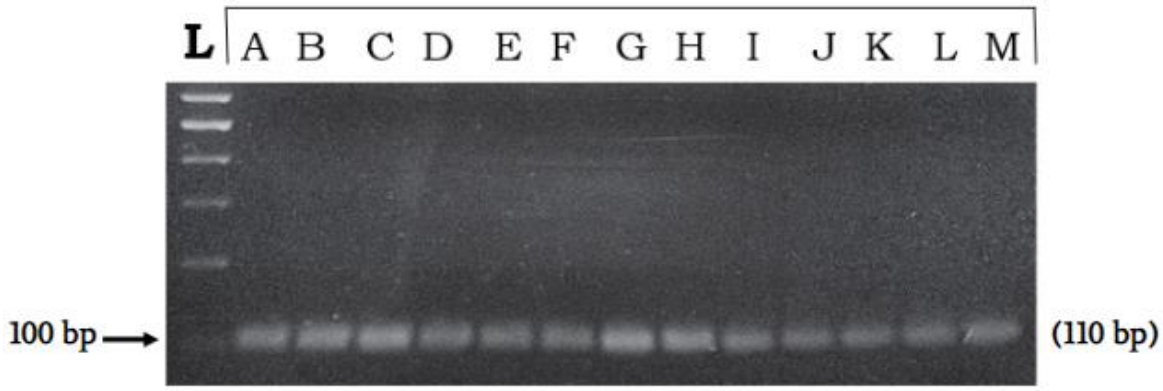


Figure 25: Visual representation of verification of amplified cDNA products using Beta Actin primers. *The house keeping gene of cDNA samples was amplified via conventional PCR to ensure if samples could be proceeded for quantitative PCR.*

After verification of cDNA synthesis, primer efficiency for target host genes was checked. Primers used to amplify FAM26F, Thioredoxin and Beta Actin were diluted according to 1:20 or 1:10 dilution factor and qpcr was run to check the primer efficiency. The primer efficiency was analysed as;

Slope	-3.150324638
R Squared	0.9991
Efficiency (%)	**107.70**

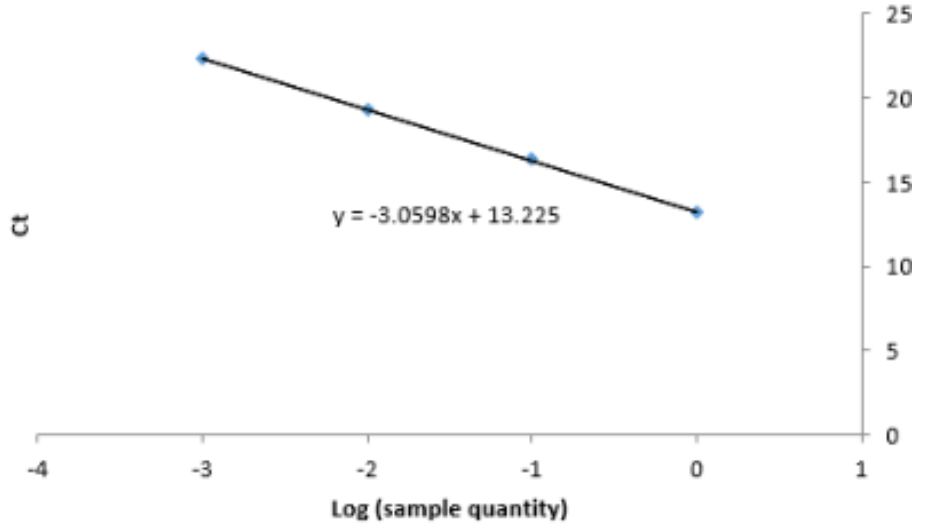


Figure 26: Optimization and Primer Efficiency Analysis of FAM26F gene.

Slope	-3.1
R Squared	0.9933
Efficiency (%)	**110.17**

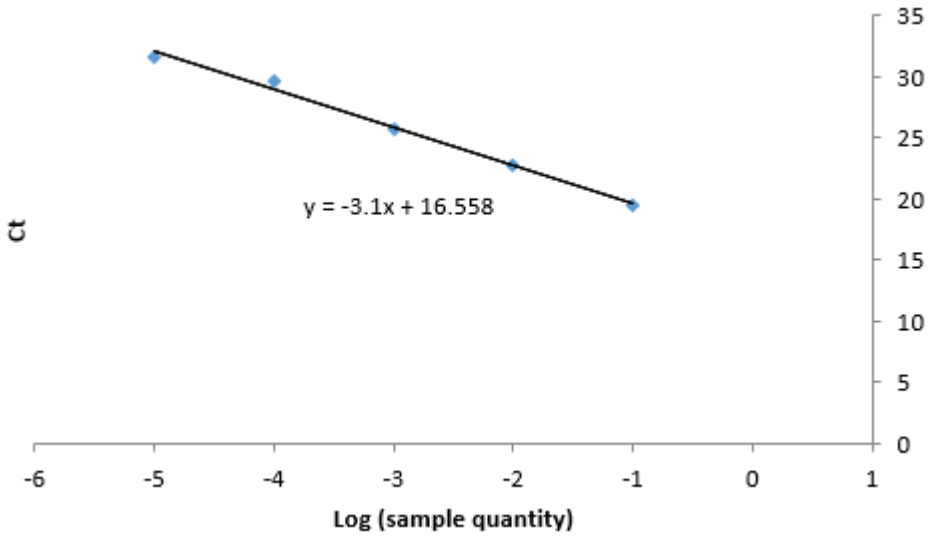


Figure 27: Optimization and Primer Efficiency Analysis of Thioredoxin gene.

After calculating primer efficiency, cDNA of patient samples were run using qpcr to check the expression of FAM26F gene and Thioredoxin (Trx) host genes in HIV positive and respective control samples with similar RNA concentrations. RNA concentrations of all patient and control samples were adjusted to a similar value in order to check the expression of host target genes with respect to house keeping gene (Beta Actin used in this experiment).

The target host genes (FAM26F, TXN along with β-actin) in both host and control cDNA samples were amplified and quantified via real time polymerase chain reaction after due amplification via conventional PCR. The amplification plots and dissociation curves were observed and analysed to ensure there were no primer dimer formation or formation of secondary structures represented below;

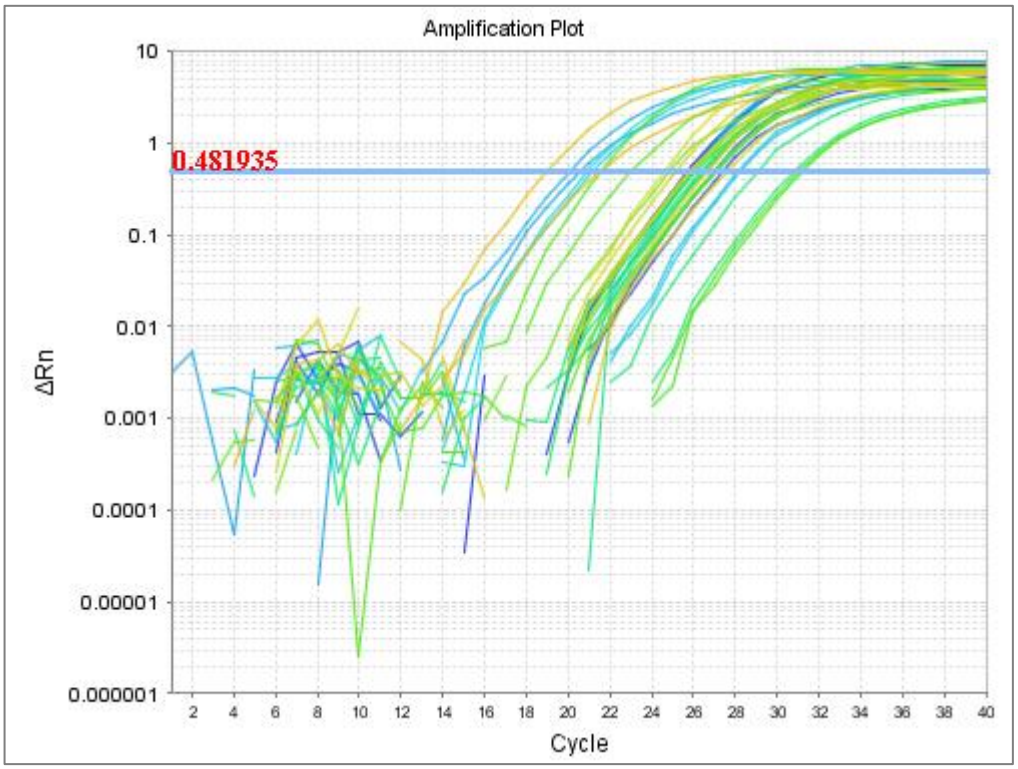


Figure 28: Representative amplification plot of qPCR (Real time PCR).

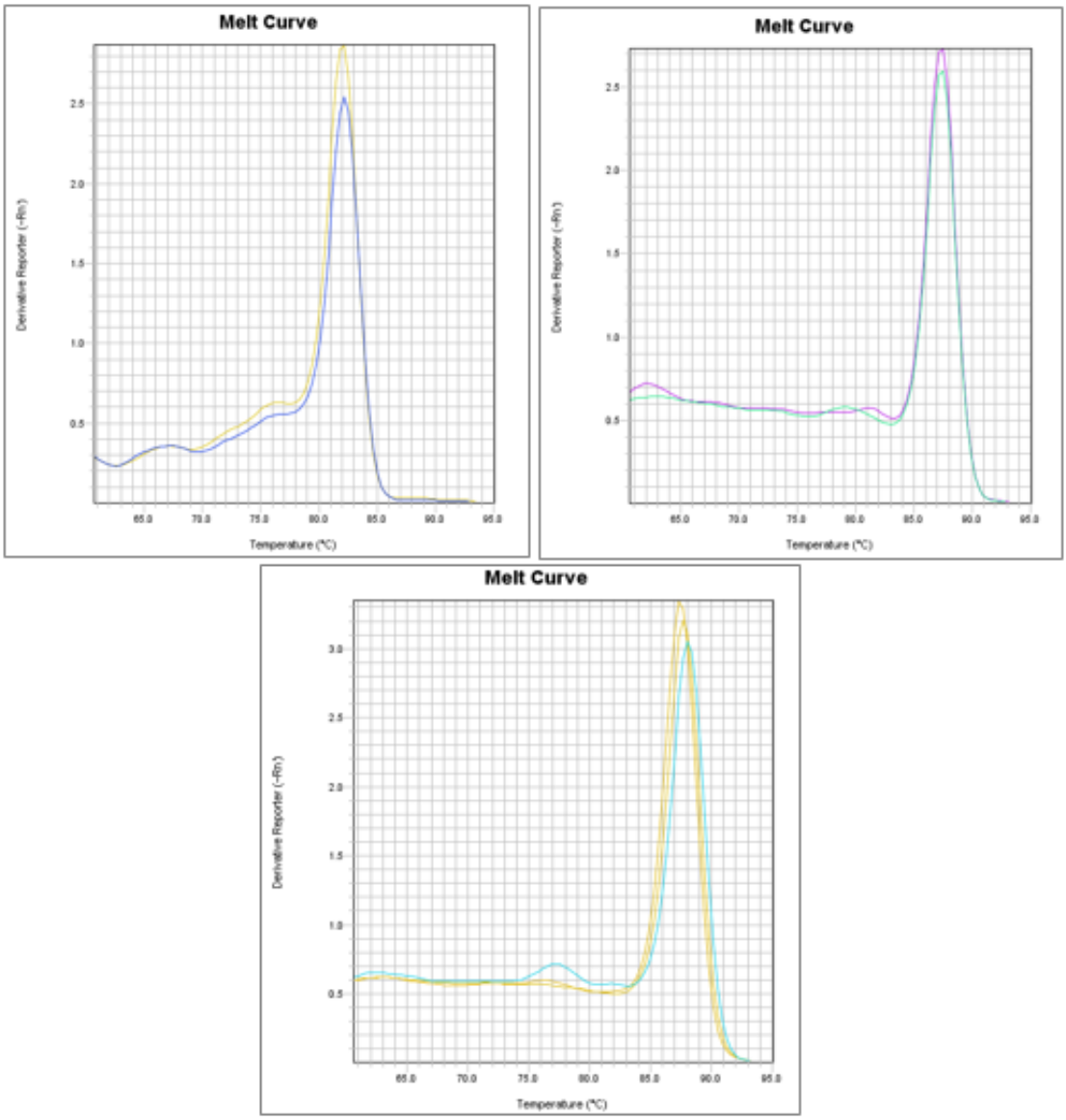


Figure 29: Representative Melt Curves of FAM26F, Thioredoxin and Beta Actin*; A single peak dissociation curve indicated primer specificity and predicted no secondary structure/ primer dimer formation in amplified genes (I): FAM26F, (II): Thioredoxin and (III): Beta Actin.*

Analysis of FAM26F and Thioredoxin (TXN) gene expression in HIV positive patients with respect to Controls

In order to evaluate the expression of target host genes FAM26F and TXN in HIV patients with respect to their expression in controls, qPCR was performed on patient and control samples. Optimized primers of FAM26F, TXN and house keeping gene Beta Actin were used. It was concluded that there was a significant increase in fold change in FAM26F and Thioredoxin as compared to controls (Figure).

The copy number on the other hand was not significant owing to its relative nature of calculation. The data affirmed the hypothesis that FAM26F and Thioredoxin are both increased in infected patients and furthered the hypothesis in concluding that their expression in HIV immunocompromised patients is also increased post infection in comparison to healthy controls (Figure).

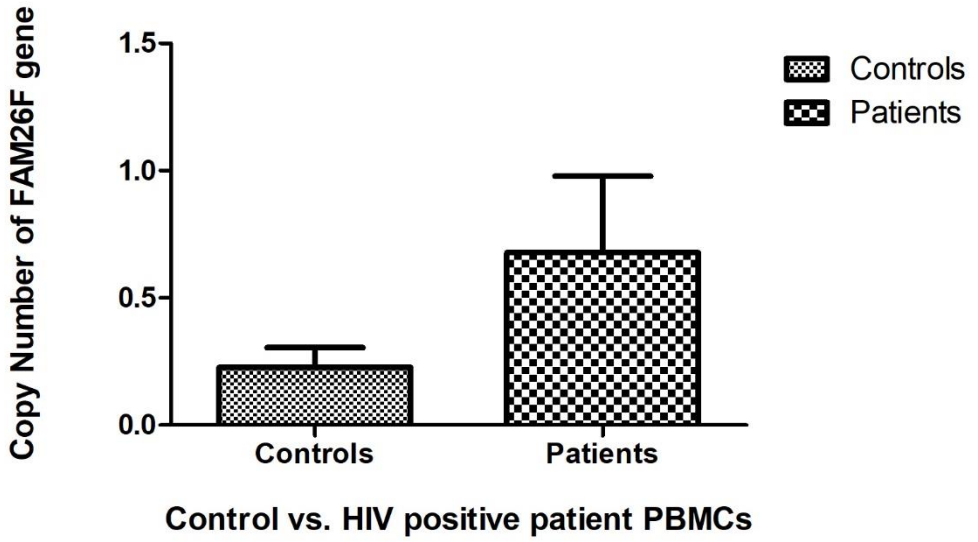


Figure 30: Copies of FAM26F gene in healthy controls vs. HIV positive patients. *An increase in copy number and expression of FAM26F was observed in patient group with a mean difference of 0.4 in copy numbers in comparison to controls.*

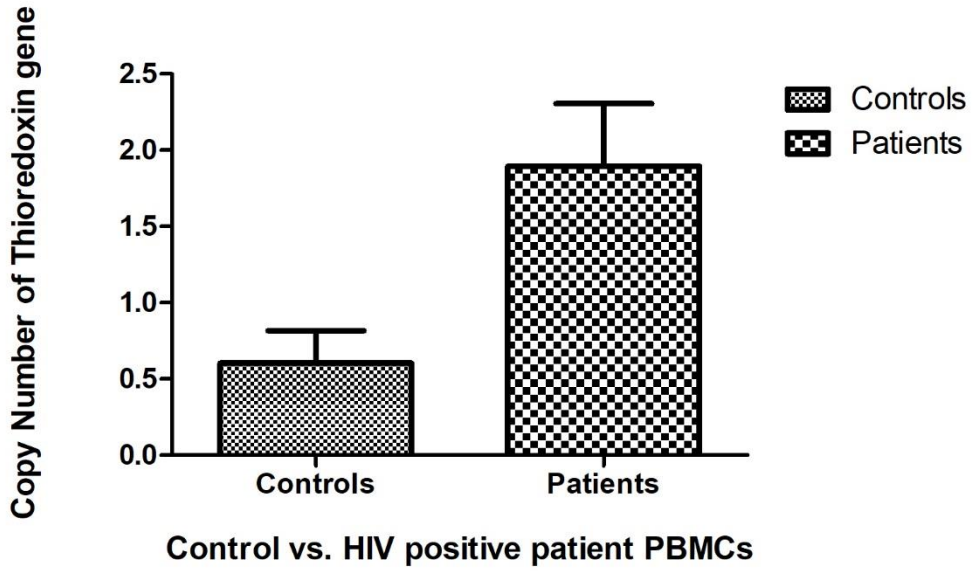


Figure 31: Copies of Thioredoxin gene in healthy controls vs. HIV positive patients. *An increase in copy number and expression of TXN was observed in patient group with a mean difference of 1.2 in copy numbers in comparison to controls.*

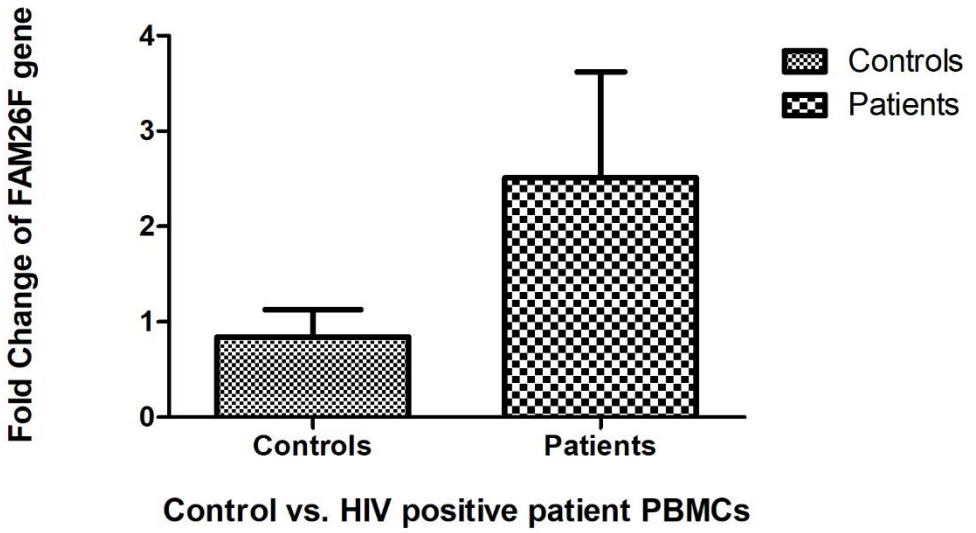


Figure 32: Fold Change of FAM26F gene in healthy controls vs. HIV positive patients. *An increase in fold change and expression of FAM26F was observed in patient group with a mean fold change difference of 1.7 in comparison to controls.*

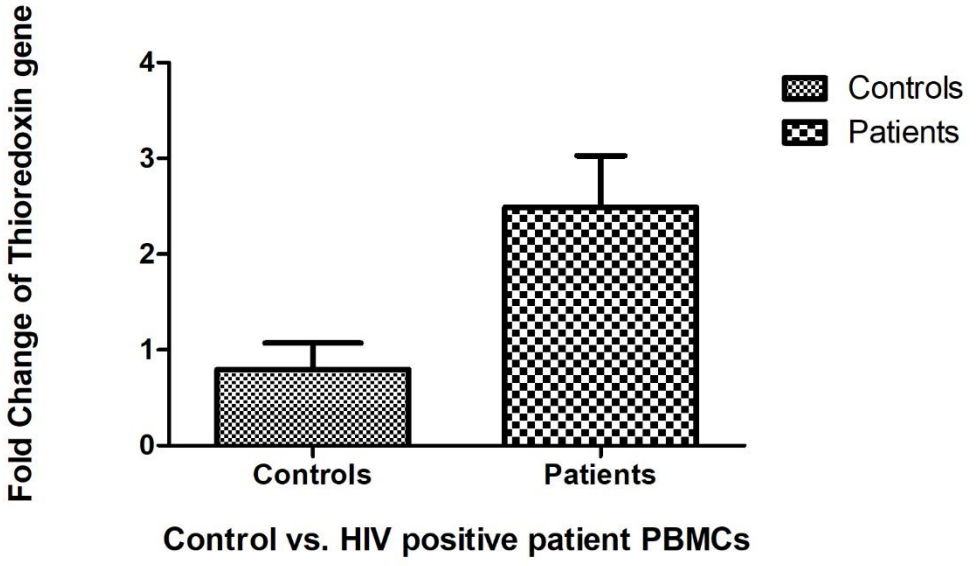


Figure 33: Fold Change of Thioredoxin gene in healthy controls vs. HIV positive patients. *An increase in fold change and expression of Thioredoxin was observed*

in patient group with a mean fold change difference of 1.7 in comparison to controls.

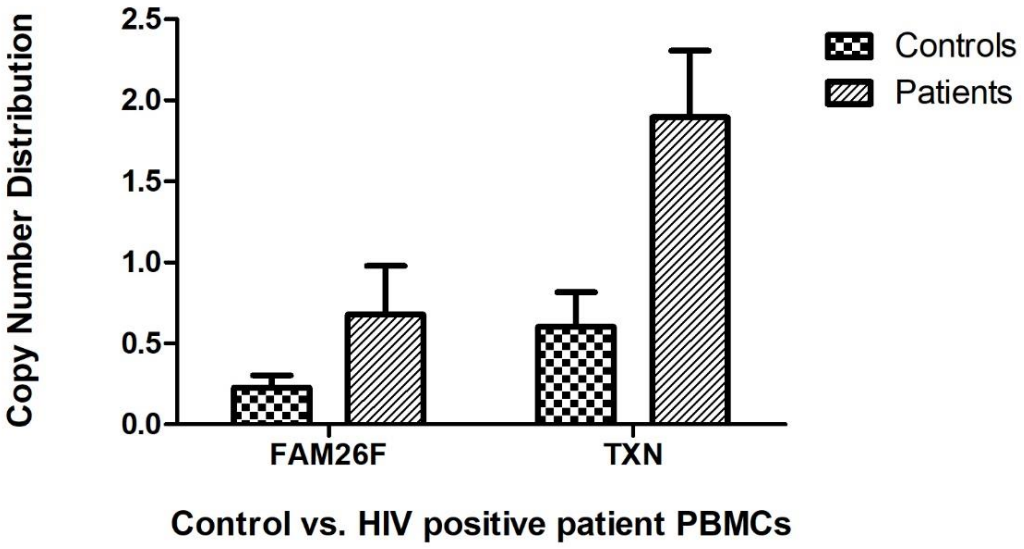


Figure 34: Correlation of copy numbers of FAM26F and Thioredoxin genes in healthy controls vs. HIV positive patients. *An increase in copy number and expression of FAM26F and Thioredoxin was observed in patient group with respect to healthy controls. In addition, copy numbers of Thioredoxin in both control and patient groups were more than FAM26F with a cumulative mean difference of 1.63.*

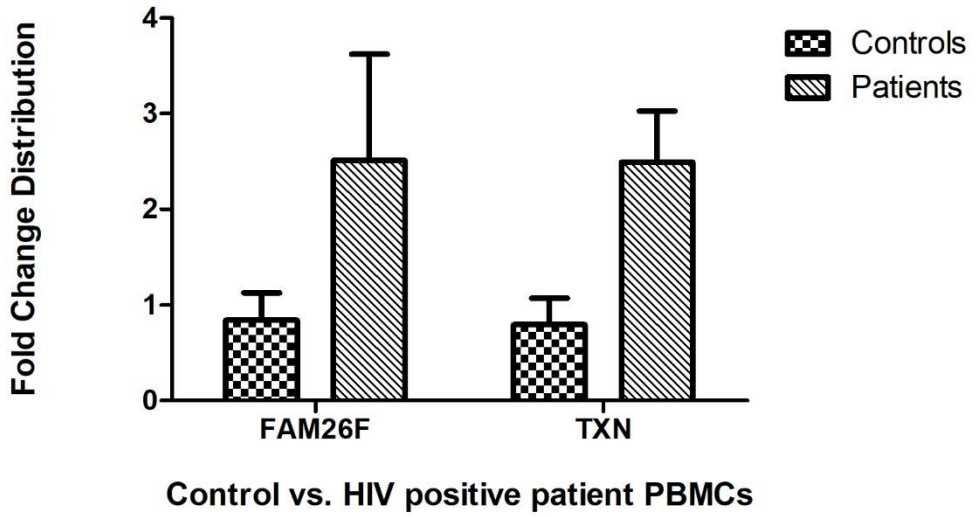


Figure 35: Correlation of fold change difference of FAM26F and Thioredoxin genes in healthy controls vs. HIV positive patients. *A significant increase (p value = 0.0152) in fold change and expression of FAM26F and Thioredoxin was observed in patient group with respect to healthy controls. In addition to that, fold change of FAM26F in both control and patient groups was more than Thioredoxin with a cumulative mean difference of 0.07.*

DISCUSSION

Human Immunodeficiency Virus Type 1 (HIV-1) is a highly intelligent retrovirus that enters inside the human body only by masking itself. It faces the enemy directly and gains entry into the body's defence line cells (CD4+ cells). HIV incorporates itself into the genome of human body and masks itself so intelligently that it could survive in human body for decades without showing any symptoms. However, more often than not it starts replicating and infects partner CD4+ cells and a cascade of infection continues leaving body vulnerable to other infections. When this CD4+ cell count in human body drops down to 200 or below, the host is said to suffer from Acquired Immunodeficiency Syndrome (AIDS) – the end stage of HIV infection ultimately (in absence of appropriate treatment) leading to death.

The virus has rendered more than 39 million people susceptible to opportunistic infections worldwide and the only cure from this massively lethal entity is a deep and insightful study about its genome and its mode of interaction with human body. The genome of HIV comprises of three major and several other genes. The major genes involve gag, pol and env. One of the many modes of HIV masking itself from host and remaining resistant to treatment regimes is the constantly evolving viral genome. It has been identified that the virus has the astonishing capacity of possessing 361025 mutations per nucleotide per replication cycle mostly without impairing its virulence or function. This leaves the mankind bewildered to analyze complete viral genome and synthesize drugs to cater to the virus. HIV-1 is also population specific and subtypes common in one region may not be in another region of the World. HIV-1 in the Europe, Americas and Africa has been sequenced and analyzed. Drug regimen specific to major populations of the world are recommended by WHO.

Pakistan, however, has been unable to completely sequence and analyze its HIV-1 genome (except for <1000 partially characterized sequences reported). Henceforth, a drug regimen and related measures to cater to virus are not population specific. This study has aimed to identify, collect and process patient samples from Pakistan Institute of Medical Sciences (PIMS – National HIV treatment center of Pakistan) located in Islamabad (capital territory of Paksitan). HIV positive patient samples were collected, processed and amplified. Gag and Pol gene of (treatment naïve and treatment failure) patient samples were successfully amplified. Sequence characterization, in order to verify amplified PCR products, was also performed on few samples. This study has served to be a step closer to the goal of sequencing HIV-1 specific to Paksitani population.

Furthermore, certain host factors were also observed and analyzed for (if any) differential expression in HIV positive patients in comparison to healthy controls. FAM26F (family with sequence similarity 26, member F) is also another important host factor that has been recently studied in different propsects. FAM26F is a member of FAM26 gene family and basically encodes a 34 kilo dalton (kD) membrane glycoprotein (predicted by TMHMM server). It has been reported that a differential expression of FAM26F is present in priming macrophages, early stage liver draft failure, in several bacterial infections and has been sporadically expressed in cancers of various body organs (such as breast, uterus, cervix etc). FAM26F is reported and postulated to be a crucial innate immunity entity. Thioredoxin (Trx), another gene involved in this study, on the other hand belongs to a class of small redox proteins. It is a 12 kDa protein that inherently possesses a disulfide reducing activity. Like FAM26F, Thioredoxin is not completely

understood but it surely is a crucial activator proteins in humans that plays an antioxidant role in several key regulatory pathways. It has been widely reported that plasma Trx levels rise in cases of HIV infection and several other cancers.

This study has been able to identify and affirm the hypothesis of elevated expression of plasma FAM26F and Trx levels in HIV positive patients in comparison to control samples. Fold change difference in both proteins has been found to be significantly improved in HIV positive patients in comparison to controls. Role of these host factors in HIV infection needs to be investigated to find a potential biomarker of disease and possibly to control the viral load.

FUTURE PROSPECTS

This study paves a way for future investigations in HIV-1 sequence characterization and role of essential host factors involved in HIV pathogenesis. This study can be further proceeded to full genome amplification of HIV-1 followed by Next Generation Sequencing (NGS) of the amplified full-length sequences. Sample size and regional range within Pakistan can be expanded in order to attain sequences of all circulating viral subtypes. The study can be further proceeded by identifying conserved mutations in the viral genome that allow development of resistance against drug regimens and targeting those regions with better and effective drugs.

In terms of host response to HIV-1, FAM26F and Thioredoxin among the many host factors involved in HIV-1 pathogenesis can be studied. These factors have a potentially significant role in innate immunity against virus and the role can be further investigated after reconfirmation of their expression using more number of HIV positive and healthy control samples.

REFERENCES

An, Ping, and Cheryl A. Winkler. 2010. "Host genes associated with HIV/AIDS: advances in gene discovery." *Trends in Genetics* 119-31.

Avert. 2018. *GLOBAL HIV AND AIDS STATISTICS.* https://www.avert.org/global-hiv-and-aids-statistics.

Benhar, M, I. L Shytaj, J. S Stamler, and A Savarino. 2016. "Dual targeting of the thioredoxin and glutathione systems in cancer and HIV." *The Journal of Clinical Investigation.*

Blattner, W., R.C. Gallo, and H.M. Temin. 1988. "HIV Causes AIDS." *Science* 514-17.

Brass, AL, DM Dykxhoorn, Y Benita, N Yan, A Engelman, RJ Xavier, J Lieberman, and SJ Elledge. 2008. "Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen." *Science* 921-926.

Burgener, Adam, Ian McGowan, and Nichole, R. Klatt. 2015. "HIV and mucosal barrier interactions: consequences for transmission and pathogenesis." *Current Opinion in Immunology* 22-30.

Burke, D.S. 1997. "Recombination in HIV: an important viral evolutionary strategy." *Emerging Infectious Diseases* 253-59.

CCM Pakistan. 2018. *Country Coordinating Mechanism Pakistan.* http://www.ccmpakistan.org.pk/disease_hiv-aids.php.

Clutter, Dana, S., Patricia, Rojas Sánchez, Soo-Yon Rhee, and Robert, W. Shafer. 2016. "Genetic Variability of HIV-1 for Drug Resistance Assay Development." *Viruses* 48.

DiPerri, Giovanni, Maria, Chiara Danzi, Giovanna, De Checchi, Sergio Pizzighella, Maurizio Solbiati, Mario Cruciani, Roberto Luzzati, et al. 1989. "NOSOCOMIAL EPIDEMIC OF ACTIVE TUBERCULOSIS AMONG HIV-INFECTED PATIENTS." *THE LANCET* 1502-1504.

Fauci, Anthony S. 2003. "HIV and AIDS: 20 years of science." *Nature Medicine* 839-843.

Fenner, T.E., J. Garweg, F.T. Hufert, M Boehnke, and H Schmitz. 1991. "Diagnosis of human cytomegalovirus-induced retinitis in human immunodeficiency virus type 1-infected subjects by using the polymerase chain reaction." *Journal of Clinical Microbiology* 2621-2622.

Fisseh, B, B Petros, and T WoldeMichael. 1998. "Cryptosporidium and other parasites in Ethiopian AIDS patients with chronic diarrhoea." *East African Medical Journal* 100-101.

Hemelaar, Joris, Eleanor Gouws, Peter D. Ghys, and Saladin Osmanov. 2011. "Global trends in molecular epidemiology of HIV-1 during 2000–2007." *AIDS* 679-689.

Hemelaar, Joris, Ramyiadarsini Elangovan, Jason Yun, Leslie Dickson-Tetteh, Isabella Fleminger, Shona Kirtley, Brian Williams, Eleanor Gouws-Williams, and Peter, D. Ghys. 2019. "Global and regional molecular epidemiology of

HIV-1, 1990–2015: a systematic review, global survey, and trend analysis." *The Lancet Infectious Diseases* 143-155.

Hirsch, Vanessa M., George Dapolito, Robert Goeken, and Barbara J. Campbell. 1995. "Phylogeny and natural history of the primate lentiviruses, SIV and HIV." *Current Opinion in Genetics & Development* 798-806.

Kandathil, A. J., S. Ramalingam, R. Kannangai, Shoba David, and G. Sridharan. 2005. "Molecular epidemiology of HIV." *Indian Journal of Medical Research* 333-344.

Kohárýová M, Kolárová M. 2008. "Oxidative stress and thioredoxin system." *General Physiology and Biophysics* 71-84.

Lifson, Alan R. 1988. "Do Alternate Modes for Transmission of Human Immunodeficiency Virus Exist?" *JAMA* 1353-56.

Malik, U, A Javed, A Ali, and K Asghar. 2017. "Structural and functional annotation of human FAM26F: A multifaceted protein having a critical role in the immune system." *Gene* 66-75.

Malik, Uzma, and Aneela Javed. 2016. "FAM26F: An Enigmatic Protein Having a Complex Role in the Immune System." *International Reviews of Immunology.*

McLaren, Paul, J., and Mary Carrington. 2015. "The impact of host genetic variation on infection with HIV-1." *Nature Immunology* 577-583.

Nakamura, H, S, D Rosa, M Roederer, M, T Anderson, J, G Dubs, J Yodoi, A Holmgren, L, A Herzenberg, and L, A Herzenberg. 1996. "Elevation of plasma

thioredoxin levels in HIV-infected individuals." *International Immunology* 603-611.

Pereira-Chioccola, Vera, Lucia, José, Ernesto Vidal, and Chunlei Su. 2009. "Toxoplasma gondii infection and cerebral toxoplasmosis in HIV-infected patients." *Future Medicine.*

Phair, John, Alvaro Muñoz, Roger Detels, Richard Kaslow, Charles Rinaldo, and Alfred Saah. 1990. "The Risk of Pneumocystis carinii Pneumonia among Men Infected with Human Immunodeficiency Virus Type 1." *The New England Journal of Medicine* 161-165.

Read, JS, R Tuomala, E Kpamegan, C Zorrilla, S Landesman, G Brown, M Vajaranant, H Hammill, and B Thompson. 1999. "Mode of delivery and postpartum morbidity among HIV-infected women: the women and infants transmission study." *Journal of Acquired Immune Deficiency Syndromes* 236-245.

Royce, Rachel A., Arlene Seña, Willard Cates, and Myron S. Cohen. 1997. "Sexual Transmission of HIV." *The New England Journal of Medicine* 1072-78.

Sharp P.M, Hahn B.H. 2011. "Origins of HIV and the AIDS pandemic." *Cold Spring Harbor Perspectives in Medicine.*

Sinoussi, Prof Françoise Barré. 1996. "HIV as the cause of AIDS." *THE LANCET* 31-35.

The World Bank. 2012. *HIV/AIDS in Pakistan.* July 10. https://www.worldbank.org/en/news/feature/2012/07/10/hiv-aids-pakistan.

Vijay, N.N.V, Ajmani, R Vasantika, A.S Perelson, and N.M Dixit. 2008. "Recombination increases human immunodeficiency virus fitness, but not necessarily diversity." *Journal of General Virology* 1467-1477.

World Health Organization. 2018. *Global Health Observatory (GHO) data - HIV/AIDS.* https://www.who.int/gho/hiv/en/.