8-12-2014 12:00 AM

Evolutionary and in silico analysis of the antiviral TRIM22 gene

Jenna Kelly, *The University of Western Ontario*

Supervisor: Stephen Barr, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Microbiology and Immunology

EVOLUTIONARY AND IN SILICO ANALYSIS OF THE ANTIVIRAL TRIM22 GENE

(Thesis format: Integrated Article)

by

Jenna N. Kelly

Graduate Program in Microbiology and Immunology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

Tripartite motif protein 22 (TRIM22) is an evolutionarily ancient interferon-induced protein that been shown to potently inhibit human immunodeficiency virus (HIV), hepatitis B virus (HBV), and influenza A virus (IAV) replication. Altered TRIM22 expression levels have also been linked to autoimmune disease, cancer, and cellular proliferation. Despite its important role in a number of biological processes, the factors that influence TRIM22 expression and/or antiviral activity remain largely unknown. To identify key functional sites in TRIM22, we performed extensive evolutionary and *in silico* analyses on the *TRIM22* coding region. These tools allowed us to pinpoint multiple sites in TRIM22 that have evolved under positive selection during mammalian evolution, including one site that coincides with the location of a common non-synonymous SNP (nsSNP) in the human *TRIM22* gene (*TRIM22* rs1063303:G>C). We found that the frequency of *TRIM22* rs1063303:G>C varied considerably among different ethnic populations and African (AFR), American (AMR), and European (EUR) populations contained an excess of intermediate frequency *TRIM22* rs1063303:G>C alleles when compared to a neutral model of evolution. The latter is typically indicative of balancing selection, a non-neutral selective process that maintains polymorphism in a population. Interestingly, we also found that the *TRIM22* nsSNP rs1063303:G>C had an inverse impact on TRIM22 function. *TRIM22* rs1063303:G>C increased TRIM22 expression levels, but decreased its anti-HIV activity and altered its subcellular localization pattern. In addition to these studies, we used a variety of *in silico* methods to prioritize and delineate other functional sites in TRIM22. We showed that the majority of positively selected sites in the C-terminal B30.2 domain of TRIM22 are located in one of four surface-exposed variable loops that are critical for the anti-HIV effects of the closely-related TRIM5α protein. Moreover, we used six different *in silico* nsSNP prediction programs to screen all of the nsSNPs in the *TRIM22* gene and identified 14 high-risk nsSNPs that are predicted to be highly deleterious to TRIM22 function. Finally, to examine the *TRIM22* nsSNP rs1063303:G>C in a more isolated population, we genotyped this nsSNP in two Inuit populations (Canadian and Greenlandic Inuit). We found that the *TRIM22* rs1063303:C allele is inordinately prevalent in the Inuit compared to non-Inuit populations and that these two populations do not contain an excess of

intermediate frequency *TRIM22* rs1063303:G>C alleles compared to a neutral model of evolution, indicating that site *TRIM22* rs1063303:G>C has not evolved under balancing selection in the Inuit. Lastly, we found an interesting association between the *TRIM22* rs1063303:C allele and serum levels of triglycerides (TG) and high-density lipoprotein (HDL). Taken together, the results presented here identify a number of pertinent sites in the TRIM22 protein that likely influence its biological and/or antiviral functions.

Keywords

Tripartite motif protein 22, host restriction factors, evolution, interferon, innate immune system, single nucleotide polymorphism, *in silico* analysis

Co-Authorship Statement

Chapter 2 of this thesis was published in Human Mutation 35: 1072-1081, June 2014 doi: 10.1002/humu.22595. I was involved in performing all of the experiments with the exception of those described in Fig 2.4d, which were primarily performed by M. W. Woods.

Chapter 3 of this thesis was published in PLoS One 9 (7): e101436, July 2014. I was involved in performing all of the experiments.

Acknowledgements

I am fortunate to have many remarkable and inspiring people in my life who have made the completion of this dissertation possible. I would like to thank my supervisor, Dr. Stephen Barr, for his guidance and support throughout my PhD studies. His positive attitude and encouragement helped me overcome many of the challenging situations I encountered during my research. I am also thankful to him for giving me the freedom to explore my own ideas and forge my own path as a researcher. I would like to thank my advisory committee members, Dr. Greg Dekaban and Dr. Joe Mymryk, for their many insightful comments and suggestions. Their guidance, support, and 'open-door policy' have been invaluable through both successes and difficulties. I would also like to thank Dr. Robert Hegele for allowing me to work in his lab and for sharing his genotyping expertise with me. His knowledge, guidance, and commitment to the highest research standards inspired and motivated me.

I am deeply indebted to my amazing family and friends for their constant support and encouragement throughout my studies. I am forever grateful to my incredible parents – Neta & Scott Leonard and Jim & Judy Kelly – and to my siblings – Mike, Kate, Erica, and Tyler – for their constant love and support over the years. Without their wisdom, strength, and encouragement, none of this would have been possible. I am so lucky to have you all in my life. Thank you for always believing in me! I am also immensely grateful to my boyfriend, Jamie, for his love, encouragement, and patience during my studies. Thank you for always making me smile through the best and worst of times. I would also like to thank my friends and colleagues for making graduate school a truly unforgettable and enjoyable experience. I owe a particularly enormous 'thank you' to Stacey, Matt, and Teresa, for always being there to discuss science, life, and just about everything else in between. You have all been wonderful friends and I could not have done this without you! Sincere thanks as well to my extended family – grandparents, aunts, uncles, and cousins. You have all been instrumental contributors to my success.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIC	Akaike Information Content
AP-1	Activator protein 1
APOBEC3G	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide 3G
APR	Acute phase response
ARF	ADP ribosylation factor-like domain
BB	B-box domain
BR	Bromodomain
BST-2	Bone marrow stromal antigen 2
CB	Cajal bodies
CC	Coiled-coil domain
CD4	Cluster of differentiation 4
COS	C-terminal subgroup one signature domain
CP	Core promoter
Cys	Cysteine residue
E1	Ubiquitin activating enzyme
E2	Ubiquitin conjugating enzyme
E3	Ubiquitin ligating enzyme
eISRE	5’ extended IFN-stimulated response element
EMCV	Encephalomyocarditis virus

FIL	Filamin-type immunoglobulin domain
FN3	Fibronectin type 3 domain
Fv1	Friend virus susceptibility 1 locus
G1	Group 1 TRIM
G2	Group 2 TRIM
GAS	Gamma interferon activation site
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HDL	High density lipoprotein
HECT	Homologous to the E6-AP carboxyl terminus
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HOM	Hominoids
huTRIM5α	Human TRIM5α
IFN	Interferon
IFNAR	Interferon alpha/beta receptor
IL-1	Interleukin 1
IRF-1	Interferon regulatory factor 1
IRF-3	Interferon regulatory factor 3

IRF-7	Interferon regulatory factor 7
IRF-8	Interferon regulatory factor 8
ISG	Interferon stimulated gene
ISG15	Interferon stimulated gene 15
ISGF3	Interferon stimulated gene factor 3
ISRE	Interferon stimulated response element
JAK/STAT	Janus kinases/signal transducers and activators of transcription
LDL	Low density lipoprotein
LPS	Lipopolysaccharide
LTA	Lipoteichoic acid
LTR	Long-terminal repeat
Lys	Lysine residue
MATH	Merpin and tumor-necrosis factor receptor-associated factor homology domain
MDM	Monocyte derived macrophages
NHL	NHL repeats
N-MLV	N-tropic murine leukemia virus
NB	Nuclear bodies
NF-κB	Nuclear factor-kappa B
NLR	Nucleotide binding oligomerization domain (NOD)-like receptor

NLS	Nuclear localization signal
NP	Nucleoprotein
nsSNP	Non-synonymous single nucleotide polymorphism (SNP)
NWM	New world monkey
OWM	Old world monkey
PAMP	Pathogen-associated molecular pattern
PHD	Plant homeodomain
PIC	Preintegration complex
PML	Promyelocytic leukemia protein (also called TRIM19)
PRR	Pattern recognition receptor
PTM	Post-translational modification
qPCR	Quantitative polymerase chain reaction
RBCC	RING, BB1 and/or BB2, and CC domains
rhTRIM5α	Rhesus macaque TRIM5α
RING	Really interesting new gene domain
RLR	Retinoic acid inducible gene (RIG-I)-like receptor
SAMHD1	Sterile alpha motif and histidine/aspartic acid domain-containing protein
SIFT	Sorting intolerant from tolerant
SIM	SUMO interacting motif

SIV	Simian immunodeficiency virus
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SPRY	SplA/ryanodine receptor domain
SUMO	Small ubiquitin-like modifier
TAK1	Transforming growth factor beta-activated kinase 1
TG	Triglyceride
TLR	Toll-like receptor
TM	Transmembrane domain
TNFα	Tumor necrosis factor alpha
TRIM	Tripartite motif protein
TRIM22	Tripartite motif protein 22
TRIM5α	Tripartite motif protein 5 alpha
UBE2D	Ubiquitin conjugating enzyme 2/E2 class D
UBE2E	Ubiquitin conjugating enzyme 2/E2 class E
v1-v4	Variable loops 1-4
VLDL	Very low density lipoprotein
VSV	Vesicular stomatitis virus

Chapter 1

1 Introduction

The immune system has evolved over millennia to provide effective defense measures against invading pathogens and promote host survival. The vertebrate immune system can be divided into innate and adaptive arms, which together comprise both germline-encoded and acquired immune responses. The innate immune system is evolutionarily more ancient than the adaptive immune system and constitutes the first line of defense against infectious agents. It is also essential for initiating subsequent adaptive immune responses, which unlike innate immune responses, are characterized by specificity and the generation of immunological memory [1]. Innate immunity is activated upon pathogen recognition by a limited repertoire of germline-encoded pattern recognition receptors (PRR). PRRs can be grouped into several distinct families: 1) the membrane bound toll-like receptors (TLR), 2) the cytosolic retinoic acid inducible gene (RIG-I)-like receptors (RLR), 3) the nucleotide binding oligomerization domain (NOD)-like receptors (NLR), and 4) DNA sensors [2,3]. PRRs recognize conserved products of microbial metabolism called pathogen-associated molecular patterns (PAMP). PAMPs, which are not usually present in host cells, bind to specific PRRs to activate innate immunity. For example, specific TLR family members recognize genomic material or replication intermediates from viruses (i.e. TLRs 3, 7, 8, and 9), while other TLRs recognize diverse microbial products from parasites, fungi, and/or bacteria [4–8].

PRR-PAMP binding triggers a complex downstream signaling cascade that induces the activation of multiple cytosolic transcription factors, including nuclear factor-kappa B (NF-κB), activator protein 1 (AP-1), and interferon regulatory factors 3 (IRF-3) and 7 (IRF-7). Activation of NF-κB and AP-1 results in the upregulation of proinflammatory cytokines [e.g. interleukin 1 (IL-1) and tumor necrosis factor alpha (TNF-α)], whereas activation of IRF-3 and/or IRF-7 leads to the production of type I interferons (IFN) [9]. Type I IFNs are one of the three major classes of the IFN family, which is a group of

secreted cytokines that are essential for host antiviral defense and immune activation. Type I IFNs are produced in direct response to viral infection and comprise multiple subtypes of IFN-α plus IFN-β and several additional IFNs (δ, ε, κ, τ, ω) with less-definitive roles in antiviral immunity. The other two major classes of IFNs, type II and type III, each contain only one subtype. IFN-γ (type II IFN) is generally produced by activated T lymphocytes or natural killer cells in response to cytokine secretion from infected cells or macrophages. Similar to type I IFNs, IFN-λ (type III IFN) is typically produced in direct response to viruses; however type III IFN binds to different cognate receptors than type I IFNs. Type III IFN receptors are found mainly on epithelial cells, whereas type I IFN receptors are present on all cell types [10–13].

Canonical type I IFN production is essential for the host antiviral response to infection and critical for host survival. Following synthesis and secretion, type I IFNs (IFN-α/β) bind to the IFN-α/β receptor (IFNAR) complex on the cell surface in an autocrine and paracrine manner, which induces signal transduction through the janus kinases/signal transducers and activators of transcription (JAK/STAT) pathway. Specifically, IFN-α/β binding to the IFNAR complex leads to the phosphorylation of receptor-bound tyrosine kinases janus kinase 1 (JAK1) and tyrosine kinase 2 (TYK2), which in turn leads to the phosphorylation of signal transducers and activators of transcription 1 (STAT1) and 2 (STAT2). STAT1/STAT2 phosphorylation results in STAT1/STAT2 dimerization and recruitment of interferon regulatory factor 9 (IRF-9) to form the IFN-stimulated gene factor 3 (ISGF3) complex. The ISGF3 complex, which acts as a transcription factor, translocates to the nucleus where it upregulates a large number of IFN-stimulated genes (ISG) (Fig 1.1). Many ISGs have antiviral properties, especially host restriction factors, which directly interfere with specific stages of viral replication. Collectively, ISGs establish a powerful cell-intrinsic antiviral state that is essential for withstanding and controlling many pathogens [14,15].

Figure 1.1: ISG induction by type I IFN signaling.

Type I IFNs bind to the IFNAR1/2 heterodimer complex on the surface of the cell and activate the JAK/STAT signaling pathway. IFN-IFNAR1/2 binding triggers JAK1 and TYK2 phosphorylation, which leads to the recruitment and phosphorylation of STAT1 and STAT2. Phosphorylated STAT1 and STAT2 form a heterodimer and bind to IRF-9, creating the ISGF3 complex. ISGF3 translocates to the nucleus where it binds to IFN-stimulated response elements (ISRE) in the promoters of multiple ISGs. This induces a powerful antiviral state in the cell that inhibits many pathogens.

Figure 1.1

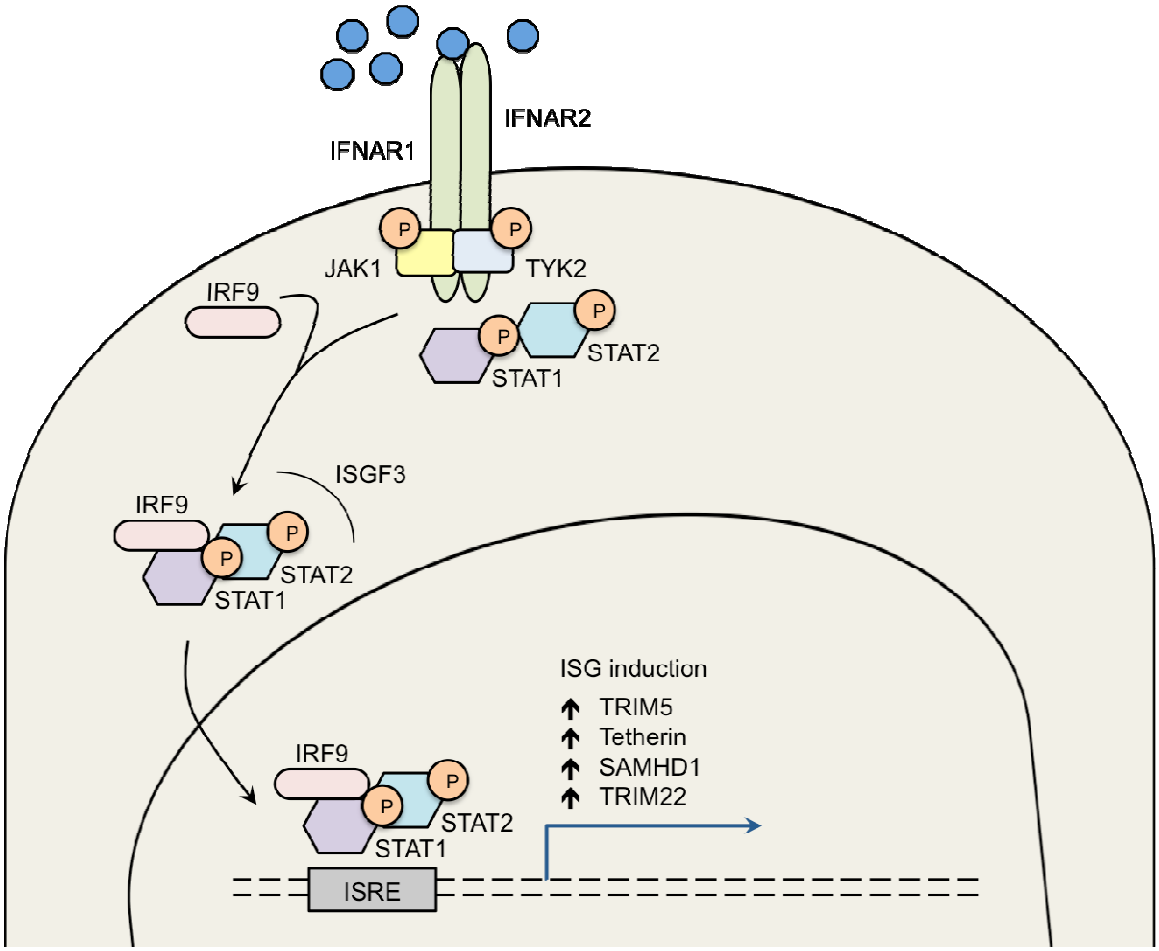


Figure 1.1: ISG induction by type I IFN signaling.

1.1 Host restriction factors

Host restriction factors are IFN-induced proteins that inhibit specific steps in the viral replication lifecycle. Researchers first coined the term 'restriction factor' in the 1970's following the characterization of the mouse *friend virus susceptibility 1* (*Fv1*) locus, which conferred resistance to murine retroviruses [16]. Since this discovery, restriction factors have been identified in many vertebrate species, and today the term is used to describe host proteins that inhibit the replication of any animal virus [17,18]. Several major characteristics distinguish restriction factors from other host proteins. Typically, host restriction factors: 1) have antiviral activity as their main biological function, 2) are induced by IFN signaling and/or viral infection, 3) contain evolutionary signatures of positive selection, and 4) are antagonized by a viral protein [18,19]. Characteristics for a number of well-known host restriction factors, including the apolipoprotein B mRNA-editing enzyme catalytic polypeptide 3G (APOBEC3G), tripartite motif protein 5 alpha (TRIM5α), bone marrow stromal antigen 2 (BST-2 or tetherin), and sterile alpha motif and histidine/aspartic acid domain-containing protein (SAMHD1), are summarized in Table 1.1. at the end of this section.

Type I IFNs have long been recognized as potent inhibitors of human immunodeficiency virus (HIV) replication [20]. As such, HIV has been used for many years as a system to identify and characterize the IFN-induced host restriction factors responsible for this inhibition. Type I IFNs block both early and late stages of the HIV lifecycle [21–25]. Early stages of HIV replication include HIV binding and fusion to the host cell via its CD4 receptor, uncoating of the viral core in the cell cytoplasm, reverse transcription of the single-stranded RNA viral genome, and integration of the viral cDNA into the host cell genome. Late stages include transcription of the integrated HIV provirus from the 5' long-terminal repeat (LTR) promoter, viral RNA export, translation of viral proteins, trafficking/assembly of new virions, virion budding/release, and maturation of virions into fully infectious HIV particles (Fig 1.2). Host restriction factors, which form the effector arm of the IFN response, target specific stages (both early and late) in the HIV replication life cycle. However, HIV also encodes viral antagonists to counteract host restriction factors. These antagonists are typically HIV accessory proteins, namely Vif,

Nef, Vpu, and Vpr, which are only required for viral replication in the presence of host restriction factors [26]. Whereas HIV type 1 (HIV-1) encodes only the latter accessory proteins, a number of strains of simian immunodeficiency virus (SIV) and HIV type 2 (HIV-2) also encode an additional accessory protein called Vpx [27]. Stages of HIV-1 replication (and/or SIV or HIV-2 replication) targeted by the APOBEC3G, TRIM5α, BST-2/tetherin, and SAMHD1 host restriction factors plus their viral antagonists (if known) are shown in Fig 1.2 and outlined in further detail in the text below.

1.1.1 Inhibition of HIV replication by host restriction factors

Multiple host restriction factors that inhibit HIV replication have been identified within the last decade, some of which have been studied extensively. One such protein is the host restriction factor APOBEC3G, which was first identified as a HIV-1 restriction factor in 2002 [28]. APOBEC3G is targeted by the HIV-1 Vif accessory protein; however when Vif is not present, APOBEC3G is packaged into assembling HIV-1 virions and released into the cytoplasm of newly infected target cells. APOBEC3G is a cytidine deaminase enzyme or an enzyme that converts deoxycytidine into deoxyuridine in a nucleic acid sequence. When APOBEC3G is present in newly infected target cells, it converts deoxycytidine to deoxyuridine in the nascent viral single-stranded negative-sense cDNA. This results in deoxyguanine to deoxyadenine hypermutation and loss of genetic integrity in the HIV-1 plus-strand sequence and inhibits HIV-1 replication at the level of reverse transcription (Fig 1.2) [29–32]. When Vif is present in cells, APOBEC3G becomes polyubiquitylated and is targeted for proteasomal degradation. Vif facilitates this process by binding directly to both APOBEC3G and the Cul5 E3 ligase ubiquitin complex. Interestingly, Vif antagonism is not always fully efficient and hypermutated HIV-1 sequences are readily recovered from HIV-1 infected individuals. Moreover, hypermutation frequency has been shown to correlate inversely with plasma viremia in several sizable HIV-1 cohorts [33–37].

Another well-studied HIV-1 restriction factor is the TRIM5α protein. Unlike other host restriction factors, TRIM5α typically only restricts retroviruses that have been isolated

Figure 1.2: Stages of HIV replication targeted by host restriction factors.

Host restriction factors, such as APOBEC3G, TRIM5α, BST-2/tetherin, and SAMHD1, target multiple stages of HIV-1 (or SIV/HIV-2) replication. Post-entry, TRIM5α targets the assembled HIV-1 capsid protein on the viral core and facilitates its disassembly. In the absence of Vif, the APOBEC3G protein creates deoxycytidine to deoxyuridine mutations in the nascent viral cDNA, which inhibits HIV-1 replication at the level of reverse transcription. SAMHD1, which is countered by Vpx (SIV/HIV-2), also inhibits reverse transcription; however, it does so by decreasing the cellular pool of dNTPs required for viral cDNA synthesis. BST-2/tetherin inhibits HIV-1 replication at the budding/release stage by 'tethering' HIV-1 virions to the cell membrane. BST-2/tetherin is antagonized by the HIV-1 Vpu protein.

Figure 1.2

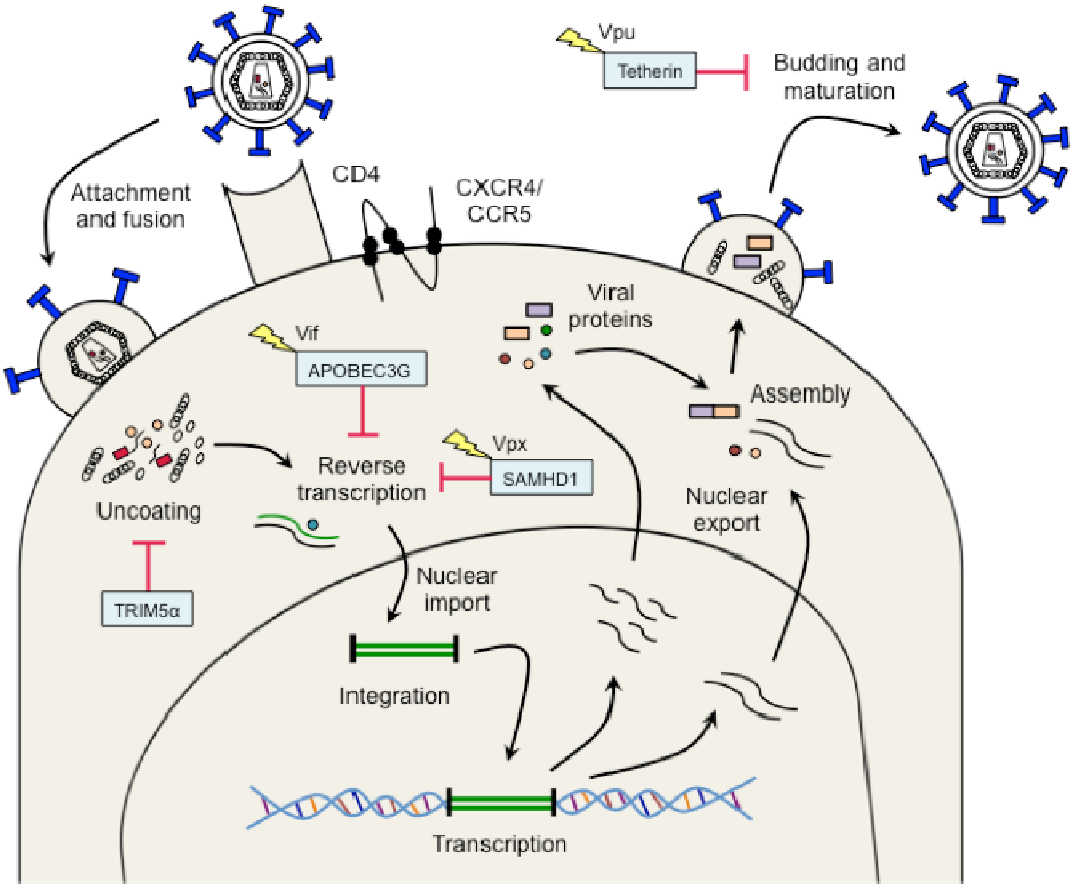


Figure 1.2: Stages of HIV replication targeted by host restriction factors.

from a host species other than its own. For example, the human TRIM5α (huTRIM5α) protein strongly inhibits N-tropic murine leukemia virus (N-MLV), but has very weak activity against HIV-1. Similarly, rhesus monkey TRIM5α (rhTRIM5α) does not inhibit SIV_{mac} (SIV strain isolated from rhesus macaques that were experimentally inoculated with SIV from asymptomatic sooty mangabey monkeys [SIV_{smm}][38]). The majority of studies have shown that rhTRIM5α inhibits HIV-1 by targeting the assembled HIV-1 capsid protein found on the mature viral core. Capsid is the main structural protein of HIV-1 and forms the inner viral protein shell that surrounds the HIV-1 genome and its core proteins in mature virions [39]. To inhibit HIV-1 replication, rhTRIM5α binds to the incoming mature viral core and facilitates its premature disassembly. This leads to the inhibition of reverse transcription and subsequent HIV-1 replication (Fig 1.2) [40–42]. Of interest, rhTRIM5α, which localizes in subcellular structures called cytoplasmic bodies, has been shown to mediate its own polyubiquitylation and proteasomal degradation during the HIV-1 restriction process. Self-ubiquitylation of rhTRIM5α may destabilize the HIV-1 capsid lattice and induce its disassembly. Notably, other studies have shown that proteasome inhibitors do prevent inhibition of reverse transcription by rhTRIM5α, but do not disrupt overall rhTRIM5α-mediated HIV-1 restriction [43–46]. Thus, it has also been proposed that rhTRIM5α may inhibit HIV-1 replication in two or more redundant ways, such as by blocking the nuclear translocation of the viral preintegration complex (PIC). The viral PIC includes the viral cDNA plus both viral and host proteins and its nuclear translocation is required for proper integration of HIV-1 into the host genome [47]. To date, there have been no HIV-1 antagonists to rhTRIM5α identified.

Tetherin or BST-2 was only recently identified as a HIV-1 restriction factor, but since this time has been studied in great detail. BST-2/tetherin is antagonized by the HIV-1 Vpu protein and targets a later stage of viral replication than APOBEC3G or TRIM5α. When Vpu is absent, BST-2/tetherin 'tethers' fully formed HIV-1 virions to the plasma membrane of infected cells (Fig 1.2). Captured HIV-1 virions are later internalized by endocytosis and accumulate in $CD63^+$ endosomes, where they are likely degraded [48–53]. BST-2/tetherin contains two membrane anchors, the N-terminal transmembrane domain and the C-terminal glycosylphosphatidylinositol domain, which are both essential for BST-2/tetherin-mediated virion retention and endocytosis [54–56]. Although several models

for BST-2/tetherin configuration have been proposed, experimental evidence strongly supports a parallel-membrane-spanning model. This models postulates that one end of two BST-2/tetherin monomers (either the N-terminus or C-terminus) is anchored to the cell membrane and the other end of both monomers is anchored to the viral membrane. BST-2/tetherin is downregulated from the cell surface when Vpu is present and likely degraded via the proteasomal and/or lysosomal pathway. However, some studies have shown that Vpu may simply sequester BST-2/tetherin in the cytoplasm. These methods are unlikely to be mutually exclusive and may work together to counteract the BST-2/tetherin protein. The mechanism of Vpu antagonism may also depend on the cellular context of HIV-1 restriction [19,57].

SAMHD1, the most recently identified host restriction factor, restricts SIV/HIV-2 viral replication in non-dividing cell types, including $CD4^+$ T cells, dendritic cells, and macrophages. Vpx, an accessory protein encoded by HIV-2 and certain strains of SIV, targets SAMHD1 and induces its proteasomal degradation. In the absence of Vpx, SAMHD1 inhibits SIV/HIV-2 replication by decreasing the intracellular pool of dNTPs to below the level required for viral cDNA synthesis. This results in early post-entry restriction at the level of reverse transcription (Fig 1.2) [58–61]. While HIV-1 is unable to neutralize SAMHD1, one recent study demonstrated that a subset of Vpr proteins from ancestral lentiviruses without Vpx are able to disrupt SAMHD1-mediated restriction. Since Vpr is hypothesized to have originally given rise to Vpx via a gene duplication event, the ability of Vpx to antagonize SAMHD1 may have originated in Vpr before it was lost following Vpx acquisition. Interestingly, distinct orthologues of Vpx and Vpr have evolved multiple ways to recognize SAMHD1. For example, certain orthologues interact with the N-terminus of SAMHD1, while others interact with the C-terminus; however, they all target SAMHD1 for proteasomal degradation. Thus, even though the method of Vpx/Vpr antagonism toward SAMHD1 has remained the same, the site of antagonism has changed back and forth throughout viral evolution [62,63].

1.1.2 Innate immune regulation by host restriction factors

In addition to their direct antiviral effects, a growing number of host restriction factors indirectly influence the antiviral response by regulating important innate immune signaling pathways. For example, TRIM5α was recently shown to function as a PRR for the HIV-1 capsid lattice. Upon capsid recognition, TRIM5α activates the transforming growth factor beta-activated kinase 1 (TAK1), AP-1 and NF-κB signaling, and the transcription of proinflammatory cytokines [64]. This signaling is required for TRIM5α-mediated retroviral restriction. Thus, in addition to its direct antiviral effects, TRIM5α also inhibits HIV-1 replication by activating innate immune signaling pathways. BST-2/tetherin and SAMHD1 have also recently been implicated in the activation of innate immune and proinflammatory signaling pathways [65,66].

1.2 The TRIM family

The tripartite motif or TRIM family is a large group of evolutionarily ancient proteins that are involved in multiple biological processes and have been linked to a number of different human diseases. Many TRIM proteins are upregulated by type I and type II IFNs and some, such as the host restriction factor TRIM5α, have antiviral properties. TRIM proteins are particularly adept at targeting retroviruses and to date, ~20 TRIM proteins have been shown to inhibit retrovirus replication [67–70]. Recent studies have also identified several TRIM proteins that regulate key proteins involved in innate immune signaling, including IRF-3, IRF-8, RIG-I, and NF-κB [71,72]. In addition to their role in antiviral immunity, TRIM proteins have also been implicated in numerous Mendelian inherited disorders and autoimmune diseases, cell cycle progression, development, and several types of cancer [73–76].

1.2.1 Evolution and classification

TRIM proteins are conserved throughout the metazoan kingdom; however, the TRIM family has expanded rapidly during vertebrate evolution. For example, the invertebrate

Table 1.1

Table 1.1: Characteristics of well-studied host restriction factors.

Restriction factor	IFN induced?	Viral target	Stage of life cycle inhibited	Viral antagonists [1]	Positive selection?
Fv1 (mouse)	No	Retroviruses	Capsid uncoating [77]	None known	Yes [78]
MxA	Yes	Orthomyxoviruses Paramyxoviruses Hepadnaviruses Rhabdoviruses Alphaviruses Bunyaviruses Togaviruses Picornaviruses	Nucleocapsid transport or another early life cycle step [79]	None known	Yes [80]
MxB	Yes	Retroviruses	Nuclear import or integration [81]	None known	ND
IFITM1, IFITM2, IFITM3	Yes	Orthomyxoviruses Flaviviruses Coronavirus Rhabdoviruses Alphaviruses Bunyaviruses Filoviruses	Endosomal fusion or uncoating [82,83]	None known	ND
TRIM5α	Yes	Retroviruses Hepadnaviruses	Capsid uncoating [84]	None known	Yes [85]
APOBEC3G	No	Retroviruses Hepadnaviruses Adenoviruses Paramyxoviruses Retrotransposons	Reverse transcription [31,33,35,77,86–90]	Vif (HIV-1) Bet (PFV) glycol-Gag (MLV)	Yes [91]
SAMHD1	Yes	Retroviruses Herpesviruses	Reverse transcription [58,60,92,93]	Vpx (HIV-2) Vpr (SIV)	Yes [94]

		Poxviruses			
BST-2/Tetherin	Yes	Retroviruses Flaviviruses Herpesviruses Rhabdoviruses Paramyxoviruses Arenaviruses Filoviruses	Budding [48,51,53,54,95–98]	Vpu (HIV-1) Nef (SIV) Env (HIV-2) VP40 (Ebola) K5 (KSHV) NA (IAV) gM (Herpes)	Yes [99]
PKR	Yes	Retroviruses Orthomyxoviruses Filoviruses Paramyxoviruses Bunyaviruses Herpesviruses Poxviruses	Viral protein translation [100–102]	NS1 (IAV, IBV) NSs (RVFV) VP35 (Ebola)	Yes [103]
HERC5	Yes	Retroviruses Orthomyxoviruses	Nuclear export, assembly [104–106]	None known	Yes [106]

[1] Viral antagonists: this is not an exhaustive list, particularly for PKR. Please refer to [107] for a more comprehensive list of PKR antagonists.

Abbreviations: Myxovirus resistance (Mx) proteins A and B, Interferon induced transmembrane (IFITM) proteins 1, 2, and 3, Prototype foamy virus (PFV), Glycosylated (glyco) Gag, Murine leukemia virus (MLV), Kaposi's sarcoma-associated herpesvirus (KSHV), Influenza A/B virus (IAV, IBV), Rift valley fever virus (RVFV), HECT domain and RCC1-like domain-containing protein 5 (HERC5). ND: not determined.

Drosophila melanogaster and *Caenorhabditis elegans* genomes contain 7 and 18 TRIM genes, respectively. In stark contrast, the vertebrate *Mus musculus* genome contains 64 *TRIM genes* and 100 *TRIM* genes have now been identified in the human genome. Notably, the TRIM family may still be expanding in humans. One study recently identified 11 *TRIM* genes that are specific to humans and African apes and another 7 *TRIM* genes that are found only in humans. According to this study, the novel *TRIM* genes were acquired through multiple segmental duplication events, the majority of which originated from a single chromosomal locus. One Han Chinese woman with 12 extra copies of these *TRIM* genes was also identified, documenting *TRIM* copy number variation in humans for the first time [108].

Several classification systems have been proposed for organizing the TRIM family of proteins in humans. Initially, the TRIM family was divided into 9 distinct subgroups according to their C-terminal domains [109]. This classification system was subsequently modified to include 2 new subgroups of TRIMs with previously mischaracterized C-terminal domains (C-I to C-XI) [68]. Another study divided the TRIM family into two major groups based on genomic organization, evolutionary properties, and domain structure. This study determined that 'Group 1' TRIM proteins contained two B-box (BB) domains, a variable C-terminal domain, and were found in both vertebrates and invertebrates. In contrast, 'Group 2' TRIM proteins had only one BB domain, a C-terminal splA/ryanodine receptor (SPRY) domain or PRY/SPRY (B30.2) domain, and were found only in vertebrates [110]. The most recent report, which performed a more stringent phylogenetic analysis of the TRIM family, divided TRIM proteins into 9 distinct subgroups (A-I). 'Group 2' TRIM proteins were placed in one subgroup (G); however, 'Group 1' TRIMs were further subdivided into 8 different phylogenetic subgroups (A-F, H, I). Importantly, this study consolidated data from previous reports and compared the phylogenetic and domain-based classification systems for the TRIM family (Fig 1.3) [111].

Figure 1.3: Subgroup classification of the TRIM family.

Subgroup classifications (A-I) are based on the most recent phylogenetic analysis of the TRIM family [111]. Previous C-terminal domain-based TRIM classifications (C-I to C-XI) are shown in parentheses [68]. TRIM proteins contain an N-terminal RING domain (blue square), one or two BB domains (dark/light pink circles), and a CC domain (orange rectangle). In addition, TRIM proteins contain one or more of 10 different C-terminal motifs (MATH, COS, FN3, PRY/SPRY, FIL [filamin], NHL, ARF, PHD, BR, and/or TM). Group 2 TRIM proteins (subgroup G) are labeled and encased in a red box. Other subgroups (A-E, H, I) are considered Group 1 TRIMs [110].

Figure 1.3

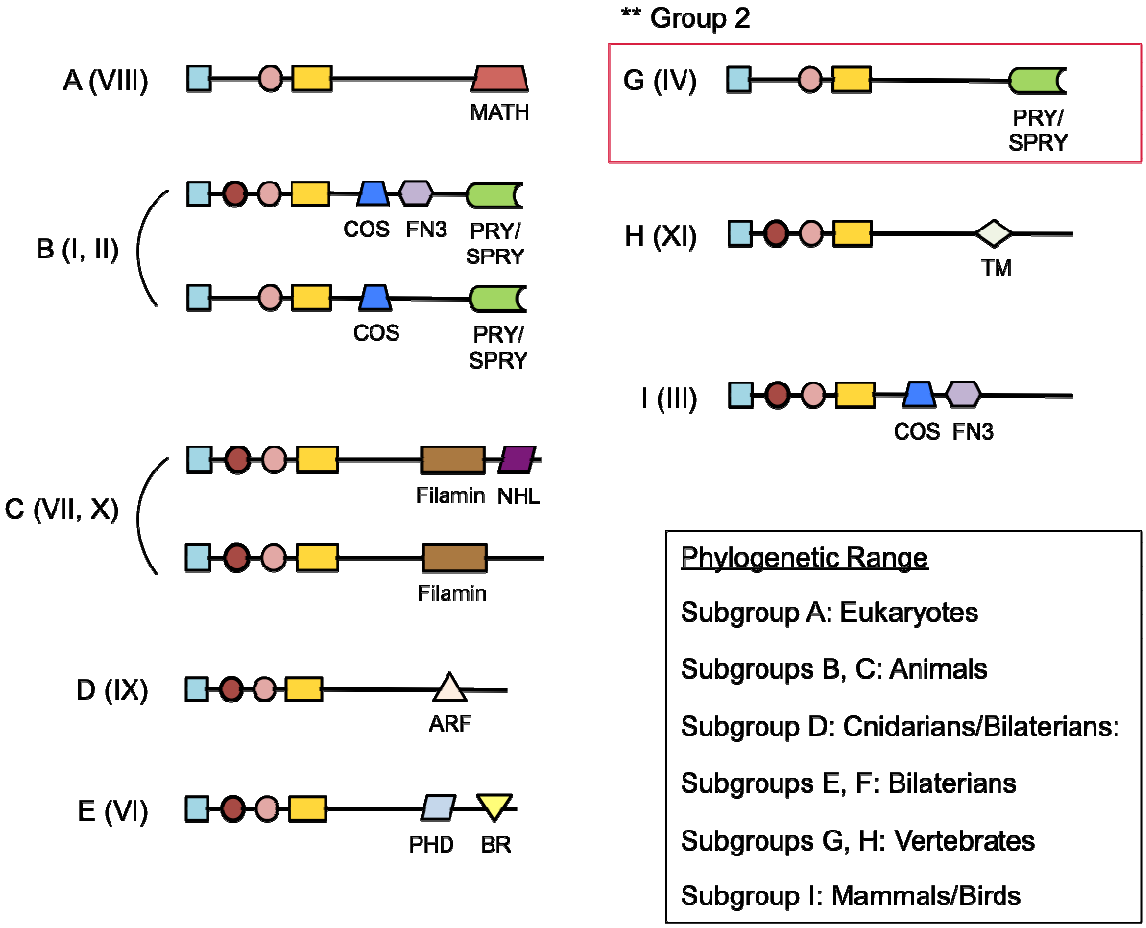


Figure 1.3: Subgroup classification of the TRIM family.

1.2.2 The RBCC motif

Members of the TRIM family are defined by their N-terminal TRIM or RBCC motif, which is comprised of a 'Really Interesting New Gene' (RING) domain, one or two BB domains, and a predicted coiled-coil (CC) region. The entire RBCC motif is present in most TRIM proteins; however, in TRIM proteins that lack one of these domains, the other domains are conserved in both order and spacing [68,112]. The C-terminal domain of TRIM family members is variable and contains any of 10 distinct motifs alone or in combination. The majority of TRIM proteins have a C-terminal SPRY or PRY/SPRY domain (also called a B30.2 domain). Other possible C-terminal domains include ARF (ADP ribosylation factor-like), the COS (C-terminal subgroup one signature) box, PHD (plant homeodomain), MATH (merpin and tumor-necrosis factor receptor-associated factor homology), FN3 (fibronectin type 3), and/or FIL (filamin-type immunoglobulin) (Fig 1.3) [112,113].

The first domain in the RBCC motif, the RING domain, is characterized by a canonical sequence that consists of a number of conserved cysteine and histidine residues. These residues form a specialized zinc finger and coordinate two zinc atoms in a cross-braced arrangement [114,115]. Studies have shown that RING finger proteins interact directly with ubiquitylation enzymes and often function as E3 ubiquitin ligases [116–119]. Consistent with these studies, many RING-containing TRIM proteins have been shown to possess E3 ubiquitin ligase activity [113,120,121]. In many (but not all) cases, this E3 ubiquitin ligase activity is required for TRIM-mediated antiviral activity. Previous studies have also shown that RING-mediated E3 ligase activity is necessary for activation of antiviral signaling by TRIMs 1, TRIM5, TRIM13, TRIM25, TRIM32, and TRIM62 [64,122–124].

Ubiquitylation is a three-step process that involves ubiquitin-activating (E1), ubiquitin-conjugating (E2), and ubiquitin-ligating (E3) enzymes. Ubiquitylation begins when the E1-activating enzyme activates ubiquitin and transfers it to an E2-conjugating enzyme. Following transfer, an E3 ligating enzyme facilitates ubiquitin attachment to the target substrate protein, which occurs via the formation of an isopeptide bond between the C-terminal glycine residue of ubiquitin and an internal lysine (Lys) residue of the target protein (Fig 1.4). There are two main families of E3 ubiquitin ligases: RING and HECT

(homologous to the E6-AP carboxyl terminus). Unlike HECT-containing E3 ligase proteins, RING E3 ligases bind directly to both the E2 conjugating enzyme and the target protein and this increased proximity facilitates the direct transfer of ubiquitin from the E2 enzyme to the target protein [23,29]. TRIM proteins represent the largest group of RING-containing E3 ubiquitin ligases and multiple TRIM family members were recently shown to interact with the UBE2D and UBE2E (ubiquitin conjugating enzymes 2/E2 class D and class E) classes of E2 conjugating enzymes; however, other combinations of TRIM-E2 proteins have also been identified [64,113,120,121,126].

Substrate proteins can be modified by a single ubiquitin moiety (monoubiquitylation) or can be tagged with a chain of ubiquitin (polyubiquitylation). Ubiquitin chains are linked together via the formation of an isopeptide bond between the C-terminal glycine residue of one ubiquitin and one of seven internal lysine residues of a second ubiquitin moiety. Notably, ubiquitin chains linked through distinct Lys residues have different cellular functions. Lys48 ubiquitin chains mark proteins for degradation by the 26S proteasome, whereas Lys63, other Lys-based ubiquitin chains, and monoubiquitylation often serve non-proteolytic functions. Depending on the substrate proteins and enzymes involved, these modifications can serve as signals for diverse cellular processes, including DNA repair, transcription, signal transduction, and/or intracellular trafficking [127]. Thus far, most well-characterized TRIM proteins have been shown to facilitate either Lys48 or Lys63 polyubiquitylation. Many TRIM proteins can also undergo self-ubiquitylation, including TRIM5α (both huTRIM5α and rhTRIM5α proteins), which induces its own proteasomal degradation and rapid turnover. Several groups have proposed that self-ubiquitylation of rhTRIM5α indirectly leads to the proteasomal degradation of HIV-1 virions. While the precise mechanism of TRIM5α-mediated restriction is still not fully understood, some studies have shown that HIV-1 restriction is dependent on TRIM5α's E3 ubiquitin ligase activity [43,45,46,128–131].

In addition to their function as E3 ubiquitin ligases, several TRIM proteins have been shown to function as E3 ligases for other ubiquitin-like proteins, such as SUMO (small ubiquitin-like modifier) and ISG15 (interferon stimulated gene 3) [132–134]. Ubiquitin-like

Figure 1.4: The RING E3 ligase-mediated ubiquitylation pathway.

Ubiquitin is activated by the E1 activating enzyme to form a thioester linkage with the active-site cysteine of E1. Activated E1-bound ubiquitin is transferred to the active-site cysteine of the E2 conjugating enzyme. E2 transfer also requires the formation of a thioester linkage. E2-bound ubiquitin then interacts with a RING E3 ligase enzyme, but does not directly transfer ubiquitin to the E3. Instead, the RING E3 ligase binds to both the appropriate protein substrate and the E2. The increased proximity between the E2-bound ubiquitin and the protein substrate facilitates the transfer of ubiquitin to a Lys residue on the substrate. The substrate can undergo monoubiquitylation (addition of one ubiquitin moiety) at one or more Lys residues or polyubiquitylation (addition of a chain of ubiquitin moieties linked together via internal Lys residues). Chains that are linked together by the Lys48 residue of ubiquitin typically target substrates for 26S proteasomal degradation.

Figure 1.4

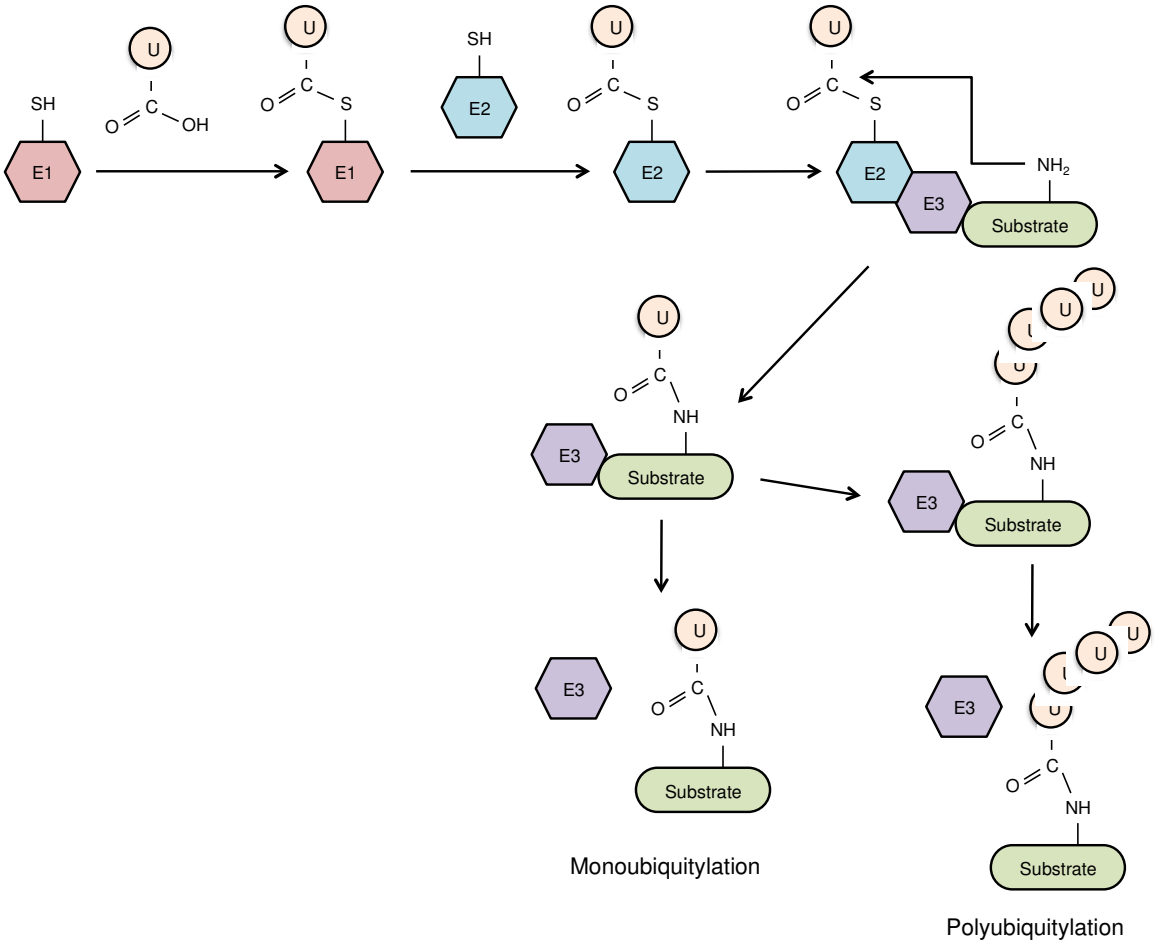


Figure 1.4: The RING E3 ligase-mediated ubiquitylation pathway.

proteins are activated and conjugated to target proteins using a distinct set of enzymes; however, the overall process is analogous to ubiquitylation [135]. Notably, unlike other SUMO E3 ligase proteins, the SUMO E3 ligase activity of TRIM proteins is dependent on both RING and BB domains [132]. Some TRIM family members have been shown to undergo SUMO modification themselves (self-sumoylation) and several TRIM proteins contain SUMO-interacting motifs (SIMs). SIMs bind to other proteins that have been modified with SUMO. Interestingly, two SIMs in huTRIM5α and rhTRIM5α play an important role in N-MLV and HIV-1 restriction [136–139].

Following the RING finger domain, TRIM proteins contain one or two BB domains. Similar to the RING domain, BB domains are zinc binding motifs with a number of conserved cysteine and histidine residues; however, unlike the RING domain, BBs are found exclusively in the TRIM family. There are two types of BBs, BB1 and BB2, that share similar but distinct consensus motifs. When both BB are present in TRIMs, BB1 always precedes BB2. If only one BB is present it is always BB2 [68,113,120]. Structural studies of several human TRIM proteins have revealed that TRIM BBs have a similar ternary structure as the RING finger domain [140–142]. Studies on the BB2 domains of human TRIM63 (MuRF1) and TRIM5α identified two unusual conserved clusters of hydrophobic residues on the surface of the domain. Two hydrophobic clusters that are flanked by a number of charged residues were identified on the surface of the BB2 domain of TRIM5α. A number of these residues (e.g. W115, L116, and R119) were required for major TRIM5α functions, including TRIM5α turnover, higher order self-association, formation of cytoplasmic bodies, HIV-1 capsid binding, and/or HIV-1 restriction [143]. TRIM63 had a similar cluster of solvent-exposed hydrophobic residues located at the surface of its BB2 domain, which form a dimer interface and mediate TRIM63 self-association [142].

The final component of the RBCC motif is the predicted CC region. The CC domain contains many predicted hyper-secondary structures and intertwined α-helices, which mediate homomeric and heteromeric interactions among TRIM proteins and other binding partners. A large number of TRIM proteins have been shown to self-associate through the CC domain [112]. The CC domain also promotes the formation of higher-

molecular-weight complexes that among other functions, define specific subcellular structures. For example, the CC domain of TRIM19 (also referred to as promyelocytic leukemia protein [PML]) is essential for the proper assembly and maintenance of macromolecular called nuclear bodies (NB). Moreover, the CC domain of TRIM5α facilitates TRIM5α trimerization. Mutations in CC that disrupt TRIM5α trimerization impair TRIM5α-mediated HIV-1 restriction [144–147]. Interestingly, it has been proposed that the pleiotropic effects of TRIM proteins may be due to the ability of the CC to facilitate diverse homomeric and heteromeric interactions [120,148].

1.2.3 The SPRY and PRY/SPRY (B30.2) domains

The SPRY and PRY/SPRY (B30.2) domains are the most common C-termini found in TRIM family proteins. The evolutionarily ancient SPRY domain (~140 amino acids) is present alone or fused to a related domain called the PRY domain (~60 amino acids), which always precedes the SPRY domain. Unlike the SPRY-only domain, the fused PRY/SPRY domain is only found in vertebrates and PRY/SPRY-containing proteins (including TRIMs) have expanded rapidly during vertebrate evolution. The reasons for this expansion are still unclear; however, it has been proposed that the PRY/SPRY domain has been selected and maintained in vertebrates as a component of immune defense [68,149]. Indeed, the PRY/SPRY domain in TRIM proteins is often critical for TRIM-mediated virus inhibition. For example, rhTRIM5α interacts with the HIV-1 capsid protein via its PRY/SPRY domain and both PRY and SPRY portions of the domain are necessary for this interaction. Furthermore, the PRY/SPRY domain in the TRIM25 protein, which activates the RIG-I signaling cascade, is both necessary and sufficient for its interaction with RIG-I. These PRY/SPRY-mediated protein-protein interactions are required for both rhTRIM5α and TRIM25 to execute their antiviral functions [40,122,150–153].

The SPRY and PRY/SPRY domain superfamily is among one of the largest families of protein interaction modules in humans. In addition to the TRIM family, the SPRY and PRY/SPRY domains are found in several other protein families, including a number of

families that contain proteins involved in ubiquitylation processes. As mentioned in the previous section, the TRIM family also contains a sizeable number of proteins that are involved in ubiquitylation. Thus, although SPRY and PRY/SPRY-containing proteins have diverse biological functions, it has been proposed in the literature that they may function primarily as target substrate recognition modules for E3 ubiquitin ligases [154,155]. This is consistent with experimental evidence for some members of the TRIM family, such as TRIM27, which interacts with the NOD2 (nucleotide-binding oligomerization domain-containing protein 2) protein through its PRY/SPRY domain and subsequently facilitates NOD2 polyubiquitylation and degradation using its RING domain [156]. However, in the majority of cases, the mechanistic details and/or interacting substrate proteins required for TRIM-mediated biological activities are unknown.

Several SPRY and PRY/SPRY domain structures have now been resolved, including a limited number of structures in complex with their binding partners. These studies have revealed that the SPRY and PRY/SPRY domain structures are extremely versatile and can interact with diverse ligands [157,158]. The core fold of the PRY/SPRY domain is a bent β-sandwich comprised of two antiparallel β sheets. The majority of conserved residues are located in the hydrophobic core between the antiparallel β sheets, whereas loops of variable length and sequence protrude from the core β-sandwich. In many PRY/SPRY-containing proteins, these variable loops form protein binding surfaces that determine substrate binding specificity [154,155,159–162]. For example, four variable (v) loops in the recently solved structure of the rhTRIM5α PRY/SPRY domain comprise the HIV-1 capsid binding site. The binding surface is dominated by one variable loop (v1) that is highly flexible and interacts weakly with multiple capsid epitopes. Interestingly, the authors of this study suggests that capsid recognition by rhTRIM5α may function in a similar manner as IgM-mediated antigen recognition [163].

1.1.3 Antiviral activity

Most TRIM proteins are upregulated by type I IFNs and multiple TRIMs have been shown to possess antiviral activity. As mentioned previously, TRIM5α restricts the

replication of diverse retroviruses, including HIV-1 and N-MLV. A comprehensive screen for antiretroviral activity of 55 human and mouse TRIMs revealed that 19 additional TRIM proteins inhibit the entry or release of HIV-1, MLV, and/or avian leucosis virus. Interestingly, unlike TRIM5α, most of the additional TRIM proteins inhibited late stages of the viral life cycle [69]. Multiple TRIM proteins have also been shown to have antiviral activity against hepatitis B virus (HBV). These TRIM proteins have been shown to significantly reduce the HBV transcription in HepG2 cells [164,165]. Some TRIM proteins, such as TRIM19/PML, have been implicated in the inhibition of additional viruses, including herpes simplex virus, human cytomegalovirus, vesicular stomatitis virus, and influenza A virus (IAV) [166]. Recently, TRIM56 was shown to inhibit the replication of bovine viral diarrhea virus *in vitro* and mouse-specific TRIM79α was identified as a potent inhibitor of tick-borne encephalitis virus [167,168]. In addition, TRIM21 was previously shown to act as an intracellular IgG receptor that neutralizes antibody-coated virus in the cytoplasm by targeting it for degradation by the 26S proteasome [169]. These studies suggest that TRIM family proteins restrict evolutionarily diverse viruses and target a variety of stages in the viral replication life cycle.

1.2 The TRIM22 protein

TRIM22 is an evolutionarily ancient IFN-induced antiviral protein that inhibits HIV-1, HBV, IAV, and encephalomyocarditis virus (EMCV) replication. Similar to other host restriction factors, the major biological activity of TRIM22 is its antiviral activity; However, altered TRIM22 expression levels have also been associated with multiple sclerosis, several cancers, and a number of autoimmune diseases. While TRIM22 is clearly an important and dynamic protein, the key factors that influence its expression and antiviral activity remain largely unknown [170]. An overview of what is currently known about TRIM22 evolution, structure, and function is provided below.

1.2.1 Origins and evolution

Human *TRIM22* is located on chromosome 11 within a cluster of closely-related *TRIM* genes that also includes *TRIM5*, *TRIM6*, and *TRIM34* [171,172]. The origins of the entire *TRIM5/6/22/34* gene cluster can be traced back to the Cretaceous period, or more specifically, to approximately 90-180 million years ago (Fig 1.5). Studies have shown that *TRIM5/6/22/34* is absent in Metatherian (marsupial) mammals (e.g. opossum and chicken), but present in all major Eutherian (placental) groups (e.g. cow, dog, and human) [172]. As such, the *TRIM5/6/22/34* gene cluster must have emerged after the divergence of Metatherian and Eutherian mammals, but before the separation of major Eutherian groups. Of interest, several groups have proposed that *TRIM5/6/22/34* likely arose through tandem gene duplication, as *TRIM5*, *TRIM6*, *TRIM22,* and *TRIM34* are close human paralogs, and because major gene re-arrangements have been observed in this chromosomal region [108,172,173]. Gene duplication plays an important role in evolution and a number of *TRIM* genes have been shown to undergo gene duplication in both primates and teleost fish [108,174].

Within the *TRIM5/6/22/34* gene cluster, *TRIM22* and *TRIM5* have an interesting and dynamic evolutionary relationship. In some Eutherian groups, such as cow, there are multiple copies of the *TRIM5* gene but no *TRIM22* gene. However, in other Eutherian groups, such as dog, the *TRIM22* gene is present but the *TRIM5* gene is not [172]. In addition, *TRIM22* and *TRIM5* have evolved in a mutually exclusive manner, whereby positive selection has acted on either *TRIM22* or *TRIM5* (but not both) in different primate lineages. This striking anti-correlative pattern of evolution is thought to occur due to tight genetic linkage between the two genes [172]. Both *TRIM22* and *TRIM5* are classified as Subgroup G (also Group 2) *TRIM* genes according to the most recent phylogenetic-based classification system (*TRIM6* and *TRIM34* are also Subgroup G, Group 2 *TRIM* genes) (Fig 1.3) [110,111].

Figure 1.5: TRIM22 evolution, genomic organization, and protein structure.

An approximate timeline of Metatherian and Eutherian mammalian evolution is shown on the top panel. The estimated time (~90-180 million years ago) that *TRIM22* (and the rest of the *TRIM5/6/22/34* gene cluster) emerged in Eutherian mammals is indicated by two arrows. The middle panel shows the genomic organization of the *TRIM5/6/22/34* gene cluster on human chromosome 11. The bottom panel illustrates the TRIM22 protein domains (RING, BB2, CC, and PRY/SPRY or B30.2) and the approximate location of its E3 ligase activity and nuclear localization signal (NLS).

Figure 1.5

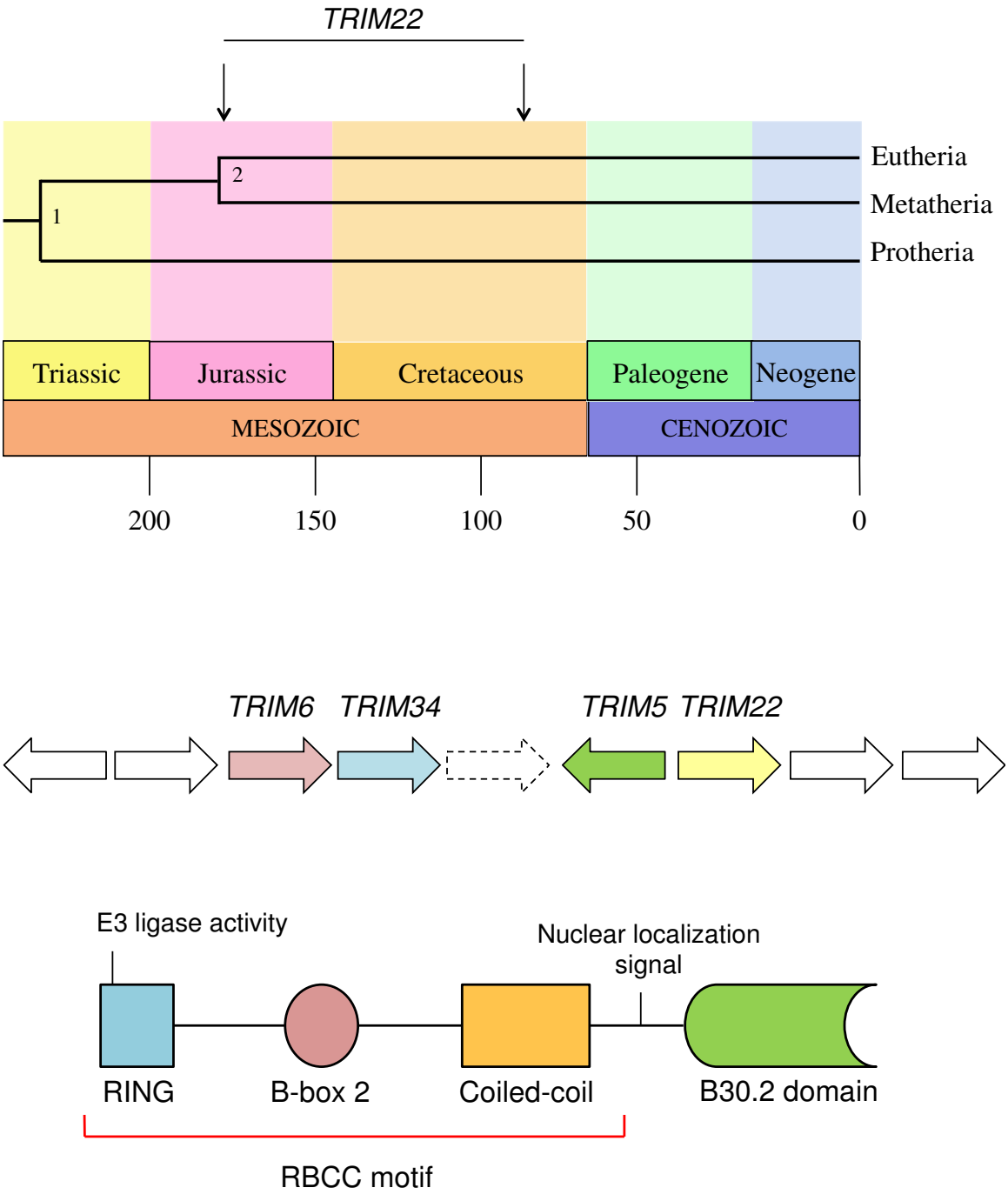


Figure 1.5: TRIM22 evolution, genomic organization, and protein structure.

1.2.2 Protein structure

Similar to other Subgroup G (and Group 2) members, the TRIM22 protein is comprised of an N-terminal RBCC motif that includes RING, BB2, and CC domains and a C-terminal B30.2 domain (Figs. 1.3 and 1.5). The RING domain of TRIM22 has been shown to possess both E3 ubiquitin and E3 SUMO ligase activity [132,175]. TRIM22's E3 ubiquitin ligase activity is dependent on two catalytic cysteine (Cys) residues (i.e. Cys15 and Cys18) in its RING domain. These Cys residues help stabilize the zinc finger motif and facilitate the transfer of ubiquitin to the appropriate substrate protein [113,117]. TRIM22 also mediates its own ubiquitylation and 26S proteasomal degradation when combined with the E2 conjugating enzyme UbcH5B (also referred to as UBC2D2) [175]. TRIM22 inhibits viral replication by both E3 ubiquitin ligase dependent and independent mechanisms. For example, TRIM22's E3 ubiquitin ligase activity is required for TRIM22-mediated inhibition of HBV, IAV, and EMCV [164,176,177]. However, HIV-1 inhibition occurs in both the presence and absence of TRIM22's E3 ubiquitin ligase activity [178,179].

The second domain in the TRIM22 protein is the BB2 domain. Similar to the RING domain, the BB2 domain contains a zinc finger motif; however, the BB2 domain only coordinates one zinc ion (the RING domain coordinates two zinc ions) [180]. While the function of TRIM22's BB2 domain is still unclear, BB2 mutations in other TRIM proteins have been shown to affect viral recognition or inhibition. For example, in the TRIM5α protein, the RING and BB2 domains work together to promote TRIM5α dimerization, which is important for higher-order self-association of rhTRIM5α on the HIV-1 capsid lattice and subsequent HIV-1 restriction [181]. It is unknown whether the TRIM22 BB2 domain is required for higher-order self-association; however, the BB2 does play a role in TRIM22's nuclear localization [182]. The BB2 domain of TRIM22 is followed by the predicted CC region, which contains many putative hyper-secondary structures and α-helices [183]. In TRIM5α, the CC region promotes TRIM5α trimerization and may be involved in TRIM5α cytoplasmic body formation. Similar to dimerization, TRIM5α trimerization drives its interaction with the HIV-1 capsid lattice [146,181]. The CC region also promotes higher-order self-association in a number of additional TRIM

proteins. While TRIM22 has been shown to undergo trimerization, the biological and antiviral significance of TRIM22 trimerization is unknown [112].

The TRIM22 protein contains a C-terminal B30.2 (PRY/SPRY) domain. Although the structure of this domain has not been resolved for TRIM22, several domain-deletion studies have shown that it is integral for many TRIM22 functions. For example, one study showed that B30.2 mutants of TRIM22 were no longer able to inhibit HBV replication and localized exclusively to the cytoplasm in HepG2 cells [164]. Several additional studies have also reported the TRIM22's B30.2 domain is necessary for its nuclear localization and the formation of nuclear bodies (NB), including a study that showed that amino acids 491-494 are critical for nuclear localization and amino acids 493-494 are critical for NB formation [184,185]. TRIM22's B30.2 domain may also be required for EMCV restriction, self-ubiquitylation, and/or the ubiquitylation of other substrate proteins. One study that investigated TRIM22-mediated EMCV restriction showed that a C-terminal TRIM22 mutant lacking both CC and B30.2 domains was unable to inhibit EMCV replication or facilitate ubiquitylation of target proteins even though its RING domain was still intact and bound to the ubiquitylated E2 enzyme [176]. An N-terminal TRIM22 mutant lacking both RING and BB2 domains lost the ability to undergo self-ubiquitylation, suggesting that the B30.2 domain of TRIM22 may be ubiquitylated by its own RING finger domain [176]. While the role of TRIM22's B30.2 domain in HIV-1 restriction is still unclear, rhTRIM5α's B30.2 domain is required for HIV-1 restriction [146]. Moreover, several hyper-variable regions (v1-v4) in the B30.2 domain of rhTRIM5α form the binding surface for HIV-1 capsid and confer virus specificity for TRIM5α-mediated restriction of retroviruses [186]. The B30.2 domain of TRIM22 also contains these hyper-variable regions (v1-v4); however, it is unknown whether they play a role in its antiviral activity or specificity.

1.2.3 Induction and expression

Several reports have demonstrated that TRIM22 is basally expressed in multiple human tissues and is highly upregulated in response to type I and type II IFN (Table 1.2)

Table 1.2

Table 1.2: Factors that alter TRIM22 protein expression levels.

	Change	**Cell type and/or Tissue** [1]
Cytokines		
IFN-α	Increase	Primary MDM; CEM, Jurkat, THP-1, H9, HepG2, U937 U-937-4, Daudi, and HeLa cells
IFN-β	Increase	HOS, Daudi, and HeLa cells
IFN-	Increase	HeLa, HepG2, and MCF7 cells
IL-1-β	Increase	Coronary artery endothelium
IL-2	Increase	CD4+, CD8+, NK cells
IL-15	Increase	CD4+, CD8+, NK cells
Progesterone	Increase	ABC28 and T47D cells
TNF-α	Increase	Coronary artery endothelium
Antigens/Infection		
EBV infection	Increase	BL41-EBV cells
EBV LMP-1	Increase	DG75 cells
HBV infection	Increase	Liver tissue
HCV infection	Increase	Liver tissue
HIV-1 infection	Increase	Immature DCs, Primary MDMs
HIV-1 Tat	Increase	Immature DCs
HPV infection	Decrease	Human keratinocytes
KSHV infection	Increase	KSHV lesion
KSHV LANA	Increase	BJAB cells

LPS	Increase	Primary MDMs
Rubella infection	Increase	ECV304 cells
Activation/Differentiation/Cell cycle		
1α,25-dihydroxyvitamine D33	Increase	Primary MDMs
Anti-CD2	Increase	Primary T cells
Anti-CD2/CD28	Decrease	Primary T cells
Anti-CD2/CD28/CD3	Decrease	CD4+, CD8+, NK cells
All-trans retinoic acid	Increase	Primary MDMs; HL60 and NB4 cells
p53	Increase	K562 and U-937-4 cells
p73	Increase	U-937-4 cells
Pioglitazone	Increase	Primary MDMs
UV-irradiation	Increase	MCF-7 cells
Disease		
Systemic Lupus Erythematosus	Increase	CD4+ T cells from SLE patient
Wilms Tumor	Decrease	Primary Wilms Tumor
Neuroblastoma	Decrease	Primary Neuroblastoma
Breast Cancer	Decrease	Primary Breast cancer and 10 Breast cancer cell lines

[1] Please refer to [170] for comprehensive list of all references

Abbreviations: Monocyte-derived macrophages (MDM), Dendritic cells (DC), Epstein-Barr virus (EBV), Human papillomavirus (HPV), Kaposi's sarcoma-associated herpesvirus (KSHV)

[164,172,178,179,187–193]. TRIM22 expression is also induced by several viral antigens (e.g. Epstein-Barr virus [EBV], HIV-1, HBV, Kaposi's sarcoma-associated herpes virus [KSHV], and Rubella), cytokines, and hormones. Studies have shown that TRIM22 contains two IFN-stimulated response elements (ISRE-1, ISRE-2) and one IFN-γ activation site (GAS) in its 5' promoter region. Notably, ISRE-1 and GAS are not required for IFN-γ induction of TRIM22. Instead, induction of TRIM22 by IFN-γ requires ISRE-2 plus six upstream nucleotides (referred to as the 5' extended ISRE or eISRE) [190]. This induction is dependent on the chromatin remodeling enzyme brahma-related gene 1, which recruits IFN regulatory factor 1 (IRF-1) to the eISRE, and histone deacetylase activity, which prevents the proteasomal degradation of IRF-1 [192,194]. JAK, phosphatidylcholine-phospholipase C, and protein kinase C are also required for induction of TRIM22 by IFN-γ. p300 enhances IFN-γ induced expression of TRIM22 and is also required for the recruitment of RNA polymerase II to the 5' TRIM22 promoter. IRF-1 binding to eISRE appears to also be required for IFN-α induced and basal TRIM22 expression [195].

1.2.4 Sub-cellular localization

Conflicting reports have been published on the sub-cellular localization of TRIM22. Some reports show that TRIM22 localizes predominantly in the cytoplasm or in the nucleus, whereas others show that TRIM22 localizes in both the cytoplasm and the nucleus [112,164,175,182,184,196–198]. These studies also report different TRIM22 localization patterns (i.e. diffuse, speckled, and/or aggregated). Although the reasons for these inconsistencies are unclear, several possible explanations have been given in the literature. These include whether TRIM22 was expressed endogenously (e.g. IFN treatment of cells) or exogenously (e.g. overexpression of TRIM22), the method of fixation, and the epitope tag. Other potential explanations include cell-type specific factors or induction by different cellular signals. Consistent with the latter, TRIM22 localizes in NBs in progesterone-treated, but not IFN-γ-treated, MCF7 and HeLa cell lines [197].

Several determinants of TRIM22 sub-cellular localization have been identified. The TRIM22 protein contains a bipartite nuclear localization signal (NLS) in the Spacer 2 domain (SP2), which was previously shown to be necessary, but not sufficient, for nuclear localization [197]. Although the B30.2 domain does not contain a known NLS, multiple groups have shown that it is required for nuclear localization [164,184,196,197]. One group in particular showed that the amino acids 491-494 are essential for TRIM22 nuclear localization and that the amino acids 493-494 are critical for the formation of TRIM22 NBs in MCF7 cells [197]. Another group reported that the amino acids 395, 396, and 400 are required for the cytoplasmic localization of TRIM22 in COS-7 (African green monkey) cells [196]. However, this group did not investigate how these amino acids influence TRIM22 localization in human cells. Notably, some studies that observed cytoplasmic TRIM22 localization used a shorter form of TRIM22 (442 instead of 498 amino acids) that is translated from a 1329 mRNA coding sequence. As such, in these studies the lack of amino acids 491-494 likely contribute to the cytoplasmic localization of TRIM22.

In many cell types, TRIM22 localizes in punctate NBs, which have been shown to partially co-localize with Cajal bodies (CB) [197]. CBs are distinct nuclear organelles that are involved in RNA processing and modification, assembly/modification of splicing machinery, and cell cycle progression [199]. TRIM22 interacts with p80-coilin, which is a major component of CBs, in stably-transfected MCF7 cells and progesterone-treated ABC28 cells [197]. Similar to CBs, TRIM22 localization may change throughout the cell cycle. One report showed that during G0/G1 TRIM22 localizes in NBs, in S-phase it localizes diffusely in the nucleus, and in mitosis it assumes a diffuse pattern in the nucleus and cytoplasm (exogenous TRIM22 expression in HeLa cells) [197]. However, a second study reported that TRIM22 co-localizes with the centrosome independently of the cell cycle (exogenous TRIM22 expression in U2OS cells as well as endogenous TRIM22 expression in peripheral blood mononuclear cells) [198]. Notably, these two localization patterns may not be mutually exclusive because TRIM22 did not localize exclusively to centrosomes. These data suggest that multiple factors influence the sub-cellular localization of TRIM22, indicating that TRIM22 may have several biological roles.

1.2.5 Antiviral function

Human TRIM22 was first identified in 1995 during a search for IFN-induced genes in Daudi cells. Following sequence analysis, which revealed that TRIM22 was highly homologous to the mouse *Rpt-1* gene, exogenous TRIM22 expression was shown to downregulate transcription from the HIV-1 LTR [187,200]. Although this experiment was performed using a luciferase reporter gene instead of the entire HIV-1 genome, it was first to report an antiviral (and anti-HIV-1) function for TRIM22. In 2006, another independent study showed that TRIM22 was highly up-regulated in primary monocyte-derived macrophages (MDM) in response to HIV-1 infection, IFN-α treatment, or stimulation with lipopolysaccharide (LPS). In addition, they showed that exogenous TRIM22 expression inhibited HIV-1 infection by 50-90% in 293T cells and primary MDMs [189]. In 2008, Barr and colleagues showed that TRIM22 was a key mediator of type I IFN-induced inhibition of HIV-1 replication [178]. Two different methods of HIV-1 inhibition were observed. In HOS and HeLa cell lines, TRIM22 expression inhibited HIV-1 particle production by preventing the trafficking of the Gag polyprotein to the plasma membrane. This effect was dependent on the E3 ligase activity of TRIM22. Because TRIM22 was also shown to interact with Gag it was thought that TRIM22-mediated post-translational modification of Gag may be responsible for altered Gag trafficking. However, to date, TRIM22 has never been shown to modify Gag post-translationally. Unlike in HOS and HeLa cells, in U2OS and 143B cells, TRIM22 inhibited the accumulation of intracellular Gag protein. Although the mechanism of restriction was not identified in these cell lines, several potential explanations were suggested, including inhibition of LTR-driven transcription or degradation of the Gag RNA and/or polyprotein [178]. Notably, these experiments provided the first mechanistic data for restriction of HIV-1 replication by TRIM22.

It has since been confirmed that TRIM22 can restrict HIV-1 transcription [179]. In 2011, TRIM22 was identified as the sole factor expressed in clones of U937 cells that were non-permissive to HIV-1 replication, but absent in permissive U937 clones. Using a luciferase reporter plasmid, the authors tested LTR-driven transcription in the non-

permissive and permissive clones. LTR transcription levels in non-permissive clones were decreased 7-10 fold compared to permissive clones; however, non-permissive transcription levels were increased when shRNA was used to knockdown TRIM22. Moreover, when TRIM22 was expressed in permissive clones transcription levels decreased to those observed in non-permissive clones. Further examination of these clones revealed that TRIM22 inhibited basal LTR-driven HIV-1 transcription and that these effects were independent of NF-κB binding sites in the LTR, Tat-mediated LTR transactivation, and TRIM22 E3 ligase activity [179].

In addition to these *in vitro* studies, there is also *in vivo* evidence to support a role for TRIM22 as an anti-HIV-1 effector. A 2011 study demonstrated that increased TRIM22 expression was associated with significantly lower viral loads and significantly higher CD4+ T-cell counts in HIV-1 positive individuals in the primary phase of infection [201]. Recently, a follow-up study showed that TRIM22 expression was also associated with significantly lower viral loads in HIV-1 positive individuals in the chronic phase of infection [202]. These data suggest that TRIM22 expression contributes to HIV-1 disease progression in infected individuals.

TRIM22's antiviral activity is not limited to HIV-1. Exogenous expression of TRIM22 has also been shown to inhibit the replication of EMCV, IAV, and HBV. For EMCV restriction, TRIM22 was shown to interact with the 3C viral protease and mediate its ubiquitylation and subsequent degradation [176]. The 3C protease has a number of important roles, including processing the viral polyprotein, and is essential for successful EMCV replication[176]. A similar mechanism of restriction was reported for IAV inhibition. Specifically, TRIM22 was shown to interact with IAV's viral nucleoprotein (NP), which plays a critical structural function and is required for IAV replication. Following TRIM22-NP interaction, TRIM22's E3 ubiquitin ligase activity facilitates ubiquitylation and subsequent proteasomal degradation of NP [177]. Several studies have recently shown that TRIM22 may be an important antiviral factor in the liver. In 2009, one study demonstrated that TRIM22 expression inhibited HBV gene expression and replication in cultured cells and mice [164]. Further experiments revealed that TRIM22 inhibited the activity of the HBV core promoter (CP), which plays a key role in HBV

replication. HBV CP inhibition was dependent on TRIM22's B30.2 domain and its E3 ligase activity, even though a ubiquitylation target was not identified [164]. Interestingly, TRIM22 expression is also significantly upregulated during clearance of HBV and hepatitis C virus (HCV) in chimpanzees [203,204]. In human HCV infection, TRIM22 is significantly upregulated in the cirrhotic liver of HCV positive individuals and individuals with mild chronic HCV with no fibrosis [205].

1.2.6 Other functions

Several reports have implicated TRIM22 in other biological processes, including cell cycle regulation and cell proliferation/differentiation. One study identified a functional p53 response element in intron 1 of the *TRIM22* gene and demonstrated that upon p53 binding this element activates TRIM22 expression. Moreover, the same study showed that overexpression of TRIM22 in U937 cells led to decreased clonogenic growth and that endogenous TRIM22 was upregulated during induced differentiation in NB4 cells [188]. A later study investigating TRIM22 expression during hematopoietic differentiation showed that TRIM22 is highly expressed in $CD34^+$ bone marrow progenitor cells, but declines in mature populations. Notably, although TRIM22 expression was inversely correlated with differentiation in both lineages, its expression pattern differed during erythroid versus granulocytic differentiation. Decreased TRIM22 expression was more pronounced and lasting during erythroid differentiation and undetectable in nucleated erythroid populations [193].

TRIM22 has also been linked to other human diseases, including certain cancers and autoimmune diseases. Two studies have shown that decreased TRIM22 expression is associated with increased progression, relapse, and mortality in cases of Wilms tumor [206,207]. In addition, a recent study reported that TRIM22 expression is significantly decreased in both a panel of 10 breast cancer cell lines and breast tumors compared to three non-malignant cell lines and normal breast tissue. Interestingly, TRIM22 protein levels correlated with p53 protein levels in the normal breast tissue, but not in breast cancer tissue. Additional experiments showed that TRIM22 expression was no longer

p53-inducible in breast cancer cells, suggesting that a defect in p53-mediated TRIM22 regulation was responsible for TRIM22 downregulation in breast cancer [208]. Two gene profiling studies have also implicated TRIM22 in the pathogenesis of systemic lupus erythematosus (SLE). The first study showed that TRIM22 was overexpressed in the $CD4^+$ T cells of individuals with active SLE compared to individuals with non-active SLE. The second study also reported that TRIM22 was upregulated in SLE-positive $CD4^+$ T cells and that, along with a number of other IFN-induced genes, TRIM22 was significantly hypomethylated in SLE cells compared to healthy controls [209,210].

1.3 Rationale and Experimental Approach

Genetic variation in immune genes plays an integral role in host susceptibility to and progression of infection and disease. Much of this variation is due to single nucleotide polymorphisms (SNPs), which are defined as single base changes in a DNA sequence. While many SNPs are phenotypically neutral, non-synonymous or amino acid altering SNPs (nsSNP) often have deleterious effects on protein structure and/or function [211–218]. Previous studies have shown that nsSNPs do not appear randomly in the genome, but emerge based on genomic location and selective pressures. Since innate immune genes are located at the interface of the microbial environment, they tend to be exposed to a wide range of selective pressures. As such, innate immune genes often contain genetic signatures of positive (directional) and/or balancing selection, whereas most other host genes are dominated by negative (purifying) selection [219–221].

Multiple host restriction factors, including APOBEC3G, TRIM5α, BST-2/tetherin, and SAMHD1, have evolved under positive and/or balancing selection [18]. Host restriction factors often interact directly with viral antagonists and/or other viral proteins and as a result, they tend to be subjected to intense pathogenic pressures. In order to effectively adapt to an ever-changing pathogenic environment, host restriction factors must evolve new advantageous mutations and/or maintain genetic flexibility to counteract emerging pathogenic threats. Since new beneficial mutations and/or sites of genetic diversity are typically located at critical functional sites, performing evolutionary analyses on host restriction factors can be a powerful way to delineate integral sites in these proteins. It

follows that a number of studies on host restriction factors have used an evolutionary approach to identify essential residues in these proteins [62,85,91,99,106,222]. In addition, other studies have used *in silico* nsSNP prediction programs to identify functionally and/or clinically relevant polymorphisms in innate immune genes [19–24]. Further, *in silico* methods have also been used to identify important functional regions in these genes (e.g. post-translational modifications) and support evolutionary analyses [229–232]. One of the major advantages of performing *in silico* analyses is that it allows for the systematic analysis and prioritization of specific functional sites. This is becoming increasingly important given the tremendous number of SNPs in the human genome and the vast amount of genetic data that is generated on a daily basis. Collectively, evolutionary and *in silico* analyses provide valuable insight into protein function and are powerful tools that can help identify key structural and/or functional sites in a protein.

Given the important role played by TRIM22 in multiple biological processes, including the host antiviral response, and the paucity of information about key functional sites in the TRIM22 protein, we conducted an extensive evolutionary and *in silico* analysis to identify critical amino acid residues that mediate TRIM22 function. This approach has previously been used to pinpoint specific amino acid residues that are essential for the activities of other host restriction factors, including the APOBEC3G, TRIM5α, BST-2/tetherin, and SAMHD1 proteins. We hypothesized that evolutionary forces have been acting on *TRIM22* and have selected for specific amino acids that impact TRIM22 function. To address this hypothesis, we characterized the evolutionary forces acting on *TRIM22* and used a number of *in silico* methods to delineate and prioritize potential functional sites in human TRIM22 that may be relevant to its overall antiviral and/or biological functions. We then investigated how various *TRIM22* nsSNPs may affect human infection and disease and identified a functional nsSNP (rs1063303:G>C) that influences two diverse TRIM22-mediated biological activities.

1.4 References

1. Hoebe, K., Janssen, E. & Beutler, B. The interface between innate and adaptive immunity. *Nat. Immunol.* **5,** 971–4 (2004).

2. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124,** 783–801 (2006).

3. Brennan, K. & Bowie, A. G. Activation of host pattern recognition receptors by viruses. *Curr. Opin. Microbiol.* **13,** 503–7 (2010).

4. Medzhitov, R. & Janeway, C. A. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91,** 295–8 (1997).

5. Janeway, C. A. & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20,** 197–216 (2002).

6. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.* **5,** 730–7 (2004).

7. Thompson, M. R., Kaminski, J. J., Kurt-Jones, E. A. & Fitzgerald, K. A. Pattern recognition receptors and the innate immune response to viral infection. *Viruses* **3,** 920–40 (2011).

8. Lester, S. N. & Li, K. Toll-Like Receptors in Antiviral Innate Immunity. *J. Mol. Biol.* **426,** 1246–1264 (2014).

9. Kawai, T. & Akira, S. Antiviral signaling through pattern recognition receptors. *J. Biochem.* **141,** 137–45 (2007).

10. Levy, D. E., Marié, I. J. & Durbin, J. E. Induction and function of type I and III interferon in response to viral infection. *Curr. Opin. Virol.* **1,** 476–86 (2011).

11. Platanias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5,** 375–86 (2005).

12. Boehm, U., Klamp, T., Groot, M. & Howard, J. C. Cellular responses to interferon-gamma. *Annu. Rev. Immunol.* **15,** 749–95 (1997).

13. Randall, R. E. & Goodbourn, S. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J. Gen. Virol.* **89,** 1–47 (2008).

14. Stetson, D. B. & Medzhitov, R. Type I interferons in host defense. *Immunity* **25,** 373–81 (2006).

15. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annu. Rev. Immunol.* **32,** 140220103029009 (2013).

16. Pincus, T., Hartley, J. W. & Rowe, W. P. A major genetic locus affecting resistance to infection with murine leukemia viruses. Tissue culture studies of naturally occurring viruses. *J. Exp. Med.* **133,** 1219–33 (1971).

17. Liu, L. *et al.* A whole genome screen for HIV restriction factors. *Retrovirology* **8,** 94 (2011).

18. Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12,** 687–95 (2012).

19. Zheng, Y.-H., Jeang, K.-T. & Tokunaga, K. Host restriction factors in retroviral infection: promises in virus-host interaction. *Retrovirology* **9,** 112 (2012).

20. Meylan, P. R., Guatelli, J. C., Munis, J. R., Richman, D. D. & Kornbluth, R. S. Mechanisms for the inhibition of HIV replication by interferons-alpha, -beta, and -gamma in primary human macrophages. *Virology* **193,** 138–48 (1993).

21. Poli, G., Orenstein, J. M., Kinter, A., Folks, T. M. & Fauci, A. S. Interferon-alpha but not AZT suppresses HIV expression in chronically infected cell lines. *Science* **244,** 575–7 (1989).

22. Agy, M. B., Acker, R. L., Sherbert, C. H. & Katze, M. G. Interferon treatment inhibits virus replication in HIV-1- and SIV-infected CD4+ T-cell lines by distinct mechanisms: evidence for decreased stability and aberrant processing of HIV-1 proteins. *Virology* **214,** 379–86 (1995).

23. Fernie, B. F., Poli, G. & Fauci, A. S. Alpha interferon suppresses virion but not soluble human immunodeficiency virus antigen production in chronically infected T-lymphocytic cells. *J. Virol.* **65,** 3968–71 (1991).

24. Coccia, E. M., Krust, B. & Hovanessian, A. G. Specific inhibition of viral protein synthesis in HIV-infected cells in response to interferon treatment. *J. Biol. Chem.* **269,** 23087–94 (1994).

25. Shirazi, Y. & Pitha, P. M. Alpha interferon inhibits early stages of the human immunodeficiency virus type 1 replication cycle. *J. Virol.* **66,** 1321–8 (1992).

26. Malim, M. H. & Emerman, M. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* **3,** 388–98 (2008).

27. Schaller, T., Bauby, H., Hué, S., Malim, M. H. & Goujon, C. New insights into an X-traordinary viral protein. *Front. Microbiol.* **5,** 126 (2014).

28. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–50 (2002).

29. Zhang, H. *et al.* The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424,** 94–8 (2003).

30. Harris, R. S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113,** 803–9 (2003).

31. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424,** 99–103 (2003).

32. Yu, Q. *et al.* Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.* **11,** 435–42 (2004).

33. Marin, M., Rose, K. M., Kozak, S. L. & Kabat, D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat. Med.* **9,** 1398–403 (2003).

34. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302,** 1056–60 (2003).

35. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.* **9,** 1404–7 (2003).

36. Pace, C. *et al.* Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. *J. Virol.* **80,** 9259–69 (2006).

37. Land, A. M. *et al.* Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. *J. Virol.* **82,** 8172–82 (2008).

38. Kaur, A. *et al.* Diverse host responses and outcomes following simian immunodeficiency virus SIVmac239 infection in sooty mangabeys and rhesus macaques. *J. Virol.* **72,** 9597–611 (1998).

39. Adamson, C. S. & Jones, I. M. The molecular basis of HIV capsid assembly--five years of progress. *Rev. Med. Virol.* **14,** 107–21

40. Stremlau, M., Owens, C. M. & Perron, M. J. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. **427,** (2004).

41. Sebastian, S. & Luban, J. TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. *Retrovirology* **2,** 40 (2005).

42. Stremlau, M. *et al.* Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5514–9 (2006).

43. Diaz-Griffero, F. *et al.* Rapid turnover and polyubiquitylation of the retroviral restriction factor TRIM5. *Virology* **349,** 300–15 (2006).

44. Yamauchi, K., Wada, K., Tanji, K., Tanaka, M. & Kamitani, T. Ubiquitination of E3 ubiquitin ligase TRIM5 alpha and its potential role. *FEBS J.* **275,** 1540–55 (2008).

45. Lukic, Z. *et al.* TRIM5α associates with proteasomal subunits in cells while in complex with HIV-1 virions. *Retrovirology* **8,** 93 (2011).

46. Rold, C. J. & Aiken, C. Proteasomal degradation of TRIM5alpha during retrovirus restriction. *PLoS Pathog.* **4,** e1000074 (2008).

47. Wu, X., Anderson, J. L., Campbell, E. M., Joseph, A. M. & Hope, T. J. Proteasome inhibitors uncouple rhesus TRIM5alpha restriction of HIV-1 reverse transcription and infection. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7465–70 (2006).

48. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–30 (2008).

49. Yang, H. *et al.* Structural insight into the mechanisms of enveloped virus tethering by tetherin. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18428–32 (2010).

50. Mitchell, R. S. *et al.* Vpu antagonizes BST-2-mediated restriction of HIV-1 release via beta-TrCP and endo-lysosomal trafficking. *PLoS Pathog.* **5,** e1000450 (2009).

51. Douglas, J. L. *et al.* Vpu directs the degradation of the human immunodeficiency virus restriction factor BST-2/Tetherin via a {beta}TrCP-dependent mechanism. *J. Virol.* **83,** 7931–47 (2009).

52. Miyakawa, K. *et al.* BCA2/Rabring7 promotes tetherin-dependent HIV-1 restriction. *PLoS Pathog.* **5,** e1000700 (2009).

53. Neil, S. J. D., Sandrin, V., Sundquist, W. I. & Bieniasz, P. D. An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. *Cell Host Microbe* **2,** 193–203 (2007).

54. Perez-Caballero, D. *et al.* Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* **139,** 499–511 (2009).

55. Swiecki, M. *et al.* Structural and biophysical analysis of BST-2/tetherin ectodomains reveals an evolutionary conserved design to inhibit virus release. *J. Biol. Chem.* **286,** 2987–97 (2011).

56. Hinz, A. *et al.* Structural basis of HIV-1 tethering to membranes by the BST-2/tetherin ectodomain. *Cell Host Microbe* **7,** 314–23 (2010).

57. Malim, M. H. & Bieniasz, P. D. HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb. Perspect. Med.* **2,** a006940 (2012).

58. Descours, B. *et al.* SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4(+) T-cells. *Retrovirology* **9,** 87 (2012).

59. Baldauf, H.-M. *et al.* SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nat. Med.* **18,** 1682–7 (2012).

60. Lahouassa, H. *et al.* SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.* **13,** 223–8 (2012).

61. Lim, E. S. *et al.* The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe* **11,** 194–204 (2012).

62. Laguette, N. *et al.* Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host Microbe* **11,** 205–17 (2012).

63. Fregoso, O. I. *et al.* Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the restriction factor SAMHD1. *PLoS Pathog.* **9,** e1003496 (2013).

64. Pertel, T. *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472,** 361–365 (2011).

65. Hotter, D., Sauter, D. & Kirchhoff, F. Emerging role of the host restriction factor tetherin in viral immune sensing. *J. Mol. Biol.* **425,** 4956–64 (2013).

66. Sze, A., Olagnier, D., Lin, R., van Grevenynghe, J. & Hiscott, J. SAMHD1 host restriction factor: a link with innate immune sensing of retrovirus infection. *J. Mol. Biol.* **425,** 4981–94 (2013).

67. Nisole, S., Stoye, J. P. & Saïb, A. TRIM family proteins: retroviral restriction and antiviral defence. *Nat. Rev. Microbiol.* **3,** 799–808 (2005).

68. Ozato, K., Shin, D.-M., Chang, T.-H. & Morse, H. C. TRIM family proteins and their emerging roles in innate immunity. *Nat. Rev. Immunol.* **8,** 849–60 (2008).

69. Uchil, P. D., Quinlan, B. D., Chan, W.-T., Luna, J. M. & Mothes, W. TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathog.* **4,** e16 (2008).

70. Carthagena, L. *et al.* Human TRIM gene expression in response to interferons. *PLoS One* **4,** e4894 (2009).

71. Kawai, T. & Akira, S. Regulation of innate immune signalling pathways by the tripartite motif (TRIM) family proteins. *EMBO Mol. Med.* **3,** 513–27 (2011).

72. McNab, F. W., Rajsbaum, R., Stoye, J. P. & O'Garra, A. Tripartite-motif proteins and innate immune regulation. *Curr. Opin. Immunol.* **23,** 46–56 (2011).

73. Hatakeyama, S. TRIM proteins and cancer. *Nat. Rev. Cancer* **11,** 792–804 (2011).

74. Petrera, F. & Meroni, G. TRIM proteins in development. *Adv. Exp. Med. Biol.* **770,** 131–41 (2012).

75. Jefferies, C., Wynne, C. & Higgs, R. Antiviral TRIMs: friend or foe in autoimmune and autoinflammatory disease? *Nat. Rev. Immunol.* **11,** 617–25 (2011).

76. Horie-Inoue, K. TRIM proteins as trim tabs for the homoeostasis. *J. Biochem.* **154,** 309–312 (2013).

77. Wolf, D. & Goff, S. P. Host restriction factors blocking retroviral replication. *Annu. Rev. Genet.* **42,** 143–63 (2008).

78. Yan, Y., Buckler-White, A., Wollenberg, K. & Kozak, C. A. Origin, antiviral function and evidence for positive selection of the gammaretrovirus restriction gene Fv1 in the genus Mus. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 3259–63 (2009).

79. Haller, O. & Kochs, G. Human MxA protein: an interferon-induced dynamin-like GTPase with broad antiviral activity. *J. Interferon Cytokine Res.* **31,** 79–87 (2011).

80. Mitchell, P. S. *et al.* Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe* **12,** 598–604 (2012).

81. Haller, O. Dynamins are forever: MxB inhibits HIV-1. *Cell Host Microbe* **14,** 371–3 (2013).

82. Smith, S., Weston, S., Kellam, P. & Marsh, M. IFITM proteins-cellular inhibitors of viral entry. *Curr. Opin. Virol.* **4C,** 71–77 (2014).

83. Diamond, M. S. & Farzan, M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat. Rev. Immunol.* **13,** 46–57 (2013).

84. Grütter, M. G. & Luban, J. TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. *Curr. Opin. Virol.* **2,** 142–50 (2012).

85. Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2832–7 (2005).

86. Lei, Y.-C. *et al.* Inhibition of hepatitis B virus replication by APOBEC3G in vitro and in vivo. *World J. Gastroenterol.* **12,** 4492–7 (2006).

87. Zhao, D. *et al.* APOBEC3G directly binds Hepatitis B virus core protein in cell and cell free systems. *Virus Res.* **151,** 213–9 (2010).

88. Jaguva Vasudevan, A. A. *et al.* Prototype foamy virus Bet impairs the dimerization and cytosolic solubility of human APOBEC3G. *J. Virol.* **87,** 9030–40 (2013).

89. Fehrholz, M. *et al.* The innate antiviral factor APOBEC3G targets replication of measles, mumps and respiratory syncytial viruses. *J. Gen. Virol.* **93,** 565–76 (2012).

90. Stavrou, S. *et al.* Murine leukemia virus glycosylated Gag blocks apolipoprotein B editing complex 3 and cytosolic sensor access to the reverse transcription complex. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 9078–83 (2013).

91. Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* **2,** E275 (2004).

92. Kim, E. T., White, T. E., Brandariz-Núñez, A., Diaz-Griffero, F. & Weitzman, M. D. SAMHD1 restricts herpes simplex virus 1 in macrophages by limiting DNA replication. *J. Virol.* **87,** 12949–56 (2013).

93. Hollenbaugh, J. A. *et al.* Host factor SAMHD1 restricts DNA viruses in non-dividing myeloid cells. *PLoS Pathog.* **9,** e1003481 (2013).

94. Laguette, N. *et al.* Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host Microbe* **11,** 205–17 (2012).

95. Blondeau, C. *et al.* Tetherin restricts herpes simplex virus 1 and is antagonized by glycoprotein M. *J. Virol.* **87,** 13124–33 (2013).

96. Jouvenet, N. *et al.* Broad-spectrum inhibition of retroviral and filoviral particle release by tetherin. *J. Virol.* **83,** 1837–44 (2009).

97. Serra-Moreno, R., Zimmermann, K., Stern, L. J. & Evans, D. T. Tetherin/BST-2 antagonism by Nef depends on a direct physical interaction between Nef and tetherin, and on clathrin-mediated endocytosis. *PLoS Pathog.* **9,** e1003487 (2013).

98. Yondola, M. A. *et al.* Budding capability of the influenza virus neuraminidase can be modulated by tetherin. *J. Virol.* **85,** 2480–91 (2011).

99. McNatt, M. W. *et al.* Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. *PLoS Pathog.* **5,** e1000300 (2009).

100. García, M. A., Meurs, E. F. & Esteban, M. The dsRNA protein kinase PKR: virus and cell control. *Biochimie* **89,** 799–811

101. Munir, M. & Berg, M. The multiple faces of proteinkinase R in antiviral defense. *Virulence* **4,** 85–9 (2013).

102. Pfaller, C. K., Li, Z., George, C. X. & Samuel, C. E. Protein kinase PKR and RNA adenosine deaminase ADAR1: new roles for old players as modulators of the interferon response. *Curr. Opin. Immunol.* **23,** 573–82 (2011).

103. Rothenburg, S., Seo, E. J., Gibbs, J. S., Dever, T. E. & Dittmar, K. Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. *Nat. Struct. Mol. Biol.* **16,** 63–70 (2009).

104. Woods, M. W. *et al.* Human HERC5 restricts an early stage of HIV-1 assembly by a mechanism correlating with the ISGylation of Gag. *Retrovirology* **8,** 95 (2011).

105. Tang, Y. *et al.* Herc5 attenuates influenza A virus by catalyzing ISGylation of viral NS1 protein. *J. Immunol.* **184,** 5777–90 (2010).

106. Woods, M. W. *et al.* Interferon-induced HERC5 is evolving under positive selection and inhibits HIV-1 particle production by a novel mechanism targeting Rev/RRE-dependent RNA nuclear export. *Retrovirology* **11,** 27 (2014).

107. George, C. X. *et al.* Tipping the balance: antagonism of PKR kinase and ADAR1 deaminase functions by virus gene products. *J. Interferon Cytokine Res.* **29,** 477–87 (2009).

108. Han, K., Lou, D. I. & Sawyer, S. L. Identification of a genomic reservoir for new TRIM genes in primate genomes. *PLoS Genet.* **7,** e1002388 (2011).

109. Short, K. M. & Cox, T. C. Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *J. Biol. Chem.* **281,** 8970–80 (2006).

110. Sardiello, M., Cairo, S., Fontanella, B., Ballabio, A. & Meroni, G. Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties. *BMC Evol. Biol.* **8,** 225 (2008).

111. Marín, I. Origin and diversification of TRIM ubiquitin ligases. *PLoS One* **7,** e50030 (2012).

112. Reymond, A. *et al.* The tripartite motif family identifies cell compartments. *EMBO J.* **20,** 2140–51 (2001).

113. Meroni, G. & Diez-Roux, G. TRIM/RBCC, a novel class of "single protein RING finger" E3 ubiquitin ligases. *Bioessays* **27,** 1147–57 (2005).

114. Saurin, A. J., Borden, K. L., Boddy, M. N. & Freemont, P. S. Does this have a familiar RING? *Trends Biochem. Sci.* **21,** 208–14 (1996).

115. Freemont, P. S. RING for destruction? *Curr. Biol.* **10,** R84–7 (2000).

116. Deshaies, R. J. & Joazeiro, C. a P. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78,** 399–434 (2009).

117. Lorick, K. L. *et al.* RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11364–9 (1999).

118. Borden, K. L. RING domains: master builders of molecular scaffolds? *J. Mol. Biol.* **295,** 1103–12 (2000).

119. Kentsis, A. & Borden, K. L. Construction of macromolecular assemblages in eukaryotic processes and their role in human disease: linking RINGs together. *Curr. Protein Pept. Sci.* **1,** 49–73 (2000).

120. Rajsbaum, R., García-Sastre, A. & Versteeg, G. A. TRIMmunity: The Roles of the TRIM E3-Ubiquitin Ligase Family in Innate Antiviral Immunity. *J. Mol. Biol.* **426,** 1265–1284 (2014).

121. Napolitano, L. M., Jaffray, E. G., Hay, R. T. & Meroni, G. Functional interactions between ubiquitin E2 enzymes and TRIM proteins. *Biochem. J.* **434,** 309–19 (2011).

122. Gack, M. U. *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446,** 916–920 (2007).

123. Uchil, P. D. *et al.* TRIM protein-mediated regulation of inflammatory and innate immune signaling and its association with antiretroviral activity. *J. Virol.* **87,** 257–72 (2013).

124. Versteeg, G. A. *et al.* The E3-ligase TRIM family of proteins regulates signaling pathways triggered by innate immune pattern-recognition receptors. *Immunity* **38,** 384–98 (2013).

125. Metzger, M. B., Hristova, V. A. & Weissman, A. M. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J. Cell Sci.* **125,** 531–7 (2012).

126. Zeng, W. *et al.* Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. *Cell* **141,** 315–30 (2010).

127. Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81,** 203–29 (2012).

128. Roa, A. *et al.* RING domain mutations uncouple TRIM5α restriction of HIV-1 from inhibition of reverse transcription and acceleration of uncoating. *J. Virol.* **86,** 1717–27 (2012).

129. Li, X. *et al.* Virus-specific effects of TRIM5α(rh) RING domain functions on restriction of retroviruses. *J. Virol.* **87,** 7234–45 (2013).

130. Lienlaf, M. *et al.* Contribution of E3-ubiquitin ligase activity to HIV-1 restriction by TRIM5alpha(rh): structure of the RING domain of TRIM5alpha. *J. Virol.* **85,** 8725–37 (2011).

131. Streich, F. C., Ronchi, V. P., Connick, J. P. & Haas, A. L. Tripartite motif ligases catalyze polyubiquitin chain formation through a cooperative allosteric mechanism. *J. Biol. Chem.* **288,** 8209–21 (2013).

132. Chu, Yaya and Yang, X. SUMO E3 ligase activity of TRIM proteins. *Oncogene* **30,** 1108–1116 (2011).

133. Liang, Q. *et al.* Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *J. Immunol.* **187,** 4754–63 (2011).

134. Zou, W. & Zhang, D.-E. The interferon-inducible ubiquitin-protein isopeptide ligase (E3) EFP also functions as an ISG15 E3 ligase. *J. Biol. Chem.* **281,** 3989–94 (2006).

135. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458,** 422–9 (2009).

136. Maroui, M. A. *et al.* Requirement of PML SUMO interacting motif for RNF4- or arsenic trioxide-induced degradation of nuclear PML isoforms. *PLoS One* **7,** e44949 (2012).

137. Zhong, S. *et al.* Role of SUMO-1-modified PML in nuclear body formation. *Blood* **95,** 2748–52 (2000).

138. Lukic, Z., Goff, S. P., Campbell, E. M. & Arriagada, G. Role of SUMO-1 and SUMO interacting motifs in rhesus TRIM5α-mediated restriction. *Retrovirology* **10,** 10 (2013).

139. Arriagada, G., Muntean, L. N. & Goff, S. P. SUMO-interacting motifs of human TRIM5α are important for antiviral activity. *PLoS Pathog.* **7,** e1002019 (2011).

140. Massiah, M. A., Simmons, B. N., Short, K. M. & Cox, T. C. Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. *J. Mol. Biol.* **358,** 532–45 (2006).

141. Massiah, M. A. *et al.* Solution structure of the MID1 B-box2 CHC(D/C)C(2)H(2) zinc-binding domain: insights into an evolutionarily conserved RING fold. *J. Mol. Biol.* **369,** 1–10 (2007).

142. Mrosek, M. *et al.* Structural analysis of B-Box 2 from MuRF1: identification of a novel self-association pattern in a RING-like fold. *Biochemistry* **47,** 10722–30 (2008).

143. Diaz-Griffero, F. *et al.* A B-box 2 surface patch important for TRIM5alpha self-association, capsid binding avidity, and retrovirus restriction. *J. Virol.* **83,** 10737–51 (2009).

144. Everett, R. D. & Chelbi-Alix, M. K. PML and PML nuclear bodies: implications in antiviral defence. *Biochimie* **89,** 819–30

145. Bernardi, R. & Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.* **8,** 1006–16 (2007).

146. Mische, C. C. *et al.* Retroviral restriction factor TRIM5alpha is a trimer. *J. Virol.* **79,** 14446–50 (2005).

147. Javanbakht, H. *et al.* Characterization of TRIM5alpha trimerization and its contribution to human immunodeficiency virus capsid binding. *Virology* **353,** 234–46 (2006).

148. Napolitano, L. M. & Meroni, G. TRIM family: Pleiotropy and diversification through homomultimer and heteromultimer formation. *IUBMB Life* **64,** 64–71 (2012).

149. Rhodes, D. A., de Bono, B. & Trowsdale, J. Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence? *Immunology* **116,** 411–7 (2005).

150. Gack, M. U. *et al.* Roles of RIG-I N-terminal tandem CARD and splice variant in TRIM25-mediated antiviral signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16743–8 (2008).

151. Kono, K., Bozek, K., Domingues, F. S., Shioda, T. & Nakayama, E. E. Impact of a single amino acid in the variable region 2 of the Old World monkey TRIM5alpha SPRY (B30.2) domain on anti-human immunodeficiency virus type 2 activity. *Virology* **388,** 160–8 (2009).

152. Li, Y., Li, X., Stremlau, M., Lee, M. & Sodroski, J. Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. *J. Virol.* **80,** 6738–44 (2006).

153. Stremlau, M., Perron, M., Welikala, S. & Sodroski, J. Species-Specific Variation in the B30 . 2 (SPRY) Domain of TRIM5α Determines the Potency of Human Immunodeficiency Virus Restriction. **79,** 3139–3145 (2005).

154. Perfetto, L. *et al.* Exploring the diversity of SPRY/B30.2-mediated interactions. *Trends Biochem. Sci.* **38,** 38–46 (2013).

155. D'Cruz, A. A., Babon, J. J., Norton, R. S., Nicola, N. A. & Nicholson, S. E. Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. *Protein Sci.* **22,** 1–10 (2013).

156. Zurek, B. *et al.* TRIM27 negatively regulates NOD2 by ubiquitination and proteasomal degradation. *PLoS One* **7,** e41255 (2012).

157. James, L. C., Keeble, A. H., Khan, Z., Rhodes, D. A. & Trowsdale, J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6200–5 (2007).

158. Woo, J.-S. *et al.* Structural and functional insights into the B30.2/SPRY domain. *EMBO J.* **25,** 1353–63 (2006).

159. Weinert, C., Grütter, C., Roschitzki-Voser, H., Mittl, P. R. E. & Grütter, M. G. The crystal structure of human pyrin b30.2 domain: implications for mutations associated with familial Mediterranean fever. *J. Mol. Biol.* **394,** 226–36 (2009).

160. D'Cruz, A. A. *et al.* Crystal structure of the TRIM25 B30.2 (PRYSPRY) domain: a key component of antiviral signalling. *Biochem. J.* **456,** 231–40 (2013).

161. Grütter, C. *et al.* Structure of the PRYSPRY-domain: implications for autoinflammatory diseases. *FEBS Lett.* **580,** 99–106 (2006).

162. Rhodes, D. A. & Trowsdale, J. TRIM21 is a trimeric protein that binds IgG Fc via the B30.2 domain. *Mol. Immunol.* **44,** 2406–14 (2007).

163. Biris, N. *et al.* Structure of the rhesus monkey TRIM5α PRYSPRY domain, the HIV capsid recognition module. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13278–13283 (2012).

164. Gao, B., Duan, Z., Xu, W. & Xiong, S. Tripartite motif-containing 22 inhibits the activity of hepatitis B virus core promoter, which is dependent on nuclear-located RING domain. *Hepatology* **50,** 424–33 (2009).

165. Zhang, S., Guo, J.-T., Wu, J. Z. & Yang, G. Identification and characterization of multiple TRIM proteins that inhibit hepatitis B virus transcription. *PLoS One* **8,** e70001 (2013).

166. Geoffroy, M.-C. & Chelbi-Alix, M. K. Role of promyelocytic leukemia protein in host antiviral defense. *J. Interferon Cytokine Res.* **31,** 145–58 (2011).

167. Wang, J. *et al.* TRIM56 is a virus- and interferon-inducible E3 ubiquitin ligase that restricts pestivirus infection. *J. Virol.* **85,** 3733–45 (2011).

168. Taylor, R. T. *et al.* TRIM79α, an interferon-stimulated gene product, restricts tick-borne encephalitis virus replication by degrading the viral RNA polymerase. *Cell Host Microbe* **10,** 185–96 (2011).

169. Mallery, D. L. *et al.* Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19985–90 (2010).

170. Hattlmann, C. J., Kelly, J. N. & Barr, S. D. TRIM22: A Diverse and Dynamic Antiviral Protein. *Mol. Biol. Int.* **2012,** 153415 (2012).

171. Tissot, C., Taviaux, S. A., Diriong, S. & Mechti, N. Localization of Staf50, a member of the Ring finger family, to 11p15 by fluorescence in situ hybridization. *Genomics* **34,** 151–3 (1996).

172. Sawyer, S. L., Emerman, M. & Malik, H. S. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathog.* **3,** e197 (2007).

173. Zhang, J. *et al.* The pericentromeric region of human chromosome 11: evidence for a chromosome-specific duplication. *Cytogenet. Cell Genet.* **94,** 137–41 (2001).

174. Van der Aa, L. M. *et al.* A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. *BMC Biol.* **7,** 7 (2009).

175. Duan, Z., Gao, B., Xu, W. & Xiong, S. Identification of TRIM22 as a RING finger E3 ubiquitin ligase. *Biochem. Biophys. Res. Commun.* **374,** 502–6 (2008).

176. Eldin, P. *et al.* TRIM22 E3 ubiquitin ligase activity is required to mediate antiviral activity against encephalomyocarditis virus. *J. Gen. Virol.* **90,** 536–45 (2009).

177. Di Pietro, A. *et al.* TRIM22 inhibits influenza A virus infection by targeting the viral nucleoprotein for degradation. *J. Virol.* **87,** 4523–33 (2013).

178. Barr, S. D., Smiley, J. R. & Bushman, F. D. The interferon response inhibits HIV particle production by induction of TRIM22. *PLoS Pathog.* **4,** e1000007 (2008).

179. Kajaste-Rudnitski, A. *et al.* TRIM22 inhibits HIV-1 transcription independently of its E3 ubiquitin ligase activity, Tat, and NF-kappaB-responsive long terminal repeat elements. *J. Virol.* **85,** 5183–96 (2011).

180. Torok, M. & Etkin, L. D. Two B or not two B? Overview of the rapidly expanding B-box family of proteins. *Differentiation.* **67,** 63–71 (2001).

181. Li, X., Yeung, D. F., Fiegen, A. M. & Sodroski, J. Determinants of the higher order association of the restriction factor TRIM5alpha and other tripartite motif (TRIM) proteins. *J. Biol. Chem.* **286,** 27959–70 (2011).

182. Sivaramakrishnan, G., Sun, Y., Rajmohan, R. & Lin, V. C. L. B30.2/SPRY domain in tripartite motif-containing 22 is essential for the formation of distinct nuclear bodies. *FEBS Lett.* **583,** 2093–9 (2009).

183. Parry, D. A. D., Fraser, R. D. B. & Squire, J. M. Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* **163,** 258–69 (2008).

184. Yu, S., Gao, B., Duan, Z., Xu, W. & Xiong, S. Identification of tripartite motif-containing 22 (TRIM22) as a novel NF-κB activator. *Biochem. Biophys. Res. Commun.* **410,** 247–51 (2011).

185. Sivaramakrishnan, G., Sun, Y., Rajmohan, R. & Lin, V. C. L. *B30.2/SPRY domain in tripartite motif-containing 22 is essential for the formation of distinct nuclear bodies*. *FEBS Lett.* **583,** 2093–2099 (2009).

186. Ohkura, S., Yap, M. W., Sheldon, T. & Stoye, J. P. All three variable regions of the TRIM5alpha B30.2 domain can contribute to the specificity of retrovirus restriction. *J. Virol.* **80,** 8554–65 (2006).

187. Tissot, C. & Mechti, N. Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus type 1 long terminal repeat expression. *J. Biol. Chem.* **270,** 14891–8 (1995).

188. Obad, S. *et al.* Staf50 is a novel p53 target gene conferring reduced clonogenic growth of leukemic U-937 cells. *Oncogene* **23,** 4050–9 (2004).

189. Bouazzaoui, A. *et al.* Stimulated trans-acting factor of 50 kDa (Staf50) inhibits HIV-1 replication in human monocyte-derived macrophages. *Virology* **356,** 79–94

190. Gao, B., Wang, Y., Xu, W., Duan, Z. & Xiong, S. A 5' extended IFN-stimulating response element is crucial for IFN-gamma-induced tripartite motif 22 expression via interaction with IFN regulatory factor-1. *J. Immunol.* **185,** 2314–23 (2010).

191. Gongora, C., Tissot, C., Cerdan, C. & Mechti, N. The interferon-inducible Staf50 gene is downregulated during T cell costimulation by CD2 and CD28. *J. Interferon Cytokine Res.* **20,** 955–61 (2000).

192. Wang, Y., Gao, B., Xu, W. & Xiong, S. BRG1 is indispensable for IFN-γ-induced TRIM22 expression, which is dependent on the recruitment of IRF-1. *Biochem. Biophys. Res. Commun.* **410,** 549–54 (2011).

193. Obad, S., Olofsson, T., Mechti, N., Gullberg, U. & Drott, K. Regulation of the interferon-inducible p53 target gene TRIM22 (Staf50) in human T lymphocyte activation. *J. Interferon Cytokine Res.* **27,** 857–64 (2007).

194. Gao, B. *et al.* Inhibition of histone deacetylase activity suppresses IFN-γ induction of tripartite motif 22 via CHIP-mediated proteasomal degradation of IRF-1. *J. Immunol.* **191,** 464–71 (2013).

195. Gao, B. *et al.* p300, but not PCAF, collaborates with IRF-1 in stimulating TRIM22 expression independently of its histone acetyltransferase activity. *Eur. J. Immunol.* **43,** 2174–84 (2013).

196. Herr, A.-M., Dressel, R. & Walter, L. Different subcellular localisations of TRIM22 suggest species-specific function. *Immunogenetics* **61,** 271–80 (2009).

197. Sivaramakrishnan, G., Sun, Y., Tan, S. K. & Lin, V. C. L. Dynamic localization of tripartite motif-containing 22 in nuclear and nucleolar bodies. *Exp. Cell Res.* **315,** 1521–1532 (2009).

198. Petersson, J. *et al.* The human IFN-inducible p53 target gene TRIM22 colocalizes with the centrosome independently of cell cycle phase. *Exp. Cell Res.* **316,** 568–79 (2010).

199. Cioce, M. & Lamond, A. I. Cajal bodies: a long history of discovery. *Annu. Rev. Cell Dev. Biol.* **21,** 105–31 (2005).

200. Patarca, R. *et al.* rpt-1, an intracellular protein from helper/inducer T cells that regulates gene expression of interleukin 2 receptor and human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2733–7 (1988).

201. Singh, R. *et al.* Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. *J. Virol.* **85,** 208–16 (2011).

202. Singh, R. *et al.* TRIM5α and TRIM22 are differentially regulated according to HIV-1 infection phase and compartment. *J. Virol.* (2014). doi:10.1128/JVI.03603-13

203. Wieland, S., Thimme, R., Purcell, R. H. & Chisari, F. V. Genomic analysis of the host response to hepatitis B virus infection. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6669–74 (2004).

204. Su, A. I. *et al.* Genomic analysis of the host response to hepatitis C virus infection. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 15669–74 (2002).

205. Folkers, M. E. *et al.* ENCODE tiling array analysis identifies differentially expressed annotated and novel 5' capped RNAs in hepatitis C infected liver. *PLoS One* **6,** e14697 (2011).

206. Wittmann, S. *et al.* New prognostic markers revealed by evaluation of genes correlated with clinical parameters in Wilms tumors. *Genes. Chromosomes Cancer* **47,** 386–95 (2008).

207. Zirn, B. *et al.* Expression profiling of Wilms tumors reveals new candidate genes for different clinical parameters. *Int. J. Cancer* **118,** 1954–62 (2006).

208. Sun, Y. *et al.* Down-regulation of tripartite-motif containing 22 expression in breast cancer is associated with a lack of p53-mediated induction. *Biochem. Biophys. Res. Commun.* **441,** 600–6 (2013).

209. Deng, Y. J. *et al.* Gene profiling involved in immature CD4+ T lymphocyte responsible for systemic lupus erythematosus. *Mol. Immunol.* **43,** 1497–507 (2006).

210. Coit, P. *et al.* Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naïve CD4+ T cells from lupus patients. *J. Autoimmun.* **43,** 78–84 (2013).

211. Bochud, P.-Y., Bochud, M., Telenti, A. & Calandra, T. Innate immunogenetics: a tool for exploring new frontiers of host defence. *Lancet Infect. Dis.* **7,** 531–42 (2007).

212. Daley, D. *et al.* Associations and interactions of genetic polymorphisms in innate immunity genes with early viral infections and susceptibility to asthma and asthma-related phenotypes. *J. Allergy Clin. Immunol.* **130,** 1284–93 (2012).

213. Azad, A. K., Sadee, W. & Schlesinger, L. S. Innate immune gene polymorphisms in tuberculosis. *Infect. Immun.* **80,** 3343–59 (2012).

214. Netea, M. G., Wijmenga, C. & O'Neill, L. A. J. Genetic variation in Toll-like receptors and disease susceptibility. *Nat. Immunol.* **13,** 535–42 (2012).

215. Arora, P. *et al.* Genetic polymorphisms of innate immunity-related inflammatory pathways and their association with factors related to type 2 diabetes. *BMC Med. Genet.* **12,** 95 (2011).

216. Sobieszczyk, M. E., Lingappa, J. R. & McElrath, M. J. Host genetic polymorphisms associated with innate immune factors and HIV-1. *Curr. Opin. HIV AIDS* **6,** 427–34 (2011).

217. Doniger, S. W. *et al.* A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **4,** e1000183 (2008).

218. Ramensky, V. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30,** 3894–3900 (2002).

219. Subramanian, S. & Kumar, S. Higher intensity of purifying selection on >90% of the human genes revealed by the intrinsic replacement mutation rates. *Mol. Biol. Evol.* **23,** 2283–7 (2006).

220. Ferrer-Admetlla, A. *et al.* Balancing selection is the main force shaping the evolution of innate immunity genes. *J. Immunol.* **181,** 1315–22 (2008).

221. Vallender, E. J. & Lahn, B. T. Positive selection on the human genome. *Hum. Mol. Genet.* **13 Spec No,** R245–54 (2004).

222. Cagliani, R. *et al.* Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. *Hum. Genet.* **128,** 577–88 (2010).

223. George Priya Doss, C. & Rajith, B. Computational refinement of functional single nucleotide polymorphisms associated with ATM gene. *PLoS One* **7,** e34573 (2012).

224. Mah, J. T. L., Low, E. S. H. & Lee, E. In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery. *Drug Discov. Today* **16,** 800–9 (2011).

225. Armitage, A. E. *et al.* APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete "all or nothing" phenomenon. *PLoS Genet.* **8,** e1002550 (2012).

226. Sunyaev, S. Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10,** 591–597 (2001).

227. Wu, J. & Jiang, R. Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases. *Sci. World J.* **2013,** 675851 (2013).

228. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7,** e46688 (2012).

229. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–33 (2010).

230. Jensen, L. J. *et al.* Prediction of human protein function from post-translational modifications and localization features. *J. Mol. Biol.* **319,** 1257–65 (2002).

231. Radivojac, P. *et al.* Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* **78,** 365–80 (2010).

232. Ren, J. *et al.* Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. *Proteomics* **9,** 3409–3412 (2009).

Chapter 2

2 Ancient and recent adaptive evolution in the antiviral *TRIM22* gene: identification of a single nucleotide polymorphism that impacts TRIM22 function [1]

TRIM22 is a novel IFN-induced protein that potently inhibits the replication of evolutionarily diverse viruses, including HIV-1, HBV, and IAV. Altered TRIM22 expression is also associated with diseases such as multiple sclerosis, cancer, and autoimmunity. To date, the factors that influence TRIM22 expression and antiviral activity are largely unknown. Here, we used an evolutionary approach to identify potential genetic determinants of TRIM22 function. Evolutionary analysis using 29 mammalian TRIM22 sequences revealed that TRIM22 evolution has been shaped by ancient and variable selective forces. Positive selection has operated on a number of TRIM22 sites, many of which cluster together in putative protein-protein interaction motifs. Interestingly, we found that the second most prevalent nsSNP in the human *TRIM22* gene (rs1063303:G>C) is located at one of these positively selected sites. In addition, the frequency of this nsSNP varied up to 10-fold between different ethnic populations. We found that nsSNP rs1063303:G>C had an inverse functional impact, whereby it increased TRIM22 expression and decreased TRIM22 antiviral activity. Taken together, our data describes the extensive genetic variation in TRIM22 and identifies nsSNP rs1063303:G>C as a highly prevalent nsSNP that impacts TRIM22 function.

---

[1] The material contained in this chapter was published in: Kelly, J.N., Woods, M.W., Xhiku, S., and Barr, S.D. Ancient and recent adaptive evolution in the antiviral *TRIM22* gene: identification of a single nucleotide polymorphism that impacts TRIM22 function. *Hum Mutat*. (2014).

2.1 Introduction

The TRIM family is a large group of proteins involved in diverse cellular processes. TRIM proteins are increasingly becoming recognized as key regulators of the innate immune response and a number of TRIM proteins have been shown to possess antiviral activity [1–6]. According to a recent evolutionary study, TRIM proteins can be divided into two main groups based on genomic organization, evolutionary properties, and domain structure. Group 1 TRIM proteins (G1) are present in both invertebrate and vertebrate species and have variable C-terminal domains, while Group 2 TRIM proteins (G2) are only found in vertebrates and have a C-terminal B30.2 domain [7]. All TRIM proteins (G1 and G2) have a conserved N-terminal RBCC motif, which consists of a RING domain, one or two BB domains, and a predicted CC region. The RING domain often has E3 ligase activity, allowing some TRIMs to modify other proteins (including viral proteins) with ubiquitin or ubiquitin-like molecules [8–10].

In the human genome, four G2 *TRIM* genes cluster together on chromosome 11: *TRIM6*, *TRIM34*, *TRIM5,* and *TRIM22* (Online Mendelian Inheritance in Man #606559). Evolutionary analysis of this gene cluster has shown that *TRIM6* and *TRIM34* have evolved under purifying selection in primates, and that *TRIM5* and *TRIM22* have a dynamic evolutionary relationship in primates that includes episodes of gene expansion and gene loss [11,12]. It was previously shown that *TRIM5* and *TRIM22* sequences have undergone positive selection in primates, with primate lineages showing positive selection in either *TRIM5* or *TRIM22*, but not both [12–15]. Moreover, the *TRIM5* and *TRIM22* genes of Haplorhini primates have evolved species-specific differences in transcriptional regulation, mediated by transposable element sequences in their non-coding regions [16]. Both *TRIM5* and *TRIM22* encode proteins with antiviral activity against retroviruses. TRIM5α inhibits HIV-1 replication in non-human primate cells; however, this activity is weak to absent in human cells [17–19]. Conversely, TRIM22 inhibits HIV-1 replication in several human cell lines and primary MDMs [20–23]. The TRIM22 protein also inhibits the replication of EMCV, HBV, and IAV [24–26].

Genetic conflict between host and viral genomes can lead to the rapid accumulation of amino acid replacement changes (dN) relative to synonymous changes (dS). This

phenomenon of positive selection (i.e. dN/dS >1) is one hallmark of the evolutionary 'battle' that occurs between host antiviral factors and their pathogen antagonists [27,28]. Although positive selection is not typically observed in other host genes, multiple antiviral factors contain genetic signatures of positive selection [29,30]. For example, well-known host restriction factors, including APOBEC3G, BST2/tetherin, and TRIM5α all contain positively selected sites that play key roles in their antiviral activities [13–15].

A previous evolutionary study of TRIM22 in primates spanning ~33 million years of primate divergence identified a strong signature of positive selection for TRIM22 in hominoids and old world monkeys (OWM) [12]. The positions of amino acid residues found to be under positive selection were located predominantly in the β2-β3 surface loop of B30.2 domain. In the present study, we analyzed TRIM22 sequences from evolutionarily diverse mammals spanning ~100 million years and used an evolution-guided functional approach to identify residues that may dictate TRIM22-mediated antiviral activity. Using three paired evolutionary models, we identified residues in several TRIM22 domains that are predicted to be under strong positive selection in mammals. We demonstrated that the human *TRIM22* gene exhibits remarkable genetic diversity and characterized one highly prevalent nsSNP (rs1063303:G>C) that is located at a site evolving under strong positive selection. Notably, we showed that nsSNP rs1063303:G>C has an inverse functional impact on TRIM22, whereby it increased TRIM22 expression levels and decreased TRIM22 antiviral activity. Our findings in this study indicate that genetic variation in the TRIM22 gene is both prevalent and ancient, and that positively selected sites in TRIM22 may influence its antiviral activity.

2.2 Materials and methods

2.2.1 Sequence analysis

TRIM22 sequences were aligned in COBALT and a phylogenetic tree was generated using EvolView software [31,32]. Positive selection was evaluated using the Selecton program, as previously described [33,34]. Briefly, sequences were analyzed using three paired models, two nested (M8a and M8; M7 and M8) and one non-nested (M8a and

MEC). The models that allowed positive selection to occur (M8 and MEC) fit the data better than models that did not allow positive selection to occur (M8a and M7). The non-synonymous (Ka) to synonymous (Ks) ratio (i.e. Ka/Ks) was also calculated for each codon using Selection. Codons with a Ka/Ks ratio >1.5 exhibited high probabilities of having evolved under positive selection and were highlighted in Fig 2.1b. Primate TRIM22 sequences in Fig 2.2c were also aligned in COBALT. Frequency data for the rs1063303:G>C nsSNP was extracted from the 1000 Genomes database and includes individuals from the following regions: African (AFN) population includes Nigeria, Kenya, The Gambia, Sierra Leone, Americans of African ancestry in southwestern United States, African Caribbean in Barbados; American (AMR) population includes Puerto Rico, Colombia, Peru, Mexican ancestry from Los Angeles, United States; European (EUR) population includes Italy, United Kingdom, Finland, Spain, Northern and Western European ancestry from Utah, United States; Asian (ASN) population includes China, Vietnam, and Japan [35].

2.2.2 Cells, plasmids, and transfections

Cells were maintained at 37°C with 5% CO_2 in standard DMEM growth medium (Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin). 293T and HeLa cell lines were obtained from the ATCC. HOS-CD4+ cell line was obtained from NIH AIDS Reagents. The empty vector plasmid (pGL3) was purchased from Promega and the pR9 plasmid encoding replication-competent HIV-1 was kindly provided by Dr. F. Bushman (University of Pennsylvania, USA). The coding region of the wild type *TRIM22* gene (GenBank Accession NM_006074.4) was subcloned into p3xFLAG-CMV-10 (Sigma) using HindIII and XbaI restriction sites to generate pWT-T22. The flag-tagged nsSNP rs1063303:G>C plasmid (pSNP-T22) was created using PCR mutagenesis. pSNP-T22 contains C in place of the ancestral allele G at nucleotide position 725 of the *TRIM22* coding region (GenBank Accession NM_006074). The following primers were used to amplify pWT-T22 and generate PCR fragment 1: "Forward WT" (5' ACG TAA GCT TAT GGA TTT CTC AGT AAA GG 3') and "Reverse SNP" (5' GAC GAT CCC GTC AAC CTC CGC TGG AGA 3'). Similarly, PCR fragment 2 was generated using the

primers: “Forward SNP” (5’ TCT CCA GCG GAG GTT GAC GGG ATC GTC 3’) and “Reverse WT” (5’ ACG TTC TAG ATC AGG AGC TCG GTG GGC ACA CAG 3’). Following amplification, a 1:25 dilution of PCR fragment 1 and PCR fragment 2 was added with “Forward WT” and “Reverse WT” primers together in a PCR reaction. The amplified *TRIM22* coding region containing the nsSNP was subcloned into p3xFLAG-CMV-10 (Sigma) using HindIII and XbaI to generate pSNP-T22. The entire coding region of pSNP-T22 was sequenced and no other PCR-introduced variations were detected. 293T and HeLa cells were seeded in 12-well or 6-well plates and transfected using Lipofectamine 2000 (Invitrogen) with 2 μg or 5 μg of plasmid DNA (pEV, pWT-T22, or pSNP-T22), respectively. As a control for transfection efficiency, a plasmid encoding enhanced green fluorescent protein (peGFP) (Clontech) was included in the transfections at a concentration of one tenth the total amount of transfected DNA. The percent GFP+ cells were measured using standard flow cytometry. Unless otherwise stated, all co-transfections of pR9 with pEV, pWT-T22, and pSNP-T22 were performed at a ratio of 10:1.

2.2.3 RNA isolation and real-time PCR

Total RNA was extracted from cells using the R&A-BLUE Total RNA Extraction kit (Frogga Bio). To ensure that no detectable genomic or plasmid DNA was carried over during the RNA purification, 0.1μg of each RNA sample was subjected to PCR (35 cycles) and real-time PCR using primers within the *TRIM22* coding region or primers specific to the TRIM22 3’ untranslated region. 1μg of DNase-treated RNA was then reverse transcribed to cDNA using the M-MLV reverse transcriptase and Oligo(dT) primers (Invitrogen). Prior to real-time PCR, cDNA samples were diluted 1:10 with water. Each PCR reaction consisted of 10μl of SYBR Green Master Mix, 2μl of the appropriate primers (1μl of 10μM forward primer (5’ CAT CTG CCT GGA GCT CCT GAC 3’) and 1μl of 10μM reverse primer (5’ AGA TGA TCA CTG ACT CCT TGA TCT TTG C 3’), 1μl of diluted cDNA, and water to a total volume of 20μl. Real-time PCR was run on the Rotor-Gene 6000 real-time PCR machine (Corbett Life Science) under the following cycling conditions: 10 min at 95°C and 40 cycles of 10 sec at 95°C, 15 sec at 60°C, and 20 sec at 72°C. The Rotor-Gene 6000 series software (version 1.7)

was used to determine the C_T for each PCR reaction. All samples were amplified in triplicate with no template controls and the mean was used for further analysis.

2.2.4 Western Blotting

Cells were pelleted by centrifugation (350 x g for 10 minutes), washed twice with PBS, and lysed in 1X RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1x Complete Protease Inhibitor (Roche), 1% Triton X-100, 0.1% SDS). Virus released into the supernatant was clarified by centrifugation, pelleted by centrifugation (21,000 x g for 2 hours) over a cushion of 20% sucrose and lysed with 1X RIPA buffer. Protein was separated on a 12% SDS-PAGE gel and then transferred to a FluorTransW membrane (Pall) by semi-dry transfer. Following transfer, the membrane was blocked for 1 hour in LI-COR Blocking Buffer (LI-COR Biosciences) and incubated overnight with primary antibody at 4°C (1:1000 dilution with LI-COR Blocking Buffer). Detection was carried out using an IRDye-labeled secondary antibody (1:20,000 dilution with LI-COR Blocking Buffer for 30 minutes) and the LI-COR Odyssey Detection System (LI-COR Biosciences).

2.2.5 Confocal immunofluorescence microscopy

Cells were seeded in 12-well plates on 18mm coverslips. 24 hours post-transfection, cells were washed with PF buffer (1x PBS + 1% FBS), fixed (1x PBS + 5% formaldehyde + 2% sucrose for 10 minutes), and permeabilized (1x PBS, 5% NP-40, 10% sucrose for 10 minutes). Cells were incubated with primary antibodies for 1 hour (1:1000 dilution of mouse anti-Flag antibody in PF buffer), washed thoroughly with PF buffer, and then incubated with secondary antibodies for 1 hour (1:1000 dilution of AlexaFluor 546 anti-mouse antibody in PF buffer). Coverslips containing the cells were mounted on glass slides with Vectashield mounting media (Vector laboratories) and slides were examined using a Zeiss LSM 510 confocal fluorescence microscope. DAPI staining was used to visualize the nuclei (blue). Images shown represent optical slices taken through the center of cells from a series of *z*-stack images with a 63x Plan-Apochromat oil immersion objective lens (Numerical Aperture = 1.4).

2.2.6 Neutrality tests

Tajima's *D* and Fu's F_S neutrality tests were performed to distinguish between neutrally evolving sequences under mutation-drift equilibrium and sequences evolving under non-neutral processes such as balancing selection [36–38]. These two tests are based on the principle that a recent population expansion associated with a non-neutral process will detect a shift in the allele frequency spectrum compared to a neutral Wright-Fisher model. A negative Tajima's *D* signifies an excess of low frequency variants relative to expectation, indicating population size expansion (e.g. after a bottleneck or selective sweep) and/or purifying selection. A positive Tajima's *D* signifies low levels of both low and high frequency (i.e. an excess of intermediate frequency) variants, indicating a decrease in population size and/or balancing selection. A negative value of Fu's F_S is evidence for an excess number of rare alleles, as would be expected from a recent population expansion or from genetic hitchhiking. A positive value of F_S is evidence for a deficiency of alleles, as would be expected from a recent population bottleneck or from overdominant selection. The F_S statistic was considered significant at the 5% level if its *P* value was below 0.02 [38]. The analyses were performed using the software Arlequin 3.5.1.3 [39]. Simulated *P* values were generated using 10,000 simulations under a model of selective neutrality.

2.2.7 Statistical analysis

Unless stated otherwise, statistical analyses were performed using GraphPad Prism (Version 6.0). *P* values of less than 0.05 were considered statistically significant.

2.3 Results

2.3.1 Positive selection in multiple TRIM22 domains among mammals

To better understand the evolution of TRIM22, and to examine how selective pressures have shaped its antiviral properties, we obtained the *TRIM22* coding sequence from 29 evolutionarily diverse mammalian species (Table 2.1). We aligned these sequences and

Table 2.1

Table 2.1: Mammalian TRIM22 coding sequences used for Selecton analysis.

Common name	Scientific name	Type [1]	NCBI Reference	Verified [2]
Human	*Homo sapiens*	HOM	NP_006065.2	Yes
Chimpanzee	*Pan troglodytes*	HOM	NP_001106867.1	Yes
Bonobo	*Pan paniscus*	HOM	XP_003819095.1	Yes
Gorilla	*Gorilla gorilla*	HOM	XP_004050610.1	Yes
Sumatran Orangutan	*Pongo abelii*	HOM	NP_001153280.1	No
Bornean Orangutan	*Pongo pygmaeus*	HOM	EU124707	Yes
Island Siamang	*Hylobates syndactylus*	HOM	EU124708	Yes
White-Cheeked Gibbon	*Nomascus leucogenys*	HOM	NP_001267028.1	Yes
African Green	*Cercopithecus aethiops*	OWM	EU124709	Yes
Patas	*Erythrocebus patas*	OWM	EU124696	Yes
Sooty Mangabey	*Cercocebus atys*	OWM	EU124695	Yes
Baboon	*Papio anubis*	OWM	NP_001162333.1	No
Rhesus Macaque	*Macaca mulatta*	OWM	NP_001106830.1	Yes
Kikuyu Colobus	*Colobus guereza kikuyuensis*	OWM	EU124712	Yes
Red-Shanked Douc	*Pygathrix nemaeus*	OWM	EU124710	Yes
Bolivian Squirrel	*Saimiri boliviensis*	NWM	XM_003923356.1	No
Common Squirrel	*Saimiri sciureus sciureus*	NWM	EU124716	Yes
White-Faced Saki	*Pithecia pithecia pithecia*	NWM	EU124715	Yes
Dusky Titi	*Callicebus moloch*	NWM	EU124692	Yes
Common Marmoset	*Callithrix jacchus*	NWM	XM_002754905.2	No

Pygmy Marmoset	*Callithrix pygmaea*	NWM	EU124714	Yes
Bolivian Red Howler	*Alouatta sara*	NWM	EU124713	Yes
Southern White Rhinoceros	*Ceratotherium simum simum*	OM	XM_004418316.1	No
Cat	*Felis catus*	OM	XM_003992898.1	No
Pacific Walrus	*Odobenus rosmarus divergens*	OM	XM_004407583.1	No
Giant Panda	*Ailuropoda melanoleuca*	OM	XM_002926071.1	No
Dog	*Canis lupus familiaris*	OM	XM_542402.4	No
Florida Manatee	*Trichechus manatus latirostris*	OM	XM_004389072.1	No
African Elephant	*Loxodonta Africana*	OM	NW_003573536.1	No

[1] Mammals are classified as hominoids (HOM), old world monkeys (OWM), new world monkeys (NWM) or other mammals (OM).
[2] Sequences were independently verified in [12]

generated a phylogenetic tree, representing ~100 million years of evolution (Fig 2.1a). Our primate phylogeny was consistent with previously published studies and only orthologous non-primate sequences were included in our analysis (Table 2.1) [12]. Using this dataset, we employed the software program Selecton to evaluate TRIM22 evolution among mammals. Selecton uses several standard evolutionary models, including two nested paired models (M8a and M8; M7 and M8) and one non-nested paired model (M8a and MEC) [33,34]. We compared the nested pairs using the likelihood ratio test and found that in both cases, the model that allowed sites to evolve under positive selection (M8) fit the data significantly better than the models that did not (M8a and M7) (Table 2.2). The non-nested MEC model differs from the nested model in that it accounts for differences in amino acid replacement rates. A position with radical amino acid replacements will obtain a higher Ka value than a position with more moderate replacements. Akaike Information Content (AIC_c) scores are compared between the MEC and the M8a models. The lower the AIC_c score, the better the fit of the model to the data, and hence the model is considered more justified [33]. Comparison of the AICc scores of the MEC and M8a models were indicative of positive selection (M8a: 25211.8, MEC: 25079.2) and that the data was more congruent with the model that allowed for positive selection (MEC) than the one that did not (M8a) (Table 2.2).

The Selecton analysis identified several codons with high probabilities of having evolved under positive selection (Fig 2.2). Eleven of the 28 codons predicted to be evolving under positive selection were located in the B30.2 domain (Fig 2.1b Table 2.3). Six of these eleven codons (K324, R327, T330, K332, S334, C337) clustered together in the one region, whereas the other five (S377, S395, G471, L488, V489) were more dispersed. The codon numbering uses the translation initiation codon as codon 1. Interestingly, many of the positively selected codons corresponded in location and spacing to previously identified positively selected codons in the TRIM5α protein (K324, P325, G330, R332, R335, Q337, F339, V340, K389, Q471, G483) (Fig 2.1c) [15]. In TRIM5α, most positively selected codons are located in one of four variable regions (v1-v4) in the B30.2 domain. These variable regions are also found in a number of other TRIM proteins, including TRIM22 [15,19]. The v1 region of TRIM5α (or the ‘antiviral patch’

Figure 2.1: Mapping positively selected sites in the TRIM22 protein.

(a) Phylogenetic tree showing the evolutionary relationship among 29 mammalian species for the TRIM22 protein. Tree was created using COBALT and EvolView software. Numbers shown represent arbitrary distances from each branch point on the tree. Hominoids (HOM) are highlighted in red, old world monkeys (OWM) in purple, new world monkeys (NWM) in blue, and additional mammals in green. **(b)** Bayesian analysis of mammalian TRIM22 coding sequences. Ka/Ks values for each codon are plotted on y-axis. Ka/Ks ratio >1 indicates positive selection, Ka/Ks ratio <1 indicates purifying selection, and Ka/Ks ratio of 1 indicates neutral selection. Asterisks show approximate location of codons under strong positive selection. TRIM22 domains are shown above the graph, along with the approximate location of functional motifs (C15/18: two cysteine residues required for E3 ligase activity; C97, H100: zinc finger motif in the BB2 domain; predicted CC region; NLS). The 'antiviral patch' refers to several residues previously shown to be a major specificity determinant for TRIM5α-mediated anti-HIV activity. **(c)** Comparative model showing the putative B30.2 3D structures for both TRIM5α and TRIM22. The blue and red colored regions represent structurally conserved regions (blue) or regions with no correspondence in structural proximities (red) after structural alignment. The Q^H value is a metric for structural homology (an adaptation of the Q value that measures structural conservation). Q=1 implies that two structures are identical. When Q has a low score (0.1-0.3), structures are not aligned well (i.e. only a small fraction of the C_α atoms superimpose). Colored balls correspond to residues (van der Waals radii) predicted to evolve under strong positive selection (colored yellow for TRIM5α and green for TRIM22). **(d)** Molecular model of TRIM22's B30.2 domain. The colored regions in the image on the left show the four variable regions (v1: red, v2: blue, v3: magenta, and v4: green). The image on the right shows only the variable regions (v1-v4) plus the location of sites predicted to evolve under strong positive selection (yellow). Molecular models were created with Visual Molecular Dynamics (VMD) software (v1.9.1) support. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

Figure 2.1

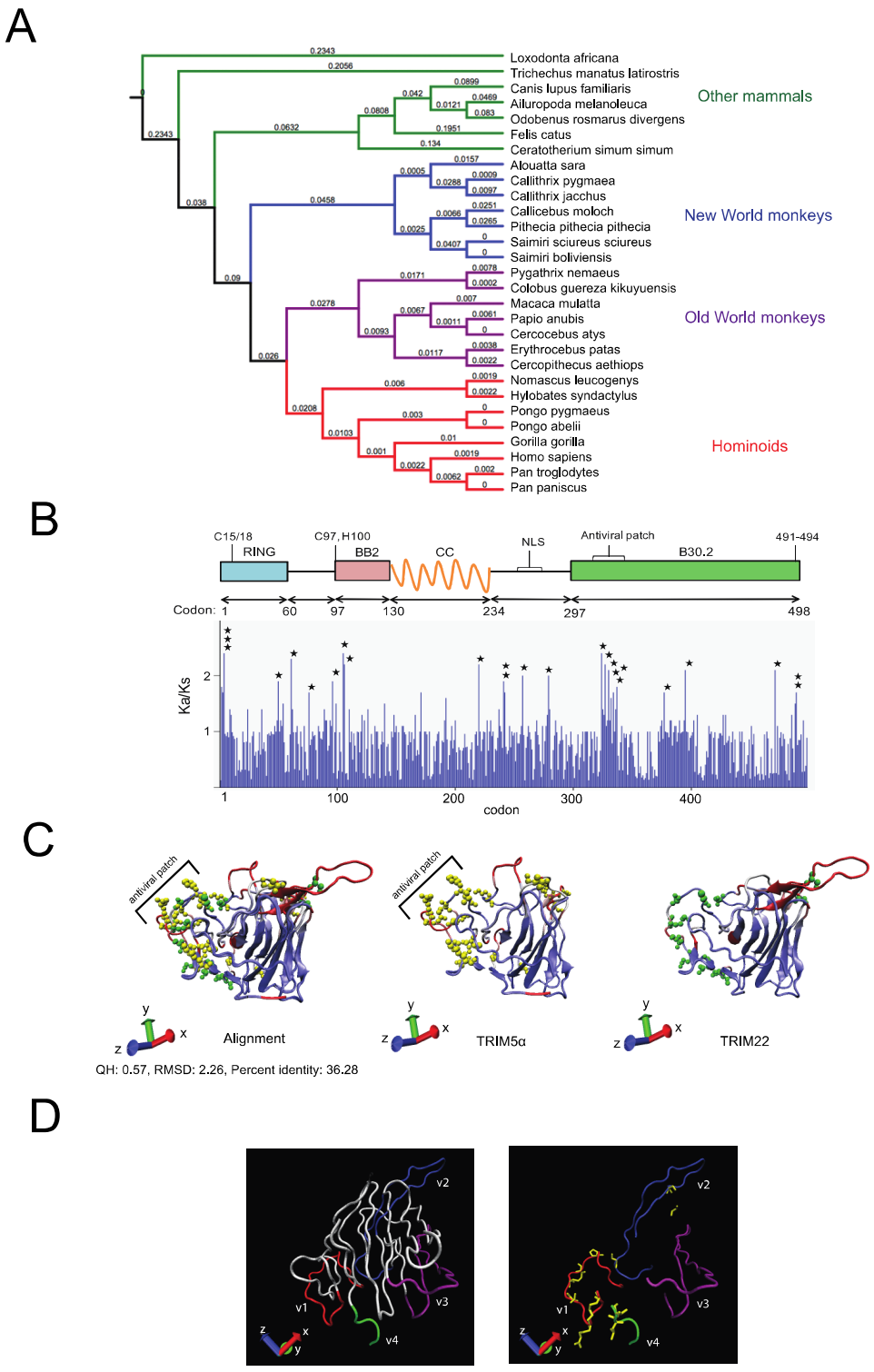


Figure 2.1: Mapping positively selected sites in the TRIM22 protein.

Table 2.2

Table 2.2: Models of evolution applied to TRIM22 coding sequences.

Nested	**Log-likelihood**	**2(Ln_1-Ln_0)**	**Degrees of Freedom**	**X^2**
M8a, M8	-12601.9, -12599.1	5.6	1	<0.001
M7, M8	-12608.1, -12599.1	18	2	<0.001
Non-nested	**Log-likelihood**	**AIC_c Score$^{\Delta}$** [1]	**Parameters**	
M8a, MEC	-12601.9, -12534.6	25211.8, 25079.2	4, 5	

[1] AIC_c Score$^{\Delta}$ = -2*log$L$+2p*(N/N-p-1), where L is the likelihood of the model given the data, p is the number of free parameters, and N is the sequence length. The lower the AIC_C score, the better the fit of the model to the data, and hence the model is considered more justified.

Figure 2.2: Selecton analysis of mammalian TRIM22 coding sequences.

Selecton results for 29 mammalian TRIM22 coding sequences using the MEC model. Positively selected sites are colored orange or yellow (Ka/Ks > 1), whereas sites undergoing purifying selection are colored shades of purple (Ka/Ks <1). Sites colored white have a Ka/Ks ratio = 1.

Figure 2.2

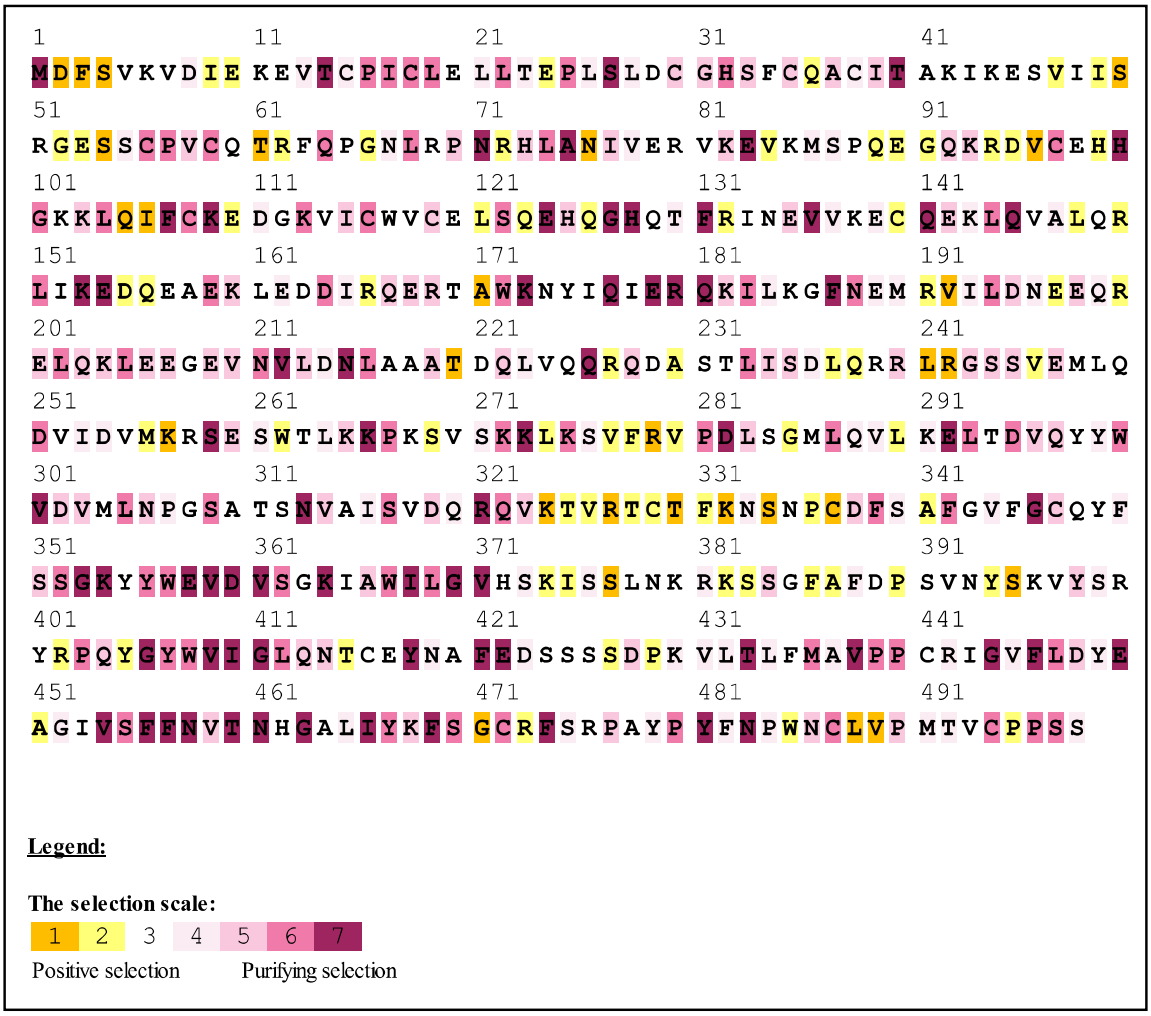


Figure 2.2: Selecton analysis of mammalian TRIM22 coding sequences.

Table 2.3

Table 2.3: Positive selection in TRIM22 protein domains.

Domain (AA [1])	Percent PS [2]	Ka/Ks [3]	Positively selected codons
RING (1-59)	1.00	1.86	D2, F3, S4, S50, S54
SP1 (60-96)	0.60	1.97	T61, N76, V96
BB2 (97-129)	0.40	2.30	Q105, I106
CC (130-233)	0.60	1.83	A171, V192, T220
SP2 (234-296)	0.80	1.90	L241, R242, K257, R279
B30.2 (297-498)	2.21	1.90	K324, R327, T330, K332, S334, C337, S377, S395, G471, L488, V489

[1] AA: Amino acids included in each domain
[2] Percent PS: Percent of positively selected sites in each domain compared to the total number of sites
[3] Ka/Ks: Number of non-synonymous substitutions per non-synonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks)

region) was previously shown to be a major determinant for species-specific HIV-1 restriction in primates [18]. Mutations in other variable regions (v2-v4) of TRIM5α have also been shown to alter HIV-1, SIV, and N-MLV restriction [40–44]. Remarkably, all of the positively selected codons in TRIM22's B30.2 domain (except G471) were located in one of its v1-v4 regions. Six positively selected codons were found in the v1 region, two in the v2 region, and two in the v4 region. Unlike TRIM5α, there were no positively selected codons in TRIM22's v3 region (Fig 2.1d).

Outside of the B30.2 domain, we also found evidence for positively selected codons in each of TRIM22's other domains (Table 2.3). Many of these codons clustered together around putative functional motifs in the TRIM22 protein. For example, three positively selected codons (V96, Q105, I106) surrounded a zinc finger motif in the BB2 domain (Fig 2.1b, Table 2.3). Zinc finger motifs are critical for proper domain folding and often contribute to protein-protein interactions [45]. We also identified four positively selected codons (L241, R242, K257, R279) that flanked the nuclear localization signal (NLS) in TRIM22's SP2 domain (Fig 2.1b, Table 2.3). Along with amino acids 491-494 in the B30.2 domain, TRIM22's NLS was previously shown to be required for its nuclear localization [46,47]. Notably, we also identified two positively selected codons (L488, V489) in the B30.2 domain that are located directly upstream of amino acids 491-494 (Fig 2.1b Table 2.3).

2.3.2 Genetic Variation in the Human *TRIM22* Gene

Since TRIM22 is an integral component of the host antiviral response, and since it has been linked to a number of diseases, we also had an interest in its evolutionary pattern among humans. To examine genetic variation in the human *TRIM22* gene, we compiled a list of known nsSNPs in the TRIM22 coding region using the National Center for Biotechnology Information (NCBI) dbSNP database. To date, 64 nsSNPs and 2 indels have been identified in the exons of human TRIM22, most of which are found in its B30.2 domain. The majority of these nsSNPs result in missense variations; however, several generate truncated versions of the TRIM22 protein (nonsense and frameshift

variations). To identify nsSNPs that may play important roles in the activity of human TRIM22, we used a program called *Sorting Intolerant From Tolerant* (SIFT) to predict their functional effects [48]. SIFT analysis predicted that 23 nsSNPs were deleterious to TRIM22 function and 33 nsSNPs were tolerated. Paradoxically, two of the potentially deleterious nsSNPs (rs192306924:C>A, rs1063303:G>C) were located at amino acid sites evolving under strong positive selection in mammals (Table 2.2). Due to its high prevalence in the human population, we selected nsSNP rs1063303:G>C for further analysis and characterization.

The guanine (G) to cytosine (C) nucleotide change in nsSNP rs1063303:G>C results in an arginine (R) to threonine (T) amino acid change at position 242 in the TRIM22 protein (Fig 2.3a). To determine its prevalence in the human population, we obtained frequency data from four different ethnic populations using the 1000 Genomes database [35]. Interestingly, we observed large differences in the frequency of nsSNP rs1063303:G>C among various ethnic groups. For example, 31% of European (EUR) individuals were homozygous for the nsSNP (C/C) genotype; however, only 3% of Asian (ASN) individuals were homozygous for C/C. African (AFN) and American (AMR) populations had intermediate nsSNP frequencies, with 15% and 19% of individuals who were homozygous for the C/C genotype, respectively (Fig 2.3b). An alignment of amino acid site 242 in mammals revealed that the ancestral allele (G) was highly conserved in hominoids (HOM) and Old World monkeys (OWM), but variable in New World monkeys (NWM) and other mammals (Fig 2.3c). The reason for these frequency differences among humans for nsSNP rs1063303:G>C is unknown.

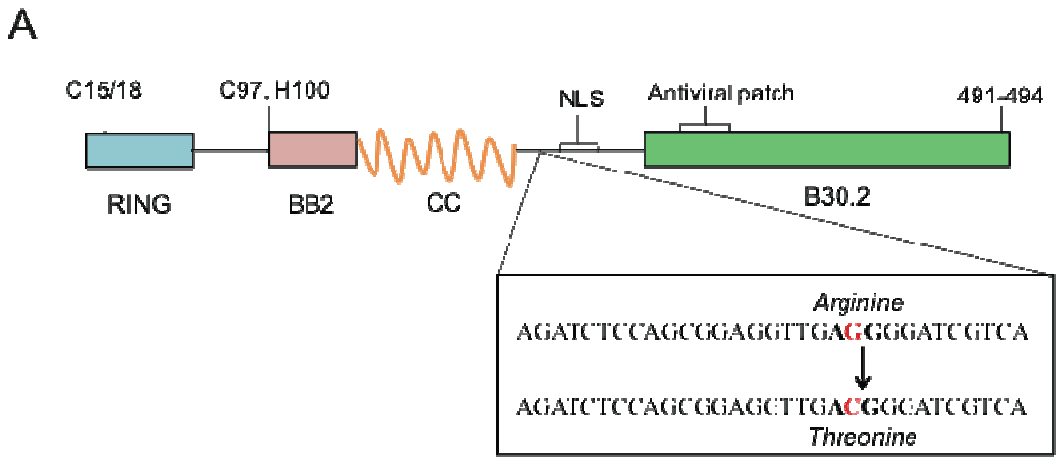
Tests for Selective Neutrality

A recent genome-wide scan of two ethnic populations identified *TRIM22* as one of 60 'extreme' genes undergoing balancing selection in humans [49]. To distinguish between a neutrally evolving site under mutation-drift equilibrium and a site evolving under non-neutral processes such as balancing selection, we performed both the Tajima's *D* and Fu's F_S tests for neutrality on the *TRIM22* rs1063303 variant. These two tests are based on the principle that a recent population expansion associated with a non-neutral process

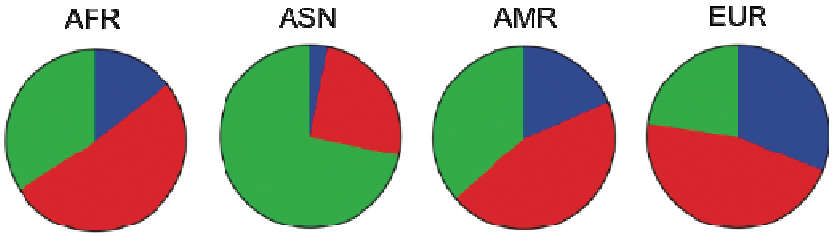
Figure 2.3: Genetic variation at amino acid site 242 in the TRIM22 protein.

(a) Schematic showing the approximate location of nsSNP rs1063303:G>C in the human TRIM22 protein. Inset shows the nucleotide (red) and amino acid changes for nsSNP rs1063303:G>C. C15/18: two cysteine residues required for E3 ligase activity; C97, H100: zinc finger motif in the BB2 domain; predicted CC region; NLS. The 'antiviral patch' refers to several residues previously shown to be a major specificity determinant for TRIM5α-mediated anti-HIV activity [18]. **(b)** Genotype frequency data for nsSNP rs1063303:G>C in African (AFR) (n= 246), American (AMR) (n= 181), Asian (ASN) (n= 286), and European (EUR) (n= 379) 1000 Genomes populations [35]. For each population, the percentage of wild type homozygotes (GG), heterozygotes (GC), and nsSNP homozygotes (CC) are colored green, red, or blue, respectively. **(c)** Sequence alignment of nsSNP rs1063303:G>C in HOM and OWM (left) as well as NWM and additional mammals (right). Amino acid changes at site 242 (location of nsSNP rs1063303:G>C in human TRIM22) are denoted by an asterisk and highlighted in bold.

Figure 2.3



B



Genotype frequencies for SNP rs1063303 (R242T)

Genotype	AFR	ASN	AMR	EUR
C/C	0.146	0.031	0.188	0.311
G/C	0.512	0.252	0.448	0.462
G/G	0.341	0.717	0.365	0.227

C

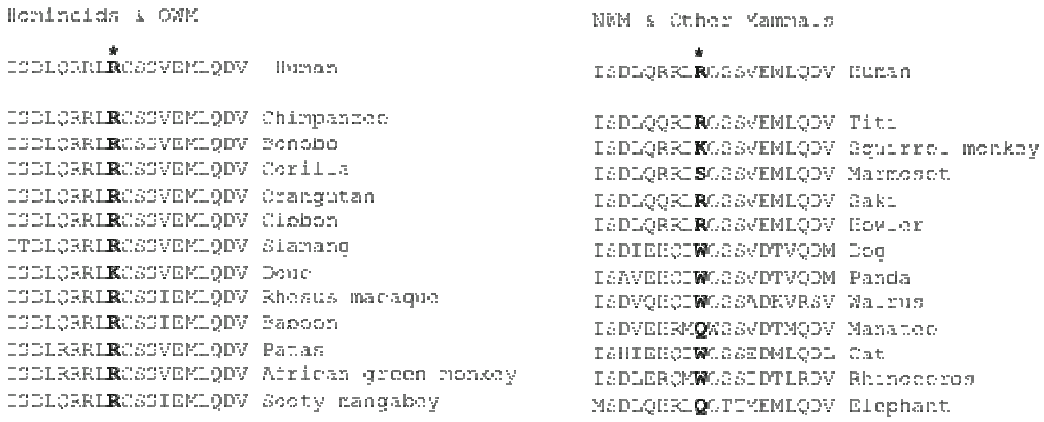


Figure 2.3: Genetic variation at amino acid site 242 in the TRIM22 protein.

will detect a shift in the allele frequency spectrum compared to a neutral Wright-Fisher model. As shown in Table 2.4, Tajima's *D* values were positive for all populations, indicating an excess of intermediate frequency nucleotide site variants compared to the expectation under a neutral model of evolution. These deviations from neutrality were significant (P <0.05) for the AFR, AMR and EUR populations, but not for the ASN population. Similarly, the results of Fu's F_S test also showed largely positive values for AFR, AMR and EUR populations, indicating an excess of intermediate haplotypes over that expected under neutrality. None of the F_S coefficients were considered significant at the 5% level (P <0.02); however, the F_S coefficients for AFR, AMR and EUR populations approached significance.

2.3.3 Functional Analysis of the rs1063303:G>C Polymorphism

To investigate whether nsSNP rs1063303:G>C was deleterious to TRIM22 function, we transfected human cells with an empty vector control plasmid (pEV), a plasmid encoding flag-tagged wild type *TRIM22* (pWT-T22), or a plasmid encoding flag-tagged *TRIM22* with the R242T variation (pSNP-T22). Total RNA was isolated 24 hours post-transfection, reverse-transcribed into cDNA and subjected to quantitative polymerase chain reaction (qPCR). Surprisingly, cells transfected with pSNP-T22 exhibited an average 40-fold increase in *TRIM22* mRNA compared to cells transfected with pWT-T22 (p= 0.0001, unpaired Student's t test) (Fig 2.4a). Whole cell lysates from similarly transfected cells were subjected to Western blot analysis using anti-flag to analyze TRIM22 levels. Densitometric analysis of Western blots after normalization to β-actin levels and transfection efficiency revealed that TRIM22 protein levels were 10.3 fold higher in cells transfected with pSNP-T22 compared to cells transfected with pWT-T22 (Fig 2.4b). We also examined the effect of nsSNP rs1063303:G>C on the sub-cellular localization of TRIM22 using confocal microscopy. Consistent with previous reports, WT-T22 protein localized predominantly in the nucleus and formed punctate bodies [25,50,51] (Fig 2.4c). In contrast, SNP-T22 protein localized diffusely in both the cytoplasm and the nucleus.

Table 2.4

Table 2.4: Results of Tajima's *D* and Fu's F_S neutrality tests.

Population	**Tajima's *D***	**p-value**	**Fu's F_S**	**p-value**
African (AFR)	2.56	0.028	2.06	0.050
Asian (ASN)	0.82	0.251	0.29	0.439
American (AMR)	2.49	0.031	1.83	0.065
European (EUR)	2.79	0.017	2.51	0.028

Figure 2.4: nsSNP rs1063303:G>C alters TRIM22 expression and antiviral activity.

(a) Total RNA was harvested from HeLa cells transfected with empty vector control plasmid (pEV), plasmid expressing wild type TRIM22 (pWT-T22) (rs1063303:G), or TRIM22 plasmid containing nsSNP rs1063303:G>C (pSNP-T22). RNA was reverse transcribed into cDNA and quantified using qPCR. Data shown as the fold change relative to wild type TRIM22 mRNA levels after normalization to β-actin levels. P= 0.0001 unpaired Student's t test. **(b)** HeLa or HOS cells were transfected with equivalent amounts of one of the following plasmids: pEV, flag-tagged pWT-T22, or flag-tagged pSNP-T22. TRIM22 protein was detected via Western blotting using an anti-flag antibody. β-actin was used as a loading control. **(c)** Confocal immunofluorescence microscopy of HeLa cells transiently transfected with pEV, pWT-T22 or pSNP-T22. HeLa cells were fixed 24 hours post-transfection and TRIM22 localization was detected using an anti-flag antibody (green). Images shown represent optical slices taken through the center of cells from a series of z-stack images with a 63x Plan-Apochromat oil immersion objective lens (Numerical Aperture = 1.4). DAPI staining was used to visualize the nuclei (blue). Scale bars = 10 μm. **(d)** Lysates from HeLa cells co-expressing the full-length replication-competent HIV-1 plasmid (pR9) and increasing concentrations of either pWT-T22 or pSNP-T22 were resolved by SDS-PAGE and subjected to Western blotting using anti-p24CA or anti-β-actin (β-actin was used as a loading control. Data shown are representative of at least three independent experiments.

Figure 2.4

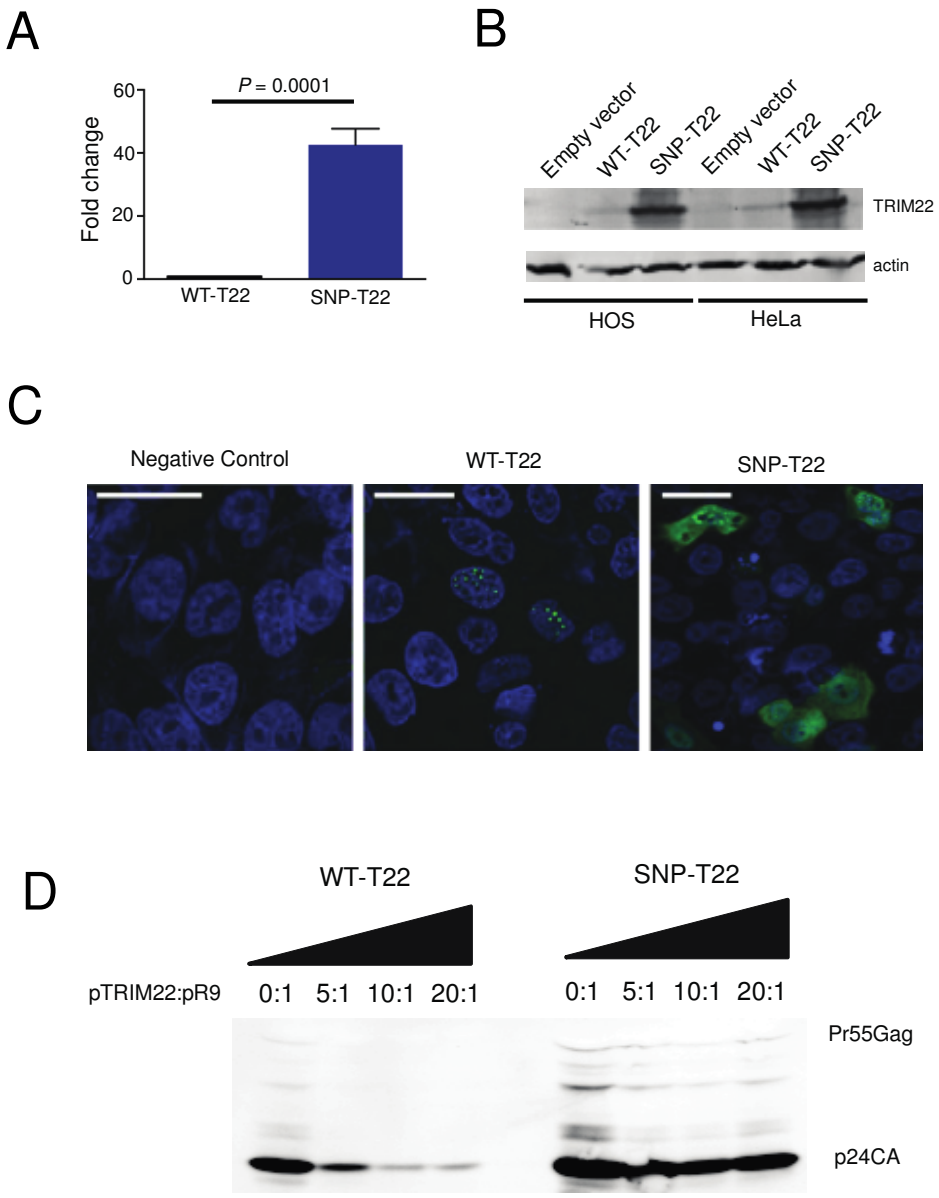


Figure 2.4: nsSNP rs1063303:G>C alters TRIM22 expression and antiviral activity.

Previous studies have demonstrated that TRIM22 can inhibit HIV-1 particle production in human cells [20–23]. To assess the impact of nsSNP rs1063303:G>C on TRIM22's antiviral activity, human osteosarcoma (HOS) cells or HeLa cells were co-transfected with a plasmid encoding replication-competent proviral HIV-1 (pR9) and either pWT-T22 or pSNP-T22. HIV-1 Gag is expressed as the Pr55Gag polyprotein, which includes the matrix (p17MA), capsid (p24CA) and nucleocapsid (p7NC) proteins. Pr55Gag is proteolytically cleaved into MA, CA and NC proteins during the assembly/budding process. Intracellular Gag protein levels were measured by quantitative Western blotting using anti-p24CA. Cells expressing increasing concentrations of WT-T22 exhibited decreasing levels of intracellular HIV-1 Gag protein expression. Cells expressing increasing concentrations of SNP-T22 did not exhibit a substantial reduction in intracellular Gag protein production (Fig 2.4d). Collectively, these data show that the rs1063303:G>C variant had an inverse functional impact where it increased TRIM22 expression and decreased the antiviral activity of TRIM22.

2.4 Discussion

Recent studies on host restriction factors have used an evolutionary approach to identify amino acid residues that are required for their antiviral activity. These studies exploit the evolutionary 'arms race' that occurs between host restriction factors and viruses as they compete to gain an evolutionary advantage over each other. As a result of this evolutionary 'battle', many host restriction factors contain genetic signatures of positive selection, particularly at amino acid sites that interact with viral antagonists [13–15]. Here, we used a similar evolutionary approach to analyze *TRIM22* sequences spanning >100 million years of evolution. In addition to the positively selected residues previously identified in the coiled-coil and B30.2 domains [12], we identified positively selected residues in several other domains of TRIM22. We also showed that the human *TRIM22* gene contains multiple nsSNPs with the potential to alter TRIM22 function and identified a highly prevalent nsSNP in *TRIM22* that alters its expression and antiviral activity against HIV-1 replication.

Positively selected residues are often located near protein binding sites and tend to be solvent-exposed. In addition, these residues typically occur in clusters with other positively selected residues [52]. Although the tertiary structure of TRIM22 has not yet been resolved, its primary amino acid sequence contains a number of putative protein binding motifs. For example, RING domains with E3 ligase activity bind E2 conjugating enzymes via their zinc finger motifs to mediate ubiquitin transfer to a substrate protein [45]. TRIM22 contains a zinc finger motif in its RING domain that facilitates E3 ligase activity in combination with the E2 enzyme UbcH5b [9]. TRIM22 also contains a second zinc finger motif in its BB2 domain, a bipartite nuclear localization signal in its Spacer 2 domain, and a short sequence in its B30.2 domain that is required for its nuclear localization (amino acids 491-494) [46,47]. Remarkably, 13 of the 28 positively selected residues that we identified in TRIM22 are located within 15 amino acids of one of these motifs (Figure 2.1c). Most are not located directly within the motifs, but instead cluster around them in groups of three or four. Although in TRIM22 none of these motifs have been explicitly shown to interact with other proteins, this type of evolutionary pattern suggests that they may indeed function as protein binding sites.

We identified a number of positively selected residues in the B30.2 domain of TRIM22 that correspond in location and spacing to positively selected residues in TRIM5α. In addition, similar to TRIM5α, we showed that the majority of these residues are located in one of four variable regions (v1-v4). In rhesus monkey TRIM5α, v1-v4 form flexible loops that map to the structurally divergent face of the protein [53]. Multiple residues within TRIM5α v1-v4 regions are critical for virus restriction, including several positively selected sites. For example, residues 324 and 332 in the v1 region and residues 385 and 389 in the v2 region are required for TRIM5α-mediated inhibition of HIV-1 and/or SIV in hominoids [18,41,42]. Moreover, in the v3 region, residues 409 and 410 are required for N-MLV restriction by human TRIM5α [43]. Unlike TRIM5α, there are no residues evolving under positive selection in the v3 region of TRIM22. Thus, it is possible that this region of TRIM22 is not subject to strong evolutionary pressures (e.g. from viral antagonists). Consistent with this, TRIM22 does not inhibit N-MLV replication in human cells [23]. Similar to TRIM5α, TRIM22 has been shown to interact with the HIV-1 Gag and/or capsid protein [23,54,55]. Because the v1-v4 regions of rhesus monkey TRIM5α form an

extensive HIV-1 capsid binding interface, it is possible that these regions are also important for TRIM22 and HIV-1 Gag/capsid binding. It will be interesting to learn whether the v1, v2, and v4 regions of TRIM22 are targeted by an HIV-1 antagonist, thereby placing evolutionary pressure on TRIM22 to, for example, maintain its interaction with the HIV-1 Gag protein.

Genes that evolve under positive selection during interspecies evolution also tend to be highly polymorphic in humans [56]. We demonstrate here that the human *TRIM22* gene contains multiple nsSNPs and that many of these nsSNPs are predicted to be deleterious to protein function. We characterized one potentially deleterious nsSNP (rs1063303:G>C) that is highly prevalent in the human population and show that it is located at an amino acid site in TRIM22 that has undergone strong positive selection in mammals. Of interest, the frequency of nsSNP rs1063303:G>C varies considerably among different ethnic populations. Ethnic differences in nsSNP frequencies have been reported for many genes, including several nsSNPs in toll-like receptor genes and a number nsSNPs associated with autoimmune disease [57,58]. It is possible that the differences in nsSNP rs1063303:G>C frequency are due to differential prevalence of certain diseases in different geographic locations. For example, one nsSNP in the β-globin gene is highly prevalent in regions that are endemic for malaria, but not in non-endemic regions. This nsSNP produces an abnormal version of the β-globin protein called hemoglobin S (HbS), which causes sickle-cell disease in homozygotes (HbSS), but affords protection from malaria in heterozygotes (HbAS) [59]. As such, the 'deleterious' HbS allele is maintained at higher than expected frequencies in certain populations. This type of evolution, whereby polymorphism is maintained in a population because it confers a selective advantage, is referred to as balancing selection [30].

Several host restriction factors have been shown to undergo balancing selection in primates, including human TRIM5α and primate OAS1 [60–62]. Moreover, a recent genome-wide scan of two ethnic populations identified TRIM22 as one of 60 'extreme' genes undergoing balancing selection in humans [49]. Given its high frequency in certain human populations and history of positive selection in mammals, it is possible that nsSNP rs1063303 confers some selective advantage in heterozygotes and is maintained by

balancing selection in these populations. For example, the nsSNP allele may be more advantageous against a virus other than HIV-1 that is particularly prevalent in certain geographic locations. The overall positive values resulting from the Tajima's *D* and Fu's F_S tests suggest that there is an excess of intermediate frequency of the *TRIM22* rs1063303 alleles, which can imply a decrease in population size and/or balancing selection. Together, our analyses reveal a complex and multifaceted scenario for the evolution of nsSNP rs1063303 and the *TRIM22* gene.

Other retroviral restriction factors, such as APOBEC3G and TRIM5α, contain nsSNPs that alter their antiviral activity against HIV-1. For example, the H186R variant in APOBEC3G is strongly associated with $CD4^+$ T-cell decline and accelerated disease progression in African Americans. Interestingly, this association is not present in Caucasian individuals or Europeans [63–65]. A number of TRIM5α nsSNPs, including H43Y, R136Q, and G249D, also correlate with notable differences in HIV-1 acquisition and disease progression [66,67]. We showed here that a TRIM22 clone containing the ancestral allele rs1063303:G potently inhibited particle production of full-length replication-competent HIV-1; whereas a TRIM22 clone containing the derived allele rs1063303:C drastically increased the amount of *TRIM22* mRNA and protein in human cells, altered its sub-cellular localization, and failed to inhibit HIV-1 particle production. This finding contrasts a previous study by Ghezzi and colleagues [68] who showed that a *TRIM22* clone containing the derived allele rs1063303:C inhibited expression of luciferase from a reporter construct containing luciferase under control from the HIV-1 long terminal repeat (LTR) promoter. A likely reason for this difference could be attributed to differences between the two systems where we used full-length HIV-1 and they used only the HIV-1 LTR, implying that other HIV-1 proteins (e.g. antagonists) may affect the antiviral activity of TRIM22. In addition, Ghezzi and colleagues [68] showed that SNP rs1063303 alone was not associated with disease progression, however a TRIM22 haplotype involving specific SNP alleles of rs1063303 and rs7935564 was found more frequently in advanced progressors than in long-term non-progressors. Together, these findings highlight the importance of function-altering *TRIM22* SNPs and haplotypes, warranting further investigation into their clinical significance.

It remains unclear how TRIM22 rs1063303:G>C increases *TRIM22* mRNA and protein levels. Non-synonymous SNPs in the gene coding region are typically expected to alter protein function, expression, conformation or stability. It is possible that the amino acid change associated with rs1063303 may affect TRIM22 protein stability or its ability to undergo self-ubiquitination and proteasomal degradation. TRIM22 has been shown to inhibit expression from the promoters of HIV-1 and hepatitis B virus [20,22,25]. The different TRIM22 SNP rs1063303 variants may exhibit reduced or enhanced TRIM22 activity, thereby affecting gene expression from TRIM22 target promoters such as the HIV-1 LTR and/or the promoter controlling expression of TRIM22 itself. Other explanations for the observed difference in TRIM22 RNA and protein levels include, among others, structural changes in the RNA transcript that can in turn influence splicing, stability or translational regulation. Further studies are required to determine precisely how TRIM22 rs1063303:G>C affects TRIM22 mRNA and protein levels.

Interestingly, *TRIM22* mRNA expression levels have previously been shown to influence HIV infection *in vivo*. In the Centre for the AIDS Programme of Research in South Africa (CAPRISA) study cohort, expression of *TRIM22* mRNA was positively correlated with $CD4^+$ T-cell count and negatively correlated with viral load [69]. Conversely, a study in the Swiss HIV study cohort found that expression of *TRIM22* mRNA was positively correlated with HIV-1 RNA levels at the viral set point [70]. Given the differences in nsSNP rs1063303 frequency among different ethnic populations, it is interesting to note that the CAPRISA cohort is comprised entirely of black African females, whereas the Swiss cohort is comprised entirely of Caucasian individuals. It is tempting to speculate that the high prevalence of nsSNP rs1063303:G>C in Caucasians is at least partially responsible for the increased viral load in these individuals. Further studies that include populations from different geographic regions are needed to determine if and how nsSNP rs1063303 influences HIV-1 infection in these different groups and if other SNPs are involved.

Previous studies on host restriction factors have shown that they are rapidly evolving genes that must remain genetically ‘flexible’ to adapt to changing pathogenic landscapes. Although their flexibility is integral to host survival, it can also be detrimental to protein

function. Our study characterizes the effects of nsSNP rs1063303:G>C on TRIM22 protein function and provides insight into the ancient and extensive genetic variation within the *TRIM22* gene.

2.5 References

1. Carthagena, L. *et al.* Human TRIM gene expression in response to interferons. *PLoS One* **4,** e4894 (2009).

2. Rajsbaum, R., Stoye, J. P. & O'Garra, A. Type I interferon-dependent and -independent expression of tripartite motif proteins in immune cells. *Eur. J. Immunol.* **38,** 619–30 (2008).

3. Ohmine, S. *et al.* The antiviral spectra of TRIM5α orthologues and human TRIM family proteins against lentiviral production. *PLoS One* **6,** e16121 (2011).

4. Uchil, P. D., Quinlan, B. D., Chan, W.-T., Luna, J. M. & Mothes, W. TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathog.* **4,** e16 (2008).

5. Towers, G. J. The control of viral infection by tripartite motif proteins and cyclophilin A. *Retrovirology* **4,** 40 (2007).

6. Kajaste-Rudnitski, A. *et al.* Restriction factors of retroviral replication: the example of Tripartite Motif (TRIM) protein 5 alpha and 22. *Amino Acids* **39,** 1–9 (2010).

7. Sardiello, M., Cairo, S., Fontanella, B., Ballabio, A. & Meroni, G. Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties. *BMC Evol. Biol.* **8,** 225 (2008).

8. Gack, M. U. *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446,** 916–920 (2007).

9. Duan, Z., Gao, B., Xu, W. & Xiong, S. Identification of TRIM22 as a RING finger E3 ubiquitin ligase. *Biochem. Biophys. Res. Commun.* **374,** 502–6 (2008).

10. Meroni, G. & Diez-Roux, G. TRIM/RBCC, a novel class of "single protein RING finger" E3 ubiquitin ligases. *Bioessays* **27,** 1147–57 (2005).

11. Tareen, S. U., Sawyer, S. L., Malik, H. S. & Emerman, M. An expanded clade of rodent Trim5 genes. *Virology* **385,** 473–483 (2009).

12. Sawyer, S. L., Emerman, M. & Malik, H. S. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathog.* **3,** e197 (2007).

13. Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* **2,** E275 (2004).

14. McNatt, M. W. *et al.* Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. *PLoS Pathog.* **5,** e1000300 (2009).

15. Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2832–7 (2005).

16. Diehl, W. E., Johnson, W. E. & Hunter, E. Elevated rate of fixation of endogenous retroviral elements in Haplorhini TRIM5 and TRIM22 genomic sequences: impact on transcriptional regulation. *PLoS One* **8,** e58532 (2013).

17. Stremlau, M., Owens, C. M. & Perron, M. J. The cytoplasmic body component TRIM5 a restricts HIV-1 infection in Old World monkeys. **427,** (2004).

18. Stremlau, M., Perron, M., Welikala, S. & Sodroski, J. Species-Specific Variation in the B30 . 2 (SPRY) Domain of TRIM5α Determines the Potency of Human Immunodeficiency Virus Restriction. **79,** 3139–3145 (2005).

19. Song, B. *et al.* The B30 . 2 (SPRY) Domain of the Retroviral Restriction Factor TRIM5α Exhibits Lineage-Specific Length and Sequence Variation in Primates. *J. Virol.* **79,** 6111–6121 (2005).

20. Tissot, C. & Mechti, N. Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus type 1 long terminal repeat expression. *J. Biol. Chem.* **270,** 14891–8 (1995).

21. Bouazzaoui, A. *et al.* Stimulated trans-acting factor of 50 kDa (Staf50) inhibits HIV-1 replication in human monocyte-derived macrophages. *Virology* **356,** 79–94

22. Kajaste-Rudnitski, A. *et al.* TRIM22 inhibits HIV-1 transcription independently of its E3 ubiquitin ligase activity, Tat, and NF-kappaB-responsive long terminal repeat elements. *J. Virol.* **85,** 5183–96 (2011).

23. Barr, S. D., Smiley, J. R. & Bushman, F. D. The interferon response inhibits HIV particle production by induction of TRIM22. *PLoS Pathog.* **4,** e1000007 (2008).

24. Eldin, P. *et al.* TRIM22 E3 ubiquitin ligase activity is required to mediate antiviral activity against encephalomyocarditis virus. *J. Gen. Virol.* **90,** 536–45 (2009).

25. Gao, B., Duan, Z., Xu, W. & Xiong, S. Tripartite motif-containing 22 inhibits the activity of hepatitis B virus core promoter, which is dependent on nuclear-located RING domain. *Hepatology* **50,** 424–33 (2009).

26. Di Pietro, A. *et al.* TRIM22 inhibits influenza A virus infection by targeting the viral nucleoprotein for degradation. *J. Virol.* **87,** 4523–33 (2013).

27. Emerman, M. & Malik, H. S. Paleovirology--modern consequences of ancient viruses. *PLoS Biol.* **8,** e1000301 (2010).

28. Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12,** 687–95 (2012).

29. Subramanian, S. & Kumar, S. Higher intensity of purifying selection on >90% of the human genes revealed by the intrinsic replacement mutation rates. *Mol. Biol. Evol.* **23,** 2283–7 (2006).

30. Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437,** 1153–7 (2005).

31. Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23,** 1073–9 (2007).

32. Zhang, H., Gao, S., Lercher, M. J., Hu, S. & Chen, W.-H. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* **40,** W569–72 (2012).

33. Doron-Faigenboim, A., Stern, A., Mayrose, I., Bacharach, E. & Pupko, T. Selecton: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics* **21,** 2101–3 (2005).

34. Stern, A. *et al.* Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res.* **35,** W506–11 (2007).

35. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

36. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–95 (1989).

37. Tajima, F. The effect of change in population size on DNA polymorphism. *Genetics* **123,** 597–601 (1989).

38. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915–925 (1997).

39. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10,** 564–7 (2010).

40. Li, Y., Li, X., Stremlau, M., Lee, M. & Sodroski, J. Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. *J. Virol.* **80,** 6738–44 (2006).

41. Ohkura, S., Yap, M. W., Sheldon, T. & Stoye, J. P. All three variable regions of the TRIM5alpha B30.2 domain can contribute to the specificity of retrovirus restriction. *J. Virol.* **80,** 8554–65 (2006).

42. Kono, K., Bozek, K., Domingues, F. S., Shioda, T. & Nakayama, E. E. Impact of a single amino acid in the variable region 2 of the Old World monkey TRIM5alpha SPRY (B30.2) domain on anti-human immunodeficiency virus type 2 activity. *Virology* **388,** 160–8 (2009).

43. Perron, M. J., Stremlau, M. & Sodroski, J. Two surface-exposed elements of the B30.2/SPRY domain as potency determinants of N-tropic murine leukemia virus restriction by human TRIM5alpha. *J. Virol.* **80,** 5631–6 (2006).

44. Yap, M. W. & Stoye, J. P. A Single Amino Acid Change in the SPRY Domain of Human Trim5 Leads to HIV-1 Restriction. **15,** 73–78 (2005).

45. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.* **32,** 63–70 (2007).

46. Herr, A.-M., Dressel, R. & Walter, L. Different subcellular localisations of TRIM22 suggest species-specific function. *Immunogenetics* **61,** 271–80 (2009).

47. Sivaramakrishnan, G., Sun, Y., Rajmohan, R. & Lin, V. C. L. B30.2/SPRY domain in tripartite motif-containing 22 is essential for the formation of distinct nuclear bodies. *FEBS Lett.* **583,** 2093–9 (2009).

48. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–81 (2009).

49. Andrés, A. M. *et al.* Targets of balancing selection in the human genome. *Mol. Biol. Evol.* **26,** 2755–64 (2009).

50. Sivaramakrishnan, G., Sun, Y., Rajmohan, R. & Lin, V. C. L. *B30.2/SPRY domain in tripartite motif-containing 22 is essential for the formation of distinct nuclear bodies*. *FEBS Lett.* **583,** 2093–2099 (2009).

51. Sivaramakrishnan, G., Sun, Y., Tan, S. K. & Lin, V. C. L. Dynamic localization of tripartite motif-containing 22 in nuclear and nucleolar bodies. *Exp. Cell Res.* **315,** 1521–1532 (2009).

52. Lin, Y.-S., Hsu, W.-L., Hwang, J.-K. & Li, W.-H. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol. Biol. Evol.* **24,** 1005–11 (2007).

53. Biris, N. *et al.* Structure of the rhesus monkey TRIM5α PRYSPRY domain, the HIV capsid recognition module. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13278–13283 (2012).

54. Sebastian, S. & Luban, J. TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. *Retrovirology* **2,** 40 (2005).

55. Pertel, T. *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472,** 361–365 (2011).

56. Liu, J., Zhang, Y., Lei, X. & Zhang, Z. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. *Genome Biol.* **9,** R69 (2008).

57. Cheng, P.-L., Eng, H.-L., Chou, M.-H., You, H.-L. & Lin, T.-M. Genetic polymorphisms of viral infection-associated Toll-like receptors in Chinese population. *Transl. Res.* **150,** 311–8 (2007).

58. Mori, M., Yamada, R., Kobayashi, K., Kawaida, R. & Yamamoto, K. Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. *J. Hum. Genet.* **50,** 264–6 (2005).

59. Rees, D. C., Williams, T. N. & Gladwin, M. T. Sickle-cell disease. *Lancet* **376,** 2018–31 (2010).

60. Ferguson, W., Dvora, S., Fikes, R. W., Stone, A. C. & Boissinot, S. Long-term balancing selection at the antiviral gene OAS1 in Central African chimpanzees. *Mol. Biol. Evol.* **29,** 1093–103 (2012).

61. Alex A. Compton, Vanessa M. Hirsch, and M. E. The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe.* **11,** 91–98 (2012).

62. Newman, R. M. *et al.* Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 19134–9 (2006).

63. An, P. *et al.* APOBEC3G Genetic Variants and Their Influence on the Progression to AIDS. *J. Virol.* **78,** 11070–11076 (2004).

64. Kavidha Reddy1, Cheryl Winkler2, Lise Werner3, Koleka Mlisana3, Salim Abdool Karim3, Thumbi Ndung'u1, 3,*, and the C. A. I. S. T. APOBEC3G expression is dysregulated in primary HIV-1 infection and a polymorphic variant influences CD4+ T-cell counts and plasma viral load. *AIDS* **24,** 195–204 (2010).

65. Do, H. *et al.* Exhaustive genotyping of the CEM15 (APOBEC3G) gene and absence of association with AIDS progression in a French cohort. *J. Infect. Dis.* **191,** 159–63 (2005).

66. Sobieszczyk, M. E., Lingappa, J. R. & McElrath, M. J. Host genetic polymorphisms associated with innate immune factors and HIV-1. *Curr. Opin. HIV AIDS* **6,** 427–34 (2011).

67. Sawyer, S. L., Wu, L. I., Akey, J. M., Emerman, M. & Malik, H. S. High-frequency persistence of an impaired allele of the retroviral defense gene TRIM5alpha in humans. *Curr. Biol.* **16,** 95–100 (2006).

68. Ghezzi, S. *et al.* Identification of TRIM22 single nucleotide polymorphisms associated with loss of inhibition of HIV-1 transcription and advanced HIV-1 disease. *AIDS* **27,** 2335–44 (2013).

69. Singh, R. *et al.* Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. *J. Virol.* **85,** 208–16 (2011).

70. Rotger, M. *et al.* Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS Pathog.* **6,** e1000781 (2010).

Chapter 3

3 *In silico* analysis of functional single nucleotide polymorphisms in the human *TRIM22* gene [2]

TRIM22 is an evolutionarily ancient protein that plays an integral role in the host innate immune response to viruses. The antiviral TRIM22 protein has been shown to inhibit the replication of a number of viruses, including HIV-1, HBV and IAV. TRIM22 expression has also been associated with multiple sclerosis, cancer, and autoimmune disease. In this study, multiple *in silico* computational methods were used to identify nsSNPs that are deleterious to TRIM22 structure and/or function. A sequence homology-based approach was adopted for screening nsSNPs in TRIM22, including six different *in silico* prediction algorithms and evolutionary conservation data from the ConSurf web server. In total, 14 high-risk deleterious nsSNPs were identified in TRIM22, most of which are located in a protein-protein interaction module called the B30.2 domain. Additionally, 9 of the top high-risk deleterious nsSNPs altered the putative structure of TRIM22's B30.2 domain, particularly in the surface-exposed v2 and v3 regions. These same regions are critical for retroviral restriction by the closely-related TRIM5α protein. A number of putative structural and functional residues, including several sites that undergo post-translational modification (PTM), were also identified in TRIM22. This study is the first extensive *in silico* analysis of the highly polymorphic *TRIM22* gene and will be a valuable resource for future targeted mechanistic and population-based studies.

---

[2] The material contained in this chapter was published in: Kelly, JN and Barr, SD. *In silico* analysis of functional single nucleotide polymorphisms in the human *TRIM22* gene. *PLoS One*. 9 (7): e101436 (2014).

3.1 Introduction

SNPs, defined as single base changes in a DNA sequence, are responsible for the majority of genetic variation in the human population. Although many SNPs are phenotypically neutral, nsSNPs often have deleterious effects on protein structure or function. nsSNPs are located in protein coding regions and generate an amino acid substitution in their corresponding protein product. As such, nsSNPs can alter the structure, stability, and/or function of proteins, and are often associated with human disease. Indeed, previous studies have shown that approximately 50% of the mutations involved in inherited genetic disorders are due to nsSNPs [1–3]. Recently, a number of genetic studies have focused on nsSNPs in innate immune genes. These studies have identified multiple nsSNPs that influence susceptibility to infection, as well as the development of inflammatory disorders and autoimmune diseases [4–9]. Nonetheless, because innate immune genes are often highly polymorphic, many nsSNPs in these genes remain uncharacterized.

Members of the TRIM protein family are involved in a wide range of biological processes related to innate immunity [10–12]. TRIM proteins are defined by their RBCC motif, which consists of a RING domain, one or two BB domains, and a predicted CC region. Most TRIM proteins also contain a protein-protein interaction module called a B30.2 domain at their C-terminus [13–15]. Many TRIM proteins are induced by IFN signaling and several possess antiviral activity, in particular against the *Retroviridae* family of viruses. Recent studies have implicated TRIM proteins in the regulation of pathogen-recognition and important immune signaling pathways, a finding that has sparked considerable interest in understanding how TRIM family proteins contribute to the innate immune response [16–21].

One well-studied member of the TRIM family, TRIM5α, is required for the species-specific block against HIV-1 replication in primate cells [22–24]. Recently, TRIM5α was also shown to promote innate immune signaling and to function as an innate immune sensor for the retrovirus capsid lattice *in vitro*. Previous studies have established that TRIM5α binds to the HIV-1 capsid protein in the mature viral core via four variable regions (v1-v4) in its B30.2 domain [25,26]. The v1 or 'antiviral patch' region was

previously shown to be the major determinant for species-specific HIV-1 restriction by TRIM5α. Mutations in the other variable regions (v2-v4) have also been shown to interfere with TRIM5α-mediated restriction of HIV-1, SIV, and/or N-MLV [22,26–29]. Notably, analogous variable regions are found in several other B30.2-containing TRIM proteins [30–32].

Human *TRIM5* is located on chromosome 11 within a cluster of four closely-related *TRIM* genes that also includes *TRIM6*, *TRIM22*, and *TRIM34*. *TRIM5* and *TRIM22* have an ancient and dynamic evolutionary relationship, whereby both genes have evolved under positive selection for millions of years in a mutually exclusive manner [33]. Similar to TRIM5α, TRIM22 has also been shown to inhibit HIV-1 replication in a number of human cell lines and primary monocyte-derived macrophages [34–37]. TRIM22 expression levels have also been shown to influence HIV-1 infection *in vivo* [38–40]. Interestingly, nsSNPs in TRIM5α, including H43Y, R136Q, and G249D, significantly alter HIV-1 acquisition and disease progression in humans [41–44]. Despite TRIM22's highly polymorphic nature, it is unknown how nsSNPs affect its biological and/or antiviral functions. Here, multiple *in silico* computational methods were used to identify nsSNPs in the *TRIM22* gene that are predicted to be highly deleterious to TRIM22 structure and/or function. A total of 14 high-risk deleterious nsSNPs were identified, including 9 that altered the putative structure of TRIM22's B30.2 domain. A number of sites predicted to undergo PTM (ubiquitylation, sumoylation, phosphorylation) were also identified. This is the first extensive *in silico* analysis of the *TRIM22* gene and will establish a strong foundation for future structure-function and population-based studies.

3.2 Materials and methods

3.2.1 Retrieval of nsSNP data

Polymorphism data for the *TRIM22* gene was retrieved from the following databases: UniProt (http://www.uniprot.org), NCBI dbSNP (https://www.ncbi.nlm.nih.gov/SNP/), 1000 Genomes (http://www.1000genomes.org/), and the Ensembl genome browser (http://www.ensembl.org/index.html) [45–47].

3.2.2 *In silico* nsSNP analysis

Functional effects of nsSNPs were predicted using the following *in silico* algorithms: Polymorphism Phenotyping v 2 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pp2) [48], Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org/) [49], nsSNP Analyzer (http://snpanalyzer.uthsc.edu/) [50], Predictor of Human Deleterious Single Nucleotide Polymorphisms (PhD-SNP) (http://snps.biofold.org/phd-snp/phd-snp.html) [51], SNPs and Gene Ontology v 3 (SNPs&GO) (http://snps-and-go.biocomp.unibo.it/snps-and-go/) [52], and PMut (mmb2.pcb.ub.es:8080/PMut) [53]. nsSNPs predicted to be deleterious by at least 4 *in silico* algorithms were categorized as high-risk nsSNPs and were selected for further analysis.

3.2.3 Phylogenetic analysis

Evolutionary conservation of amino acid residues in TRIM22 was determined using the ConSurf web server (consurf.tau.ac.il/) [54]. In ConSurf, 14 TRIM22 homologues were aligned and position-specific conservation scores were calculated using an empirical Bayesian algorithm (Conservation Scores: 1-4 Variable, 5-6 Intermediate, and 7-9 Conserved). Putative functional and structural residues were also predicted using ConSurf by combining evolutionary conservation scores with solvent accessibility predictions (Fig 3.1). Highly conserved amino acids that were located at high-risk nsSNP sites were selected for further analysis.

3.2.4 Comparative molecular modeling

3D-Jigsaw (v 2.0) was used to generate 3D structural models for the B30.2 domain of wild type TRIM22 (UniProtKB Q8IYM9) and the 9 high-risk nsSNPs. For each model, only the B30.2 domain sequence was submitted. 3D-Jigsaw searches multiple sequence databases (e.g. PFAM and PDB) and builds structures based on homologues of known structure [55]. Models were viewed using the Swiss-PdbViewer (v 4.1) [56]. Tm-Align was used to calculate Tm-scores and root mean square deviation (RMSD) [57]. The total energy after minimization (TEM) was calculated via the L-BFGF method using the NOMAD-Ref Gromacs server (http://lorentz.immstr.pasteur.fr/nomad-ref.php) [58].

3.2.5 Prediction of post-translational modification sites

Putative ubiquitylation sites were predicted using the UbPred (www.ubpred.org) and BDM-PUB (bdmpub.biocuckoo.org) programs [2]. For UbPred, lysine residues with a score of ≥ 0.62 were considered ubiquitylated. For BDM-PUB, the balanced cut-off option was selected. Putative sumoylation sites were predicted using the SUMOplot (http://www.abgent.com/sumoplot) and SUMOsp v 2.0 (http://sumosp.biocuckoo.org/) programs [59]. For SUMOplot, only high probability motifs with a score > 0.5 were considered sumoylated. Medium level threshold with a 2.64 cut-off value was selected for SUMOsp 2.0 analysis. Putative phosphorylation sites were predicted using GPS 2.1 (http://gps.biocuckoo.org/) and NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) [60,61]. For GPS 2.1 analysis, high level threshold with cut-off values ranging from 0.776-11 were selected. For NetPhos 2.0, serine, threonine, and tyrosine residues with a score of > 0.5 were considered phosphorylated. Sumo-interacting motifs (SIM) were identified manually and compared to experimentally verified SIMs in the scientific literature [62,63].

3.2.6 Protein stability analysis

I-Mutant version 2.0, an online support vector machine tool based on the ProTherm database, was used to evaluate nsSNP-induced changes in protein stability [64]. nsSNP protein-coding sequences were submitted to I-Mutant 2.0 for 2 high-risk nsSNPs that coincide with putative PTM sites, 5 low-risk nsSNPs that coincide with putative PTM sites, and 12 additional high-risk nsSNPs that do not coincide with predicted PTM sites. I-Mutant 2.0 estimates the free energy change value (DDG) by calculating the unfolding Gibbs free energy value (ΔG) for the wild type protein and subtracting it from that of the mutant protein (DDG or ΔΔG = ΔG mutant – ΔG wild type). It also predicts the sign (increase or decrease) of the free energy change value (DDG), along with a reliability index for the results (RI: 0-10, where 0 is the lowest reliability and 10 is the highest reliability). A DDG < 0 corresponds to a decrease in protein stability, whereas a DDG > 0 corresponds to an increase in protein stability. However, according to the ternary classification system (SVM3), a large decrease in protein stability corresponds to a DDG < -0.5 and a large increase in protein stability corresponds to a DDG > 0.5. In contrast, DDG values that fall between -0.5 and 0.5 correspond to relatively neutral protein

stability [64,65]. The pH was set to 7 and the temperature was set to 25°C for all submissions.

3.3 Results and Discussion

3.3.1 SNP dataset

SNP data for the *TRIM22* gene was retrieved from the NCBI dbSNP database, the Ensembl genome browser, and the UniProt database [45–47]. According to these databases, the *TRIM22* gene contains a total of 64 nsSNPs plus 8 SNPs in its 5' UTR and 32 SNPs in its 3' UTR. Of the 64 nsSNPs, 10 generate truncated versions of the TRIM22 protein (nonsense and frameshift mutations), whereas 54 introduce single amino acid changes (missense mutations) into TRIM22 (Table 3.1). To determine whether a given missense mutation affected TRIM22 function, we subjected the latter 54 nsSNPs to multiple *in silico* SNP prediction algorithms. The results, which are summarized in Table 3.2, identified a number of nsSNPs with a high probability of being deleterious to TRIM22 structure and/or function.

3.3.2 *In silico* nsSNP analysis

Our analyses included the following six *in silico* SNP prediction algorithms: Polyphen-2, SIFT, nsSNP Analyzer, PhD-SNP, PMUT, and SNPs&GO [48–52,66]. According to our Polyphen-2 results, 13 nsSNPs (23%) are damaging to TRIM22 function, whereas 33 nsSNPs (59%) are benign. An additional 10 nsSNPs (18%) are predicted to be 'possibly damaging' by Polyphen-2 (Table 3.2). Our SIFT analysis predicted that 19 nsSNPs (34%) are deleterious to TRIM22 function and 37 nsSNPs (66%) are tolerated. On the contrary, the nsSNP Analyzer predicted that 21 nsSNPs (38%) cause disease and 35 nsSNPs (62%) are neutral (Table 3.2). Both PhD-SNP and PMUT predicted that 25 (45%) nsSNPs are pathological and 31 (55%) nsSNPs are neutral (Table 3.2). SNPs&GO analysis, which

Table 3.1

Table 3.1 *In silico* prediction results for nsSNPs in TRIM22.

nsSNP ID	Mutation [1]	Polyphen-2	SIFT	nsSNP Analyzer	PhD-SNP	PMUT	SNPs&Go	# Del [2]
rs368531868	D2E	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs200816458	V7A	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs372487646	E10D	Benign	Deleterious	Neutral	Neutral	Neutral	Neutral	1
rs375540431	I43M	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs11541920	V47E	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs192306924	T61N	Benign	Deleterious	Neutral	Neutral	Neutral	Neutral	1
rs201847190	L68R	Probably damaging	Deleterious	Disease	Disease	Neutral	Disease	5
rs182619286	R69Q	Possibly damaging	Tolerated	Neutral	Disease	Pathological	Neutral	2.5
rs199625192	H73R	Probably damaging	Deleterious	Disease	Disease	Pathological	Neutral	5
rs200668710	E83K	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs141180305	E90K	Benign	Tolerated	Neutral	Neutral	Pathological	Neutral	1
rs370446835	D95A	Possibly damaging	Tolerated	Neutral	Neutral	Pathological	Neutral	1.5
rs373103298	H99Y	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs371543745	L104F	Probably damaging	Deleterious	Disease	Neutral	Neutral	Neutral	3
rs187628129	Q105K	Benign	Tolerated	Neutral	Neutral	Pathological	Neutral	1
rs368058642	E135K	Possibly damaging	Tolerated	Disease	Disease	Pathological	Disease	4.5
rs200924168	R150K	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs375795798	I152K	Benign	Tolerated	Neutral	Disease	Pathological	Neutral	2
rs7935564	D155N	Benign	Deleterious	Disease	Neutral	Neutral	Neutral	2
rs201531661	K185Q	Benign	Tolerated	Neutral	Neutral	Pathological	Neutral	1

rs368924880	N196S	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs2291843	T232A	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs374292901	I234N	Probably damaging	Deleterious	Disease	Disease	Pathological	Neutral	5
rs201523218	R239Q	Benign	Tolerated	Neutral	Disease	Pathological	Neutral	2
rs1063303	R242T	Benign	Tolerated	Disease	Neutral	Pathological	Neutral	2
rs61735273	S244L	Probably damaging	Deleterious	Disease	Disease	Neutral	Disease	5
rs370736499	I253T	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs143605305	V255I	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs112606816	L264M	Possibly damaging	Tolerated	Neutral	Neutral	Neutral	Neutral	0.5
rs181298463	R279Q	Benign	Tolerated	Neutral	Neutral	Pathological	Neutral	1
rs73404240	T294K	Benign	Tolerated	Disease	Neutral	Pathological	Neutral	2
rs368682946	A315V	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs12364019	R321K	Possibly damaging	Deleterious	Neutral	Neutral	Neutral	Neutral	1.5
rs372042006	R327C	Benign	Tolerated	Neutral	Disease	Pathological	Neutral	2
rs75023388	R327H	Benign	Tolerated	Neutral	Neutral	Pathological	Neutral	1
rs201494620	T330I	Benign	Tolerated	Neutral	Disease	Neutral	Neutral	1
rs368220166	K332N	Benign	Tolerated	Neutral	Disease	Neutral	Neutral	1
rs199987600	A341T	Benign	Tolerated	Neutral	Neutral	Neutral	Neutral	0
rs371728648	G346S	Probably damaging	Deleterious	Disease	Disease	Pathological	Neutral	5
rs200243523	V359I	Possibly damaging	Tolerated	Disease	Disease	Neutral	Disease	3.5
rs191847788	K364N	Possibly damaging	Deleterious	Disease	Disease	Neutral	Disease	4.5
rs375595000	P403T	Probably damaging	Deleterious	Disease	Disease	Pathological	Disease	6
rs369734227	I410V	Possibly damaging	Tolerated	Disease	Neutral	Neutral	Neutral	1.5
rs150095329	T415I	Benign	Tolerated	Disease	Neutral	Neutral	Neutral	1
rs370495523	L432W	Probably damaging	Deleterious	Disease	Neutral	Pathological	Neutral	4
rs200915295	F435C	Benign	Deleterious	Neutral	Disease	Neutral	Neutral	2
rs187416296	R442C	Probably damaging	Tolerated	Disease	Disease	Pathological	Disease	5

rs370420711	A451V	Possibly damaging	Deleterious	Neutral	Disease	Pathological	Neutral	3.5
rs377529439	F456I	Probably damaging	Deleterious	Disease	Disease	Neutral	Disease	5
rs371028900	T460I	Probably damaging	Deleterious	Disease	Disease	Pathological	Disease	6
rs371139090	G471E	Benign	Tolerated	Neutral	Disease	Neutral	Neutral	1
rs138529937	C472S	Probably damaging	Tolerated	Neutral	Neutral	Pathological	Neutral	2
rs368256788	R473H	Benign	Tolerated	Neutral	Disease	Neutral	Neutral	1
rs267603016	P480S	Probably damaging	Deleterious	Disease	Disease	Pathological	Disease	6
rs200638791	P484S	Possibly damaging	Deleterious	Neutral	Disease	Pathological	Disease	4.5
rs61735327	M491I	Benign	Tolerated	Disease	Disease	Pathological	Neutral	3
rs200148337	C494F	Probably damaging	Deleterious	Disease	Disease	Pathological	Disease	6

[1] Mutation: Wild type residue/amino acid position/nsSNP residue.
[2] # Del: Total number of deleterious predictions; A score of 0.5 was assigned to amino acids that were predicted to be ‘possibly damaging’ by Polyphen-2.

Table 3.2

Table 3.2 Summary of *in silico* prediction results for all nsSNPs in TRIM22.

Prediction	**Total number of nsSNPs (%)** [1]					
	Polyphen-2	**SIFT**	**nsSNP Analyzer**	**PhD-SNP**	**PMUT**	**SNPs&GO**
Deleterious	13 (23)	19 (34)	-	-	-	-
PD [2]	10 (18)	-	-	-	-	-
Benign	33 (59)	37 (66)	-	-	-	-
Disease	-	-	21 (38)	25 (45)	25 (45)	11 (20)
Neutral	-	-	35 (62)	31 (55)	31 (55)	45 (80)

[1] Percentage of total nsSNPs (56) are shown in parentheses for each prediction program.
[2] PD: nsSNPs predicted to be 'possibly deleterious' by Polyphen-2

includes information from the Gene Ontology annotation, predicted that 11 nsSNPs (20%) cause disease and 45 nsSNPs (80%) are neutral (Table 3.2). Interestingly, we found that the majority of potentially deleterious nsSNPs were located in the B30.2 domain, including 3 nsSNPs that were predicted to be damaging by all six SNP prediction algorithms (P403T, T460I, and C494F). Because each algorithm uses different parameters to evaluate the nsSNPs, nsSNPs with more positive results are more likely to be truly deleterious. Here, we classified nsSNPs as high-risk if they were predicted to be deleterious by four or more SNP prediction algorithms. 14 nsSNPs met this criteria and were selected for further analysis (Table 3.3).

3.3.3 Conservation Profile of High-Risk nsSNPs

Amino acids that are involved in important biological processes, such as those located in enzymatic sites or required for protein-protein interactions, tend to be more conserved than other residues. As such, nsSNPs that are located at highly conserved amino acid positions tend to be more deleterious than nsSNPs that are located at non-conversed sites [3,67]. To further investigate the potential effects of the 14 high-risk nsSNPs in Table 3.3, we calculated the degree of evolutionary conservation at all amino acid sites in the TRIM22 protein using the ConSurf web server. ConSurf employs an empirical Bayesian method to determine evolutionary conservation and identify putative structural and functional residues [54]. For the purpose of this study, we focused on amino acid sites that coincide in location with the 14 high-risk nsSNPs; however, ConSurf also identified a number of other residues that may be functionally relevant (Fig 3.1).

ConSurf analysis revealed that residues L68, H73, E135, I234, S244, G346, K364, P403, L432, R442, F456, T460, and C494 are highly conserved (Conservation Score of 7-9). In addition, ConSurf predicted that T460 was an important structural residue (highly conserved and buried) and that L68, K364, and P403 were important functional residues (highly conserved and exposed) (Table 3.4). To identify putative structural and functional sites, ConSurf combines evolutionary conservation data with solvent accessibility predictions. Highly conserved residues are predicted to be either structural or functional based on their location relative to the protein surface or protein core [68]. Remarkably, two of the three high-risk nsSNPs that were predicted to be deleterious by all six SNP

Table 3.3

Table 3.3: TRIM22 nsSNPs predicted to be functionally significant by four or more SNP prediction algorithms.

nsSNP ID	Mutation	Domain	MAF	# Del [1]
rs201847190	L68R	Spacer 1	N/A	5
rs199625192	H73R	Spacer 1	0.0005	5
rs368058642	E135K	Coiled-coil	0.0001	4.5
rs374292901	I234N	Spacer 2	0.0001	5
rs61735273	S244L	Spacer 2	0.0354	5
rs371728648	G346S	B30.2	0.0001	5
rs191847788	K364N	B30.2	0.0005	4.5
rs375595000	P403T	B30.2	0.0001	6
rs370495523	L432W	B30.2	0.0001	4
rs187416296	R442C	B30.2	0.0041	5
rs377529439	F456I	B30.2	0.0001	5
rs371028900	T460I	B30.2	0.0003	6
rs200638791	P484S	B30.2	0.0008	4.5
rs200148337	C494F	B30.2	N/A	6

[1] # Del: number of deleterious predictions.

Figure 3.1: ConSurf analysis of amino acid residues in the TRIM22 protein.

Schematic showing ConSurf results for the human TRIM22 protein. Amino acids were ranked on a conservation scale of 1-9 and are highlighted as follows: blue residues (1-4) are variable, white residues (5) are average, and purple residues (6-9) are conserved. Residues predicted to be exposed to the surface of the protein are indicated via an orange letter ‘e’, while residues predicted to be buried are indicated via a green letter ‘b’. Putative structural residues are demarcated with a blue letter ‘s’ (highly conserved and buried), whereas putative functional residues are demarcated with a red letter ‘f’ (highly conserved and exposed).

Figure 3.1

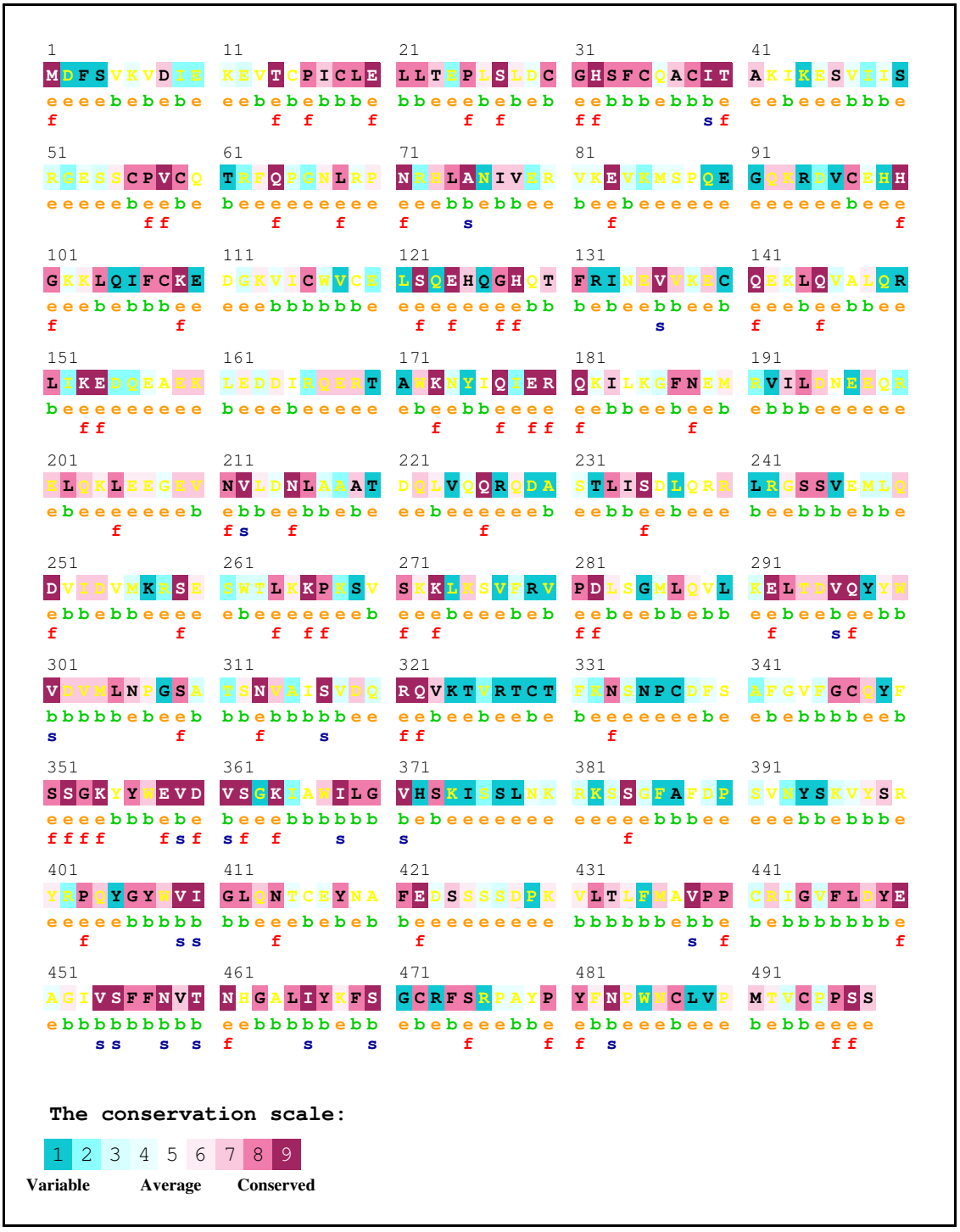


Figure 3.1: ConSurf analysis of amino acid residues in the TRIM22 protein.

Table 3.4

Table 3.4 Conservation profile of amino acids in TRIM22 that coincide with the location of high-risk nsSNPs.

nsSNP ID	Amino Acid	CS [1]	ConSurf prediction
rs201847190	L68	8	Highly conserved and exposed (f [2])
rs199625192	H73	7	Exposed
rs368058642	E135	7	Exposed
rs374292901	I234	7	Buried
rs61735273	S244	8	Buried
rs371728648	G346	8	Buried
rs191847788	K364	9	Highly conserved and exposed (f [2])
rs375595000	P403	8	Highly conserved and exposed (f [2])
rs370495523	L432	8	Buried
rs187416296	R442	7	Exposed
rs377529439	F456	8	Buried
rs371028900	T460	9	Highly conserved and buried (s [3])
rs200638791	P484	6	Exposed
rs200148337	C494	8	Buried

[1] CS: Conservation score (1-4= variable, 5-6= intermediate, 7-9= conserved).
[2] f: predicted functional residue
[3] s: predicted structural residue

prediction algorithms (P403T and T460I) were also identified as important structural or functional residues by ConSurf (Table 3.1, Table 3.3). Taken together, our data strongly suggest that the nsSNPs P403T and T460I are deleterious to TRIM22 structure and/or function.

3.3.4 Comparative Modeling of High-Risk nsSNPs

To examine whether P403T and T460I altered the 3D structure of TRIM22's B30.2 domain, we individually substituted each nsSNP into the wild type TRIM22 sequence and submitted the sequences to 3D-Jigsaw for structural analysis. We also submitted sequences for the remaining 7 high-risk nsSNPs in the B30.2 domain (i.e. G346S, K364N, L432W, R442C, F456I, P484S, and C494F) since our *in silico* and ConSurf results indicated that these nsSNPs were also highly likely to be deleterious. Theoretical structural models were generated for each nsSNP using the 3D-Jigsaw program, which constructs 3D models for proteins based on homologues of known structure [55]. We then used Swiss-PdbViewer to compare each nsSNP model to the predicted 3D-Jigsaw model of wild type TRIM22 [56]. All of the nsSNPs altered the putative 3D structure of wild type TRIM22's B30.2 domain. G346S, P403T, L432W, F456I, and C494F introduced an alpha helix into the v2 region, whereas the other 4 nsSNPs introduced beta strands into the v2 region (Fig 3.2). With the exception of P484S, which introduced an alpha helix into the v3 region, all of the nsSNP models contained elongated and/or additional beta strands in the v3 region. Only G346S and F456I altered the v1 region (both introduced an alpha helix); however, all 9 nsSNPs altered the length and/or number of beta strands in non-variable regions of the B30.2 domain. Notably, P484S was the only nsSNP model that contained fewer beta strands than wild type TRIM22 in certain regions (Fig 3.2). The majority of nsSNP models contained a greater number of beta strands than wild type TRIM22, resulting in overall net increase in beta strand formation.

To extend our structural analysis, we used Tm-Align to calculate the Tm-score and root mean square deviation (RMSD) for each nsSNP model. Tm-score is used to assess topological similarity between wild type and mutant models, whereas RMSD is used to measure average distance between the α-carbon backbones of wild type and mutant

Figure 3.2: Structural models for wild type TRIM22 and high-risk nsSNPs in the B30.2 domain.

Putative structural models for the B30.2 domains of wild type TRIM22 and the 9 high-risk nsSNPs located in the B30.2 domain. Variable regions (v1-v4) are highlighted as follows: v1 blue, v2 orange, v3 magenta, and v4 green. Non-variable regions are shown in white and mutated amino acids are shown in yellow. Left image: Enlarged reference image that illustrates the color and location of each variable region and the color of mutated amino acids (image shown is the v1-v4 regions of wild type TRIM22 and the P403 amino acid). Each of the 9 nsSNP images (small images on the right) show the putative 3D structure of wild type TRIM22's B30.2 domain on the left and the putative 3D structure of TRIM22's B30.2 domain with the mutated amino acid (nsSNP) on the right. The location of the amino acid in question is shown (yellow) on both wild type and nsSNP structures. All models were generated using the 3D-JigSaw protein comparative modeling server and SPDBV (v4.1).

Figure 3.2

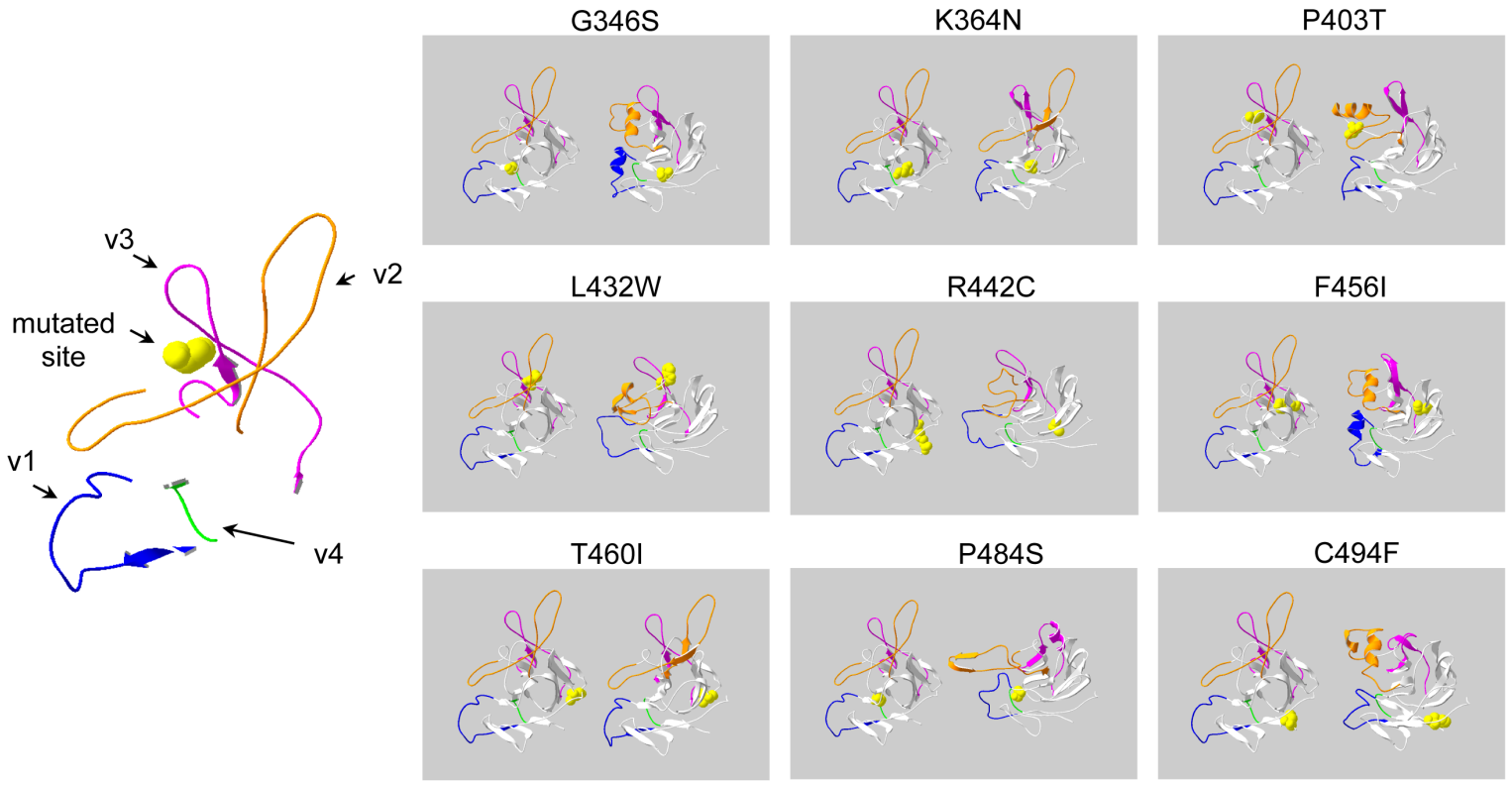


Figure 3.2: Structural models for wild type TRIM22 and high-risk nsSNPs in the B30.2 domain.

models [57,69]. A higher RMSD typically indicates greater deviation between wild type and mutant structures. The Tm-score and RMSD for each nsSNP model is listed in Table 3.4. The maximum RMSD was 3.04 (R442C), followed by 3.03 (F456I), 3.00 (L432W), 2.96 (G346S), and 2.80 (P484S). RMSD for nsSNPs K364N, P403T, T460I, and C494F ranged from 1.58 to 1.99 Å. These results indicate that 9 high-risk nsSNPs markedly alter the putative structure of TRIM22's B30.2 domain, in particular the surface-exposed v2 and v3 regions, and that they likely induce severe structural changes in the TRIM22 protein.

Importantly, these nsSNPs may decrease flexibility in the v2 and v3 regions of TRIM22. The v2/v3 regions of wild type TRIM22 are predicted to form relaxed loop segments, similar to the loops in the recently solved 3D structure of rhesus monkey TRIM5α's B30.2 domain [26]. In contrast, the v2 and v3 regions of the nsSNP models contain more rigid secondary structures, such as alpha helices or beta strands (Fig 3.2). Since loop flexibility in rhesus monkey TRIM5α is thought to facilitate restriction of divergent retroviruses and to increase resistance to mutations in the HIV-1 capsid protein, it is possible that these nsSNPs may impair the antiviral activity and/or breadth of TRIM22. Further experiments, such as the resolution of wild type TRIM22's tertiary structure, are required to address these possibilities.

3.3.5 Prediction of Post-Translational Modification Sites in TRIM22

To investigate how nsSNPs may influence the post-translational modification (PTM) of TRIM22, we used a variety of *in silico* prediction tools to identify putative PTM sites in the TRIM22 protein. PTMs are involved in many biological processes, including a number of canonical innate immune pathways, and are essential for the regulation of protein structure and function [59,70–72]. To analyze residues in TRIM22 that may undergo ubiquitylation or sumoylation, we used the UbPred, BDM-PUB, SUMO-plot, and SUMOsp 2.0 programs. The GPS 2.1 and NetPhos 2.0 servers were used to predict serine, threonine, and tyrosine phosphorylation sites in the TRIM22 protein [2,60,61,73].

Table 3.5

Table 3.5: RMSD (Å) and TM-score for the 9 high-risk nsSNPs in the B30.2 domain of TRIM22.

nsSNP ID	Mutation	RMSD (Å)	TM-Score
rs371728648	G346S	2.96	0.75184
rs191847788	K364N	1.72	0.93911
rs375595000	P403T	1.99	0.85389
rs370495523	L432W	3.00	0.70821
rs187416296	R442C	3.04	0.68305
rs377529439	F456I	3.03	0.73743
rs371028900	T460I	1.76	0.94873
rs200638791	P484S	2.80	0.75981
rs200148337	C494F	1.58	0.95645

RMSD and Tm-scores were calculated using Tm-Align.

UbPred predicted that 6 lysine residues in TRIM22 undergo ubiquitylation. In contrast, BDM-PUB predicted that 19 lysine residues undergo ubiquitylation. Both UbPred and BDM-PUB predicted that residues K63, K160, and K173 undergo ubiquitylation (Table 3.5). According to ConSurf, these 3 lysine residues are highly conserved and exposed to the protein surface. ConSurf also predicted that K173 was a functional residue (Fig 3.1). SUMOplot predicted that 4 lysine residues in TRIM22 undergo sumoylation, whereas SUMOsp 2.0 predicted that 2 lysine residues undergo sumoylation. Both programs predicted that K153 undergoes sumoylation (Table 3.5). Similar to K173, ConSurf showed that K153 is highly conserved and exposed to the protein surface. ConSurf also predicted that K153 was a functional residue (Fig 3.1).

In addition to putative sumoylation sites, we also identified 7 potential sumo-interacting motifs (SIM) (Fig 3.3a). SIMs are short hydrophobic motifs that interact non-covalently with other sumoylated proteins. The best characterized SIMs have the consensus sequence V/I/L-x-V/I/L-V/I/L or V/I/L-V/I/L-x-V/I/L [63]. Notably, 5 of the putative SIMs are highly conserved in multiple TRIM22 orthologues and 3 are also present in the human and rhesus monkey TRIM5α proteins (Fig 3.3b). In addition, 2 TRIM5α SIMs (ILGV and VIGL) were previously shown to be required for TRIM5α-mediated antiviral activity. SIM mutations in the rhesus monkey TRIM5α protein abolished HIV-1 restriction and disrupted TRIM5α trafficking to SUMO-1 nuclear bodies. Moreover, SIM mutations in the human TRIM5α protein abrogated N-MLV restriction by preventing TRIM5α binding to the sumoylated N-MLV capsid protein [62,74]. More studies are needed to determine the role that SIMs play in TRIM22-mediated antiviral activity.

To identify putative phosphorylation sites in TRIM22, we used GPS 2.1 and NetPhos 2.0 servers. The GPS 2.1 server predicted that there were 31 serine-specific phosphorylation sites, 13 threonine-specific sites, and 11 tyrosine-specific sites in the TRIM22 protein. Conversely, NetPhos 2.0 predicted that there were 19 serine-specific phosphorylation sites, 4 threonine-specific sites, and 2 tyrosine-specific sites (Table 3.6). 16 serine residues, 3 threonine residues, and 2 tyrosine residues were predicted to be phosphorylated by both GPS 2.1 and NetPhos 2.0 servers. Many of these putative phosphorylation sites are highly conserved among multiple TRIM22 orthologues and

Table 3.6

Table 3.6 Putative ubiquitylation and sumoylation sites in the TRIM22 protein.

Ubiquitylation		**Sumoylation**	
UbPred	**BDM-PUB**	**SUMOplot**	**SUMOsp 2.0**
93 (7e)*	6 (3e)	6 (3e)	85 (2e)
160 (7e)*	44 (1e)	**153 (9e)***	**153 (9e)***
173 (9e)*	85 (2e)	185 (4e)	
204 (6e)	93 (7e)*	265 (6e)	
257 (1e)	103 (6e)		
430 (6e)	**109 (9e)**		
	160 (7e)*		
	173 (9e)*		
	265 (6e)		
	266 (9e)		
	268 (2e)		
	272 (6e)		
	273 (9e)		
	275 (7e)		
	324 (1e)		
	332 (2e)#		
	374 (1e)		
	380 (3e)		
	382 (1e)		

Conservation scores are shown in parentheses following amino acid site; Putative functional residues are indicated with bold text, whereas putative structural residues are indicated with italicized text (ConSurf results Fig 3.1); Residues predicted to undergo ubiquitylation or sumoylation by both programs are indicated with an asterisk; Residues predicted to undergo ubiquitylation or sumoylation that also coincide with the location of nsSNPs are indicated with a hashtag.

Figure 3.3: Putative sumo-interacting motifs (SIM) in TRIM22.

(a) List of putative SIMs in the TRIM22 protein, including the sequence and domain location for each SIM (amino acids are indicated in parentheses); Red and blue amino acids are predicted functional and structural residues, respectively (ConSurf analysis Fig 3.1); Asterisk: SIMs that are conserved in all mammalian TRIM22 orthologues except elephant; Double asterisk: SIMs that are not found in TRIM5α, but are replaced by a different SIM (e.g. VLTL, IVPL). **(b)** Alignment of mammalian TRIM22, human TRIM5α, and rhesus monkey TRIM5α amino acid sequences (amino acids 350-444 of the B30.2 domain are shown). Conserved SIMs are highlighted in magenta and other SIMs are highlighted in light blue. Conserved amino acids are indicated with an asterisk.

Figure 3.3

A

		Conserved in:		
SIM	**Domain**	**Primates**	**Mammals**	**TRIM5α**
LELL (19-22)	RING	Y	Y	Y
VNVL (210-213)	Coiled-coil	Y	N*	N
VIDV (252-255)	Spacer 2	Y	N	N
LQVL (287-290)	Spacer 2	N	N	N
ILGV (368-371)	B30.2	Y	Y	Y
VIGL (409-412)	B30.2	Y	Y	Y
VLTL (431-434)	B30.2	Y	N*	N**

B

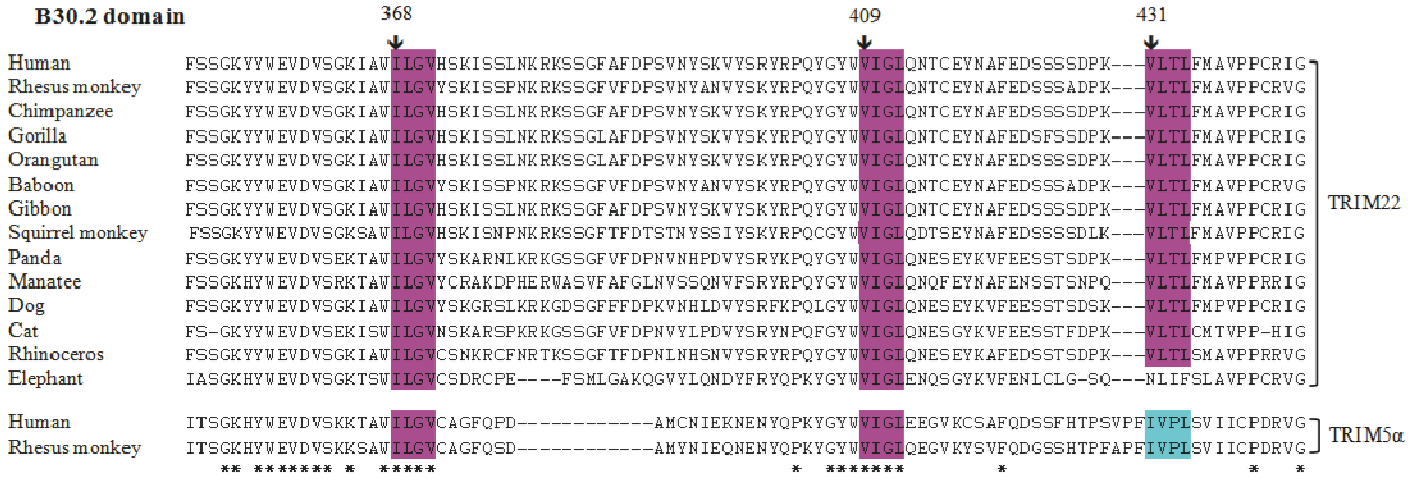


Figure 3.3: Putative sumo-interacting motifs (SIM) in TRIM22.

Table 3.7

Table 3.7 Putative phosphorylation sites in the TRIM22 protein.

GPS 2.1			**NetPhos 2.0**		
Serine	**Threonine**	**Tyrosine**	**Serine**	**Threonine**	**Tyrosine**
4 (1e)	23 (7e)	175 (1b)	46 (7e)*	130 (7b)	356 (8b)*
27 (9e)	61 (1b)#	298 (1e)	50 (1e)	263 (3e)*	479 (5b)*
46 (7e)*	170 (1e)	299 (6b)	54 (3e)*	325 (1e)*	
54 (3e)*	220 (1e)	355 (5b)	87 (4e)*	330 (1e)*#	
87 (4e)*	232 (1e)#	356 (8b)*	244 (8b)*#		
122 (9e)	263 (3e)*	394 (1b)	245 (8b)*		
231 (4e)	294 (7e)#	398 (7b)	**259 (9e)***		
235 (9e)	311 (2b)	418 (8b)	261 (2e)*		
244 (8b)*#	325 (1e)*	467 (8b)	269 (1e)*		
245 (8b)*	330 (1e)*#	479 (5b)*	**271 (8e)***		
259 (9e)*	433 (7b)	**481 (8e)**	276 (5e)*		
261 (2e)*	*460 (9b)#*		284 (5e)*		
269 (1e)*	492 (6e)		373 (8b)*		
271 (8e)*			383 (3e)*		
276 (5e)*			**384 (9e)***		
284 (5e)*			399 (7b)		
309 (8e)			425 (6e)*		
312 (6e)			426 (4e)*		
317 (9b)			**475 (8e)**		
373 (8b)*					
376 (2e)*					
377 (1e)					
383 (3e)*					
384 (9e)*					
391 (3e)					
424 (7e)					
425 (6e)*					
426 (4e)*					
455 (9b)					
497 (9e)					
498 (7e)					

Conservation scores are shown in parentheses following amino acid site; Putative functional residues are indicated with bold text, whereas putative structural residues are indicated with italicized text (ConSurf results Fig 3.1); Residues predicted to undergo phosphorylation by both programs are indicated with an asterisk; Residues predicted to undergo phosphorylation that also coincide with the location of nsSNPs are indicated with a hashtag.

several were predicted to be important structural or functional residues by ConSurf (Table 3.6, Fig 3.1). Although TRIM22 phosphorylation has never been demonstrated experimentally, our results suggest that it may undergo phosphorylation at a number of sites. Of interest, other TRIM proteins have been shown to undergo phosphorylation, including the antiviral TRIM19 and TRIM21 proteins [75–79].

Several putative PTMs coincide in location with nsSNPs in the *TRIM22* gene (T61, T232, S244, T294, T330, K332, and T460). S244 and T460 are particularly interesting because both sites are highly conserved among TRIM22 orthologues and S244L and T460I were predicted to be deleterious by 5 and 6 *in silico* algorithms, respectively (Table 3.2, Table 3.3). In addition, T460 was predicted to be a critical structural residue by ConSurf. Although the consequences of TRIM22 phosphorylation are currently unknown, the mutation of phosphorylation sites in other proteins has been shown to profoundly alter protein function by, for example, altering protein stability, localization, or protein-protein interactions. To this end, we used I-Mutant to predict whether S244L and T460I altered the stability of the TRIM22 protein. I-Mutant is a support vector machine-based tool that predicts changes in protein stability following single site mutations by estimating free energy changes as well as the direction of the change (increase or decrease) [64]. Both S244L and T460I were predicted to be less stable than the wild type protein, with free energy change values of -0.83 and -1.38, respectively (Table 3.7). The I-Mutant results for the 12 high-risk nsSNPs that do not coincide with putative PTM sites, plus the results for the 5 low-risk nsSNPs that do coincide with putative PTM sites, are also shown in Table 3.7.

It is possible that the phosphorylation of TRIM22 at sites S244 and/or T460 is required for some integral TRIM22 function and that the nsSNPs S244L and T460I impair this function; however, these nsSNPs may also impair protein stability, which would likely amplify any detrimental of PTM impairment. Many additional high-risk nsSNPs, plus several low-risk nsSNPs located at putative PTM sites, also decreased TRIM22 protein stability (Table 3.7). A number of studies have shown that decreased protein stability leads to increased protein misfolding, aggregation, and degradation. Accordingly, decreased stability typically results in decreased net function [80–83]. Future in-depth studies

are required to investigate the effects of these nsSNPs on the structure and function of TRIM22's B30.2 domain. Pertinent TRIM22 sites that are predicted to be highly deleterious and/or undergo PTMs are depicted in Figure 3.4.

3.4 Conclusions

Our results demonstrate that multiple nsSNPs in the antiviral *TRIM22* gene may be deleterious to TRIM22 structure and/or function. Most of these high-risk nsSNPs are located at highly conserved amino acid sites in a protein-protein interaction module called the B30.2 domain. In this study, we show that 9 of the top high-risk nsSNPs disrupt the putative structure of TRIM22's B30.2 domain, particularly the surface-exposed v2 and v3 regions. In the closely-related TRIM5α protein, these same regions were previously shown to play a key role in retroviral restriction. In addition to these findings, we also identify several TRIM22 sites that may undergo post-translational modification, including sites that coincide with the location of high-risk nsSNPs. This study is the first systematic and extensive *in silico* analysis of functional SNPs in the *TRIM22* gene.

Table 3.8

Table 3.8: I-Mutant results for selected nsSNPs in the TRIM22 protein.

nsSNP ID	Mutation	# Del. Pred.	DDG	Sign of DDG	PTM	ConSurf
rs192306924	T61N	1	0.56	Decrease (1)	Yes	1b
rs201847190	**L68R**	**5**	**-1.02**	**Decrease (7)***	**No**	**8e**
rs199625192	H73R	5	0.23	Decrease (3)	No	7e
rs368058642	E135K	4.5	-1.00	Decrease (9)*	No	7e (9b)
rs2291843	T232A	0	-0.53	Decrease (5)	Yes	1e
rs374292901	I234N	5	-0.80	Decrease (1)	No	7b (9e)
rs61735273	S244L	5	-0.83	Decrease (2)	Yes	8b
rs73404240	T294K	2	-0.63	Decrease (5)	Yes	7e
rs201494620	T330I	1	-2.14	Decrease (7)*	Yes	1e
rs368220166	K332N	1	-0.42	Decrease (2)	Yes	2e
rs371728648	G346S	5	-0.27	Decrease (7)	No	8b
rs191847788	**K364N**	**4.5**	**-1.09**	**Decrease (4)**	**No**	**9e**
rs375595000	**P403T**	**6**	**-2.64**	**Decrease (8)**	**No**	**8e**
rs370495523	L432W	4	0.08	Decrease (6)	No	8b
rs187416296	R442C	5	-1.23	Decrease (6)*	No	7e
rs377529439	F456I	5	-1.59	Decrease (8)*	No	8b (9b)
rs371028900	*T460I*	*6*	*-1.38*	*Decrease (5)**	*Yes*	*9b*
rs200638791	P484S	4.5	-2.97	Decrease (9)*	No	6e (9b)
rs200148337	C494F	6	-0.21	Decrease (4)	No	8b

Figure 3.4: Putative functional sites in the TRIM22 protein.

Schematic depicting the approximate location of the top predicted PTM sites (ubiquitylation, sumoylation, and phosphorylation), the 14 high-risk nsSNPs in TRIM22, the 3 sumo-interacting motifs (SIMs), and the 2 high-risk nsSNP sites (S244L and T460I) predicted to undergo phosphorylation in the wild type TRIM22 protein. Several sites of known functional importance are marked on the TRIM22 protein (top image), including the C15/C18 residues (required for TRIM22 E3 ligase activity), the C97/H100 residues (part of the zinc-binding motif in BB2), and the nuclear localization signal (NLS) [84–86]. The 'antiviral patch' region, which was previously shown to be integral for the antiviral activity of TRIM5α, is shown in the B30.2 domain, as well as the approximate location of each variable region (v1-v4, bright blue areas) [29,33]. Amino acids 491-494 were previously shown to be required for the nuclear localization of TRIM22 [87]. RING, B-box 2 (BB2), coiled-coil (CC), and B30.2 (PRY/SPRY) domains are listed.

Figure 3.4

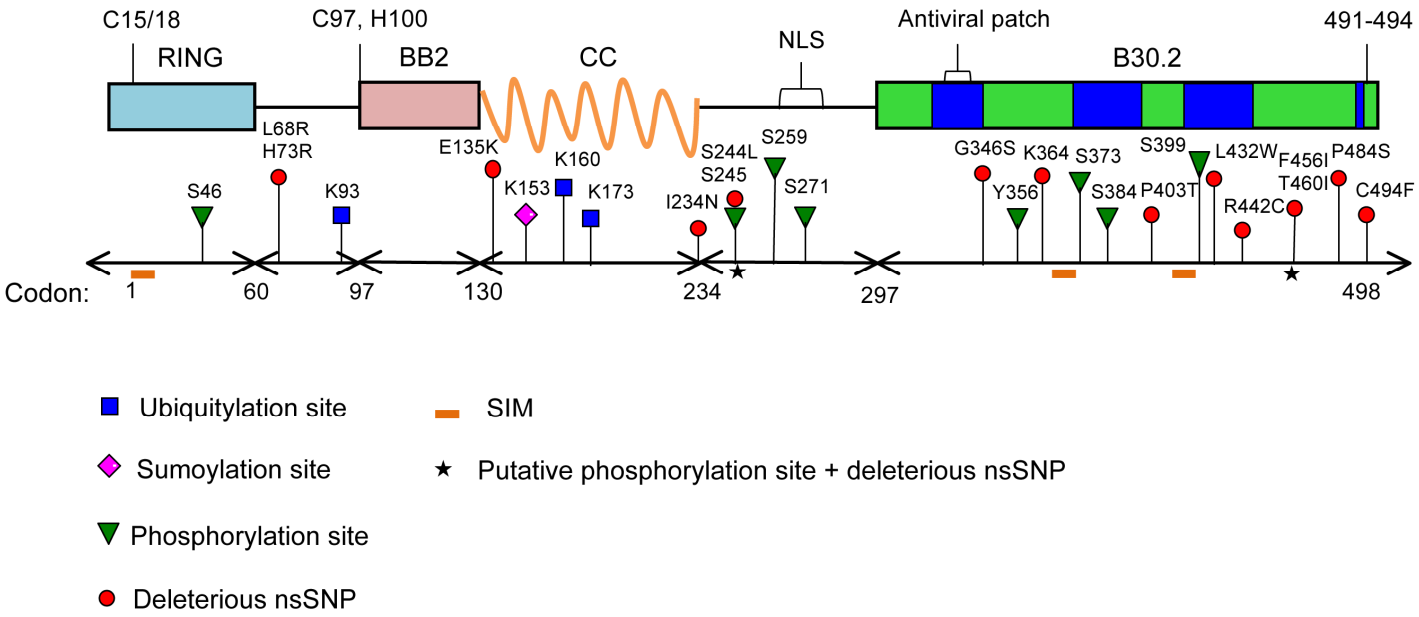


Figure 3.4 Putative functional sites in the TRIM22 protein.

3.5 References

1. Ramensky, V. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30,** 3894–3900 (2002).

2. Radivojac, P. *et al.* Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* **78,** 365–80 (2010).

3. Doniger, S. W. *et al.* A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **4,** e1000183 (2008).

4. Daley, D. *et al.* Associations and interactions of genetic polymorphisms in innate immunity genes with early viral infections and susceptibility to asthma and asthma-related phenotypes. *J. Allergy Clin. Immunol.* **130,** 1284–93 (2012).

5. Azad, A. K., Sadee, W. & Schlesinger, L. S. Innate immune gene polymorphisms in tuberculosis. *Infect. Immun.* **80,** 3343–59 (2012).

6. Netea, M. G., Wijmenga, C. & O'Neill, L. A. J. Genetic variation in Toll-like receptors and disease susceptibility. *Nat. Immunol.* **13,** 535–42 (2012).

7. Sobieszczyk, M. E., Lingappa, J. R. & McElrath, M. J. Host genetic polymorphisms associated with innate immune factors and HIV-1. *Curr. Opin. HIV AIDS* **6,** 427–34 (2011).

8. Heim, M. H. Innate immunity and HCV. *J. Hepatol.* **58,** 564–74 (2013).

9. Santana-de Anda, K., Gómez-Martín, D., Díaz-Zamudio, M. & Alcocer-Varela, J. Interferon regulatory factors: beyond the antiviral response and their link to the development of autoimmune pathology. *Autoimmun. Rev.* **11,** 98–103 (2011).

10. Nisole, S., Stoye, J. P. & Saïb, A. TRIM family proteins: retroviral restriction and antiviral defence. *Nat. Rev. Microbiol.* **3,** 799–808 (2005).

11. Hattlmann, C. J., Kelly, J. N. & Barr, S. D. TRIM22: A Diverse and Dynamic Antiviral Protein. *Mol. Biol. Int.* **2012,** 153415 (2012).

12. Jefferies, C., Wynne, C. & Higgs, R. Antiviral TRIMs: friend or foe in autoimmune and autoinflammatory disease? *Nat. Rev. Immunol.* **11,** 617–25 (2011).

13. Reymond, A. *et al.* The tripartite motif family identifies cell compartments. *EMBO J.* **20,** 2140–51 (2001).

14. Meroni, G. & Diez-Roux, G. TRIM/RBCC, a novel class of "single protein RING finger" E3 ubiquitin ligases. *Bioessays* **27,** 1147–57 (2005).

15. Marín, I. Origin and diversification of TRIM ubiquitin ligases. *PLoS One* **7,** e50030 (2012).

16. Kawai, T. & Akira, S. Regulation of innate immune signalling pathways by the tripartite motif (TRIM) family proteins. *EMBO Mol. Med.* **3,** 513–27 (2011).

17. Carthagena, L. *et al.* Human TRIM gene expression in response to interferons. *PLoS One* **4,** e4894 (2009).

18. McNab, F. W., Rajsbaum, R., Stoye, J. P. & O'Garra, A. Tripartite-motif proteins and innate immune regulation. *Curr. Opin. Immunol.* **23,** 46–56 (2011).

19. Rajsbaum, R., Stoye, J. P. & O'Garra, A. Type I interferon-dependent and -independent expression of tripartite motif proteins in immune cells. *Eur. J. Immunol.* **38,** 619–30 (2008).

20. Uchil, P. D., Quinlan, B. D., Chan, W.-T., Luna, J. M. & Mothes, W. TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathog.* **4,** e16 (2008).

21. Ohmine, S. *et al.* The antiviral spectra of TRIM5α orthologues and human TRIM family proteins against lentiviral production. *PLoS One* **6,** e16121 (2011).

22. Li, Y., Li, X., Stremlau, M., Lee, M. & Sodroski, J. Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. *J. Virol.* **80,** 6738–44 (2006).

23. Stremlau, M., Owens, C. M. & Perron, M. J. The cytoplasmic body component TRIM5 a restricts HIV-1 infection in Old World monkeys. **427,** (2004).

24. Stremlau, M., Perron, M., Welikala, S. & Sodroski, J. Species-Specific Variation in the B30 . 2 (SPRY) Domain of TRIM5α Determines the Potency of Human Immunodeficiency Virus Restriction. **79,** 3139–3145 (2005).

25. Pertel, T. *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472,** 361–365 (2011).

26. Biris, N. *et al.* Structure of the rhesus monkey TRIM5α PRYSPRY domain, the HIV capsid recognition module. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13278–13283 (2012).

27. Yap, M. W. & Stoye, J. P. A Single Amino Acid Change in the SPRY Domain of Human Trim5 □ Leads to HIV-1 Restriction. **15,** 73–78 (2005).

28. Perron, M. J., Stremlau, M. & Sodroski, J. Two surface-exposed elements of the B30.2/SPRY domain as potency determinants of N-tropic murine leukemia virus restriction by human TRIM5alpha. *J. Virol.* **80,** 5631–6 (2006).

29. Ohkura, S., Yap, M. W., Sheldon, T. & Stoye, J. P. All three variable regions of the TRIM5alpha B30.2 domain can contribute to the specificity of retrovirus restriction. *J. Virol.* **80,** 8554–65 (2006).

30. Kono, K., Bozek, K., Domingues, F. S., Shioda, T. & Nakayama, E. E. Impact of a single amino acid in the variable region 2 of the Old World monkey TRIM5alpha SPRY (B30.2) domain on anti-human immunodeficiency virus type 2 activity. *Virology* **388,** 160–8 (2009).

31. Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2832–7 (2005).

32. Song, B. *et al.* The B30 . 2 (SPRY) Domain of the Retroviral Restriction Factor TRIM5α Exhibits Lineage-Specific Length and Sequence Variation in Primates. *J. Virol.* **79,** 6111–6121 (2005).

33. Sawyer, S. L., Emerman, M. & Malik, H. S. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathog.* **3,** e197 (2007).

34. Bouazzaoui, A. *et al.* Stimulated trans-acting factor of 50 kDa (Staf50) inhibits HIV-1 replication in human monocyte-derived macrophages. *Virology* **356,** 79–94 (2006).

35. Barr, S. D., Smiley, J. R. & Bushman, F. D. The interferon response inhibits HIV particle production by induction of TRIM22. *PLoS Pathog.* **4,** e1000007 (2008).

36. Tissot, C. & Mechti, N. Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus type 1 long terminal repeat expression. *J. Biol. Chem.* **270,** 14891–8 (1995).

37. Kajaste-Rudnitski, A. *et al.* TRIM22 inhibits HIV-1 transcription independently of its E3 ubiquitin ligase activity, Tat, and NF-kappaB-responsive long terminal repeat elements. *J. Virol.* **85,** 5183–96 (2011).

38. Kajaste-Rudnitski, A. *et al.* TRIM22 inhibits HIV-1 transcription independently of its E3 ubiquitin ligase activity, Tat, and NF-kappaB-responsive long terminal repeat elements. *J. Virol.* **85,** 5183–96 (2011).

39. Singh, R. *et al.* Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. *J. Virol.* **85,** 208–16 (2011).

40. Singh, R. *et al.* TRIM5α and TRIM22 are differentially regulated according to HIV-1 infection phase and compartment. *J. Virol.* (2014). doi:10.1128/JVI.03603-13

41. Van Manen, D. *et al.* The effect of Trim5 polymorphisms on the clinical course of HIV-1 infection. *PLoS Pathog.* **4,** e18 (2008).

42. Nakajima, T. *et al.* Impact of novel TRIM5alpha variants, Gly110Arg and G176del, on the anti-HIV-1 activity and the susceptibility to HIV-1 infection. *AIDS* **23,** 2091–100 (2009).

43. Goldschmidt, V. *et al.* Role of common human TRIM5alpha variants in HIV-1 disease progression. *Retrovirology* **3,** 54 (2006).

44. Sawyer, S. L., Wu, L. I., Akey, J. M., Emerman, M. & Malik, H. S. High-frequency persistence of an impaired allele of the retroviral defense gene TRIM5alpha in humans. *Curr. Biol.* **16,** 95–100 (2006).

45. The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42,** D191–8 (2014).

46. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res.* **41,** D48–55 (2013).

47. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–11 (2001).

48. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–9 (2010).

49. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–81 (2009).

50. Bao, L., Zhou, M. & Cui, Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* **33,** W480–2 (2005).

51. Capriotti, E., Calabrese, R. & Casadio, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **22,** 2729–34 (2006).

52. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. & Casadio, R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* **30,** 1237–44 (2009).

53. Ferrer-Costa, C. *et al.* PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* **21,** 3176–8 (2005).

54. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–33 (2010).

55. Bates, P. A., Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins* **Suppl 5,** 39–46 (2001).

56. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18,** 2714–23 (1997).

57. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33,** 2302–9 (2005).

58. Lindahl, E., Azuara, C., Koehl, P. & Delarue, M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.* **34,** W52–6 (2006).

59. Gill, G. Post-translational modification by the small ubiquitin-related modifier SUMO has big effects on transcription factor activity. *Curr. Opin. Genet. Dev.* **13,** 108–13 (2003).

60. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294,** 1351–62 (1999).

61. Xue, Y. *et al.* GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Eng. Des. Sel.* **24,** 255–60 (2011).

62. Arriagada, G., Muntean, L. N. & Goff, S. P. SUMO-interacting motifs of human TRIM5α are important for antiviral activity. *PLoS Pathog.* **7,** e1002019 (2011).

63. Hecker, C.-M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **281,** 16117–27 (2006).

64. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33,** W306–10 (2005).

65. Mavroconstanti, T., Johansson, S., Winge, I., Knappskog, P. M. & Haavik, J. Functional properties of rare missense variants of human CDH13 found in adult attention deficit/hyperactivity disorder (ADHD) patients. *PLoS One* **8,** e71445 (2013).

66. Ferrer-Costa, C., Orozco, M. & de la Cruz, X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *J. Mol. Biol.* **315,** 771–86 (2002).

67. Miller, M. P. & Kumar, S. Understanding human disease mutations through the use of interspecific genetic variation. *Hum. Mol. Genet.* **10,** 2319–28 (2001).

68. Berezin, C. *et al.* ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* **20,** 1322–4 (2004).

69. Carugo, O. & Pongor, S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci.* **10,** 1470–3 (2001).

70. Dai, C. & Gu, W. p53 post-translational modification: deregulated in tumorigenesis. *Trends Mol. Med.* **16,** 528–36 (2010).

71. Gallego, M. & Virshup, D. M. Post-translational modifications regulate the ticking of the circadian clock. *Nat. Rev. Mol. Cell Biol.* **8,** 139–48 (2007).

72. Shiloh, Y. & Ziv, Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat. Rev. Mol. Cell Biol.* **14,** 197–210 (2013).

73. Ren, J. *et al.* Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. *Proteomics* **9,** 3409–3412 (2009).

74. Lukic, Z., Goff, S. P., Campbell, E. M. & Arriagada, G. Role of SUMO-1 and SUMO interacting motifs in rhesus TRIM5α-mediated restriction. *Retrovirology* **10,** 10 (2013).

75. Gresko, E. *et al.* PML tumor suppressor is regulated by HIPK2-mediated phosphorylation in response to DNA damage. *Oncogene* **28,** 698–708 (2009).

76. Hayakawa, F. & Privalsky, M. L. Phosphorylation of PML by mitogen-activated protein kinases plays a key role in arsenic trioxide-mediated apoptosis. *Cancer Cell* **5,** 389–401 (2004).

77. Stacey, K. B., Breen, E. & Jefferies, C. A. Tyrosine phosphorylation of the E3 ubiquitin ligase TRIM21 positively regulates interaction with IRF3 and hence TRIM21 activity. *PLoS One* **7,** e34041 (2012).

78. Roberts, J. D., Chiche, J.-D., Kolpa, E. M., Bloch, D. B. & Bloch, K. D. cGMP-dependent protein kinase I interacts with TRIM39R, a novel Rpp21 domain-containing TRIM protein. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **293,** L903–12 (2007).

79. Valiyeva, F. *et al.* Characterization of the oncogenic activity of the novel TRIM59 gene in mouse cancer models. *Mol. Cancer Ther.* **10,** 1229–40 (2011).

80. Prior, T. W. *et al.* A missense mutation in the dystrophin gene in a Duchenne muscular dystrophy patient. *Nat. Genet.* **4,** 357–60 (1993).

81. Singh, S. M., Kongari, N., Cabello-Villegas, J. & Mallela, K. M. G. Missense mutations in dystrophin that trigger muscular dystrophy decrease protein stability and lead to cross-beta aggregates. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15069–74 (2010).

82. Mayer, S., Rüdiger, S., Ang, H. C., Joerger, A. C. & Fersht, A. R. Correlation of levels of folded recombinant p53 in escherichia coli with thermodynamic stability in vitro. *J. Mol. Biol.* **372,** 268–76 (2007).

83. Du, K., Sharma, M. & Lukacs, G. L. The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat. Struct. Mol. Biol.* **12,** 17–25 (2005).

84. Duan, Z., Gao, B., Xu, W. & Xiong, S. Identification of TRIM22 as a RING finger E3 ubiquitin ligase. *Biochem. Biophys. Res. Commun.* **374,** 502–6 (2008).

85. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.* **32,** 63–70 (2007).

86. Borden, K. L. RING fingers and B-boxes: zinc-binding protein-protein interaction domains. *Biochem. Cell Biol.* **76,** 351–8 (1998).

87. Sivaramakrishnan, G., Sun, Y., Rajmohan, R. & Lin, V. C. L. B30.2/SPRY domain in tripartite motif-containing 22 is essential for the formation of distinct nuclear bodies. *FEBS Lett.* **583,** 2093–9 (2009).

Chapter 4

4 The *TRIM22* nsSNP rs1063303:G>C is not evolving under balancing selection in the Inuit and is associated with low serum TG and high serum HDL levels in the Canadian Inuit

TRIM22 is an IFN-induced antiviral protein that plays a key role in the host antiviral response and inhibits the replication of diverse viruses, such as HIV-1, HBV, and IAV. Altered TRIM22 expression has also been linked to cancer, autoimmune disease, and cellular proliferation/differentiation. We previously identified a nsSNP in the *TRIM22* gene (rs1063303:G>C) that had an inverse impact on TRIM22 function. The *TRIM22* rs1063303:C allele increased TRIM22 expression, but decreased its anti-HIV activity. Interestingly, we found that the frequency of rs1063303:G>C varied markedly among ethnic populations and that the AFR, AMR, and EUR cohorts from the 1000 Genomes project contained an excess of intermediate frequency *TRIM22* rs1063303:G>C alleles. The latter is typically indicative of balancing selection, a non-neutral selective force that maintains polymorphism in a population. In this study, we determined the frequency of the *TRIM22* nsSNP rs1063303:G>C in two different Inuit populations and one non-Inuit population and calculated the selective forces acting on this site. Interestingly, we found that the *TRIM22* rs1063303:C allele is significantly more prevalent in Inuit compared to non-Inuit populations and that unlike AFR, AMR, and EUR cohorts, the Inuit do not contain an excess of intermediate frequency *TRIM22* rs1063303:G>C alleles, indicating that *TRIM22* rs1063303:G>C has not evolved under balancing selection in these Inuit populations. Surprisingly, we also found that *TRIM22* rs1063303:G>C was associated with significantly lower serum triglycerides (TG) levels and significantly higher high-density lipoprotein (HDL) levels in the Canadian Inuit population. TRIM22's effect on TG and HDL is unprecedented in the TRIM protein family and may represent an exciting new research avenue for TRIM22 and other TRIM proteins.

4.1 Introduction

Infection and inflammation induce the acute-phase response (APR), an early, complex host reaction to injurious stimuli, that leads to specific changes in lipid and lipoprotein metabolism [1]. During the APR in humans and other primates, serum triglyceride (TG) levels are markedly increased, a phenomenon that is largely due to increased levels of circulating very-low-density lipoprotein (VLDL). Conversely, both serum low-density lipoprotein (LDL) and serum high-density lipoprotein (HDL) levels, along with serum total cholesterol levels, are decreased (Appendix A). These changes are mediated by a number of proinflammatory cytokines, including TNF, IL-1, IL-2, IL-6, and IFN [2–8]. Accordingly, SNPs in these cytokines and the genes they induce have been associated with altered levels of serum lipoproteins, a modified APR to infection, and multiple inflammatory and metabolic disorders [9–15].

Interestingly, in addition to their well-documented role in lipid transport, an increasing body of evidence suggests that lipoproteins participate in host defense as agents of the innate immune response. For example, lipoproteins bind to and neutralize a variety of DNA and RNA viruses, such as Vesicular stomatitis virus (VSV), Rubella virus, HIV, SIV, Herpes simplex virus, Vaccinia virus, Japanese encephalitis virus, Poliovirus, and Epstein-Barr virus [16–21]. Moreover, during the IFN response to viral infection a soluble antiviral form of the LDL receptor is produced that inhibits the replication of VSV and rhinovirus [22–24]. Multiple lipoproteins also bind to and neutralize the bacterial cell wall components lipopolysaccharide (LPS) and lipoteichoic acid (LTA) [25–29]. Thus, the APR and its cytokine-induced lipoprotein changes are initially beneficial to the host because they help resolve infection and inflammation; however, they can also promote disease pathogenesis if they are prolonged (e.g. during chronic infection, metabolic syndrome, inflammatory disorders, diabetes, obesity, or heart failure) [11–13,30–38].

TRIM family proteins have been implicated in a wide range of biological processes, including the innate immune response to viral infection. Multiple TRIM proteins are upregulated by type I and type II IFNs and recent studies have identified a number of TRIMs that can: 1) restrict viral replication directly by targeting specific viral proteins (e.g. TRIM5α inhibits HIV replication by targeting the HIV capsid protein), and/or 2)

restrict viral replication indirectly by augmenting innate immune signaling pathways (e.g. TRIM5α activates the TAK1 kinase complex, which induces downstream NF-κB and AP-1 signaling and the transcription of inflammatory cytokines) [39–44]. TRIM proteins contain a conserved N-terminal RBCC motif, which consists of RING, BB, and CC domains, and a variable C-terminal domain [45]. The RING domain typically confers E3 ligase activity and this activity is often critical for TRIM protein function. Multiple TRIMs have been shown to function as E3 ubiquitin and/or E3 sumo ligases and can post-translationally modify specific host/viral proteins with ubiquitin and/or sumo moieties [46,47]. Most TRIM proteins contain a C-terminal B30.2 domain, which is a flexible, highly plastic structure that typically functions as a protein-protein interaction module [48,49].

The TRIM22 protein is a B30.2-containing member of the TRIM family that plays an integral role in the host antiviral response. TRIM22 is upregulated by type I and type II IFNs, several proinflammatory cytokines, and a number of different viral infections [50]. Studies have shown that TRIM22 inhibits the replication of multiple viruses, including HIV-1, HBV, and IAV [51–54]. TRIM22 has also been linked to cancer, autoimmune disease, and cellular proliferation/differentiation [55–63]. We previously identified a common nsSNP in the *TRIM22* gene (rs1063303:G>C) that had an inverse impact on TRIM22 function [64]. rs1063303:G>C increased TRIM22 expression, but decreased its anti-HIV activity and altered its sub-cellular localization pattern. Interestingly, the frequency of rs1063303:G>C varied markedly among different ethnic groups (1000 Genomes AFN, AMR, ASN, and EUR populations) and we found an excess of intermediate frequency rs1063303:G>C alleles in AFN, AMR, and EUR cohorts. An excess of intermediate frequency alleles is often indicative of balancing selection, an evolutionary process whereby polymorphism is maintained in a population because it confers some selective advantage [65].

Although it is rare in other host genes, balancing selection has been described in many immune genes and may be beneficial when, for example, different alleles are effective against distinct pathogens [66]. Selecting for multiple alleles at a single site (maintaining heterozygosity) may also make it more difficult for pathogens to evolve suitable escape

mutants against immune genes [67]. Here, we genotyped *TRIM22* nsSNP rs1063303:G>C in two distinct Inuit populations (Canadian and Greenlandic Inuit) as well as one non-Inuit population (Canadian population of European descent). In contrast to some ethnic groups (including the AFR, AMR, and EUR cohorts mentioned above), the Inuit have been geographically isolated for many years. As such, they have likely evolved against unique selective pressures and distinct pathogenic landscapes [68,69]. We determined the frequency of a functionally important *TRIM22* nsSNP (rs1063303:G>C) in these Inuit populations and examined the selective forces acting on this site. Of interest, we found that rs1063303:G>C is significantly more prevalent in the Inuit compared to non-Inuit populations, and that unlike AFR, AMR, and EUR cohorts, there is not an excess of intermediate frequency rs1063303:G>C alleles in the two Inuit populations, suggesting that this site has not evolved under balancing selection in these groups. We also found an unexpected, but striking, association between the *TRIM22* nsSNP rs1063303:G>C and serum levels of TG and HDL. The *TRIM22* rs1063303:C allele was associated with significantly lower serum TG levels and significantly higher serum HDL levels in the Canadian Inuit population. The effect of TRIM22 on serum TG and HDL levels is unprecedented in the TRIM family.

4.2 Materials and methods

4.2.1 Study subjects

Study subjects included Inuit individuals from Canada (Kivalliq region, Nunavut) and Greenland (Nuuk, Sisimiut, Qasigiannguit, and four villages in Uummannaq region), as well as Canadians of European descent (European Caucasian population). 42 subjects in the Canadian Inuit population self-reported as being of European background and were included in the analysis as a regional Caucasian control group [70]. All study subjects had previously been participants in one of the following population studies: 1) the Keewatin Health Assessment Study [71]; 2) the Greenland Population Study [72]; and 3) the Study of Health Assessment and Risk in Ethnic Groups [73]. The details of these studies have been described previously [71–73]. Signed informed consent was obtained from all participants

and the studies were approved by the Universities of Manitoba and Toronto, and from Western University.

4.2.2 Clinical characteristics and biochemical analyses

Body weight, height, and plasma lipoprotein analyses [fasting plasma concentrations of total cholesterol (TChol), TG, HDL-cholesterol, LDL-cholesterol, and apolipoproteins (Apo) A-I and B] were determined as previously described [71–73].

4.2.3 Genotype analyses

Detection of nsSNP rs1063303:G>C was carried out using a custom designed TaqMan genotyping assay (Applied Biosystems, Foster City, CA). The nsSNP genotyping was performed using an allelic discrimination assay with the 7900HT Fast Real-time PCR system and read using automated software (SDS 2.3 Applied Biosystems, Foster City, CA). PCR reactions were run in 5 μL volumes using an amplification protocol of 95°C for 10 minutes, followed by 50 cycles of 95°C for 15 sec, then 60°C for 1.3 minutes.

4.2.4 Neutrality tests

Tajima's *D* and Fu's F_S neutrality tests were performed to detect signals of selection at site rs1063303:G>C in the Canadian Inuit, Greenlandic Inuit, and European Caucasian populations. These tests differentiate between neutrally evolving sites under mutation-drift equilibrium and sites evolving under non-neutral processes such as balancing selection. Negative Tajima's *D* and/or Fu's F_S values are evidence of an excess of low frequency variants relative to expectation, whereas positive Tajima's *D* and/or Fu's F_S values are indicative of an excess of intermediate frequency variants. A *P* value of less than 0.02 was considered significant for the F_S statistic [74–76]. Statistical significance was assessed by coalescent simulations using Arlequin software (v 3.5.1.3) with 10,000 iterations [77].

4.2.5 Statistical analyses

Statistical analyses were performed using the SAS version 9.2 software (SAS Institute, Cary, NC). A χ^2 test was used to examine deviation in genotype frequency from the

Hardy-Weinberg equilibrium. ANOVA was used to determine sources of variation for study subjects' levels of TChol, TG, HDL-cholesterol, LDL-cholesterol, Apo-AI, and Apo-B. *TRIM22* rs1063303:G>C was used as the independent class variable, while sex and body mass index (BMI) were entered as covariates. *P* values less than 0.05 were considered statistically significant.

4.3 Results

4.3.1 Baseline phenotypic characteristics of study subjects

Baseline phenotypic characteristics for the Canadian Inuit, Greenlandic Inuit, and European Caucasian populations whose *TRIM22* rs1063303:G>C genotype and serum lipoprotein levels were determined are summarized in Table 4.1, including the baseline levels of total cholesterol (TChol), TG, HDL-cholesterol, LDL-cholesterol, Apo-AI, and Apo-B. Mean age, body mass index (BMI), and percentage of female study subjects are also shown in Table 4.1.

4.3.2 *TRIM22* rs1063303:G>C genotype and allele frequencies

TRIM22 rs1063303:G>C genotype and allele frequencies were calculated for the three populations (Canadian Inuit, Greenlandic Inuit, and European Caucasian) and are shown in Table 4.2. There were 21 G/G homozygotes (0.091), 62 heterozygotes (0.267), and 147 C/C homozygotes (0.634) in the Canadian Inuit population. Similarly, there were 20 G/G homozygotes (0.078), 104 heterozygotes (0.408), and 131 C/C homozygotes (0.514) in the Greenlandic Inuit population. Thus, rs1063303:C was the major allele in both the Canadian and Greenlandic Inuit populations, which had allele frequencies of 0.84 and 0.72, respectively. In contrast, the rs1063303:C allele was much less prevalent in the European Caucasian population as well as in a small subset of European Caucasian individuals within the Canadian Inuit (Table 4.2). Similar low rs1063303:C allele frequencies were reported previously for the European Caucasian population from the 1000 Genomes project (EUR) [64]. Moreover, rs1063303:C was even less prevalent in

Table 4.1

Table 4.1 Baseline phenotypic characteristics (mean ± SD) of Canadian Inuit, Greenland Inuit, and European Caucasian populations.

	Canadian Inuit	Greenland Inuit	European Caucasian
Number	232	254	67
Percent female (%)	50.4	62.9	59.7
Age (years)	38.1 ± 14.8	48.6 ± 13.3	57.7 ± 18.1
BMI (kg/m^2)	26.9 ± 4.84	27.1 ± 5.18	27.6 ± 5.26
Total cholesterol (mmol/L)	5.06 ± 1.01	6.35 ± 1.03	5.29 ± 0.86
Triglycerides (mmol/L)	1.17 ± 0.77	1.14 ± 0.68	1.20 ± 0.40
HDL cholesterol (mmol/L)	1.37 ± 0.39	1.67 ± 0.47	1.21 ± 0.33
LDL cholesterol (mmol/L)	3.16 ± 0.92	4.15 ± 1.04	3.73 ± 0.79
Apo-AI (g/L)	N/A	1.81 ± 0.30	1.41 ± 0.25
Apo-B (g/L)	N/A	0.99 ± 0.24	1.07 ± 0.21

Table 4.2

Table 4.2 Genotype frequencies for *TRIM22* rs1063303:G>C in the Canadian Inuit, Greenlandic Inuit, and European Caucasian populations.

Population	n	G/G	G/C	C/C
Canadian Inuit	232	0.091	0.267	0.634
Inuit	188	0.053	0.208	0.734
Caucasian	44	0.273	0.523	0.205
Greenland Inuit	255	0.078	0.408	0.514
European Caucasian	67	0.397	0.397	0.206
European (1000 Genomes*)	193	0.228	0.507	0.265
African (1000 Genomes*)	133	0.329	0.541	0.130
American (1000 Genomes*)	84	0.365	0.464	0.171
Asian (1000 Genomes*)	66	0.745	0.231	0.024

* 1000 Genomes populations: European (EUR) cohort includes British individuals from Scotland and England, Iberian individuals from Spain, Finnish individuals and Italian individuals; African (AFR) cohort includes Yoruba individuals from Nigeria, Luhya individuals from Kenya, and African Americans from the southwestern United States; American (AMR) cohort includes Colombian individuals, Puerto Rican individuals, and Mexican Americans from Los Angeles, United States; Asian (ASN) cohort includes Japanese individuals, Han Chinese individuals from Beijing, and Southern Han Chinese.

AFR, AMR, and ASN populations from the 1000 Genomes project [64]. Genotype frequencies determined for the Canadian Inuit, Greenlandic Inuit, and European Caucasian populations did not show significant deviation from expectations based on Hardy-Weinberg equilibrium ($P > 0.05$).

4.3.3 TRIM22 rs1063303:G>C is not evolving under balancing selection in the Canadian or Greenlandic Inuit

Since the Inuit have lived in isolation for many years, we hypothesized that within the Inuit, *TRIM22* rs1063303 may contain a unique evolutionary signature. We previously identified an excess of intermediate frequency rs1063303:C alleles in the AFR, AMR, and EUR, but not ASN, cohorts from the 1000 Genomes project. This excess differed significantly from what was expected under a neutral model of evolution ($P < 0.05$) [64]. We performed the Tajima's *D* and Fu's F_S neutrality tests to establish whether *TRIM22* rs1063303:G>C has evolved under neutral or non-neutral processes in the Inuit. As shown in Table 4.3, the Tajima's *D* values for both the Canadian and Greenlandic Inuit populations did not differ significantly ($P > 0.05$) from what was expected under a neutral model of evolution, suggesting that rs1063303:G>C has not evolved under balancing selection in these populations. While the Tajima's *D* value was higher in the European Caucasian population (1.61) compared to the Canadian and Greenlandic Inuit populations (1.20 and 1.29), it did not reach statistical significance. Within the Canadian Inuit, a smaller Tajima's *D* value (0.60) was achieved when the small subset of European Caucasian individuals was removed from the population. Similarly, the results of Fu's F_S test also showed non-significant values for the Canadian Inuit, Greenlandic Inuit, and European Caucasian populations (Table 4.3).

4.3.4 Association between *TRIM22* rs1063303:G>C and plasma lipoproteins in the Canadian Inuit

We previously demonstrated that the *TRIM22* rs1063303:C allele is associated with a significant increase in TRIM22 expression levels and a significant reduction in antiviral

Table 4.3

Table 4.3: Results of Tajima's *D* and Fu's F_s neutrality tests.

Population	**Tajima's *D***	**p-value**	**Fu's F_s**	**p-value**
Canadian Inuit	1.20	0.179	1.83	0.126
Inuit	0.60	0.272	1.05	0.294
Caucasian	1.11	0.19	1.17	0.167
Greenland Inuit	1.29	0.150	1.96	0.111
European Caucasian	1.61	0.114	1.78	0.092

activity [64]. Since TRIM22 has been implicated in a plethora of diverse biological roles, we hypothesized that a significant increase in TRIM22 expression due to the *TRIM22* rs1063303:C allele may contribute to other biological functions. We took advantage of the extensive clinical data associated with the Canadian and Greenlandic Inuit populations used in this study to determine if the *TRIM22* rs1063303:C allele was associated with a particular clinical phenotype. ANOVA was used to determine sources of variation for various serum lipoproteins, including TChol, TG, HDL, LDL, Apo-AI, and Apo-B, in Canadian Inuit, Greenlandic Inuit, and European Caucasian populations. The ANOVA results for the three populations are summarized in Table 4.4. Significant associations were found in the Canadian Inuit population between the *TRIM22* rs1063303:C allele and serum levels of both TG ($P < 0.0008$) and HDL ($P < 0.05$). TG and HDL associations were not found in the Greenlandic Inuit or the European Caucasian populations. Sex was found to significantly associate with HDL in all populations ($P < 0.005$), while BMI was significantly associated with most lipoproteins in all three populations. Sex was also significantly associated with Apo-AI ($P < 0.05$) and LDL ($P < 0.05$) in the Greenlandic Inuit population (Table 4.4).

TRIM22 rs1063303:G>C genotypes and their corresponding TG and HDL levels are shown in Table 4.5 for the Canadian Inuit population. A significant association between the presence of the rs1063303:C allele and decreased serum TG levels was detected, with C/C homozygotes having lower serum TG levels than heterozygotes and heterozygotes having lower serum TG levels than G/G homozygotes. This pattern suggests that the rs1063303:C allele shows incomplete dominance. A similar significant association was found between the rs1063303:C allele and elevated levels of plasma HDL (Table 4.5).

4.4 Discussion

Here, we showed that the *TRIM22* nsSNP rs1063303:G>C is inordinately prevalent in two Inuit populations (Canadian and Greenlandic) and that the rs1063303:C allele is

Table 4.4

Table 4.4: ANOVA results

Table 4.4: Summary of ANOVA results showing determinants of serum lipoproteins in the Canadian Inuit, Greenlandic Inuit, and European Caucasian populations.

		TChol		TG		HDL		LDL		Apo-A1		Apo-B	
Source of Variation	***df***	F	*P*	F	*P*	F	*P*	F	*P*	F	*P*	F	*P*
Canadian Inuit													
Sex	1	0.76	n.s.	0.51	n.s.	23.42	**<0.0001**	0.07	n.s.	1.19	n.s.	0.16	n.s.
BMI	1	3.89	**0.0454**	26.48	**<0.0001**	8.34	**0.0043**	0.75	n.s.	19.29	**<0.0001**	6.78	**0.0031**
TRIM22 rs1063303:C	2	2.21	n.s.	7.31	**0.0008**	3.49	**0.0323**	1.43	n.s.	0.67	n.s.	0.33	n.s.
Greenlandic Inuit													
Sex	1	0.07	n.s.	0.04	n.s.	36.37	**<0.0001**	11.93	**0.0006**	4.96	**0.0268**	0.44	n.s.
BMI	1	4.95	**0.0271**	51.94	**<0.0001**	2.97	n.s.	0.08	n.s.	10.42	**0.0014**	9.26	**0.0026**
TRIM22 rs1063303:C	2	0.65	n.s.	0.07	n.s.	0.18	n.s.	0.69	n.s.	0.24	n.s.	0.13	n.s.
European Caucasian													
Sex	1	0.47	n.s.	0.27	n.s.	8.76	**0.0043**	0.95	n.s.	0.17	n.s.	1.09	n.s.
BMI	1	5.11	**0.0112**	71.42	**<0.0001**	6.54	**0.0012**	1.22	n.s.	10.21	**0.0005**	2.32	n.s.
TRIM22 rs1063303:C	2	0.89	n.s.	0.54	n.s.	2.90	n.s.	1.79	n.s.	0.34	n.s.	0.71	n.s.

P value indicates probability of a greater between-group F value using ANOVA; n.s. indicates not significant (i.e. $P > 0.05$)

Table 4.5

Table 4.5: Significant associations between the *TRIM22* rs1063303:G>C genotype and plasma lipoproteins in the Canadian Inuit.

Genotype	TG (LSM ± SEM) (mmol/L)	HDL (LSM ± SEM) (mmol/L)
G/G	1.72 ± 0.09	1.18 ± 0.08
G/C	1.15 ± 0.08	1.34 ± 0.05
C/C	1.07 ± 0.06	1.41 ± 0.03

* LSM: least squares means, SEM: standard error of the mean

associated with significantly lower levels of serum TG and significantly higher levels of serum HDL in the Canadian Inuit. Unlike in several non-Inuit populations (i.e. AFR, AMR, and EUR cohorts from the 1000 Genomes project [64]), the two Inuit populations (Canadian and Greenlandic Inuit) did not contain a significant excess of intermediate frequency *TRIM22* rs1063303:G>C alleles when compared to a neutral evolutionary model. The latter, which is not indicative of balancing selection, suggests that this site (*TRIM22* rs1063303:G>C) has evolved in response to distinct evolutionary pressures in the Canadian and Greenlandic Inuit.

One possible explanation for this disparate evolutionary footprint in the Inuit is lack of exposure to certain pathogens. We previously showed that *TRIM22* rs1063303:G>C is located at site in TRIM22 that has been subject to strong positive selection during the evolution of mammals. Moreover, certain human populations (AFR, AMR, and EUR) contained an excess of intermediate frequency *TRIM22* rs1063303:G>C alleles, which indicates a decrease in population size and/or balancing selection [64]. In innate immune genes, amino acid sites that are subject to positive and/or balancing selection are often functionally relevant, typically because they are located at the host-pathogen interface where they are exposed to diverse pathogenic pressures [66]. Indeed, we demonstrated in our previous study that *TRIM22* rs1063303:G>C increased the expression of TRIM22, but disrupted its ability to restrict HIV-1 replication [64]. While the reason for this effect is currently unclear, these results combined with the fact that this site has evolved under multiple non-neutral evolutionary forces, strongly suggests that its evolution has been driven by HIV-1 and/or other pathogens. Canadian and Greenlandic Inuit populations, who have lived in isolation for many years with little exposure to 'modern' infectious diseases, may not have been subjected to the same pathogenic pressures at this site as other populations.

Despite advances in other non-Inuit populations, relatively few innate immune nsSNPs have been documented in the Inuit. Notably, there is evidence that the Inuit (and other indigenous populations) suffer disproportionately from infectious diseases and have a much higher rate of infection when compared to their non-indigenous counterparts [68]. While several socio-economic factors certainly contribute to this increased prevalence,

differences in their immune response may also play a role. Interestingly, the Inuit have been shown to possess a unique distribution of human leukocyte antigen (HLA) alleles when compared to non-Inuit populations [78–81]. Because specific HLA alleles have been linked with susceptibility to a number of viral infections, it has been suggested that the Inuit's unique distribution of HLA alleles may make them more vulnerable to certain infections [82–85]. Since TRIM22 and other important antiviral genes, such as TRIM5α, APOBEC3G, BST-2/tetherin, and SAMHD1, are located at the host-pathogen interface where they often interact directly with viral pathogens, it is possible that the Inuit also contain a unique distribution of alleles in these genes and that this affects their antiviral response. It would be interesting to clone several of these antiviral genes from the Inuit population and test their antiviral activity against the spectrum of viruses known to be inhibited by these proteins. One could also compare various innate immune responses among the Inuit and other non-Inuit populations.

Surprisingly, we found an association between *TRIM22* rs1063303:G>C and the serum levels of specific lipoproteins. In the Canadian Inuit population, the rs1063303:C allele was associated with significantly lower serum TG levels and significantly higher serum HDL levels. These TG and HDL associations were not found in the Greenlandic Inuit or the European Caucasian populations. The effect of TRIM22 on serum lipoprotein levels is unprecedented in the TRIM protein family. The molecular details underlying this effect are currently unexplored and may be explained by a number of diverse factors given the wide ranging functions of TRIM22. It is possible, for example, that TRIM22 modulates serum lipoprotein levels indirectly by influencing innate immune and/or inflammatory signaling pathways. Recent studies have implicated multiple TRIM proteins in the regulation of key innate signaling pathways, particularly pathways that control NF-κB signaling, the IFN response, and cytokine production following PRR activation [40,42,44]. PRRs, which include the membrane-bound TLRs and cytoplasmic receptors such as RLRs, are activated by diverse microbial products termed PAMPs. TRIM proteins have been shown to regulate NF-κB activation, either positively or negatively, by targeting various stages of the NF-κB signaling pathway. For example, TRIM19/PML sequesters NF-κB in PML nuclear bodies to prevent it from activating downstream transcription, whereas TRIM30α and TRIM27 repress NF-κB signaling by targeting TAB2/TAB3 or

IKK family members, respectively [86–88]. TRIM23, which is essential for viral-induced NF-κB activation, enhances downstream antiviral signaling by modifying NEMO with ubiquitin [89]. Although it remains unclear whether endogenous levels of TRIM22 can regulate NF-κB signaling, one recent study demonstrated that TRIM22 overexpression inhibited the TRAF6-stimulated NF-κB pathway by facilitating TAB2 degradation [69]. However, TRIM22 overexpression has also been shown to activate NF-κB signaling [70].

Since NF-κB activates the transcription of proinflammatory cytokines, and because the APR is induced by proinflammatory cytokines, TRIM22 expression may inadvertently alter the APR by regulating NF-κB signaling. In this scenario, nsSNP rs1063303:G>C would either: 1) disrupt TRIM22-mediated NF-κB activation, or 2) amplify TRIM22-mediated NF-κB inhibition. Both possibilities would theoretically lead to a decrease in proinflammatory cytokine production and a less potent APR. Given the integral role played by TRIM proteins in the antiviral response, and the link between the immune system and the APR, other TRIM family members may have a similar effect on serum lipoproteins. Future studies are needed to address these possibilities and others and to determine the molecular details underlying TRIM22-mediated effects on serum TG and HDL levels.

4.5 References

1. Gabay, C. & Kushner, I. Acute-phase proteins and other systemic responses to inflammation. *N. Engl. J. Med.* **340,** 448–54 (1999).

2. Feingold, K. R. *et al.* Effect of interleukin-1 on lipid metabolism in the rat. Similarities to and differences from tumor necrosis factor. *Arterioscler. Thromb.* **11,** 495–500

3. Feingold, K. R. *et al.* Multiple cytokines stimulate hepatic lipid synthesis in vivo. *Endocrinology* **125,** 267–74 (1989).

4. Grunfeld, C. *et al.* Lipids, lipoproteins, triglyceride clearance, and cytokines in human immunodeficiency virus infection and the acquired immunodeficiency syndrome. *J. Clin. Endocrinol. Metab.* **74,** 1045–52 (1992).

5. Sammalkorpi, K., Valtonen, V., Kerttula, Y., Nikkilä, E. & Taskinen, M. R. Changes in serum lipoprotein pattern induced by acute infections. *Metabolism.* **37,** 859–65 (1988).

6. Rosenzweig, I. B. *et al.* Effects of interleukin-2 (IL-2) on human plasma lipid, lipoprotein, and C-reactive protein. *Biotherapy* **2,** 193–8 (1990).

7. Memon, R. A., Grunfeld, C., Moser, A. H. & Feingold, K. R. Tumor necrosis factor mediates the effects of endotoxin on cholesterol and triglyceride metabolism in mice. *Endocrinology* **132,** 2246–53 (1993).

8. Ettinger, W. H., Miller, L. D., Albers, J. J., Smith, T. K. & Parks, J. S. Lipopolysaccharide and tumor necrosis factor cause a fall in plasma concentration of lecithin: cholesterol acyltransferase in cynomolgus monkeys. *J. Lipid Res.* **31,** 1099–107 (1990).

9. Vandenbroeck, K. Cytokine gene polymorphisms and human autoimmune disease in the era of genome-wide association studies. *J. Interferon Cytokine Res.* **32,** 139–51 (2012).

10. Markovic, O. *et al.* Role of single nucleotide polymorphisms of pro-inflammatory cytokine genes in the relationship between serum lipids and inflammatory parameters, and the lipid-lowering effect of fish oil in healthy males. *Clin. Nutr.* **23,** 1084–95 (2004).

11. Joffe, Y. T., Collins, M. & Goedecke, J. H. The relationship between dietary fatty acids and inflammatory genes on the obese phenotype and serum lipids. *Nutrients* **5,** 1672–705 (2013).

12. Bonnet, E. *et al.* Impact of genetic polymorphisms on the risk of lipid disorders in patients on anti-HIV therapy. *Clin. Chem. Lab. Med.* **45,** 815–21 (2007).

13. Cuda, C., Garcia-Bailo, B., Karmali, M., El-Sohemy, A. & Badawi, A. A common polymorphism near the interleukin-6 gene modifies the association between dietary fat intake and insulin sensitivity. *J. Inflamm. Res.* **5,** 1–6 (2012).

14. Fontaine-Bisson, B., Wolever, T. M. S., Connelly, P. W., Corey, P. N. & El-Sohemy, A. NF-kappaB -94Ins/Del ATTG polymorphism modifies the association between dietary polyunsaturated fatty acids and HDL-cholesterol in two distinct populations. *Atherosclerosis* **204,** 465–70 (2009).

15. Fontaine-Bisson, B. *et al.* Genetic polymorphisms of tumor necrosis factor-alpha modify the association between dietary polyunsaturated fatty acids and fasting HDL-cholesterol and apo A-I concentrations. *Am. J. Clin. Nutr.* **86,** 768–74 (2007).

16. Shortridge, K. F., Ho, W. K., Oya, A. & Kobayashi, M. Studies on the inhibitory activities of human serum lipoproteins for Japanese encephalitis virus. *Southeast Asian J. Trop. Med. Public Health* **6,** 461–6 (1975).

17. Huemer, H. P. *et al.* Herpes simplex virus binds to human serum lipoprotein. *Intervirology* **29,** 68–76 (1988).

18. Singh, I. P., Chopra, A. K., Coppenhaver, D. H., Ananatharamaiah, G. M. & Baron, S. Lipoproteins account for part of the broad non-specific antiviral activity of human serum. *Antiviral Res.* **42,** 211–8 (1999).

19. Halonen, P. E., Toivanen, P. & Nikkari, T. Non-specific serum inhibitors of activity of haemagglutinins of rabies and vesicular stomatitis viruses. *J. Gen. Virol.* **22,** 309–18 (1974).

20. Ho, W. K. Serum lipoproteins as inhibitors of haemafflutination by rubella virus. *Lipids* **12,** 85–91 (1977).

21. Martin, I., Dubois, M. C., Saermark, T. & Ruysschaert, J. M. Apolipoprotein A-1 interacts with the N-terminal fusogenic domains of SIV (simian immunodeficiency virus) GP32 and HIV (human immunodeficiency virus) GP41: implications in viral entry. *Biochem. Biophys. Res. Commun.* **186,** 95–101 (1992).

22. Fischer, D. G., Tal, N., Novick, D., Barak, S. & Rubinstein, M. An antiviral soluble form of the LDL receptor induced by interferon. *Science* **262,** 250–3 (1993).

23. Marlovits, T. C. *et al.* Recombinant soluble low density lipoprotein receptor fragment inhibits minor group rhinovirus infection in vitro. *FASEB J.* **12,** 695–703 (1998).

24. Hofer, F. *et al.* Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 1839–42 (1994).

25. Harris, H. W., Eichbaum, E. B., Kane, J. P. & Rapp, J. H. Detection of endotoxin in triglyceride-rich lipoproteins in vitro. *J. Lab. Clin. Med.* **118,** 186–93 (1991).

26. Harris, H. W., Grunfeld, C., Feingold, K. R. & Rapp, J. H. Human very low density lipoproteins and chylomicrons can protect against endotoxin-induced death in mice. *J. Clin. Invest.* **86,** 696–702 (1990).

27. Barcia, A. M. & Harris, H. W. Triglyceride-rich lipoproteins as agents of innate immunity. *Clin. Infect. Dis.* **41 Suppl 7,** S498–503 (2005).

28. Grunfeld, C. *et al.* Lipoproteins inhibit macrophage activation by lipoteichoic acid. *J. Lipid Res.* **40,** 245–52 (1999).

29. Van Lenten, B. J., Fogelman, A. M., Haberland, M. E. & Edwards, P. A. The role of lipoproteins and receptor-mediated endocytosis in the transport of bacterial lipopolysaccharide. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 2704–8 (1986).

30. Hansson, G. K., Libby, P., Schönbeck, U. & Yan, Z.-Q. Innate and adaptive immunity in the pathogenesis of atherosclerosis. *Circ. Res.* **91,** 281–91 (2002).

31. Steinhardt, A. P. *et al.* A functional nonsynonymous toll-like receptor 4 gene polymorphism is associated with metabolic syndrome, surrogates of insulin resistance, and syndromes of lipid accumulation. *Metabolism.* **59,** 711–7 (2010).

32. Li, J. H. *et al.* Interferon-lambda genotype and low serum low-density lipoprotein cholesterol levels in patients with chronic hepatitis C infection. *Hepatology* **51,** 1904–11 (2010).

33. Arora, P. *et al.* Genetic polymorphisms of innate immunity-related inflammatory pathways and their association with factors related to type 2 diabetes. *BMC Med. Genet.* **12,** 95 (2011).

34. Fontaine-Bisson, B. *et al.* Tumor necrosis factor alpha -238G>A genotype alters postprandial plasma levels of free fatty acids in obese individuals with type 2 diabetes mellitus. *Metabolism.* **56,** 649–55 (2007).

35. Feingold, K. R. & Grunfeld, C. Role of cytokines in inducing hyperlipidemia. *Diabetes* **41 Suppl 2,** 97–101 (1992).

36. Nikolic, V. N. *et al.* An inverse correlation between TNF alpha serum levels and heart rate variability in patients with heart failure. *J. Cardiol.* **62,** 37–43 (2013).

37. Celis, R., Torre-Martinez, G. & Torre-Amione, G. Evidence for activation of immune system in heart failure: is there a role for anti-inflammatory therapy? *Curr. Opin. Cardiol.* **23,** 254–60 (2008).

38. Bieghs, V. & Trautwein, C. The innate immune response during liver inflammation and metabolic disease. *Trends Immunol.* **34,** 446–52 (2013).

39. Markus G Grütter and Jeremy Luban. TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. *Curr Opin Virol* **2,** 142–150 (2012).

40. McNab, F. W., Rajsbaum, R., Stoye, J. P. & O'Garra, A. Tripartite-motif proteins and innate immune regulation. *Curr. Opin. Immunol.* **23,** 46–56 (2011).

41. Carthagena, L. *et al.* Human TRIM gene expression in response to interferons. *PLoS One* **4,** e4894 (2009).

42. Kawai, T. & Akira, S. Regulation of innate immune signalling pathways by the tripartite motif (TRIM) family proteins. *EMBO Mol. Med.* **3,** 513–27 (2011).

43. Versteeg, G. A. *et al.* The E3-ligase TRIM family of proteins regulates signaling pathways triggered by innate immune pattern-recognition receptors. *Immunity* **38,** 384–98 (2013).

44. Rajsbaum, R., García-Sastre, A. & Versteeg, G. A. TRIMmunity: The Roles of the TRIM E3-Ubiquitin Ligase Family in Innate Antiviral Immunity. *J. Mol. Biol.* **426,** 1265–1284 (2014).

45. Reymond, A. *et al.* The tripartite motif family identifies cell compartments. *EMBO J.* **20,** 2140–51 (2001).

46. Meroni, G. & Diez-Roux, G. TRIM/RBCC, a novel class of "single protein RING finger" E3 ubiquitin ligases. *Bioessays* **27,** 1147–57 (2005).

47. Chu, Yaya and Yang, X. SUMO E3 ligase activity of TRIM proteins. *Oncogene* **30,** 1108–1116 (2011).

48. James, L. C., Keeble, A. H., Khan, Z., Rhodes, D. A. & Trowsdale, J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6200–5 (2007).

49. Biris, N. *et al.* Structure of the rhesus monkey TRIM5α PRYSPRY domain, the HIV capsid recognition module. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13278–13283 (2012).

50. Hattlmann, C. J., Kelly, J. N. & Barr, S. D. TRIM22: A Diverse and Dynamic Antiviral Protein. *Mol. Biol. Int.* **2012,** 153415 (2012).

51. Barr, S. D., Smiley, J. R. & Bushman, F. D. The interferon response inhibits HIV particle production by induction of TRIM22. *PLoS Pathog.* **4,** e1000007 (2008).

52. Singh, R. *et al.* Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. *J. Virol.* **85,** 208–16 (2011).

53. Gao, B., Duan, Z., Xu, W. & Xiong, S. Tripartite motif-containing 22 inhibits the activity of hepatitis B virus core promoter, which is dependent on nuclear-located RING domain. *Hepatology* **50,** 424–33 (2009).

54. Di Pietro, A. *et al.* TRIM22 inhibits influenza A virus infection by targeting the viral nucleoprotein for degradation. *J. Virol.* **87,** 4523–33 (2013).

55. Obad, S., Olofsson, T., Mechti, N., Gullberg, U. & Drott, K. Regulation of the interferon-inducible p53 target gene TRIM22 (Staf50) in human T lymphocyte activation. *J. Interferon Cytokine Res.* **27,** 857–64 (2007).

56. Obad, S. *et al.* Staf50 is a novel p53 target gene conferring reduced clonogenic growth of leukemic U-937 cells. *Oncogene* **23,** 4050–9 (2004).

57. Nexø, B. A. *et al.* Restriction genes for retroviruses influence the risk of multiple sclerosis. *PLoS One* **8,** e74063 (2013).

58. Deng, Y. J. *et al.* Gene profiling involved in immature CD4+ T lymphocyte responsible for systemic lupus erythematosus. *Mol. Immunol.* **43,** 1497–507 (2006).

59. Coit, P. *et al.* Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naïve CD4+ T cells from lupus patients. *J. Autoimmun.* **43,** 78–84 (2013).

60. Zirn, B. *et al.* Expression profiling of Wilms tumors reveals new candidate genes for different clinical parameters. *Int. J. Cancer* **118,** 1954–62 (2006).

61. Wittmann, S. *et al.* New prognostic markers revealed by evaluation of genes correlated with clinical parameters in Wilms tumors. *Genes. Chromosomes Cancer* **47,** 386–95 (2008).

62. Carotenuto, M. *et al.* Neuroblastoma tumorigenesis is regulated through the Nm23-H1/h-Prune C-terminal interaction. *Sci. Rep.* **3,** 1351 (2013).

63. Sun, Y. *et al.* Down-regulation of tripartite-motif containing 22 expression in breast cancer is associated with a lack of p53-mediated induction. *Biochem. Biophys. Res. Commun.* **441,** 600–6 (2013).

64. Kelly, J. N., Woods, M. W., Xhiku, S. & Barr, S. D. Ancient and Recent Adaptive Evolution in the Antiviral TRIM22 Gene: Identification of a Single Nucleotide Polymorphism that Impacts TRIM22 Function. *Hum. Mutat.* (2014). doi:10.1002/humu.22595

65. Andrés, A. M. *et al.* Targets of balancing selection in the human genome. *Mol. Biol. Evol.* **26,** 2755–64 (2009).

66. Ferrer-Admetlla, A. *et al.* Balancing selection is the main force shaping the evolution of innate immunity genes. *J. Immunol.* **181,** 1315–22 (2008).

67. Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12,** 687–95 (2012).

68. Gracey, M. & King, M. Indigenous health part 1: determinants and disease patterns. *Lancet* **374,** 65–75 (2009).

69. Bjerregaard, P., Young, T. K., Dewailly, E. & Ebbesson, S. O. E. Indigenous health in the Arctic: an overview of the circumpolar Inuit population. *Scand. J. Public Health* **32,** 390–5 (2004).

70. Hegele, R. a, Young, T. K. & Connelly, P. W. Are Canadian Inuit at increased genetic risk for coronary heart disease? *J. Mol. Med. (Berl).* **75,** 364–70 (1997).

71. Moffatt, M. E. *et al.* The Keewatin Health Assessment Study, NWT, Canada. *Arctic Med. Res.* **52,** 18–21 (1993).

72. Bjerregaard, P. *et al.* Inuit health in Greenland: a population survey of life style and disease in Greenland and among Inuit living in Denmark. *Int. J. Circumpolar Health* **62 Suppl 1,** 3–79 (2003).

73. Anand, S. S. *et al.* The Study of Health Assessment and Risk in Ethnic groups (SHARE): rationale and design. The SHARE Investigators. *Can. J. Cardiol.* **14,** 1349–57 (1998).

74. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–95 (1989).

75. Tajima, F. The effect of change in population size on DNA polymorphism. *Genetics* **123,** 597–601 (1989).

76. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915–925 (1997).

77. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10,** 564–7 (2010).

78. Metcalfe, S. *et al.* The frequency of HLA alleles in a population of Inuit women of northern Quebec. *Int. J. Circumpolar Health* **72,** (2013).

79. Welinder, L., Graugaard, B. & Madsen, M. HLA antigen and gene frequencies in Eskimos of East Greenland. *Eur. J. Immunogenet.* **27,** 93–7 (2000).

80. Wiseman, M. C., Orr, P. H., Macdonald, S. M., Schroeder, M. L. & Toole, J. W. Actinic prurigo: clinical features and HLA associations in a Canadian Inuit population. *J. Am. Acad. Dermatol.* **44,** 952–6 (2001).

81. Harvald, B. Genetic epidemiology of Greenland. *Clin. Genet.* **36,** 364–7 (1989).

82. Quiñones-Parra, S. *et al.* Preexisting CD8+ T-cell immunity to the H7N9 influenza A virus varies across ethnicities. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 1049–54 (2014).

83. Hertz, T. *et al.* HLA targeting efficiency correlates with human T-cell response magnitude and with mortality from influenza A infection. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13492–7 (2013).

84. Duvvuri, V. R. *et al.* Preexisting CD4+ T-cell immunity in human population to avian influenza H7N9 virus: whole proteome-wide immunoinformatics analyses. *PLoS One* **9,** e91273 (2014).

85. Metcalfe, S. *et al.* The association between human leukocyte antigen (HLA)-G polymorphisms and human papillomavirus (HPV) infection in Inuit women of northern Quebec. *Hum. Immunol.* **74,** 1610–5 (2013).

86. Zha, J. *et al.* The Ret finger protein inhibits signaling mediated by the noncanonical and canonical IkappaB kinase family members. *J. Immunol.* **176,** 1072–80 (2006).

87. Shi, M. *et al.* TRIM30 alpha negatively regulates TLR-mediated NF-kappa B activation by targeting TAB2 and TAB3 for degradation. *Nat. Immunol.* **9,** 369–77 (2008).

88. Wu, W.-S. *et al.* Promyelocytic leukemia protein sensitizes tumor necrosis factor alpha-induced apoptosis by inhibiting the NF-kappaB survival pathway. *J. Biol. Chem.* **278,** 12294–304 (2003).

89. Arimoto, K. *et al.* Polyubiquitin conjugation to NEMO by triparite motif protein 23 (TRIM23) is critical in antiviral defense. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15856–61 (2010).

Chapter 5

5 Discussion

5.1 Summary of results

5.1.1 Ancient and recent adaptive evolution in the antiviral *TRIM22* gene: identification of a single nucleotide polymorphism that impacts TRIM22 function

The evolution of TRIM22 in mammals was examined using 29 evolutionarily diverse mammalian TRIM22 sequences. These sequences were aligned using COBALT and a phylogenetic tree was generated using EvolView software. Site-specific evolutionary analysis of TRIM22 with the Selecton program identified multiple amino acid sites in TRIM22 that have evolved under strong positive selection. The majority of these sites were located in the PRY/SPRY (B30.2) domain; however, there was also evidence for positive selection in other TRIM22 domains. Many positively selected sites clustered around putative functional motifs in TRIM22, such as the zinc-finger motif in the BB2 domain and the NLS in the SP2 domain. Moreover, a number of sites corresponded in location and spacing to sites that have evolved under positive selection in the closely-related TRIM5α protein. For example, in both TRIM22 and TRIM5α, many positively selected sites are located in one of four variable regions (v1-v4) in the B30.2 domain.

In addition to sites undergoing positive selection among mammals, nsSNPs in human TRIM22 were investigated. A total of 64 nsSNPs and 2 indels were obtained from the NCBI dbSNP database for the human *TRIM22* gene, including 56 missense mutation-inducing nsSNPs and 8 frameshift mutation-inducing nsSNPs. To identify nsSNPs in TRIM22 that may be functionally relevant, an *in silico* prediction program (SIFT) was used to analyze the 56 missense mutation-inducing nsSNPs. SIFT predicted that 23 of these nsSNPs were deleterious to TRIM22 function and 33 were tolerated. Notably, 2 potentially deleterious nsSNPs were located at sites that evolved under strong positive

selection in mammals. One of these nsSNPs, rs1063303:G>C, was selected for further analysis because of its high prevalence in the human population.

Large frequency differences were observed for nsSNP rs1063303:G>C among distinct ethnic populations, including AFR, AMR, ASN, and EUR 1000 Genomes cohorts. Of interest, an excess of intermediate frequency rs1063303 alleles was identified in AFR, AMR, and EUR populations, indicating a decrease in population size and/or balancing selection. To assess potential functional consequences of nsSNP rs1063303:G>C, its RNA and protein expression, sub-cellular localization pattern, and anti-HIV-1 activity were determined and compared to the wild type TRIM22 protein. Surprisingly, nsSNP rs1063303:G>C significantly increased both RNA and protein expression of TRIM22, but disrupted its anti-HIV-1 activity. nsSNP rs1063303:G>C also obstructed TRIM22 sub-cellular localization (localized diffusely in both the cytoplasm and nucleus, not in punctate NBs). Taken together, these results describe multiple sites that have evolved under positive selection in TRIM22 and identify a highly prevalent functional nsSNP (rs1063303:G>C) with a complex evolutionary history.

5.1.2 *In silico* analysis of functional single nucleotide polymorphisms in the human *TRIM22* gene

To identify additional nsSNPs that may alter TRIM22 function, and to examine amino acid sites in TRIM22 that may be subject to post-translational modification (PTM), an extensive *in silico* analysis was performed on the protein coding region of the *TRIM22* gene. All missense mutation-inducing nsSNPs (56) in TRIM22 were analyzed using 6 different nsSNP prediction algorithms, including Polyphen-2, PhD-SNP, SIFT, nsSNP Analyzer, PMUT, and SNPs&GO. Since these algorithms use different parameters to evaluate and rank nsSNPs, multiple algorithms were used to increase the accuracy and power of prediction. A total of 14 nsSNPs were predicted to be deleterious to TRIM22 function by ≥ 4 nsSNP prediction algorithms. These 14 nsSNPs (L68R, H73R, E135K, I234K, S244L, G346S, K364N, P403T, L432W, R442C, F456I, T460I, C494F) were classified as high-risk deleterious and selected for further *in silico* analysis.

Highly conserved amino acids tend to be required for important protein functions. As such, nsSNPs that are located at highly conserved sites are often deleterious to protein function [1]. ConSurf analysis revealed that 13 of the 14 sites occupied by the high-risk nsSNPs were highly conserved (conservation score of 7-9). In addition, by combining evolutionary conservation data and solvent accessibility predictions, the ConSurf web server predicted that T460 was a key structural residue and that L68, K364, and P403 were key functional residues. Structural analysis of the 9 high-risk deleterious nsSNPs located in the B30.2 domain, including K364N, P403T, and T460I, showed that all of these nsSNPs markedly altered the putative 3D structure of TRIM22's B30.2 domain, particularly the surface-exposed v2 and v3 regions. These same regions are critical for HIV-1 restriction in the closely-related TRIM5α protein [2,3].

A number of putative PTM sites were also identified in the TRIM22 protein, including multiple ubiquitylation (3), sumoylation (1), and phosphorylation (21) sites that were predicted to undergo PTM by two or more *in silico* programs. Moreover, 7 SIMs were identified in TRIM22, 2 of which are also present in TRIM5α (ILGV and VIGL) and were previously shown to be required for its antiviral activity [4,5]. These 2 SIMs, plus 3 additional SIMs (5/7), are highly conserved among TRIM22 orthologues. Of interest, several PTM sites coincide with the location of nsSNPs, including 2 high-risk nsSNPs (S244L and T460I, which are both predicted to undergo phosphorylation). This study comprises the first systematic *in silico* analysis of functional sites in the *TRIM22* gene and will be a valuable resource for future studies.

5.1.3 The *TRIM22* nsSNP rs1063303:G>C is not evolving under balancing selection in the Inuit and is associated with low serum TG and high serum HDL levels in the Canadian Inuit

To determine the frequency of *TRIM22* nsSNP rs1063303:G>C in the Inuit and examine the selective forces acting on this site, we genotyped *TRIM22* rs1063303:G>C in two different Inuit populations and one non-Inuit population (Canadian Inuit, Greenlandic Inuit, and Canadian population of European Caucasian descent). Interestingly, we found that the *TRIM22* rs1063303:C allele is inordinately prevalent in both Inuit populations

and that unlike in the AFR, AMR, and EUR cohorts from the 1000 Genomes project, the Canadian and Greenlandic Inuit populations do not contain an excess of intermediate frequency *TRIM22* rs1063303:G>C alleles. The latter indicates that in these two Inuit populations *TRIM22* nsSNP rs1063303:G>C is not evolving under balancing selection. We also found an unexpected, but interesting association between the *TRIM22* nsSNP rs1063303:G>C and serum lipoprotein levels. Specifically, the *TRIM22* rs1063303:C allele was associated with significantly lower serum TG levels and significantly higher serum HDL levels in the Canadian Inuit population. The effect on TRIM22 on TG and HDL levels is unprecedented in the TRIM family and may represent an exiting new avenue of research for TRIM22 and other TRIM family members.

5.2 Multiple sites in TRIM22 have evolved under positive and/or balancing selection

5.2.1 Positive selection

Genetic conflict between host and viral genomes often results in the accumulation of a large number of non-synonymous (dN) relative to synonymous (dS) mutations in host and/or viral genes. While the majority of host genes evolve under negative (purifying) selection, which removes deleterious nsSNPs from genes to preserve protein function, host restriction factors tend to evolve under positive (directional) selection (defined as dN/dS > 1) [6–8]. This is largely due to the evolutionary 'arms race' that occurs between host restriction factors and viruses as they attempt to gain evolutionary superiority over each other. Host restriction factors, for example, are often targeted by viral antagonists and/or interact directly with viral proteins as part of their antiviral mechanism. For this reason, they tend to select for novel nsSNPs that allow them to evade viral antagonists and/or enhance their antiviral capacity (e.g. nsSNPs that increase viral protein binding affinity). However, since viruses mutate frequently and select for their own beneficial nsSNPs (e.g. nsSNPs that help circumvent the immune response and/or augment viral replication), both host restriction factors and viruses are under tremendous pressure to evolve new, more effective ways to counteract each other [9]. As a result, multiple host

restriction factors contain genetic signatures of positive selection, particularly at amino acid sites that interact with viral antagonists. It follows that a number of recent studies have conducted evolutionary analyses on host restriction factors to pinpoint amino acid sites that are important for their antiviral activity [10–12].

In this work, we used a similar approach to examine 29 evolutionarily diverse *TRIM22* sequences spanning >100 million years of evolution in mammals. While we identified codons evolving under positive selection in all TRIM22 protein domains, 11 of the 28 codons were located in the B30.2 domain (Table 2.3). Moreover, 10 of these 11 codons were located within one of four variable loops (v1-v4) in the B30.2 (Fig 2.1d). v1-v4, which are also found in the closely-related TRIM5α protein, are required for TRIM5α-mediated retroviral restriction [2,13]. In the rhTRIM5α protein, v1-v4 form an extensive, highly flexible, HIV-1 capsid binding interface [3]. The v1-v4 interface, which includes several positively selected codons in the v1 and v2 loops (324, 332, 385, and 389), are critical TRIM5α-mediated inhibition of HIV and/or SIV in hominoids [2,14–16]. There are also two positively selected codons in the v3 loop of huTRIM5α (409 and 410) that are required for N-MLV restriction [17]. We identified 6 positively selected codons (K324, R327, T330, K332, S334, and C337) in the v1 loop of TRIM22, 2 in the v2 loop (S377 and S395), and 2 in the v4 loop (i.e. L488 and V489). Many of these codons mirrored both the location and spacing of positively selected codons in the TRIM5α protein (Fig 2.1c). Similar to TRIM5α, TRIM22 has been shown to interact with the HIV-1 Gag or capsid protein; however, the binding site for capsid is unknown [18–20]. It is possible that the v1-v4 loops also comprise an HIV-1 capsid binding interface in TRIM22 and that some of the positively selected codons in v1, v2, and/or v4 are important for TRIM22-mediated HIV-1 restriction. Codons K324 and K332 are particularly interesting given their importance in TRIM5α-mediated retroviral restriction and because both sites have also evolved under positive selection in TRIM5α. Some of the codons evolving under positive selection in TRIM22's v1-v4 loops may alternatively be required for TRIM22-mediated inhibition of other viruses, such as HBV and/or IAV, or help form a binding site for interacting with viral proteins other than Gag. Previous studies have shown that TRIM22 interacts with the EMCV 3C protease and the IAV NP; however, the binding sites for these viral proteins are unknown [21,22]. One interesting difference between the

TRIM22 and TRIM5α proteins is not there are no positively selected codons in the v3 loop of TRIM22 (Fig 2.1d). In contrast, there are several positively selected codons in TRIM5α's v3 loop, some of which are required for inhibition of N-MLV [17]. Thus, it is possible that TRIM22, which does not inhibit N-MLV replication, does not contain any positively selected codons in its v3 loop because it has not been subject to evolutionary pressure from N-MLV [18]. Future studies are needed to address these possibilities and to establish the functional implications of the positively selected sites found in TRIM22's B30.2 domain.

Outside of the B30.2 domain, we identified 5 positively selected codons in the RING domain (D2, F3, S4, S50, and S54), 3 in the SP1 domain (T61, N76, and V96), 2 in the BB2 domain (Q105, I106), 3 in the CC domain (A171, V192, and T220), and 4 in the SP2 domain (L241, R242, K257, R279) (Table 2.3). Of interest, many of these codons are located close to putative protein binding sites in the TRIM22 protein. For example, codons V96, Q105, and I106 cluster around the zinc finger motif in the BB2 domain. Codons L241, R242, K257, and R279 flank the bipartite NLS in the SP2 domain (Fig 2.1b, Fig 2.2). Consistent with this, previous studies have shown that codons evolving under positive selection tend to be located near protein binding sites and are typically solvent-exposed [23]. Although the zinc finger motif and bipartite NLS in TRIM22 have never explicitly been shown to function as protein binding sites, these motifs are often involved in protein-protein interactions in other proteins [24,25]. Further, the presence of positively selected codons near these motifs suggests that they may indeed function as protein binding sites in TRIM22. Interestingly, our ConSurf results (Chapter 3) showed that 24 of the 28 positively selected codons are solvent-exposed (codons I106, V192, L241, and S395 are predicted to be buried) (Fig 3.1).

Future studies should focus on determining the functional significance of the positively selected codons that we identified in TRIM22. For example, it will be important to test whether these codons are involved in TRIM22's antiviral activity against HIV-1, HBV, IAV, and/or EMCV. In addition, examining whether specific codons affect TRIM22's ability to interact with viral proteins, such as the HIV-1 Gag and/or capsid protein, the IAV NP, and the EMCV 3C protease, will be of great interest and may provide further

insight into TRIM22's antiviral mechanism. Several codons within the B30.2 domain, such as K324 and K332, may be particularly relevant in this regard. Other codons may also be interesting to examine, including two codons (T61 and R242) that coincide in location with the *TRIM22* nsSNPs rs192306924:C>A and rs1063303:G>C. Curiously, these nsSNPs were predicted to be deleterious by some of the prediction programs in Chapter 3; however, they are located at amino acid sites that have evolved under strong positive selection in mammals (Tables 2.3, 3.1). Due to its high prevalence in humans, we examined the *TRIM22* nsSNP rs1063303:G>C in more detail in Chapters 2 and 4 (discussed in sections below). Finally, some codons may be involved in protein-protein interactions. For example, codons V96, Q105, and I106 surround a zinc finger motif in the BB2 domain and are also located amongst 2 putative hydrophobic surface patches (also in the BB2). Zinc finger motifs are often involved in protein-protein interactions and the zinc finger motifs found in the BB2 domains of other TRIM proteins have been shown to facilitate homo- and/or heterodimerization [24,26–28]. Notably, V96, Q105, and I106 are located amongst 2 putative hydrophobic surface patches in TRIM22. Similar patches were previously identified in the BB2 of TRIM5α and are required for proper TRIM5α self-association, capsid binding, and HIV-1 restriction [29]. While TRIM22 has been shown to undergo trimerization, the residues responsible for self-association are unknown and it is unclear whether self-association is required for TRIM22 function or antiviral activity [30]. For these reasons, it will be interesting to examine if V96, Q105, and/or I106 are involved in TRIM22 self-association and if self-association influences TRIM22 function and/or antiviral activity.

5.2.2 Balancing selection

In contrast to positive selection, which involves transient genetic diversity and drives advantageous alleles to fixation in a population, balancing selection actively maintains polymorphism (multiple alleles) at selected loci in a population [31,32]. While it is rare in other host genes, balancing selection has been identified in a number of immune genes and may be beneficial in specific pathogenic environments [33]. For example, one well-known nsSNP in the β-globin gene induces sickle-cell disease in homozygotes (sickle-

cell mutation), but affords protection from malaria in heterozygotes. Even though this mutation is deleterious in homozygous individuals, it is inordinately prevalent in high-malaria environments because it lowers mortality in the overall population [34]. In other words, balancing selection selects for heterozygosity in the population because in this environment it leads to fewer deaths. Heterozygosity may also be beneficial when, for example, different alleles are effective against different pathogens. In addition, having multiple alleles at specific loci in immune genes may make it harder for pathogens to evolve suitable escape mutants and evade immune surveillance [6].

Several host restriction factors, including huTRIM5α and OAS1, have been shown to undergo balancing selection in primates [35–37]. In Chapters 2 and 4, we found that there was an excess of intermediate frequency *TRIM22* rs1063303:G>C alleles in the AFN, AMR, and EUR cohorts from the 1000 Genomes project, but not in the ASN cohort or two different Inuit populations (Tables 2.4, 4.3). An excess of intermediate frequency alleles is typically indicative of balancing selection. Interestingly, the frequency of the *TRIM22* nsSNP rs1063303:G>C varied markedly among different ethnic groups. The Canadian and Greenlandic Inuit populations had the highest rs1063303:C frequencies, whereas the ASN population had the lowest (Fig 2.3, Table 4.2). While the reasons for these differences in frequency are currently unknown, it is possible that they are due to differential prevalence of specific diseases in distinct geographic locations. Indeed, we have shown in this work that the *TRIM22* nsSNP rs1063303:G>C influences both the antiviral activity of TRIM22 and its novel effects on serum TG and HDL levels in the Canadian Inuit. Given the excess of intermediate frequency *TRIM22* rs1063303:G>C alleles in AFR, AMR, and EUR cohorts, and its disease-related functional effects, it is possible that *TRIM22* rs1063303:C confers some selective advantage in heterozygotes and that the nsSNP is being maintained by balancing selection in these populations. In contrast, the lack of balancing selection at the *TRIM22* rs1063303:G>C site in the two Inuit populations (Canadian and Greenlandic Inuit), may be due to lack of exposure to certain pathogens. These populations have lived in isolation for many years with little exposure to modern infectious diseases [38].

In the future, it will be important to establish which pathogens (if any) are targeting the *TRIM22* rs1063303:G>C site in AFN, AMR, and EUR populations and to investigate why these pathogenic pressures are not present in ASN and Inuit populations. This will likely provide insight into the antiviral mechanism of TRIM22 and help elucidate why the *TRIM22* rs1063303:G>C site has been targeted by multiple evolutionary forces. Of interest, a recent genome-wide scan of two ethnic populations (African Americans and European Americans) identified TRIM22 as one of 60 'extreme' genes evolving under balancing selection in humans [39]. While we have identified one site in TRIM22 that is likely targeted by balancing selection, additional sites may also be targeted. As such, it would be interesting to perform a more thorough analysis of the entire *TRIM22* gene (coding and non-coding regions) to search for additional sites that may be undergoing balancing selection.

5.3 TRIM22 contains 14 high-risk deleterious nsSNPs and numerous putative PTM sites

5.3.1 High-risk deleterious nsSNPs

Studies have shown that genes that evolve under positive selection during interspecies evolution also tend to be highly polymorphic in humans [40]. In Chapter 3, we found that the human *TRIM22* gene is highly polymorphic and contains multiple nsSNPs that are likely deleterious to TRIM22 structure and/or function (Table 3.1). Using 6 different *in silico* prediction programs, we analyzed all of the missense mutation-inducing nsSNPs (56) in *TRIM22* and identified 14 high-risk deleterious nsSNPs (L68R, H73R, E135K, I234K, S244L, G346S, K364N, P403T, L432W, R442C, F456I, T460I, C494F) (Table 3.3). These nsSNPs were predicted to be deleterious by ≥ 4 of the prediction programs and were located at highly conserved amino acid sites (Fig 3.1). Conserved sites are often involved in important biological processes and thus, nsSNPs located at these sites are often deleterious [1,41].

Of interest, our ConSurf analysis, which combines evolutionary conservation data with solvent accessibility predictions, identified codon T460 as a critical structural site and

codons L68, K364, and P403 as critical functional sites in the TRIM22 protein (Table 4.3). These sites coincide in location with 4 high-risk deleterious nsSNPs (i.e. L68R, K364N, P403T, and T460I). While it is currently unknown why these sites are critical, one possibility is that they are involved in key protein-protein interactions. Sites K364 and P403, for example, are located in the B30.2 domain just upstream of the v2 and v3 loops, respectively. These loops are essential for the antiviral activity of the rhTRIM5α protein and help form an extensive HIV-1 capsid binding interface that is required for HIV-1 restriction [3,20]. Since TRIM22 also interacts with the HIV-1 Gag and/or capsid protein, its v2 and v3 loops may help form a similar interface in the B30.2 domain for binding to capsid or other host and/or viral proteins [18]. K364N and P403T may disrupt this interface and impair these interactions. Indeed, our molecular modeling results in Chapter 3 showed that K364N and P403T significantly altered the putative structure of the v1-v4 loops (Fig 3.2). Molecular modeling of the 7 additional high-risk deleterious nsSNPs located in the B30.2 domain also altered the putative structure of these loops, particularly within the v2 and v3 loops. Importantly, these 9 nsSNPs may decrease the flexibility of v2 and v3 (and/or v1 and v4 for some nsSNPs) by introducing more rigid secondary structures, such as alpha helices and/or beta strands, into these regions (Fig 3.2). In rhTRIM5α, v1-v4 flexibility is thought to facilitate the restriction of divergent retroviruses and increase resistance to mutations in the HIV-1 capsid protein [3]. nsSNPs that interfere with v1-v4 flexibility in TRIM22 may be equally important to restriction and may impair the antiviral activity and/or breath of TRIM22. Further studies, such as the resolution of TRIM22's tertiary structure, will be critical for addressing these and other possibilities.

In Chapter 3, we analyzed and prioritized all 56 missense mutation-inducing nsSNPs in TRIM22 and identified 14 high-risk deleterious nsSNPs that likely disrupt its structure and/or function. Future in-depth studies on these 14 nsSNPs are required to establish if they alter, for example, TRIM22 stability, localization, and/or antiviral activity. It will also be important to investigate the clinical implications (if any) of these nsSNPs. This will be essential for further prioritizing the 14 nsSNPs and elucidating their functional consequences.

5.3.2 Putative PTM sites

Although TRIM22 is involved in a variety of important biological processes, very few studies have investigated the molecular determinants of its function. In Chapter 3, we performed an extensive *in silico* analysis of TRIM22 and identified numerous putative PTM sites (sites predicted to undergo ubiquitylation, sumoylation, or phosphorylation, and several SIMs) that likely influence its function. PTMs are involved in a number of biological processes, including many immune pathways, and are often essential for the regulation of protein structure and function [42–45].

Previous studies have shown that TRIM22 can mediate both its own polyubiquitylation and monoubiquitylation. These modifications are dependent on its RING-mediated E3 ligase activity; however, the specific sites within TRIM22 that undergo ubiquitylation have not been identified [21,46]. One study demonstrated that a TRIM22 mutant lacking residues 201-498 (comprises approximately one-third of the CC domain and the entire B30.2 domain), but not a TRIM22 mutant lacking residues 1-200 (entire RING domain and BB2 domain plus approximately two-thirds of the CC domain), was susceptible to self-mediated TRIM22 polyubiquitylation [21]. Our results in Chapter 3, which revealed that K93, K160, and K173 were the only TRIM22 lysine residues predicted to undergo ubiquitylation by both UbPred and BDM-PUB, are consistent with this finding (Table 3.5). In addition, K173 was predicted to be an important functional residue by ConSurf (Fig 3.1). Further studies on K93, K160, and K173 are required to establish how they may influence TRIM22 turnover and/or function.

Our work also identified a number of putative sumoylation sites in TRIM22, as well as several highly conserved SIM domains. K153 was the only lysine residue predicted to undergo sumoylation by both SUMOplot and SUMOsp 2.0 (Table 3.6). While TRIM22 has been shown to function as a E3 sumo ligase for Mdm2, it has never been shown to undergo sumoylation itself [47]. Notably, sumoylation is often associated with nuclear or sub-nuclear targeting. For example, TRIM19/PML contains 8 sumoylation sites plus 1 SIM domain that are required for proper formation, maintenance, and function of sub-

nuclear structures called PML nuclear bodies (NB). PML NBs recruit diverse proteins involved in many cellular processes, including transcriptional regulation, DNA damage repair, cell cycle control, apoptosis, and the host antiviral response. Thus, it is perhaps not surprising that these sumoylation and SIM sites are also required for TRIM19/PML stability and protein-protein interactions [48,49]. Of interest, the number and size of PML NBs increases in response to IFN treatment and PML NBs contribute to innate defense against a number of viruses. It follows that following entry into host cells many viruses induce the modification and/or disassembly of PML NBs [50–52].

Although TRIM22 does not co-localize with TRIM19/PML in PML NBs, it does form NBs and partially co-localizes with p80-coilin in Cajal bodies (CB) [53,54]. Interestingly, several recent studies have implicated sumoylation in CB biogenesis and function. For example, survival motor neuron protein (SMN), one of the major components found in CBs, is targeted by sumoylation and contains a SIM-like domain that is integral to CB assembly and SMN's interaction with p80-coilin [55]. In addition, the sumo isopeptidase UPL1 co-localizes with p80-coilin in CBs and its depletion leads to striking p80-coilin mislocalization and defects in cell proliferation [56]. Given these studies, and the critical role of sumoylation in PML NB assembly and function, it would be interesting to test whether K153 and the SIMs in TRIM22 are necessary for its co-localization with p80-coilin in CBs. Indeed, previous studies have shown that the B30.2 domain of TRIM22, which contains 3 of the putative SIMs we identified in Chapter 3, is required for proper TRIM22 localization in NBs [53]. Finally, because K153 is located in the CC domain of TRIM22, which is typically required for self-association, it is possible that SIMs in one TRIM22 monomer interact with sumoylated K153 in another TRIM22 monomer [57,58]. This may help facilitate TRIM22 trimerization and/or its localization in CBs. It will be interesting to investigate these possibilities in future studies.

Interestingly, 3 of the SIMs we identified in TRIM22 are also present in TRIM5α (Fig 3.3). Previous studies have demonstrated that 2 of these 3 SIMs (i.e. ILGV and VIGL) are required TRIM5α-mediated antiviral activity. Mutating these SIMs in huTRIM5α abolished its ability to interact with the sumoylated N-MLV capsid protein. As a result, SIM mutation prevented N-MLV restriction [4]. Similarly, SIM mutation in rhTRIM5α

abrogated its ability to inhibit HIV-1 replication [5]. Although the role of these SIMs in TRIM22 is currently unknown, it is tempting to speculate that they are involved in its antiviral activities. Further studies are necessary to determine whether these 2 SIMs (or any of the other SIMs in TRIM22) are important for its antiviral activity against HIV-1 and/or other viruses.

Our *in silico* analyses in Chapter 3 also identified multiple phosphorylation sites in the TRIM22 protein (Table 3.6). A total of 21 sites were predicted to undergo serine (16), threonine (3), or tyrosine (2) phosphorylation by both GPS 2.1 and NetPhos 2.0. Many sites were also predicted to be important structural or functional residues by ConSurf (Table 3.6, Fig 3.1). TRIM22 phosphorylation has never been demonstrated; however, several other TRIM proteins have been shown to undergo phosphorylation, including TRIM19/PML and TRIM21 [59–63]. A few particularly interesting sites exist in TRIM22, such as S244, S259, S271, and T460. S259 and S271, both of which were identified as key functional sites by ConSurf, are located within the NLS in TRIM22's SP2 domain (Table 3.6, Fig 3.1). Importantly, phosphorylation sites within or adjacent to NLSs are often found in key regulatory proteins (e.g. transcription factors) and typically regulate nuclear import, which directly affects gene expression [64,65]. Sites S244 and T460 both coincide with the location of high-risk deleterious nsSNPs (S244L and T460I) (Tables 3.2, 3.3). In the future, it will be important to determine whether S259 and/or S271 are required for the nuclear localization of TRIM22 and if the nsSNPs S244L and/or T460I disrupt functionally relevant phosphorylation sites.

Collectively, this work comprises the first systematic *in silico* analysis of the *TRIM22* gene and will be a valuable resource for many future studies. Although we identified a number of putative PTM sites in TRIM22, these PTMs must be verified experimentally and it remains unknown how specific PTMs influence TRIM22 function. Future work using a panel of PTM mutants (clones of wild-type TRIM22 mutated at specific PTM sites) could be used to confirm our *in silico* results and to establish how (or if) verified PTMs affect TRIM22 stability, localization, and/or antiviral activity. Moreover, it may also be interesting to investigate if certain PTM sites in TRIM22 are required for the recruitment of specific E3 ligase target proteins. For example, sumo-targeted ubiquitin

ligases (STUb), a novel class of E3 ubiquitin ligases, only recognize sumoylated target proteins. STUbs that have been identified thus far are characterized by the presence of multiple SIMs (facilitate the recruitment of the sumoylated target protein) and a RING domain (mediates ubiquitylation of the sumoylated target protein) [66].

5.4 The *TRIM22* nsSNP rs1063303:G>C influences diverse TRIM22-mediated biological functions

5.4.1 The *TRIM22* nsSNP rs1063303:G>C increases TRIM22 expression levels

Several interesting functional consequences of the *TRIM22* nsSNP rs1063303:C were identified in this work. In Chapter 2, we found that exogenous expression of *TRIM22* rs1063303:C significantly increased TRIM22 mRNA and protein levels in human cells (Fig 2.4a,b). nsSNPs often alter protein function, expression, conformation, or stability and there are multiple examples of this in the literature [67–71]. Fewer studies have reported nsSNP-induced changes in mRNA expression; however, there are a number of notable examples [72–75]. In these studies, the nsSNPs altered the secondary structure of the RNA transcript, which led to changes in mRNA expression. Changes in RNA structure have been shown to alter mRNA stability, splicing and/or the rate of translation [76,77]. While *TRIM22* rs1063303:C may alter the secondary structure of the RNA transcript, further studies are needed to determine whether this is the case. Since TRIM22 has previously been shown to inhibit gene expression from the HIV-1 and HBV promoters, it is also possible that TRIM22 targets the promoter controlling its own expression [78–80]. In this scenario, the resultant TRIM22 rs1063303:C protein may exhibit reduced or enhanced activity, thereby increasing its own mRNA expression. It will be important to address these possibilities in future research.

Given its effect on TRIM22 mRNA expression, it is likely that *TRIM22* rs1063303:C increases TRIM22 protein expression by increasing TRIM22 mRNA levels. However, other possibilities for increased TRIM22 protein levels also exist. For example, since TRIM22 undergoes self-ubiquitylation and proteasomal degradation, it is possible that

TRIM22 rs1063303:C impairs TRIM22's ability to regulate its own turnover [46]. Other possible explanations include changes in TRIM22 protein stability and/or structure. Of interest, one recent study demonstrated that α-helices in the SP2 domain of rhTRIM5α govern cytoplasmic body formation and HIV-1 restriction. Specific residues in the SP2 domain, including residues 240-242 (RLQ), were critical for the formation of these α-helices [81]. Notably, the study showed that rhTRIM5α forms antiparallel dimers that are stabilized by interactions between three α-helices (H1, H2, H3) found in the CC/SP2 domains of each monomer. H1, which contains the CC domain and the N-terminus of SP2, makes contacts with H2 and together they form a hairpin structure that allows the monomer to double back along the dimer. Residues surrounding the hairpin structure, including residues 240-242, interact with each other and with other residues in the N-terminus of H1' in the second monomer [81]. Because TRIM22 and TRIM5α are closely-related, it is tempting to speculate that TRIM22 also forms antiparallel dimers that are linked via the CC and SP2 domains. Since *TRIM22* rs1063303:C induces a 'R' to 'T' amino acid change at site 242 in the TRIM22 protein, it is also possible that this nsSNP disrupts critical intra- and/or intermolecular interactions that are needed for antiparallel dimer formation. While the tertiary structure of TRIM22 has not been resolved, other TRIM proteins (e.g. TRIM25) have been shown to form antiparallel dimers and studies have noted that this structure may be a common feature among all TRIM proteins [82].

5.3.2 The *TRIM22* nsSNP rs1063303:G>C decreases TRIM22-mediated antiviral activity against HIV-1

In addition to its effects on mRNA and protein expression, *TRIM22* rs1063303:C also altered the sub-cellular localization of TRIM22 and abrogated its anti-HIV-1 activity in human cells. Other host restriction factors, such as APOBEC3G and TRIM5α, contain nsSNPs that have been shown to decrease their ability to inhibit HIV-1. For example, several nsSNPs in TRIM5α, including H43Y, R136Q, and G249D are associated with marked changes in HIV-1 acquisition and disease progression *in vivo* [83,84]. In addition, nsSNP H186R in APOBEC3G is strongly associated with reduced $CD4^+$ T-cell counts and accelerated disease progression in African American individuals. Interestingly, this

effect is not present in European Caucasian individuals [85–87]. In Chapter 2, we showed that a TRIM22 clone containing the ancestral allele rs1063303:G (wild-type TRIM22) potently inhibited HIV-1 replication, whereas a TRIM22 clone containing the derived allele rs1063303:C (nsSNP TRIM22) failed to restrict HIV-1 replication (Fig 2.4d). In contrast, a previous report by Ghezzi and colleagues [88] showed that a TRIM22 clone containing the derived allele rs1063303:C inhibited HIV-1 LTR-driven expression of a luciferase reporter gene. This discrepancy is likely due to differences between the two systems where we used full-length HIV-1 and they used a reporter construct controlled by the HIV-1 LTR. Of interest, these differences imply that other HIV-1 proteins (e.g. antagonists) may influence TRIM22's antiviral activity. Ghezzi and colleagues [88] also found that *TRIM22* rs1063303:C alone was not associated with disease progression in HIV-1 infected individuals; however, a TRIM22 haplotype including nsSNP alleles of rs1063303 and rs7935564 was found more often in advanced progressors compared to long-term non-progressors.

It is interesting to note that two previous studies have shown that TRIM22 expression levels influence HIV-1 infection *in vivo*. One study examined the Centre for the AIDS Programme of Research in South Africa (CAPRISA) study cohort and found that the expression of TRIM22 mRNA was positively correlated with $CD4^+$ T-cell counts and negatively correlated with viral load [89]. In contrast, the other study examined the Swiss HIV study cohort and found that TRIM22 mRNA expression was positively correlated with HIV-1 RNA levels at the viral set point [90]. In Chapters 2 and 4, we found striking differences in *TRIM22* rs1063303:G>C allele frequency among different ethnic groups (Fig 2.3b, Table 4.2). Along these lines, it is worth noting that the CAPRISA cohort is comprised entirely of African females, while the Swiss cohort is comprised entirely of Caucasian individuals. It is possible that the high prevalence of *TRIM22* rs1063303:C found in Caucasian individuals is at least partially responsible for the increased HIV-1 RNA levels in the Swiss cohort. More studies that include populations from different geographic regions are required to establish how *TRIM22* rs1063303:C impacts HIV-1 infection and if other nsSNPs are involved.

5.3.3 The *TRIM22* nsSNP rs1063303:G>C is associated with low serum TG and high serum HDL levels in the Canadian Inuit

In Chapter 4, we demonstrated that the *TRIM22* rs1063303:C allele is associated with significantly lower serum TG levels and significantly higher serum HDL levels in the Canadian Inuit. Notably, these associations were not found in the Greenlandic Inuit or European Caucasian populations (Table 4.4). Our results showed that homozygotes for the *TRIM22* rs1063303:C allele (Canadian Inuit population) had lower levels of serum TG and higher levels of serum HDL than heterozygotes, while the *TRIM22* rs1063303:G allele homozygotes had higher serum TG levels and lower serum HDL levels (Table 4.5). It remains unclear how the *TRIM22* nsSNP rs1063303:C alters serum lipoprotein levels; however, one possible explanation is that it alters them indirectly by regulating innate immune signaling pathways. Recent studies have identified many TRIM proteins that modulate key innate immune pathways, including NF-κB signaling, the IFN response, and cytokine production following PRR activation [91–93]. Although TRIM22 has never been explicitly shown to regulate these pathways, there is some evidence that it affects NF-κB signaling. For example, one study showed that exogenous TRIM22 expression inhibited the TRAF6-stimulated NF-κB pathway by degrading TAB2 [94]. However, a serparte study showed that TRIM22 activated NF-κB signaling [70]. Because activation of the NF-κB signaling pathway leads to proinflammatory cytokine production, and since proinflammatory cytokines induce the APR, it is possible that TRIM22 influences TG and HDL levels by modulating NF-κB signaling. Future studies are needed to address these possibilities and others.

5.5 Concluding remarks

In this body of work, we have identified a number of pertinent sites in the TRIM22 protein the likely contribute to its overall biological and/or antiviral functions. Further, we have shown that specific amino acid sites in TRIM22 have been subjected to strong positive and/or balancing selection. One particular site, which coincides in location with the *TRIM22* nsSNP rs1063303:G>C, has evolved under positive selection during the

evolution of mammals and has a disparate evolutionary footprint among distinct ethnic populations. Moreover, TRIM22 nsSNP rs1063303:G>C increased TRIM22 expression levels, altered its subcellular localization, decreased its antiviral activity against HIV, and was associated with significantly lower serum TG levels and significantly higher serum HDL levels in the Canadian Inuit population. Collectively, this research has paved the way for multiple follow-up studies to further characterize highly relevant sites in the TRIM22 protein and help identify novel factors that may regulate TRIM22 expression, subcellular localization, and antiviral activity.

5.5 References

1. Miller, M. P. & Kumar, S. Understanding human disease mutations through the use of interspecific genetic variation. *Hum. Mol. Genet.* **10,** 2319–28 (2001).

2. Ohkura, S., Yap, M. W., Sheldon, T. & Stoye, J. P. All three variable regions of the TRIM5alpha B30.2 domain can contribute to the specificity of retrovirus restriction. *J. Virol.* **80,** 8554–65 (2006).

3. Biris, N. *et al.* Structure of the rhesus monkey TRIM5α PRYSPRY domain, the HIV capsid recognition module. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13278–13283 (2012).

4. Arriagada, G., Muntean, L. N. & Goff, S. P. SUMO-interacting motifs of human TRIM5α are important for antiviral activity. *PLoS Pathog.* **7,** e1002019 (2011).

5. Lukic, Z., Goff, S. P., Campbell, E. M. & Arriagada, G. Role of SUMO-1 and SUMO interacting motifs in rhesus TRIM5α-mediated restriction. *Retrovirology* **10,** 10 (2013).

6. Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12,** 687–95 (2012).

7. Subramanian, S. & Kumar, S. Higher intensity of purifying selection on >90% of the human genes revealed by the intrinsic replacement mutation rates. *Mol. Biol. Evol.* **23,** 2283–7 (2006).

8. Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437,** 1153–7 (2005).

9. Meyerson, N. R. & Sawyer, S. L. Two-stepping through time: mammals and viruses. *Trends Microbiol.* **19,** 286–94 (2011).

10. Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2832–7 (2005).

11. McNatt, M. W. *et al.* Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. *PLoS Pathog.* **5,** e1000300 (2009).

12. Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* **2,** E275 (2004).

13. James, L. C., Keeble, A. H., Khan, Z., Rhodes, D. A. & Trowsdale, J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6200–5 (2007).

14. Kono, K., Bozek, K., Domingues, F. S., Shioda, T. & Nakayama, E. E. Impact of a single amino acid in the variable region 2 of the Old World monkey TRIM5alpha SPRY (B30.2) domain on anti-human immunodeficiency virus type 2 activity. *Virology* **388,** 160–8 (2009).

15. Yap, M. W. & Stoye, J. P. A Single Amino Acid Change in the SPRY Domain of Human Trim5 • Leads to HIV-1 Restriction. **15,** 73–78 (2005).

16. Stremlau, M., Perron, M., Welikala, S. & Sodroski, J. Species-Specific Variation in the B30 . 2 (SPRY) Domain of TRIM5α Determines the Potency of Human Immunodeficiency Virus Restriction. **79,** 3139–3145 (2005).

17. Perron, M. J., Stremlau, M. & Sodroski, J. Two surface-exposed elements of the B30.2/SPRY domain as potency determinants of N-tropic murine leukemia virus restriction by human TRIM5alpha. *J. Virol.* **80,** 5631–6 (2006).

18. Barr, S. D., Smiley, J. R. & Bushman, F. D. The interferon response inhibits HIV particle production by induction of TRIM22. *PLoS Pathog.* **4,** e1000007 (2008).

19. Sebastian, S. & Luban, J. TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. *Retrovirology* **2,** 40 (2005).

20. Pertel, T. *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472,** 361–365 (2011).

21. Eldin, P. *et al.* TRIM22 E3 ubiquitin ligase activity is required to mediate antiviral activity against encephalomyocarditis virus. *J. Gen. Virol.* **90,** 536–45 (2009).

22. Di Pietro, A. *et al.* TRIM22 inhibits influenza A virus infection by targeting the viral nucleoprotein for degradation. *J. Virol.* **87,** 4523–33 (2013).

23. Lin, Y.-S., Hsu, W.-L., Hwang, J.-K. & Li, W.-H. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol. Biol. Evol.* **24,** 1005–11 (2007).

24. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.* **32,** 63–70 (2007).

25. Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.* **284,** 478–85 (2009).

26. Borden, K. L. RING fingers and B-boxes: zinc-binding protein-protein interaction domains. *Biochem. Cell Biol.* **76,** 351–8 (1998).

27. Borden, K. L. *et al.* In vivo and in vitro characterization of the B1 and B2 zinc-binding domains from the acute promyelocytic leukemia protooncoprotein PML. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 1601–6 (1996).

28. Cao, T., Duprez, E., Borden, K. L., Freemont, P. S. & Etkin, L. D. Ret finger protein is a normal component of PML nuclear bodies and interacts directly with PML. *J. Cell Sci.* **111 (Pt 1,** 1319–29 (1998).

29. Diaz-Griffero, F. *et al.* A B-box 2 surface patch important for TRIM5alpha self-association, capsid binding avidity, and retrovirus restriction. *J. Virol.* **83,** 10737–51 (2009).

30. Li, X. *et al.* Unique features of TRIM5α among closely related human TRIM family members. *Virology* **360,** 419–433 (2007).

31. Charlesworth, D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* **2,** e64 (2006).

32. Olson, M. V. Human genetic individuality. *Annu. Rev. Genomics Hum. Genet.* **13,** 1–27 (2012).

33. Ferrer-Admetlla, A. *et al.* Balancing selection is the main force shaping the evolution of innate immunity genes. *J. Immunol.* **181,** 1315–22 (2008).

34. Rees, D. C., Williams, T. N. & Gladwin, M. T. Sickle-cell disease. *Lancet* **376,** 2018–31 (2010).

35. Ferguson, W., Dvora, S., Fikes, R. W., Stone, A. C. & Boissinot, S. Long-term balancing selection at the antiviral gene OAS1 in Central African chimpanzees. *Mol. Biol. Evol.* **29,** 1093–103 (2012).

36. Alex A. Compton, Vanessa M. Hirsch, and M. E. The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. *Cell Host Microbe.* **11,** 91–98 (2012).

37. Newman, R. M. *et al.* Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 19134–9 (2006).

38. Gracey, M. & King, M. Indigenous health part 1: determinants and disease patterns. *Lancet* **374,** 65–75 (2009).

39. Andrés, A. M. *et al.* Targets of balancing selection in the human genome. *Mol. Biol. Evol.* **26,** 2755–64 (2009).

40. Liu, J., Zhang, Y., Lei, X. & Zhang, Z. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. *Genome Biol.* **9,** R69 (2008).

41. Doniger, S. W. *et al.* A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **4,** e1000183 (2008).

42. Dai, C. & Gu, W. p53 post-translational modification: deregulated in tumorigenesis. *Trends Mol. Med.* **16,** 528–36 (2010).

43. Gallego, M. & Virshup, D. M. Post-translational modifications regulate the ticking of the circadian clock. *Nat. Rev. Mol. Cell Biol.* **8,** 139–48 (2007).

44. Gill, G. Post-translational modification by the small ubiquitin-related modifier SUMO has big effects on transcription factor activity. *Curr. Opin. Genet. Dev.* **13,** 108–13 (2003).

45. Shiloh, Y. & Ziv, Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat. Rev. Mol. Cell Biol.* **14,** 197–210 (2013).

46. Duan, Z., Gao, B., Xu, W. & Xiong, S. Identification of TRIM22 as a RING finger E3 ubiquitin ligase. *Biochem. Biophys. Res. Commun.* **374,** 502–6 (2008).

47. Chu, Yaya and Yang, X. SUMO E3 ligase activity of TRIM proteins. *Oncogene* **30,** 1108–1116 (2011).

48. Cheng, X. & Kao, H.-Y. Post-translational modifications of PML: consequences and implications. *Front. Oncol.* **2,** 210 (2012).

49. Kamitani, T. *et al.* Identification of three major sentrinization sites in PML. *J. Biol. Chem.* **273,** 26675–82 (1998).

50. Lavau, C. *et al.* The acute promyelocytic leukaemia-associated PML gene is induced by interferon. *Oncogene* **11,** 871–6 (1995).

51. Regad, T. & Chelbi-Alix, M. K. Role and fate of PML nuclear bodies in response to interferon and viral infections. *Oncogene* **20,** 7274–86 (2001).

52. Everett, R. D. & Chelbi-Alix, M. K. PML and PML nuclear bodies: implications in antiviral defence. *Biochimie* **89,** 819–30

53. Sivaramakrishnan, G., Sun, Y., Rajmohan, R. & Lin, V. C. L. B30.2/SPRY domain in tripartite motif-containing 22 is essential for the formation of distinct nuclear bodies. *FEBS Lett.* **583,** 2093–9 (2009).

54. Sivaramakrishnan, G., Sun, Y., Tan, S. K. & Lin, V. C. L. Dynamic localization of tripartite motif-containing 22 in nuclear and nucleolar bodies. *Exp. Cell Res.* **315,** 1521–32 (2009).

55. Tapia, O. *et al.* The SMN Tudor SIM-like domain is key to SmD1 and coilin interactions and to Cajal body biogenesis. *J. Cell Sci.* **127,** 939–46 (2014).

56. Schulz, S. *et al.* Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. *EMBO Rep.* **13,** 930–8 (2012).

57. Napolitano, L. M. & Meroni, G. TRIM family: Pleiotropy and diversification through homomultimer and heteromultimer formation. *IUBMB Life* **64,** 64–71 (2012).

58. Parry, D. A. D., Fraser, R. D. B. & Squire, J. M. Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* **163,** 258–69 (2008).

59. Gresko, E. *et al.* PML tumor suppressor is regulated by HIPK2-mediated phosphorylation in response to DNA damage. *Oncogene* **28,** 698–708 (2009).

60. Hayakawa, F. & Privalsky, M. L. Phosphorylation of PML by mitogen-activated protein kinases plays a key role in arsenic trioxide-mediated apoptosis. *Cancer Cell* **5,** 389–401 (2004).

61. Stacey, K. B., Breen, E. & Jefferies, C. A. Tyrosine phosphorylation of the E3 ubiquitin ligase TRIM21 positively regulates interaction with IRF3 and hence TRIM21 activity. *PLoS One* **7,** e34041 (2012).

62. Roberts, J. D., Chiche, J.-D., Kolpa, E. M., Bloch, D. B. & Bloch, K. D. cGMP-dependent protein kinase I interacts with TRIM39R, a novel Rpp21 domain-containing TRIM protein. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **293,** L903–12 (2007).

63. Valiyeva, F. *et al.* Characterization of the oncogenic activity of the novel TRIM59 gene in mouse cancer models. *Mol. Cancer Ther.* **10,** 1229–40 (2011).

64. Harreman, M. T. *et al.* Regulation of nuclear import by phosphorylation adjacent to nuclear localization signals. *J. Biol. Chem.* **279,** 20613–21 (2004).

65. Nardozzi, J. D., Lott, K. & Cingolani, G. Phosphorylation meets nuclear import: a review. *Cell Commun. Signal.* **8,** 32 (2010).

66. Sriramachandran, A. M. & Dohmen, R. J. SUMO-targeted ubiquitin ligases. *Biochim. Biophys. Acta* **1843,** 75–85 (2014).

67. Ramensky, V. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30,** 3894–3900 (2002).

68. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–81 (2009).

69. Alshatwi, A. A., Hasan, T. N., Syed, N. A., Shafi, G. & Grace, B. L. Identification of Functional SNPs in BARD1 Gene and In Silico Analysis of Damaging SNPs□: Based on Data Procured from dbSNP Database. *PLoS One* **7,** 1–10 (2012).

70. De Carvalho, M. D. C. & De Mesquita, J. F. Structural modeling and in silico analysis of human superoxide dismutase 2. *PLoS One* **8,** e65558 (2013).

71. Alanazi, M. *et al.* In silico analysis of single nucleotide polymorphism (SNPs) in human β-globin gene. *PLoS One* **6,** e25876 (2011).

72. Bartoszewski, R. A. *et al.* A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J. Biol. Chem.* **285,** 28741–8 (2010).

73. Capon, F. *et al.* A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. *Hum. Mol. Genet.* **13,** 2361–8 (2004).

74. Drysdale, C. M. *et al.* Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10483–8 (2000).

75. Nackley, A. G. *et al.* Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314,** 1930–3 (2006).

76. Hiller, M., Zhang, Z., Backofen, R. & Stamm, S. Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.* **3,** e204 (2007).

77. Wen, J.-D. *et al.* Following translation by single ribosomes one codon at a time. *Nature* **452,** 598–603 (2008).

78. Tissot, C. & Mechti, N. Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus type 1 long terminal repeat expression. *J. Biol. Chem.* **270,** 14891–8 (1995).

79. Kajaste-Rudnitski, A. *et al.* TRIM22 inhibits HIV-1 transcription independently of its E3 ubiquitin ligase activity, Tat, and NF-kappaB-responsive long terminal repeat elements. *J. Virol.* **85,** 5183–96 (2011).

80. Gao, B., Duan, Z., Xu, W. & Xiong, S. Tripartite motif-containing 22 inhibits the activity of hepatitis B virus core promoter, which is dependent on nuclear-located RING domain. *Hepatology* **50,** 424–33 (2009).

81. Sastri, J. *et al.* Restriction of HIV-1 by rhesus TRIM5α is governed by alpha-helices in the Linker2 region. *J. Virol.* (2014). doi:10.1128/JVI.01134-14

82. Sanchez, J. G. *et al.* The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 2494–9 (2014).

83. Sobieszczyk, M. E., Lingappa, J. R. & McElrath, M. J. Host genetic polymorphisms associated with innate immune factors and HIV-1. *Curr. Opin. HIV AIDS* **6,** 427–34 (2011).

84. Sawyer, S. L., Wu, L. I., Akey, J. M., Emerman, M. & Malik, H. S. High-frequency persistence of an impaired allele of the retroviral defense gene TRIM5alpha in humans. *Curr. Biol.* **16,** 95–100 (2006).

85. An, P. *et al.* APOBEC3G Genetic Variants and Their Influence on the Progression to AIDS. *J. Virol.* **78,** 11070–11076 (2004).

86. Kavidha Reddy1, Cheryl Winkler2, Lise Werner3, Koleka Mlisana3, Salim Abdool Karim3, Thumbi Ndung'u1, 3,*, and the C. A. I. S. T. APOBEC3G expression is dysregulated in primary HIV-1 infection and a polymorphic variant influences CD4+ T-cell counts and plasma viral load. *AIDS* **24,** 195–204 (2010).

87. Do, H. *et al.* Exhaustive genotyping of the CEM15 (APOBEC3G) gene and absence of association with AIDS progression in a French cohort. *J. Infect. Dis.* **191,** 159–63 (2005).

88. Ghezzi, S. *et al.* Identification of TRIM22 single nucleotide polymorphisms associated with loss of inhibition of HIV-1 transcription and advanced HIV-1 disease. *AIDS* **27,** 2335–44 (2013).

89. Singh, R. *et al.* Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. *J. Virol.* **85,** 208–16 (2011).

90. Rotger, M. *et al.* Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS Pathog.* **6,** e1000781 (2010).

91. McNab, F. W., Rajsbaum, R., Stoye, J. P. & O'Garra, A. Tripartite-motif proteins and innate immune regulation. *Curr. Opin. Immunol.* **23,** 46–56 (2011).

92. Kawai, T. & Akira, S. Regulation of innate immune signalling pathways by the tripartite motif (TRIM) family proteins. *EMBO Mol. Med.* **3,** 513–27 (2011).

93. Rajsbaum, R., García-Sastre, A. & Versteeg, G. A. TRIMmunity: The Roles of the TRIM E3-Ubiquitin Ligase Family in Innate Antiviral Immunity. *J. Mol. Biol.* **426,** 1265–1284 (2014).

94. Qiu, H., Huang, F., Xiao, H., Sun, B. & Yang, R. TRIM22 inhibits the TRAF6-stimulated NF-κB pathway by targeting TAB2 for degradation. *Virol. Sin.* **28,** 209–15 (2013).

Appendix A

A.1 Overview of serum lipoproteins

A.1.1 Introduction to serum lipoproteins

Serum lipoproteins are complex aggregates of lipids and proteins that are formed in the liver and intestine, secreted into the circulation, and then taken up by peripheral tissues for energy and storage [1]. The formation of lipoproteins is necessary because it enables lipid transport in the aqueous plasma environment. There are 4 major groups of serum lipoproteins: 1) chylomicrons (CM), 2) very-low-density lipoproteins (VLDL), 3) low-density lipoproteins (LDL), and 4) high-density lipoproteins (HDL). These groups each contain different ratios of specific lipid constituents, namely triacylglycerols (TG), free cholesterol, cholesterol esters, and phospholipids (Table A.1). The groups also contain several lipid-binding proteins called apolipoproteins that surround the lipids and create soluble lipoprotein particles (Fig A.1). All lipoproteins are involved in lipid transport; however, CM and VLDL function primarily to transport TG, whereas LDL and HDL function primarily to transport cholesterol [2].

A.1.2 Triacylglycerol transport

Serum TG levels are often measured clinically, in which case they are defined as all of the TGs present in one deciliter of plasma (mg/dL). This definition includes both CMs and VLDLs; however, because CMs are very short-lived (half-life of ~1 hour) most of the TGs transported in plasma are associated with VLDLs. CMs are synthesized in the small intestine after meals and function primarily to transport dietary (exogenous) TGs from the intestine to other tissues in the periphery [3]. VLDLs, which are synthesized by the liver, transport endogenous TGs to peripheral tissues following their excretion into the bloodstream. VLDLs are carried into tissue capillaries via the bloodstream, where they encounter an enzyme called lipoprotein lipase (LpL). LpL hydrolyzes some of the

Table A.1

Table A.1: Physical properties and lipid composition of serum lipoproteins

	CM	VLDL	LDL	HDL
Density (g/ml)	< 0.94	0.94-1.006	1.006-1.063	1.063-1.210
Total lipids (wt. %) *	99	91	80	44
Triacylglycerols	85	55	10	6
Cholesterol esters	3	18	50	40
Cholesterol	2	7	11	7
Phospholipids	8	20	29	46

* The remaining material is mainly composed of apolipoproteins

Figure A.1: General structure of serum lipoprotein.

Schematic showing the general structure of serum lipoproteins (CM, VLDL, LDL, and HDL). Lipoproteins are spherical in shape with a hydrophobic core comprised of both triacylglycerols (TG) and cholesterol esters. This core is surrounded by a monolayer of phospholipids, free cholesterol, and apolipoproteins.

Figure A.1

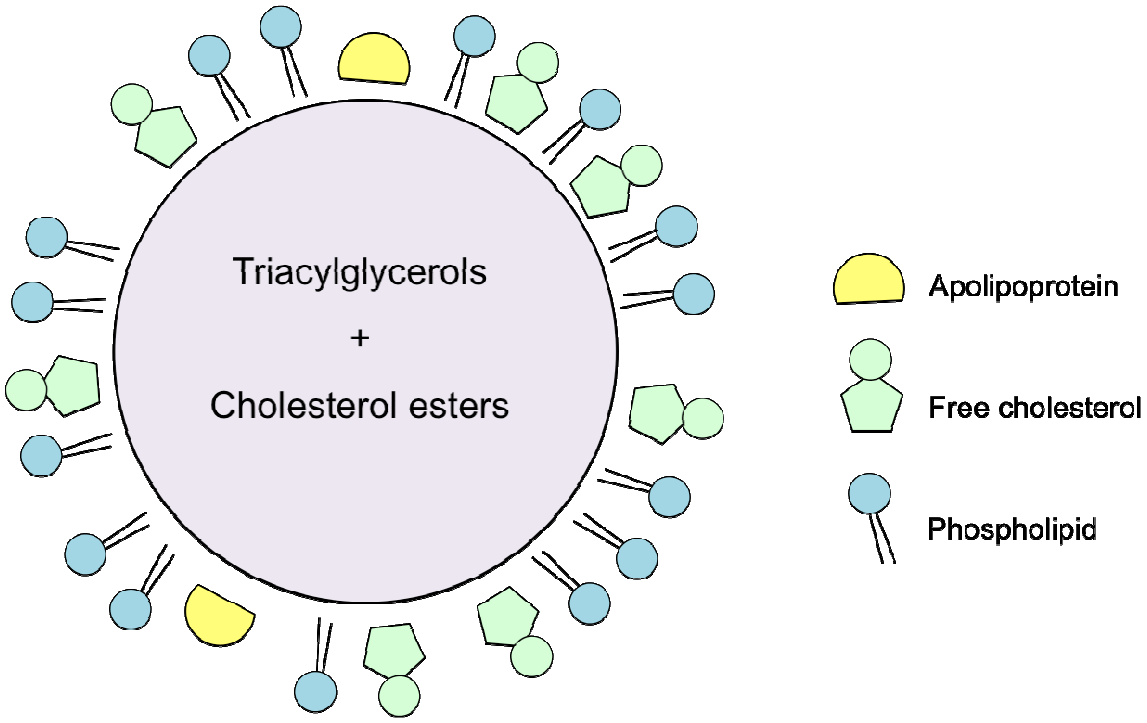


Figure A.1 General structure of serum lipoprotein.

TGs from VLDL into their core components (TGs consist of 1 glycerol molecule and 3 fatty acid molecules) so that they can be taken up by adjacent cells. For example, they may be taken up by muscle cells for energy or fat cells for storage. VLDLs continue to lose TGs as they travel through the blood and decrease in size, eventually turning into intermediate density lipoproteins (IDL). Some IDLs are subjected to further lipolysis to become LDL particles; however, most are reabsorbed by the liver (Fig A.2) [4].

A.1.3 Cholesterol transport

As mentioned above, LDLs are formed from VLDLs as they lose TGs to surrounding tissues. As such, LDLs contain very few TGs compared to VLDLs, but have the same protein shell and amount of cholesterol (Table A.1). Cells internalize LDL particles via receptor-mediated endocytosis and hydrolyze LDL-bound cholesterol esters to release free cholesterol. Cholesterol can be used for a variety of cellular functions, such as in cellular membranes or for the synthesis of steroid hormones. Although LDL is integral for cholesterol transport and cholesterol is necessary for proper cellular functions, too much LDL can be harmful. Multiple studies have shown that higher levels of LDL are associated with an increased risk of cardiovascular disease [5,6].

In contrast to LDLs, nascent HDL particles are formed in the liver and small intestine. HDLs interact with multiple apolipoproteins (Apo) to facilitate cholesterol efflux from cells; however, ApoA1 is particularly important in the initial stages of HDL assembly. ApoA1 forms the scaffold for HDL assembly and is also required for activation of the lecithin:cholesterol acyltransferase (LCAT) enzyme. LCAT, which is essential for the formation of cholesterol esters, allows HDL to extract free cholesterol from LDLs and IDLs. Cholesterol is also removed by HDL from cell surface membranes via a specific transporter molecule (ABCG1) and delivered back to the liver for excretion (Fig A.2). This process, whereby HDL extracts excess cellular cholesterol and transports it to the liver, is called reverse cholesterol transport [7]. A large number of epidemiological and clinical studies have demonstrated that low HDL levels are a major risk factor for the development of cardiovascular disease [8].

Figure A.2: Serum lipoprotein synthesis and transport.

CMs are synthesized in the small intestine and enter the bloodstream to deliver dietary (exogenous) TGs to various peripheral tissues. VLDLs are synthesized in the liver and deliver endogenous TGs to the periphery. LpL hydrolyzes TGs from both VLDLs and CMs so they can be taken up by cells. IDLs are formed from VLDLs as they lose TGs. IDLs continue to lose TGs to peripheral tissues and are eventually either reabsorbed by the liver or turned into LDL particles. LDLs transport cholesterol to cells in peripheral tissues. In contrast, HDL, which is synthesized in the liver, removes excess cholesterol from cells in the periphery. This excess cholesterol is transported back to the liver to be excreted in a process called reverse cholesterol transport. The path for CM, VLDL, and IDL transport is outlined in red, LDL in blue, and HDL in green.

Figure A.2

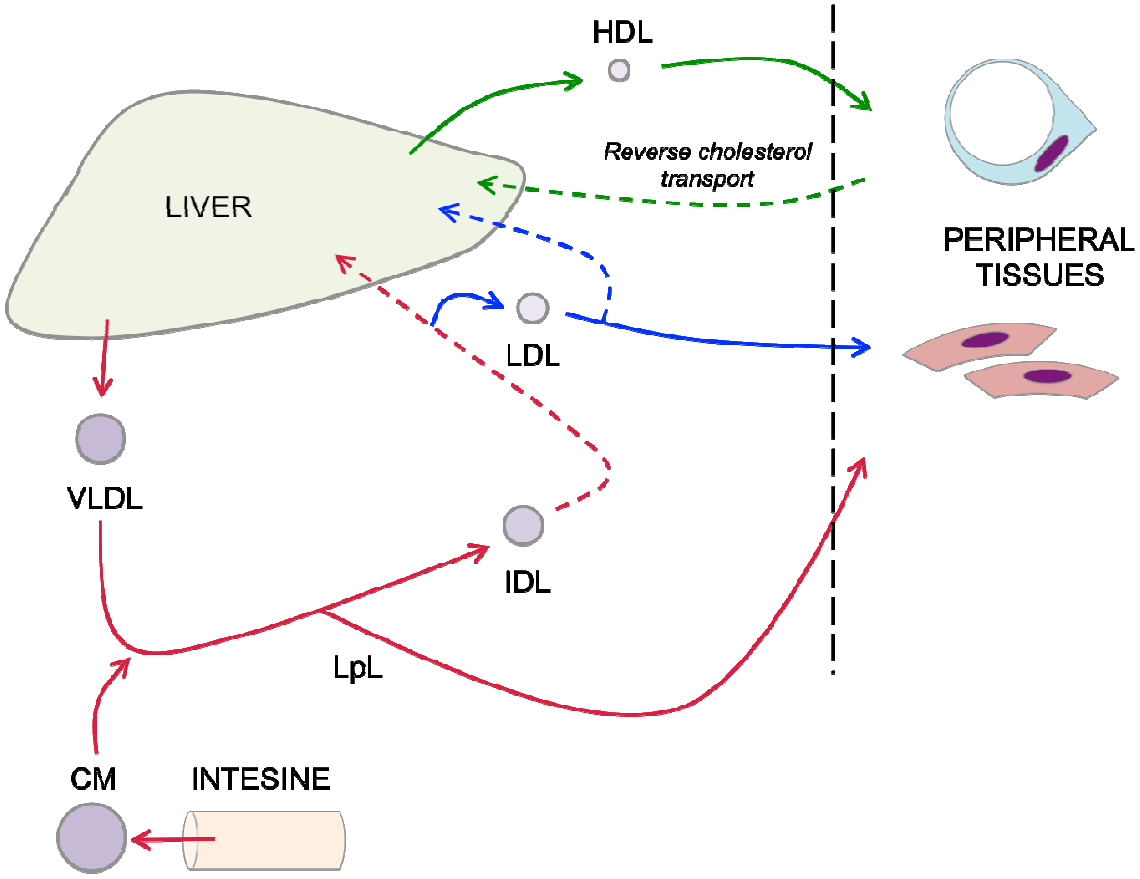


Figure A.2 Serum lipoprotein synthesis and transport.

A.2 Serum lipoproteins during the APR

A.2.1 Overview of the APR

The acute phase response (APR) is an early, complex host reaction that is generated by infection and inflammation. The APR in animals leads to significant changes in serum lipoprotein metabolism. For example, in primates, serum TG levels increase and serum HDL levels decrease during the APR [9]. The increase in serum TGs is largely due to a concomitant increase in serum VLDL levels. Multiple proinflammatory cytokines (e.g. TNF, IL-1, IL-2, IL-6, and IFN) induce this increase in serum VLDL levels via one or more of the following metabolic changes: 1) increase in hepatic fatty acid synthesis, 2) decrease in hepatic fatty acid oxidation, 3) induction of adipose tissue lipolysis, and 4) decrease in hepatic VLDL clearance. Changes 1-3 increase the overall production and secretion of VLDL into the circulation by increasing the amount of fatty acid substrate available for TG re-esterification. Change 4, which often only occurs at higher doses of LPS, decreases TG catabolism by reducing the activity of the LpL enzyme.

A.3 References

1. Scanu, A.M. & Wisdom, C. Serum lipoproteins structure and function. *Annu. Rev. Biochem.* **41,** 703–30 (1972).
2. Eisenberg, S. & Levy, R. Lipoprotein metabolism. *Adv. Lipid Res.* **13,** 1–89 (1975).
3. Kindel, T., Lee, D.M. & Tso, P. The mechanism of the formation and secretion of chylomicrons. *Atheroscler. Suppl.* **11,** 11–6 (2010).
4. Gibbons GF, Wiggins D, Brown AM, H.A. Synthesis and function of hepatic very-low-density lipoprotein. *Biochem. Soc. Trans.* **32,** 59–64 (2004).
5. Castelli, W.P. *et al.* Incidence of coronary heart disease and lipoprotein cholesterol levels. The Framingham Study. *JAMA* **256,** 2835–8 (1986).
6. Glass, C.K. & Witztum, J.L. Atherosclerosis. the road ahead. *Cell* **104,** 503–16 (2001).
7. Lund-Katz, S. & Phillips, M.C. High density lipoprotein structure-function and role in reverse cholesterol transport. *Subcell. Biochem.* **51,** 183–227 (2010).
8. Assmann, G. & Gotto, A.M. HDL cholesterol and protective factors in atherosclerosis. *Circulation* **109,** III8–14 (2004).
9. Khovidhunkit, W. *et al.* Effects of infection and inflammation on lipid and lipoprotein metabolism: mechanisms and consequences to the host. *J. Lipid Res.* **45,** 1169–96 (2004).

Curriculum Vitae

Name Jenna Nicole Kelly

Education

2007	BSc	Western University, London, Ontario, Canada Honors Medical Sciences
2008-2014 (Expected)	PhD	Western University, London, Ontario, Canada Department of Microbiology and Immunology Supervisor: Dr. Stephen Barr

Honors and Awards

2008-2009	Roche Diagnostics Award
2009-2010	Queen Elizabeth II Graduate Scholarship in Science and Technology
2010-2011	Queen Elizabeth II Graduate Scholarship in Science and Technology
2010-2011	Graduate Thesis Research Award
2011-2012	Ontario Graduate Scholarship
2011-2012	Graduate Entrance Award

Research Publications

1. **Jenna N. Kelly** and Stephen D. Barr. (2014). *In silico* analysis of functional single nucleotide polymorphisms in the human *TRIM22* gene. PLoS ONE.
2. **Jenna N. Kelly**, Matthew W. Woods, Sintia Xhiku, and Stephen D. Barr. (2013). Adaptive evolution in the antiviral *TRIM22* gene: identification of a single nucleotide polymorphism that profoundly impacts TRIM22 function. Human Mutation.
3. Clayton J. Hattlmann, **Jenna N. Kelly**, and Stephen D. Barr. (2012). TRIM22: a diverse and dynamic antiviral protein. Invited Review article from Drs. Abraham Brass (MIT) and Abdul Waheed (NCI Frederick) for a special issue on Host-Pathogen Interactions of Retroviruses. Molecular Biology International Volume 2012:153415.
4. Matthew S. Miller, Peter Pelka, Greg Fonseca, Michael Cohen, **Jenna N. Kelly**, Stephen D. Barr, Roger Grand, Andrew Turnell, Peter Whyte, Joe S. Mymryk. (2012). Characterization of the 55 residue E1A protein encoded by the 9S E1A mRNA of species C adenovirus. Journal of Virology 86(8):4222-4233.
5. Matthew W. Woods, **Jenna N. Kelly**, Clayton J. Hattlmann, Jessica G.K. Tong, Li S. Xu, Macon D. Coleman, Graeme R. Quest, James R. Smiley, Stephen D.

Barr. (2011). Human HERC5 restricts an early stage of HIV-1 assembly by a mechanism correlating with the ISGylation of Gag. Retrovirology 8:95. Rated *highly accessed* by the journal.

6. **Jenna N. Kelly,** Jessica G.K. Tong, Clayton J. Hattlmann, Matthew W. Woods, and Stephen D. Barr. (2011). ISBN 978-953-307-665-2. Book Title: HIV and AIDS: Updates on Biology, Immunology, Epidemiology and Treatment Strategies. Chapter Title: Cellular Restriction Factors: Can We Exploit the Body's Natural Antiviral proteins to Combat HIV/AIDS? *Downloaded 1098 times since October 26, 2011.*

Related Work Experience

2010-2012	Teaching Assistant Western University, London, Ontario, Canada
2011-2012	Developer of Neglected Tropical Diseases Teaching Modules Consortium of Universities for Global Health (CUGH)
2011-2012	Graduate Chair of Research Stream, Global Health Conference Western University, London, Ontario, Canada
2008	Internship at Child Family Health International Mumbai, India

Committees and Activities

2009-2014	Volunteer at the Regional HIV/AIDS Connection London, Ontario, Canada
2011-2014	Member of the Global Health Speaker Series Western University, London, Ontario, Canada
2010-2012	Member of the Infection and Immunity Research Forum Western University, London, Ontario, Canada