VARIATION IN SIX INNATE IMMUNE GENES DOES NOT DETERMINE PROTECTIVE IMMUNITY AGAINST SIV IN RHESUS MACAQUES

by

Blair Morgan Gleeson

B.S., Allegheny College, 2010

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Blair M. Gleeson

It was defended on

March 29, 2012

and approved by

Thesis Advisor: Jeremy Martinson, DPhil
Assistant Professor
Infectious Diseases and Microbiology
Graduate School of Public Health
University of Pittsburgh

Todd A. Reinhart, ScD
Professor
Infectious Diseases and Microbiology
Graduate School of Public Health
University of Pittsburgh

Michael Murphey-Corb, PhD
Professor
Department of Microbiology and Molecular Genetics
School of Medicine
University of Pittsburgh

Copyright © by Blair M. Gleeson

2012

VARIATION IN SIX INNATE IMMUNE GENES DOES NOT DETERMINE PROTECTIVE IMMUNITY AGAINST SIV IN RHESUS MACAQUES

Blair M. Gleeson, M.S.

University of Pittsburgh, 2012

HIV vaccine trials in humans have shown variability in vaccine responses among individuals. The SIV:macaque model can be used to understand the outcome of HIV vaccine trials in humans, specifically, the role of the innate immune system in inducing protective immunity. *CXCL13*, *CCL20*, *CCL1*, *CCL2*, *CCL5*, and *TRIM5α*, molecules of the innate immune system, were selected as part of a study into the relationship between these molecules and vaccine response. Sequence variation in the innate immune genes was determined in 25 rhesus macaques that were part of a SIV vaccine trial, a multiple low dose exposure study, or a control group. PCR primers were designed, based on the published rhesus macaque genome, to amplify approximately 2 kilobases of promoter region and the exons of the genes. Each amplicon was sequenced in at least 8 animals, and the sequences were visually screened for single nucleotide polymorphisms (SNPs). Numerous SNPs were detected in all of these genes. Statistical analysis showed that there was no correlation between polymorphisms in these genes and protection against SIV challenge. Copy number variation (CNV) was also examined in these genes using either custom TaqMan copy number assays or SYBR Green assays and quantitative real-time PCR. The results of the TaqMan assay indicate that *CXCL13* demonstrates CNV. Genetic variation in these innate immune genes is likely not important in rhesus macaque response to SIV challenge, although larger studies are needed to provide more definitive results. The public Health significance of this study is that understanding genetic factors that influence macaque

responses to vaccination or SIV challenge will help us to understand HIV in humans with the goal of developing a more effective vaccine and the ability to predict individual’s susceptibility to disease or response to vaccination.

TABLE OF CONTENTS

PREFACE XIII

1.0 INTRODUCTION 1

1.1 HUMAN IMMUNODEFICIENCY VIRUS 1

1.1.1 Epidemiology 1

1.1.2 Infection 1

1.1.3 Origin 2

1.2 SIMIAN IMMUNODEFICIENCY VIRUS 3

1.2.1 Pathogenic vs. Non-pathogenic Infection 3

1.2.1.1 Acute Infection 3

1.2.1.2 Key Differences between Pathogenic and Non-pathogenic SIV infection 4

1.2.1.3 Chronic Infection 4

1.2.1.4 Relevance to HIV research 5

1.2.2 SHIV 6

1.2.3 Vaccine Studies 6

1.2.4 Multiple Low Dose Exposure Studies 7

1.3 HUMAN VACCINE TRIALS 8

1.4 HOST GENETICS AND DISEASE 9

1.4.1 Known Genetic Factors that Influence HIV-1 9
1.4.2 Single Nucleotide Polymorphisms (SNPs) 10
1.4.3 Copy Number Variation (CNV) 11
1.5 INVESTIGATING GENETIC VARIATION IN RHESUS MACAQUE INNATE IMMUNE GENES 12
2.0 STUDY OVERVIEW 14
2.1 HYPOTHESIS 14
2.1.1 Specific Aim 1: Characterize genetic variation 14
2.1.2 Specific Aim 2: Determine relationship between genetic variation and protective immunity 15
3.0 MATERIALS AND METHODS 16
3.1 SAMPLES 16
3.2 WHOLE GENOME AMPLIFICATION 18
3.3 POLYMERASE CHAIN REACTION (PCR) 19
3.3.1 Primer Design 19
3.3.2 PCR 19
3.4 GEL ELECTROPHORESIS 21
3.5 SANGER SEQUENCING 22
3.5.1 ExoSAP 22
3.5.2 BigDye 23
3.5.3 Ethanol/EDTA precipitation 24
3.6 SNP DETECTION 24
3.7 STR GENOTYPING 25

3.8 COPY NUMBER VARIATION ANALYSIS ... 25
3.8.1 TaqMan method ... 25
3.8.2 SYBR Green method ... 26
4.0 RESULTS ... 28
4.1 AIM 1: CHARACTERIZE GENETIC VARIATION IN RHESUS MACAQUE INNATE IMMUNE GENES ... 28
4.1.1 SNP detection ... 28
4.1.1.1 *CXCL13* ... 28
4.1.1.2 *CCL20* ... 30
4.1.1.3 *CCL1* ... 31
4.1.1.4 *CCL2* ... 32
4.1.1.5 *CCL5* ... 33
4.1.1.6 *TRIM5α* ... 34
4.1.1.7 Amino Acid Changes ... 39
4.1.2 CNV ... 39
4.2 AIM 2: DETERMINE THE RELATIONSHIP BETWEEN VARIATION IN INNATE IMMUNE GENES AND VACCINE RESPONSE ... 42
4.2.1 SNPs ... 42
4.2.2 CNV ... 42
5.0 DISCUSSION ... 43
5.1 OVERVIEW OF CURRENT STUDY ... 43
5.2 SNP DETECTION ... 43
5.3 COPY NUMBER VARIATION ... 45

5.4 FUTURE DIRECTIONS ... 46

5.5 PUBLIC HEALTH IMPLICATIONS ... 47

APPENDIX A: PCR PRIMERS AND PROBES ... 48

APPENDIX B: SNP DISTRIBUTIONS ... 50

BIBLIOGRAPHY ... 59

LIST OF TABLES

Table 1: Original 25 Study Animals 17

Table 2: Newly Included 19 Study Animals 18

Table 3: PCR Reagents 20

Table 4: PCR Gradient Program 20

Table 5: ExoSAP Reagents 22

Table 6: ExoSAP Program 22

Table 7: BigDye Reagents 23

Table 8: BigDye Program 23

Table 9: TaqMan Assay Reagents 26

Table 10: SYBR Green Assay Reagents 27

Table 11: *CXCL13* SNPs 29

Table 12: *CCL20* SNPs 31

Table 13: *CCL1* SNPs 32

Table 14: *CCL2* SNPs 33

Table 15: *CCL5* SNPs 34

Table 16: *TRIM5α* SNPs 35

Table 17: *TRIM5α* allelic classes 38

Table 18: Chemokine SNP summary 39

Table 19: CNV Primers and Probes 48

Table 20: PCR primers 49

Table 21: *CXCL13* SNP Distribution 50

Table 22: *CCL20* SNP Distribution 52

Table 23: *CCL1* SNP Distribution 53

Table 24: *CCL2* SNP Distribution 54

Table 25: *CCL5* SNP Distribution 55

Table 26: *TRIM5α* SNP Distribution 56

LIST OF FIGURES

Figure 1: Rhesus macaque *CXCL13* gene 29

Figure 2: Rhesus macaque *CCL20* gene 30

Figure 3: Rhesus macaque *CCL1* gene 32

Figure 4: Rhesus macaque *CCL2* gene 33

Figure 5: Rhesus macaque *CCL5* gene 34

Figure 6: Rhesus macaque *TRIM5α* gene 35

Figure 7: CNV Quantitative Real-Time PCR 40

Figure 8: *CXCL13* Copy Number Variation 41

PREFACE

First, I would like to thank my advisor, Dr. Jeremy Martinson, and the rest of my thesis committee, Dr. Todd Reinhart and Dr. Michael Murphey-Corb for their support, guidance, and advice over the past two years. I would also like to thank the students, faculty, and staff of the department of Infectious Diseases and Microbiology for providing me with the opportunity and environment to conduct this research. Specifically, I would like to acknowledge the members of the Martinson Lab: Matthew Nicholaou, Rebecca Bosko Marino, and Gregory Joseph for their help, support, and friendship. Finally, I would like to thank my parents Robert and Linda Gleeson, my fiancé Mark Erdeljac, and the rest of my family: Garrett Gleeson, Travis Petruso, and Racheal Keener for all of their love and support.

1.0 INTRODUCTION

1.1 HUMAN IMMUNODEFICIENCY VIRUS

1.1.1 Epidemiology

Human immunodeficiency virus (HIV) is the causative agent of acquired immunodeficiency syndrome (AIDS), which is one of the most important global epidemics of our time. Currently, over 34 million people live with HIV worldwide. In 2010, 2.7 million people became infected with HIV, and 1.8 million people died of AIDS(61). It is the leading cause of death in Africa, and the fourth leading cause of death worldwide(53).

1.1.2 Infection

HIV is a blood-borne pathogen and can be transmitted via several modes including mother to child, needle sharing among injection drug users, receiving contaminated blood products, or sexual transmission(17, 53). Initially the viral glycoprotein, gp120, attaches to the cellular receptor CD4 which is primarily found on helper T cells and macrophages. This binding is necessary, but it is insufficient for viral entry. The virus must also bind to a coreceptor. Most sexually transmitted HIV strains use the CCR5 coreceptor that is found on memory CD4+ T cells and macrophages. In some cases, the virus will use CXCR4 as a coreceptor which is found on

many T cells and macrophages. CXCR4 becomes more important later in infection as the virus shifts its tropism to this coreceptor. Once the virus has established infection, there is a period of rapid viral replication during which CD4+ T cells of the mucosa are depleted(17, 28, 53, 56). Peak viremia usually occurs about 10-14 days after infection. At about 14-21 days, a robust antiviral cytotoxic T lymphocyte (CTL) response is observed, and plasma viral titers decrease. Virus specific antibodies emerge later around 6-8 weeks after infection. This HIV-specific immune response helps set a plasma viral load, or set point, that is a predictor of the rate of progression to AIDS. Although the virus elicits a strong cellular and humoral immune response, it is not enough to clear the host of the virus. The highly error-prone reverse transcriptase enzyme of HIV introduces mutations into the viral genome resulting in escape mutants that are able to evade the immune system and antiretroviral drugs(28, 53, 56).

1.1.3 Origin

It is thought that the origin of HIV was from lentiviruses that infected primate species in Africa. There are two strains of HIV that exist today, HIV-1 and HIV-2. HIV-2 infection is far less common and less pathogenic(53). It is thought that SIVsm, which naturally infects sooty mangabeys (*Cercocebus atys*) in West Africa, infected humans about 50-60 years ago and evolved into HIV-2. This hypothesis is supported because sooty mangabeys that are naturally infected by SIVsm live in West Africa, HIV-2 infection of humans is prevalent in West Africa, and the genetic sequences of SIVsm and HIV-2 share a high degree of homology(21). HIV-1 infection of humans is more common than HIV-2 and is the primary cause of AIDS(53). The exact origin of HIV-1 is less obvious, but it is widely accepted that it also originated from an African primate lentivirus(53). Because HIV-1 can be further divided into different subtypes, it

is hypothesized that HIV-1 actually originated via several independent infections of humans by primate lentiviruses. Strains of SIVcpz from chimpanzees in Gabon are genetically similar to HIV-1, thus it is accepted that HIV-1 originated through cross-species transmission of SIVcpz from chimpanzees to humans(57).

1.2 SIMIAN IMMUNODEFICIENCY VIRUS

1.2.1 Pathogenic vs. Non-pathogenic Infection

Lentivirus infections of nonhuman primates have provided us with models to understand HIV pathogenesis better in humans. With regard to disease, two models of SIV infection exist: nonpathogenic and pathogenic. In Africa, dozens of species of primates are infected with species-specific lentiviruses(32). SIV infection of these natural hosts, including sooty mangabeys, African green monkeys, chimpanzees, and mandrills is typically nonpathogenic. Although these animals become infected and sustain high viral replication, they remain healthy and do not progress to AIDS. However, infection of Asian-origin primates (non-natural hosts), such as rhesus macaques, with SIV is pathogenic and often leads to the progression of AIDS(6, 7, 51).

1.2.1.1 Acute Infection

Although pathogenic and non-pathogenic SIV models represent infection that results in different disease outcomes, they share many similarities, especially during acute infection. In both models, acute infection is associated with rapid viral replication with peak virus titers occurring

about 1-2 weeks post infection. Virus levels then dramatically decline, usually by 2 log_{10} or more(7, 51). Acute infection is also characterized by a strong innate immune response to the virus. In the blood and in lymphoid tissues there is an upregulation of interferon-stimulated genes (ISGs)(6, 32). Both models also show elevated levels of CD8+ T cell apoptosis(32). Additionally, activated CD4+ T cells that express the coreceptor CCR5 are the main target cells that support viral replication, and subsequently become depleted. In both models, the majority of the depletion of these CD4+ T cells occurs in the mucosa. Thus, viral replication kinetics and CD4+ T cell deletion during acute infection does not explain the difference in disease outcome between African and Asian nonhuman primates (7, 51).

1.2.1.2 Key Differences between Pathogenic and Non-pathogenic SIV infection

There are, however, key differences between pathogenic and non-pathogenic SIV infection. During non-pathogenic infection peripheral CD4+ T cell levels do not decline, there is less immune activation, and there is less expression of CCR5 on CD4+ T cells(7). In the vast majority of natural SIV infections the host maintains near normal levels of peripheral CD4+ T cells, although mucosal CD4+ T cells are depleted during acute infection(6, 7, 32, 51). Even in the small fraction of natural SIV infections that result in severe peripheral CD4+ T cell depletion, the animals do not progress to AIDS(51). Of note, natural hosts are able to preserve near normal levels of Th17 T cells in the gut, which may play an important role in preventing progression to AIDS(7, 32).

1.2.1.3 Chronic Infection

Although immune responses during acute infection is strikingly similar in pathogenic and non-pathogenic SIV infection, the level of immune activation during chronic infection is quite

different. In the pathogenic model, animals maintain a high level of immune activation throughout the chronic stage of infection. There are high levels of T cell turnover, activation, and apoptosis(7). This chronic immune activation is thought to affect CD4+ T cell homeostasis and immune function by increasing the number of CD4+ T cells that can be targeted by the virus, exhaustion of CD8+ T cells, abnormal T cell apoptosis, and suppression of lymphocyte regeneration(6). Natural hosts, on the other hand, demonstrate little immune activation after acute infection. Specifically, they show lower levels of bystander apoptosis and cell cycle dysregulation(51). The mechanisms that underlie this low level of immune activation in the face of viral replication is not well understood, but it is suspected that it is one of the key features of natural infection that protects these animals from progressing to AIDS. The proportion of CD4+ T cells expressing CCR5 is quite different between pathogenic and non-pathogenic models. In non-natural hosts (and humans) about 10-20% of CD4+ T cells in the blood and more than 50% in the mucosa express CCR5. In natural hosts, however, only about 1-5% of CD4+ T cells in both the blood and mucosa express CCR5. Although this lower expression of CCR5 does not protect natural hosts from becoming infected, nor does it prevent viral replication, it may play a role in maintaining CD4+ T cell homeostatsis(7, 32, 51).

1.2.1.4 Relevance to HIV research

Both pathogenic and non-pathogenic models of SIV infection are important in understanding HIV pathogenesis and disease progression. The natural host model can help us to better understand the immunological factors that promote a non-progressive form of infection, while non-natural hosts (such as rhesus macaques) provide a valuable animal model that closely mimics HIV-1 infection in humans. Macaque infection by SIV, and HIV-1 infection in humans share numerous similarities including infection by mainly CCR5-tropic viruses, severe and rapid

depletion of CD4+ T cells in the gut, viral replication in both activated and resting T cells, resolution of acute infection by antigen-specific immune response, viral immune evasion, sustained viral load in plasma after acute infection predicting disease progression, and peripheral CD4+ T cell depletion leads to progression to AIDS(22).

1.2.2 SHIV

Although the macaque:SIV model provides us with a good model of HIV-1 infection in humans, it is not a perfect representation. For example, HIV-1 vaccines cannot be directly assessed in this model because at the DNA level SIV is about 80% homologous to HIV-2, and about 40-50% homologous to HIV-1. Additionally, the strains of SIV used in macaque studies have been selected by researchers to be highly virulent and result in rapid progression to AIDs. To help ameliorate some of these problems and make vaccine testing in macaques more relevant, HIV-1/SIV chimeric viruses (SHIV) have been created. Desired genes from HIV-1, such as *tat*, *rev*, *vpu*, and *env*, are introduced into the genome of a SIV strain. Such SHIV strains usually demonstrate similar pathogenesis to SIV in the macaque model, but SHIV offers the advantage of allowing for HIV-1 vaccine testing(22).

1.2.3 Vaccine Studies

Macaques are also used in SIV vaccine studies. These vaccine studies are designed to study the efficacy of different vaccine strategies that could be employed by future HIV-1 vaccines. One vaccine strategy that has been tested in the macaque model is live-attenuated vaccines. The vaccine of this model that has probably shown the most success is SIVmac239Δnef in which the

nef gene has been deleted. SIVmac239Δnef was able to protect adult animals from infection upon challenge with SIVmac239 or SIVmac251. It was much less successful, however, in preventing infection by heterologous strains of SIV. There are also safety concerns with live-attenuated vaccines that make them a less than ideal candidate for HIV-1 vaccines (22). Several prime-boost vaccine strategies have also been tested in macaques with varying success. In the prime-boost strategy, the animal is first given a recombinant virus vector and then boosted with either protein subunits or DNA(22). Other studies have shown promising results with the use of particle-mediated (PMED) DNA vaccines. For example, Fuller and colleagues demonstrated that a SIV DNA vaccine resulted in protective immunity upon mucosal challenge with SIV in a number of monkeys. This PMED DNA vaccine contained *gag-pol-env* sequences from SIV strain 17E-Fr. Rhesus macaques were given multiple immunizations via the epidermis 12 – 14 weeks apart. Immunizations resulted in increased levels of cytotoxic T lymphocytes. When challenged intrarectally with a high dose of SIV strain DeltaB670, four of seven vaccinated animals were protected against infection, while only one of 10 control animals did not become infected(16).

1.2.4 Multiple Low Dose Exposure Studies

Rhesus macaques have been used in several different experimental designs to study HIV-1 in humans. Multiple low-dose exposure studies provide a useful model for sexual transmission of the virus in humans. In most macaque vaccine studies, the monkeys are inoculated with high titer inocula to ensure consistent infection, because such animal studies are expensive and thus utilize a limited number of animals. Such high titer inocula are not representative of sexual transmission, making multiple low dose exposure studies important to understanding sexual

transmission of HIV(22, 33). Similarly, rhesus macaques have been shown to acquire protective immunity following multiple mucosal exposures to SIV. In one multiple low dose challenge study, researchers administered several low dose challenges intrarectally to rhesus macaques. When these monkeys were then challenged with a high dose of SIV intrarectally, four out of eleven were resistant to infection compared to zero of four control animals(34).

1.3 HUMAN VACCINE TRIALS

Vaccine trials conducted in both macaques and humans have resulted in protective immunity in up to about 30% of individuals(24). To date, there have been over 163 HIV-1 vaccine clinical trials conducted in humans(33). Very few, however, have yielded promising results. One such study was an HIV vaccine efficacy trial in Thailand where 16,402 healthy adults were either administered a placebo, or four priming shots of a recombinant canarypox vector vaccine and two booster shots of a recombinant glycoprotein 120 subunit vaccine (ALVAC-HIV [Vcp1521], and AIDSVAX B/E, respectively). HIV-1 infection was monitored in this cohort at regular time intervals for three years. At the conclusion of the study, the investigators found vaccine efficacy to be about 30%. Although the vaccine showed a modest effect in preventing HIV-1 infection, it did not influence viral load or CD4+ T cell counts in those who became infected(33, 40). The immunological mechanisms underlying the moderate success of this vaccine are unclear. Although neutralizing antibodies were not elicited by this vaccine, binding antibodies against the HIV-1 envelope were produced. These antibodies may have promoted antibody-dependent cell-mediate cytotoxicity, which may have afforded some protection against HIV-1. The vaccine also did not elicit a CD8+ T cell response, although it did stimulate a strong CD4+ T cell

response(33). Although the exact mechanisms of its action are unclear, this vaccine remains the most effective HIV-1 vaccine to date, eliciting protective immunity in about 30% of individuals.

1.4 HOST GENETICS AND DISEASE

1.4.1 Known Genetic Factors that Influence HIV-1

In both humans and macaques, vaccines have conferred protective immunity in about one third of individuals. The reason for this variability in vaccine response is not well understood, but one explanation may be host genetic factors. Although host genetics are often overlooked in infectious diseases, they can greatly influence factors such as immunity and virus replication. For example, the CCR5Δ32 mutation in humans has been shown to profoundly influence HIV disease outcome. CCR5 is the coreceptor used to gain entry into target cells by most transmitted strains of HIV-1, and is found on the surface of activated T cells as well as macrophages, dendritic cells, and microglial cells(52). The CCR5Δ32 mutation is a 32 basepair deletion in the *CCR5* gene that results in a frameshift that introduces a premature stop codon. As a result, the CCR5 protein is truncated and not expressed on the cell surface, eliminating the coreceptor used by CCR5 tropic viruses. People who are homozygous for this mutation are virtually resistant to infection by HIV while people who are heterozygous for the mutation tend to progress slower to AIDS by about 2-4 years(1, 9, 37, 52). The CCR5Δ32 mutation is most common among northern Europeans at a frequency of about 15%(9) (although only 1% are homozygous for the mutation(1)). It is also found at a frequency of 0-12% in different Asian populations, but is nearly absent among Africans(1, 9, 37, 52). Alternatively, mutations within the same *CCR5*

gene are associated with more rapid progression to AIDS. The promoter region of the gene contains a series of single nucleotide polymorphisms (SNPs) that have been organized into distinct haplotypes. Individuals that are homozygous for the *CCR5P1* allele progress to AIDS more quickly(9, 52).

Chemokines produced by the host may also play a role in disease outcome. These small proteins, produced by cells in the immune system, regulate the trafficking of other cells via chemotaxis. Their targets are seven-transmembrane G-protein coupled receptors on cells of both the innate and adaptive immune systems including neutrophils, monocytes, and lymphocytes(39, 59, 60). Some chemokines, specifically β-chemokines, have been shown to greatly influence the outcome of HIV exposure. The β-chemokines, MIP-1α (CCL3), MIP-1β (CCL4), and RANTES (CCL5) actually prevent the infection of CD4+ T cells by HIV. These chemokines are the natural ligands of CCR5, the main coreceptor used by HIV-1 in viral entry. Because they are ligands of CCR5, they compete with CCR5-tropic viruses for the receptor, thus blocking viral entry. Several polymorphisms in these genes contribute to their levels of expression which correlate with rate of progression to AIDS(1, 37, 41).

1.4.2 Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are a common type of genetic variation in the human genome that can influence disease. SNPs are single basepair mutations in the DNA. Typically they are biallelic, meaning there is either one of two possible nucleotides at the particular position. A SNP can arise by one of two kinds of mutations, transitions or transversions. Transitions are purine to purine (adenine and guanine) or pyrimidine to pyrimidine (cytosine and thymine) mutations, while transversions are purine to pyrimidine or pyrimidine to purine

mutations. Transition mutations are more common. Most of these mutations are found in introns or between genes, and are not important. If a SNP lies in an exon of a gene, it may result in an amino acid change, altering the protein. Similarly, SNPs in promoter regions may affect the binding of transcription factors, varying the expression of the protein. Although single basepair mutations are fairly common in the human genome, they are only considered SNPs if the minor allele frequency is at least 1% in the population(23). So far, over 1.42 million SNPs have been detected in the human genome(23, 46). Several of these SNPs have been shown to affect HIV-1 susceptibility and progression to AIDS. The CCR2-V64I mutation is a guanine to adenine transition in the DNA that changes an amino acid from a valine to an isoleucine. Although the exact mechanism of action is unclear, this mutation is associated with a slower progression to AIDS. Several SNPs within the gene encoding *CCL5*, a natural ligand for CCR5, alter the expression of the protein and affect the rate of progression to AIDS. The In1.1C SNP in the intronic regulatory region of the gene decreases the expression of the protein and results in a more rapid progression to AIDS(1, 9, 37, 41). These are just a few of the many SNPs that have been found to be associated with susceptibility to infection and progression to disease.

1.4.3 Copy Number Variation (CNV)

Another type of genetic variation that has been shown to influence disease in humans is copy number variation (CNV). CNV is characterized by large deletions or duplications that involve DNA fragments of at least 1 kilobase that alter the number of copies of a gene in the genome(15). Such CNV has been detected in the human genome, especially in genes involved in immunity. CNV in some of these genes has been found to have effects on disease outcome. Probably the best characterized example in humans is CNV in *CCL3L1* and the effect on HIV-1

infection and progression to AIDS(18). As mentioned earlier, CCL3 (MIP-1α) is one of the natural ligands for CCR5, the primary coreceptor used by HIV-1(1, 37, 41). *CCL3L1*, or *MIP-1αP*, is the copied isoform of the gene that encodes MIP-1α. Like CCL3, CCL3L1 is also a ligand for CCR5, and is actually the most potent of the known ligands for this receptor. Examination of CNV in *CCL3L1* revealed both interindividual and interpopulation variation. Low copy number of *CCL3L1*, relative to the geographic ancestral population mean, was associated with increased susceptibility to HIV-1 infection, and more rapid progression to AIDS. Alternatively, additional copies of the *CCL3L1* gene, compared to the geographic ancestral population mean, were associated with a decreased risk of HIV-1 infection, and a slower progression to AIDS(18). Similarly, CNV has been detected within the rhesus macaque genome(14, 29). Like humans, CNV in *CCL3*-like genes correlates with rate of progression to AIDS(14).

1.5 INVESTIGATING GENETIC VARIATION IN RHESUS MACAQUE INNATE IMMUNE GENES

The innate immune system is now recognized as a factor that may exert a strong influence over the host response to infection and to vaccines. In the current study, we are using the SIV:macaque model to explore the role of the innate immune system in inducing protective immunity. *CXCL13*, *CCL20*, *CCL1*, *CCL2*, *CCL5*, and *TRIM5α*, molecules of the innate immune system, have been selected from a list of other innate immune genes to begin investigating the relationship between these molecules and vaccine response. Although the rhesus macaque is a good model system in which to study vaccine response, little is known about

the extent of genetic variation in this species. This contrasts strongly with the extensive data available on genetic variation in humans. We have begun to address this deficiency by resequencing many innate immune response genes in a cohort of rhesus macaques previously studied for two models of vaccine responsiveness.

1. A model of **mucosal viral challenge**, in which animals were exposed mucosally to a low dose of SIV, followed later by a higher dose of virus. In this study, animals were divided into those who were protected as a result of the earlier low-dose challenge, and those for whom this did not confer protection

2. A vaccine model, using a particle-mediated (**PMED**) DNA vaccine. In this study, animals who received the vaccine were divided into ones that were protected against subsequent high-dose challenge and those that were not protected. This study also included a control group that did not receive a vaccine.

This study of variation in rhesus macaque innate immune genes is the first of its kind. Hopefully, the characterization of genes in this study can lay the groundwork for future studies to further elucidate the relationship between variation in the innate immune system and immune response. Understanding genetic factors that influence macaque responses to SIV vaccines will help us to better understand HIV in humans with the goal of developing a more effective vaccine and the ability to predict an individual's susceptibility to disease or response to vaccination.

2.0 STUDY OVERVIEW

2.1 HYPOTHESIS

Very little is known about genetic variation in rhesus macaques, making this study of variation in macaque chemokine genes, the first of its kind. Thus, the primary goal of this study is to characterize genetic variation by resequencing many innate immune response genes in a cohort of rhesus macaques. We hypothesize that variation in rhesus macaque chemokine genes influences protective immunity against SIV challenge.

2.1.1 Specific Aim 1: Characterize genetic variation

Characterize variation in rhesus macaque chemokine genes by identifying single nucleotide polymorphisms (SNPs) and determining copy number variation (CNV). Genomic DNA from 25 rhesus macaques will be used to sequence genes of the innate immune system and identify polymorphisms which may control expression. For these same genes, quantitative real time PCR will be used to determine CNV in the same cohort of 25 macaques.

2.1.2 Specific Aim 2: Determine relationship between genetic variation and protective immunity

Determine association between variation in innate immune genes and vaccine response. The polymorphisms and copy number variation these innate immune genes will be analyzed in the cohort of 25 rhesus macaques that were part of a SIV vaccine trial, a multiple low dose exposure study, or a control group with the goal of determining the relationship between variation in these genes and protective immunity.

3.0 MATERIALS AND METHODS

3.1 SAMPLES

DNA obtained from venous blood drawn from 25 outbred Indian-origin rhesus macaques was provided by Dr. Murphey-Corb. The macaques were part of a SIV vaccine trial (PMED), a multiple low dose exposure study (Mucosal), or a control group (Table 1). This genomic DNA was initially used to make a working stock of DNA for each sample at a final concentration of 5ng/µL. Because genomic DNA for these samples was limited, whole genome amplification was performed, using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare), to create a larger quantity of DNA for each animal. Although this GenomiPhi DNA is reliable for sequencing and SNP detection, it is not ideal for use in CNV assays. Thus, we obtained venous blood samples from 19 newly recruited Chinese-origin rhesus macaques (Table 2) from which genomic DNA was isolated. This genomic DNA was used for CNV assays. As these animals were newly recruited, their disease outcome is not yet available for analysis.

Table 1: Original 25 Study Animals

Study	Group	Animal #	Code #	Exposure Outcome	Challenge Outcome
PMED	Vaccinates	8097	SIV04	NA	Not protected
		11797	SIV06	NA	Not protected
		11897	SIV03	NA	Not protected
		12097	SIV05	NA	Protected
		12297	SIV07	NA	Protected
		7697	SIV02	NA	Protected
		7797	SIV01	NA	Protected
	Controls	7097	SIV14	NA	Infected
		7297	SIV15	NA	Infected
		8497	SIV16	NA	Infected
		698	SIV08	NA	Infected
		1798	SIV09	NA	Infected
		2298	SIV10	NA	Infected
		1698	SIV11	NA	Not Infected
Mucosal	MAC239/Colonic	N255	SIV17	Transient	Not Protected
		M224	SIV18	Infected	Not Protected
		M029	SIV19	Transient	Not Protected
		L870	SIV20	Transient	Protected
	17E-FR/Colonic	N041	SIV13	Negative	Protected
		N385	SIV12	Infected	Not Protected
		N138	SIV21	Infected	Not Protected
		M155	SIV22	Uninfected	Not Protected
		L987	SIV23	Uninfected	Not Protected
		M259	SIV24	Infected	Not Protected
		M577	SIV25	Uninfected	Protected

Original study animals were a cohort of 25 Indian origin rhesus macaques that were either part of an SIV vaccine trial (16) or a multiple low dose exposure study (35)

Table 2: Newly Included 19 Study Animals

Animal #	Code #
R0070	m01
R0081	m02
R0063	m03
R0078	m04
R0563	m05
R0701	m06
R0702	m07
R0072	m08
R0705	m09
R0704	m10
R0564	m11
R0697	m12
R0698	m13
R0700	m14
R0037	m15
	m16
R0024	m17
R0074	m18
R0069	m19

Newly Included animals were a cohort of 19 Chinese origin rhesus macaques

3.2 WHOLE GENOME AMPLIFICATION

Because genomic DNA from the original 25 rhesus macaques samples was limited, a whole genome amplification was performed to create a substantial amount of DNA for each sample that could be used for PCR and sequencing. The whole genome amplification was carried out using the general protocol that accompanied the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare). For each sample, 9µL of sample buffer was combined with 1µL DNA (10ng) and

heated to 95°C for 3 minutes and then cooled to 4°C. 10μL of master mix (9μL of reaction buffer, 1μL enzyme mix) was added. The tubes were then incubated at 30°C for 1.5 hours before the enzyme was inactivated by heating to 65°C for 10 minutes and then cooled to 4°C. Gel electrophoresis was used to confirm the success of the amplification. The samples were then diluted 1:100 for use in later PCR.

3.3 POLYMERASE CHAIN REACTION (PCR)

3.3.1 Primer Design

The published rhesus macaque genomic sequence data was used to design PCR and sequencing primers for the *CXCL13*, *CCL20*, *CCL1*, *CCL2*, *CCL5,* and *TRIM5α* genes using Primer3 software. Each pair of primers was designed to yield a PCR product about 400-600 basepairs in length. Enough primers were designed for each gene to cover all exons and about 2 kilobases of upstream promoter region. PCR, sequencing, and SYBR green primers were ordered through Integrated DNA Technologies (IDT).

3.3.2 PCR

For each pair of primers, the optimal magnesium concentration and annealing temperature was determined by running a PCR with a temperature gradient and a magnesium titration. First, 4 PCR master mixes were made, each with a different magnesium concentration (1.0mM, 1.5mM, 2.0mM, and 2.5mM) using AmpliTaq Gold reagents (Applied Biosystems) and dNTPs

(Invitrogen) (Table 3). Each master mix was aliquoted into an 8-tube strip (Phenix) with the final volume in each tube being 25 µL.

Table 3: PCR Reagents

	Final MgCl2 concentration in master mix			
	1.0mM	1.5mM	2.0mM	2.5mM
Reagents	**Volume per reaction (µL)**			
DNA (5ng/µL)	2.0	2.0	2.0	2.0
10X PCR Buffer	2.5	2.5	2.5	2.5
MgCl2 (25mM)	1.0	1.5	2.0	2.5
dNTPs (25mM)	0.2	0.2	0.2	0.2
Primers (10mM)	0.4	0.4	0.4	0.4
AmpliTaq Gold	0.2	0.2	0.2	0.2
Sterile Distilled Water	18.7	18.2	17.7	17.2
Total Volume (µL)	25.0	25.0	25.0	25.0

The four 8-tube strips were put into an Eppendorf Mastercycler and a temperature gradient cycle (Table 4) was run so that each tube in an 8-tube strip was run with a different annealing temperature. The annealing temperatures ranged from 52°C to 62°C.

Table 4: PCR Gradient Program

Step	Temperature (°C)	Time
1	94	2 min
2	94	30 sec
3	52	15 sec
annealing temperature gradient of 10°C		
4	72	30 sec
steps 2-4, repeat 34 times		

Gel electrophoresis was then used to separate the DNA in the PCR products. After UV light was used to visualize the products, optimal PCR conditions were determined by assessing at which magnesium concentration and annealing temperature the cleanest and brightest bands were produced. The optimized PCR conditions were used to amplify DNA in 24 of the 25 macaque

DNA samples with each pair of primers. Twenty-four (instead of all 25) samples were used because 24 is divisible by 8, making it far more convenient for amplification in 8-tube strips and 96-well plates. After each PCR, gel electrophoresis was used to confirm the presence of amplified DNA.

3.4 GEL ELECTROPHORESIS

PCR products were run on a 2% agarose gel, which was made by combining 4g of GenePure LE agarose powder (ISC BioExpress) in 200mLs of 0.5X TBE (ISC BioExpress) buffer and heating in a microwave until all agarose was dissolved. Once removed from the microwave, 10µL ethidium bromide was added to the agarose mixture. The liquid mixture was poured into a form, combs were inserted, and it was allowed to cool and set. 5µL of the PCR product was combined with 5µL 6X loading dye and loaded into the gel, along with one well of a ladder, Phi-X174 RF DNA *Hae* III digest (Thermo Scientific). The loaded gel was run at 100V for 30 minutes in a running buffer of 300mL 0.5X TBE buffer and 15µL ethidium bromide. After the gel was run, it was taken to a gel documentation system and the ethidium bromide-stained DNA was visualized using UV light.

3.5 SANGER SEQUENCING

3.5.1 ExoSAP

After PCR was confirmed to have worked via gel electrophoresis, 10µL of the remaining PCR product was used for sequencing. First, an ExoSAP reaction was run, using an Exonuclease enzyme (NewEngland BioLabs) and an alkaline phosphatase (Roche), to remove any unconsumed dNTPs and primers, which could interfere with the sequencing reaction. For this reaction, a master mix was created with the ExoSAP reagents. For each amplified sample, 10µL of PCR product was pipetted into a well of a 96-well 0.2mL, low profile PCR plate (ISC BioExpress). 10µL of ExoSAP master mix was added to each well with a repeat dispenser. The plate was loaded into an Eppendorf Mastercycler and the ExoSAP program was run.

Table 5: ExoSAP Reagents

Reagents	Volume per reaction (µL)
Exonuclease I	0.05
rAPid Alkaline Phosphatase	1.0
rAPid Alkaline Phosphatase buffer (10X)	1.0
Sterile Distilled Water	7.95
Total Volume (µL)	10.0

Table 6: ExoSAP Program

Step	Temperature (°C)	Time
1	37	60 min
2	85	15 min

3.5.2 BigDye

After the ExoSAP reaction was complete, the clean PCR products were transferred to clean semi-skirted optical 96-well plates (Applied Biosystems) using an 8-channel pipette. From each row of PCR products in the ExoSAP plate, 5µL were pipetted into 2 different rows in the clean plates such that every clean PCR product was represented twice in order to run both a forward and reverse sequencing reaction. A BigDye master mix was created for each individual primer using the BigDye reagents (Applied Biosystems) and a 1:500 dilution of the primer from the original stock, giving a final primer concentration of 10mM. 5µL of the master mix was added to the 5µL PCR product. The plates were loaded into the Eppendorf Mastercyclers and the sequencing reaction was run.

Table 7: BigDye Reagents

Reagents	Volume per reaction (µL)
Primers (10mM)	2.5
5X Dilution Buffer	2.0
BigDye Terminator	0.5
Total volume (µL)	5.0

Table 8: BigDye Program

Step	Temperature (°C)	Time
1	96	2 min
2	96	10 sec
3	50	5 sec
4	60	4 min

steps 2-4, repeat 24 times

3.5.3 Ethanol/EDTA precipitation

After the BigDye reaction was complete, each 10μL sequencing reaction product was transferred to a clean 96-well optical plate using an 8-channel pipette. To each well 5μL of 125mM ethylenediaminetetraacetic acid (EDTA) and 60μL 100% ethanol were added. The plate was sealed and inverted 4 times before incubation at room temperate for 15 minutes in a dark lab drawer, as the BigDye reagents can be degraded by light. Plates were loaded into the centrifuge and spun at 2500 x g for 30 minutes at 4°C. The supernatant was then removed by inverting the plate over the sink and using a flick of the wrist. 60μL of 70% ethanol was then added to each well. The plates were again sealed and spun at 1650 x g for 15 minutes at 4°C. The supernatant was again removed, and the plates were inverted onto a stack of about 5 paper towels and spun in the centrifuge for a few seconds. The plates were then allowed to air dry for about 20 minutes in a dark lab drawer. They were then sealed and sent to the University of Pittsburgh Genomics and Proteomics Core Laboratories facility (GPCL) for capillary electrophoresis.

3.6 SNP DETECTION

To analyze the DNA for polymorphisms, the sequencing data from GPCL was imported into Sequencher software along with the published rhesus macaque reference sequence for each gene. Sequencing data for each gene was organized into its own contiguous assembly with the appropriate reference sequence. The software aligns all of the forward and reverse reads from each sample with the reference sequence and highlights any ambiguities. Each ambiguity was then visually analyzed to determine if it was the result of a genuine SNP, messy sequencing data,

or poor call made by the software. For each SNP, the allele frequencies were compared between the different exposure groups using a t-test for proportions.

3.7 STR GENOTYPING

At the end of the *CXCL13* gene there is a Short Tandem Repeat (STR). In order to characterize this variant region, genotyping was used. Primers were designed, as described previously, to amplify the STR region at the end of the gene. These primers, however, were fluorescently labeled with FAM. PCR was run as described previously, and the PCR products were sent to GPCL in an optical 96-well plate for capillary electrophoresis.

3.8 COPY NUMBER VARIATION ANALYSIS

3.8.1 TaqMan method

Fluorescently-labeled custom TaqMan primers and probes were designed using the Primer3 software program for *CXCL13* and a control gene *STAT6*, which is known to be present at 2 copies per diploid cell. Custom probes have to be designed because Applied Biosystems does not sell ready-made assays for the rhesus macaque, only for human and mouse. To optimize the assay, a standard curve was first prepared, where *STAT6* was present at 300,000 copies, 30,000 copies, 3,000 copies, 300 copies, and 30 copies. To do this, the calculations in the Applied Biosystems worksheet, Creating Standard Curves with Genomic DNA or Plasmid DNA

Templates for Use in Quantitative PCR, were followed. Using these calculations, 5 dilutions of DNA were made from the same DNA sample. A master mix was created for each gene, using Sso Fast reagents (BioRad) and the custom TaqMan probes (Applied Biosystems), and 10µL of this master mix was added to 5µL DNA in a semi-skirted, green 96-well plate (Eppendorf).

Table 9: TaqMan Assay Reagents

Reagents	Volume per reaction (µL)
SsoFast Probes Supermix	7.5
20X probe	0.375
Sterile Distilled Water	2.125
Total Volume (µL)	10

The plate was loaded into an Eppendorf realplex^4 Mastercycler. Each DNA dilution was run in triplicate for both *STAT6* and *CXCL13*. After the PCR program was run, the efficiencies of the probes were assessed and the magnesium concentration was increased in future runs for *CXCL13* in order to bring its efficiency closer to that of *STAT6*. Once the assay was optimized, it was run again using 16 genomic macaque DNA samples diluted to 19.8ng/µL. Each sample was run in at least triplicate for each gene. The results were analyzed using CopyCaller software (Applied Biosystems).

3.8.2 SYBR Green method

The SYBR Green method was used to analyze CNV for the rest of the genes, because custom TaqMan probes were both expensive and took weeks for the company to synthesize. The SYBR Green method uses non-fluorescently labeled primers that are inexpensive and quick to synthesize. The DNA is amplified with these primers in the presence of SYBR Green dye, a fluorescent dye that non-specifically binds double stranded DNA. First, primers were designed,

as described previously, for *CXCL13*, *CCL1*, *STAT6* (control gene, present at 2 copies per diploid genome), and *CCL3L* (known to show copy number in rhesus macaques). It is important that the primers for each gene are designed such that the amplified product from each gene is the same size so that the amount of fluorescence is comparable between genes. For the SYBR green assay, 8μL of master mix, created using SYBR Green reagents (Qiagen), was added to 2μL of DNA in a semi-skirted, green 96-well plate (Eppendorf).

Table 10: SYBR Green Assay Reagents

Reagents	Volume per reaction (μL)
SYBR Green ROX qPCR MasterMix	5.0
Primers	0.2
DNA (5ng/μL)	2.0
Sterile Distilled Water	2.8
Total Volume (μL)	10.0

The same 16 genomic macaque DNA samples that were used in the TaqMan method were used in the SYBR green assay, except they were diluted to 5ng/μL. Each sample was run in triplicate for each gene.

4.0 RESULTS

4.1 AIM 1: CHARACTERIZE GENETIC VARIATION IN RHESUS MACAQUE INNATE IMMUNE GENES

4.1.1 SNP detection

4.1.1.1 *CXCL13*

The *CXCL13* chemokine, also referred to as B lymphocyte chemoattractant (BLC) or B cell-attracting chemokine 1 (BCA-1), is highly chemotactic for B-cells. It is important for proper B-cell trafficking and germinal center formation(2, 30). Although it is known that *CXCL13* is important for B-cell homing and antibody production, its role in SIV pathogenesis is less clear. *CXCL13* is upregulated during infection, and may play a role in trafficking other cell types, such as dendritic cells and follicular helper T cells, to the germinal center(39).

Seven different PCR amplicons were used to examine variation in the exons and promoter region of *CXCL13*. Twenty-nine SNPs were detected in *CXCL13* (Figure 1, Table 11). No SNPs were found in the amplicon containing exons 3 and 4. Overall, 15 SNPs were detected in the promoter region and 2 synonymous SNPs were detected in exons (Table 18). The available genome sequence of *CXCL13* contains a gap of 851bp, which we have sequenced and

will be depositing on Genbank. Genotyping of the STR, CTTn, at the end of the gene revealed that the number of repeats is highly variable among animals.

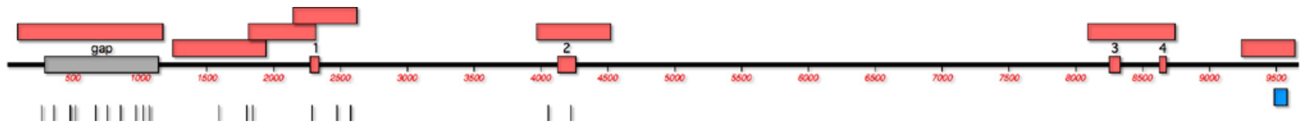

Figure 1: Rhesus macaque *CXCL13* gene

Exons are shown on the scale line. PCR amplicons generated for sequencing are shown above the line, and SNPs discovered are shown as tick marks below the gene. The blue box below the line denotes the CTT STR region.

Table 11: *CXCL13* SNPs

SNP	# animals sequenced	allele frequency
A-2012C	22	1
T-1921C	22	0.55
G-1797A	20	0.4
C-1758T	20	0.53
G-1671A	17	0.06
G-1607A	17	0.53
C-1521T	16	0.44
A-1421T	25	0.52
G-1308G	25	0.5
G-1251A	25	0.54
G-1208T	25	0.52
G-1176A	25	0.54
G-1132C	25	0.96
G-1060T	24	0.44
C-993T	24	0.02
C-956T	24	0.81
ins-932G	20	0.15
G-684A	25	0.56
G-476C	24	0.42
G-432A	24	0.6
G15A	25	0.62
C200T	24	0.58
C302T	24	0.4
C1780A	18	0.47
G1949A	20	0.55

Table 11 continued

C2213T	12	0.54
G6012A	8	0.06
C6232T	8	0.13
C6330T	10	0.9

4.1.1.2 *CCL20*

CCL20 is also known as liver and activation-regulated chemokine (LARC), macrophage inflammatory protein-3α (MIP-3α), and Exodus(49). This chemokine is expressed in the liver and mucosal tissue(20) and is chemotactic for immature dendritic cells and T lymphocytes, subsets of which express the *CCL20* receptor, CCR6(4, 49). The primary role of *CCL20* is the trafficking of lymphocytes and immature dendritic cells to sites in mucosal tissue(10, 11, 55). During pathogenic SIV infection of non-natural hosts, *CCL20* is upregulated in lymph nodes(38).

Seven different amplicons were designed to examine variation in the promoter and exons of *CCL20*. Eighteen SNPs were detected in *CCL20* (Figure 2, Table 12). Eight of these SNPs were found in the promoter and 1 nonsynonymous SNP was detected in an exon (4.1.1.7). No SNPs were found in the amplicon containing exon 4.

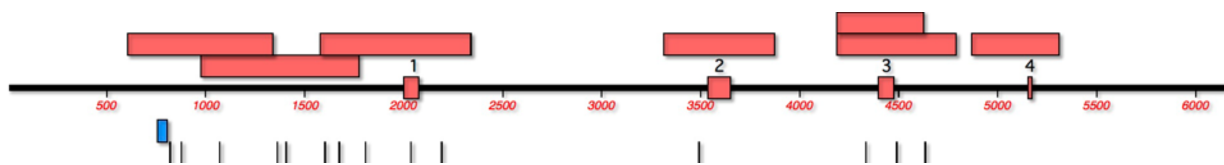

Figure 2: Rhesus macaque *CCL20* gene

Exons are shown on the scale line. PCR amplicons generated for sequencing are shown above the line, and SNPs discovered are shown as tick marks below the gene. The blue box below the line denotes a repeat region identified by the RepeatMasker software.

Table 12: ***CCL20*** **SNPs**

SNP	# animals sequenced	allele frequency
G-1181C	23	0.33
G-1123A	22	0.3
G-1034C	25	0.04
A-930C	25	0.02
C-639T	24	0.06
T-595C	24	0.44
T-448C	25	0.02
A-398G	24	0.42
G-326T	25	0.04
A-251G	25	0.02
G-193T	19	0.21
G-179T	25	0.02
ins-46G	21	0.88
G36T	23	0.11
A191G	22	0.3
A1491G	22	0.55
A2333G	19	0.08
A2489G	16	0.09

4.1.1.3 *CCL1*

CCL1, or human cytokine I-309, is secreted by activated T cells(31). It is chemotactic for monocytes(31), which express the receptor CCR8(43).

Originally, 3 amplicons were designed to examine variation in the promoter region and 2 exons. Ten SNPs were detected in *CCL1* (Figure 3, Table 13). Two of these SNPs were found to be nonsynonymous SNPs in exons (4.1.1.7). The available genome sequence of *CCL1* contains a gap of about 3000bp, which we are currently sequencing and will deposit on Genbank. As sequencing data for this region becomes available, more SNPs are likely to be found.

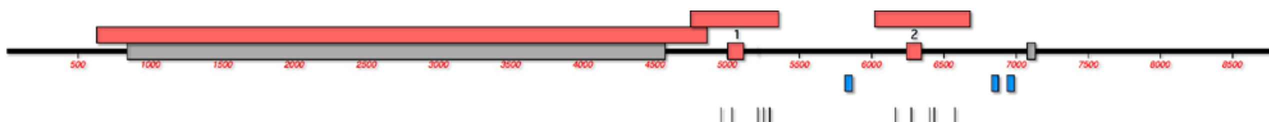

Figure 3: Rhesus macaque *CCL1* gene

Exons are shown on the scale line. PCR amplicons generated for sequencing are shown above the line, and SNPs discovered are shown as tick marks below the gene. Blue boxes below the line denote repeat regions identified by the RepeatMasker software.

Table 13: *CCL1* SNPs

SNP	# animals sequenced	allele frequency
C-45T	23	0.04
T98C	23	0.89
T131G	22	0.16
C257T	23	0.07
C366G	22	0.02
T1238C	20	0.1
G1280A	23	0.24
C1320T	23	0.07
A1501G	23	0.87
C1576T	23	0.65

4.1.1.4 *CCL2*

CCL2, also known by other names including monocyte chemoattractant protein-1 (MCP-1), is chemotactic for monocytes, T cells, and NK cells(8, 19). The receptor for this chemokine is CCR2(27).

Six amplicons were used to examine variation in the promoter region and exons of *CCL2*. Seventeen SNPs were detected in *CCL2* (Figure 4, Table 14), all in the promoter and introns (Table 18).

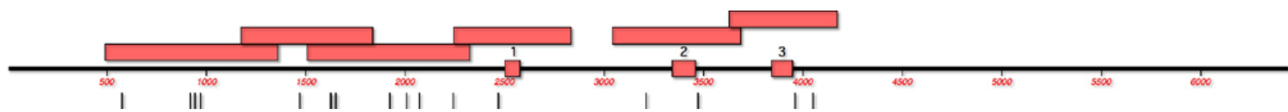

Figure 4: Rhesus macaque *CCL2* gene

Exons are shown on the scale line. PCR amplicons generated for sequencing are shown above the line, and SNPs discovered are shown as tick marks below the gene.

Table 14: *CCL2* SNPs

SNP	# animals sequenced	allele frequency
A-1925G	23	0.93
T-1581C	24	0.6
C-1558T	24	0.58
T-1530C	24	0.04
C-1032T	24	0.04
C-878T	24	0.02
G-874T	24	0.54
C-850T	24	0.04
G-578A	24	0.6
G-495C	24	0.6
G-430A	24	0.1
T-79C	23	0.61
G-33T	20	0.63
C711A	24	0.29
C971T	24	0.29
C1459T	24	0.31
A1550C	23	0.3

4.1.1.5 *CCL5*

CCL5 is also known as RANTES (regulated upon activation, normal T cell expressed and secreted). As its name suggests, this chemokine is expressed by T cells and is also chemotactic for T cells, eosinophils, and basophils(12, 44, 47). During pathogenic SIV infection of non-natural hosts, *CCL5* is upregulated in lymph nodes as well as in the intestinal tissue(39).

Six amplicons were designed to examine variation in the promoter region and exons of *CCL5*. Seven SNPs were detected in *CCL5* (Figure 5, Table 15), all in the promoter and introns (Table 18).

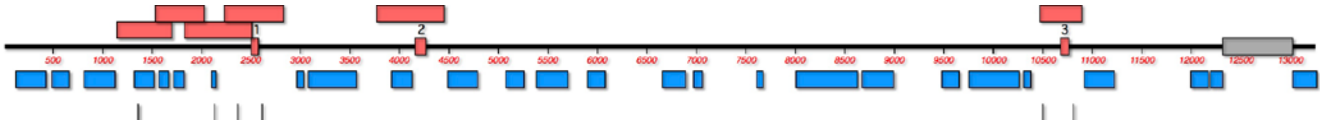

Figure 5: Rhesus macaque *CCL5* gene

Exons are shown on the scale line. PCR amplicons generated for sequencing are shown above the line, and SNPs discovered are shown as tick marks below the gene. Blue boxes below the line denote repeat regions identified by the RepeatMasker software.

Table 15: *CCL5* SNPs

SNP	# animals sequenced	allele frequency
C-1142T	23	0.26
G-1141A	23	0.3
A-339G	22	0.91
A-135G	24	0.81
C113T	23	0.28
G8000T	21	0.55
T8309A	23	0.04

4.1.1.6 *TRIM5α*

TRIM5α (tripartite motif-5alpha) is a cellular restriction factor. This factor is found in most primates and is considered to be an important innate immune defense against retroviruses. *TRIM5α* demonstrates species-specific restriction. In other words, *TRIM5α* of different primate

species prevents infection by different retroviruses. Human *TRIM5α* is able to restrict retroviruses that infect other species. *TRIM5α* of African monkeys is able to prevent infection by HIV-1, but human *TRIM5α* is not(1, 3, 26, 41, 54, 58). The difference between *TRIM5α* that can restrict HIV-1 and *TRIM5α* that cannot is a single amino acid at position 332. In humans, this amino acid position is filled by an arginine while in monkeys it is a proline(58). Although the exact mechanism is unknown, *TRIM5α* is thought to act by interacting with and degrading capsid proteins, inhibiting viral replication(1, 3, 26, 41, 54, 58).

Seven PCR amplicons were designed to examine variation in the exons and untranslated regions of *TRIM5α*. Forty-eight SNPs were detected in *TRIM5α* (Figure 6, Table 16).

Figure 6: Rhesus macaque *TRIM5α* gene

Exons are shown on the scale line in red. 5' and 3' untranslated regions are shown on the scale line in white. PCR amplicons generated for sequencing are shown above the line, and SNPs discovered are shown as tick marks below the gene. Blue boxes below the line denote repeat regions identified by the RepeatMasker software.

Table 16: *TRIM5α* SNPs

SNP	# animals sequenced	allele frequency
G-11T	23	0.98
G42C	23	0.46
A9340C	8	0.25
C9433T	19	0.13
A9491T	13	0.38
G9834A	22	0.18
T10162C	22	0.66
T10479C	22	0.48
G10679T	22	0.27
C10686T	22	0.5

Table 16 continued

A10687G	22	0.5
C11215T	22	0.5
C11219T	22	0.18
C11233T	22	0.48
G11252T	22	0.45
T11324C	13	0.04
C11338A	14	0.04
C11343T	13	0.12
C11350T	13	0.54
T11378C	13	0.12
C11437T	13	0.12
G11443A	13	0.12
G11456A	13	0.12
A11480G	21	0.26
A11500T	21	0.48
C23693T	22	0.02
G23737A	21	0.21
G25393A	16	0.5
A25478G	17	0.47
C25527T	17	0.65
T25651C	19	0.18
G25667T	19	0.18
C25680T	19	0.18
A25689G	19	0.18
C25722T	20	0.18
A25770G	22	0.23
A25823C	22	0.2
C25916T	22	0.18
C25973A	22	0.18
G25987A	22	0.18
C25992T	22	0.21
C26033A	14	0.21
A26049G	12	0.02
A26094T	8	0.06
G26140A	8	0.06
T26207C	8	0.06
A26275G	8	0.44
T26319C	8	0.5

Although many *TRIM5α* alleles exist, they can be classified into 3 allelic classes (TFP, Q, CypA) based on variation in the C-terminal domain(42). These alleles demonstrate different restrictive

capabilities against different strains of SIV in macaques. The TFP/CypA genotype tends to be most restrictive, while the Q/Q genotype tends to be the least restrictive. Thus, animals with Q/Q tend to demonstrate poor virus control(25, 42). The TFP and Q allele types are differentiated by the presence of a 6 base pair insertion or deletion at positions 1015-1020 (amino acids 339-340). The TFP allele contains the 6 base pair insertion, and as a result, that region encodes threonine, phenylalanine, and proline. The Q allele, however, lacks this 6 base pair insertion and instead encodes a glutamine(25, 35). The CypA allele type is unique in that it is a splicing mutation that results in part of the TRIM5 C-terminal domain to be replaced by part of the *cyclophilin A* gene, resulting in a TRIM5-cyclophilin A chimeric protein. The mutation responsible for this alternative splicing is a G/T SNP in last position of intron 6. A guanine at this position results in wildtype *TRIM5α*, either a TFP or a Q allele. A threonine at this position, however, results in alternative splicing where exons 7 and 8 are replaced by a *cyclophilin A*-pseudogene that has been retrotransposed into the 3'UTR of the *TRIM5α* gene(36). This chimeric protein is still functional because cyclophilin A is also a capsid-binding protein(45).

The *TRIM5α* sequencing data that was used to detect SNPs was also used to categorize the *TRIM5α* of each monkey into the 3 allelic classes. First, the sequences were examined for the 6 base pair insertion/deletion to differentiate between TFP and Q alleles. The sequences were also examined for the G/T SNP at the end of intron 6 to determine the presence or absence of the CypA allele. The 6 base pair insertion/deletion and the G/T SNP were used to determine the *TRIM5α* genotype in the original cohort of monkeys (Table 17). Individuals that were heterozygous for cypA could not be fully typed because it is difficult to determine which allele contains the mutation using genomic DNA alone. Interestingly, the CypA mutation was found on both TFP and Q allelic backgrounds. The *TRIM5α* genotype, however, did not appear to

relate to disease outcome in this cohort of rhesus macaques. Other factors may influence this relationship, including the strains of virus used, the titer of inocula, and other host genetic factors.

Table 17: *TRIM5α* allelic classes

monkey	study	outcome	339-340	cypA
SIV01	PMED	protected	TFP/Q	wildtype/wildtype
SIV02	PMED	protected	-	-
SIV03	PMED	not protected	TFP/Q	wildtype/wildtype
SIV04	PMED	not protected	TFP/Q	wildtype/wildtype
SIV05	PMED	protected	TFP/Q	cypA/cypA
SIV06	PMED	not protected	-	wildtype/wildtype
SIV07	PMED	protected	TFP/Q	wildtype/wildtype
SIV08	PMED	infected	-	-
SIV09	PMED	infected	TFP/Q	wildtype/wildtype
SIV10	PMED	infected	Q/Q	wildtype/wildtype
SIV11	PMED	not infected	Q/Q	-
SIV12	MUCOSAL	not protected	TFP/Q	wildtype/wildtype
SIV13	MUCOSAL	protected	TFP/Q	cypA/wildtype
SIV14	PMED	infected	TFP/Q	-
SIV15	PMED	infected	TFP/Q	cypA/wildtype
SIV16	PMED	infected	TFP/Q	cypA/wildtype
SIV17	MUCOSAL	not protected	TFP/TFP	wildtype/wildtype
SIV18	MUCOSAL	not protected	TFP/Q	cypA/cypA
SIV19	MUCOSAL	not protected	TFP/Q	wildtype/wildtype
SIV20	MUCOSAL	protected	Q/Q	wildtype/wildtype
SIV21	MUCOSAL	not protected	TFP/Q	-
SIV22	MUCOSAL	not protected	-	wildtype/wildtype
SIV23	MUCOSAL	not protected	TFP/Q	wildtype/wildtype
SIV24	MUCOSAL	not protected	TFP/TFP	wildtype/wildtype
SIV25	MUCOSAL	protected	Q/Q	cypA/cypA

4.1.1.7 Amino Acid Changes

If a SNP falls within an exon of a gene, it could result in an amino acid change, potentially altering the function of the protein. Of the 81 SNPs detected in the 5 chemokine genes, 5 were found in exons. Two of these SNPs were synonymous, while the other 3 were predicted to result in an amino acid change. The 2 exonic SNPs in *CXCL13* were not predicted to change the amino acid at those positions. In both cases, a serine would remain a serine. For *CCL20*, the exonic SNP is predicted to change a leucine to a phenylalanine. Lastly, the two exonic SNPs in *CCL1* are predicted to change a tyrosine to histidine and a leucine to a phenylalanine (Table 18).

Table 18: Chemokine SNP summary

Gene	SNPs				
	Total	Promoter	Introns	Exons	Amino Acid
CXCL13	29	15	12	2	S→S, S→S
CCL20	18	8	9	1	L→F
CCL1	10	1	7	2	Y→H, L→F
CCL2	17	13	4	0	
CCL5	7	4	3	0	

4.1.2 CNV

Because copy number assays yield more reliable results with the use of genomic DNA, compared to whole genome amplified DNA, all CNV assays were performed using genomic DNA isolated from 19 PBMC samples. These 19 rhesus macaques were only recently included, thus their disease outcome is not yet available for analysis. First, custom TaqMan probes and quantitative real time PCR were used to determine copy number variation (CNV) in *CXCL13*

(Figure 7). Analysis of this PCR using CopyCaller software revealed that this innate immune gene does show copy number variation (Figure 8).

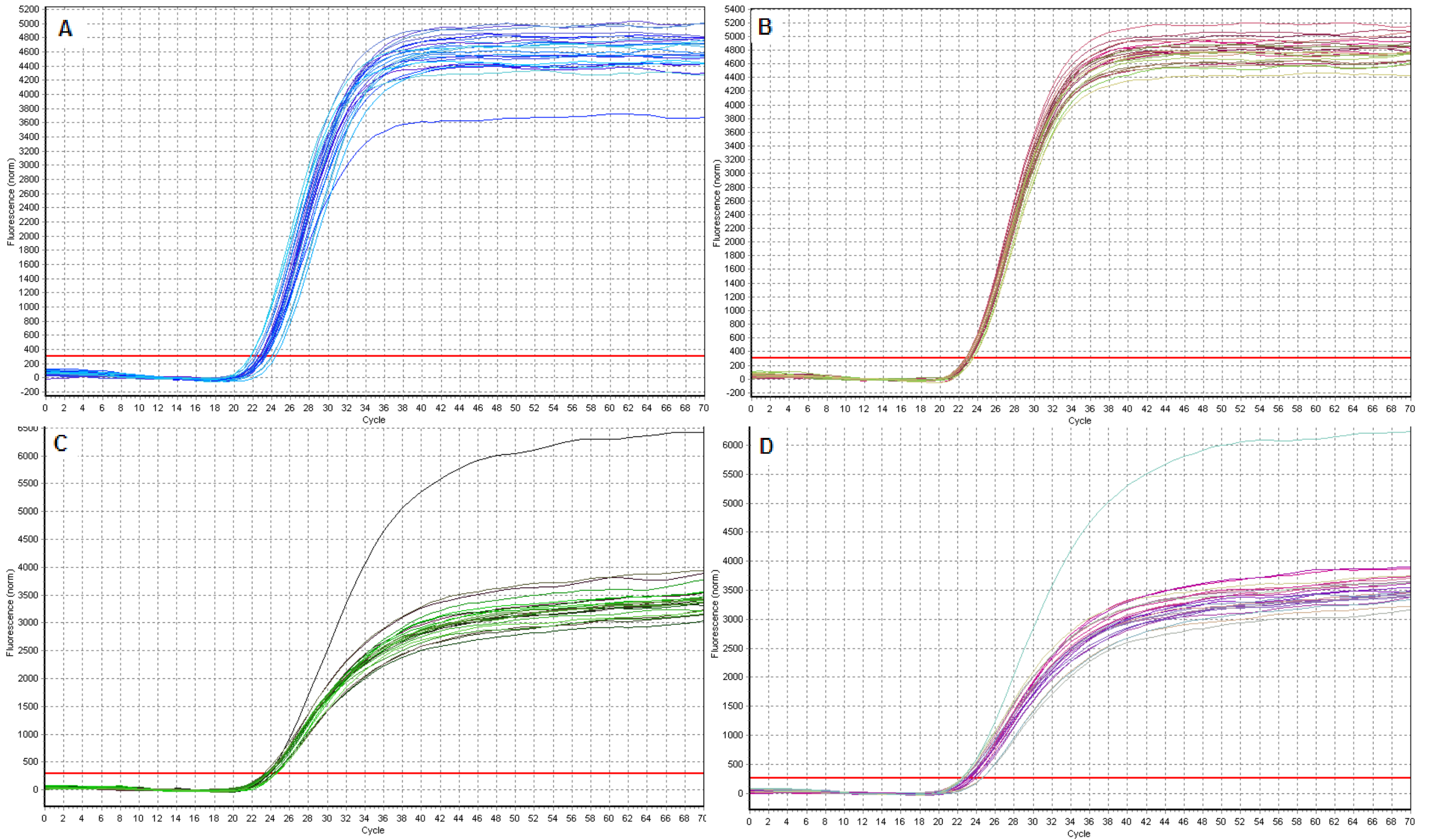

Figure 7: CNV Quantitative Real-Time PCR

PCR curves of 16 genomic DNA samples in triplicate using custom TaqMan Copy Number Probes for *CXCL13* (top panels) and a control gene, *STAT6* (bottom panels). A: *CXCL13*, monkeys 01-08; B: *CXCL13*, monkeys 09-16; C: *STAT6*, monkeys 01-08; D: *STAT6*, monkeys 09-16.

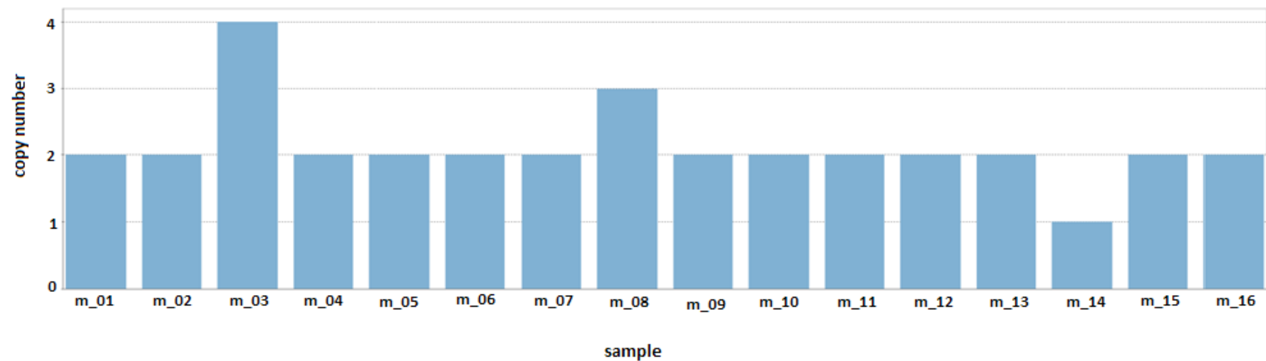

Figure 8: *CXCL13* Copy Number Variation

Copy Number Variation (CNV) was examined in 16 genomic DNA samples using custom TaqMan Copy Number probes for *CXCL13*. *STAT6* was used as a control gene (present at 2 copies per diploid genome).

Although the TaqMan assay seemed to work well, it did have its drawbacks including the lengthy amount of time for probes to be synthesized and the high cost of the reagents. Thus, we decided to use SYBR Green to analyze CNV in the remaining genes, as well as to replicate the results from the TaqMan assay for *CXCL13*. *CCL3L* was also included because it is known to show CNV in rhesus macaques, and could serve as a kind of positive control to demonstrate that the SYBR Green assay was working. Initially, the SYBR Green assay was performed using primers for *CXCL13*, *CCL1*, *CCL3L*, and *STAT6* (control gene, shown to be present at 2 copies per diploid genome). Although the PCR worked well and was consistent, analysis revealed no copy number variation in any of the genes, including *CXCL13* which was believed to show CNV based on the TaqMan assay, and *CCL3L* which is known to be highly variable among individuals.

4.2 AIM 2: DETERMINE THE RELATIONSHIP BETWEEN VARIATION IN INNATE IMMUNE GENES AND VACCINE RESPONSE

4.2.1 SNPs

Numerous SNPs were detected in all of the innate immune genes studied. None of the SNPs identified showed significant difference in frequency between the responder and non-responder groups, based on a t-test of proportions. It is important to note, however, that the small sample sizes of each group mean that these tests do not have adequate statistical power to detect anything other than extreme effect sizes. A larger study would yield more definitive results. However, a study such as the current one, that characterizes variation in these innate immune genes, is necessary to lay the groundwork for larger studies.

4.2.2 CNV

Copy number variation was examined in these innate immune genes by two different methods. Because the TaqMan assay and analysis software are specifically designed to detect CNV, it is likely the more reliable of the two methods. Thus, *CXCL13* most likely shows copy number variation. As the genomic DNA from the 19 newly recruited animals was used for the CNV assays, statistical analysis cannot be conducted because the disease outcome data are not yet available for these animals.

5.0 DISCUSSION

5.1 OVERVIEW OF CURRENT STUDY

Very little is known about genetic variation in rhesus macaques, making this study of variation in macaque chemokine genes, the first of its kind. Thus, the primary goal of this study was to characterize genetic variation in rhesus macaque innate immune response genes. The immune system is comprised of a vast number of proteins that could play a role in SIV vaccine responses and disease outcomes. To focus the study, a list of genes with clear antimicrobial and/or chemotactic activities or a role in inflammation was made. One by one, genes were selected from this list and examined for variation. We hypothesized that variation in rhesus macaque innate immune genes influences protective immunity against SIV challenge.

5.2 SNP DETECTION

Although numerous SNPs were detected in all of the innate immune genes studied, none of them correlated with disease outcome in the cohort of 25 rhesus macaques available for this study. A sample size of 25 animals is, however, quite small to detect the impact of genetic variation that does not have a profound effect on disease outcome. The experimental design of the current study, however, was appropriate for characterizing previously undiscovered genetic variation in

rhesus macaques. Thus, the SNPs detected in this study provide the basis for future studies that may examine their effects on protective immunity in greater detail.

SNPs in innate immune genes have previously been shown to be associated with response to vaccines. In a study of a DC based HIV-1 immuno-treatment, most people demonstrated an induction of HIV specific cellular immunity while a small portion of study participants did not(50). The researchers found that two SNPs are significantly associated with response to this DC-based vaccine. One SNP is located in the *NOS1* gene which is involved in the production of nitric oxide, a pro-inflammatory and antimicrobial molecule of the innate immune system. The second of the two SNPs is located in the *MBL2* (mannose-binding lectin) gene, involved in the activation of the complement system(50). Similarly, a SNP in another innate immune gene was found to be associated with response to the hepatitis B vaccine. Although HLA (human leukocyte antigen) plays a role, a SNP in the *FOXP1* gene, encoding a transcription factor that is involved in B-cell development, also influences vaccine response(13).

It is usually not the case that one SNP in one gene results in a profound effect on disease outcome or vaccine response. More often, disease outcomes are influenced by the cumulative effect of variation in several different genes(9, 37, 54). It is likely that the SNPs characterized in this study do not have a profound effect on SIV vaccine response or disease outcome, but any given SNP may play a minor role and be part of the larger picture. Even if the SNPs characterized in this study have no effect on the outcome of SIV challenge, they may play a role in the outcome of other diseases.

5.3 COPY NUMBER VARIATION

Copy number variation was detected in *CXCL13* using the TaqMan method (Figure 8). Because *CXCL13* is a gene of the immune system, this finding was not surprising. Segmental duplications that can lead to copy number variation do not occur randomly throughout the genome. Genes involved in immunity and defense, drug detoxification, and growth and development are particularly subject to these segmental duplications and thus often show copy number variation(5).

Although CNV was detected in *CXCL13*, the impact of copy number in this gene on disease outcome is not able to be determined at this time. Because a substantial volume of genomic DNA was needed for the CNV assays, DNA from 19 newly recruited animals was used. The disease outcome data are not yet available for these animals, putting statistical analysis on hold. Copy number variation in other innate immune genes has previously been shown to be associated with the outcome of HIV-1 or SIV exposure. In humans, more copies of *CCL3L*, relative to the geographical ancestry, is associated with lower susceptibility to HIV-1 infection and a slower progression to AIDS(18). Similarly, in rhesus macaques infected with SIV, a lower number of copies of *CCL3L* is associated with more rapid progression to disease(14). Thus, it is possible that CNV in *CXCL13* plays a role in vaccine response and/or disease outcome. Again, the role of this copy number on disease outcome may not be able to be revealed with such a small sample size. Additionally, CNV in this gene may only be part of the puzzle, and may play a minor role on its own.

Although CNV was detected in *CXCL13* using the TaqMan method, these results were not supported by the SYBR Green method. According to the results of the SYBR Green assay, none of the genes tested demonstrated copy number variation, including *CCL3L* which is known

to be highly variable in copy number among individuals(14, 29). These results may be a function of the SYBR Green technology itself. Compared to TaqMan assays, SYBR Green assays have lower specificity, and are not as sensitive. Additionally, TaqMan assays can be specifically designed for detection of copy number of a gene, and the results can be analyzed using software designed to interpret results of TaqMan assays and detect copy number. Thus, the results of the TaqMan assay are likely more reliable and if so, then *CXCL13* does show copy number variation.

5.4 FUTURE DIRECTIONS

It is important to recognize that the main focus of this study was to characterize genetic variation in a small cohort of rhesus macaques. Because little research has been dedicated to this topic in the past, this study was starting from square one. Thus, to begin such an endeavor, it seemed logical to begin by resequencing genes of the innate immune system to detect and curate polymorphisms in these genes. It is unlikely that the impact of such variation on protective immunity could be fully elucidated with such a small cohort of animals. That subject can be addressed in future studies, for which this study lays the groundwork. Specifically, this study has developed a list of novel candidate SNPs that can be applied to other studies. SNPs of interest could be rapidly screened in a large group of samples, from both previous and current studies, using a TaqMan SNP genotyping assay.

In addition to the innate immune genes examined in this study, it would beneficial to study other cellular restriction factors such as APOBEC3G and Tetherin. APOBEC3G (apolipoprotein beta mRNA-editing enzyme catabolic polypeptide 1-like protein 3G) is a cellular

deaminase that gets packaged in newly formed virions, and helps to block viral replication. After entry of the virion, during reverse transcription, APOBEC3G works by deaminating cytosines to uridines in the negative DNA strand. During positive strand synthesis, this deamination results in a guanine to adenine mutation. The accumulation of such mutations can lead to DNA breakage or result in the introduction of premature stop codons(1, 3, 26, 28, 41, 48, 52, 54, 58). The HIV-1 Vif protein, however, counteracts the effects of APOBEC3G by marking it for degradation via polyubiquitination(58). Tetherin is yet another cellular restriction factor that works by inhibiting the release of newly formed virions from the plasma membrane of the infected cell(3, 17). Similar to the scenario with APOBEC3G, HIV-1 encodes a protein, in this case Vpu, that counters the effects of tetherin(17).

5.5 PUBLIC HEALTH IMPLICATIONS

Rhesus macaques provide a valuable model that have been used in numerous studies and have greatly advanced our understanding of HIV and AIDS, which remains a major public health problem. Understanding genetic factors that influence macaque response to vaccine or SIV challenge will help us to better understand HIV in humans with the goal of developing a more effective vaccine and the ability to predict individual’s susceptibility to disease or response to vaccine.

APPENDIX A

PCR PRIMERS AND PROBES

Table 19: CNV Primers and Probes

Gene	Source	Primer	Sequence	Length/bp
Stat6	Degenhardt et al., 2009	Forward	CCAGATGCCTACCATGGTGC	129
		Reverse	CCATCTGCACAGACCACTCC	
		Probe	CTGATTCCTCCATGAGCATGCAGCTT	
CCL3L	Degenhardt et al., 2009	Forward	CCAGTGCTTAACCTTCCTCC	122
		Reverse	TCAGGCACTCAGCTCCAGGT	
		Probe	AGGCCGGCAGGTCTGTGCTGACC	
CXCL13	This study	Forward	GAGTCTGGAAGAAGAACAAGTCAGT	104
		Reverse	CCCCTTGTTATGCCTGACTTACTTT	
		Probe	ACCCTCAAGCTGAATGGATACAG	
CCL1	This study	Forward	TTCTAAACCCAGAGTTGCCAGC	100
		Reverse	GCACAGATGCGTCCTTCCTTAC	

Table 20: PCR primers

Gene	Amplicon	Forward	Reverse	Length/bp
CCL1	Exon 1	GGTGAAGGCACATCTTCCTCAAG	TTGGCACCTTGGAGTTTCAGG	608
CCL1	Exon 2	AGGGGTCCAGGAAGCTCAATTC	CAAAGGCAGTCACTGAGAGGTATTG	659
CCL1	Promoter	TCAGTCCATCCTATGCCATCTGTC	TGTCTCATCAGCCACCTCGTTC	4,228
CCL1	Prom seqF	TCCCAGCTTGCTCAGTTACA	-	-
CCL1	Prom seqF2	AGAAGTCTGTGCCCTTGG	-	-
CCL1	Prom seqF3	AGAGAAGTCTGTGCCCTTGG	-	-
CCL1	Prom seqR	-	TAAAGGCTCATCCGGTTTTG	-
CCL1	Prom seqR2	-	GGTAAAGAGTCAGAAGATTGGG	-
CCL2	Exon 1	TGGAAAAAGTGTCTTGTCCTGACC	AGAGTAACTGTGCTGAGTGTGTCCC	589
CCL2	Exon 2	GCTGGTCATCTGGATTATTGGTCC	TGGAGCCATCCTGACTTCAAAC	645
CCL2	Exon 3	CACCTGGGTGCCTATTCAGAAC	GAACTGTGGTTCAAGAGGAAAAGC	544
CCL2	Promoter 1	CCTTTAGCTGTCTGCCCATTAGG	TGCTGAGCAAATGCAGCGTC	818
CCL2	Promoter 2	ACTCAGCTCAGCCATTTACATGC	GCTGGTGTCATACTCACAGAGTGC	662
CCL2	Promoter 3	TGAGCAGAGGACTGAGCCAAAC	AGCAGAGGAGGAGTATTTCGGG	867
CCL5	Exon 1	CAAAGGGGAAACTGATGAGC	CTTGGGGGTGTAGACCTTGA	594
CCL5	Exon 2	GTTATGGAGCCTGAGCCTTG	TGGGCAGGATTCTCTAAGGA	681
CCL5	Exon 3	ATTTGTTTCTGGCTTGGAGC	GTGGTAGAATCTGGGCCCTT	419
CCL5	Promoter 1	CCTATGGCCAGGATGAAAGA	CCTTCATGGTACCTGTGGGA	680
CCL5	Promoter 2	GGCCTTCAGTGAGATGGGAT	CTTTCCCTCATCCACGGTTA	491
CCL5	Promoter 3	AGCCCAGATCAAATTGTTGC	CTCTTACAGCATACTCTACC	550
CCL20	Exon 1	TGTGTGGGGCTGACCTTTGTATC	GCACTTTGGCTTTGTCCTGAAG	760
CCL20	Exon 2	TCTCAGCACACAAACTGTTGTTCC	GGAAGTTTTTGCCTTATTCCAGTCC	558
CCL20	Exon 3	TTTGTTACATTGCAGAGAAAGGGAG	CATTCTTTGTGTTTGTGGTCCAAC	438
CCL20	Exon 3 seq	ACCTCGTCAACTGAAGTGTTTAGT	-	-
CCL20	Exon 3a	TTTGTTACATTGCAGAGAAAGGGAG	CAACCAAATGCTGGATTGCC	604
CCL20	Exon 4	CACTTTATCAACGGATGGAAGGC	TTGGACAAGTCCCAGTGAGGC	440
CCL20	Promoter 1	TGATATAGGCATCACCAACTCCG	TGGGAAGATGCAAAGGTCAGG	799
CCL20	Promoter 2	GGCAAAAATTATCATGGTGCATGAC	GGCTCAAACCTCATCTTCATCTTG	733
CCL20	Promoter 2a	GGCAACAAGAGCAAAATTCTGTCTC	GGCTCAAACCTCATCTTCATCTTG	607
CXCL13	3'STR	FAM-AGCATTCAAAGATCCCCAGA	GTGATTGTGCCACTGCACTC	400
CXCL13	Exon 1	TGATGCAGCTGACCTACTGG	TGGAATCTGAAAAAGGCTGG	479
CXCL13	Exon 1a	TGACCTACTGGAGACAAAGGCAG	TGAAAAAGGCTGGGAATATACAAGC	462
CXCL13	Exon 2	AGCCAAGGCAAGCAGAAATA	GCCTCAGGCTGTAAGGAACA	554
CXCL13	Exon 3 & 4	TGACCCCATGCTACAAATCA	GCACAAAACTAAAACCCGGA	655
CXCL13	Promoter 1	GCAGTTCAGTGGGTCTGGTT	GCTATTGGTGATGGTGGCTT	696
CXCL13	Promoter 2	CCCTTCCACATGCTAGGGAT	CAGCAGCATGAGAAGCAGAG	503
CXCL13	Promoter 3	GGATGGATAGATGGGGGAGT	ATCCCAGGTGCCTCAGTATG	802
CXCL13	Promoter 3a	CTTCCTCCTTCTGTCTATTGCAGG	TGGGCTTCCTGGAGTGATTTCC	491

APPENDIX B

SNP DISTRIBUTIONS

Table 21: *CXCL13* SNP Distribution

SNP	A-2012C	T-1921C	G-1797A	C-1758T	G-1671A	G-1607A	C-1521T	A-1421T	G-1308G	G-1251A	G-1208T	G-1176A	G-1132C	G-1060T	C-993T
Reference	**A**														
SIV01	C	Y	G	T				A	G	R	K	R	C	K	C
SIV02	C	C						A	R	R	K	A	C	K	Y
SIV03	C	T	A	C	G	G	C	T	A	A	T	G	C	G	C
SIV04	C	Y	G	Y	G	R	Y	A	R	R	K	A	C	K	C
SIV05	C	C	R	Y	G	A	Y	A	R	R	K	A	C	K	C
SIV06	C	C			G			A	G	G	G	A	C	T	C
SIV07	C	T	A	C	G	G	C	T	R	R	G	A	C	K	C
SIV08	C	C	G	T				W	G	R	K	R	S	K	C
SIV09	C	C	G	T	G	A	T	T	A	A	T	G	C	G	C
SIV10	C	T	A	C	G	G	C	T	A	A	T	G	C	G	C
SIV11	C	Y	G	T	G	A	T	A	G	R	K	R	S	K	C
SIV12	C	C	G	T	A	A		A	G	G	G	A	C	T	C
SIV13	C	T						T	A	A	T	G	C	G	C
SIV14	C	T	A	C	G	G	C	T	A	R	K	R	C	K	C
SIV15	C	T	A	C	G	G	C	T	G	G	G	A	C	K	C
SIV16	C	T	A	C	G	G	C	T	A	A	T	G	C	G	C
SIV17	C	C	G	T	G	A	T	T	A	A	T	G	C	G	C
SIV18	C	Y	R	Y	G	R	Y	T	A	A	T	G	C	G	C
SIV19	C	C	G	T	G	A	T	T	A	A	T	G	C	G	C
SIV20	C	Y	R	Y	G	R	Y	A	G	G	G	A	C	T	C
SIV21	C	Y	R	Y	G	R	Y	W	R	R	K	R	C	K	C
SIV22	C	C	G	T		A	T	A	G	G	G	A	C	T	C
SIV23								W	R	R	K	R	C	K	C
SIV24								W	R	R	K	R	C	T	C
SIV25								A	G	G	G	A	C		

C-956T	ins-932G	G-684A	G-476C	G-432A	G15A	C200T	C302T	C1780A	G1949A	C2213T	G6012A	C6232T	C6330T
		G	**C**	**G**	**A**	**T**	**C**	**C**	**A**	**T**	**G**	**C**	**T**
T	T	R	S	R	·	Y	·	M	R		·	·	·
Y	T	·	S	R	R	·		·	G	Y	·	·	·
T		A	G	A	G			A	·		·	·	·
T	T	R	S	R	·	·	·	·	N		·	·	·
T	T	R	S	R	·	·	·						
Y	T	·	·	·	·	·	·						
T	T	R	S	R	·	·	·						
T	T	R	S	R	·	·	·	M	R	Y			
T		A	G	A	G	C	T	·	G	C			
T		A	G	A	G	C	T	·	G	Y			
T	T	R	S	A	·	Y	Y	·	R	Y			
T	T	·	·	·	·	·	·	A	·	·	R	S	C
T	C	A	G	A	G	C	T	A	·	·	·	·	·
Y	Y	R	S	R	·	Y	·	A	R		·	·	·
Y	T	·	S	R	·	Y	Y	·	G	C			·
T		A	G	A	G	C	T	A	·				
T	T	A	G	A	G	C	T		R				
T	T	A	G	A	G	C	T		N	Y			·
T	C	A	G	A	G	C	T		·	·			
Y	T	·			·	·	·		·				
Y	T	R	S	R	·	·	Y	A	·	·			
C	T	·	·	·	·	·	·	A	·				
Y	T	A	S	R	R	Y	Y	·	G	C			
T		A	S	R	R	Y	Y	M	R		·	S	·
	T	·	·	·	·	·	·	·	G				

Table 22: *CCL20* SNP Distribution

SNP Reference	G-1181C G	G-1123A G	G-1034C G	A-930C A	C-639T C	T-595C T	T-448C T	A-398G A	G-326T G	A-251G A	G-193T :	G-179T G	ins-46G :	G36T G	A191G A	A1491G A	A2333G A	A2489G A
SIV01	·	A	·	·	·	·	·	·	·	·	G	·	G	T	·	·		
SIV02	·		·	·	·	·	·	·	·	·	G	·	G	T	·	·		
SIV03	S	·	·	·	·	C	·	G	·	·	G	·	G	·	·	R	·	
SIV04	·	·	·	·	·	Y	·	R	·	·	K	·	G	K	·	R	·	·
SIV05	C	A	·	·	Y	Y	·	R	·	·	K	·	G	·	R	·	·	R
SIV06	C	A	·	·	·	·	·	·	·	·	G	·	G	·	·	·	·	·
SIV07	·	·	·	·	·	Y	·	R	·	·	G	·	G	·	R	R		
SIV08			·	·			·		·	·		·	G					
SIV09			·	·	·	Y	·	R	·	·	·	·	G	·	R	R	·	·
SIV10	·	·	·	·	·	Y	·	R	·	·	K	·	G			R	·	·
SIV11	·	·	·	·	·	Y	·	·	K	·	K	·	G	·	R	R	·	·
SIV12	·	R	·	·	·	·	·	·	K	·	G	·	G	·	·		·	·
SIV13	S	·	·	·	·	Y	·	R	·	·	G	·	G	·	·	·		
SIV14	S	R	·	M	Y	C	·	G	·	·	·	·	G	·	G	R	·	·
SIV15	·	·	·	·	·	Y	·	R	·	·	K	·	G	·	R	R		
SIV16	·	A	·	·	·	·	·	·	·	·	G	·	G	·	·	·	·	·
SIV17	C	A	·	·	·	·	·	·	·	·	G	·	G	·		·	·	·
SIV18	·	·	·	·	·	C	·	G	·	·	·	·	G	·	G	G	·	
SIV19	·	·	·	·	·	Y	·	R	·	·	K	·	G	·	R	R	·	·
SIV20	C	·	·	·	·	Y	·	R	·	·	·	·	G	·	·	·	R	R
SIV21	·	·	C	·	·	·	Y	·	·	R	·	K	G	·	·	·	·	·
SIV22	C	R	·	·	Y	C	·	G	·	·	K	·	·	·	R	·	R	
SIV23	·	·	·	·	·	Y	·	R	·	·	G	·	·	·	R	R	·	·
SIV24	·	·	·	·	·	Y	·	R	·	·	K	·	·	·	R	R	·	·
SIV25	C	·	·	·	·	Y	·	R	·	·	G	·	·	·	·		R	R

Table 23: *CCL1* SNP Distribution

SNP		C-45T	T98C	T131G	C257T	C366G	T1238C	G1280A	C1320T	A1501G	C1576T
Reference		**C**	**T**	**T**	**C**	**C**	**T**	**G**	**C**	**A**	**C**
	SIV01	.	C	.	.	.	.	A	.	G	T
	SIV02										
	SIV03	.	C	G	.	.	N	.	.	G	T
	SIV04	.	C	.	.	.	.	R	.	G	Y
	SIV05	.	C	K	.	.	.	R	.	G	T
	SIV06	.	Y		.	.	.	.	.	R	Y
	SIV07	.	C	.	.	.	.	.	.	G	Y
	SIV08										
	SIV09	.	C	.	.			.	.	G	Y
	SIV10	.	C	.	Y	.	.	R	Y	G	T
	SIV11	.	C	.	.	.	Y	.	.	G	Y
	SIV12	.	C	G	.	.	Y	.	.	G	T
	SIV13	.	C	.	.	.	.	R	.	G	Y
	SIV14	.	.	.	.	.	.	.	.	.	.
	SIV15	.	C	.	.	.	Y	.	.	G	Y
	SIV16	.	Y	.	.	.	.	R	.	R	Y
	SIV17	.	C	K	.	.	.	.	.	G	T
	SIV18	.	C	.	.	.	.	R	.	G	Y
	SIV19	.	C	.	.	.	.	R	.	G	Y
	SIV20	.	Y	K	.	.	.	.	.	R	Y
	SIV21	T	C	.	.	T		.	.	G	T
	SIV22	.	C	.	Y	.	.	R	Y	G	T
	SIV23	.	C	.	Y	.	.	.	Y	G	T
	SIV24	.	C	.	.	.	Y	.	.	G	Y
	SIV25	.	C	.	.	.	.	.	.	R	.

Table 24: *CCL2* SNP Distribution

SNP	A-1925G	T-1581C	C-1558T	T-1530C	C-1032T	C-878T	G-874T	C-850T	G-578A	G-495C	G-430A	T-79C	G-33T	C711A	C971T	C1459T	A1550C
Reference	**A**	**T**	**C**	**T**	**C**	**C**	**G**	**C**	**G**	**G**	**G**	**T**	**G**	**C**	**C**	**C**	**A**
SIV01	G	C	T	.	.	.	T	.	A	C	.	C	T	A	.	T	C
SIV02		C	T	.	.	.	K	.	A	C	.	C	T	.	T	.	.
SIV03	.	Y	Y	.	.	.	K	.	R	S	.	Y	K	M	.	Y	M
SIV04	G	Y	Y	.	.	.	K	.	R	S	.	Y		M	.	Y	M
SIV05	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SIV06	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SIV07	G	C	T	.	.	.	T	.	A	C	.	C	T	.	T	.	.
SIV08																	
SIV09	G	C	T	.	.	.	T	.	A	C	.	C	T	A	.	T	C
SIV10	G	Y	Y	.	.	.	K	.	R	S	.			M	.	Y	M
SIV11	G	Y	Y	.	.	.	K	.	R	S	.	Y	K	M	.	Y	M
SIV12	G	Y	Y	.	.	.	.	.	R	S	R	Y	K	.	Y	.	.
SIV13	G	Y	Y	Y	Y	.	K	Y	R	S	.	Y		M	.	Y	M
SIV14	R	Y	Y	.	.	.	K	.	R	S	.	Y		M	.	Y	M
SIV15	G	C	T	.	.	.	T	.	A	C	.	C	T	M	Y	Y	M
SIV16	G	Y	Y	.	.	.	K	.	R	S	.	Y	K	.	Y	.	.
SIV17	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SIV18	G	C	T	Y	Y	.	T	Y	A	C	.	C	T	M	Y	Y	M
SIV19	G	Y	.	.	.	Y	.	.	R	S	.	Y	K	.	.	Y	
SIV20	G	Y	Y	.	.	.	K	.	R	S	.	Y	K	.	Y	.	.
SIV21	G	C	T	.	.	.	T	.	A	C	R	C	T	M	Y	Y	M
SIV22	G	Y	Y	.	.	.	K	.	R	S	R	Y	K	.	Y	.	.
SIV23	G	C	T	.	.	.	T	.	A	C	R	C	T	.	T	.	.
SIV24	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SIV25	G	C	T	.	.	.	T	.	A	C	R	C	T	M	Y	Y	M

Table 25: *CCL5* SNP Distribution

SNP		C-1142T	G-1141A	A-339G	A-135G	C113T	G8000T	T8309A
Reference		**C**	**G**	**G**	**G**	**C**	**G**	**T**
	SIV01	.	.	.	.	.	.	.
	SIV02	.	.	.	.	.		
	SIV03	.	R	.	A	.	T	.
	SIV04	Y	.	.	.	.		.
	SIV05	Y	.	.	.	.	.	.
	SIV06	.	R	.	R	Y	T	.
	SIV07	.	R	.	.	Y	T	.
	SIV08							
	SIV09	.	R	.	R	Y	K	.
	SIV10	.	R	.	R	Y	K	.
	SIV11	Y	R	.	.	Y	K	.
	SIV12	T	.	.	.	.	.	.
	SIV13	.	R	.	.	Y	K	.
	SIV14	.	R		R	Y	T	.
	SIV15	Y	.	R	R	.	K	A
	SIV16	Y	.	.	.	.	.	.
	SIV17			.	.		K	.
	SIV18	.	R	.	.	Y	K	.
	SIV19	Y	.	.	.	.	.	.
	SIV20	Y	.	.	R	.		.
	SIV21	Y	.	R	R	.	K	.
	SIV22	Y	R	.	.	Y	K	.
	SIV23	.	R	.	.	Y	T	.
	SIV24	Y	R		.	Y	T	.
	SIV25	.	A	.	.	T	T	.

Table 26: *TRIM5α* SNP Distribution

SNP		G-11T	G42C	A9340C	C9433T	A9491T	G9834A	T10162C	T10479C	G10679T	C10686T	A10687G	C11215T	C11219T	C11233T	G11252T	T11324C
Reference		**T**	**C**	**A**	**C**	**A**	**A**	**C**	**C**	**G**	**C**	**A**	**C**	**T**	**C**	**G**	**T**
	SIV01	K	G	·	·	T	G	T	·	·	·	·	·	C	·	·	·
	SIV02	·	G														
	SIV03	·	S		·	A	·	·	·	T	·	·	·	C	T	T	·
	SIV04	·	S		·	N	·	Y	·	K	·	·	·	C	Y	K	·
	SIV05	·	S			N	R	·	Y	K	Y	R	Y	Y	Y	K	
	SIV06	·	G		·	A	·	·	Y	T	Y	R	Y	C	T	T	
	SIV07	·	·		·	A	·	T	·	·	·	·	·	C	·	·	·
	SIV08																
	SIV09	·	·	·	·	N	·	Y	·	K	·	·	·	C	Y	K	·
	SIV10	·	S	·	·	·	·	T	·	·	·	·	·	C	·	·	·
	SIV11	·	S	·	Y	N	·	·	T	T	T	G	T	C	T	T	Y
	SIV12	·	G		·	N	R	Y	·	·	·	·	·	Y	·	·	·
	SIV13	·	S		Y	N	·	·	Y	T	Y	R	Y	C	T	T	
	SIV14	·	S		Y	A	·	·	Y	T	Y	R	Y	C	T	T	
	SIV15	·	S		Y	·	·	Y	Y	K	Y	R	Y	C	Y	K	
	SIV16	·	·	·	·	·	·	T	·	·	·	·	·	C	·	·	·
	SIV17	·	S		·	A	A	·	·	·	·	·	·	·	·	·	·
	SIV18	·	G	C		N	R	·	Y	K	Y	R	Y	Y	Y	·	
	SIV19	·	·	·	·	N	R	Y	·	·	·	·	·	Y	·	·	·
	SIV20	·	S		·	N	·	Y	Y	K	Y	R	Y	C	Y	K	·
	SIV21	·	·		Y	·	·	Y	Y	K	Y	R	Y	C	Y	K	
	SIV22	·	G	C	·	A	·	·	Y	T	Y	R	Y	C	T	T	
	SIV23	·	G			A	·	·	Y	T	Y	R	Y	C	T	T	
	SIV24	·	S		·	A	A	·	·	·	·	·	·	·	·	·	·
	SIV25	·	·		T	·	·	·	T	T	T	G	T	C	T	T	C

Table 26 continued

C11338A	C11343T	C11350T	T11378C	C11437T	G11443A	G11456A	A11480G	A11500T	C23693T	G23737A	G25393A	A25478G	C25527T	T25651C	G25667T
C	**C**	**C**	**T**	**C**	**G**	**G**	**A**	**A**	**C**	**A**	**A**	**G**	**T**	**T**	**G**
.	.	Y	.	.	.	.	.	.			G	A	C	.	.
									.	G					
.	.	T	.	.	.	.	.	T	.	.	R	R	.	.	.
.	.	Y	.	.	.	.	.	W	.	G	R	R	Y	.	.
.							R	W	.	G	.	.	.	C	T
							R	T	.	G	R	R	.	.	.
.	.	.	.	.	.	.	.	.	.	.	R	R	Y	.	.
.	.	Y	.	.	.	.	.	W	.	G	R	R	Y	.	.
.	.	.	.	.	.	.	.	.	.	.	G	A	C	.	.
M	T	T	C	T	A	A	G	T	.	G					
.	.	Y	.	.	.	.	.	.	.		G	A	C	.	.
							R	T	.	G		.	.	Y	K
							R	T	Y	G					
									.	G				Y	K
.	.	.	.	.	.	.	.	.	.	G				Y	K
.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.
							R	W	.	G	.	.	.	C	T
.	.	Y	.	.	.	.	.	.	.	G	G	A	C	.	.
.	Y	Y	Y	Y	R	R	R	W	.	.	G	A	Y	.	.
							R	W	.	G					
							R	T	.	G	.	R	.	.	.
							R	T	.	R	R	R	.	.	.
.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.
A	T	T	C	T	A	A	G	T	.	G	.	.	.	C	T

Table 26 continued

C25680T	A25689G	C25722T	A25770G	A25823C	C25916T	C25973A	G25987A	C25992T	C26033A	A26049G	A26094T	G26140A	T26207C	A26275G	T26319C
C	**A**	**C**	**A**	**A**	**C**	**C**	**G**	**C**	**C**	**A**	**A**	**G**	**T**	**G**	**C**
.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	Y
.	.	.	.	.	.	.	.	.	M	.					
.	.	.	.	.	.	.	.	.	.	.					
T	G	Y	C	M	Y	M	R	Y							
.	.	.	.	.	.	.	.	.	M	.					
.	.	.	.	.	.	.	.	.	.	.					
.	.	.	.	.	.	.	.	.	.	.					
.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T
			R	M	Y	.	R	Y	M	R	W	R	Y	A	Y
.	.	.	.	.	.	M	.	.	.	.	.	.	.	R	Y
Y	R	Y	R	M	Y	.	R	Y							
		Y	R	M	Y	.	R	Y							
Y	R	Y	R	M	Y	.	R	.							
Y	R	Y	R	M	Y	.	R	Y							
.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
T	G	Y	G	C	Y	M	R	Y							
.	.	.	.	.	.	M	.	.			.	.	.	R	Y
.	.	.	.	.	.	.	.	.	M	.	.	.	.	A	T
		Y	R	M	Y	.	R	Y							
.	.	.	.	.	.	.	.	.	M						
.	.	.	.	.	.	.	.	.	M						
.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
T	G	T	G	C	T	.	A	T	.	G	T	A	C	.	.

BIBLIOGRAPHY

1. **An, P., and C. A. Winkler**. 2010. Host genes associated with HIV/AIDS: advances in gene discovery. Trends in genetics: TIG **26**:119-31.

2. **Ansel, K. M., R. B. S. Harris, and J. G. Cyster**. 2002. CXCL13 is required for B1 cell homing, natural antibody production, and body cavity immunity. Immunity **16**:67-76.

3. **Arhel, N., and F. Kirchhoff**. 2010. Host proteins involved in HIV infection: new therapeutic targets. Biochimica et biophysica acta. Elsevier B.V. **1802**:313-21.

4. **Baba, M., T. Imai, M. Nishimura, M. Kakizaki, S. Takagi, K. Hieshima, H. Nomiyama, and O. Yoshie**. 1997. Identification of CCR6, the specific receptor for a novel lymphocyte-directed CC chemokine LARC. The Journal of biological chemistry **272**:14893-8.

5. **Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li, and E. E. Eichler**. 2002. Recent segmental duplications in the human genome. Science (New York, N.Y.) **297**:1003-7.

6. **Bosinger, S. E., D. L. Sodora, and G. Silvestri**. 2011. Generalized immune activation and innate immune responses in simian immunodeficiency virus infection. Current opinion in HIV and AIDS **6**:411-8.

7. **Brenchley, J. M., and M. Paiardini**. 2011. Immunodeficiency lentiviral infections in natural and nonnatural hosts. Blood 847-854.

8. **Carr, M. W., S. J. Roth, E. Luther, S. S. Rose, and T. A. Springer**. 1994. Monocyte chemoattractant protein 1 acts as a T-lymphocyte chemoattractant. Proceedings of the National Academy of Sciences of the United States of America **91**:3652-6.

9. **Chatterjee, K.** 2010. Host genetic factors in susceptibility to HIV-1 infection and progression to AIDS. Journal of genetics **89**:109-16.

10. **Cyster, J. G.** 1999. Chemokines and the homing of dendritic cells to the T cell areas of lymphoid organs. The Journal of experimental medicine **189**:447-50.

11. **Cyster, J. G.** 1999. Chemokines and cell migration in secondary lymphoid organs. Science (New York, N.Y.) **286**:2098-102.

12. **Dahinden, C. A., T. Geiser, T. Brunner, V. von Tscharner, D. Caput, P. Ferrara, A. Minty, and M. Baggiolini**. 1994. Monocyte chemotactic protein 3 is a most effective basophil- and eosinophil-activating chemokine. The Journal of experimental medicine **179**:751-6.

13. **Davila, S., F. E. M. Froeling, A. Tan, C. Bonnard, G. J. Boland, H. Snippe, M. L. Hibberd, and M. Seielstad**. 2010. New genetic associations detected in a host response study to hepatitis B vaccine. Genes and immunity. Nature Publishing Group **11**:232-8.

14. **Degenhardt, J. D., P. de Candia, A. Chabot, S. Schwartz, L. Henderson, B. Ling, M. Hunter, Z. Jiang, R. E. Palermo, M. Katze, E. E. Eichler, M. Ventura, J. Rogers, P. Marx, Y. Gilad, and C. D. Bustamante**. 2009. Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus Macaques (Macaca mulatta). PLoS genetics **5**:e1000346.

15. **Freeman, J. L., G. H. Perry, L. Feuk, R. Redon, S. A. McCarroll, D. M. Altshuler, H. Aburatani, K. W. Jones, C. Tyler-Smith, M. E. Hurles, N. P. Carter, S. W. Scherer, and C. Lee**. 2006. Copy number variation: new insights in genome diversity. Genome research **16**:949-61.

16. **Fuller, D. H., P. A. Rajakumar, L. A. Wilson, A. M. Trichel, J. T. Fuller, T. Shipley, M. S. Wu, K. Weis, C. R. Rinaldo, J. R. Haynes, M. Murphey-corb, and J. V. Irol**. 2002. Induction of Mucosal Protection against Primary , Heterologous Simian Immunodeficiency Virus by a DNA Vaccine. Society **76**:3309-3317.

17. **Girard, M. P., S. Osmanov, O. M. Assossou, and M. P. Kieny**. 2011. Human immunodeficiency virus (HIV) immunopathogenesis and vaccine development: A review. Vaccine. Elsevier Ltd **29**:6191-6218.

18. **Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R. J. Nibbs, B. I. Freedman, M. P. Quinones, M. J. Bamshad, K. K. Murthy, B. H. Rovin, W. Bradley, R. A. Clark, S. A. Anderson, R. J. O'connell, B. K. Agan, S. S. Ahuja, R. Bologna, L. Sen, M. J. Dolan, and S. K. Ahuja**. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science (New York, N.Y.) **307**:1434-40.

19. **Gu, L., S. C. Tseng, and B. J. Rollins**. 1999. Monocyte chemoattractant protein-1. Chemical immunology **72**:7-29.

20. **Hieshima, K., T. Imai, G. Opdenakker, J. Van Damme, J. Kusuda, H. Tei, Y. Sakaki, K. Takatsuki, R. Miura, O. Yoshie, and H. Nomiyama**. 1997. Molecular cloning of a novel human CC chemokine liver and activation-regulated chemokine (LARC) expressed in liver. Chemotactic activity for lymphocytes and gene localization on chromosome 2. The Journal of biological chemistry **272**:5846-53.

21. **Hirsch, V. M., R. A. Olmsted, M. Murphey-Corb, R. H. Purcell, and P. R. Johnson**. 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature **339**:389-92.

22. **Hu, S. L.** 2005. Non-human primate models for AIDS vaccine research. Current drug targets. Infectious disorders **5**:193-201.

23. **Ignal, A. V., and D. M. Ilan**. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Review Literature And Arts Of The Americas **34**:275-305.

24. **Johnston, M. I.** 2000. The role of nonhuman primate models in AIDS vaccine development. Molecular medicine today **6**:267-70.

25. **Kirmaier, A., F. Wu, R. M. Newman, L. R. Hall, J. S. Morgan, S. O'Connor, P. A. Marx, M. Meythaler, S. Goldstein, A. Buckler-White, A. Kaur, V. M. Hirsch, and W. E. Johnson**. 2010. TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. PLoS biology **8**.

26. **Komano, J., Y. Futahashi, E. Urano, K. Miyauchi, T. Murakami, Z. Matsuda, and N. Yamamoto**. 2005. The interaction of HIV-1 with the host factors. Japanese journal of infectious diseases **58**:125-30.

27. **Kurihara, T., G. Warr, J. Loy, and R. Bravo**. 1997. Defects in macrophage recruitment and host defense in mice lacking the CCR2 chemokine receptor. The Journal of experimental medicine **186**:1757-62.

28. **Lackner, A. A., and R. S. Veazey**. 2007. Current concepts in AIDS pathogenesis: insights from the SIV/macaque model. Annual review of medicine **58**:461-76.

29. **Lee, A. S., M. Gutiérrez-Arcelus, G. H. Perry, E. J. Vallender, W. E. Johnson, G. M. Miller, J. O. Korbel, and C. Lee**. 2008. Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. Human molecular genetics **17**:1127-36.

30. **Legler, D. F., M. Loetscher, R. S. Roos, I. Clark-Lewis, M. Baggiolini, and B. Moser**. 1998. B cell-attracting chemokine 1, a human CXC chemokine expressed in lymphoid tissues, selectively attracts B lymphocytes via BLR1/CXCR5. The Journal of experimental medicine **187**:655-60.

31. **Miller, M. D., and M. S. Krangel**. 1992. The human cytokine I-309 is a monocyte chemoattractant. Proceedings of the National Academy of Sciences of the United States of America **89**:2950-4.

32. **Mir, K. D., M. A. Gasper, V. Sundaravaradan, and D. L. Sodora**. 2011. SIV infection in natural hosts: resolution of immune activation during the acute-to-chronic transition phase. Microbes and infection / Institut Pasteur **13**:14-24.

33. **Munier, C. M. L., C. R. Andersen, and A. D. Kelleher**. 2011. HIV vaccines: progress to date. Drugs **71**:387-414.

34. **Murphey-Corb, M., L. A. Wilson, A. M. Trichel, D. E. Roberts, K. Xu, S. Ohkawa, B. Woodson, R. Bohm, and J. Blanchard**. 1999. Selective induction of protective MHC class I-restricted CTL in the intestinal lamina propria of rhesus monkeys by transient SIV infection of the colonic mucosa. Journal of immunology (Baltimore, Md.: 1950) **162**:540-9.

35. **Newman, R. M., L. Hall, M. Connole, G. L. Chen, S. Sato, E. Yuste, W. Diehl, E. Hunter, A. Kaur, G. M. Miller, and W. E. Johnson**. 2006. Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. Proceedings of the National Academy of Sciences of the United States of America **103**:19134-9.

36. **Newman, R. M., L. Hall, A. Kirmaier, L. A. Pozzi, E. Pery, M. Farzan, S. P. O'Neil, and W. Johnson**. 2008. Evolution of a TRIM5-CypA splice isoform in old world monkeys. PLoS pathogens **4**:e1000003.

37. **O'Brien, S. J., and G. W. Nelson**. 2004. Human genes that limit AIDS. Nature genetics **36**:565-74.

38. **Pegu, A., J. L. Flynn, and T. A. Reinhart**. 2007. Afferent and efferent interfaces of lymph nodes are distinguished by expression of lymphatic endothelial markers and chemokines. Lymphatic research and biology **5**:91-103.

39. **Reinhart, T. A., S. Qin, and Y. Sui**. 2009. Multiple roles for chemokines in the pathogenesis of SIV infection. Current HIV research **7**:73-82.

40. **Rerks-Ngarm, S., P. Pitisuttithum, S. Nitayaphan, J. Kaewkungwal, J. Chiu, R. Paris, N. Premsri, C. Namwat, M. de Souza, E. Adams, M. Benenson, S. Gurunathan, J. Tartaglia, J. G. McNeil, D. P. Francis, D. Stablein, D. L. Birx, S. Chunsuttiwat, C. Khamboonruang, P. Thongcharoen, M. L. Robb, N. L. Michael, P. Kunasol, and J. H. Kim**. 2009. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. The New England journal of medicine **361**:2209-20.

41. **Restrepo, C., N. I. Rallón, J. Carrillo, V. Soriano, J. Blanco, and J. M. Benito**. 2011. Host factors involved in low susceptibility to HIV infection. AIDS reviews **13**:30-40.

42. **Reynolds, M. R., J. B. Sacha, A. M. Weiler, G. J. Borchardt, C. E. Glidden, N. C. Sheppard, F. A. Norante, P. A. Castrovinci, J. J. Harris, H. T. Robertson, T. C. Friedrich, A. B. McDermott, N. A. Wilson, D. B. Allison, W. C. Koff, W. E. Johnson, and D. I. Watkins**. 2011. The TRIM5{alpha} genotype of rhesus macaques affects

acquisition of simian immunodeficiency virus SIVsmE660 infection after repeated limiting-dose intrarectal challenge. Journal of virology **85**:9637-40.

43. **Roos, R. S., M. Loetscher, D. F. Legler, I. Clark-Lewis, M. Baggiolini, and B. Moser**. 1997. Identification of CCR8, the receptor for the human CC chemokine I-309. The Journal of biological chemistry **272**:17251-4.

44. **Rot, A., M. Krieger, T. Brunner, S. C. Bischoff, T. J. Schall, and C. A. Dahinden**. 1992. RANTES and macrophage inflammatory protein 1 alpha induce the migration and activation of normal human eosinophil granulocytes. The Journal of experimental medicine **176**:1489-95.

45. **Song, B., B. Gold, C. O'Huigin, H. Javanbakht, L. Xing, M. Stremlau, C. Winkler, M. Dean, and J. Sodroskil**. The B30.2(SPRY) domain of the retroviral restriction factor TRIM5α exhibits lineage-specific length and sequence variation in primates. Journal of virology. American Society for Microbiology **79**:6111-6121.

46. **Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, and D. Altshuler**. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409**:928-33.

47. **Schall, T. J., K. Bacon, K. J. Toy, and D. V. Goeddel**. 1990. Selective attraction of monocytes and T lymphocytes of the memory phenotype by cytokine RANTES. Nature **347**:669-71.

48. **Scherer, E., D. Douek, and A. McMichael**. 2008. 25 years of HIV research on virology, virus restriction, immunopathogenesis, genes and vaccines. Clinical and experimental immunology **154**:6-14.

49. **Schutyser, E., S. Struyf, P. Menten, J. P. Lenaerts, R. Conings, W. Put, A. Wuyts, P. Proost, and J. Van Damme**. 2000. Regulated production and molecular diversity of human liver and activation-regulated chemokine/macrophage inflammatory protein-3 alpha from normal and transformed cells. Journal of immunology (Baltimore, Md. : 1950) **165**:4470-7.

50. **Segat, L., L. A. C. Brandão, R. L. Guimarães, A. Pontillo, E. Athanasakis, R. M. de Oliveira, L. C. Arraes, J. L. de Lima Filho, and S. Crovella**. 2010. Polymorphisms in innate immunity genes and patients response to dendritic cell-based HIV immuno-treatment. Vaccine **28**:2201-6.

51. **Silvestri, G.** 2008. AIDS pathogenesis: a tale of two monkeys. Journal of medical primatology **37 Suppl 2**:6-12.

52. **Singh, P., G. Kaur, G. Sharma, and N. K. Mehra**. 2008. Immunogenetic basis of HIV-1 infection, transmission and disease progression. Vaccine **26**:2966-80.

53. **Sleasman, J.** 2003. 13. HIV-1 infection*1. Journal of Allergy and Clinical Immunology **111**:582-592.

54. **Sobieszczyk, M. E., J. R. Lingappa, and M. J. McElrath**. 2011. Host genetic polymorphisms associated with innate immune factors and HIV-1. Current opinion in HIV and AIDS **6**:427-34.

55. **Sozzani, S., P. Allavena, A. Vecchi, and A. Mantovani**. 1999. The role of chemokines in the regulation of dendritic cell trafficking. Journal of leukocyte biology **66**:1-9.

56. **Stevenson, M.** 2003. HIV-1 pathogenesis. Nature Medicine **9**:853-860.

57. **Vanden Haesevelde, M. M., M. Peeters, G. Jannes, W. Janssens, G. van der Groen, P. M. Sharp, and E. Saman**. 1996. Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee. Virology **221**:346-50.

58. **Williams, K. C., and T. H. Burdo**. 2009. HIV and SIV infection: the role of cellular restriction and immune responses in viral replication and pathogenesis. APMIS□: acta pathologica, microbiologica, et immunologica Scandinavica **117**:400-12.

59. **Zlotnik, A., O. Yoshie, and H. Nomiyama**. 2006. The chemokine and chemokine receptor superfamilies and their molecular evolution. Genome biology **7**:243.

60. **Zlotnik, A., and O. Yoshie**. 2000. Chemokines: a new classification system and their role in immunity. Immunity **12**:121-7.

61. 2011. Global HIV/AIDS Response: Progress Report 2011. at <http://whqlibdoc.who.int/publications/2011/9789241502986_eng.pdf>