Understanding the mucosal fluid proteome in rectal susceptibility to HIV infection

By

Laura Romas

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba

In partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Medical Microbiology

University of Manitoba

Winnipeg

Table of Contents

Abstract

Objective: The rectal mucosa is highly susceptible to HIV infection. Mucosal fluid contains soluble immune proteins that influence HIV infection, and previous studies have shown unique mucosal protein expression in HIV-exposed seronegative (HESN) populations, which may contribute to reduced HIV susceptibility. However, the key correlates of susceptibility at the rectal mucosa have not been well defined, which is a critical knowledge gap for our understanding of HIV pathogenesis. **Methods:** Rectal lavage from low risk men was screened for HIV-neutralizing activity in a TZM-bl reporter cell line against an R5-tropic HIV virus. Label-free tandem mass spectrometry was used to characterize soluble proteins within rectal lavage samples from a low-risk cohort of men (n=15), and HESN men who have sex with men (MSM; n=25). Protein expression between populations was compared using adjusted t tests ($p<0.05$), and was interpreted using hierarchical clustering and DAVID biofunctional analysis. Protein expression was further analyzed using survey data on sexual behaviours. Proteins associated with the HESN population were screened for antiviral activity in TZM-bl and PBMC culture against an R5- and X4-tropic virus. **Major Results:** Rectal mucosal fluid was able to inhibit HIV infection *in vitro* by 40% ($p<0.05$). Mass spectrometry identified 30/341 (9%) proteins differentially expressed (DE) in HESN MSM. DE proteins held functions in immunity ($p=6.68x10^{-6}$, $p=0.001$) and epithelial barrier development ($p=1.81x10^{-4}$; $p=0.01$); notably, specific antiproteases were elevated in HESN secretions, two of which were screened for antiviral activity. Serpin B4 (+2.52 L2FD; $p=1.09x10^{-5}$), showed significant inhibition of HIV in TZM-bl (45% BaL, 34% IIIB; $p<0.05$) and PBMC culture (37% BaL, 49% IIIB; $p<0.05$); cystatin A (+1.52 L2FD; $p=1.40x10^{-3}$) showed no inhibitory effects. Serpin B4 expression was not associated with frequency of oral intercourse ($p=0.32$), partner viral load ($r=0.16$; $p=0.29$) or presence of HIV neutralizing IgA in secretions ($p=0.52$). **Conclusions:** This thesis reports the use of proteomics to understand HIV-susceptibility at the rectal mucosa, and identified serpin B4 as a novel antiviral immune correlate in a population of HESN MSM. These results may help guide future studies of prevention technologies, such as microbicides or vaccines, which would ultimately help limit the spread of HIV.

Acknowledgments

I would like to first thank my primary supervisor and mentor, Dr. Adam Burgener for the incredible opportunity to be a part of his team and contribute to his research. Adam has been an outstanding mentor, and has provided monumental advice in all aspects of career development from pipetting, to networking, and publishing, but most importantly, in selecting a good wine. It has been my pleasure to learn from such a dedicated, ambitious and talented scientist.

Secondly, I would like to thank my co-supervisor, Dr. Blake Ball, for his thoughtful guidance and encouragement throughout my studies. It has been a privilege to work with such a brilliant and accomplished HIV immunologist.

I would also like to thank my committee members Dr. Kevin Coombs and Dr. Emmanuel Ho for their insightful advice and critiques of my work throughout the years.

My sincerest thanks go to the rest of the HIV proteomics team for their support throughout my project. In particular, thank you to Kenzie Birse, my partner in data analysis and fateful travelling companion, and to Max Abou for his technical assistance and answering my many questions in the lab.

This work would not be possible without the Dr. Garrett Westmacott, Stuart McCorrister and Dr. Patrick Chong from the Proteomics Core of the National Microbiology Laboratory, and I'd like to express my sincere gratitude for their help in this project.

I'd like to thank our collaborators at in Dr. Kristina Broliden's Lab at the Karolinska Institutet, Dr. Peter Anton's group at University of Los Angeles California, and Dr. Robin Shattock's lab at the Imperial College London. I need to especially thank Dr. Klara Hasselrot, Dr. Carolina Herrera and Dr. Jennifer Fulchur for all their help in this study.

Thank you as well to Syeda Rahman, Christine Mesa, and Margot Plews for their technical assistance, and of course to Sue Ramdahin for all her help in getting me set up in the lab and for sharing her expertise in cell culture and virus infections.

I would also like to thank the many members (past and present) of the National Laboratory for HIV Immunology for their friendship and many scientific discussions and demonstrations that have

helped me along in one way or another. In particular, thank you senior students Lindsay, Melissa, Aida, Derek, and Were for their mentorship through graduate studies.

I am deeply indebted to Angela Nelson, Jude Zieske, Sharon Tardi and Cheryl Reimer for helping me navigate the administrative components a University of Manitoba graduate degree.

I'd like to thank the men of the Venhälsan and Multicenter AIDS Cohort Study (MACS) cohorts for their dedication and providing the clinical samples to help further this research.

Lastly, I would like to thank my friends and family for their support in accomplishing this degree. To the lovely David Noël who has been my rock throughout the highs and lows of this project; your nomination for sainthood is in the mail. Also to my parents, Randy and Peggy Romas, for their love and support, and finally to my grandparents Peter and Adeline Romas, for not only love and support, but reading my publications despite probably never really wanting to know that much about rectal mucus.

Dedication

This thesis is dedicated to those who have been marginalized by the HIV pandemic. The story of discrimination against homosexual men in North America during the 1980s due to a lack of scientific understanding is a story that continues in many parts of the world today, and it is what inspired me to undertake this research. To those who have felt like their health was ignored because of who they are, this work is hopefully one of many small steps towards a better tomorrow.

List of abbreviations

Abbreviation	Description	Abbreviation	Description
AIDS	Acquired Immune Deficiency Syndrome	L2FD	Log_2 Fold Difference
AIII	Antithrombin III	LC	Langerhan Cells
APOBEC3G	Apolipoprotein B mRNA-editing enzyme 3G	LC	Liquid Chromatography
ART	Antiretroviral Therapy	LTR	Long Terminal Repeat
ARV	Antiretroviral	LXR/RXR	Liver X Receptor/Retinoid X Receptor
AZT	Azidothymidine	MHC	Major Histocompatability Complex
BCA	Bicinochronic Acid	MIP	Macrophage Inflammatory Protein
BST-2	Bone marrow Stromal Cell Antigen 2	MMP	Matrix Metalloprotease
cART	Combination Antiretroviral Therapy	MPO	Myeloperoxidase
CCR5	C-C Chemokine Receptor Type 5	MS	Mass Spectrometry
CD4	Cluster of Differentiation 4	MSM	Men who have Sex with Men
CDC	Centers for Disease Control and Prevention	N9	Nonoxyl-9
CLR	C Type Lectin Receptor	NK	Natural Killer
CPM	Cyclic Peptide	PAMP	Pathogen associated molecular pattern
CRF	Circulating Recombinant Forms	PBMC	Peripheral Blood Mononuclear Cells
CROI	Conference on Retroviruses and Opportunistic Infections	PBS	Phosphate Buffer Saline
CSW	Commercial Sex Worker	PCD	Programmed Cell Death
CTL	Cytotoxic T Lymphocyte	PCP	*Pneumocystis carnii* pneumonia
CVL	Cervicovaginal Lavage	PR	HIV Protease Enzyme
CXCR4	C-X-C Chemokine Receptor Type 4	Pol	Polymerase
DC	Dendritic Cell	PRR	Pattern Recognition Receptor
DMBT1	Deleted in Malignant Brain Tumors type 1	RANTES	Regulated on Activation, Normal T Cell Expressed and Secreted
ECM	Extracellular Matrix	RLU	Relative Light Units
ELISA	Enzyme-Linked Immunosorbent Assay	RLV	Rectal Lavage
Env	Envelope	RNA	Ribonucleic Acid
FBS	Fetal Bovine Serum	RSE	Rectal Sponge Eluate
FDA	Food and Drug Administration	RT	Reverse Transcriptase Enzyme
FTC	Emtricitabine	SAMHD1	SAM domain and HD domain-containing protein 1
FGT	Female Genital Tract	Serpin	Serine Protease Inhibitor
GALT	Gut Associated Lymphoid Tissue	SHIV	Simian Human Immunodeficiency Virus
GP	Glycoprotein	SIV	Simian Immunodeficiency Virus
GI	Gastrointestinal System	SLPI	Secretory Leukocyte Protease Inhibitor
HAART	Highly Active Antiretroviral Therapy	SLV	Saliva
HESN	HIV-Exposed Seronegative	DNA	Deoxyribonucleic Acid
HHV-8	Human Herpesvirus 8	TAP	Treatment as Prevention
IDU	Intravenous Drug Users	TDF	Tenofovir Disoproxil Furmarate
IN	HIV Integrase Enzyme	URAI	Unprotected Receptive Anal Intercourse
IVR	Intravaginal Ring	URF	Unusual Recombinant Form
KS	Kaposi's Sarcoma	VIRIP	Virus Inhibitory Peptide

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Origins of the HIV pandemic

The current HIV pandemic originated from several zoonotic infections from monkeys carrying the Simian Immunodeficiency Virus (SIV). SIVsm, most similar to HIV-2, infects sooty mangabeys in West Africa, while SIVcpz, related to the more prominent HIV-1 strain, infects chimpanzees. Three different primate-to-humans transmission events occurred within the late 19th to early 20th century that derived three HIV-1 groups: major group (M), outlier (O) and non-major/non-outlier (N)[1-3]. The oldest HIV isolate on record originated from a plasma sample taken in 1959 from the Niger-Congo region of South-Central Africa, and resembles M group strains[4]. It is commonly thought that humans first became infected with SIVcpz through contact with infected chimpanzee tissue or fluids during the hunting process. Supporting evidence has shown that the transmitted SIV mutated into a virus that could efficiently infect human tissue, and was then referred to as HIV[5,6]. Furthermore, SIV can undergo crossover events to produce progeny virions similar to the HIV-1 strains that currently infect humans, providing further evidence that HIV could have evolved from SIV[7]. Although the exact origin of HIV/AIDS is unclear, as the earliest known isolates are from African samples, that region generally accepted as the virus' geographical origin.

HIV is thought to have spread to the United States in the 1970s. A subtype B virus spread from Africa to Haiti through a single infected individual in 1960, which expanded HIV incidence throughout the Caribbean and then disseminated to North America[8]. In 1981, the Centers for Disease Control (CDC) reported several cases of Kaposi's Sarcoma (KS) cancer and *Pneumocystis carinii pneumonia* (PCP) lung infections among young, previously healthy, gay men in Los Angeles; the epidemic rapidly spread, with 270 reported cases of immune deficiency in gay men by the end of 1981[9]. Cases were later reported in an infant who received a blood transfusion and in females with male sexual partners, expanding the outbreak to other populations[10,11]. In 1982 "AIDS," or acquired immune deficiency syndrome, was used by the CDC to describe the affliction for the first time[12]. Research by Robert Gallo at the National Cancer

Institute (US) and Luc Montagner of the Pasteur Institute in France identified the etiological agent of AIDS to be a retrovirus, which they named Human T-Lymphotropic Virus III (HTLV-III) and Lymphadenopahty Associated Virus (LAV), respectively; the virus was officially named Human Immunodeficiency Virus (HIV) in 1986[13-15]. At the peak of the US epidemic in the mid 1990s, HIV-related deaths rose to 300,000 and the number of those infected was nearly 1.2 million. In 1995, HIV protease inhibitors and use of combination antiretroviral therapy (cART) or highly active antiretroviral therapy (HAART) facilitated a drastic decline in new HIV infections and the number of AIDS related deaths[16].

1.2 HIV Virus

HIV infection is responsible for the development of AIDS; a syndrome characterized by a gradual destruction of the immune system. HIV is a *Lentivirus* of the family *Retroviridae*, and can be divided into three groups: major (M) or outlier (O) and non-major/non-outlier (N) strains[1,2]. The M group is comprised of the major circulating viruses, and is divided into 9 subtypes with A and F having sub-subtypes (A1, A2, A3, A4, B, C, D, F1, F2, G, H, J and K) based on genetic diversity[17]. Further, novel strains formed through hybridization of two concurrent HIV infections within the same individual are preliminarily defined as unique or unusual circulating recombinant forms (URCs), and complete sequencing of URFs from three epidemiologically unlinked individuals designates novel strains as circulating recombinant forms (CRFs)[18,19]. Clade C viruses circulate largely in India, Eastern and Southern Africa, and are the most prevalent, accounting for roughly half (49.9%) of all global infections. Clade A (East and Central Africa, Central Asia and Eastern Europe) and clade B (North and South America, Western Europe, East Asia and Indonesia/Australia) are other highly prevalent strains that account for 12.3% and 10.2% of infections, respectively[5,17,20]. Other subtypes and CRFs make up a smaller proportion of global infections (<10% each) and are largely circulating within East and West Africa, as well as Southeast Asia. Clade specificity poses a challenge for HIV prevention and treatment strategies, as genetic diversity of the

strains may increase likelihood of drug resistant mutations, or differences may alter antigens targeted in a vaccine response.

1.3 HIV life cycle and replication

HIV primarily infects CD4+ T cells to establish an infection, and replicates through seven stages: cell membrane fusion, capsid release, reverse transcription, integration into host DNA, replication and protein production, virus assembly and budding. The virus capsule is surrounded by a membrane with host and viral proteins embedded. To initiate viral-host membrane fusion, the viral envelope proteins gp41 and gp120 interact with the T cell receptor CD4, and utilize the chemokine receptors CCR5 or CXCR4 as a co-receptor in order to initiate fusion of the viral and host membranes. Viruses capable of using the CCR5 receptor are referred to as R5 viruses, while strains that use the CXCR4 receptor for entry are called R4 viruses[21]; however, viral tropism varies by HIV clade. The majority of circulating strains (B and C) establish infection using the CCR5 receptor, with R4 tropic viruses emerging later in infection, while clade A uses mostly CCR5 infection at all stages, and clade D uses CXCR4 throughout[5]. Upon membrane fusion, the viral matrix and capsid are released into the cytoplasm. The viral matrix is composed of p17 protein, and surrounds the viral capsid, containing p24 protein. Both the matrix and capsid are digested, releasing two copies of a positive sense, single-stranded RNA genome, and viral proteins integrase (IN), protease (PR) and reverse transcriptase (RT) inside the cell. The viral RNA genome is converted into ssDNA, and subsequently, dsDNA during two rounds of reverse transcription. The HIV RT enzyme has poor proofreading ability during DNA synthesis, and has a mutation rate of 3.4×10^{-5} mutations/base pair or 1 mutation and 7-30 crossover events per replication cycle[5]. The IN enzyme transports the dsDNA to the nucleus and fuses the viral genome with host chromosomes. This allows the viral genome to be transcribed using host cell machinery. Transcribed viral RNA can then be transported out of the nucleus, where it is either directly transported to the viral membrane for assembly, or translated into viral proteins.

The viral genome contains 9 genes, which are transcribed into 19 proteins. Viral genes from the 5´- to 3´ ends of a ssRNA include *gag* (group-specific antigen), *pol* (polymerase), and *env* (envelope glycoproteins), which are flanked at each end by long terminal repeat (LTR) regions that act as transcriptional regulators. Transcribed *gag* produces a Pr55Gag precursor polyprotein, which is cleaved by the viral PR enzyme into the structural proteins p17 (matrix protein), p24 (capsid protein), p7 (nucleocapsid) and p6 (viral assembly protein)[22]. The *pol* gene segment is transcribed into the Pr160GagPol precursor polypeptide, which is cleaved by PR into viral replication enzymes, PR, RT and IN. The *env* portion of the genome is transcribed into the gp160 precursor, which is processed by host proteases into the transmembrane glycoproteins, gp120 and gp41, that facilitate host receptor binding and viral fusion[23]. HIV also produces several accessory proteins that help facilitate viral transcription. Tat and Rev proteins are transcribed from a frame-shift of the env gene; the Tat protein assists in transcription of the HIV genome from the LTR, while Rev transports viral RNA out of the nucleus for translation and assembly. Additional auxiliary genes *vif*, *vpr*, *vpx*, *vpu* and *nef* produce proteins with no currently known enzymatic activity. Accessory proteins can act as ligands for host restriction factors such as apolipoprotein B mRNA-editing enzyme 3G (APOBEC3G), bone marrow stromal cell antigen 2 (BST-2) and SAM domain and HD domain-containing protein 1 (SAMHD1) that normally counter HIV replication, and facilitate sequestering of such host antiviral factors to counter the host immune response and promote HIV replication[24]. Pressure from the host immune system, combined with an error-prone reverse transcriptase enzyme creates a significant level of genetic diversity, yet limits the amount of viable progeny virus that is produced during infection. Approximately only 1 in 1000 viruses that bud from a cell is able to establish new infection[25]. Despite its inefficiency, HIV is able to overcome this limitation through massive amount of viral production; during acute infection, an individual will produce as many as 10^9 total virions each day, which allows for a high error rate while still producing sufficient virus to establish infection[5,26]. HIV will continue to expand into a local founder population within the submucosa until disseminating to the lymphatic system to replicate in other sites of the body and establish disease.

1.4 HIV pathogenesis and disease

The major pathology of HIV is the killing of CD4 lymphocytes, leading to systemic immune suppression. Following local expansion, HIV travels to the gut associated lymphoid tissue (GALT), which is the major site of virus replication during acute infection. Replication leads to high levels of apoptosis and massive depletion of CD4 lymphocytes within the lamina propria within the first 3-6 weeks of infection[27]. Induction of systemic cytotoxic T lymphocyte (CTL) responses in the host helps to control viral replication into a balanced state of low-grade replication and immune cell recovery (viral set point)[28]; however, the CTL response is not enough to completely clear the infection and all established reservoirs, which allows HIV to establish a chronic infection that leaves the patient asymptomatic for a period of years. In the absence of ART, blood CD4+ T cells will slowly decline, and host immunity will become exhausted[29]. Clinical AIDS patients are characterized as having <200 CD4+ T cells per cubic millimeter of blood, which is a significant reduction relative to the 1000 CD4+ T cells in healthy individuals[30]. This depletion of immune cells leaves the patient open to opportunistic infections and non-AIDS defining conditions (cardiovascular disease, kidney and liver dysfunction and neurodegeneration) that lead to morbidity and mortality.

Co-infections are largely responsible for AIDS-related deaths as immunosuppressed individuals are more susceptible to other communicable diseases. Tuberculosis (TB) can result from infection of the lung mucosal tissue with *Mycobacterium tuberculosis*; it is the most common presenting illness among HIV positive individuals, and is the leading cause of mortality in AIDS patients[31]. Co-infections with hepatitis B and C are also common as these viruses share similar routes of transmission as HIV, and can result in liver dysfunction and disease[32,33]. AIDS patients have an increased susceptibility to oncogenic virus infections, such as human herpesvirus 8 (HHV-8), which can cause cervical cancers, or Kaposi's sarcoma, which largely affects the skin. Long-term viral infection within the lymphatic system can lead to transformation of lymphatic cells, causing lymphoma or Hodgkin's lymphoma; co-infection with Epstein-Barr virus (EBV) can also result in non-Hodgkin's lymphoma[34]. Although many of these morbidities can

be managed through effective treatment while on ART, lifelong and extensive treatment is usually needed.

Despite the success of antiretroviral drugs against the HIV epidemic, there are still 35 million people infected with HIV, and 2.3 million of new infections each year[35]. Men who have sex with men (MSM) remain one of the highest risk groups for HIV acquisition worldwide, and are a major source of new infections each year. Developing new prevention technologies for this group is a critical research priority in order to prevent new infections. HIV positive patients depend on lifelong treatment, which is a large financial responsibility, and is a major public healthcare burden in both treatment and subsidization of drug costs; furthermore, access to ARVs remains a critical issue in resource limited settings. Currently, there is no cure for HIV, though recent studies have provided hope for a functional cure with early treatment intervention[36]. In the absence of a cure, HIV prevention remains the most effective means of stemming the global pandemic.

1.5 Current progress in HIV prevention strategies

1.5.1 HIV Vaccine Development

Since the beginning of the epidemic, there have been more than 250 vaccine clinical trials conducted[37]. There is currently no completely effective vaccine against HIV. Extensive genetic variability associated with HIV replication due to the high mutation rate of the reverse transcriptase enzyme has resulted in antigenic diversity in HIV Env proteins that is resistant to the traditional vaccine induction of neutralizing antibodies[37]. Early vaccine trials by VaxGen failed in their attempts to produce broadly neutralizing antibodies (bNAbs) against the Env glycoprotein gp120[38,39]. Due to a lack of success in the field of antibodies, vaccine development focused on cellular immunity and the induction of a strong cytotoxic response against HIV. Vaccines targeted at inducing protective cellular immunity through stimulation with inactivated SIV or a replication-incompetent adenovirus vector vaccine failed to protect the animals from infection, but did show a lower viral set point and prolonged survival[40]. Mild success in the area of HIV specific T cell immunity prompted the Step Study trial, which vaccinated MSM and high-

risk women with a recombinant adenovirus serotype 5 (rAD5) vector containing the HIV proteins Gag, Pol and Nef. Despite promising efficacy results in preclinical studies, and clinical safety in several phase 1 and 2 trials, the STEP trial was discontinued due to safety concerns and a risk of enhanced HIV acquisition in some treated individuals[41,42]. A second trial of the rAd5 vaccine in heterosexual individuals in South Africa (HVTN 503/Phambili study) was also terminated due to lack of efficacy[41,42]. Subsequent gp120 vaccines tested in the RV144 Thai trial, utilized a canarypox vector, and revitalized the field of antibody vaccines. RV144 showed 31.2% efficacy, and cited non-neutralizing antibodies against the V1V2 loop of gp120 and antibody-dependent cell-mediated cytotoxicity (ADCC) as possible correlates of protection, with Env-specific IgA contributing to HIV infectivity[43,44]. Though recent trials have showed some promise for the field, the overall lack of success in developing an effective HIV vaccine highlights our lack of understanding of immune protection against the virus. Only recently have vaccine trials begun to collect mucosal samples, and it is possible that retrospective studies of these samples may identify mucosal immune factors that influence vaccine efficacy[45]; thus, increasing our understanding of mucosal correlates of susceptibility is important for vaccine success. However, in the absence of a vaccine, alternative HIV prevention strategies are needed.

1.5.2 HAART: Treatment as prevention

Highly active antiretroviral therapy (HAART) has traditionally been given as a post-exposure therapy for the control of disease progression; however, new insights have demonstrated the benefits of HAART in control of HIV transmission. Antiretroviral treatment of HIV positive patients improves their clinical outcomes while greatly reducing viral load in both the blood and mucosal secretions, which reduces their infectiousness[46,47]; therefore, increasing the proportion of HIV positive individuals on ART will be a major factor in pandemic containment. However, barriers such as unknown HIV status and access to the proper treatment have limited the success of treatment as prevention (TAP). Coverage of ART is widespread in high income countries, and as of 2012, 9.7 million people in low and middle income countries were on some form of antiretroviral therapy, though many people remain undiagnosed

and untreated[48]. In 2014, UNAIDS released its goal to reach 90-90-90 by 2030; this mandate would realize 90% of all HIV-positive people aware of their HIV positive status, 90% of people receiving ARV treatment, and 90% of people on ARVs having an undetectable viral load[49]. Treatment as prevention targets HIV positive individuals as the source of new infections, but highlights the logistical difficulty of coordinated, global programs. In order to effectively limit new infections, strategies that target HIV susceptible populations must be used in conjunction with limiting the infection source.

1.5.3 Pre-exposure prophylaxis (PrEP)

Several clinical trials have shown the use of antiviral drugs by HIV negative individuals greatly reduces HIV acquisition. Pre-exposure prophylaxis (PrEP) requires the administration of antiviral therapy to high-risk, HIV uninfected individuals to establish effective concentrations of antiviral drugs within the plasma or at mucosal tissues. PrEP is currently approved by the FDA for use as a once-daily oral pill, which is formulated as combination of the nucleoside reverse transcriptase inhibitor, emtricitabine (FTC), and the nucleotide reverse transcriptase inhibitor tenofovir disoproxil furmarate (TDF), or as TDF alone[50]. The iPrEx trial of daily oral FTC/TDF showed 44% overall efficacy among MSM[51]; however, drug adherence was not complete within the trial. Subsequent analyses showed that the ability of the drug to prevent infection was drastically affected by the number of doses taken on a weekly basis, with efficacy rising to >90% when the drug was taken more than four times per week[52]. FTC/TDF has also been shown to be effective in injection drug users (IDUs), where a Bangkok trial demonstrated 48.9% reduction in HIV incidence[53]. Success with PrEP has also been demonstrated in heterosexual transmission. The TDF2 study group showed a 45.7% reduction of incidence in women using PrEP[54]. The Partners PrEP study of heterosexual serodiscordant couples from Kenya and Uganda showed 71% (TDF) and 66% (FTC/TDF) efficacy among women, and 63% (TDF) and 84% (FTC-TDF) effective among men, though these results were not significantly different between genders[55].

Though clinical trials have shown promising results for PrEP, adherence to a daily pill appears to be critical for its success. Both the Bangkok IDU and FDT2 trials showed increased adverse events of

nausea with treatment relative to placebo controls, which may limit uptake. Recently published results from the VOICE trial showed a complete lack of protection from both oral and topical gel forms of tenofovir PrEP due to low adherence, and showed that adherence was lowest in high-risk populations of women under 25 years of age[56]. The VOICE trial clearly demonstrated the need for prevention therapies to be better integrated with sexual behaviour in order to improve uptake within key populations. Further, there is concern that increased drug resistance in the circulating HIV strains will occur with increased ARV use. Daily oral PrEP appears to be biologically effective, but cannot overcome difficulties of adherence. Promising results from Ipergay trial of on-demand PrEP at the 2015 Conference of Retrovirus Infections (CROI) showed that intermittent prep use around the time of exposure (one pill taken at least two hours prior to unprotected intercourse and two within the following 72 hours) was 86% effective in reducing transmission among MSM[57]. The requirement of using PrEP around time of exposure will likely improve adherence, reduce the risk of drug resistant mutations, and will greatly reduce the cost burden of daily therapy. Injectable PrEP may also have drastic benefits for adherence. Long-acting, injectable formulations of the non-nucleoside RT inhibitor, rilpivirine, and integrase inhibitor, S/GSK1265744, have been shown to be stably expressed for three months after injection in several animal models[58,59], and have shown protection against high-dose SHIV intravaginal and intrarectal challenge[60,61]. Rilpivirine and S/GSK1265744 are an attractive HIV prevention strategy for at-risk women taking the hormonal contraceptive, Depo-Provera, which is also administered as an injection every three months. Both on-demand PrEP and long-acting injectables are promising alternatives to daily pills that may allow for better acceptability in high-risk populations. Though major success in PrEP development has been made, adherence is still a major barrier to successful implication of this therapy, and adequate uptake may not be uniform across all at risk populations; thus, the additional use of rectal microbicides with PrEP has been suggested to increase the success of these new HIV prevention tools[62]. However, major knowledge gaps regarding the safety and efficacy of rectal microbicides at this mucosal surface have limited their development.

1.5.4 HIV microbicide trials

Microbicides are topical products containing anti-HIV drugs that are designed to be applied directly to mucosal surfaces. They have been formulated as gels, creams, films, fast dissolving tablets, and intravaginal rings[63,64]. Microbicides are designed to be applied around the time of sexual intercourse, or on a daily basis to consistently administer the drug. These vectors provide a direct route to administer therapeutics to the site of exposure, circumventing the need for drug to travel through the upper GI and peripheral blood system. This dramatically improves the pharmacokinetics of treatment, as direct application increases the concentration at mucosal surfaces and requires a lower concentration to be used, thereby reducing cost of therapy and lowering the risk of negative side effects in bystander tissues. The use of microbicides as an alternative form of HIV prevention is a promising condom alternative, as it allows the receptive partner control over their own protection, and is an alternative to lubricants, which are already commonly used during receptive anal, and other forms of intercourse[65].

Microbicides have demonstrated limited efficacy against vaginal transmission of HIV. The spermicide, nonoxynol-9 (N-9), was initially tested as a microbicide against HIV and other STIs. However, the benefits of N-9 use were conflicting, with studies showing no protective effect against HIV, Chlamydia or gonorrhea[66,67], and increased incidence of genital lesions and epithelial toxicity[68,69] at both the vaginal, rectal and penile mucosae. Interest in the field was renewed with the CAPRISA 004 trial of a 1% vaginal gel formulation of tenofovir use in women from South Africa. The trial demonstrated a 39% reduction in HIV incidence among all women, with protection increasing to 54% among high adherers, and no adverse events associated with use[70,71]. Nearly every clinical trial of prevention therapy has provided evidence that the success of microbicides and PrEP is contingent on adherence. To improve uptake of therapy, intravaginal rings (IVR) are now being developed. IVRs are inserted monthly, and allow for slow and controlled release of antiretroviral drugs directly to the vaginal mucosa[72], and will likely be used synergistically with current IVR contraceptive technology. TDF containing IVRs have shown protection against Simian Human Immunodeficiency Virus (SHIV) in pre-clinical macaque trials, and are expected to show increased protection and adherence in ongoing clinical evaluations (IPM-072

and MTN-020)[73]. Though progress has been made in the field of prevention therapy for vaginal transmission, the development of rectal microbicides is lagging.

Lubricant use is common among MSM and others practicing URAI due to the lack of natural secretions at this mucosal surface relative to the urogenital or oral mucosae[74]; thus, developing rectal microbicides presents a promising opportunity to integrate HIV prevention strategies into these high risk populations. Current progress in rectal microbicide research has not fully elucidated the safety and acceptability of these products, and rectal microbicides have not yet been assessed for efficacy against transmission. The UC781 non-nucleoside RT inhibitor showed efficacy against HIV in a vaginal explant model, but has not moved on to clinical trials[75,76]. Maraviroc, a CCR5 co-receptor antagonist, is a promising candidate; however, non human primate models have shown maraviroc was not protective against SHIV challenge despite high CCR5 binding in rectal tissue[77]. Currently, maraviroc is being investigated in combination with tenofovir for safety and acceptability in the CHARM program[76]. The first formulation of rectal gels were tailored for vaginal use; however, the hyperosmolarity and acidic pH needed for vaginal acceptability caused mucosal damage within the rectal compartment[78,79], which is adapted for an isotonic osmolarity and a neutral pH. Such damage would compromise the integrity of the mucosal barrier, overriding any protective effect of an antiviral. A reduced-glycerine formulation of a 1% TFV gel has shown improved rectal safety and acceptability in phase 1 study (MTN-007) using epithelial sloughing, fecal calprotectin, inflammatory cytokine levels and cellular immune markers as indicators of acceptability in the rectum[80]. Conversely, a systems biology analysis of the rectal immune environment in MTN-007 associated TFV usage with mitochondrial dysfunction and altered cell proliferation pathways[81]. Though increased investigation is needed to elucidate any adverse effects of a TFV rectal microbicide, if the formulation does result in activation of potentially cancerous pathways, it would overshadow any antiviral effects of the product in terms of both efficacy and uptake. Alternatively, it may be possible to model rectal microbicides after mechanisms of natural resistance to HIV, which would likely cause minimal disruption of the mucosa. The lack of success in rectal microbicide development underscores the

need to increase our knowledge of the healthy rectal mucosa, and define natural correlates of HIV immunity, which may provide promising alternatives to the current microbicide candidate formulations.

1.6 HIV-exposed seronegative immunity

Research into HIV-exposed seronegative (HESN) populations has previously used a reductionist approach to investigate mechanisms of natural immunity against HIV transmission and disease pathogenesis. The most renowned correlate of protection discovered from HIV resistant populations is the homozygous Δ32 CCR5 genetic polymorphism, which modulates the CCR5 co-receptor on susceptible immune cells, preventing viral entry[82]. To date, the only person to be successfully cured of HIV received a stem cell transplant from a CCR5(-)Δ32 donor, and study of this mutation helped the development of maraviroc, which is widely used as ARV treatment[83]. Cellular immune functions, such as HIV specific CD8+ and CD4+ T cell responses have been associated with protection in HESN populations[84-87]. Specific HIV-neutralizing IgA responses found in the sera and mucosal secretions of HESN individuals have become a target for vaccine development[88-92]. Further, polymorphisms within the IRF-1 gene identified in a HESN population of commercial sex workers have been thought to modulate T regulatory cell development and mucosal cytokine expression to favour host defense[93]. Several studies of HESN women have pointed to reduced immune activation as a critical mechanism of protection against HIV acquisition; reduced expression of T cell activation markers, lower baseline expression of pro-inflammatory cytokines and increased expression of regulatory T cells (Tregs) that suppress cellular immune activation may lead to decreased target cell availability at mucosal surfaces and lowered risk of HIV infection[86]. Characterization of these singular immune responses has provided critical information on immune protection against HIV; however, novel approaches to understand HESN immunity should focus on multifactor immune responses within biological systems that could elucidate complex mechanisms of protection.

Proteomic analysis tools can be used to provide a deeper understanding of immune responses within HESN populations. Proteomics analysis is unbiased towards previously characterized factors of HIV

immunity, and can elucidate complex biological interactions at mucosal surfaces to gain a more complete insight into natural immunity[94,95]. Analyses of mucosal secretions have been able to investigate hundreds of soluble immune factors at these frontline barriers to HIV. Several cohorts have associated increased expression of single soluble antiviral proteins (RANTES, SLPI, MIP1-α/β, and other cc chemokines and cytokines) with reduced susceptibility to HIV[84,93,96,97]. Indeed, heightened levels of antiviral factors within secretions may lead to protection against infection; however, immune pathways as well as other factors with unknown immune properties may be equally important in HIV immunity. Our lab has been investigating mucosal secretions using proteomic tools to uncover novel biomarkers of HESN immunity. Previous proteomic investigation of HESN women from the Pumwani sex worker cohort in Nairobi Kenya identified elevated serine and cysteine antiproteases within their female genital tract secretions, while similar analysis of oral secretions from a population of HESN MSM found heightened levels of basic proline rich protein 2 (bPRP2)[98,99]. While these proteins have known functions in general immunity, many had not previously been characterized in the context of HIV infection at mucosal surfaces; therefore, an unbiased proteomics approach to understanding HESN mucosal immunity is powerful in its ability to uncover novel biomarkers, biological functions and immune pathways associated with reduced HIV susceptibility, and has greatly contributed to our knowledge of HIV transmission through mucosal surfaces.

1.7 HIV transmission across mucosal barriers

Greater than 90 percent of HIV infections occur across mucosal surfaces during sexual intercourse, which include vaginal, penile, foreskin and rectal tissues[100]. The majority of our knowledge of HIV transmission through the mucosa originates from the study of the female genital tract (FGT). The lower FGT is the primary site of microbial exposure, and consists of upper (endocervix) and lower (vagina and ectocervix) regions[101]. The FGT is lined by the stratum cornea, the outermost layer of the epithelium, consisting of multiple layers of dead cells. Underneath, the vagina and ectocervix contain an unkeratinized multilayered squamous epithelium[25,102,103], which thins into a single layer of columnar cells

at the ectocervix[104]. It is unclear if HIV preferentially infects any sub-compartment of the FGT. The relatively thin epithelium of the endocervix may have increased likelihood of infection; however, increased surface area within the vaginal and ectocervix may offer more opportunities for virus to breach the outer barrier[105]. Further, the small transformation zone between the ecto- and endo- cervix may be particularly at risk for infection due to an increased population of CD4+ T cells[106]. The FGT epithelia may be penetrated by virus due to natural disruptions between epithelial cells or through breaches that occur during physical damage during intercourse. Virus within the epithelium causes inflammatory stimulation of the genital epithelial cells (GECs), which can lead to destruction of tight junction proteins and disruption of the epithelial barrier[107]. Phagocytic langerhan cells within the epithelia may become infected or facilitate trans-infection of lymphocytes CD4+ T cells, macrophages and dendritic cells are considered to be the primary target cells for HIV replication, and increased presence of these cells at the mucosa during inflammation increases the probability of acquisition[102]. Inflammation is also characterized by increased immune stimulation of front-line immune cells such as phagocytic cells (macrophages or DCs) that secrete pro-inflammatory chemokines and cytokines, which leads to the recruitment and activation of CD4 T+ cells and increased likelihood of established viral infection[108]. Maintenance of the epithelial barrier and control of inflammation are critical to reduce susceptibility to HIV at the FGT mucosa. There is increasing evidence overlying mucosal fluid barrier may act to facilitate these immunological responses.

In addition to the epidermis, the FGT is covered in a layer of cervicovaginal mucosa (CVM) to protect against pathogenic invasion and is the first site of exposure during HIV-infection. CVM is replete with soluble protein factors secreted by the underlying epithelial and immune cells, and these proteins hold important functions in regulation of homeostatic functions within the vaginal compartment, as well as immune properties that protect against pathogenic invasion. Antimicrobial peptides (AMPs) such as defensins, lactoferrin and lysozymes have antiviral or antibacterial properties to kill these microbes, and may also be able to stimulate inflammation and lead to an adaptive immune response[99]. Specific to HIV infection, inhibitory chemokines SDF-1, MIP-α/β and RANTES block viral entry via CXCR4 and CCR5

co-receptors and can recruit interferon-producing cells to restrict viral replication at the site of infection[104,109]. Much of the research into HIV pathogenesis and susceptibility in mucosal tissues is limited to FGT, and to a lesser extent the oral mucosa, and foreskin tissues. However, our knowledge of the rectal mucosa in the context of HIV is lacking. Compartment- and gender-specific differences between these exposure sites may limit the translation of our knowledge of the FGT to the rectum. Thus, independent studies of rectal HIV acquisition and susceptibility are needed to fully understand this highly susceptible site of HIV exposure.

1.7.1 HIV transmission and susceptibility in the rectum

The risk of HIV acquisition through the rectum is significantly higher than other sites of mucosal exposure. The increased susceptibility of the colo-rectal mucosa to HIV infection is likely a consequence of its biological requirements to facilitate digestion and fluid reabsorption, rather than act as a reproductive compartment. Points of increased susceptibility in the rectum can be associated with each stage of early infection events, including crossing the epithelia barrier, establishing a founder population within the submucosa, and dissemination to lymphatic tissues, which are summarized in Figure 1, and are described in detail below.

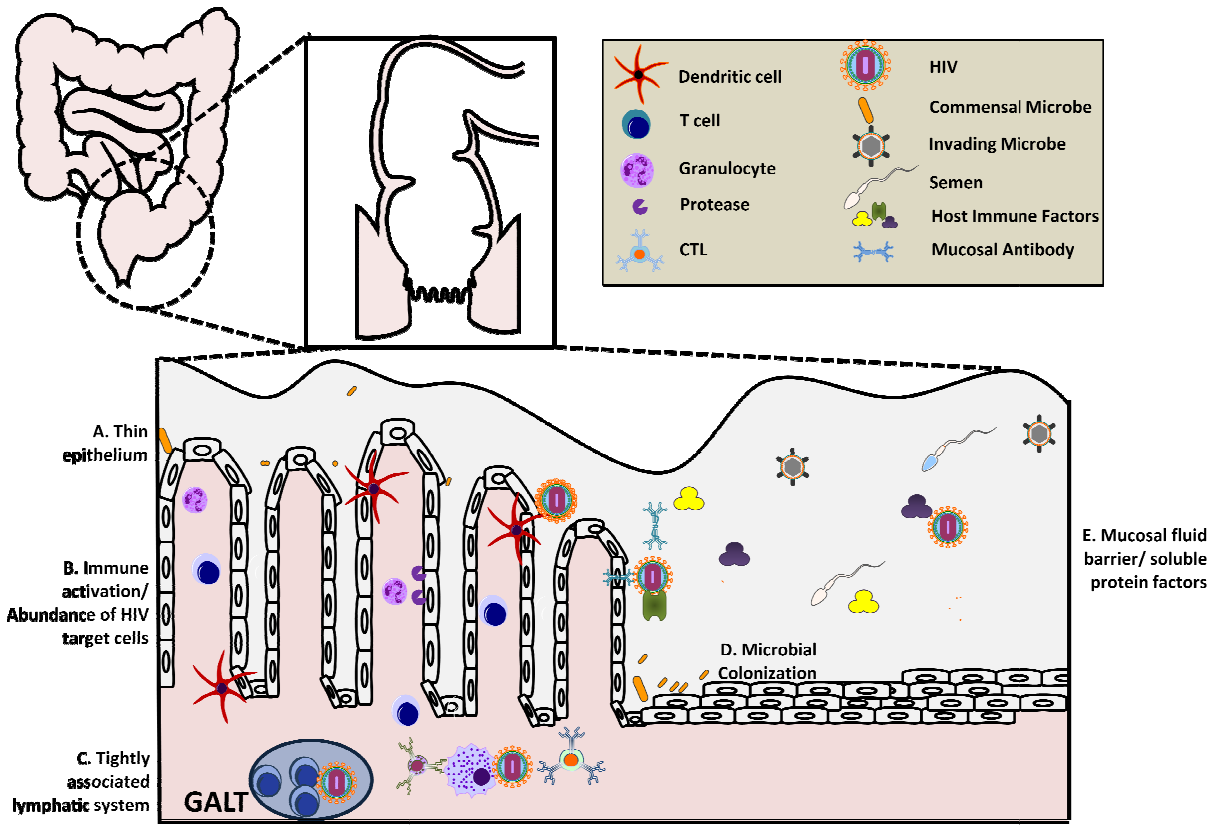


Figure 1: Innate functions of the ano-rectal compartment contribute to HIV transmission dynamics and susceptibility. The lower gastrointestinal lumen that is vulnerable during rectal exposure to virus includes the anal compartment and the recto-sigmoid colon. Susceptibility of the rectal mucosa to HIV infection has been associated with thin, damage prone layer of columnar epithelial cells (A), an extensive population of activated CD4+ T cells within the submucosa relative to other sites of HIV exposure (B), and a gut associated lymphatic system near the gastrointestinal lumen that allows for rapid viral dissemination (C). The external surface of the rectal mucosa is home to an extensive microbiome that may shape the immune environment to favour HIV infection (D). Further, the mucosal epithelium is lined by fluid secretions that contain a plethora of immune proteins, such as antimicrobial factors and inflammatory proteins that may play a role in HIV acquisition through the rectal mucosa (E).

HIV can penetrate the rectal mucosal barrier via direct infection of epithelial cells, breaches in the epithelial barrier, or transmigration through dendritic cells (DCs)[110]. The anal mucosal tissue is composed of multicellular squamous epithelia, much like the oral or vaginal compartment, and contains a large amount of muscular tissue, which is not optimal for HIV infection; however, the recto-sigmoid colon is lined with thin, single cell columnar epithelia to optimize fluid re-absorption during digestion. The columnar epithelium is prone to micro-trauma events during intercourse, allowing HIV direct access to

the HIV target cells of the submucosa. Though mucosal tears are responsible for a large proportion of rectal susceptibility, SIV-macaque models have shown that even with an intact barrier, the rectal mucosa has the highest rate of viral transmission[111]. Alternate mechanisms of barrier permeation include transcytosis of virions through epithelial cells, and uptake by dendritic cells (DCs)[110,112]. DCs are important mediators of HIV infection at mucosal surfaces. They lie at the outer surface of the submucosa and extend into the rectal lumen where they are specialized at antigen capture via surface C-type lectin receptors (DC-SIGN, langerin and CLEC4A)[113]. Rhesus macaques exposed to SIV show that 90% of initial infection occurs in DCs[114]; though much of this infection is non-productive, some virions are not neutralized upon phagocytosis. Infected DCs can migrate to lymphatic tissues to facilitate trans-infection of susceptible T cells within lymph tissues to establish productive infection[104,115]. DCs also express numerous pattern recognition receptors (PRRs), which interact with pathogen associated molecular patterns (PAMPs) on the HIV virion and induce cytokine signaling to increase T cell recruitment, which may either benefit the immune response or increase target cell availability.

Once past the rectal epithelium, HIV has access to a relatively dense population of activated CD4+ T cells compared to other mucosal sites of exposure[116,117]. This is due to the transient nature of the gastrointestinal lumen that is continuously exposed to bacterial antigens from human waste, which leads to a heightened basal level of activated CD4+ T cells that are highly susceptible to HIV infection relative to their quiescent counterparts[118]. Activated T cells have increased metabolic activity, amplifying the production of glucose and other mediators within the cell that are critical energy requirements for the virus to progress through each stage of replication[119]. Furthermore, CD4+ lymphocytes within the gut also harbor a higher percentage of cells expressing the CCR5 HIV co-receptors (70%) relative to tonsillar tissue that is exposed during unprotected oral sex (15%)[120]. The rectum also harbours CD4+ T cells with the X4-tropic CXCR4 receptor, yet the ability of these cells to respond to viral infection differs than that of the oral compartment. *Ex vivo* tissue studies have demonstrated that infected CXCR4+ T cells in tonsillary tissue secrete an array of HIV-inhibitory chemokines (MIP-1α, MIP-1β and RANTES) that are able to block further uptake of virus in surrounding CCR5 + T cells; however, CXCR4+ T cells from the

rectum are unable to produce this blockade response and remains susceptible to secreted virus[120]. The highly active, CCR5-dominant T cell populations within the rectal mucosa strongly favour the predominant mucosal transmission of R5 virus and contributes to susceptibility at this site.

The gastrointestinal mucosa is tightly associated with the largest organized lymphoid system in the body. Apart from the skin, the gastrointestinal mucosa is the largest organ that interacts with the external environment[121], and requires an extensive immune repertoire in order to maintain a homeostatic function with transient contents of the GI tract. Gut associated lymphoid tissue (GALT) houses the largest population of immunocompetent cells in the body[121], and is the major site of viral replication during acute infection[27,122]. Apart from enhanced target cell availability, the close proximity of the GALT and the colorectal mucosal surface increases the likelihood of disseminated infection, as virus transmitted via the rectal mucosa would not need to travel through the blood and risk elimination through humoral immune functions. However, to date, no comparative analyses have been completed to assess susceptibility associated with viral dissemination between sites of exposure. Indeed, despite the obvious risk associated with rectal susceptibility in HIV acquisition and contributions to the epidemic, HIV transmission through the rectum is grossly understudied in the literature relative to oral, penile and vaginal transmission.

1.8 Mucosal fluid as an immune barrier

Mucosal tissue surfaces are covered by a fluid layer that acts as an additional immune barrier against HIV entry. Mucosal fluids are densely comprised of sticky glycoproteins, such as mucins, that trap invading microbes, which are then removed from the body cavity. Mucosal fluids have been repeatedly shown to have HIV-neutralization activity *in vitro*. A comprehensive investigation of the inhibitory capacity of several mucosal fluids by Kazim *et al.* 2006 demonstrated that whole, submandibular/subliungual and parotid saliva, breast milk, seminal plasma and cervicovaginal secretions from HIV-1 negative individuals significantly limit HIV infection. Indeed, the ability of saliva to limit HIV infection *in vitro* has been well defined in the literature[123-125]. Apart from the physical properties of this barrier, immune activity of these fluids was attributed to the presence of soluble immune proteins[126-

[129]. In addition to innate immune factors, mucosal fluids also contain immunoglobulins that facilitate ADCC activity, virus neutralization and immune induction[130]. Though the inhibitory activity of mucosal fluids associated with oral, vaginal and mother-to-child transmission have been well-investigated in the literature, the inhibitory capacity of rectal mucosal fluid, and the soluble factors within have not been characterized.

1.9 Soluble immune factors of mucosal fluid

Soluble immune proteins play an important role in limiting pathogenic invasion at mucosal surfaces. Studies of oral and vaginal mucosal fluid have shown that these secretions contain a plethora of antimicrobial peptides, inflammatory molecules and immunoglobulins that aid in the innate immune response. Several antimicrobial peptides isolated from oral and vaginal secretions, such as SLPI, Human β-defensins 2 and 3, LL-37 have demonstrated inhibitory activity against HIV-infection *in vitro*[131-134]. The CC family of chemokines may also play a role in mucosal HIV-pathogenesis. Regulated upon activation normal T cell expressed and secreted (RANTES) and macrophage inflammatory protein-1α and -1β (MIP1α/β) act as general inflammatory mediators and can competitively inhibit the HIV-1 Env protein (gp120) binding at the CCR5 co-receptor[135]. Further, mucosal fluid contains high levels of immunoglobulins IgA and IgM, known to be important for the binding and clearing of pathogens, increased phagocytosis of microbes and complement activation[136,137]. HIV-neutralizing IgA isolated from sera, oral and vaginal fluid has been correlated with HIV-protection in resistant individuals[89,90], and the potent antiviral activity of HIV specific IgA has made it an attractive vaccine target. The role of these factors in HIV pathogenesis have been somewhat defined in the context of oral and cervicovaginal infection and have been summarized in Table 1 (adapted from Romas *et al*. 2014)[138]. However, the role of these factors in HIV susceptibility through the rectum is understudied, which represents a major gap in our knowledge of HIV pathogenesis and is a major barrier to understanding HIV transmission through URAI.

Table 1. Selected list of immune proteins having described roles in HIV defense found in mucosal fluids

Mucosal Protein	Antimicrobial Activity	Proposed anti-HIV mechanism	Ref.
Mucins	Physical entrapment, sequestering and clearing of pathogens	Binding inhibitor, inflammation regulator	139,140
Cathelicidins	Disrupts pathogen cell membrane integrity	Early replication inhibitor, may increase HIV infection	141-144
Thrombospondin	Physical entrapment, sequestering and clearing of pathogens, regulation of neutrophil inflammation	Competitive CD4 binding inhibitor, inhibits HIV transcription through Tat	145,146
Complement Components	Acute-phase response, opsonisation of pathogens, inflammation	Binding inhibitor (CIq) of HIV-env gp	147,148
MIP1α/β (CCL3/4)	Inflammation	Competitive CCR5 binding inhibitor	149
SLPI	Disrupts pathogen cell membrane integrity	Binds HIV-cofactor annexin a2, inflammation regulator; epithelial maintenance	150-153
DMBT1	Bacterial agglutination, Influenza A agglutination	Facilitates contact of HIV-1 to protein-bound antimicrobial proteins and IgA	
Mucosal IgA	Physical entrapment, sequestering and clearing of pathogens, complement activation, regulation of bacterial colonization	Direct virus neutralization, Fc-mediated inhibition, NK cell mediated lysis (ADCC), enhanced phagocytosis, epithelial transcytosis	89,154,155
Basic Proline Rich Proteins (bPRP)	Soluble bPRPs bind dietary tannins and viruses to facilitate clearing. Adherent bRPR2 may promote bacterial infection.	Binding inhibitor	156-159
Human α-defensins	Disrupts pathogen cell membrane integrity	Binding inhibitor, HαD-4 modulates CXC4 expression in target cells, increases HIV susceptibility with prior bacterial infection	160-163 164-166
Human β-defensin	Disrupts pathogen cell membrane integrity	Chemotactic activity, HβD-2/-3 modulates CXCR4 expression in target cells.	160,161 133
Lactoferrin	Iron sequestering, disrupts pathogen cell membrane integrity	Fusion Inhibitor	167-169
Lysozymes	Disrupts pathogen cell membrane integrity	Cell killing, antimicrobial defense	170,171
RANTES (CCL5)	Inflammation	Competitive binding inhibition through CCR5 co-receptor	149
Elafin/Trappin-2	Bacteriocidal activity	Inflammation regulator	172-174
Serpin Antiproteases	Regulation of protease activity, regulate inactivation of host defense factors, regulate inflammation, and promote epithelial maintenance	α-1-antitrypsin (Serpin A1), alters NF-κB signaling to inhibit HIV replication in T cells, inhibits fusion via gp41. α-1-antichymotrypsin (Serpin A3) inhibits proteolysis of proteins that increase HIV susceptibility	175-178 179-183
Cystatin Antiprotease	Compliment activation; antigen presentation; inflammation	Cystatin B inhibits HIV replication via STAT-1 pathway activation in monocyte derived macrophages. Oral cystatins found to have anti-HIV activity *in vitro*	184-187 131,188

Many soluble innate factors have demonstrated anti-HIV activity *in vitro*, but their role in HIV susceptibility *in vivo* may be much more complex. Antimicrobial peptides such as defensins, SLPI and cathelicidin (LL-37) may be able to limit HIV infection; however, they have been associated with increased HIV acquisition[164,189]. The confounding role of these proteins in immunity largely comes from a combined antimicrobial and pro-inflammatory activity. Macrophage inflammatory proteins (MIP1α/β) are able to competitively bind the CCR5 co-receptor; however, increased levels of MIP proteins are also associated with their ability to recruit HIV target cells. The inconsistency of soluble immune factors to provide a protective response against infection elucidates the complexity of HIV-infection at mucosal surfaces, and highlights our limited knowledge of how factors such as immune cell availability, co-infections and inflammatory states at the mucosa determine the ability of the host response to block HIV infection. This complexity is a major limitation of investigating HIV-immunity on a single-factor basis. Using omics techniques such as proteomics and genomics to study multiple factors within a host system and how they work together to provide a response is a more powerful immune monitoring tool that may help overcome this barrier.

1.9.1 Antiproteases in mucosal immunity

1.9.1.1 Serine and cysteine proteases

Proteases or endopeptidases are enzymes that cleave peptide bonds. Serine proteases are broken up in to sub-categories based on the amino acid sites they cleave: trypsin-like proteases cleave positively charged lysine or arginine, chymotrypsin-like proteases cleave the hydrophobic residues tyrosine, phenylalanine and tryptophan, and elastase-like proteases cleave smaller residues such as alanine, glycine and valine[190]. Cysteine proteases are grouped into two superfamilies based on their enzymatic activity, and are known as interleukin 1β converting enzymes (ICE) and papain enzymes[191]; most relevant to immunity are the papain-like cysteine proteases, cathepsins, which include lysosomal cysteine proteases (B, C, H, L and S) and others (F, K O, V, W, X)[192]. Protease enzymes facilitate several critical

physiological functions in humans, such as food digestion, cell viability, blood coagulation, tissue remodelling, apoptosis, and host immunity and inflammation[190,193-195]. Their widespread activity in the immune system may influence HIV infection at mucosal surfaces.

Proteases are vital effector molecules that mediate early stages of the innate immune response at mucosal surfaces, such as microbial defense, inflammation and epithelial barrier maintenance. The complement system is a central component of antimicrobial defense, and relies heavily on proteases (C1 proteins and MASP) to cleave and subsequently activate downstream effector molecules (atypical serine protease factor B, C2, C3, C4 and C5) that facilitate microbial opsonisation, lysis of infected host cells and release of chemoattractants to enhance the innate immune response[196]. Neutrophils, mast cells and macrophages also facilitate microbial killing via proteases; post phagocytosis, cytosolic granules comprised of serine proteases (cathepsin G, elastase, and matrix metalloproteases (MMPs)) merge with phagosomes to degrade of the outer bacterial membrane[197,198]. Lysosomal cathepsins facilitate phagocyte inflammation and T cell immunity as they are responsible for degredation of engulfed microbes and antigen presentation via MHC II[199]. Proteases can also be released into the external environment to attack extracellular pathogens and can act as cytokines to further recruit of target cells to the site of infection[200]. Further, soluble proteases are also critical regulators of epithelial barrier integrity and tissue remodelling[201-203]. These protease enzymes are critical regulators of the host defense response, but also hold functions in viability of immune cells.

Proteases facilitate programmed cell death (PCD). Intracellular, “executioner” proteases are released from lysozymes or exist within the cytoplasm to facilitate chromosome breakdown and mitochondrial outer membrane permeabilization required for apoptosis; further, proteolytic caspase enzymes are members of the apoptotic signalling cascade[204]. Granzymes and cathepsins work in conjunction with caspase activity during Tumor Necrosis Factor Receptor (TNFR), B cell receptor (BCR), T cell receptor (TCR) and reactive oxygen species (ROS) stimulation events, making these proteases highly relevant in the control of immune cell viability[204]. Though the effector functions of proteases are critical for host defense, their unregulated activity can lead to adverse events associated with

host tissue damage, unwanted apoptosis, and overstimulation of inflammation. As proteases facilitate an irreversible cleavage event, their activity must be tightly controlled. Thus, regulation of protease activity is critical to maintain proper immune homeostasis at mucosal surfaces.

1.9.1.2 Serine protease inhibitors (Serpins)

Serpins (**ser**ine **p**rotease **in**hibitors) are the physiological regulators of serine protease activity. The serpin superfamily consists of over 1000 proteins across all kingdoms, with 37 human serpins divided into 9 clades (A-I) based on conserved amino acid sequence and function[205]. Only clade B contains intracellular proteins, while the remaining eight contain secreted factors. Serpins are characterized by their ability to inhibit serine proteases, though some cross-clade serpins may inhibit cysteine protease activity[175]. Inhibitory serpins are composed of three β-sheets, eight to nine α-helices, and a C-terminal reactive center loop (RCL). The RCL forms a covalent bond with the target protease in a "bait and trap" mechanism that competitively inhibits their enzymatic function[205]. Serpins are referred to as "suicide inhibitors" as the covalent bond, causes an irreversible inhibitory event between the serpin and protease[206]. Through their inhibitory activity on proteases, serpins are able to modulate a variety of physiological functions. Uncontrolled protease activity at mucosal surfaces can lead to destruction of localized cells and tissue structures, releasing inflammatory mediators into the environment and induction of innate cell recruitment and tissue damage; further, proteases themselves can act as chemoattractants, leading to mucosal inflammation. Therefore, serpin inhibition is critical for regulation of protease-mediated inflammation and host tissue damage.

Extracellular serpins have been shown to counter inflammation, and control tissue degredation through inhibition of proteases at mucosal surface, thereby having indirect anti-HIV function. Serpin A1 inhibits neutrophil elastase (ELANE), and loss of A1 in the lung mucosa has been linked to chronic pulmonary emphysema due to excessive tissue breakdown and inflammation from ELANE[207]. Serpin A1 may also act as a signaling molecule to reduce production of other pro-inflammatory compounds in macrophages and B cells[208,209]. Serpin A3 holds functions in epithelial maintenance; this antiprotease can

inhibit cathepsin G and mast cell chymases released during inflammation in order to prevent degradation of the extracellular matrix of cells within the lung mucosal epithelia[210]. As well, serpin G1 is a common acute phase protein within sera that inhibits proteases of the complement, contact and fibrinolysis systems, and has direct anti-inflammatory effects on macrophages[211,212]. Extracellular serpins at mucosal surfaces act as innate immune regulators after stimulation. In addition to their extracellular function, research into intracellular serpins from the B family is providing more evidence of a role for these serpins in immunity.

Intracellular serpins maintain many of the anti-inflammatory functions of extracellular serpins, as they largely function to regulate protease activity within intracellular granules or during phagocytosis of pathogens; however, these proteins have the additional function of preserving immune cell viability. The role of Serpin B1 in immunity is well defined, and is the best understood B family member in this context. Serpin B1 inhibits ELANE and cathepsin G within granulocytes, which store proteases within intracellular vesicles to aid in phagocytic digestion of bacteria, while intracellularly, B1 neutralizes cytoplasmic proteases released during cellular stress or phagocytosis, thereby reducing immune cell activation and/or cell death[177]. Serpin B1 has shown co-protective roles with serpin A1 in chronic inflammatory disorders such as cystic fibrosis and emphysema, where, much like its extracellular counterpart, B1 can decrease tissue damage and restore phagocyte function[213]. Serpins also protect cytotoxic lymphocytes (CTLs and NK cells) from self-inflicted damage during the release of perforin and granzyme B upon MHC binding of infected cells[204,214]. Studies have also shown serpins to be upregulated in memory T cells to prevent PCD and help preserve these highly effective immune cells[204]. The ability of serpins to inhibit cell death may preserve immune cell viability and epithelial integrity at mucosal surfaces. The regulatory functions of these proteins to increase epithelial barrier integrity and control excessive inflammation may be beneficial in host defense during HIV infection; however, the specific role of serpins in HIV infection has not been defined.

1.9.1.3 Cystatins

The cystatin superfamily of protease inhibitors consists of three families: cystatin, stefin and kininogen families, with novel families and subtypes still being identified[215,216]. Three types of cystatins exist: intracellular type 1 cystatins (A and B), extracellular type2 cystatins (C-G, M, S, SN and SA), and type 3 cystatins (L- and H-kininogens) which circulate within the blood[217]. These proteins have been isolated from epithelial cells, as well as immune cells (granulocytes, monocytes/macrophages dendritic cells)[218]. Cystatins consist of a one α-helix and a five-stranded β-pleated sheet, which wraps around the α-helix to form two β-hairpin loops. The hairpin loops form a hydrophobic projection that is complementary to the active site of cysteine proteases, which allows tight-binding, reversible inhibition to occur[219]. Their protease regulatory function is critical for the maintenance of almost all biological functions, and loss of their activity results in disorders such as neurodegeneration, cardiovascular disease, osteoporosis and arthritis, as well as cancer[216].

Cystatins are able to regulate the immune response through interference with antigen presentation, phagocytosis, nitric oxide production and cytokine expression, allowing for control of the immune response and stimulation[220]. A major function of cystatins intracellularly is the regulation of lysosomal cathepsins in polymophonuclear leukocytes to prevent cellular damage and excessive inflammation during the host defense response[221]; further, their activity within lysozymes also limits microbe digestion, interfering with antigen presentation and stimulation of an adaptive response[220]. Cystatin C has shown direct anti-inflammatory activity through the inhibition of complement activation, and inhibition of chemotaxis and apoptosis of neutrophils[184]. Further, the immunoregulatory capacity of cystatins has been exploited by several parasites, which use cysteine protease inhibition to counter host inflammation and phagocytic digestion[222,223]. Indeed, antiproteases have a diverse range of functions in cellular maintenance and the immune response; many of which are redundant with serpin activity. Antiproteases have been somewhat defined in the context of host defense at mucosal surfaces; however, there is lack of understanding of their roles in HIV transmission.

1.9.1.4 Antiproteases in HIV infection

The presence of cystatins at mucosal surfaces may be beneficial in HIV infection, as reduced inflammatory signalling and T cell activation through antigen presentation would decrease target cell availability. Apart from the indirect mechanisms of serpins in reducing HIV-susceptibility, direct HIV-inhibitory functions for these proteins have recently been discovered. Alpha-1-antitrypsin (AAT) or Serpin A1 can block replication through inhibition NF-κB transcriptional activation of the HIV genome, and can interfere with late-stage replication via sequestering HIV protease and interfering with proper Env processing and virion packaging[180,224]. The c-terminal residues of serpin A1 (virus inhibitory peptide (VIRIP) and cyclic peptide (CPM)) are also able to inhibit HIV-1 entry through gp120 or gp41-mediated fusion[182,225]. VIRIP has shown promise as a monotherapeutic in HIV-infected patients, as intravenous VIRIP is able to reduce viral load[226]. Antithrombin III (serpin C1) has demonstrated wide-spectrum antiviral activity against HIV, HCV, HSV and influenza[227-231]. Serpin C1 can modulate immune signalling pathways (ERK1/2 and NFκB) through prostaglandin synthetase-2 to promote antiviral activity during HIV infection[227]. Serpin E1 has also demonstrated recent antiviral activity against influenza via interference with influenza glycoprotein cleavage, preventing virion maturation[232]. The well described antiviral activity of serpins make this family of proteins a promising target for HIV intervention therapy in both HIV acquisition and disease. However, much of the described immune activity of these proteins is limited to their activity in blood rather than mucosal surfaces, and to date, little is known about the function of these proteins in rectal immunity.

A role for cysteine protease inhibitors in HIV infection is more complex than their serine counterparts as they have been associated with inhibition of HIV, as well as enhanced infection. Cystatin-containing fractions of salivary mucosal secretions have shown modest inhibitory activity in monocytes[131,139]. As well, recombinant forms of cystatin A and B are able to inhibit HIV infection in HeLa cells expression CD4 and CXCR4/CCR5 coreceptors[188]. These proteins may be able to limit infection at mucosal surfaces by blocking virus-epithelial interactions[233], or may be able to block replication through direct interaction with HIV Vif to interfere with Env protein maturation[234]. Cystatin B

has shown inhibitory effects in placental macrophages[186]; however, within other tissues it has been associated with HIV infection[235], and has been hypothesized to enhance HIV infection in macrophages through the inhibition of interferon via the STAT-1 pathway[187,236]. The role of cystatins in HIV acquisition is not fully understood, but their putative roles in regulation of inflammation and antiviral activity makes them intriguing targets for future studies in HIV acquisition at mucosal surfaces.

1.9.1.5 Antiproteases relevant to this thesis

Three antiproteases were found to be associated with a population of HESN MSM using a proteomic analysis of rectal mucosal secretions. A brief description of these factors is described below:

Serpin B3 and Serpin B4: Serpin B3 and B4, or squamous cell carcinoma antigens 1 and 2 (SCCA1/2) were originally isolated from cervical tumor cells. The genes for these proteins reside within the same chromosomal locus (18q21.3) and share a 91% sequence homology; however, these serpins show distinct substrate specificity[237]. Serpin B3 inhibits cysteine cathepsin (S, K, L) and papain-like cysteine proteases, while serpin B4 is a cross-clade serpin, inhibiting serine proteases cathepsin G and mast cell chymase and to a lesser extent, cysteine proteases[238]. Serpin B3 and B4 have been predominantly described in epithelial cells from the mucosa and skin, yet their expression in other immune cell subsets, such as neutrophils or mast cells has not been fully investigated[237]. Serpins B3 and B4 are generally intracellular factors found within the cytoplasm where they are able to regulate protease activity, tumor suppression/cell growth pathways, apoptosis, inflammation and angiogenesis[237]. Secretions of modest amounts of serpin B3 have been reported[239]; however, since both serpin B3/B4 lack the N-terminal signal peptide required for Golgi/ER mediated secretion, the mechanism by which these factors are still under investigation[237]. Thus the unknown function of these antiproteases within the extracellular environment may suggest novel roles in immune regulation and disease.

Cystatin A: Cystatin A has known roles in immunity and potential anti-HIV activity that may contribute to reduced susceptibility in HESN men. Cystatin A is expressed abundantly in epithelial and lymphoid tissues, and inhibits cathepsin B H and L. Cathepsins degrade tight junction protein complexes, such as demosomes and zonulins, and keratins that maintain epithelial barrier integrity at both the skin and mucosa. The regulation of these cathepsins by cystatin A has been repeatedly associated with proper epidermal cell adhesion and epithelial barrier integrity[240,241]. Cystatin A can regulate inflammation, as loss or mutated expression of the CYSA gene has been linked to the inflammatory disorders COPD, rheumatoid arthritis and psoriasis[242-244]; further, expression of cystatin A has been shown to control autoimmune reactions through regulation of the pro-inflammatory cytokine IL-8[245]. This antiprotease also has antiviral properties against HSV[246], and has recently has demonstrated anti-HIV properties[188]. The putative role of cystatin A in maintenance of the epithelial barrier, and regulation of inflammation, as well as novel properties as an antiviral immune protein would suggest that increased levels of cystatin A at the rectal mucosa would be protective against HIV acquisition.

This investigation used a comprehensive proteomics approach to characterize the soluble protein components of rectal mucosal secretions from a cohort of men from Stockholm, Sweden. It examined the composition and quantitative levels of known antiviral factors in rectal mucosa, and established rectal mucosal fluid, and the protein components within, as a factor that affects HIV infection *in vitro*. A second proteomic analysis of a cohort of HIV-exposed seronegative men who have sex with men (HESN MSM) in serodiscordant relationships characterized protein factors and pathways that may be important for reduced HIV-susceptibility in the rectum. Two proteins that correlated with reduced HIV-susceptibility in the HESN population, Serpin B4 and cystatin A, were assessed for novel antiviral activity *in vitro*; serpin B4, but not cystatin A, demonstrated novel antiviral activity. These results contribute to our understanding of the immunobiology of the rectal mucosa, which may help guide future studies of prevention technologies, such as microbicides or vaccines, and could help limit the spread of HIV during URAI and help slow the growth the HIV epidemic through new infections.

1.10 Central aims of this thesis

1.10.1 Global Hypothesis

The global hypothesis of this Master's thesis is that the fluid secretions of the rectal mucosa contains inhibitory properties against HIV, and certain immune factors within this fluid are responsible for affecting HIV-transmission and/or HIV infectivity. In order to test this hypothesis, this thesis will complete three specific objectives.

1.10.2 Objective 1: Proteomic characterization of the rectal mucosal fluid of healthy men

Rationale: Our knowledge of the immunobiology of the rectum is limited, and this represents a major barrier to understanding the pathogenesis of HIV infection at this high-risk site of exposure. This will be a proof-of-concept analysis to show that rectal mucosal fluid has the capacity to limit HIV-infection *in vitro*, establishing it as an important immune barrier and worthy of future study to understand HIV-susceptibility at the rectal mucosa. Additionally, this objective will establish methods to be used for the proteomic analysis of rectal fluid, and will determine baseline levels of immune factors within the rectal fluid of control populations.

Sub-hypothesis: Rectal mucosal fluid will be able to reduce HIV infection in an *in vitro* system. Proteomic analysis of this fluid will identify immune factors that may contribute to this activity.

1.10.3 Objective 2: Proteomic characterization of the rectal mucosal fluid of HESN MSM

Rationale: Previous studies of this and other HIV-exposed seronegative cohorts have demonstrated that these populations have unique mucosal immune responses in cervical, salivary and foreskin secretions associated with reduced HIV-susceptibility; however, rectal mucosal responses of HESN individuals have not been defined. Our novel proteomic analysis will elucidate mucosal immune differences in the rectal mucosal fluid of HESN MSM that may contribute to a reduced susceptibility to HIV-1 at this portal of entry.

Sub-hypothesis: HESN MSM will have differential expression of immune proteins within their rectal mucosal fluid that may contribute to a reduced susceptibility to HIV.

1.10.4 Objective 3: Determine the antiviral activity of proteins that are associated with reduced HIV susceptibility in HESN populations

Rationale: Though many correlates of protection have been defined in HESN populations, very few are followed up to understand their role in HIV infection. HIV-neutralization assays will be performed in a reporter cell line and *ex vivo* PBMC culture to determine if proteins that are differentially abundant in HESN MSM are able to limit HIV infection, and therefore are potentially contributing to reduced HIV susceptibility.

Sub-hypothesis: Proteins that correlate with protection in HESN MSM will demonstrate the ability to limit HIV infection *in vitro*.

Chapter 2: Materials and Methods

2.1 Ethics

The ethical committees at the University of Manitoba (HS15735 [H2012:313]) in Winnipeg, Manitoba Canada, and the Karolinska Institutet in Stockholm, Sweden, have approved this study. All participants gave written, informed consent.

2.2 Study populations

2.2.1 The Venhälsan cohort of HIV-exposed seronegative men

The Venhälsan cohort (Venereal Clinic for Homosexual and Bisexual Men) is comprised men who have sex with men in serodiscordant relationships with their HIV-positive partner. The cohort is operated through the Gay Men's Health Clinic via the Karolinska University Hospital in Stockholm, Sweden. These men have documented repeat oral exposure to their partner's virus, and their status as a HESN population has been well defined in the literature[89,90,98,247,248]. The HESN men have demonstrated unique oral and systemic responses to HIV in the form HIV-neutralizing IgA1 antibodies against their partner's virus in both the saliva and sera[248,249], and unique cc chemokine and immune protein expression in saliva[98,247]. Genetic screens for the protective CCR5 deletion have been performed, with only one individual (patient ID: 4B) possessing a homozygous deletion[250].

Criteria for selection of control and HESN men: Mucosal samples were collected from healthy male participants recruited by an advertisement at a blood donor clinic through the Gay Men's Health Clinic in Stockholm, Sweden. HESN MSM were recruited through the same clinic. Low risk individuals were determined if they had 0-1 sexual partners and were unlikely to have been exposed to HIV. Those defined as HESN MSM were homosexual males in a serodiscordant relationship with their HIV-positive partner for at least six months prior to sampling, and were followed-up with physical and psychological examinations every six months for a consecutive five visits. HESN participants answered questionnaires regarding age and sexual behaviour, such as monogamy status, number of casual partners, condom usage, and frequency of oral/anal sex. HIV-negative serostatus (IgG plasma screen) was confirmed in all study

participants, and individuals tested negative for Chlamydia (throat, urine and rectum) and gonorrhea (throat, urethra and rectum)[251].

2.2.2 The Multicenter AIDS Cohort Study (MACS) validation cohort

The MACS cohort contains a population of men who have remained HIV seronegative despite hundreds of high-risk exposures during the peak of the epidemic in the 1980s. The cohort involves approximately 7,000 gay men from the United States and is operated by four institutions: the University Of Los Angeles, California (UCLA), Northwestern University in Chicago, the University of Pittsburgh, and the Johns Hopkins University in Baltimore. These men were largely exposed prior to the discovery of HIV as sexually transmitted etiological agent for AIDS in 1986, which was a period when condom use was low among MSM and HIV positive individuals were undiagnosed, with high levels of viremia. Previous studies of this cohort lead to the discovery of a 32 nucleotide deletion (delta 32) in the CCR5 gene, which compromises the structural integrity of the CCR5 co-receptor[82]. Though this cohort has aged since the early 1980's, the resistant individuals have had some of the highest levels of HIV exposure of any HESN cohort, and may have retained their immunologic responses that made them resistant to HIV.

Criteria for selection of control and HESN men: The men defined as resistant within this cohort had at least 45 anal intercourse partners (median 92; range 46-504) in the 2.5 years prior to 1985, and remained IgG seronegative[252]. Men were recruited to the cohort beginning in 1985 and have been followed up with physical and mental examinations every 3-6 months since enrollment[253].

2.2.3 Collection of mucosal samples (saliva and rectal lavage; rectal sponge)

Venhälsan: Whole, un-stimulated saliva was collected in 50 ml vials, aliquoted and frozen at -80°C; participants were instructed not to eat or drink for two hours preceding. During the same visit, rectal lavage was collected after installing 5 ml of sterile PBS and then aspirating the fluid, which was subsequently filtered (2µM) to remove debris and immediately frozen at -70°C. Samples were shipped to the University of Manitoba on dry ice and stored at -80°C until used for proteomic analysis.

MACS: Rectal mucosal samples were collected by WEC-Cel sponges. Briefly, sponges were placed on rectal tissue through an anoscope. Sponges were placed in contact with the rectum for five minutes before removal through the anoscope and immediately frozen at -80°C until thawed and eluted for proteomic analysis.

2.2.4 Label-free proteomic analysis of rectal mucosal fluid

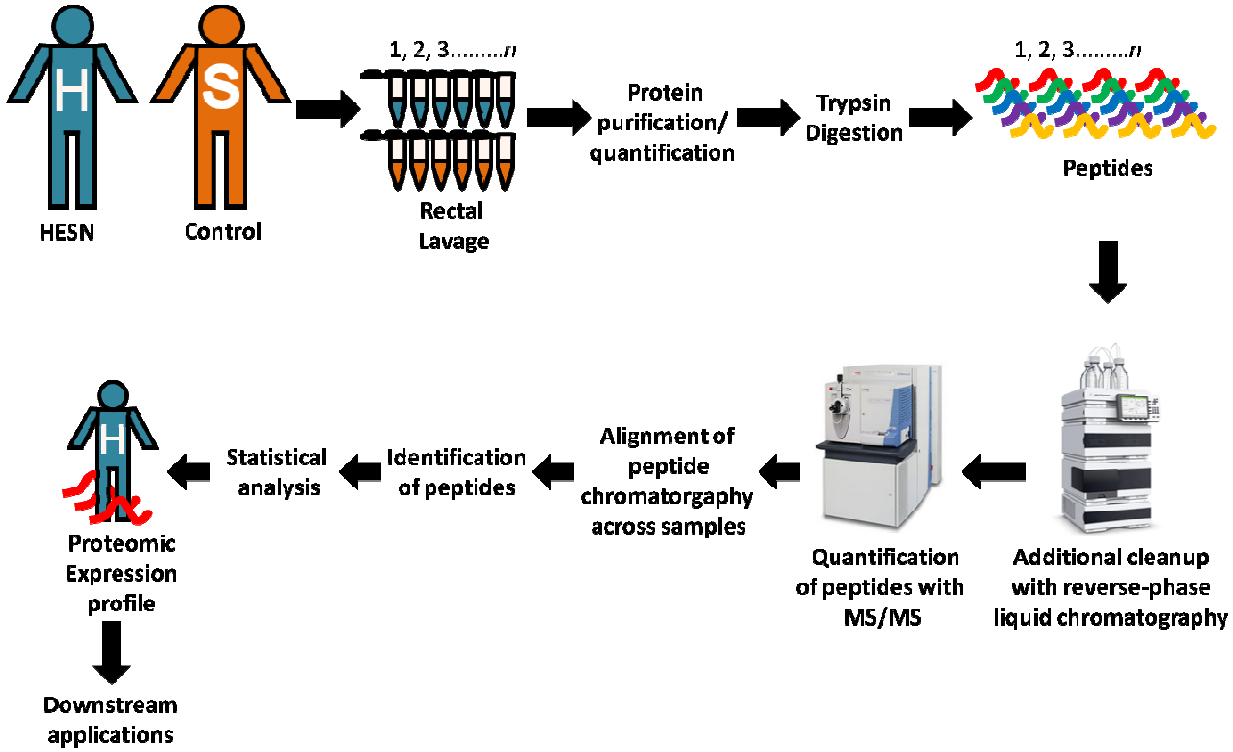


Figure 2. Overview schematic for rectal mucosal proteomics workflow. Using the HESN study as an example, the above figure shows the general workflow for analysis of the soluble mucosal proteome of the rectum using label-free, tandem mass spectrometry.

2.2.4.1 Protein purification and concentration

In order to fully purify and concentrate rectal lavage, each sample underwent ultracentrifugation for (5 min at 10,000g) to pellet out cellular debris and fecal contaminants. Supernatants were removed and subject to acetone precipitation to concentrate mucosal proteins and further clean the samples. Briefly, 100 µl aliquots of centrifuged sample were transferred into acetone-compatible 1.5 ml centrifuge

tubes. A 400 µl (4x sample) volume of cold (-20°C) acetone (Sigma-Aldrich; St. Louis, MO, USA) was added and vortexed. Acetone samples were incubated at -20°C for 60 minutes and centrifuged for 10 minutes at 13,000g to precipitate out protein. Supernatants were decanted and the proteinacious pellet was dried. Pellets were re-suspended in phosphate buffer saline (PBS) for quantification.

2.2.4.2 Qualitative assessment of mucosal protein contents via silverstain

A preliminary assessment of protein levels within rectal lavage samples was completed using the SilverQuest™ Staining kit (Invitrogen; Carlsbad, CA, USA) according to the manufacturer's "basic staining" protocol. Briefly, 5 µl aliquots of samples were added to 3.75 µl 4x LDS buffer, and 1.5 µl reducing agent (Invitrogen) in 500 µl microfuge tubes. Samples were heated at 95°C for 5 minutes, vortexed and spun down at 1,500g for 1 minute. Samples were run in on a 4-12% NuPAGE SDS-PAGE gradient gel in 5x MES Running Buffer (Novex Life Technologies; Burlington, ON, Canada) and electrophoresed at 200 V for 30 minutes using a gel electrophoresis unit (XCell4 SureLock™ Midi-Cell Runner, Invitrogen). The gels was then rinsed with reverse osmosis (RO) water, and processed using a series of buffers provided by the SilverQuest™ Staining kit. Briefly, the gel was fixed in 100 ml Fixative solution for 20 minutes with gentle rotation, followed by a wash with 30% ethanol (10 minutes, gentle rotation). The gel was incubated with 100 ml of Sensitizing solution for 10 minutes with gentle rocking, and subsequently washed with 30% ethanol (10 minutes) and 100 ml RO water (10 minutes). The gel was then incubated in 100 ml of Staining solution (15 minutes), and washed in RO water for 20-60 seconds to remove excess silver ions. Finally, the gel was developed in Developing solution for 4-8 minutes until bands started to appear. Once bands were resolved, the reaction was neutralized by adding Stopper solution directly to the Developer to avoid excessive background staining. The gel was read on Bio-Rad Universal Hood II (Bio-Rad; Hercules, CA, USA) using QuantityOne software (Bio-Rad; v4.6.9).

2.2.4.3 BCA quantification of protein

The protein concentration of each sample was determined by Novagen® BCA (Bicinochronic acid) assay (EMD Millipore; Etobicoke ,ON, Canada) according to manufacturer's instructions (96-well plate, micro-scale assay procedure). This assay detects protein levels between 20-20,000 µg/ml based on

the biuret reaction, which is the reduction of Cu^{2+} to Cu^{1+} ions in an alkaline solution. The reduced copper ions will chelate with the chromogenic Bicinochronic acid which produced a purple reaction complex which can be quantified. Briefly, a standard preparation of known concentrations of Bovine Serum Albumin (BSA) was produced over a range of 1000-0 µg/ml. A working solution of 200 µl BCA working reagent with 4 µl of 4% cupric sulfate per sample was prepared. Twenty-five µl of diluted rectal lavage samples (1:3 in PBS) were plated in duplicate in a 96-well plate, and 200 µl of the working reagent was added to each well. Plates were gently rocked by hand (60 seconds) and incubated for 30 minutes at 37°C. After incubation, plates were cooled to room temperature and read at 562nm on a Synergy HI Hybrid Reader (BioTek Instruments Inc.; Winooski, VT, USA) with Gen5 software (BioTek Instruments Inc.; v2.05.5). Protein levels within samples were quantified against the BSA standard curve.

2.2.4.4 FASP tryptic digestion

Once samples were cleaned and concentrated to workable levels, they were digested using filter-aided sample preparation (FASP) for Label Free MS/MS analysis[254]. This process involves the denaturation of proteins into peptides in three treatments (Urea, IAA, and DTT), followed by cleavage of peptide bonds by trypsin on a 10 kDa filter. Sample proteins are initially treated with urea (GE Healthcare Life Sciences; Mississauga, ON, Canada), a chaotropic agent that disrupts hydrogen bonds between water molecules in order to destabilize three-dimensional protein structures. Reduction of disulfide bonds with DL-Diotheritol (DTT; Sigma-Aldrich), and alkylation of cysteine peptide residues with iodoacetic acid (IAA; Sigma-Aldrich) to prevent their re-formation further opens up the protein to allow access of digestive enzyme to target sites within peptides. Finally, proteins are hydrolysed at carboxyl region of lysine and argentine amino acids by the serine protease, trypsin (Promega; Madison, WI, USA). Post digestion, peptides are small enough to flow through the10 kDa membrane, and can be eluted for further cleanup and MS/MS analysis. A more detailed protocol for the FASP digestion method used in the two proteomic studies for this thesis can be found below.

For the proteomic analysis of healthy men, 10 µg of protein was collected from paired saliva and rectal lavage samples from 10 different individuals to make 100 µg pools for each sample type. Each pool

was treated with 7x their initial volume of urea-exchange buffer (8M Urea, 50mM HEPES [Sigma-Aldrich], pH 8.0) for 30 minutes at room temperature. Treated sample pools were each run through a 10 kDa Nanosep filter cartridges (Pall Life Sciences; Port Washington, NY, USA) at 10,000 g for 5 minute intervals. After centrifugation, samples were reduced with 25mM DTT for 15 minutes, centrifuged at 10,000 g for 5 minutes, incubated with 50 mM IAA in the dark for 15 minutes, and washed with ammonium bicarbonate buffer (50 mM). Trypsin (catalogue no. PR-V5280, Promega) was added to pools at a concentration of 1.5 µg/100 µg protein and was incubated at 37°C overnight on the cartridge. Peptides were eluted off the filter with 50 mM HEPES and were dried via vacuum centrifugation. Samples were then and cleaned of any residual salts and detergents using reversed-phase liquid chromatography using a high pH RP, Agilent 1200 series micro-flow pump and Water XBridge column (Agilent Technologies; Santa Clara, CA, USA) with a step-function gradient that allowed all peptides for each pool to elute into a single fraction. The fractions were then dried via vacuum centrifugation and kept at -80°C and analyzed by label-free mass spectrometry.

Proteomic digestion for the Venhälsan HESN study was completed as the previous study apart from the following exceptions: each individual was treated as a separate sample, 75 µg of protein was used in each digestion, and HEPES buffer was used for the wash following IAA treatment rather than ammonium bicarbonate.

Digestion of samples from the MACS cohort study was modified to include removal of samples from rectal sponges. Briefly, samples were shaken in 300 µl PBS for 30 minutes at 4°C and 500 rpm. . Shaken sponges and excess liquid were transferred to a Costar Spin-X Centrifuge unit (Corning Inc., Corning, NY, USA) and spun at 10,000 g for 5 minutes at 4°C. Eluate was centrifuged again at 23,000g for 10 min at 4°C to remove cellular debris. Supernatants were digested using the above FASP method in equal volumes and quantified at the peptide level using a FluorProphile® Protein Quantification kit (Sigma-Aldrich) after reverse-phase liquid chromatography. Samples were re-suspended to 0.5 µg/µl in a low-organic solvent (section 2.2.4.5) prior to mass spectrometry analysis.

2.2.4.5 LC-MS/MS

Mass spectrometry analysis of all samples from the Venhälsan cohort were performed as follows: samples were re-suspended in a low-organic solvent solution of 2% acetonitrile (ThermoFisher Scientific; Waltham, MA, USA) and 0.1% formic acid (EMD; Mississauga ON, Canada) prior to LC-MS/MS. Samples were injected into a nanoflow LC system (Easy nLC; catalogue no. LC120, ThermoFisher Scientific) that was connected inline to a LTQ Orbitrap Velos (ThermoFisher Scientific) mass spectrometer. A 2 cm long, 5 µm particle-sized ReproSil-Pur C18-AQ column was used for peptide trapping and desalting. A 15 cm long, 3 µm particle-size C18 column (ThermoFisher Scientific) was used for peptide separation. The elution gradient was from 100% buffer A (2% acetonitrile [ThermoFisher Scientific], 0.1% formic acid [ThermoFisher Scientific]) to 32% buffer B (98% acetonitrile, 0.1% formic acid) in 12 minutes at a constant flow of 300 nl/min. MS spectra were acquired on the Orbitrap analyzer at 60000 resolution. After each MS spectrum, and automatic selection of the 10 most intense precursor ions were selected for fragmentation by CID, at 35% normalized collision energy. Samples from the MACS cohort were analyzed as above, with technical modifications for the Q Exactive Hybrid Quadrupole Orbitrap Mass Spectrometer (ThermoFisher Scientific).

2.2.4.6 Proteomic data analysis

Protein spectra were processed using Mascot Daemon (Matrix Science; Boston, MA, USA; v2.4) with IPIHuman database (v3.87) for the proteomic analysis of healthy men, while the HESN proteome study was completed using the UniProtKB/SwissProt (2012-05) Human (v3.87) database for human and bacterial proteins. Mascot search results were imported into Scaffold (Proteome Software; Portland, OR, USA; v4.4.1.1) to validate the protein identifications using the following criteria: 80% confidence for peptide identification, 95% confidence for protein identification, and at least two peptides identified per protein, using a decoy database with a 2% false discovery rate. The Scaffold data was imported into Progenesis LC-MS software (Nonlinear Dynamics; Durham, NC, USA; v4.0) to determine label-free protein expression levels from MS peak intensities. Feature detection, normalization, and quantification were all performed using default settings from the software. Retention time alignment was performed

automatically, and was manually reviewed for errors. Only peptides with charge states between 2+ and 10+ were included in this analysis; peptides outside of this range could not be separated from contaminants and were eliminated. Protein abundance levels were further normalized by total ion current within the software before exporting.

2.2.4.7 Statistical Analysis of Proteomic data

Fold-difference values were calculated relative to the mean expression of that protein across all samples within an experiment. Data was normalized to a Gaussian distribution using $\log_2$ transformation; normalization was confirmed using normal quantile plots. Total protein levels in each sample were variable in the HESN vs. healthy men experiment; to correct for this, samples that deviated greater than one standard deviation of the median protein abundance were removed as outliers to avoid skewing of the data.

Differentially abundant proteins were determined with two-tailed, independent t-tests (α=0.05 corrected for multiple comparisons using the Benjamini-Hochberg method) using GraphPad Prism (GraphPad Software Inc.; La Jolla, CA, USA, v6.05). Statistically significantly different proteins were further restricted to those with 2.4 or greater fold-difference (1.3 $\log_2$ FD, effect size=1.4) between groups (comparative analysis of saliva and rectal mucosal fluid of healthy men) or to those proteins with a 3.3 or greater fold-difference (1.7 $\log_2$ FD, effect size=2.3) to retain an experimental power of 80% for each proteomic analysis.

2.2.4.8 Functional Annotation

Individual proteins identified were subject to manual functional annotation using the UniProt database (http://www.uniprot.org/). To elucidate major biological functions within a group of proteins, the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Tool was used (National Institute of Allergy and Infectious Diseases; Frederick, MD, USA; v6.7; available from: http://david.abcc.ncifcrf.gov/). Statistically enriched bio-functions contained greater than 10 proteins and had p-value < 0.05 (FDR 5%, Benjamini-Hochberg correction).

2.2.4.9 Hierarchical clustering

Hierarchical clustering was performed on normalized, log transformed protein abundance values using Gene Cluster Software v3.0 (Eisen M., de Hoon M., Imoto S., Miyano S.; Standford, CA, USA; v3.0). Clustered of protein expression among individuals was generated with unsupervised complete linkage hierarchical clustering, using Pearson correlation coefficient as the distance metric. Dendrograms were viewed using Java Treeview (Saldanha, A.; Stanford, CA, USA; v1.1.6r4).

2.2.4.10 Biological Pathway analysis

Canonical pathways associations within proteomic data sets were determined using Ingenuity Pathway Analysis (Quiagen; Redwood City, CA, USA; v21901358). Significantly enriched pathways had a p-value < 0.05 and greater than five proteins associated. IPA network analysis was performed on the same protein subset for the HESN analysis.

2.2.5 HIV neutralization assays

2.2.5.1 X4- and R5-tropic virus production

The method used for of HIV-1 virus production in peripheral blood mononuclear cell (PBMC) culture is standard to Canada's National Laboratory for HIV Immunology, the University of Manitoba's HIV research group and others[255]. Briefly, eight tubes of whole blood were drawn from a single donor and combined. Lymphocyte PBMC populations were isolated on the same day using Ficoll-Hypaque density gradient centrifugation[256]. Isolated PBMCs were suspended to $2x10^6$ cells/ml in supplemented RPMI 1640 media (10% fetal bovine serum [GE Healthcare], 1% Penicillin-Streptomycin [ThermoFisher Science] and 20 IU human rIL-2 [NIH AIDS Reagent Program; Germantown, MD, USA]), and were stimulated with 5 µg/ml PHA (Phytohaemagglutinin; Sigma-Aldrich) for 3 days at 37°C and 5% CO_2. At 72 hours, PBMCs were spun out of the PHA-containing media at 600 g for 8 minutes, washed in 30 ml of RPMI, and spun down again at 600 g for 8 minutes. Washed cell pellets were re-suspended to $20x10^6$ cells/ml and1 ml of this cell suspension were infected with 3 stock tubes of HIV-BaL (10^6 titre, passage 3) or HIV-3B (10^6 titre, passage 5) in T25 flasks at 37°C/5% CO_2 for 4 hours with gentle rocking every 30

minutes. At 4 hours, 10 ml of supplemented RPMI was added to each cell infection and cells were incubated overnight. Twenty-four hours post-infection, cell suspensions were washed and fed fresh supplemented media; four days after infection, $20x10^6$ feeder cells from the same donor were added to the culture. Virus-infected cultures were monitored for virus production using p24 ELSIA, harvested 10, 19 and 28 days post-infection, aliquoted and frozen at -80°C.

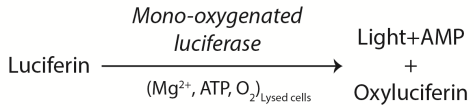
Virus titres of harvested stocks were calculated using a Tissue Culture Infectious Dose 50 (TCID50) protocol standard to our laboratory and others.[257] PHA-stimulated PBMCs are plated in flat-bottom 96 well plates at a density of $1x10^5$ cells/well. Stock virus from each harvest was plated with PMBCs in a 1:4 dilution series for 24 hours under standard incubation conditions. At 24 and 72 hours, residual input virus was removed from each culture well to assure only new p24 production is measured. The culture was continued and supernatants were harvested after 6 days for p24 quantification via enzyme-linked immunosorbent assay (p24 ELISA). P24 measurements were used to calculate 50% infectious dose using the Spearman-Kaber calculation[258]. Virus stocks with titres of $5x10^5$ or higher were used for future infectious experiments.

2.2.5.2 p24 Enzyme-Linked Immunosorbent Assay

Viral protein levels were measured as an indirect quantification of transcriptionally competent virus in culture. Currently, p24 ELISA assay is the gold standard for HIV measurement as HIV is not a plaque forming virus. ELISAs were performed as previously described in Nyambi *et al*. 2001[259]. Briefly, 96-well plates were coated with goat monoclonal antibodies against HIV p24 antigen. Plates were incubated with blocking buffer (Tris buffer) to bind any loose plate antigen and reduce background signal. Cell culture supernatants were incubated with blocked plates for 3 hours at 4 degrees to allow antigen binding. Supernatants were subsequently removed and washed six times in Tris wash buffer (washes were performed before and after each subsequent addition of reagent). The following reagents were added: rabbit anti p24, biotin anti rabbit, SAAP and SAAP substrate. Cleavage of substrate by the antibody-linked enzyme resulted in a chemiluminescent signal that was proportional to the amount of p24 in culture.

2.2.5.3 Cell viability/metabolism assay

Viability of cultured TZM-bl cells and PBMCs were measured using the CellTitre-Glo® Luminescent Cell Viability Assay (Promega), according to manufacturer's instructions. The assay estimates the number of viable cells in culture via quantification of adenosine triphosphate (ATP) as an estimate of metabolic activity. Briefly, immediately after cell culture supernatant harvest for p24 ELISA, cell pellets were incubated with CellTitreGlo reagent to lyse cells and release intracellular contents, such as ATP, into the solution. The reagent solution also contains luciferin substrate and luciferase enzymes. In the presence of intracellular contents (magnesium ion, ATP and molecular oxygen), the luciferase enzyme is activated, and catalyzes the oxidation of luciferin, yielding the following reaction:



The oxidation of luciferin produces a light signal that is proportional to the amount of ATP released. ATP is the molecular form of energy, and is used as an estimate of metabolically active cells in culture. The viability of experimental culture wells (i.e. in the presence of recombinant antiproteases) are calculated relative to negative control wells (no antiprotease added) to quantify the per cent viability of experimental wells, quantifying any toxic effect of the recombinant protein on culture that may reduce viral infection. Experiments were performed in triplicate, and one-tailed t tests ($p<0.05$) to determine any loss of cell viability in the presence of recombinant protein.

2.2.5.4 Recombinant protein production

Serpin B4 and cystatin A were produced commercially by GenScript USA Inc. Proteins were produced from a pUC57 contract containing either the Serpin B4 coding sequence (FASTA reference: sp|P48594|) or cystatin A sequence (FASTA reference: sp|P01040|). Protein was produced in 1L of bacterial culture with one-step affinity purification and endotoxin removal for a final purity of 75% and an Eu/µg of 0.2ng/µg. Recombinant protein was stored at -80°, aliquoted, and used on ice to avoid protein denaturing.

2.2.5.5 HIV inhibition assays of mucosal fluid in TZM-bl culture

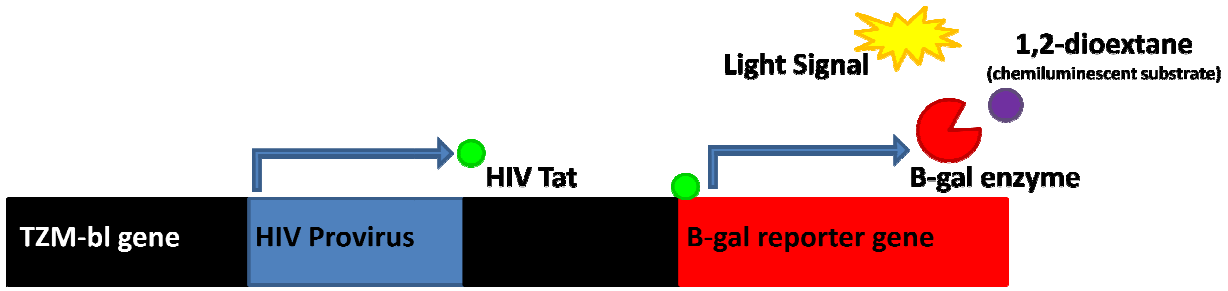


Figure 3: TZM-bl reporter genes are induced by HIV Tat protein transcribed from the integrated HIV provirus. TZM-bl cells are an HIV-permissive cell line with reporter gene activity that allows for quantification of viral replication in culture. Cells have been transfected with a β-galactosidase gene that is transcribed upon promoter binding with HIV Tat protein, which is produced from the transcribed HIV provirus from an infected cell. The presence of β-galactosidase is detected with the addition of the chemiluminescent substrate, 1,2-dioextane, which gives a luminescent signal upon cleavage. The signal is measured by a spectrophotometer in relative light units (RLUs), which is directly proportional to the amount of transcriptionally active HIV provirus.

Conditions for HIV-infection assays using a TZM-bl cell line were established previously by Aboud *et al*. (2014)[188]. TZM-bl cells are commonly used as a preliminary screening model against potential antiviral compounds as they allow rapid quantification of HIV Tat gene expression to estimate viral infection in culture, and can measure antiviral activity against both X4 and R5 HIV viruses. TZM-bl cells are modified HeLa cell line generated from the JC53-bl cell line that expresses stable amounts of CD4 and CCR5. β-galactosidase and luciferase genes under the control of the HIV-1 promoter have been integrated into the TZM-bl reporter cells for quantitative analysis of HIV infection. The reagent was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl from Dr. John C. Kappes, Dr. Xiaoyun Wu and Tranzyme Inc (Druham, NC, USA).

The TZM-bl HIV-neutralization assay was used to determine the HIV-inhibitory activity of whole, sterile filtered rectal mucosal fluid relative to salivary mucosal fluid (Chapter 3). Briefly, TZM-bl cells were maintained under standard adherent cell line procedures in Dulbecco's Modified Eagle Medium (DMEM; Life Technologies) completed with 10% Fetal Bovine Serum (GE Healthcare) and 1%

Penicillin-Streptomycin (ThermoFisher Science). Cells were grown in T75 flasks to approximately 70-80% confluence, harvested with 5 µl trypsin for 5 minutes (37°C) and washed in media. Cells were then seeded in flat-bottom, 96-well plates were seeded with $1x10^4$ cells 2 days before infection. Immediately prior to infection, culture media was removed and cells were incubated with a 1:3 dilution series of whole, sterile filtered (0.2 µM) lavage fluid diluted in PBS (39.12–0.31 µg/ml salivary fluid and 40.73–0.32 µg protein/ml rectal lavage fluid); R5-tropic HIV-1 virus (BaL) was then added at an M.O.I. of 0.2 (3.92 µl/well), and incubated with cells for 3 hours. This assays only used an R5-tropic strain of HIV as these are the major strains found to establish a founder population in mucosal lymphocyte populations[260]. Negative control wells, containing only virus, cells and PBS were included. A protein control that plated Bovine Serum Albumin (Bio-Rad) in the same concentration range as antiproteases was also included to assure that we were not overloading culture with protein to inhibit viral infection in a non-specific manner. Virus and mucosal fluid were then removed and cells were incubated in complete DMEM. TZM-bl cell cytotoxicity in the presence of mucosal fluids was measured using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega) according to manufacturer's instructions. Experimentally treated TZM-bl cells were screened for viral infection after 72 hours incubation (37°C, 5% CO_2) according to a β-Gal Screen System (Life Technologies) in relative light units (RLUs) (Figure 3). Percent infectivity of experimental wells was calculated relative to the negative control and conditions were compared using a one-tailed t-test (α=0.05).

Correlates of reduced HIV-susceptibility identified during mass spectrometry analysis of HESN rectal mucosal fluid in Chapter 5 were also screened for HIV-neutralizing activity in the TZM-bl assay prior to analysis in PBMC cell culture. Assays were performed as described above with the several exceptions. One hour prior to infection, cells were incubated with a 1:2 dilution series of sterile filtered (0.2 µM), recombinantly produced antiprotease in PBS. As well, additional viruses were analyzed: after one hour, either the R5-tropic HIV-lab strain, BaL (MOI of 0.2, 11.5 µl/well), or R4-tropic strain, IIIB (MOI of 2.0, 36.6 µl/well), were added to each infected well for 3 hours. Proceeding infection protocol and infectivity measurements were implemented as above.

2.2.5.6 HIV inhibition assays in PBMC Culture

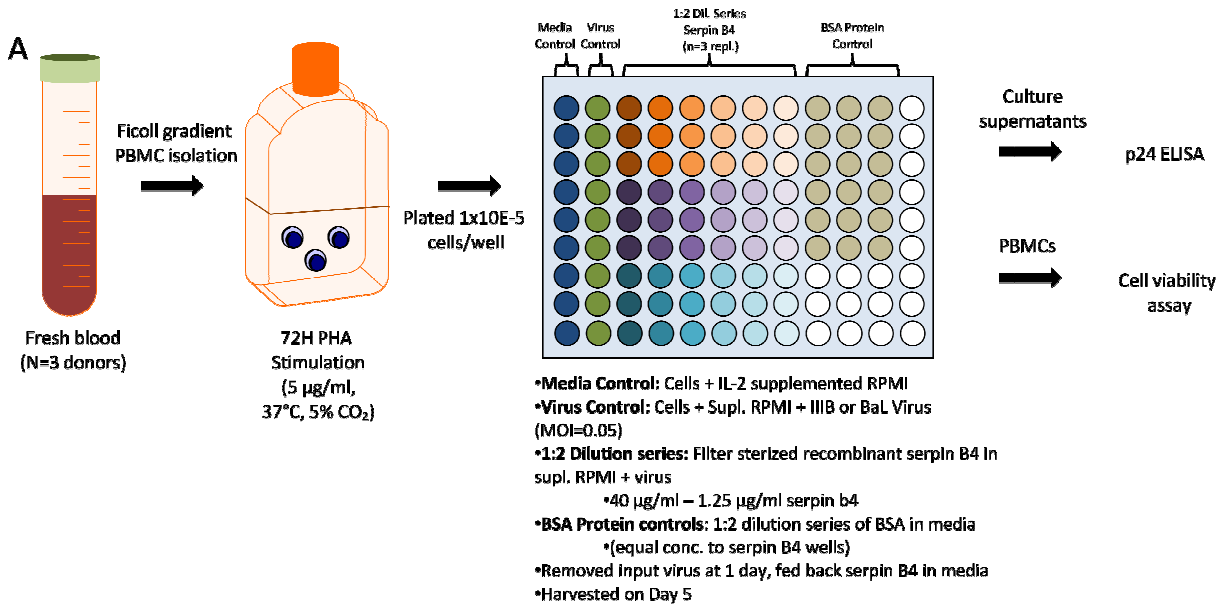


Figure 4: A schematic of HIV-neutralization assays developed in *ex vivo* PBMC culture to assess the HIV-inhibitory capacity of recombinant protein. A flow diagram shows the basic process and experimental conditions of the assessment of the inhibitory capacity of recombinant proteins against two HIV lab strains in PBMC culture, using serpin B4 as an example. Inhibitions were performed on PHA-stimulated Ficoll-Hypaque density gradient lymphocytes over 5 days. Both correlates were assessed against an R5- and X4- tropic lab strain (MOI=0.05). Cell supernatants were harvested for p24 ELISA quantification to assess viral infectivity, while cell pellets were analyzed by commercial ATP-luminescent detection assay to determine metabolic activity and overall viability of cell culture.

Peripheral blood mononuclear cell (PBMC) HIV inhibitory cell culture assays were used to analyze the antiviral capacity of a recombinantly produced serpin B4 and cystatin A protein against two laboratory strains of HIV. Briefly, lymphocyte PBMC populations were isolated from freshly drawn blood from local donors using Ficoll-Hypaque density gradient centrifugation, as mentioned previously[256]. Isolated cells were suspended in IL-2 supplemented RPMI media and stimulated with 5 µg/ml PHA for 3 days at 37°C and 5% CO_2. At 72 hours, cells were washed in media, re-suspended, and plated in 96-well plates at $1x10^{-5}$ cells/100 µl/well. Working stocks of sterile filtered (0.2 µM) serpin B4 in supplemented RPMI media were prepared in a concentration range of 40 µg/ml-1.25 µg/ml and plated in 100 µl aliquots/ well in triplicate, while cystatin A stocks were similarly prepared in a concentration

range of 100 μg/ml -0.4 μg/ml. Cells were incubated with the protein for one hour prior to infection to allow for any stimulation events to occur before virus was introduced. At time of infection, either an R5-tropic virus (BaL) was added to each well at an MOI of 0.05 (6.6 μl/well) or an X4-tropic virus (IIIB) was added at an MOI of 0.05 (1.5 μl/well). Cells were infected for 24 hours, input virus was removed and cell cultures were fed back IL-2 supplemented media with the same concentration of working stock. A set of media control wells consisting of RPMI, PBS and cells (no virus or serpin) were plated to control for background in the p24 and cell viability assays. Negative controls containing cells, supplemented media, PBS and either BaL or IIIB virus (no protease) were also included to assess levels of infection in the absence of antiprotease. Cultures were incubated at 37°C and 5% CO_2 for four additional days when cell supernatants were harvested for p24 quantification and cells were assessed for ATP production as a measure of cell viability.

HIV neutralization was quantified as a ratio of p24 production within experimental wells (diluted concentrations of serpin B4/cystatin A) relative to p24 production in positive control wells (per cent inhibition). Triplicate assays within three separate experiments (n=3 donors) were used to provide a distribution of p24 and ATP production under positive control and experimental conditions (n=9 replicates total). One tailed t tests were used to determine if a significant inhibition or decrease in cell viability was obtained with protein concentration ($p<0.05$). Increased potency of inhibition effects across virus strains were assessed using the 50% infectious dose (IC50) as calculated from dose-response curves in GraphPad Prism. Evidence for antiviral activity in PBMC culture was determined under conditions that showed significant decrease in HIV protein production, while maintaining viability in culture.

Chapter 3: Results

3.1 A proteomic analysis of the rectal mucosal fluid of healthy men

3.1.1 HIV-neutralization assays of rectal mucosal fluid in a TZM-bl reporter cell line

Results from the proteomic analysis of rectal mucosal fluid in healthy, non-HIV exposed men were published previously in Romas *et al.* 2013[138]. The capacity of rectal mucosal fluid to directly limit HIV replication *in vitro* had, to our knowledge, not been previously assessed in the literature. Within same-sex, HIV-serodiscordant relationships, receptive MSM are commonly exposed to partner virus at both the oral and rectal mucosae. The potent antiviral activity of saliva has been well established[123-125], and is reflected by an almost negligible HIV incidence through this exposure site[261]; however, despite high rates of acquisition, susceptibility relating to the rectal fluid barrier has not been as been investigated as it's oral or vaginal counterparts. Due to the considerable disparity in HIV susceptibility between these common exposure routes in MSM, we sought to determine the anti-HIV capacity of rectal mucosal fluid using saliva as a reference.

HIV inhibition assays were performed by incubating an R5-tropic HIV lab strain (BaL) with CCR5+/CXCR4+ TZM-bl reporter cells in the presence of salivary or rectal mucosal fluid. Mucosal fluid pools were constructed from matched salivary and rectal samples, with aliquots saved for mass spectrometry analysis. The data demonstrated that rectal mucosal fluid had the ability to limit HIV infection in the presence of susceptible cells. Rectal mucosal fluid inhibited HIV infection of TZM-bl cells in a dose-dependent manner (Figure 5A). Inhibition of infection was significant and reached 40% at mucosal protein concentrations of 2 μg/μl to a maximum of 61.5% at 64 μg/μl relative to a negative control (p=0.05, one-tailed t test; Figure 5A). The inhibitory capacity of rectal lavage fluid demonstrated in this assay was relatively mild compared to saliva. In agreement with the literature, our assays demonstrated that salivary fluid possessed higher antiviral activity, limiting HIV infectivity by 70-80% at as low as 2 μg/μl (p value=0.005, Figure 5B). This demonstrates that rectal mucosal fluid can inhibit HIV at physiologically relevant concentrations (6 μg/ml to 68 μg/ml), albeit with a lower capacity than saliva.

This may have relevance to what is observed *in vivo*, as demonstrated by a much higher incidence of infection upon rectal exposure than through oral exposure. Previously, the low incidence of oral HIV transmission has been attributed to high levels of soluble immune factors such as CC-chemokines, basic proline rich proteins, mucins, and the antimicrobial peptides SLPI, LL-37 and defensins. In an attempt to understand the role of these, and other soluble factors in HIV infection between sites of exposure, we used mass spectrometry to comprehensively define proteins contained within oral and rectal mucosal fluid, and define natural differences within these fluids that may be responsible for the observed discrepancy in HIV inhibition.

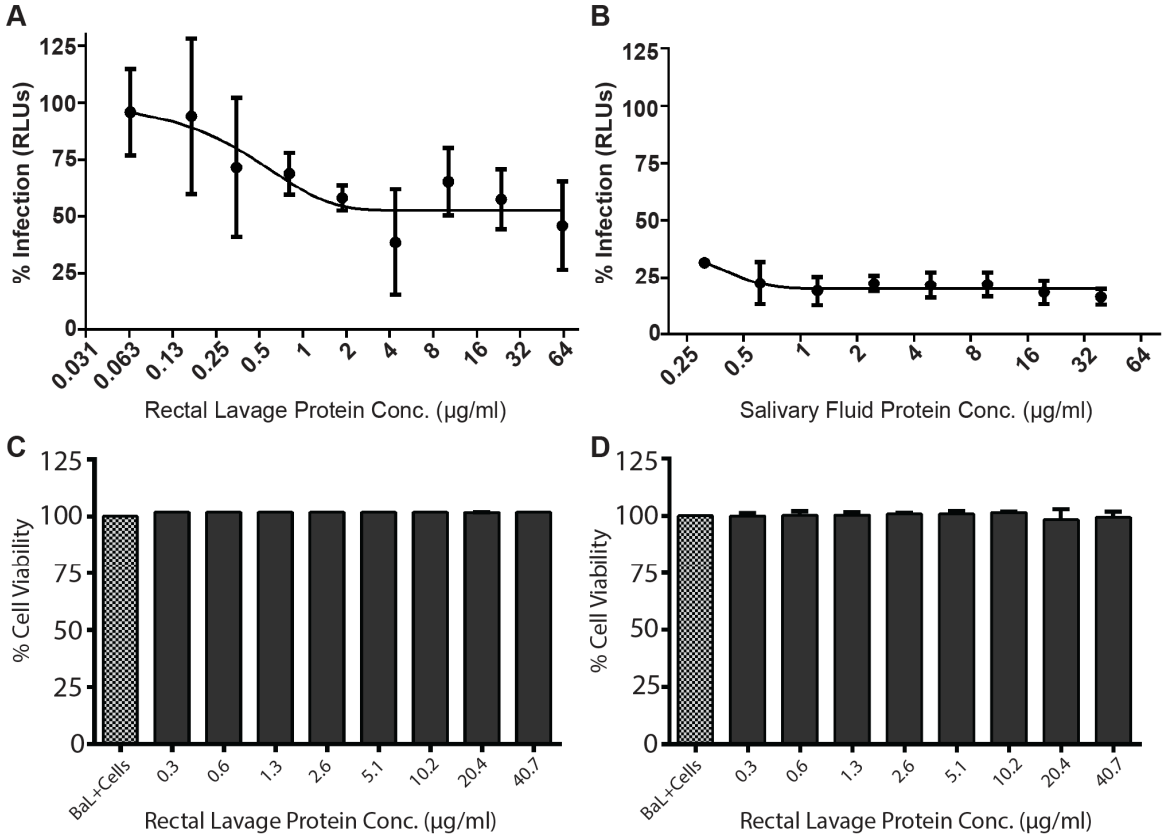


Figure 5. Rectal lavage shows inhibitory activity against HIV BaL in a reporter cell line. Rectal lavage was assessed for inhibitory activity against an R5-tropic, laboratory strain of HIV (BaL) within a CCR5+/CXCR4+ TZM-bl reporter cell line. Rectal lavage exhibited a significant inhibitory effect on HIV infection in TZM-bl cells (~40% inhibition) beginning at 2 µg/ml of protein relative to a negative control ($p=0.05$) (A). Parallel assays demonstrated that salivary fluid had a more potent anti-HIV activity (70-80% inhibition) at as low as 0.3 µg/ml ($p<0.005$) (performed by L. Aboud, B). Mucosal fluids were determined to have a negligent effect (~100% viability) on cell death based on a luciferase assay that measured the number of viable cells in each culture relative to infected cells in the absence of mucosal fluid (C and D).

3.1.2 Proteomic analysis of the rectal mucosal fluid of healthy men relative to saliva

The HIV-inactivating components of saliva were first described nearly 30 years ago[262], and since then many soluble immune components have been described in oral and vaginal mucosal fluid. However, immune proteins within rectal mucosal fluid and their contribution to HIV-susceptibility are largely unknown. Label-free mass spectrometry was performed on rectal lavage sample pools (n=10) in parallel with saliva sample pools (n=10). Aliquots of the same pools used in HIV inhibition assays in section 3.1.1 were analyzed to determine differentially expressed proteins that may be responsible for the disparity in HIV inhibitory activity observed between fluids (Figure 5).

Our proteomic analysis identified 315 human proteins expressed in both rectal and salivary mucosal fluid, with 72 proteins found to play a role in host defence and immunity (Figure 6). A small portion of proteins were unique to either fluid (one protein was unique to saliva [0.3%] and four proteins were unique to rectal mucosa [1.3%], but none had known roles in immunity. The majority of immune proteins identified across fluids hold functions in inflammation (9.6% of all 315 proteins identified) and/or antimicrobial defence (8.8%; Figure 6). Several other functional categories were identified, which included the following: antiproteases (4.0%), immunoglobulins (4.0%), wound healing (3.1%), acute phase response (2.1%), platelet activation (1.2%) and MHC Immunity (1.2%). Within the proteomic expression dataset, we identified many proteins without immune function; however, statistical analysis to determine differentially expressed protein between fluids was focused on proteins with known immune function to best determine immunological differences between compartments.

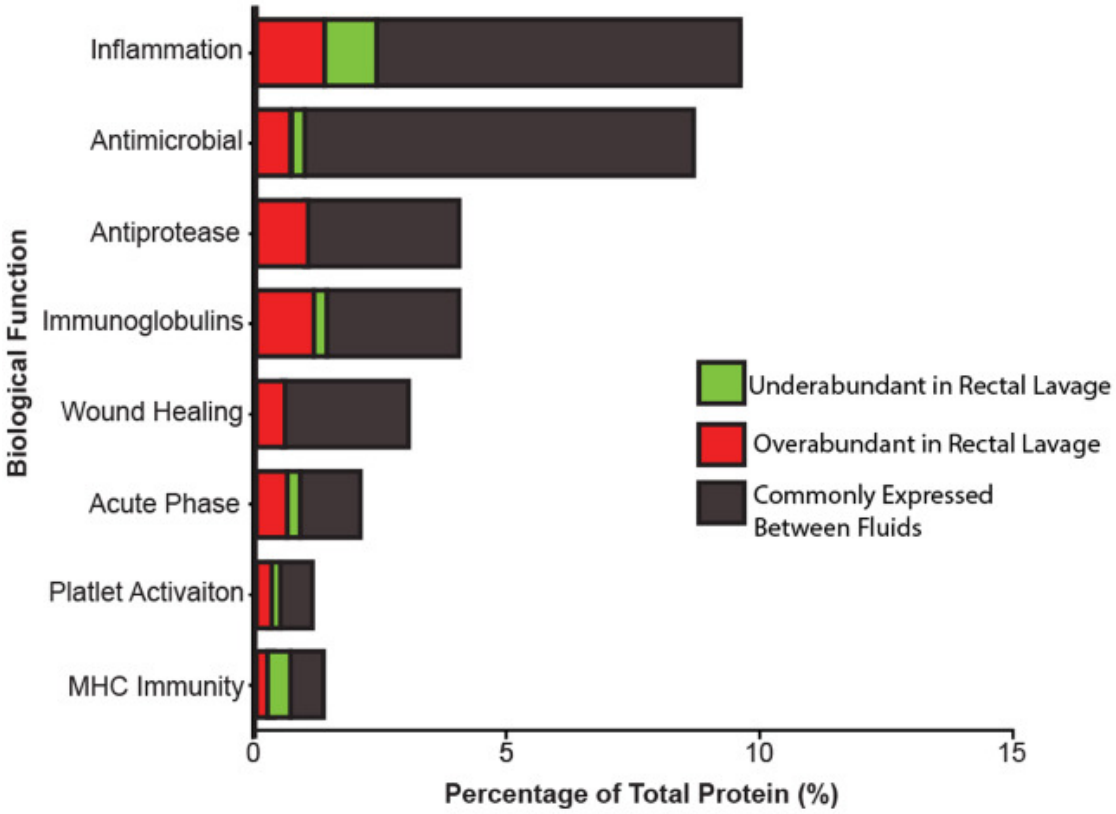


Figure 6. Functional categorization of immune factors identified in saliva and rectal lavage fluid. The 315 proteins identified by label free mass spectrometry in both rectal and salivary mucosal fluid pools were annotated by function using the UniProtKB database. Functional analysis found 72 of the 315 identified proteins had functions in immunity. The proportion of the total number of proteins known to possess each given function is displayed (proteins may be found under multiple categories if they have displayed more than one function). Differential expression analysis identified 49 immune proteins with no significant difference in expression between fluids ($p>0.05$, grey), 15 overabundant proteins in rectal lavage ($p<0.05$, red) and 8 proteins underabundant in rectal lavage ($p<0.05$, green), relative to saliva (corrected for multiple hypothesis testing using the Benjamini-Hochberg method). Differentially expressed proteins were found in multiple functional categories. The complete list of differentially expressed proteins is shown in Tables 2A and 2B.

The large majority of immune proteins identified by mass spectrometry were commonly expressed between the two fluids, with 29% (23 immune factors) differentially abundant between saliva and rectal lavage ($p<0.05$, Benjamini-Hochberg correction; Tables 2A and 2B). Certain immune factors were found to be higher in abundance in rectal lavage fluid; notably, mucosal immunoglobulins IgA and IgM (Table 2A), known to be important for the binding and clearing of pathogens, increased phagocytosis of microbes and complement activation[136,137]. As well, several antiproteases known to be important in immune regulation and tissue remodelling (serpins, inter-alpha trypsin inhibitor, and alpha-2 macroglobulin) and wound healing proteins (fibronectin) were found overabundant in rectal lavage (Table 2A). Though salivary and rectal mucosal fluid have a similar abundance of most immune proteins, several differentially expressed factors may suggest different immune mechanisms at each mucosal surface due to their unique immune requirements. Apart from this trend, the high similarity between these fluids is intuitive when considering that both compartments are a part of the gastrointestinal system and likely use similar mechanisms in maintaining mucosal homeostasis and immunity.

Table 2A. Average abundance of proteins significantly overabundant in rectal mucosa relative to salivary fluid as determined by mass spectrometry, and relative expression of these proteins in rectal mucosa compared to saliva.

Protein	Functions	Mean Abundance ($x10^3$)	SD Abundance ($x10^3$)	Mean Log2 Fold-Difference	SD log2 Fold-Difference	P-value
Calreticulin	MHC Immunity	3855.63	33.87	3.19	0.01	0.001
Protein S100-A7	Antimicrobial, Inflammation	3459.17	762.25	2.87	0.87	0.006
Myosin-reactive Ig κ-chain variable region	Ig-mediated immunity	1165.02	189.21	2.78	0.24	0.002
Serpin B3	Antiprotease, Apoptosis	745.43	159.88	2.14	0.32	0.009
Similar to VH-3 family (VH26)D/J protein	Ig-mediated immunity	231.38	25.91	1.66	0.16	0.003
Complement Protein C4-B	Complement Cascade	255.63	52.40	1.35	0.30	0.008
Fibronectin	Wound Healing, Acute Phase Response	270.48	44.52	1.30	0.24	0.004
Inter-alpha-trypsin inhibitor heavy chain H4	Antiprotease, Acute Phase Response, Inflammation	82.33	6.65	1.28	0.12	0.009
Serpin G1	Antiprotease, Complement Cascade, Acute Phase Response, Wound Healing	669.10	382.28	1.15	0.31	0.01
Lambda-chain	Ig-mediated immunity	26045.38	3313.85	1.00	0.18	0.002
Neutrophil gelatinase-associated lipocalin	Antimicrobial	1759.33	72.33	0.90	0.06	0.0006
Apolipoprotein B-100	Inflammation	18.53	1.43	0.87	0.11	0.002
IGA1	Ig-mediated immunity	54074.10	1245.78	0.87	0.03	0.0007
Alpha-2-Macroglobulin	Antiprotease, Complement Cascade, Platelet Degranulation	1952.70	26.49	0.80	0.02	0.0002
IGM	Ig-mediated immunity	2443.25	74.32	0.78	0.04	0.0007

Table 2B. Average abundance of proteins significantly underabundant in rectal mucosa relative to salivary fluid as determined by mass spectrometry, and relative expression of these proteins in rectal mucosa compared to saliva.

Protein	Functions	Mean Abundance ($x10^3$)	SD Abundance ($x10^3$)	Mean Log2 Fold-Difference	SD log2 Fold-Difference	P-value
Bactericidal/permeability-increasing protein-like 1	Antimicrobial	3.83	0.816	-2.73	0.31	0.004
Beta-2-microglobulin	MHC Immunity	20.10	0.689	-2.43	0.05	0.00005
Mucin-5B	Antimicrobial	31.85	18.16	-2.00	0.14	0.0006
Interleukin-1 receptor antagonist protein	Acute Phase Response, IL-1 Signaling, Inflammation	2.41	0.099	-1.96	0.06	0.0008
Ig lambda chain V region 4A	Ig-mediated immunity	2.48	0.615	-1.53	0.36	0.008
Zinc-alpha-2-glycoprotein	Antigen Processing	718.00	51.34	-1.33	0.10	0.0009
Metalloproteinase Inhibitor 1	Inflammation	16.24	0.30	-1.11	0.03	0.0001
Moesin	Inflammation	48.94	1.91	-0.49	0.06	0.002

3.2 A proteomic analysis of the rectal mucosal fluid of HESN MSM

3.2.1 Levels of HIV exposure in HESN MSM from the Venhälsan cohort

Risk assessment of sexual behaviour has previously been performed by Dr. Klara Hasselrot and other members of Dr. Kristina Broliden's research group at the Karolinska Institutet in Stockholm, Sweden (published in Hasselrot *et al.* AIDS 2009[248], Hasselrot *et al.* AIDS 2010[250] and Pérez *et al.* AIDS 2010[249]) and is summarized below.

Examination of epidemiological factors affecting HIV susceptibility showed that the HIV positive partners experienced a large range of viremia (range 50-475,000 viral RNA copies/ml) in the period 12 months prior to sampling (Table 3). ARV use among HIV positive partners was low prior to inclusion in this study; only two of the 25 individuals were on treatment with suppressed viral loads during the entirety of the relationship with their HESN partner. ARV use increased upon study follow up and counselling (22/25 HIV positive partners on ART and three with low viral loads and maintained CD4+ levels), but all but two HESN participants were exposed to viral loads above 50 copies/ml during at least two sampling visits during periods of no ART, or failed ART by their HIV positive partner. Thus, it is likely the study population has remained seronegative despite HIV exposure, and any reduced susceptibility observed in this cohort is not solely due to ARV use by the HIV positive partner.

Frequency of oral and anal intercourse was collected via survey and data showed that the HESN men are primarily orally exposed to their partner's virus (n=7 seldom/sometimes, n=13 frequent/often); with a small subset regularly practicing unprotected anal intercourse (n=2 receptive, n=5 insertive) (Table 3). Anal intercourse is generally under-reported in clinical survey data[263]; thus, it is possible more of these men were exposed rectally, though this could not be assessed within our study. Most men did not practice receiving ejaculate of the HIV-positive partner during oral intercourse[264], which may contribute to a lower risk of HIV exposure, though men were likely exposed through pre-ejaculatory fluid containing virus[265]. Further, 9/25 HESN men engaged in sexual acts outside of their serodiscordant relationship during the study period with men of unknown serostatus; however, sexual behaviour with additional partners was

equivalent to that of their primary relationship (predominantly unprotected oral intercourse and protected anal intercourse), though an increase in sexual partners would increase HIV risk[266].

Despite primary oral exposure, previously measured rectal HIV-specific responses suggest these HESN men are exposed rectally. Analyses performed by the Kristina Broliden's research group screened rectal mucosal fluid samples for the presence of HIV neutralizing IgA. Despite low rectal exposure, almost 30% of the men (n=7) were able to neutralize HIV infection with IgA isolated from rectal secretions. None of these men had documented receptive anal intercourse, and only one was practicing insertive anal intercourse, providing further evidence that this population has been exposed to HIV yet remains uninfected.

These HESN men are primarily exposed to HIV at the oral mucosa, but may demonstrate unique HIV immune responses at the rectal mucosa. Differences in the sexual behaviour, viral load of HIV positive partner, and presence of HIV neutralizing IgA, may influence protein expression within mucosal fluid. These variables will be considered in interpretation of global proteomic data in section 3.2.3.7 and in understanding the expression of serpin B4 in section 3.3.3.

Table 3. Relative levels of exposure in the Venhälsan cohort of HESN MSM (n=25 total).

Exposure Variable	*Estimate*
Frequency of unprotected intercourse	***Frequency (no. of individuals)***
Anal Intercourse	
Receptive	**3**
Seldom/Sometimes	**3**
Frequent/Often	**0**
Insertive	**5**
Seldom/Sometimes	**5**
Frequent/Often	**0**
Oral intercourse	
Seldom/Sometimes	**10**
Frequent/Often	**13**
Partner Viral Load (highest in 12 mo. prior to sampling)	***(copies/ml)***
Minimum	**50**
25% Percentile	**50**
Median	**50**
75% Percentile	**900**
Maximum	**475000**
Mean	**45106**
Std. Deviation	**118921**
Detected HIV neutralizing IgA in rectal lavage	***Frequency (no. of individuals)***
HIV-neutralizing IgA	
Present	**7**
Absent	**18**

3.2.2 Optimization of a proteomic analysis for individual rectal lavage samples

3.2.2.1 Analysis of protein content of rectal lavage

Methods for the label-free proteomic investigation of vaginal mucosal samples have been well established within our lab[99,267,268]. However, similar analysis of rectal lavage samples is logistically more challenging. Cervicovaginal secretions originate from several glands within the upper and lower female reproductive tract (FGT), and the fluid remains relatively stagnant within the lower FGT, making it easy to sample. The colorectal lumen is part of the larger gastrointestinal system, and is much less autonomous than the FGT. Mucosal fluid is constantly removed along with organic waste due to the transient nature of the GI system, and is also much larger than the vaginal compartment, making it difficult to get a concentrated sample of fluid when taken via lavage. Further, though both compartments contain a diverse microflora, the gut also contains food waste and other bodily excretions that exist as contaminants within the protein sample. Thus, our methods needed to be optimized to account for the dilution of the sample and utilize additional cleanup steps to purify rectal mucosal proteins.

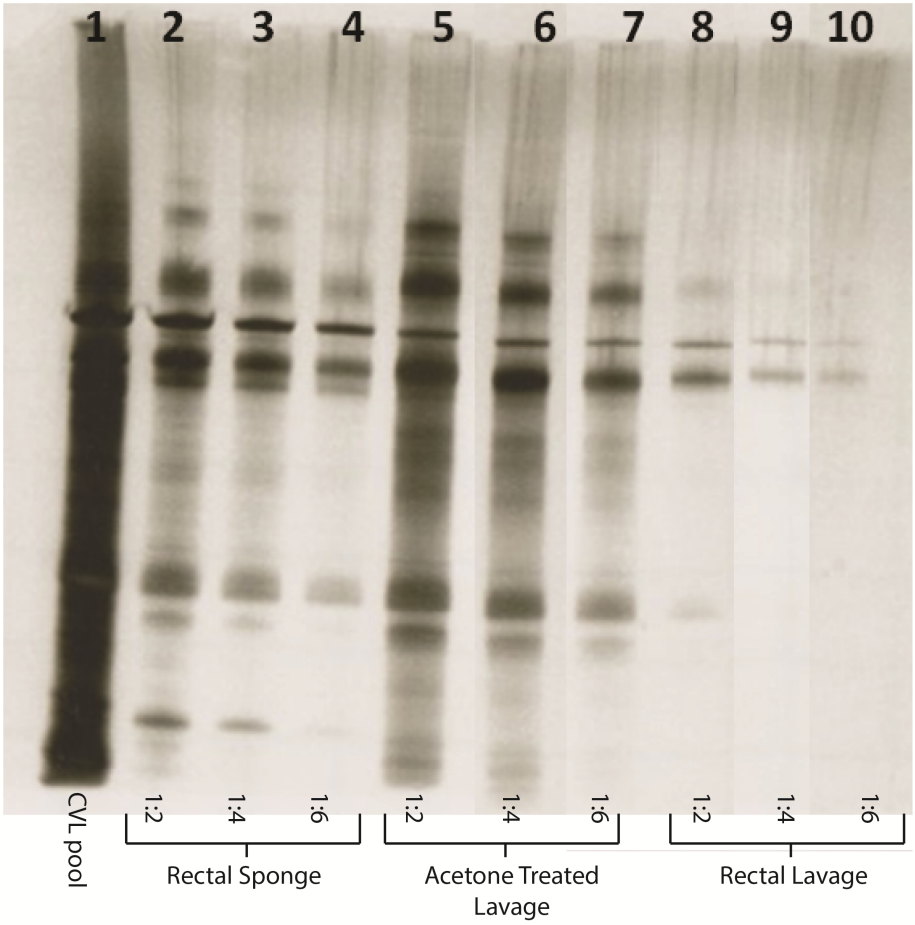


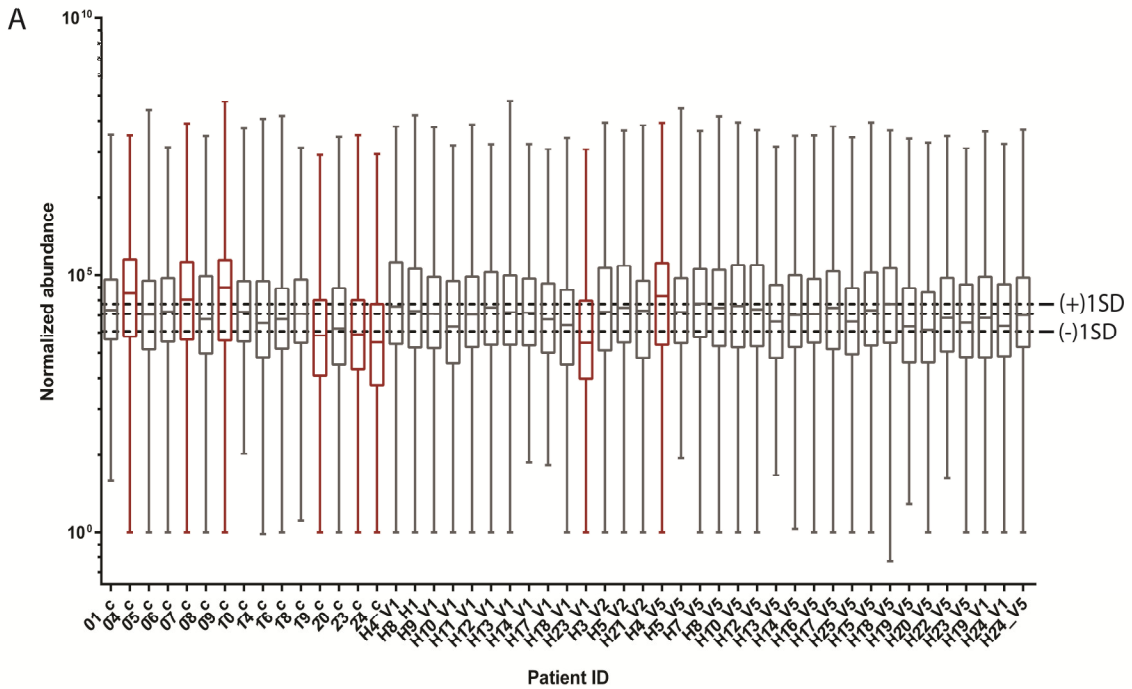
Figure 7: A comparison of protein concentrations within rectal and vaginal mucosal samples by 1-D gel electrophoresis. Each lane of this silverstain gel contains 2 µl of the following mucosal samples: Lane 1: CVL Pool; Lane 2-4: rectal sponge sample 1:2, 1:4 and 1:6 PBS dilution; Lane 5-7: acetone precipitated rectal lavage 1:2, 1:4 and 1:6 PBS dilution; Lane 8-10: centrifuged (un-concentrated) rectal lavage 1:2, 1:4, and 1:6 PBS dilution. Protein concentration is proportional to the intensity silver ions bound to the protein bands displayed in each lane. The 1:2 dilution of acetone precipitated rectal lavage sample has the closest intensity to the CVL pool, which proteomic methods are optimized for.

The previous investigation of rectal mucosal fluid (Chapter 3) pooled lavage samples prior to digestion to increase protein levels; however, pooling eliminates the possibility of analyzing variability in protein expression between samples, and would not allow us to look at the effect of differences in sexual behaviour between individuals. Individual rectal lavage samples were found to have a much lower protein concentration relative to cervicovaginal lavage, as shown in Figure 7, Lane 1 vs. Lanes 8-10. Ideally, rectal mucosal samples are taken via sponge sampling (Figure 7, Lanes 2-4), which yields a usable concentration of mucosal protein for digestion (approximately 10-30 µg/ml); however, rectal samples

collected four our studies of control and HESN men were taken via lavage (Figure 7, Lanes 8-10). By utilizing acetone precipitation of protein in rectal lavage, we were able to add an additional cleaning step to separate protein from cellular and bacterial debris, and were able to concentrate rectal lavage samples to workable levels for proteomic digestion (Figure 7, Lanes 5-7). Once precipitated, samples were re-suspended in PBS to equal concentrations as determined by BCA assay and digested in 20 μg aliquots for label-free MS analysis.

3.2.2.2 Normalization of rectal proteomic data

As mucosal samples taken via lavage can be variable compared to sponge samples, additional normalization procedures were conducted to minimize technical variability in protein content[268]. The software Progenesis QI for proteomics (Nonlinear Dynamics) normalized total protein content in each sample to total ion current (TIC) as a primary normalization step. Normalized abundance (NA) values determined by label-free LC-MS/MS were plotted to determine samples that were not adequately corrected by initial TIC normalization (Figure 8A). Criteria for determining outliers were set at one standard deviation outside of the median NA value for all proteins across samples. All samples with a median NA outside a range of 8,184.13 - 27,730.44 NA were excluded (Figure 8B). A total of 8 individuals were removed based on this criterion, 6 of which were control individuals (c) and 2 were HESN (H), though both HESNs had a lavage sample from an additional visit and could still be included in the study (Figure 8C). The remaining 9 Controls and 25 HESNs were analyzed further.



B Outlier Criteria

Statistic	Value
Median	1.80E+04
STD Deviation	9.77E+03
Upper Range	2.77E+04
Lower Range	8.18E+03

C Outlers Removed

Sample ID	Median
04_c	4.43E+04
07_c	3.83E+04
09_c	5.78E+04
19_c	6.78E+03
23_c	7.18E+03
24_c	4.98E+03
H23_V1	4.82E+03
H4_V5	3.90E+03

Figure 8. Outlier determination of rectal mucosal fluid samples by total protein content. Normalized abundance (NA) values for each protein identified by MS were plotted to view differences in total protein content between samples, with outliers highlighted in red (A). Median and standard deviation (SD) NA values for each sample were calculated and used as a measure for determining sample outliers. Outlier criteria were set at one standard deviation outside of the median NA value for all proteins across samples (B). A total of 8 individuals were removed based on this criterion, 6 of which were control samples and 2 were HESN samples (C).

3.2.3 Mass spectrometry analysis of rectal lavage fluid from HESN MSM

3.2.3.1 Longitudinal assessment of rectal proteome data in HESN individuals

Rectal lavage sample collection from HESN participants was conducted upon multiple clinic visits (visits 1 and 5; 6 month periods between each visit, 2 years total) for 11 of the 25 individuals. In order to assess protein changes over time, differential expression analysis was conducted on rectal protein levels between sample visits. This analysis would determine if repeated protein measurements for an individual could be averaged or needed to be treated separately in downstream comparisons to rectal protein expression within a low-risk control cohort. HESN MSM showed minimal changes in protein expression over time. Only 6% of factors were differentially expressed at $p<0.05$, which is approximately the proportion we would expect due to random chance (Figure 9). This evidence suggests that there are negligible protein changes over time within our sample set, and the averaging of repeated samples is an appropriate estimate of protein expression in these individuals.

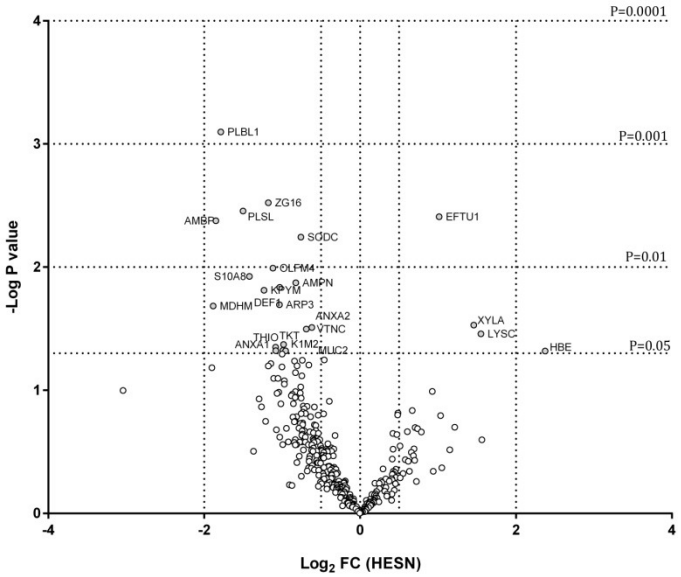


Figure 9: HESN MSM show minimal variation in protein expression over time. Duplicate rectal lavage samples from separate clinic visits (visit 1 and visit 5) for 11 of the 23 HESN MSM analyzed by mass spectrometry. Differential expression analysis was performed to determine if there were significant differences in protein expression over time within participants. Paired analysis identified no significant changes in protein expression within samples collected at different time points (two-tailed paired t tests, $p<0.05$ Benjamini-Hochberg corrected for multiple comparisons), and only 23/341 factors (6%) were significant at $p<0.05$.

3.2.3.2 Hierarchical clustering analysis of global protein expression in HESN MSM and controls in rectal lavage

Label-free mass spectrometry analysis detected 341 human and bacterial proteins in the rectal lavage fluid of HESN MSM and Control men. Unsupervised, global hierarchical clustering was used to visualize differences between HESN and healthy control men based on similarity in protein expression, and validated unique protein expression within rectal mucosal fluid samples between the two groups. Cluster analysis was able to accurately group HESN from control men with the exception of HESN H5, H16 and H25, which clustered with the control men, and the control sample, C20, that clustered with the HESNs (Figure 10). Unbiased hierarchical clustering suggests that the HESN individuals have a unique proteomic expression profile from control individuals.

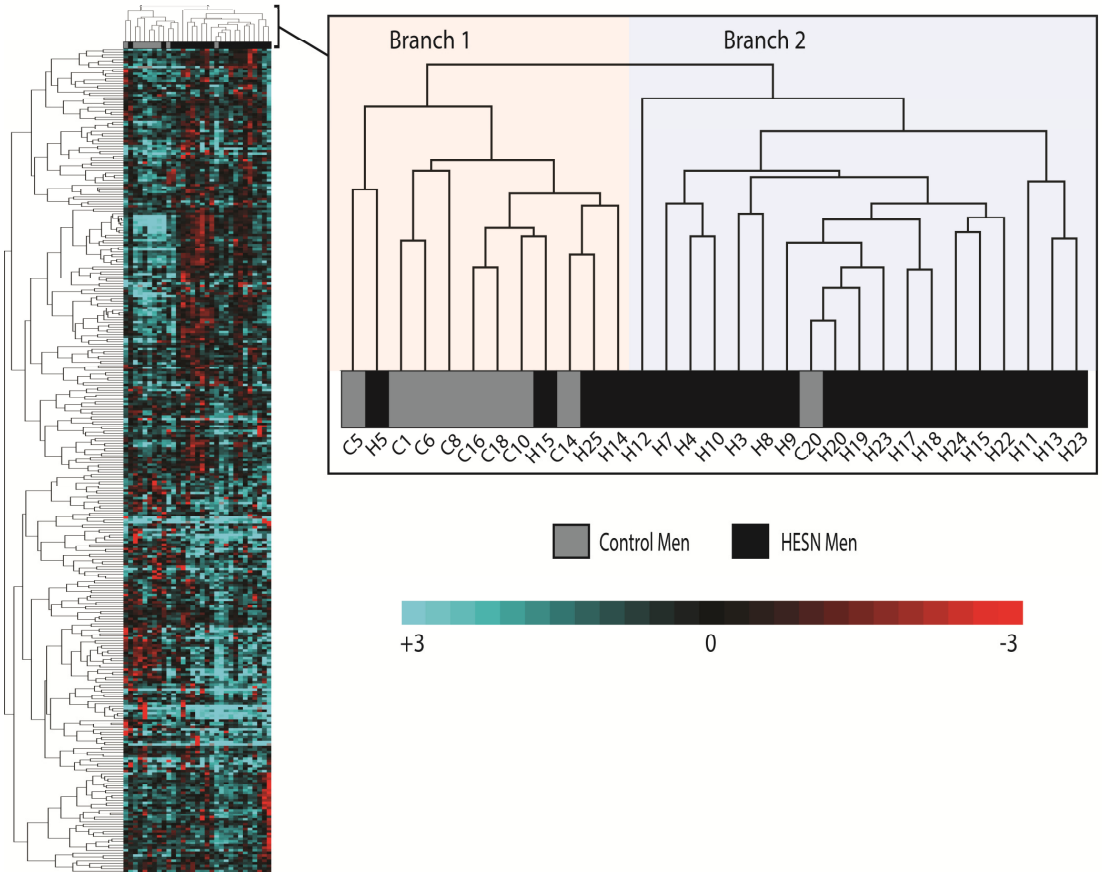


Figure 10. Global hierarchical cluster analysis of protein abundance in rectal mucosal fluid of HESN MSM relative to non-exposed control population. Protein expression patterns differentiate HESN MSM from control individuals. Proteins with a positive $\log_2$ fold-difference abundance value relative to mean protein expression across samples are coloured in red, negative expression values are in blue within the dendrogram matrix. Hierarchical clustering using centroid linkage clustering (Pearson's correlation as a distance metric) was able to accurately cluster HESN from control men with the exception of 4 HESN and 1 Control sample which did not distinctly cluster with their own phenotype. The heat map illustrates 341 proteins detected and quantified by mass spectrometry within the rectal lavage fluid of HESN MSM and control men.

3.2.3.3 Diversity of biofunctions held by proteins detected within rectal mucosal fluid

To understand the functional diversity of rectal fluid proteins identified across samples, functional annotation was performed using DAVID bioinformatics software. Twelve broad categories of biological functions spanning 68 functional categories were identified, and included immune system processes (innate immunity and inflammation, adaptive immunity, wound healing, phagocytosis, reactive oxygen species (ROS) response, and cell migration), cellular processes (cell metabolism, cytoskeletal processes, cell death), functions in tissue development (organ/tissue development, barrier development), and homeostatic processes (Figure 11). Innate immunity and inflammation, barrier development, cell metabolism and homeostatic processes held the largest proportion of unique sub-functions. This analysis allowed for a macroscopic view of protein expression within rectal lavage across samples, and suggests the soluble rectal proteome is capable of a wide range of functions in gastrointestinal maintenance and immunity that contribute to the immunobiology of the rectal mucosa.

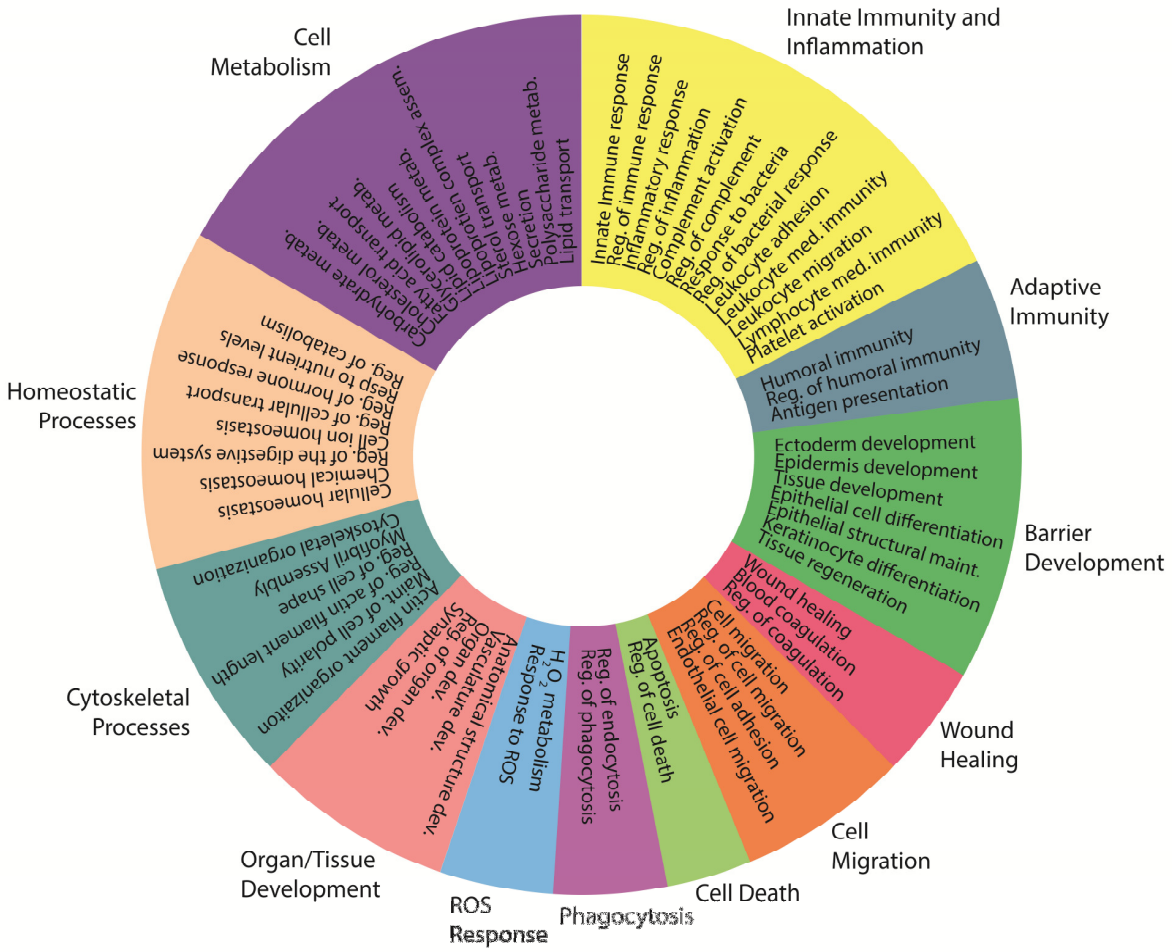


Figure 11. Rectal mucosal fluid proteins hold a diverse range biofunctions in cellular processes and immunity. Proteins identified by mass spectrometry within rectal lavage samples taken from low-risk and HESN men were analyzed using the gene ontology database DAVID biofunctional analysis (v6.7, Gene Ontology database, Goterm_bp_3). Biofunctions within the protein dataset ($p<0.05$) are represented above. A total of 68 gene ontology categories were identified, which could be grouped into 12 broader categories in organ development, cellular growth/metabolism, immune system processes and maintenance of gut homeostasis.

3.2.3.4 Differential expression analysis of protein levels in HESN MSM relative to controls

Proteomic analysis of rectal secretions from cohort of HESN MSM was used to determine soluble immune factors that may be differentially abundant compared to non-HIV exposed controls. Our proteomic analysis identified 341 proteins present in all samples. Of these, 25 (7.3%) proteins were of plant or non-human animal origin and had no human homolog in the UniProt database; as well, 3 trypsin factors were detected across samples and were subsequently removed as trypsin was exogenously added during enzymatic protein digestion. Of the 313 remaining proteins, the large majority (82%) were human, while the remainder were of bacterial or yeast origin (18%). Bacterial, yeast and human proteins were considered to be part of the innate mucosal environment and were selected for further analysis.

Differential expression analysis using two-tailed, independent t tests with Benjimani-Hochberg correction to determine that HESN MSM uniquely express immune factors within their rectal mucosal fluid. One hundred and three proteins (30%) were differentially expressed at a significance threshold of $p<0.05$ and 30 proteins (9%) were significant after multiple hypothesis testing correction where significant proteins had a p value of less than 0.0026 (Benjamini-Hochberg adjusted significance threshold, q) (Figure 12A). Rectal lavage from HESN MSM contained lower levels of bacterial proteins, and had relatively high levels of immunoglobulin, antiprotease and innate defense factors. A mild skewing towards overabundant immune factors in HESN samples was observed, which deviates from the expected observation of equivalent over- and under-abundant protein expression that accompanies immune pathway stimulation. The skewing may be an artifact of sampling bias between HESN and control populations that would not have been corrected during normalization to total protein content. However, despite this mild abnormality, differential expression analysis suggests a strong difference in protein expression in HESN MSM relative to controls, and this agrees with previous hierarchical clustering results.

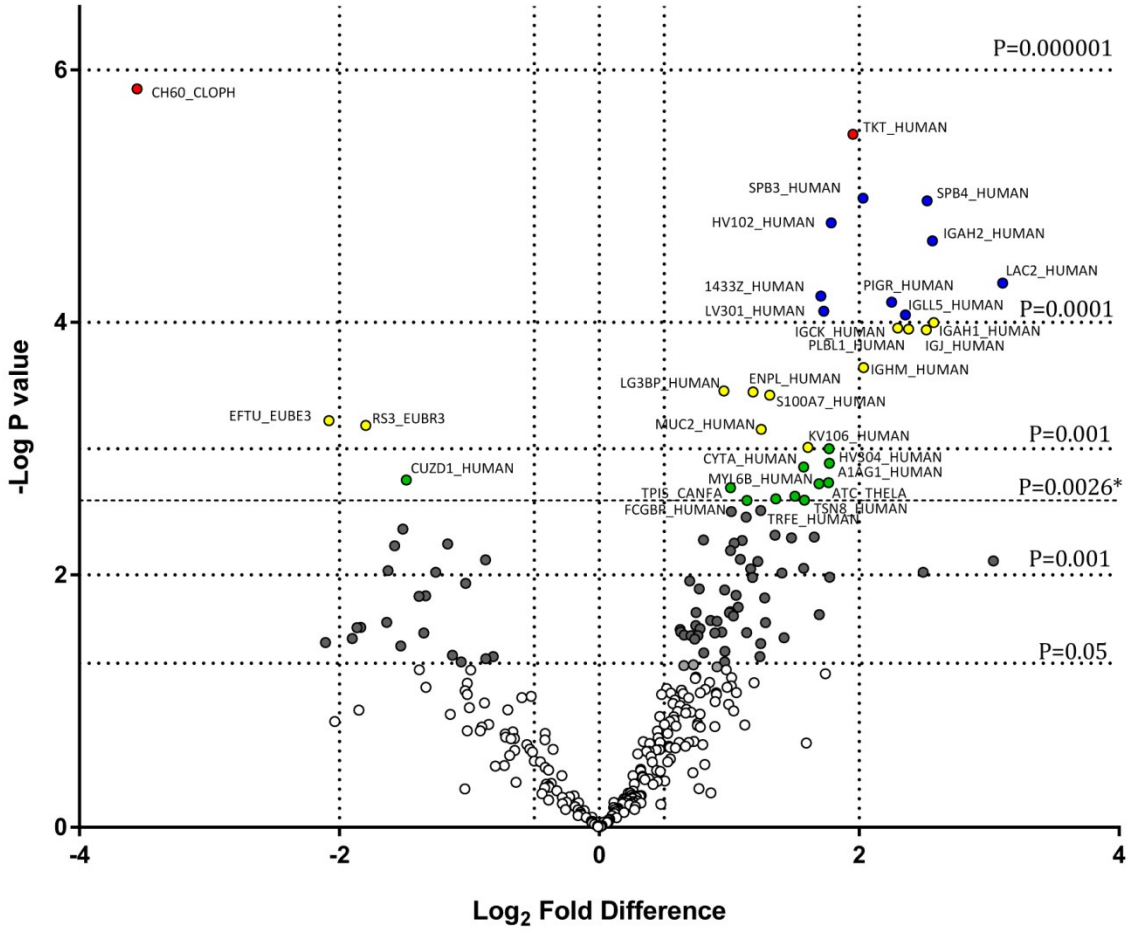


Figure 12: HESN MSM uniquely express immune factors within their rectal mucosal fluid. Differential expression analysis of rectal proteins identified by MS was performed using multiple two-tailed t tests ($p<0.05$, Benjamini-Hochberg corrected) and was visualized using volcano plots. Thirty proteins (9%) were significantly differentially abundant ($p<0.0026$*).

The two most significantly differentially expressed proteins were a 60 kDa chaperonin protein from *Clostridium phytofermentans* (-3.56 $\log_2$ fold-difference [L2FD], $p=1.42x10^{-6}$) and the human protein, transkelotase (1.95 L2FD, $p=3.25x10^{-6}$). The majority of statistically significant proteins were overabundant immunoglobulins and variable chains (12/30, 40%), which included mucosal immunoglobulins IgA1 (2.57 L2FD, p=0.001), IgA2 (2.56 L2FD, $p=2.26x10^{-5-}$), IgJ (2.52 L2FD, p=0.0001) and IgM (2.03 L2FD, p=0.0002). As well, several serine and cysteine protease inhibitors were overabundant in HESNs. The highly homologous serpin B3 (2.02 L2FD, $p=1.04x10^{-5}$) and serpin B4 (2.52 L2FD, $p=1.09x10^{-5}$) were both within the top 2% of significantly different proteins, while cystatin A (1.57 L2FD, p=0.001) was within the top 10% of significant proteins. Notably, several innate defense factors were also overabundant in HESNs: 14-3-3 protein zelta/delta (1.71 L2FD, $p=6.17x10^{5}$), phospholipase B-like 1 (2.38 L2FD, p=0.0001), S100A7 protein (1.31 L2FD, p=0.0004), mucin 2 (1.25 L2FD, p=0.0007) and the acute phase protein, alpha-1-acid-glycoprotein 1 (1.77 L2FD, p=0.001) (Table 4). Differential expression analysis was able to identify several immune proteins associated with the HESN phenotype; however, analyzing correlates of reduced susceptibility on a factor-by-factor basis does not elucidate how these proteins are working together within the biological system of the rectal compartment. Functional enrichment and pathway analysis were performed on significantly different proteins to gain insights into how these factors may be working together to facilitate an immune response in HESN individuals.

Table 4: Statistical values for differentially expressed proteins in rectal fluid samples from HESN MSM.

Accession	*Protein Name*	*P value*	*Log_2 FD*	*Stdev*
CH60_CLOPH	60 kDa chaperonin (Clostridium phytofermentans)	1.42E-06	-3.56	0.59
TKT_HUMAN	Transketolase	3.25E-06	1.95	0.34
SPB3_HUMAN	Serpin B3	1.04E-05	2.03	0.38
SPB4_HUMAN	Serpin B4	1.09E-05	2.52	0.48
HV102_HUMAN	Ig heavy chain V-I region HG3	1.63E-05	1.79	0.35
IGHA2_HUMAN	Ig alpha-2 chain C region	2.26E-05	2.56	0.51
LAC2_HUMAN	Ig lambda-2 chain C regions	4.88E-05	3.10	0.65
1433Z_HUMAN	14-3-3 protein zeta/delta	6.18E-05	1.71	0.36
PIGR_HUMAN	Polymeric immunoglobulin receptor	6.92E-05	2.25	0.49
LV301_HUMAN	Ig lambda chain V-III region SH	8.13E-05	1.73	0.38
IGLL5_HUMAN	Immunoglobulin lambda-like polypeptide 5	8.73E-05	2.36	0.52
IGHA1_HUMAN	Ig alpha-1 chain C region	1.00E-04	2.57	0.57
IGKC_HUMAN	Ig kappa chain C region	1.11E-04	2.30	0.51
PLBL1_HUMAN	Phospholipase B-like 1	1.13E-04	2.38	0.53
IGJ_HUMAN	Immunoglobulin J chain	1.15E-04	2.52	0.56
IGHM_HUMAN	Ig mu chain C region	2.28E-04	2.03	0.48
LG3BP_HUMAN	Galectin-3-binding protein	3.50E-04	0.96	0.24
ENPL_HUMAN	Endoplasmin	3.56E-04	1.18	0.29
S10A7_HUMAN	Protein S100-A7	3.77E-04	1.31	0.33
EFTU_EUBE2	Elongation factor Tu (*Eubacterium eligens*)	6.00E-04	-2.08	0.54
RS3_EUBR3	30S ribosomal protein S3 (E*ubacterium rectale*)	6.56E-04	-1.80	0.47
MUC2_HUMAN	Mucin-2	7.06E-04	1.25	0.33
KV106_HUMAN	Ig kappa chain V-I region EU	9.76E-04	1.61	0.44
HV304_HUMAN	Ig heavy chain V-III region TIL	1.00E-03	1.77	0.48
A1AG1_HUMAN	Alpha-1-acid glycoprotein 1	1.31E-03	1.77	0.50
CYTA_HUMAN	Cystatin-A	1.40E-03	1.57	0.45
CUZD1_HUMAN	CUB and zona pellucida-like domain-containing protein 1	1.77E-03	-1.49	0.43
ACT_THELA	Actin (*Thermomyces lanuginosus*)	1.86E-03	1.76	0.52
MYL6B_HUMAN	Myosin light chain 6B	1.90E-03	1.69	0.49
TPIS_HUMAN	Triosephosphate isomerase	2.04E-03	1.01	0.30

Factors passing significance criteria for differential expression are listed and colour coded based on degree of significance: red ($p<1x10^{-6}$), blue ($p<1x10^{-5}$), yellow ($p<1x10^{-4}$) and green ($p<1x10^{-3}$).

3.2.3.5 Functional annotation of differentially expressed protein in HESN MSM

Proteins that are uniquely expressed within rectal secretions from HESN MSM may be co-expressed in immune response pathways. To elucidate relationships between proteins in our dataset, differentially expressed proteins were subject to a functional enrichment analysis. A wider signature threshold was applied for functional analysis to be more inclusive of other factors, in order to gain a clearer insight of activated biological processes within this dataset. A total of 103 proteins differentially abundant between HESN MSM and controls ($p<0.05$) were screened using the DAVID Bioinformatics tool. Biofunctions defined as enriched within the dataset met a threshold of greater than five overlapping proteins and had a Benjamini-Hochberg adjusted p value of less than 0.05.

Eight biofunctions were enriched within the dataset, and contained overabundant proteins in HESN MSM that held functions in immune system processes, epithelial barrier development and cell death (Figure 13). The top two significant functions were related to immune system process ($p=6.68x10^{-6}$) and immune response ($p=1.06x10^{-5}$). This was largely due to the heightened expression of immunoglobulins described previously. Several other immune proteins were highlighted by this analysis: the antiprotease, serpin B4, was enriched along with inflammatory mediators thioredoxin (THIO), peroxiredoxin 1 (PRDX1), and matrix metalloprotease 9 (MMP9). Antimicrobial (AMPs) S100 proteins (S10A7 and S10A9) were also elevated. AMPs and inflammatory factors aid in host defence against bacteria/viruses and promote inflammation, while antiproteases regulate host cell proteases activity to limit inflammation to preserve host tissue[238]. Apoptotic proteins (regulation of apoptosis, $p=2.27x10^{-4}$; regulation of programmed cell death, $p=2.49x10^{-4}$; regulation of cell death, $p=2.58x10^{-4}$) and epidermal development factors were also significantly overabundant in the HESN group. Factors such as keratins (K1C9), small proline rich proteins (SPR1A and SPR2B) and cell cycle regulators (CASPE and CYTA) had increased abundance in lavage samples. These factors are known structural components of the epithelium[269-271], and some (CYTA) can maintain tissue integrity and facilitate repair[240]; however, the significance of these epidermal factors within mucosal fluid is unclear. Functional enrichment analysis depicted an immune

environment with immunoglobulin factors, pro-inflammatory proteins and antimicrobial defense proteins, and epithelial barrier proteins in HESN MSM.

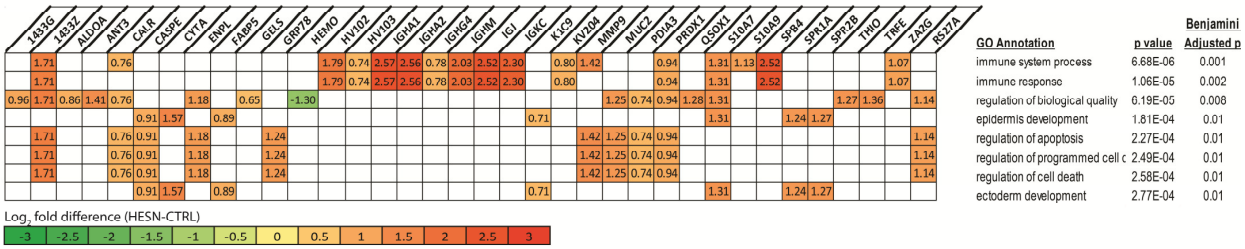


Figure 13: HESN MSM have uniquely express of immune, apoptosis and epithelial development factors. Functional enrichment of proteins was performed with DAVID Bioinformatics software using the Gene Ontology database (Goterm_bp_5) for annotation of proteins. Eight enriched functions in HESNs are displayed, with proteins associating with each function colour coded according to their $\log_2$ fold difference value in HESNs relative to controls.

3.2.3.6 Differences in microbial protein expression in HESN MSM relative to non-exposed controls

The effect of the microbiome on the immune and homeostatic state of mucosal surfaces is an increasing field of interest for HIV acquisition, as the composition of bacterial species within the gut has been linked to overall gastrointestinal health and immune function[272-274]. This proteomic analysis was able to detect 55 bacterial proteins within rectal lavage fluid samples. The most represented bacteria genera in the dataset were *Eubacterium* (11 proteins), *Bacteroides* (8 proteins), *Clostiridium* (7 proteins), *Bifidobacteria* (6 proteins) and yeast from genus *Saccharomyces* (5 proteins). Single protein detections occurred for 18 other genera, including the commensal, *Lactobiacillus planarum*, and potentially pathogenic species *Escherichia coli O157:H7* and *Helobacter hepaticus*, a less virulent member of the Helobacter genus that is able to cause inflammatory diseases in the gut[275,276]. Since there is a high diversity of bacteria in the gut, our coverage of approximately 20 bacterial genera likely did not see an accurate representation of the gut flora through proteomic method, and any trends in this data are limited to these species.

Mild differences in microbial protein expression were observed in the between study groups, with only one protein significant after multiple comparison corrections (Figure 14A). The 60 kDa chaperonin protein from *Clostiridium phytofermentans* was significantly underabundant in HESN MSM at a false discovery rate (FDR) of 5% (-3.56 L2FD, $p=1.41x10^{-6}$). Two Eubacterium proteins were also notably underabundant in HESN rectal samples: elongation factor tu protein (*E. elegans*; -2.07 L2FD, $p=6.00x10^{-4}$) and 30S ribosomal protein S3 (*E. rectale*; -1.80 L2FD, $p=6.56x10^{-4}$). When significance threshold was relaxed to include all proteins significant at $p<0.05$, 23 proteins (42%) were differentially abundant. The majority of under-abundant proteins in HESNs were anaerobic species that are normally present with in the gut flora, while two over-abundant proteins from the *Saccharomyces* yeast genus were overabundant in HESN samples (Figure 14B). Clustering of samples based on proteins differentially abundant at a relaxed threshold of $p<0.05$ showed almost complete separation between study groups, with two HESN samples (14H and 17H) clustering with the control branch, and two controls (20C and 14C) grouping with

the HESN (Figure 14B). This preliminary investigation of microbial protein expression within this HESN population highlights several bacterial and yeast species that contribute soluble protein factors into rectal mucosal fluid, and suggests mild differences in microbial protein expression within this group, which has the potential to influence the mucosal environment.

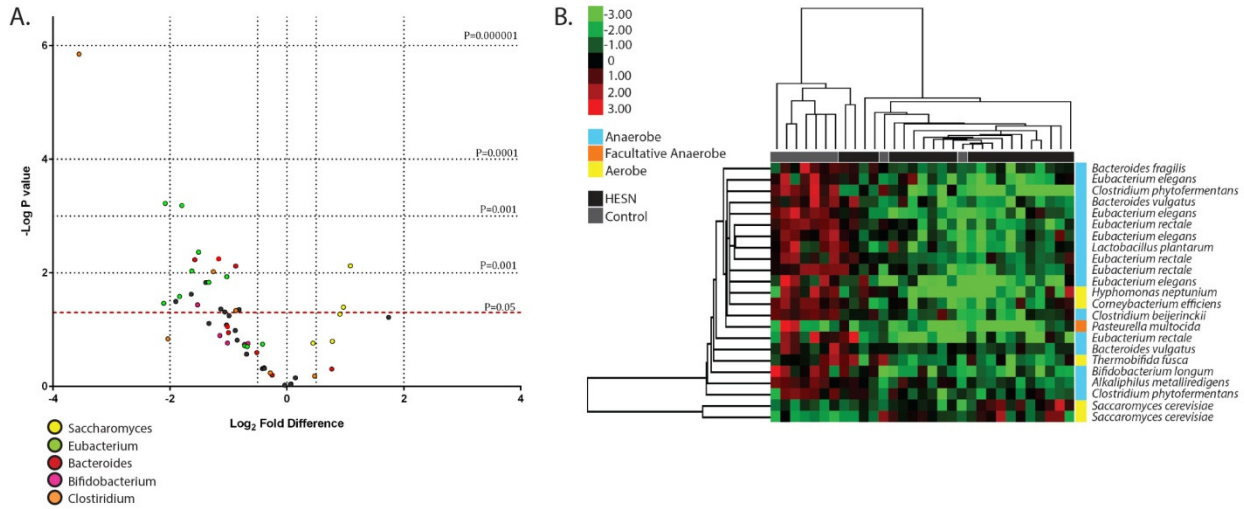


Figure 14: Differential expression analysis of bacterial proteins in rectal samples from HESN MSM. Mass spectrometry detected 55 unique bacterial and yeast proteins across mucosal samples. Differential expression of these proteins was identified by two-tailed t test ($p<0.05$) between HESN and control individuals. Two-tailed t tests were used to determine differentially abundant bacterial proteins between HESN and control samples. Statistical results are displayed in a volcano plot; only one protein (60 kDa chaperonin, *Clostiridium phytofermentans*) passed Benjimani-Hochberg multiple comparison analysis (A). The 23 proteins that trended towards significance ($p<0.05$, unadjusted) were analyzed with hierarchical cluster analysis using complete linkage with Pearson's R as the distance coefficient. Hierarchical clustering showed distinct clustering of HESN and control samples based on bacterial protein expression with the exception of two HESN (14H and 17H) and two control (20C and 14C) samples. The majority of under-abundant proteins at $p<0.05$ were anaerobic bacteria species, with two over-abundant *Saccharomyces* proteins (B).

3.2.3.7 Relationship of protein expression in HESN MSM with sexual behaviour and indicators of HIV exposure

Unprotected intercourse and subsequent exposure to HIV, semen and/or bacterial antigens may have effects upon mucosal surfaces. Semen contains an abundance of immunomodulatory proteins, including cytokines, chemokines, proteases, prostaglandins and immunoglobulins[277], and exposure of seminal plasma and commensal bacteria is able to induce pro-inflammatory signaling within the genital epithelium of the FRT, enhancing immune cell recruitment and a favorable environment for infection[278]. Sexual activity has been associated with micro-abrasions and mucosal damage that can facilitate viral penetration[105], and may stimulate inflammation and tissue remodeling pathways. Further, one of the greatest factors to promote HIV acquisition is viral load of the HIV positive partner. Currently, the effects of semen, frequency of sexual intercourse, and viral load on the rectal proteome have not been fully characterized, and may impact our interpretation of unique protein expression within rectal mucosal samples from HESN MSM.

We sought to determine if any proteins factors associated with the HESN population were related to an increase in exposure, and not a true biomarker of this phenotype. Information on sexual behavior was collected from HESN study participants by the Venhälsan clinic at the time of sampling and is described in section 4.2.1. Two tailed t tests ($p<0.05$) showed that 23 proteins were differentially expressed between HESN MSM practicing unprotected oral intercourse often/always (n=13) and sometimes/seldom (n=9); two individuals did not respond to the question (Figure 15A). Comparisons to show proteome effects due to receptive anal (n=3 total) and insertive anal intercourse (n=5 total) were underpowered and could not be assessed. Additionally, 29 proteins were differentially expressed between HESN MSM who had HIV-neutralizing IgA in their rectal lavage (n=7) and those that did not (n=18) (Figure 15A). The most dramatic proteome effects were observed when protein expression was correlated to measured partner viral load; 48 proteins showed a trending correlation ($p<0.05$, Pearson's correlation) with partner viral load >50 (n=10). The expression of two factors appeared to be affected by all collected variables: 14-3-3 zelta/delta (1433Z) and merpin 1, subunit alpha (MEP1A) (Figure 15A). The 1433Z

protein is a multifunctional signaling molecule, which influences functions ranging from cell development to innate immunity, while the biological influence of MEP1A is less known, this metalloendopeptidase largely functions in protein processing and cleavage. Functional annotation was performed to understand biological processes relating to proteins that may be affected by clinical variables.

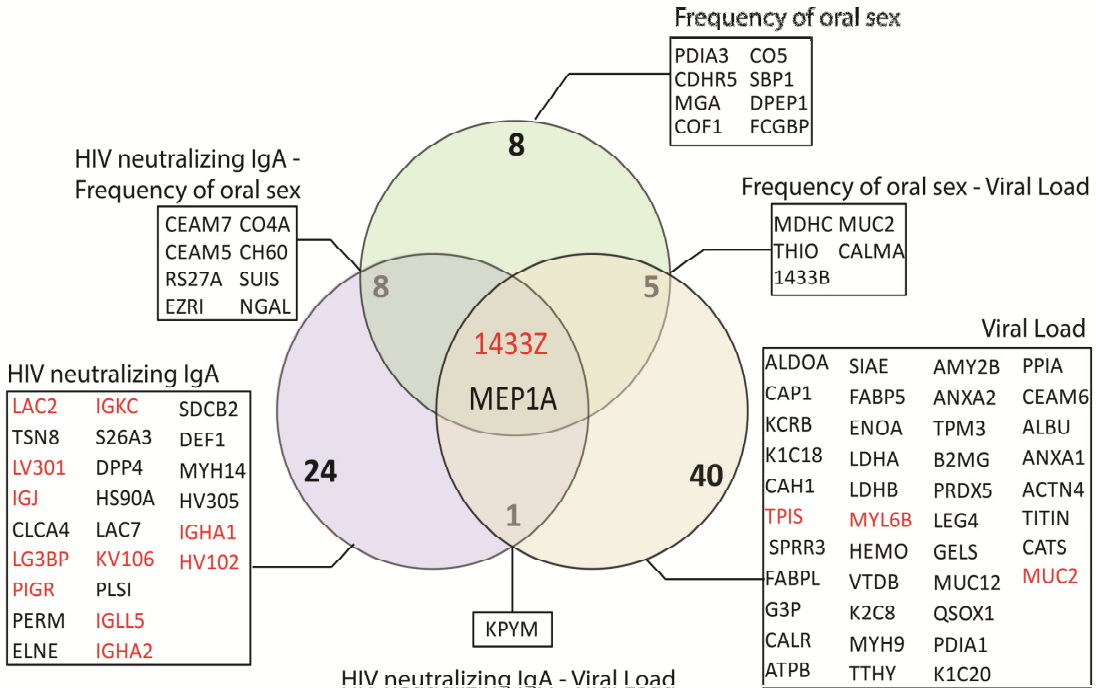


Figure 15: Relationships between rectal protein expression and clinical variables. Frequency of unprotected oral intercourse collected via survey, viral load (VL) of HIV+ partner (0-12 months prior to sampling) was measured, and the presence of HIV-neutralizing IgA in the HESN rectal mucosal sample was assayed by collaborators at the Karolinska Institutet. Two tailed t tests ($p<0.05$) showed that 23 proteins were differentially expressed between HESN MSM practicing unprotected oral intercourse often/always and sometimes/seldom. Further, 29 proteins were differentially expressed between HESN MSM who had HIV-neutralizing IgA in their rectal lavage and those that did not. Further, 48 proteins showed a trending correlation ($p<0.05$, Pearson's correlation) with partner viral load >50 (n=9). No factors passed significance when corrected for multiple comparisons. Factors highlighted in red overlap with proteins significantly differentially expressed ($p<0.0026$) in HESN samples.

Proteins that correlated with epidemiological variables were screened for functional enrichment using the gene ontology database (DAVID software). The top 3 annotations for each variable with an adjusted p value less than 0.05 (Benjamini-Hochberg correction) are shown (Table 5). Proteins that correlated with partner viral load held functions in glycolysis energy metabolism ($p=1.38x10^{-10}$). This may signify the

expansion of local immune cells within mucosal tissues upon viral exposure, which has been previously observed during early HIV infection events[279]; however, the degree of activation, if any, could not be determined from this analysis. Proteins associated with HIV-neutralizing IgA in rectal secretions were held functions in with immune system process (1.78×10^{-3}), which was largely attributed to an increased expression of total IgA and non-IgA (IgM, IgG, variant chains) immunoglobulins in individuals with a neutralizing response. This is intuitive and likely due to multiple B cell stimulation events upon HIV or other pathogen / microbial exposures within these individuals during unprotected intercourse. Interestingly, non-immunoglobulin proteins were associated with HIV-neutralizing IgA activity, including neutrophil elastase (ELNE), dipeptidyl peptidase (DPP4), 1433Z (general regulator, apoptosis, platelet activation), DEF1 (defensin) and CO4A (complement 4a). This may suggest that other IgA contributes to immune pathways outside of simple HIV-neutralization, and may have alternate mechanisms of influence on HIV susceptibility; as IgA has known functions in regulation of complement and inflammatory mediators, it is possible that the IgA response contributes to immunity in addition its bind-and-clear mechanism[280]. Finally, frequency of unprotected oral intercourse had significant association with apoptotic proteins rectal lavage samples ($p=7.30 \times 10^{-6}$). Several factors were linked to apoptotic signaling are cell cycle and proliferation factors (1433 proteins) and immune proteins (mucins and thioreductase), which many be a redundant observation of the increased cell proliferation, or immune response to exposure relating to intercourse.

Table 5: Top biofunctions associated with rectal fluid proteins related to clinical variables.

Correlation	*GO Annotation*	*Adjusted P value*	*Protein (Gene ID)*
Partner Viral Load	❖ glycolysis	1.38E-10	ALDOA KPYM G3P TPIS MDHC LDHB LDHA ENOA
	❖ generation of precursor metabolites and energy	6.19E-07	ALDOA KPYM G3P TPIS MDHC THIO LDHB LDHA ATPB ATPB ENOA
	❖ regulation of biological quality	8.30E-06	ALDOA QSOX1 1433Z GELS PDIA1 THIO ALBU HEMO MUC2 MUC2 CALR ATPB PRDX5 MYH9 KCRB 1433B ANXA2 ENOA
HIV-neutralizing IgA	❖ immune system process	1.78E-03	ELNE DPP4 1433Z IGHA1 HV102 IGHA2 IGJ DEF1 CO4A CO4A IGKC
	❖ immune response	5.32E-03	1433Z IGHA1 HV102 IGHA2 IGJ DEF1 CO4A IGKC
	❖ response to stimulus	1.71E-02	1433Z IGHA1 DPP4 PERM IGJ IGHA2 DEF1 LG3BP IGKC IGKC ELNE HS90A SUIS NGAL HV102 CO4A
Frequency of Exposure	❖ regulation of apoptosis	7.30E-06	COF1 1433Z 1433B PDIA3 MUC2 RS27A
	❖ regulation of programmed cell death	7.85E-06	COF1 1433Z 1433B PDIA3 MUC2 RS27A
	❖ regulation of biological quality	1.00E-05	COF1 1433Z 1433B THIO PDIA3 EZRI MUC2 RS27A

This sub-analysis identified several factors that are differentially expressed in HESN MSM relative to the control population, and may also be influenced by increased sexual activity or pathogen exposure in the HESNs that may not be occurring in the control group. Thus, this analysis allowed us to filter our list of HESN biomarkers to focus on proteins that are more likely to be associated with reduced susceptibility to HIV, rather than a consequence of confounding variables. In total, 15 of the original 30 identified factors were retained (Table 6). Notably, all three antiproteases (serpin B3, serpin B4, and cystatin A) were not associated with clinical variables, and were chosen for further analysis.

Table 6: Proteins associated with HESN MSM independent of clinical variables

Accession	*Protein Name*	*P value*	*Log2 FD*	*Stdev*
CH60_CLOPH	60 kDa chaperonin (Clostridium phytofermentans)	1.42E-06	-3.56	0.59
TKT_HUMAN	Transketolase	3.25E-06	1.95	0.34
SPB3_HUMAN	Serpin B3	1.04E-05	2.03	0.38
SPB4_HUMAN	Serpin B4	1.09E-05	2.52	0.48
PLBL1_HUMAN	Phospholipase B-like 1	1.13E-04	2.38	0.53
IGHM_HUMAN	Ig mu chain C region	2.28E-04	2.03	0.48
ENPL_HUMAN	Endoplasmin	3.56E-04	1.18	0.29
S10A7_HUMAN	Protein S100-A7	3.77E-04	1.31	0.33
EFTU_EUBE2	Elongation factor Tu (Eubacterium eligens)	6.00E-04	-2.08	0.54
RS3_EUBR3	30S ribosomal protein S3 (Eubacterium rectale)	6.56E-04	-1.80	0.47
HV304_HUMAN	Ig heavy chain V-III region TIL	1.00E-03	1.77	0.48
A1AG1_HUMAN	Alpha-1-acid glycoprotein 1	1.31E-03	1.77	0.50
CYTA_HUMAN	Cystatin-A	1.40E-03	1.57	0.45
CUZD1_HUMAN	CUB and zona pellucida-like domain-containing protein 1	1.77E-03	-1.49	0.43
KV113_HUMAN	Ig kappa chain V-I region Lay	2.39E-03	1.51	0.45

3.2.3.8 Serine and Cystatin antiproteases are highly abundant in rectal secretions and correlate with reduced susceptibility to HIV infection in the Venhälsan cohort of HESN MSM.

Proteomic investigation identified three antiproteases that were significantly elevated in rectal secretions of mucosal fluid from HESN MSM. Serpin B3, serpin B4 and cystatin A all showed a $\log_2$ fold increase greater than 1.3 to satisfy power requirements (>80% experimental power), and were highly significant after multiple comparison corrections (Figure 16A-C). Normal expression of the proteins of interest were confirmed using normal quantile plots, which validated the correct use of parametric statistics to define differentially expressed proteins between the two populations (Figure 16D-F). All three antiproteases have previously been identified as correlates of protection in our lab's previous proteomic investigations of FGT secretions from women of the Pumwani sex worker cohort[99,281]. Further, cystatin A was shown to significantly correlate with monokine induced by gamma-interferon (MIG), and serpin B4 correlated with IP-10 and IL-1Rα in this cohort, suggesting a role for these proteins in inflammation at mucosal surfaces[282]. Other antiproteases associated with resistance in the previous studies were not associated with rectal susceptibility in the Venhälsan cohort (Table 7). Serpin B1 and antithrombin III were trending at significance (p=0.01 and p=0.001, respectively), but did not pass significance thresholds for multiple hypothesis testing. The known antiviral serpin, alpha-1-antitrypsin, was not differentially expressed (p=0.5), nor were the previously identified correlates, A2ML1 (p=0.21) and serpin G1 (p=0.29). The differences in the antiproteases associated with the HESN phenotypes in this and the previous Pumwani cohort analysis may be due to differences in gender, the biological compartment analyzed or variables relating to exposure, as CSWs are exposed to multiple partners and may have a higher frequency of exposure, which would alter the mucosal immune response greatly. However, the finding of three antiproteases as associated with reduced HIV infection in two very distinct cohorts and biological compartments, as well as their largely undiscovered role in mucosal immunity, makes them attractive targets for further investigation into their expression in rectal mucosal fluid overall role in rectal immunity.

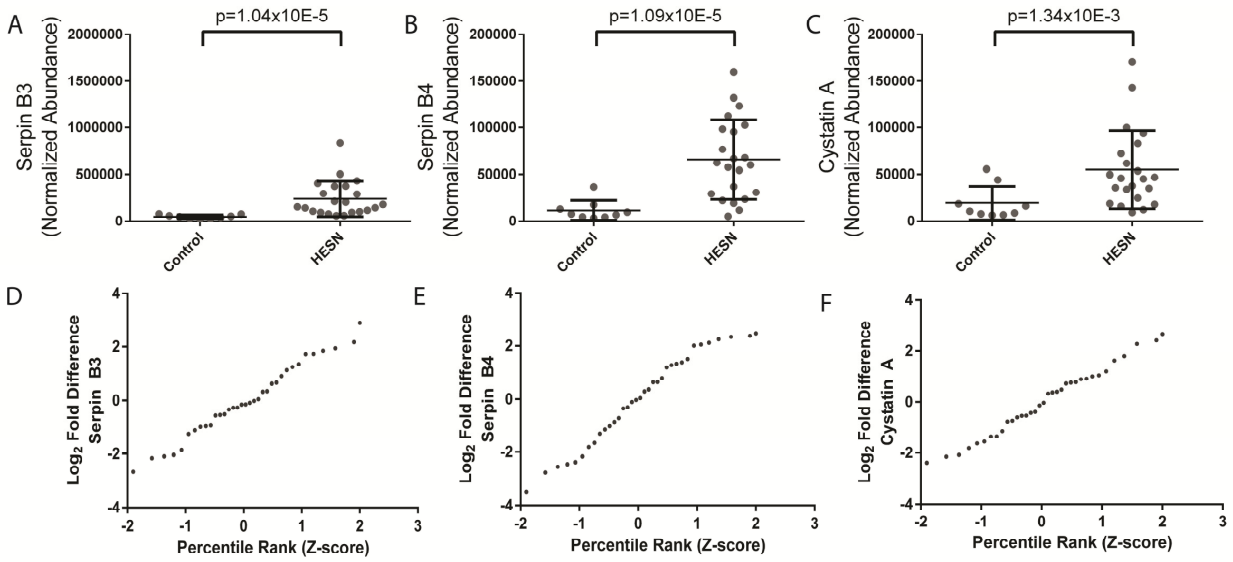


Figure 16: HIV-exposed seronegative men who have sex with men have elevated levels of antiproteases in their rectal mucosal secretions. Unbiased, label free proteomics was able to identify several antiproteases that were significantly differentially abundant in the rectal secretions of HESN MSM. Serpin B3 and serpin B4 had the most significant increased abundance in the HESN cohort (2.02 L2FD, $p=1.05x10^{-5}$ and 2.52 L2FD, $p=1.08x10^{-5}$, respectively) (A-B). The cysteine protease inhibitor, cystatin A, was also differentially expressed after multiple comparison correction (1.57 L2FD, $p=1.34x10^{-3}$) (C). Normal quantile (P-P) plots confirmed that expression of antiproteases was normally distributed in our dataset (D-F).

Table 7. Antiprotease expression in rectal secretions from HESN MSM of the Venhälsan cohort

							normalized log ratio			
Accession	*Antiprotease*	*Alternative name*	*Antiprotease activity*	*Significant (p<0.05)*	*FDR 5% significant*	*P value*	*Control*	*HESN*	*HESN-Control*	*SD of difference*
SPB3_HUMAN	Squamous cell carcinoma antigen 1 (SCCA1)	Serpin B3	serine, papain-like cysteine	•	•	1.039E-05	-2.03	0.00	2.03	0.38
SPB4_HUMAN	Squamous cell carcinoma antigen 2 (SCCA2)	Serpin B4	Serine	•	•	1.089E-05	-2.53	0.00	2.52	0.48
CYTA_HUMAN	Cystatin-A	Stefin-A	Cystiene	•	•	1.40E-03	-1.66	-0.09	1.57	0.45
ANT3_HUMAN	Antithrombin-III	Serpin C1	Serine	•		9.66E-03	-1.91	-0.51	1.41	0.51
ILEU_HUMAN	Leukocyte elastase inhibitor	Serpin B1	Serine	•		1.12E-02	-0.70	0.00	0.70	0.26
A2AP_HUMAN	Alpha-2-antiplasmin	Serpin F2	Serine	•		1.99E-02	-1.18	-0.18	1.00	0.41
ITIH1_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H1	serum-derived hyaluronan associated protein (SHAP)	Serine			7.24E-02	0.00	-1.01	-1.02	0.55
A2ML1_HUMAN	Alpha-2-macroglobulin-like protein 1 (A2ML1)	PZP-like alpha-2-macroglobulin domain-containing protein 9	Serine			2.12E-01	-0.74	-0.27	0.47	0.37
IC1_HUMAN	Plasma protease C1 inhibitor	Serpin G1	Serine			2.98E-01	-0.13	-0.63	-0.50	0.47
AACT_HUMAN	Alpha-1-antichymotrypsin	Serpin A3	Serine			3.88E-01	-0.03	-0.32	-0.29	0.33
ITIH4_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H4	Plasma kallikrein sensitive glycoprotein 120	Serine			4.12E-01	-0.77	-0.39	0.38	0.46
ITIH2_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H2	serum-derived hyaluronan associated protein (SHAP)	Serine			4.29E-01	-1.34	-0.83	0.50	0.63
A1AT_HUMAN	Alpha-1-antitrypsin	Serpin A1	Serine			5.81E-01	-0.91	-0.62	0.29	0.51
SPB5_HUMAN	Maspin	Serpin B5	Serine			7.37E-01	-0.17	-0.28	-0.12	0.34
SPB6_HUMAN	Placental thrombin inhibitor (PTI)	Serpin B6	Serine			9.90E-01	-0.33	-0.33	0.01	0.39

3.2.4 Antiprotease expression in the MACS validation cohort of HESN MSM

Proteomic investigation of a separate cohort of HESN MSM was used to validate our finding of unique antiprotease expression in the Venhälsan population. The MACS cohort consists of American MSM who have demonstrated resistance to rectal acquisition of HIV despite high levels of exposure with multiple HIV positive partners in the early 1980's during the peak of the epidemic[252]. These mechanisms of protection may still be present in the absence of high exposure, and may be visualized via proteomic analysis of their mucosal secretions. Though there are many differences in exposure and other epidemiological variables between the cohorts, unique expression of antiproteases in the rectal secretions of this cohort would provide further supporting evidence for a role for these proteins in rectal susceptibility.

The proteomic investigation of rectal secretions from HESN MSM from the MACS cohort was performed on sponge samples from 14 control MSM and four HESN MSM. As only four HESNs samples were collected at the time of investigation (2014), this analysis is restricted to a preliminary investigation of this cohort. Of note, the HESN men in this cohort are slightly older than the controls (median age HESN age 79, median control age 57, difference of 22 years); however, the statistical significance of the age difference could not be assessed due to lack of power. Mass spectrometry identified 599 proteins that passed covariance (CV) filtration (<25% CV between technical replicates). The heightened number of protein identifications (a 55% increase) in the MACS cohort analysis relative to Venhälsan was likely due to the method sample collection and the mass spectrometer used. Rectal mucosal fluid was collected via sponge in the MACS cohort, which would yield a more protein-dense sample than lavage, which was used to collect samples in the Venhälsan group. The higher yield of mucosal proteins in each sample would contribute to an increased number of identifications. Further, the Q Exactive mass spectrometer used for this analysis is an advanced instrument that allows for higher resolution of peptides, and greater coverage of the rectal mucosal proteome. Differential expression analysis of normalized values was performed using two-tailed, independent t tests ($p<0.05$, Benjamini-Hochberg multiple comparison corrected). The above dendrogram displays clustering of samples based on similarity in protein

expression, and identifies two main branches of samples (Figure 17A). Within “Branch 2”, HESN men (black) clustered together, but could not be distinguished from several control individuals that clustered within the same branch (grey). A volcano plot displays transformed p values and fold difference values for each protein. Weak effects were seen between the two populations as only 25 proteins (4% of all identifications) were significant at $p<0.05$ (navy blue) with none passing multiple hypothesis correction (Figure 17B).

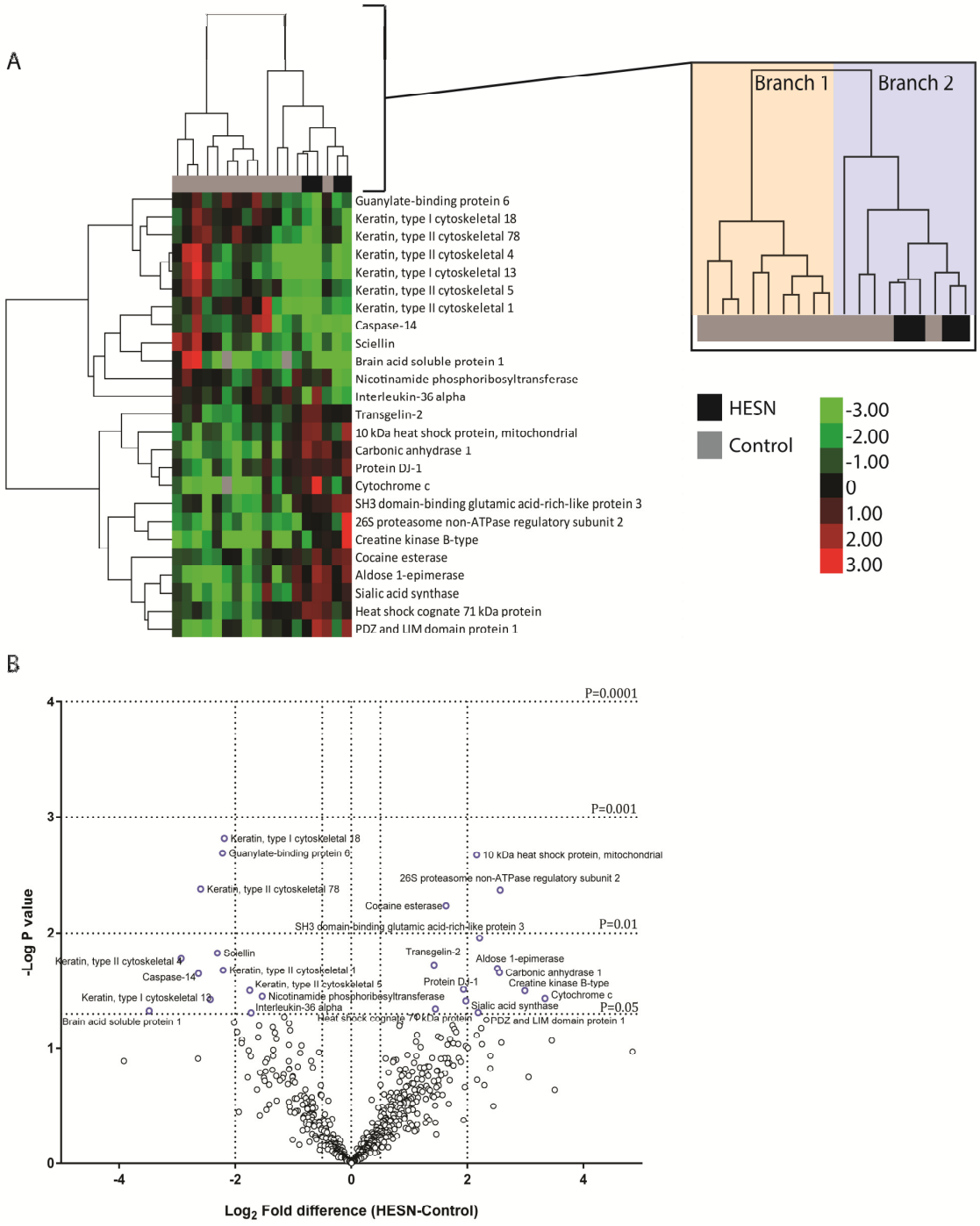


Figure 17: Antiproteases were not differentially expressed in a preliminary investigation of HIV resistant MSM from the MACS cohort. A preliminary label-free mass spectrometry analysis of rectal secretions from HESN (n=4) and control (n=14) men from the MACS cohort was performed to determine if antiproteases were uniquely expressed in the HIV resistant group. A heatmap illustrates rectal fluid proteins which were differentially abundant ($p<0.05$), between HESN and control MSM, but protein expression was not distinct between the two groups. Clustering of samples was generated by unsupervised complete linkage hierarchical clustering using Pearson correlation coefficient as the distance metric (A). Volcano plot illustrates differentially expressed proteins between groups ($p<0.05$, blue) (B).

A sub analysis focused on antiprotease expression within this cohort identified 20 antiproteases across rectal samples, but none were differentially expressed between groups (Table 8). Our finding that no proteins were significantly differentially expressed, and that there was poor clustering between the two populations does not disprove our findings in the Venhälsan populations, but it rather provides no supporting evidence to our observation. A priori power calculations based on this pilot study (599 proteins and an average experimental error estimate of 99.1% covariance in expression between proteins) estimates a required samples size of 60 HESN and 60 control men to see fold differences of two between groups and retain 80% experimental power[283]; as there are currently 80 active members of the MACS cohort, adequately powered proteomic studies of this cohort are feasible. Additional, high powered studies of the MACS cohort would be needed to fully investigate the unique responses within this group and validate antiproteases in reduced HIV susceptibility in the rectal mucosa.

Table 8. Expression of antiproteases in rectal secretions of HESN MSM from the MACS cohort

				normalized log ratio			
Accession	*Antiprotease*	*Alternative name*	*P value*	*HESN*	*Control*	*HESN-Control*	*SD of difference*
A2ML1_HUMAN	Alpha-2-macroglobulin-like protein 1	PZP-like alpha-2-macroglobulin domain-containing protein 9	0.07	-1.17	-0.12	-1.05	0.54
CYTC_HUMAN	Cystatin-3	Cystatin-C	0.09	0.68	-0.97	1.66	0.93
ILEU_HUMAN	Leukocyte elastase inhibitor	Serpin B1	0.13	0.33	-0.44	0.77	0.48
IC1_HUMAN	Plasma protease C1 inhibitor	Serpin G1	0.16	0.07	-1.40	1.47	1.00
SPB4_HUMAN	Squamous cell carcinoma antigen 2 (SCCA2)	Serpin B4	0.20	-1.81	-0.71	-1.09	0.83
SPB13_HUMAN	Hurpin	Serpin B13	0.25	-0.84	-2.41	1.57	1.32
A1AT_HUMAN	Alpha-1-antitrypsin	Serpin A1	0.29	-1.09	-0.11	0.98	0.90
CYTD_HUMAN	Cystatin-5	Cystatin-D	0.31	-1.41	-0.57	-0.84	0.79
ITIH2_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H2	serum-derived hyaluronan associated protein (SHAP)	0.36	-1.08	-0.41	-0.67	0.72
SPB3_HUMAN	Squamous cell carcinoma antigen 1 (SCCA1)	Serpin B3	0.43	-1.63	-0.85	-0.78	0.96
SPB5_HUMAN	Maspin	Serpin B5	0.43	-0.92	-0.37	-0.55	0.69
SPB6_HUMAN	Placental thrombin inhibitor (PTI)	Serpin B6	0.47	-1.04	-0.49	-0.55	0.74
CYTB_HUMAN	Stefin-B	Cystatin-B	0.52	-0.13	-0.58	0.45	0.68
CYTS_HUMAN	Salivary acidic protein 1	Cystatin-S	0.61	-0.18	-0.52	0.33	0.64
CYTA_HUMAN	Stefin-A	Cystatin-A	0.80	-1.37	-1.11	-0.26	1.00
PAI2_HUMAN	Plasminogen activator inhibitor 2		0.84	-0.71	-0.91	0.20	0.96
CYTN_HUMAN	Salivary cystatin-SA-1	Cystatin-SN	0.84	-0.71	-0.91	0.20	0.96
ANT3_HUMAN	Antithrombin-III	Serpin C1	0.97	-0.60	-0.57	-0.03	0.76
ITIH4_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H4	Plasma kallikrein sensitive glycoprotein 120	0.97	-0.73	-0.76	0.03	0.86
AACT_HUMAN	Alpha-1-antichymotrypsin	Serpin A3	0.98	-0.43	-0.44	0.01	0.61

3.3 An evaluation of the antiviral activity of serpin and cystatin antiproteases against HIV

3.3.1 Serpin B4 demonstrates novel antiviral activity in TZM-bl culture

Recombinantly produced Serpin B4 and Cystatin A were assessed for antiviral activity in a TZM-bl reporter cell line. Increasing concentrations of antiprotease were incubated with cells and either BaL (R5-tropic) or IIIB (X4-tropic) HIV lab strains for three days, followed by cellular harvest for assessment of virus infection and cell viability through quantitative assays of HIV-tat protein and ATP levels, respectively. Serpin B4 demonstrated antiviral activity in a dose-dependent manner, and had increased potency against the R5 strain relative to the X4 strain (IC50=14.64 µg/ml for BaL and IC50=32.07 µg/ml for IIIB); however, serpin B4 demonstrated a significant ($p<0.05$, one tailed t test) reduction in viral infection relative to a 100% infection control (33-89% BaL inhibition and 25-70% IIIB inhibition at 3.7-40 µg/ml; Figure 18A). Serpin B4 demonstrated significant reduction in cellular metabolism to 42% of the negative control at 40 µg/ml ($p=0.002$); however, serpin B4 was able to significantly inhibit virus infection where infected cells maintained 100% viability, which provides supporting evidence that inhibition was not due to cell toxicity or a lowered metabolic rate (Figure 18A). Cystatin A did not inhibit infection in TZM-bl culture (Figure 18B) – an observation that is contrary to recent findings published by our lab in Aboud *et al*. 2014 that showed cystatin A may have mild HIV inhibitory properties in TZM-bl cell culture[188]. Differences in virus stocks used between assays may account for differences in findings, and illustrates the need to have independent screening of factors for antiviral properties. A bovine serum albumin (BSA) protein control had no effect on viral inhibition and cell viability; thus, the inhibitory activity of serpin B4 is likely due to qualities specific to this protein (Figure 18C). TZM-bl cells provide a simplistic model to screen factors of interest for antiviral activity in the presence of susceptible cells, but are not the most physiological relevant *in vitro* assay available. TZM-bl cells are transformed cell lines that artificially express high levels of CD4 and chemokine receptors; further the cytokine production is altered relative to human immune cells, which may have added effect on our ability to determine the inhibitory capacity of antiproteases, which are hypothesized to limit infection through inflammatory

regulating capacity. HIV neutralization assays were performed in PBMC culture to attempt to provide a more physiologically relevant model to assess the antiviral activity of these factors.

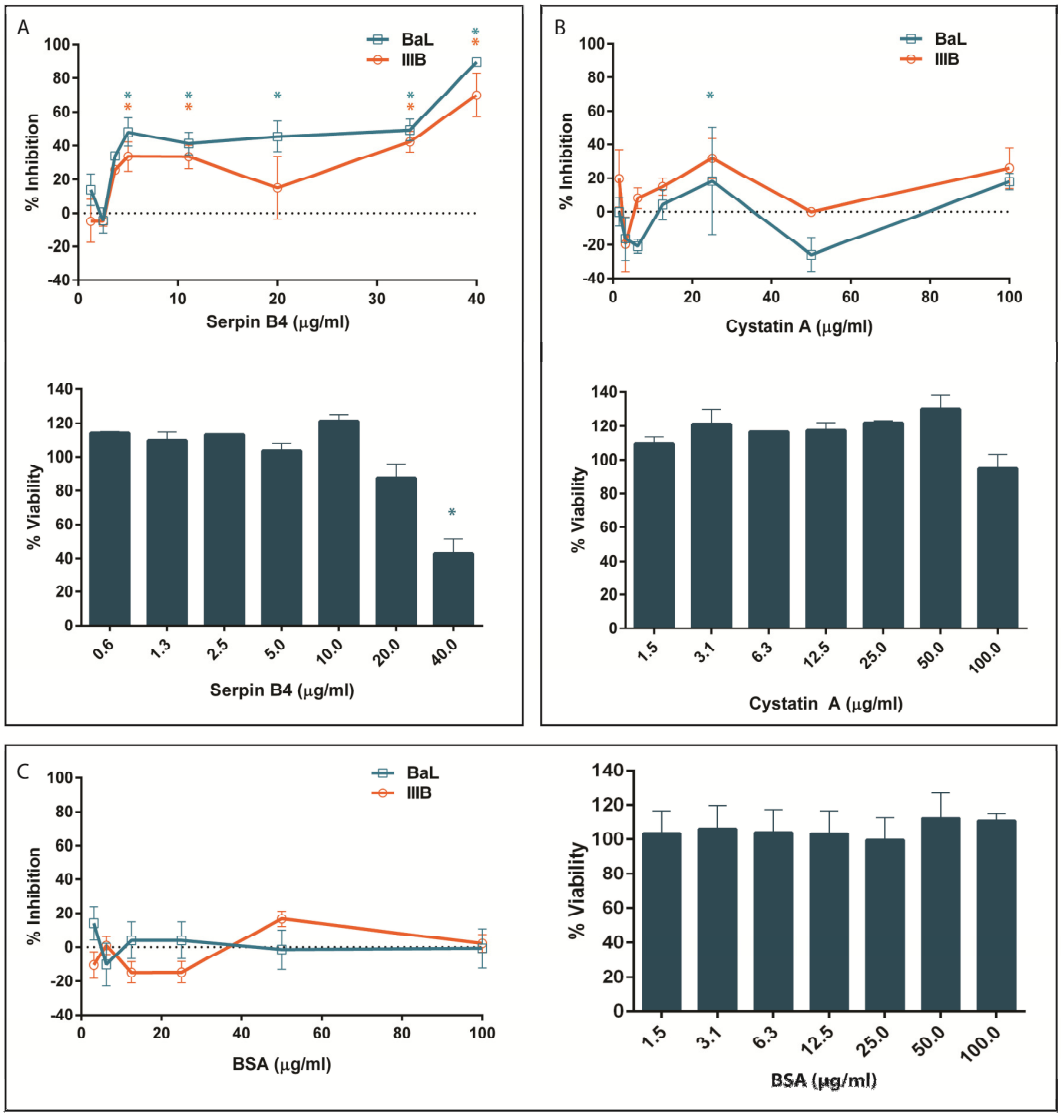


Figure 18: Serpin B4, not cystatin A, demonstrates novel antiviral activity in a TZM-bl reporter cell line. TZM-bl cells were infected with an X4-tropic (IIIB, MOI of 2.0, 36.6 µl/well) or R5-tropic (BaL, MOI of 0.2, 11.5 µl/well) in the presence of serpin B4 (0.6-40.0 µg/ml) and cystatin A (1.6-100.0 µg/ml) for three days. Percent inhibition and viability were calculated relative to a negative infection control (no protein added) via β-gal assay measuring HIV-tat production and ATP (luciferase assay) production, respectively. Serpin B4 demonstrated greater than 45% inhibition against BaL infection at as low as 3.7 µg/ml ($p<0.05$) and showed 34% inhibition at ≥3.7 µg/ml. The inhibitory effect of Serpin B4 was more potent against BaL (IC50=14.64 µg/ml) than IIIB (IC50=32.07 µg/ml). Serpin B4 had statistically significantly cytotoxic effects within cell culture at 40.0 µg/ml (42% viability, $p=0.002$) (A). Cystatin A showed no inhibitory activity or cytotoxic effect in TZM-bl cells against either virus (B). The protein control, bovine serum albumin (BSA), showed no effect on virus inhibition of cell viability within the concentration range tested for either antiprotease (100.0-1.5 µg/ml, $p>0.05$) (C).

3.3.2 Serpin B4 demonstrates novel antiviral activity in PBMC culture

Serpin B4 and cystatin A were assessed for antiviral activity against lab strains of HIV in PHA-stimulated *ex vivo* peripheral blood mononuclear cell (PBMC) culture. Quantification of virus infection was determined using p24 ELISA of culture supernatants, while cell metabolism was estimated using luciferase assay of intracellular ATP levels. Serpin B4 demonstrated significant reduction in HIV infection against IIIB (48.7% inhibition at 1.25 μg/ml, $p<0.05$) and Bal (37.3% inhibition at 5-10 μg/ml, $p<0.05$), with serpin B4 having a more potent effect on IIIB (IC50=2.9 μg/ml) than BaL (IC50= 0.03 μg/ml; Figure 19A). However, at serpin concentrations greater than 10 μg/ml, ATP production levels dropped slightly below 80% of the negative control ($p=0.02$ BaL and $p=0.02$ IIIB infected cultures), and were less than 50% above 20 μg/ml ($p=0.02$ BaL and $p=0.01$ IIIB; Figures 19A). The exact concentrations of serpin B4 within rectal mucosal fluid are unknown, but are likely not as high as 20-40 μg/ml; the high concentrations of serpins used in this assay were assayed to observe a full range of serpin B4 activity, but likely do not reflect natural concentrations. Further, the reduction in ATP may signify a cytotoxic effect due to other components in the serpin suspension media other than the protein of interest.

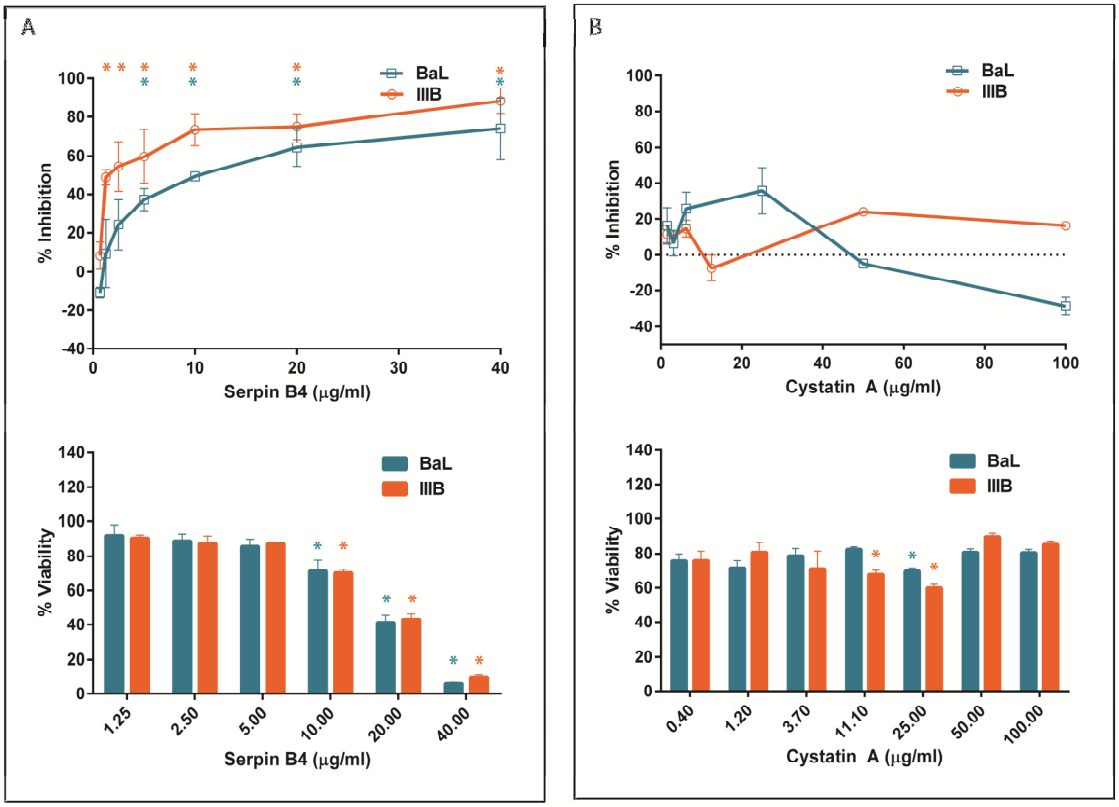


Figure 19: Serpin B4 shows HIV-inhibitory activity against laboratory strains in PBMC culture, but has significant cytotoxicity at high concentrations. HIV-inhibition assays were performed in PHA-stimulated, *ex vivo* peripheral blood mononuclear cell (PBMC) culture with BaL (MOI of 0.05, 6.6 µl/well) and IIIB (MOI 0.05, 1.5 µl) virus. PBMC infection was measured by p24 ELISA of cell culture supernatants and cell viability was assayed through luminescent detection of ATP production by cells. Serpin B4 showed significant inhibition at as low as 5 µg/ml of protein ($p<0.05$) against BaL (37.3% inhibition; IC50=2.9 µg/ml) and IIIB (48.7% inhibition at as low as 1.25 µg/ml; IC50=0.3 µg/ml). PBMCs had a significant reduction in viability below 80% at concentrations greater than 10 µg/ml for both BaL- and IIIB-infected cultures ($p<0.05$) (A). Cystatin A showed no trend in HIV inhibition or cytotoxicity at increasing concentration of protein in culture (B).

3.3.3 Serpin B4 does not correlate with clinical variables relating to exposure.

Serpin B4 levels as determined by mass spectrometry were analyzed in the context of epidemiological data relating to exposure to partner's virus, or a prior immune response to HIV and frequency of unprotected oral intercourse. Partner viral load 0-12 months prior to rectal sampling did not show any significant relationship with serpin B4 expression (Pearson's r=0.016, p=0.29; Figure 20A). No difference was observed in serpin B4 expression between individuals who had detected HIV-neutralizing IgA in their rectal mucosal fluid relative to those who did not ($p<0.52$, two-tailed t tests; Figure 20B).

Further, serpin B4 expression did not differ between individuals practicing unprotected oral intercourse with their HIV-positive partner frequently (responded "often" or "always") or infrequently (responded "sometimes" or "seldom") based on self reported data (two-tailed t tests, p<0.32) (Figure 20C). Overall, serpin B4 expression levels showed no significant association to collected data on exposure in the Venhälsan HESN cohort. Though these comparisons are not adequately powered (<10 subjects/group), they do provide supporting evidence that serpin B4 may be innately expressed in this HESN population, rather than just being expressed as an artifact of exposure. Future investigation into mechanism of these proteins at the population level should consider longitudinal expression serpin B4 in the rectal mucosa to validate that these proteins are expressed at innately high levels.

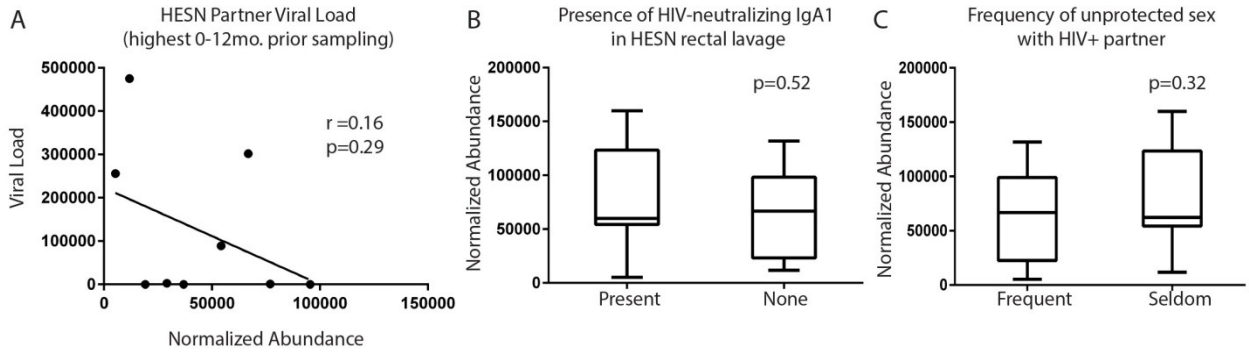


Figure 20: Serpin B4 does not correlate with epidemiological variables relating to exposure to partner virus, presence of HIV-neutralizing IgA in rectal mucosa or frequency of unprotected oral intercourse. HESN MSM reported frequency of exposure to partner's virus through oral unprotected sex (n=13 having often/always, n=10 sporadic/seldom; n=2 no response), viral load of HIV+ partner was measured within 0-12 months prior to sampling (VL>50 copies/ml, n=9) and the presence of HIV-neutralizing IgA in the HESN rectal mucosal sample was assayed (n=7 individuals with HIV-neutralizing IgA, n=18 individuals without HIV-neutralizing IgA). Serpin B4 did not correlate with measured partner viral load (Pearson's r=0.016, p=0.29) (A). There was also no significant difference in serpin B4 expression when HESNs were grouped by presence or absence of HIV-neutralizing IgA in rectal secretions (p=0.52) (B) or the frequency of unprotected oral intercourse with HIV-positive partner (p=0.32) (C).

3.3.4 Evaluation of mucosal protein factors that correlate with serpin B4

Protein correlation analyses were conducted to identify proteins that are related to serpin B4 expression. Serine protease inhibitors have previously characterized roles in immunity[188], yet this has not been fully investigated for serpin B4 in the context of mucosal fluid. A correlation analysis of serpin B4 expression compared to all identified mucosal proteins was completed using Pearson's correlation, and was restricted to proteins that met a significance threshold of $p<0.001$. Co-expression analysis identified strong relationships with other antiproteases (serpin B3 and cystatin A), as well as the antimicrobial protein S100-A7 and acute phase protein, alpha-1-acid glycoprotein 1 ($p<0.0001$, $r>0.6$; Table 7A). Other significantly associated factors included a caspase enzyme (CASPE), several keratin proteins (K2C5, K22E, K1C13, K1C9 and K1C10), antimicrobial proteins (S100A9 and THIO), immunoglobulin chains (LAC2 and IGLL5), and others (Table 7A). Co-expression of proteins may suggest a shared immunological pathway, and provide clues to a mechanism of regulation or mechanism of action of serpin B4 within mucosal fluid. To confirm these relationships, a parallel analysis was performed to correlate serpin B4 expression in rectal secretions from the MACS cohort with other soluble factors.

Correlative analysis in the MACS cohort verified co-expression of serpin B4 with other antiproteases (serpin B3 and cystatin A), but was able to identify co-expression with additional protease inhibitors, A2ML1 and serpin B2 (Table 7B). This analysis was also able to validate co-expression with caspase-14 and S100A7 (Table 7B). Caspase-14 is a cysteine protease involved in epidermal differentiation and cellular apopotosis; serpin B3/4 have been previously identified as inhibitors of intracellular caspases to prevent apoptosis and/or granzyme-mediated death of macrophages to preserve immune cell function during inflammation[284,285]. Serpins B3/4 have also been previously identified as co-expressed within S100-A7 in psoriasis, a disease characterized by chronic inflammation within epidermal tissue, and it has been hypothesised that this co-expression is a result of a continued effort to control protease digestion of epidermal components and limit inflammation associated with this pathology[286]. However, the significance of these relationships within mucosal fluid is still unknown, and additional studies are necessary to understand the role of serpin B4 in immunity within this critical immune barrier.

Table 9A. Proteins co-expressed with serpin B4 within rectal mucosa from the Venhälsan cohort

Accession	*Protein Identification*	*P value*	*R value*
SPB3_HUMAN	Serpin B3	< 0.0001	0.84
S10A7_HUMAN	Protein S100-A7	< 0.0001	0.78
A1AG1_HUMAN	Alpha-1-acid glycoprotein 1	< 0.0001	0.72
CYTA_HUMAN	Cystatin-A	< 0.0001	0.66
PDIA3_HUMAN	Protein disulfide-isomerase A3	0.0001	0.63
CASPE_HUMAN	Caspase-14	0.0001	0.63
MYH11_HUMAN	Myosin-11	0.0001	0.64
1433Z_HUMAN	14-3-3 protein zeta/delta	0.0002	0.63
PLBL1_HUMAN	Phospholipase B-like 1	0.0002	0.61
S10A9_HUMAN	Protein S100-A9	0.0004	0.60
FABP5_HUMAN	Fatty acid-binding protein, epidermal	0.0006	0.58
LAC2_HUMAN	Ig lambda-2 chain C regions	0.0008	0.57
THIO_HUMAN	Thioredoxin	0.0008	0.57
K2C5_HUMAN	Keratin, type II cytoskeletal 5	0.0015	0.54
K22E_HUMAN	Keratin, type II cytoskeletal 2 epidermal	0.0017	0.54
TKT_HUMAN	Transketolase	0.0018	0.54
K1C13_HUMAN	Keratin, type I cytoskeletal 13	0.0021	0.53
K1C9_HUMAN	Keratin, type I cytoskeletal 9	0.0024	0.52
K1C10_HUMAN	Keratin, type I cytoskeletal 10	0.0031	0.51
IGLL5_HUMAN	Immunoglobulin lambda-like polypeptide 5	0.0035	0.51

Table 9B. Proteins co-expressed with serpin B4 in rectal mucosa from the MACS cohort

Accession	*Protein Identification*	*P value*	*R value*
SPB3_HUMAN	Serpin B3	< 0.0001	0.90
A2ML1_HUMAN	Alpha-2-macroglobulin-like protein 1	0.0006	0.73
PAI2_HUMAN	Serpin B2	0.0006	0.73
DSG3_HUMAN	Desmoglein-3	0.0008	0.72
LV106_HUMAN	Ig lambda chain V-I region WAH	0.001	-0.71
NAGK_HUMAN	N-acetyl-D-glucosamine kinase	0.0012	0.70
SPRR3_HUMAN	Small proline-rich protein 3	0.0014	0.69
GGCT_HUMAN	Gamma-glutamylcyclotransferase	0.0016	0.69
S10A7_HUMAN	Protein S100-A7	0.0016	0.69
RD23B_HUMAN	UV excision repair protein RAD23 homolog B	0.0017	0.69
TALDO_HUMAN	Transaldolase	0.0017	0.69
HSPB1_HUMAN	Heat shock protein beta-1	0.0018	0.68
HEBP2_HUMAN	Heme-binding protein 2	0.0021	0.68
CAN2_HUMAN	Calpain-2 catalytic subunit	0.0026	0.67
CALB2_HUMAN	Calretinin	0.0028	0.66
GPX1_HUMAN	Glutathione peroxidase 1	0.0032	0.65
CYTA_HUMAN	Cystatin-A	0.0036	0.65
IGHM_HUMAN	Ig mu chain C region	0.0038	-0.65
CASPE_HUMAN	Caspase-14	0.0048	0.63
TTC38_HUMAN	Tetratricopeptide repeat protein 38	0.0061	-0.62

Chapter 6: Discussion

This thesis used a comprehensive proteomics approach to characterize the rectal mucosal proteome and mucosal immunological differences in HESN individuals. HIV transmission through unprotected receptive anal intercourse (URAI) holds the highest risk for acquisition relative to any other site of mucosal exposure[261]. Susceptibility in the rectum has lead to overrepresentation of high-risk populations practicing URAI within the HIV/AIDS epidemic who remain a major transmission source [287, 288,289]. There is currently no effective vaccine available against HIV, and alternative prevention therapies, such as pre-exposure prophylaxis, have showed mild success, but are still limited by uptake in high risk populations; thus, developing new prevention strategies targeted at limiting rectal acquisition of HIV is a major priority for HIV prevention science and pandemic control.

The rectal mucosa is the primary immunological barrier against HIV transmission during URAI, and contains tissue-based immunological properties that have previously been investigated for their role in HIV susceptibility. The focus of this thesis was to analyze the role of the overlying rectal mucosal fluid layer in HIV susceptibility. Rectal secretions are the first point of contact with the virus upon HIV exposure through URAI; however, their role in HIV acquisition has not been well explored beyond its limited properties as a physical immune barrier. Soluble immune factors within rectal secretions contribute to the defensive properties of mucosal fluid beyond pathogen trapping and clearance and can help shape susceptibility of the underlying tissue layers through regulation of immune activation. Though these proteins have been shown to influence HIV infection at other sites of exposure, the expression and activity of these proteins within rectal fluid has not been well investigated.

The analysis of rectal mucosal fluid from healthy, non-HIV exposed men from the Venhälsan cohort addressed many of the knowledge gaps associated with soluble factor expression within these secretions. HIV neutralization assays of matched salivary and rectal fluid pools demonstrated a significant ability of rectal fluid to limit HIV infection, albeit at a lower capacity than salivary fluid. The mild inhibitory effect of rectal fluid within the physiological range of mucosal protein expression justified further investigation

into the immune properties of this fluid. A comparative proteomic analysis of the matched salivary and rectal samples identified nearly 100 factors with known immune activity between the two fluids. Due to the large increase in rectal susceptibility relative to oral exposure, we hypothesized that rectal fluid would have lowered expression of immune factors that limit HIV infection; however, this was not shown within our study. The majority of immune factors detected showed no significant difference between fluids, which is intuitive when considering the biological requirements of the two compartments.

The oral and rectal mucosa are constantly exposed to food antigens and commensal bacteria, as well as harmful pathogens, as a part of the digestive tract; therefore, the oral and rectal mucosal defence systems must be similarly equipped to maintain a defensive barrier against pathogens while avoiding severe immunopathology from constant stimulation[290,291]. Both fluids contained defensive pro-inflammatory proteins (complement components and S100 proteins) that promote the activation and recruitment of immune cells[292,293], and regulatory anti-inflammatory factors (apolipoproteins) that act to counter inflammation through attenuation of inflammation-signaling pathways[294]. Mass spectrometry was also able to characterize many proteins with direct antimicrobial functions such as pathogen binding/clearing (mucins, deleted in malignant brain tumors 1, peptidoglycan recognition protein) or microbicidal activity (lysozyme c, lactoperoxidase and myeloperoxidase)[139,295,296]; some of these have been found to have specific anti-HIV mechanisms, such as antileukoproteinase and cathelicidin[297]. Furthermore, several antiproteases (serpins and cystatins) were found to be commonly expressed between saliva and rectal mucosal fluid. Antiproteases have an emerging role in immune defense at the mucosal surface and in the blood[238]. They have known inflammatory regulating properties against innate cell proteases in response to infection. Antiproteases have also been found to be overexpressed in an HIV-exposed yet seronegative (HESN) population of commercial sex workers and have defined anti-viral activity in the blood and lung mucosa, implicating a potential role in susceptibility to HIV infection[182,232,281].

Of the significantly differentially expressed factors between saliva and rectal mucosal fluid, many held overlapping functions host defense and inflammation, making it difficult to attribute these

differences to a mechanism of rectal susceptibility. The difference in inhibitory capacity between fluids may also be due to soluble factors below the detection threshold of our proteomic analysis; this includes CC-chemokines/cytokines which are known to be in high abundance in salivary fluid[132], and other short antimicrobial peptides[161] that are critical factors in HIV infection and general immunity, and have previously been characterized within this cohort in the context of saliva, but have not been well defined in rectal fluid[247]. Proteomic analysis did identify a trend in overabundant mucosal immunoglobulins (IgA and IgM) in rectal secretions, suggesting a possible reliance on these proteins for lower gastrointestinal function and/or pathogen defense. This analysis was an important first step in the characterization of the soluble environment of rectal mucosa, and argues that this fluid is an important immune barrier and should be further investigated in its role in gastrointestinal health.

The second proteomic analysis of rectal secretions from a cohort of HIV-exposed seronegative MSM characterized the rectal fluid proteome of individuals who have demonstrated a reduced susceptibility to HIV. We were able to define several hundred proteins involved in various stages of the innate and adaptive immune response within rectal fluid samples from this population. HESN MSM showed unique protein expression in these secretions as defined by both hierarchical clustering of global protein expression and differential expression analysis. Proteins significantly associated with HESN MSM included an increase in innate defense factors (14-3-3 protein zelta/delta, phospholipase B-like 1, S100A7 protein, mucin 2, and alpha-1-acid-glycoprotein 1), antiproteases (serpin B3, serpin B4 and cystatin A) and mucosal immunoglobulins. Differentially expressed proteins were largely of human origin; however, any differences in microbial species between two groups could greatly affect the immune properties at this barrier and throughout the immune system[298]. A sub analysis was performed to determine if there were any bacterial factors uniquely expressed in the HESN population. Mild differences in microbial protein expression were observed within the small subset of proteins. The human gut houses a large and diverse population of commensal bacteria, which can influence host immune signalling and protect colonization of the gut by invading microorganisms[299-301]. Thus, the composition of the microbial community at mucosal surfaces may be important for regulation of both innate and adaptive immunity against HIV. Our analysis using microbial protein expression as an indication of bacterial

composition suggest there could be differences in HESN microbial colonization relative to the control population; however, this is very preliminary, and additional studies into the gut microbiome and its influence of rectal susceptibility to HIV in both susceptible and HESN populations are needed to understand the relationship of these factors on mucosal immunity.

A biofunctional analysis of differentially expressed, human-derived factors described a proteomic profile of HESN men as significantly enriched in immune factors, cell cycle factors, and increased proteins involved in epithelial barrier development. An increase in cell proliferative/apoptotic factors and epidermal development factors in HENS mucosal fluid may suggest increased maintenance activity on the mucosal barrier[270], though the function of these factors extracellularly and their relevance to mucosal integrity and immune function is not well understood. Of particular interest was the increase in immune factors within HESN samples, which was largely related to a relatively high expression of immunoglobulin expression in their rectal lavage. IgA antibodies are a critical component of the gastrointestinal immune barrier. Particular IgA paratopes are able to bind the HIV virion and can aid in viral clearing, block transcytosis across the epithelial barrier or prevent infection of CD4+ T cells[90]. Further, IgA can regulate gastrointestinal immunity through neutralization of bacterial, food and toxin antigens prior to immune stimulation of the gastrointestinal mucosa, thereby limiting inflammatory stimulation within the gut[280,302]. Therefore, in addition to direct HIV-neutralization responses, maintaining a diverse range of IgA binding specificities at the rectal mucosal surface would be beneficial in reducing inflammatory events that may indirectly increase chances of HIV acquisition due to increased target cell recruitment and activation.

In addition to increased immunoglobulin expression, our proteomic investigation of HESN MSM identified antiproteases serpin B3, serpin B4 and cystatin A as associated with the HESN phenotype independent of confounding variables relating to exposure. Elevated serine and cysteine antiproteases have previously been identified in the female genital tract of HESN commercial sex workers[99], making them an appealing target for further investigation to into HIV susceptibility at the soluble barrier of the rectal mucosa. Immune activity of serpins and cystatins may be facilitated through their ability to

maintain mucosal barriers and control inflammation through regulation of protease activity. Proteases may enhance HIV susceptibility through degredation of mucosal tissues, stimulation of inflammatory pathways, degredation of antiviral defense proteins, innate cell chemotaxis, and increased activation of T cells via MHC antigen presentation[177,195]; thus, proper control of these protein are needed to maintain a healthy mucosal barrier. Antiproteases can also control inflammation outside of protease inhibition, as they can modulate immune signalling pathways and pro-inflammatory cytokine expression, suggesting that an increase in these proteins at mucosal surfaces would facilitate a controlled inflammatory state within the environment[177,245]. Several serpins have demonstrated antiviral activity against viruses such as HSV, HCV, influenza and HIV[227-230]. All mechanisms of antiprotease-mediated viral inhibition have not yet been defined; however, peptides from alpha-1-antichymotrypsin have been shown to inhibit HIV fusion[180,182], and antiproteases are able to interfere with late-stage viral processing through protease inhibition[232]. The role of cystatins in HIV immunity are less defined, but they have shown anti-inflammatory signalling properties within macrophages[236], and may have HIV inhibitory activity *in vitro*[188]. Together, the proteomic analysis of rectal mucosal fluid of HESN MSM describes a soluble immune factor environment enriched in mucosal immunoglobulins and antiproteases that, based on their previous characterizations within the literature, may be able to have direct antiviral neutralization properties, and control inflammatory stimulation within tissue to decrease HIV susceptibility. The proposed mechanisms of immune control within this cohort are summarized in Figure 21.

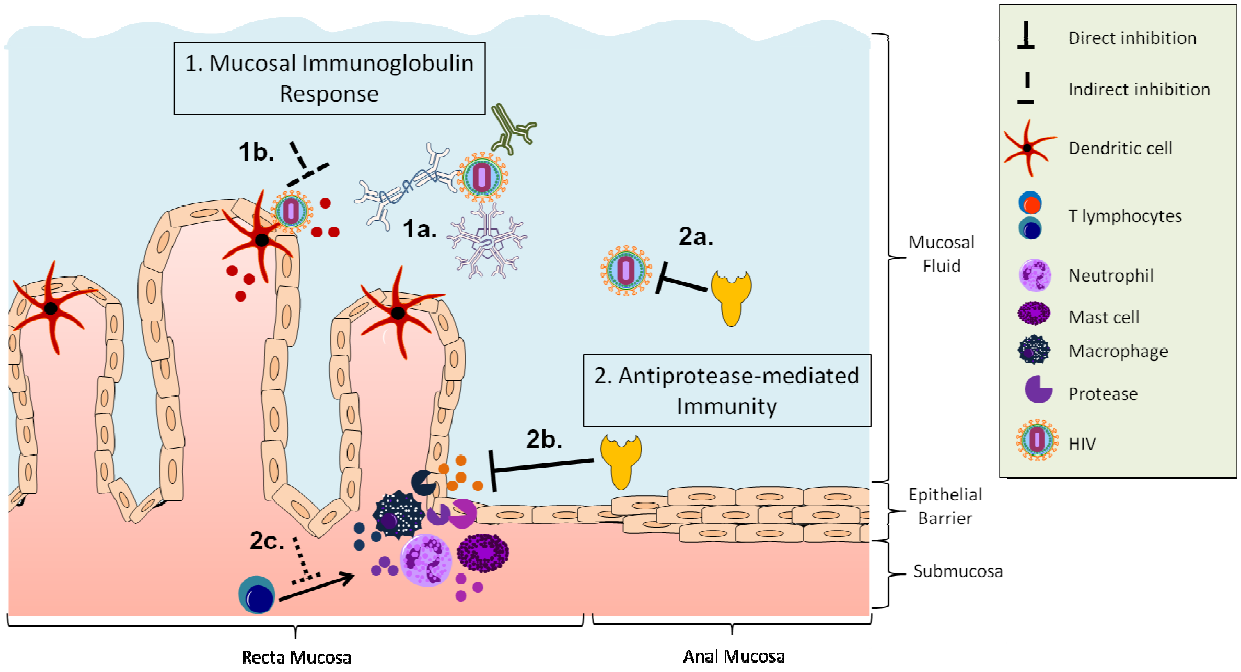


Figure 21: A model of mucosal fluid protein-mediated protection against rectal acquisition of HIV in the Venhälsan cohort of HESN MSM. HESN MSM had elevated levels of mucosal immunoglobulins and antiproteases within their rectal secretions, which may have dual roles in antimicrobial defense and inflammatory control to reduce susceptibility to HIV. Mucosal immunoglobulins are able to neutralize HIV and other invasive microbes to limit infection and colonization (1a); this has the indirect benefit of reducing antigen stimulation at the tissue level, indirectly reducing rectal inflammation (1b). Soluble antiproteases may be able to limit HIV infection through direct antiviral activity (2a). They may also restrict protease activity to prevent degredation of the epithelial barrier and activation/recruitment of front-line immune cells (2b). Reduced innate cell inflammation would subsequently reduce T cell chemotaxis and activation at the rectal mucosa, limiting target cell availability.

Though both increased immunoglobulin levels and immune protein expression suggested by this analysis may contributed to reduced HIV susceptibility, it is possible these are an artifact of sexual behaviour or microbial exposure not present in the control population[280]; thus, protein expression was analyzed in the context of clinical variables relating to sexual activity in order to understand if differentially abundant proteins were a result of exposure. Sexual activity can induce significant changes to the local immune response. Semen contains an abundance of foreign antigens (seminal, microbial or viral), as well as immunomodulatory proteins[277], which may alter immune cell populations and secreted factors at the rectal mucosa[278]. Further, receptive anal intercourse can lead to damage within mucosal tissue, which may increase tissue remodeling and inflammatory pathways[303-305]. However, the effect of intercourse on the rectal proteome has not been fully investigated. The correlative analysis of protein expression with epidemiological variables (frequency of unprotected oral intercourse, presence of HIV-neutralizing IgA and viral load) was able to provide insight into the effect of unprotected sex and pathogen exposure on the rectal mucosa. Major findings include a correlative link between partner viral load and glycolysis proteins, which may be a reflection of the ability of HIV to increase the activation state of the immune cells within tissue[119]. Furthermore, the elevated levels of mucosal immunoglobulins were significantly associated with the presence of HIV-neutralizing IgA expressed in a subset of individuals within the HESN population; this is intuitive as stimulation of HIV-specific IgA would lead to an increase in overall immunoglobulin production[280]. As this was a predominantly orally exposed cohort, the frequency of unprotected oral intercourse on protein expression was examined and showed links to apoptotic factors within secretions; however, additional studies into populations regularly practicing receptive anal intercourse and/or HIV positive populations would yield a better indication of the effects of sexual intercourse on the rectal proteome. This sub-analysis of rectal protein expression was able to highlight proteins that were differentially expressed in HESN secretions that may be associated with epidemiological variables relating to exposure. Moreover, this distinguishes a subset of differentially abundant factors that may be expressed at heightened levels in HESN individuals but are not influenced by exposure levels, and are thereby be said to be associated with reduced HIV-susceptibility in this

cohort. Of interest was a trend in increased levels of antiproteases, which have not yet been investigated for their role in rectal mucosal fluid.

A proteomic investigation of rectal mucosal fluid from the MACS cohort of HESN MSM was performed to validate increased antiprotease expression in populations with reduced susceptibility to HIV. Differential expression of antiproteases was not observed within this cohort, though this cannot be confirmed as this study was underpowered, containing only four HESN individuals. Achieving adequate experimental power remains one of the greatest challenges to studying rectal mucosal fluid by proteomics. The proteome composition of the rectum is highly variable within the colorectal compartment due to the constant movement of organic waste through the GI lumen, and additional variability is incorporated when sampling mucosal fluid vial lavage, rather than a sponge sample, which can be more regulated during the collection process[121,268]. Further, our statistical approach of performing multiple t tests requires us to severely limit our significance threshold to correct for multiple hypothesis testing. High variability within the rectum, combined with stringent statistical thresholds requires large sample sizes to adequately assess proteomic differences between populations. This is both financially and logistically challenging from both the clinical and laboratory perspectives. Analysis of proteomic differences using advanced statistical modelling such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) are novel tools being used to identify immune markers associated with different populations from proteomic data and may help in future analyses of rectal mucosal samples. Though limited, this secondary validation study was able to provide critical insight into the proper design of future proteomic analysis of rectal mucosal secretions and other clinical samples with high variability. Despite our inability to validate antiprotease expression in a second cohort of HESN MSM, two of the antiprotease biomarkers, serpin B4 and cystatin A, were evaluated for antiviral activity in two HIV neutralization assays.

HIV inhibition assays determined that serpin B4 was able to limit HIV infection (CXCR4- and CCR5-tropic) in both a TZM-bl and PBMC cell culture. The use of the transformed reporter cell line (TZM-bl assays) allowed for the rapid quantification of HIV infection in the presence of antiproteases,

while PBMC assays demonstrated a more physiologically relevant cell line. PBMCs contain freshly isolated human immune cells that are more similar to infectable cells at mucosal surfaces. Though the bulk PBMC population contains all leukocytes (lymphocytes, neutrophils, monocytes, basophils and eosinophils), after 72 hours of PHA stimulation, the majority of PBMCs consist of activated CD4+ T lymphocytes, which are the preferred target cell of HIV[306]. PBMCs are able also able to produce large amounts of cc-chemokines and cytokines upon activation, which may be affected through the presence of antiproteases in culture, allowing for a better assessment of antiprotease immune activity against HIV infection. HIV neutralization assays showed a significant inhibitory effect of serpin B4, but not cystatin A, on both an R5- and X4-tropic laboratory strain of HIV. Cytotoxic effects observed at high concentrations of serpin B4 are concerning, yet they are counter intuitive, as this factor has been largely described for its role in inhibition of cell death and promotion of cell proliferation in tumor cells[285]. Additional knockdown experiments are needed to confirm the antiviral activity of serpin B4 in the absence of cellular cytotoxicity. Despite cytotoxicity at high concentrations, serpin B4 was able to demonstrate novel antiviral activity in this thesis makes it a target for future mechanistic studies for a role in inflammation and direct inhibition of virus. Experiments to investigation cytokine expression in colorectal explants in the presence of serpin B4 are an obvious next step and will test the inflammatory modulating activity of this protein within a more physiologically relevant model.

Interestingly, despite known overlapping functions in immunity, serpin B4 and cystatin A showed differences their ability to limit HIV infection within these assays. Cell-specific roles for these factors may limit the applicability of their functions in the literature to our HIV neutralization models. Both antiproteases are known to be expressed abundantly within epithelial cells, which were not accounted for by either the TZM-bl or PBMC culture; therefore, we may be missing potential immune effects of these proteins *in vivo*. For example, known barrier maintenance functions of cystatin A suggest that its heightened expression in the HESN population may be contributing to reduced susceptibility at the rectal mucosa through preservation of tight-junction proteins; however, as epithelial cells were not present in our assays, barrier integrity could not assessed within our model. Further, antiviral activity of cystatins

has been demonstrated within monocytes culture[186]. Monocytes and macrophages are present at only minor concentrations within PBMC culture, with activated CD4+ T cells making up the bulk of infectable cells[306]; therefore, our assays would have missed the antiviral potential of cystatin A in these cell types. However, this possible cell-specificity does highlight that serpin B4 was able to inhibit infection in the predominant HIV target cell, while cystatin A was not, making serpin B4 a promising target possible HIV prevention strategies.

Serpin B4 was produced through plasmid expression in *E. coli* cell culture, which releases bacterial endotoxin in conjunction with expressed protein. Though the recombinant protein suspension was subject to endotoxin removal (<0.2 ng/μg), it may still be able to effect PBMC activation and cytotoxicity at these concentrations[307]. A caveat of this study was that stock vials of serpin B4 (0.130mg/ml) were more dilute than cystatin A (0.980 mg/ml). If endotoxin is affecting culture conditions, this would be enhanced in serpin culture as more of the protein stock media was contained within experimental wells. We would expect that it is toxic components within the media resulting in this cytotoxicity, rather than serpin B4, as B4 and other serpins can inhibit programmed cell death, and promote cell proliferation; however, parallel neutralization assays in the presence of serpin B4 neutralizing antibody, are needed to further confirm the antiviral activity of serpin B4 in the absence of significant cytotoxicity. As serpin B4 was able to inhibit HIV infection in the absence of cytotoxicity at expected physiologically relevant concentrations, it is a promising new biomarker for reduced susceptibility to rectal acquisition of HIV, and should be followed up in mechanistic studies to understand its role in HIV pathogenesis.

Co-expression analysis was performed on serpin B4 expression relative to global protein expression within individuals to gain further insight into shared immune pathways between serpin B4 and other soluble protein factors in rectal fluid. If these proteins are working together in a true immune pathway, we would expect serpin B4 expression to correlate with these factors within rectal mucosal fluid samples from both the Venhälsan and MACS cohort. Global protein correlations identified a strong linear relationship between serpin B3, cystatin A, S100A7 protein and caspase-14, independent of HESN status

or cohort. The correlation between serpin B3 and B4 is intuitive, as they often co-transcribed from the same chromosomal locus and have been well defined by overlapping molecular functions, and roles in immunity and disease[238]. Cystatin A is not transcribed from the same chromosome, but shares many regulatory functions of protease activity[217,245], and may be co-expressed in redundant physiological functions. The correlative relationship between serpin B4 and S100A7 have previously been defined in the context of psoriasis, which suggests a potential link in the immune response to this inflammatory disorder of epithelial cells. S100A7, also known as psoriasin, is a calcium-binding protein that regulates cell cycle progress, while its secreted form has antimicrobial properties at epithelial surfaces[308]. Increased expression of this protein can lead to psoriasis, which is a condition characterized by excessive skin growth and inflammation. Serpin B3 and B4 have been shown to be co-expressed with S100A7 in psoriatic skin lesions under the regulation of IL-1β and IFN-γ, and may suggest a yet undefined relationship with the antimicrobial or inflammatory response of S100 proteins within innate immunity[308]. Previous relationships have also been identified between serpin B4 and caspase-14, as serpins can inhibit caspases intracellularly in the regulation of apoptosis in immune cells[285]; their roles extracellularly in mucosal fluid, however, are unknown. Indeed, this co-expression analysis suggest a possible mechanism for serpin B4 in innate defense and inflammation in conjunction with secreted S100A7 and other granulocyte-secreted factors, such as caspases, and may have a cumulative effect through co-expression with other antiproteases. Further experiments measuring the relationship of serpin B4 with measured cytokine outcomes immune cell recruitment are needed to determine if the antiviral activity of serpin B4 is due to anti-inflammatory activity.

Understanding the role of serpin B4 in inflammation and defense within the soluble layer of the rectal mucosa may be beneficial for HIV prevention therapy. Control of infection at the outermost barrier of the mucosa, rather than within tissue, would be preferable in contrast to defense mechanisms within tissue that allow viral permeation and access to the cellular environment before mounting an attempt to eliminate virus. In the context of rectal microbicide gels, these prevention technologies are an addition to the rectal fluid barrier to provide enhanced protection against HIV during exposure; modelling the design

of these products after the rectal mucosal fluid layer of populations who demonstrate natural reduced susceptibility to HIV while maintaining overall health and controlled immune activation would be a promising strategy for the design of a safe and effective microbicide.

New therapies are being targeted at integrating inflammatory regulators into treatment. Multi-purpose prevention technologies (MPTs) are now being developed to not only target HIV infection, but prevent against concurrent STI infection, as they are a major driver of inflammation and HIV susceptibility[309]. The regulatory function of serpins has not been overlooked as a therapeutic target. Serpin such as B1 and B6 have been proposed as a treatment in chronic infections to restore balance between neutrophil function and tissue remodelling[213,310]. Further, antiviral serpin A1 showed efficacy in under clinical investigation for its use as a therapeutic to reduce viral loads of HIV positive pathients[226]; thus, the use of serpin B4 as a therapeutic is feasible, and may be a beneficial contribution to current microbicides in combination with ARVs. Though these proteins show promise as a novel HIV therapeutic, the cost-effective production and delivery of human serpins as a drug may be challenging, as synthetic production of such large proteins is expensive. As well, though HESN populations demonstrate no adverse events related to heightened levels of antiprotease expression at mucosal tissues, the consequences of increasing levels a human protease inhibitor within tissues not undergoing chronic inflammation are currently unknown; interference with normal serpin-protease interactions holds the potential to over-inhibit normal human protease function and disrupt physiological functions. However, continued research into the exact antiviral mechanisms of serpin B4 may uncover novel immune or inflammatory pathways that can be regulated through targeted therapeutics that may be more cost-effective and have decrease the risk of over-inhibition of proteases or inflammatory responses at the rectal mucosa. Though these findings clearly suggest a novel role for serpin B4 in rectal immunity, with a potential to impact rectal microbicide development, several limitations should be highlighted as they may limit the applicability of the findings within this thesis.

This thesis demonstrated several proteomic analyses to comprehensively define the role of soluble factors in rectal immunity, and identified serpin B4 as a potential novel antiviral proteins within these

secretions; however, limitations within the cohorts used, gender and limitations to studying rectal transmission may limit our interpretation of these findings to all individuals practicing URAI. As this study was largely an examination of the mucosal proteome of Caucasian men from Sweden, it is possible that underlying genetic or environmental factors may restrict our findings to this and ethnically or geographically similar populations. Apart from genetic differences, alterations in diet between populations will likely influence gastrointestinal microflora composition and secreted mucosal immune factors[298], through the exact mechanisms of these relationships are not fully understood. Variation may also exist between the male and female rectal compartment. Current research in women suggests that mucosal factors fluctuate with hormone levels during the menstrual cycle[311,312]. As in the female genital tract, the rectal compartment of both genders contains hormone receptors, such as luteinizing hormone (LH) receptor, that have been found to fluctuate with hormone changes during the menstrual cycle[313]. It is therefore plausible that sex hormone differences between genders can impact immune factor expression in the rectal compartment; however, additional research is needed to fully elucidate the role of hormones, such as estrogen, progesterone and testosterone, on immunity at mucosal surfaces. Further, studying both protected and unprotected receptive anal intercourse exposure in any population (MSM, transgender and heterosexual women) is challenging as RAI is still highly stigmatized. Significant underreporting of RAI, and to a greater extent URAI, has been found across populations due to fear of discrimination by healthcare providers and a history of oppression in MSM and transgender populations due to their sexual orientation[314,315]. Underreporting of the frequency of URAI cannot be ruled out within this study and remains a potential underlying variable in our interpretation of HESN correlates and the epidemiological sub analysis. Future studies of longitudinal analyses of persons who seroconvert during RAI would be able to validate the mucosal factors and biological pathways associated with HIV susceptibility within this study. These would not only add to our knowledge of HIV pathogenesis through rectum, but would help shape the microbicide development pipeline to incorporate such risk factors into efficacy studies, which could increase success of these products. However, despite limitation in the study of rectal

transmission of HESN MSM, this study of the Venhälsan population provides a unique opportunity to understand HIV susceptibility in high-risk MSM.

Several major findings of this thesis will greatly contribute to our knowledge of the immune environment of the rectal mucosa in the context of HIV infection. Rectal mucosal fluid in healthy individuals was able to limit HIV infection *in vitro*, and the antiviral activity within was attributed to a plethora of immune factors contained within these secretions. Analysis of rectal protein expression in the context of confounding variables associated altered levels of energy metabolism and apoptotic immune factors with HIV exposure and unprotected oral intercourse, respectively. Identification of factors artificially enhanced with HIV exposure is important for identification of proteins associated with true reduced susceptibility to viral infection. A secondary proteomic investigation of rectal secretions from a population of HESN MSM identified reduced immune pathway stimulation, and increased levels of mucosal immunoglobulins and antiproteases. Both mucosal immunoglobulins and antiproteases are unique in that they may hold dual roles in antimicrobial defense and inflammatory control, making them promising targets for HIV prevention strategies. This analysis identified serpin B3, serpin B4 and cystatin A as significant biomarkers that had elevated expression in the HESN population relative to control, non-HIV exposed men. The expression of these antiproteases could not be attributed to epidemiological variables relating to HIV exposure, suggesting increased levels were due to innate expression within this population. Of these markers, serpin B4 demonstrated novel antiviral activity *in vitro*. This activity may be due to direct inhibition of the viral life cycle or regulation of target cell activation or availability; further studies are needed to assess the role of serpin B4 as a novel host restriction factor, and as a potential immunoregulatory molecule at the rectal mucosa.

Continuing research into mechanisms of rectal immunity is needed to fully understand this critical immune barrier. Homeostatic properties of the healthy rectal mucosa as well as mucosal dysfunction are not fully understood, and are a major barrier to assessing microbicide safety at the rectal mucosa. Though our investigation into the effect of HIV exposure on the rectal mucosa identified several altered immune pathways, additional high powered studies are need to investigate these proteome changes

and further understand the consequences of HIV exposure on rectal immunity. Finally, our target molecule serpin B4 has proposed mechanisms of defense and immune regulation at mucosal surfaces. Ongoing studies of the influence of serpin B4 on inflammation in colorectal explants will provide supporting evidence for this function and will identify potential inflammatory pathways that may be targets for anti-inflammatory therapeutics within HIV prevention strategies. This thesis has provided a foundation of evidence to support the investigation of rectal mucosal fluid in HIV susceptibility, and has demonstrated the utility of using proteomics to analyze populations with reduced susceptibility to HIV in order to identify new targets for HIV prevention therapy.

Appendix A: Protein expression in rectal mucosal fluid relative to saliva

Supplemental Table A1. Average abundance of ant-inflammatory proteins found in rectal mucosa as determined by mass spectrometry, and relative expression of these proteins in rectal mucosa compared to saliva.

Protein	Mean NA ($x10^3$)	SD NA ($x10^3$)	Mean L2FD	SD L2FD	P-value
Aminopeptidase N	172.76	191.83	2.13	2.16	0.1
Annexin A1	379.22	88.65	0.81	0.03	0.03
Apolipoprotein A-I	307.73	103.33	0.64	0.49	0.08
Apolipoprotein A-II	16.79	13.08	1.12	1.26	0.1
Apolipoprotein A-IV	152.31	131.37	4.40	1.45	0.01
Apolipoprotein D	126.62	64.73	1.41	0.77	0.05
Apolipoprotein E	43.46	36.21	-0.42	1.38	0.5
Apolipoprotein H	39.91	12.61	0.72	0.46	0.1
Apolipoprotein J	558.80	507.91	1.64	1.56	0.1
CD55	116.85	82.20	-0.12	1.11	0.8
CD59 glycoprotein	43.57	44.97	0.23	1.89	0.8
Glutathione S-transferase P	4.66	1.95	-1.65	0.62	0.02

Supplemental Table A2. Average abundance of pro-inflammatory proteins found in rectal mucosa as determine by mass spectrometry, and relative expression of these proteins in rectal mucosa compared to saliva.

Protein	Mean NA ($x10^3$)	SD NA ($x10^3$)	Mean L2FD	SD L2FD	P-value
Adenylyl cyclase-associated protein 1	42.81	39.6	-0.22	1.56	0.7
Cathepsin B	416.13	542.07	1.21	3.26	0.4
CD177 antigen	7.67	1.16	-0.07	0.22	0.6
Complement factor C3	87.75	210.03	0.02	0.39	0.9
Complement factor C5	0.27	0.099	-0.11	0.53	0.7
Complement factor C8	0.99	0.69	0.37	1.10	0.5
Complement factor B	771.64	379.74	1.68	0.74	0.03
Complement factor I	5.16	0.94	0.51	0.26	0.1
Fibulin-1	8.59	6.86	0.71	1.24	0.3
Heat shock 20kDa protein 5	5262.53	7202.34	1.62	4.19	0.4
Heparin cofactor 2	15.05	7.60	3.68	0.76	0.06
IgG	695.54	839.76	0.38	2.59	0.7
Integrin beta-2	1.64	1.80	-1.05	2.11	0.3
Leukotriene A-4 hydrolase	12.71	12.71	-0.15	1.80	0.8
Peroxiredoxin-1	4299.01	6000.28	1.65	5.13	0.5
Phospholipase B-like 1	28.34	12.08	0.99	0.63	0.05
Plasminogen	43.25	16.18	0.93	0.55	0.04
Plastin-2	589.60	93.16	0.68	0.23	0.02
Protein S100-A12	31.00	1.43	-0.29	0.07	0.04
Protein S100-A2	256.25	300.54	1.27	2.42	0.3
Protein S100-A8	478.75	386.65	0.37	1.32	0.5
Protein S100-A9	979.83	705.56	0.47	1.15	0.4
Purine nucleotide phosphorylase	4.19	5.40	1.20	3.12	0.4
Thymidine phosphorylase	248.74	235.85	1.87	1.66	0.09
Ubiquitin C	1704.32	654.21	1.73	0.57	0.02

Supplemental Table A3. Average abundance of antimicrobial proteins found in rectal mucosa as determine by mass spectrometry, and relative expression of these antimicrobials in rectal mucosa compared to saliva.

Protein	Mean NA ($x10^3$)	SD NA ($x10^3$)	Mean L2FD	SD L2FD	P-value
Annexin A3	100.78	54.84	-0.55	0.83	0.2
Antileukoproteinase (SLPI)	91.02	53.40	-1.44	0.90	0.05
Cathelicidin antimicrobial peptide precursor	53.66	19.67	-0.15	0.54	0.5
Deleted in malignant brain tumors 1 protein	1051.50	1253.13	-1.42	2.51	0.3
Haptoglobin	942.93	160.676	-0.19	0.25	0.2
Lactoperoxidase	83.36	34.97	-1.30	0.62	0.03
Lysozyme C	1283.64	668.15	0.01	0.79	0.7
Mucin-2	73.22	63.81	1.62	1.47	0.09
Mucin-5AC	31.85	18.16	0.56	0.87	0.2
Mucin-7	49.71	38.35	-2.21	1.25	0.04
Myeloperoxidase	1196.32	389.22	0.90	0.48	0.03
Peptidoglycan recognition protein 1	1.67	0.68	-0.98	0.60	0.06

Supplemental Table A4. Average abundance of antiprotease proteins found in rectal mucosa as determine by mass spectrometry, and relative expression of these proteins in rectal mucosa compared to saliva.

Protein	Mean NA ($x10^3$)	SD NA ($x10^3$)	Mean L2FD	SD L2FD	P-value
Alpha-2-macroglobulin-like protein 1	794.93	235.99	0.64	0.43	0.06
Cystatin-A	29.08	13.28	0.08	0.68	0.8
Cystatin-B	1266.41	1618.59	-0.82	3.04	0.5
Cystatin-C	40.93	13264.45	-0.95	0.48	0.03
Cystatin-D	181.60	212.00	-1.52	2.39	0.2
Cystatin-S	629.27	423.78	-2.64	1.06	0.02
Cystatin-SA	1918.04	1811.97	-0.81	1.65	0.3
Cystatin-SN	159.41	2.08	-0.73	0.02	0.06
Serpin A1	135.86	36.35	0.02	0.39	0.9
Serpin A3	39.55	24.09	0.81	0.94	0.1
Serpin B1	113.39	26.87	0.09	0.35	0.5
Serpin B12	1.46	1.20	-0.56	1.35	0.4
Serpin B13	80.20	54.63	0.87	1.07	0.1
Serpin C1	1.65	0.72	-0.53	0.65	0.2

References

1 Keele, B. F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523-526, doi:10.1126/science.1126531 (2006).

2 Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789-1796 (2000).

3 Van Heuverswyn, F. *et al.* Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* **444**, 164, doi:10.1038/444164a (2006).

4 Zhu, T. *et al.* An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594-597, doi:10.1038/35400 (1998).

5 Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E. & Hammer, S. M. The challenge of HIV-1 subtype diversity. *N Engl J Med* **358**, 1590-1602, doi:10.1056/NEJMra0706737 (2008).

6 Wolfe, N. D. *et al.* Naturally acquired simian retrovirus infections in central African hunters. *Lancet* **363**, 932-937, doi:10.1016/S0140-6736(04)15787-5 (2004).

7 Bailes, E. *et al.* Hybrid origin of SIV in chimpanzees. *Science* **300**, 1713, doi:10.1126/science.1080657 (2003).

8 Gilbert, M. T. *et al.* The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* **104**, 18566-18570, doi:10.1073/pnas.0705329104 (2007).

9 Hymes, K. B. *et al.* Kaposi's sarcoma in homosexual men-a report of eight cases. *Lancet* **2**, 598-600 (1981).

10 Control, C. f. D. Vol. 31 697-698 (New York, 1983).

11 Control, C. f. D. Vol. 31 652-654 (California, 1982).

12 Control, C. f. D. Vol. 31 513-514 (United States, 1982).

13 Gallo, R. C. *et al.* Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500-503 (1984).

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

15 Schupbach, J. *et al.* Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* **224**, 503-505 (1984).

16 Holtgrave, D. R. Causes of the decline in AIDS deaths, United States, 1995-2002: prevention, treatment or both? *Int J STD AIDS* **16**, 777-781, doi:10.1258/095646205774988109 (2005).

17 Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20**, W13-23, doi:10.1097/01.aids.0000247564.73009.bc (2006).

18 Robertson, D. L. *et al.* HIV-1 nomenclature proposal. *Science* **288**, 55-56 (2000).

19 Burke, D. S. Recombination in HIV: an important viral evolutionary strategy. *Emerg Infect Dis* **3**, 253-259, doi:10.3201/eid0303.970301 (1997).

20 Hemelaar, J. *et al.* Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* **25**, 679-689, doi:10.1097/QAD.0b013e328342ff93 (2011).

21 Berger, E. A. *et al.* A new classification for HIV-1. *Nature* **391**, 240, doi:10.1038/34571 (1998).

22 Solbak, S. M. *et al.* HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein. *Biochim*

Biophys Acta **1828**, 816-823, doi:10.1016/j.bbamem.2012.11.010 (2013).

23 Freed, E. O. HIV-1 replication. *Somat Cell Mol Genet* **26**, 13-33 (2001).

24 Strebel, K. HIV accessory proteins versus host restriction factors. *Curr Opin Virol* **3**, 692-699, doi:10.1016/j.coviro.2013.08.004 (2013).

25 Hladik, F. & Hope, T. J. HIV infection of the genital mucosa in women. *Curr HIV/AIDS Rep* **6**, 20-28 (2009).

26 Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582-1586 (1996).

27 Brenchley, J. M. *et al.* CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J Exp Med* **200**, 749-759, doi:10.1084/jem.20040874 (2004).

28 Klenerman, P., Wu, Y. & Phillips, R. HIV: current opinion in escapology. *Curr Opin Microbiol* **5**, 408-413 (2002).

29 Fevrier, M., Dorgham, K. & Rebollo, A. CD4+ T cell depletion in human immunodeficiency virus (HIV) infection: role of apoptosis. *Viruses* **3**, 586-612, doi:10.3390/v3050586 (2011).

30 Selik R., M. E., Branson B., Owen M., Whitmore S., Hall I. . (ed Center for Disease Control) (2014).

31 Control, C. f. D. (2012).

32 Coppola, N. *et al.* Treatment of chronic hepatitis C in patients with HIV/HCV coinfection. *World J Virol* **4**, 1-12, doi:10.5501/wjv.v4.i1.1 (2015).

33 Ye, B. *et al.* T-cell exhaustion in chronic hepatitis B infection: current knowledge and clinical significance. *Cell Death Dis* **6**, e1694, doi:10.1038/cddis.2015.42 (2015).

34 Pinzone, M. R., Berretta, M., Cacopardo, B. & Nunnari, G. Epstein-Barr Virus- and Kaposi Sarcoma-Associated Herpesvirus-Related Malignancies in the Setting of Human Immunodeficiency Virus Infection. *Semin Oncol* **42**, 258-271, doi:10.1053/j.seminoncol.2014.12.026 (2015).

35 UNAIDS. Global Fact Sheet. (2013).

36 Taylor, B. S., Shalev, N. & Wilkin, T. J. CROI 2014: advances in antiretroviral therapy. *Top Antivir Med* **22**, 616-631 (2014).

37 Wang, H. B., Mo, Q. H. & Yang, Z. HIV vaccine research: the challenge and the way forward. *J Immunol Res* **2015**, 503978, doi:10.1155/2015/503978 (2015).

38 Flynn, N. M. *et al.* Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J Infect Dis* **191**, 654-665, doi:10.1086/428404 (2005).

39 Pitisuttithum, P. *et al.* Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *J Infect Dis* **194**, 1661-1671, doi:10.1086/508748 (2006).

40 Hirsch, V. M. *et al.* Prolonged clinical latency and survival of macaques given a whole inactivated simian immunodeficiency virus vaccine. *J Infect Dis* **170**, 51-59 (1994).

41 Gray, G. E. *et al.* Recombinant adenovirus type 5 HIV gag/pol/nef vaccine in South Africa: unblinded, long-term follow-up of the phase 2b HVTN 503/Phambili study. *Lancet Infect Dis* **14**, 388-396, doi:10.1016/S1473-3099(14)70020-9 (2014).

42 Buchbinder, S. P. *et al.* Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet* **372**, 1881-1893, doi:10.1016/S0140-6736(08)61591-3 (2008).

43 Haynes, B. F. *et al.* Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med* **366**, 1275-1286, doi:10.1056/NEJMoa1113425 (2012).

44 Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med* **361**, 2209-2220, doi:10.1056/NEJMoa0908492 (2009).

45 Day, T. A. & Kublin, J. G. Lessons learned from HIV vaccine clinical efficacy trials. *Curr HIV Res* **11**, 441-449 (2013).

46 Gray, R. H. *et al.* Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet* **357**, 1149-1153, doi:10.1016/S0140-6736(00)04331-2 (2001).

47 Quinn, T. C. *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med* **342**, 921-929, doi:10.1056/NEJM200003303421303 (2000).

48 Organization, W. H. (Geneva, Switzerland, 2013).

49 UNAIDS. (Joint United Nations Programme on HIV/AIDS (UNAIDS), 2014).

50 Perry, C. M. Emtricitabine/tenofovir disoproxil fumarate: in combination with a protease inhibitor in HIV-1 infection. *Drugs* **69**, 843-857, doi:10.2165/00003495-200969070-00005 (2009).

51 Grant, R. M. *et al.* Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N Engl J Med* **363**, 2587-2599, doi:10.1056/NEJMoa1011205 (2010).

52 Anderson, P. L. *et al.* Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. *Sci Transl Med* **4**, 151ra125, doi:10.1126/scitranslmed.3004006 (2012).

53 Choopanya, K. *et al.* Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet* **381**, 2083-2090, doi:10.1016/S0140-6736(13)61127-7 (2013).

54 Thigpen, M. C. *et al.* Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *N Engl J Med* **367**, 423-434, doi:10.1056/NEJMoa1110711 (2012).

55 Baeten, J. M. *et al.* Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N Engl J Med* **367**, 399-410, doi:10.1056/NEJMoa1108524 (2012).

56 Marrazzo, J. M. *et al.* Tenofovir-based preexposure prophylaxis for HIV infection among African women. *N Engl J Med* **372**, 509-518, doi:10.1056/NEJMoa1402269 (2015).

57 Jean-Michel Molina, C. C., Isabelle Charreau, Laurence Meyer, Bruno Spire, Gilles Pialoux, Christian Chidiac, Jean-Francois Delfraissy, Cecile Trenblay. in *Conferenc on Retroviruses and Opportunistic infections* (Seattle, WA. USA, 2015).

58 Baert, L. *et al.* Development of a long-acting injectable formulation with nanoparticles of rilpivirine (TMC278) for HIV treatment. *Eur J Pharm Biopharm* **72**, 502-508, doi:10.1016/j.ejpb.2009.03.006 (2009).

59 van 't Klooster, G. *et al.* Pharmacokinetics and disposition of rilpivirine (TMC278) nanosuspension as a long-acting injectable antiretroviral formulation. *Antimicrob Agents Chemother* **54**, 2042-2050, doi:10.1128/AAC.01529-09 (2010).

60 Andrews, C. D. *et al.* Long-acting integrase inhibitor protects macaques from intrarectal simian/human

immunodeficiency virus. *Science* **343**, 1151-1154, doi:10.1126/science.1248707 (2014).

61 Andrews, C. D. *et al.* A long-acting integrase inhibitor protects female macaques from repeated high-dose intravaginal SHIV challenge. *Sci Transl Med* **7**, 270ra274, doi:10.1126/scitranslmed.3010298 (2015).

62 McGowan, I. & Dezzutti, C. Rectal microbicide development. *Curr Top Microbiol Immunol* **383**, 117-136, doi:10.1007/82_2013_325 (2014).

63 Garg, S. *et al.* Advances in development, scale-up and manufacturing of microbicide gels, films, and tablets. *Antiviral Res* **88 Suppl 1**, S19-29, doi:10.1016/j.antiviral.2010.09.010 (2010).

64 Rohan, L. C. & Sassi, A. B. Vaginal drug delivery systems for HIV prevention. *AAPS J* **11**, 78-87, doi:10.1208/s12248-009-9082-7 (2009).

65 McGowan, I. Rectal microbicides: can we make them and will people use them? *AIDS Behav* **15 Suppl 1**, S66-71, doi:10.1007/s10461-011-9899-9 (2011).

66 Roddy, R. E. *et al.* A controlled trial of nonoxynol 9 film to reduce male-to-female transmission of sexually transmitted diseases. *N Engl J Med* **339**, 504-510, doi:10.1056/NEJM199808203390803 (1998).

67 Feldblum, P. J. & Weir, S. S. The protective effect of nonoxynol-9 against HIV infection. *Am J Public Health* **84**, 1032-1034 (1994).

68 Niruthisard, S., Roddy, R. E. & Chutivongse, S. The effects of frequent nonoxynol-9 use on the vaginal and cervical mucosa. *Sex Transm Dis* **18**, 176-179 (1991).

69 Tabet, S. R. *et al.* Safety and toxicity of nonoxynol-9 gel as a rectal microbicide. *Sex Transm Dis* **26**, 564-571 (1999).

70 Abdool Karim, Q. *et al.* Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science* **329**, 1168-1174, doi:10.1126/science.1193748 (2010).

71 Mansoor, L. E. *et al.* Adherence in the CAPRISA 004 tenofovir gel microbicide trial. *AIDS Behav* **18**, 811-819, doi:10.1007/s10461-014-0751-x (2014).

72 Kiser, P. F., Johnson, T. J. & Clark, J. T. State of the art in intravaginal ring technology for topical prophylaxis of HIV infection. *AIDS Rev* **14**, 62-77 (2012).

73 Smith, J. M. *et al.* Tenofovir disoproxil fumarate intravaginal ring protects high-dose depot medroxyprogesterone acetate-treated macaques from multiple SHIV exposures. *J Acquir Immune Defic Syndr* **68**, 1-5, doi:10.1097/QAI.0000000000000402 (2015).

74 Carballo-Dieguez, A. *et al.* Frequent use of lubricants for anal sex among men who have sex with men: the HIV prevention potential of a microbicidal gel. *Am J Public Health* **90**, 1117-1121 (2000).

75 Fletcher, P. *et al.* The nonnucleoside reverse transcriptase inhibitor UC-781 inhibits human immunodeficiency virus type 1 infection of human cervical tissue and dissemination by migratory cells. *J Virol* **79**, 11179-11186, doi:10.1128/JVI.79.17.11179-11186.2005 (2005).

76 McGowan, I. The development of rectal microbicides for HIV prevention. *Expert Opin Drug Deliv* **11**, 69-82, doi:10.1517/17425247.2013.860132 (2014).

77 Massud, I. *et al.* Lack of prophylactic efficacy of oral maraviroc in macaques despite high drug concentrations in rectal tissues. *J Virol* **87**, 8952-8961, doi:10.1128/JVI.01204-13 (2013).

78 Begay, O. *et al.* Identification of personal lubricants that can cause rectal epithelial cell damage and enhance HIV type 1 replication in vitro. *AIDS Res Hum Retroviruses* **27**, 1019-1024, doi:10.1089/AID.2010.0252 (2011).

79 Fuchs, E. J. *et al.* Hyperosmolar sexual lubricant causes epithelial damage in the distal colon: potential implication for HIV transmission. *J Infect Dis* **195**, 703-710, doi:10.1086/511279 (2007).

80 McGowan, I. *et al.* A phase 1 randomized, double blind, placebo controlled rectal safety and acceptability study of tenofovir 1% gel (MTN-007). *PLoS One* **8**, e60147, doi:10.1371/journal.pone.0060147 (2013).

81 Hladik, F. *et al.* Mucosal effects of tenofovir 1% gel. *Elife* **4**, doi:10.7554/eLife.04525 (2015).

82 Huang, Y. *et al.* The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med* **2**, 1240-1243 (1996).

83 Allers, K. *et al.* Evidence for the cure of HIV infection by CCR5Delta32/Delta32 stem cell transplantation. *Blood* **117**, 2791-2799, doi:10.1182/blood-2010-09-309591 (2011).

84 Stranford, S. A. *et al.* Lack of infection in HIV-exposed individuals is associated with a strong CD8(+) cell noncytotoxic anti-HIV response. *Proc Natl Acad Sci U S A* **96**, 1030-1035 (1999).

85 Kaul, R. *et al.* HIV-1-specific mucosal CD8+ lymphocyte responses in the cervix of HIV-1-resistant prostitutes in Nairobi. *J Immunol* **164**, 1602-1611 (2000).

86 Card, C. M., Ball, T. B. & Fowke, K. R. Immune quiescence: a model of protection against HIV infection. *Retrovirology* **10**, 141, doi:10.1186/1742-4690-10-141 (2013).

87 Card, C. M. *et al.* Decreased immune activation in resistance to HIV-1 infection is associated with an elevated frequency of CD4(+)CD25(+)FOXP3(+) regulatory T cells. *J Infect Dis* **199**, 1318-1322, doi:10.1086/597801 (2009).

88 Broliden, K. *et al.* Functional HIV-1 specific IgA antibodies in HIV-1 exposed, persistently IgG seronegative female sex workers. *Immunol Lett* **79**, 29-36 (2001).

89 Devito, C. *et al.* Cross-clade HIV-1-specific neutralizing IgA in mucosal and systemic compartments of HIV-1-exposed, persistently seronegative subjects. *J Acquir Immune Defic Syndr* **30**, 413-420 (2002).

90 Devito, C. *et al.* Mucosal and plasma IgA from HIV-exposed seronegative individuals neutralize a primary HIV-1 isolate. *AIDS* **14**, 1917-1920 (2000).

91 Hirbod, T. *et al.* HIV acquisition is associated with increased antimicrobial peptides and reduced HIV neutralizing IgA in the foreskin prepuce of uncircumcised men. *PLoS Pathog* **10**, e1004416, doi:10.1371/journal.ppat.1004416 (2014).

92 Yates, N. L. *et al.* HIV-1 gp41 envelope IgA is frequently elicited after transmission but has an initial short response half-life. *Mucosal Immunol* **6**, 692-703, doi:10.1038/mi.2012.107 (2013).

93 Ball, T. B. *et al.* Polymorphisms in IRF-1 associated with resistance to HIV-1 infection in highly exposed uninfected Kenyan sex workers. *AIDS* **21**, 1091-1101, doi:10.1097/QAD.0b013e3280ef6ae1 (2007).

94 Peretz, Y., Cameron, C. & Sekaly, R. P. Dissecting the HIV-specific immune response: a systems biology approach. *Curr Opin HIV AIDS* **7**, 17-23, doi:10.1097/COH.0b013e32834ddb0e (2012).

95 Burgener, A., Sainsbury, J., Plummer, F. A. & Ball, T. B. Systems biology-based

approaches to understand HIV-exposed uninfected women. *Curr HIV/AIDS Rep* **7**, 53-59, doi:10.1007/s11904-010-0039-3 (2010).

96 Miyazawa, M. *et al.* The 'immunologic advantage' of HIV-exposed seronegative individuals. *AIDS* **23**, 161-175, doi:10.1097/QAD.0b013e3283196a80 (2009).

97 Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-725, doi:10.1038/382722a0 (1996).

98 Burgener, A. *et al.* Salivary basic proline-rich proteins are elevated in HIV-exposed seronegative men who have sex with men. *AIDS* **26**, 1857-1867, doi:10.1097/QAD.0b013e328357f79c (2012).

99 Burgener, A. *et al.* Comprehensive proteomic study identifies serpin and cystatin antiproteases as novel correlates of HIV-1 resistance in the cervicovaginal mucosa of female sex workers. *J Proteome Res* **10**, 5139-5149, doi:10.1021/pr200596r (2011).

100 CDC. (National Center for HIV/AIDS, Viral Hepatits, STD, and TB Prevention, Centers for Disease Control and Prevention, 2014).

101 Shacklett, B. L. Cell-mediated immunity to HIV in the female reproductive tract. *Journal of reproductive immunology* **83**, 190-195, doi:10.1016/j.jri.2009.07.012 (2009).

102 Keele, B. F. & Estes, J. D. Barriers to mucosal transmission of immunodeficiency viruses. *Blood* **118**, 839-846, doi:10.1182/blood-2010-12-325860 (2011).

103 Broliden, K. Innate molecular and anatomic mucosal barriers against HIV infection in the genital tract of HIV-exposed seronegative individuals. *The Journal of infectious diseases* **202 Suppl 3**, S351-355, doi:10.1086/655964 (2010).

104 Haase, A. T. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature* **464**, 217-223, doi:10.1038/nature08757 (2010).

105 Hladik, F. & McElrath, M. J. Setting the stage: host invasion by HIV. *Nat Rev Immunol* **8**, 447-457, doi:10.1038/nri2302 (2008).

106 Pudney, J., Quayle, A. J. & Anderson, D. J. Immunological microenvironments in the human vagina and cervix: mediators of cellular immunity are concentrated in the cervical transformation zone. *Biol Reprod* **73**, 1253-1263, doi:10.1095/biolreprod.105.043133 (2005).

107 Nazli, A. *et al.* Exposure to HIV-1 directly impairs mucosal epithelial barrier integrity allowing microbial translocation. *PLoS Pathog* **6**, e1000852, doi:10.1371/journal.ppat.1000852 (2010).

108 Mirmonsef, P., Krass, L., Landay, A. & Spear, G. T. The role of bacterial vaginosis and trichomonas in HIV transmission across the female genital tract. *Curr HIV Res* **10**, 202-210 (2012).

109 Berger, E. A., Murphy, P. M. & Farber, J. M. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annual review of immunology* **17**, 657-700, doi:10.1146/annurev.immunol.17.1.657 (1999).

110 Shattock, R. J. & Moore, J. P. Inhibiting sexual transmission of HIV-1 infection. *Nat Rev Microbiol* **1**, 25-34, doi:10.1038/nrmicro729 (2003).

111 Chenine, A. L. *et al.* Relative transmissibility of an R5 clade C simian-human immunodeficiency virus across different mucosae in macaques parallels the relative risks of sexual HIV-1 transmission in humans via different routes. *J Infect Dis* **201**, 1155-1163, doi:10.1086/651274 (2010).

112 Hocini, H. & Bomsel, M. Infectious human immunodeficiency virus can

rapidly penetrate a tight human epithelial barrier by transcytosis in a process impaired by mucosal immunoglobulins. *J Infect Dis* **179 Suppl 3**, S448-453, doi:10.1086/314802 (1999).

113 Altfeld, M., Fadda, L., Frleta, D. & Bhardwaj, N. DCs and NK cells: critical effectors in the immune response to HIV-1. *Nat Rev Immunol* **11**, 176-186, doi:10.1038/nri2935 (2011).

114 Hu, J., Gardner, M. B. & Miller, C. J. Simian immunodeficiency virus rapidly penetrates the cervicovaginal mucosa after intravaginal inoculation and infects intraepithelial dendritic cells. *J Virol* **74**, 6087-6095 (2000).

115 Smed-Sorensen, A. *et al.* Differential susceptibility to human immunodeficiency virus type 1 infection of myeloid and plasmacytoid dendritic cells. *J Virol* **79**, 8861-8869, doi:10.1128/JVI.79.14.8861-8869.2005 (2005).

116 Anton, P. A. *et al.* Enhanced levels of functional HIV-1 co-receptors on human mucosal T cells demonstrated using intestinal biopsy tissue. *AIDS* **14**, 1761-1765 (2000).

117 McKinnon, L. R. & Kaul, R. Quality and quantity: mucosal CD4+ T cells and HIV susceptibility. *Curr Opin HIV AIDS* **7**, 195-202, doi:10.1097/COH.0b013e3283504941 (2012).

118 Janeway CA., T., P., Walport M, and Shlomchik M. in *Immunobiology* (Garland Science, 2001).

119 Craveiro, M., Clerc, I., Sitbon, M. & Taylor, N. Metabolic pathways as regulators of HIV infection. *Curr Opin HIV AIDS* **8**, 182-189, doi:10.1097/COH.0b013e32835fc53e (2013).

120 Grivel, J. C. *et al.* HIV-1 pathogenesis differs in rectosigmoid and tonsillar tissues infected ex vivo with CCR5- and CXCR4-tropic HIV-1. *AIDS* **21**, 1263-1272, doi:10.1097/QAD.0b013e3281864667 (2007).

121 Brandtzaeg, P. *et al.* Immunobiology and immunopathology of human gut mucosa: humoral immunity and intraepithelial lymphocytes. *Gastroenterology* **97**, 1562-1584 (1989).

122 Mehandru, S. *et al.* Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. *J Exp Med* **200**, 761-770, doi:10.1084/jem.20041196 (2004).

123 Archibald, D. W. & Cole, G. A. In vitro inhibition of HIV-1 infectivity by human salivas. *AIDS Res Hum Retroviruses* **6**, 1425-1432 (1990).

124 Bolscher, J. G. *et al.* Inhibition of HIV-1 IIIB and clinical isolates by human parotid, submandibular, sublingual and palatine saliva. *Eur J Oral Sci* **110**, 149-156 (2002).

125 Malamud, D. *et al.* Inhibition of HIV infectivity by human saliva. *Oral Dis* **3 Suppl 1**, S58-63 (1997).

126 Kazmi, S. H. *et al.* Comparison of human immunodeficiency virus type 1-specific inhibitory activities in saliva and other human mucosal fluids. *Clin Vaccine Immunol* **13**, 1111-1118, doi:10.1128/CDLI.00426-05 (2006).

127 Venkataraman, N., Cole, A. L., Svoboda, P., Pohl, J. & Cole, A. M. Cationic polypeptides are required for anti-HIV-1 activity of human vaginal fluid. *J Immunol* **175**, 7560-7567 (2005).

128 Frew, L. *et al.* Human cathelicidin production by the cervix. *PLoS One* **9**, e103434, doi:10.1371/journal.pone.0103434 (2014).

129 Levinson, P. *et al.* HIV-neutralizing activity of cationic polypeptides in cervicovaginal secretions of women in HIV-serodiscordant relationships. *PLoS One* **7**, e31996,

doi:10.1371/journal.pone.0031996 (2012).

130 Fouda, G. G. *et al.* HIV-specific functional antibody responses in breast milk mirror those in plasma and are primarily mediated by IgG antibodies. *J Virol* **85**, 9555-9567, doi:10.1128/JVI.05174-11 (2011).

131 McNeely, T. B. *et al.* Secretory leukocyte protease inhibitor: a human saliva protein exhibiting anti-human immunodeficiency virus 1 activity in vitro. *J Clin Invest* **96**, 456-464, doi:10.1172/JCI118056 (1995).

132 Shugars, D. C., Alexander, A. L., Fu, K. & Freel, S. A. Endogenous salivary inhibitors of human immunodeficiency virus. *Arch Oral Biol* **44**, 445-453 (1999).

133 Quinones-Mateu, M. E. *et al.* Human epithelial beta-defensins 2 and 3 inhibit HIV-1 replication. *AIDS* **17**, F39-48, doi:10.1097/01.aids.0000096878.73209.4f (2003).

134 Bergman, P., Walter-Jallow, L., Broliden, K., Agerberth, B. & Soderlund, J. The antimicrobial peptide LL-37 inhibits HIV-1 replication. *Curr HIV Res* **5**, 410-415 (2007).

135 Cocchi, F. *et al.* Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* **270**, 1811-1815 (1995).

136 Mantis, N. J., Rol, N. & Corthesy, B. Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunol* **4**, 603-611, doi:10.1038/mi.2011.41 (2011).

137 Woof, J. M. & Kerr, M. A. The function of immunoglobulin A in immunity. *J Pathol* **208**, 270-282, doi:10.1002/path.1877 (2006).

138 Romas, L. M. *et al.* A comparative proteomic analysis of the soluble immune factor environment of rectal and oral mucosa. *PLoS One* **9**, e100820, doi:10.1371/journal.pone.0100820 (2014).

139 Bergey, E. J. *et al.* Interaction of HIV-1 and human salivary mucins. *J Acquir Immune Defic Syndr* **7**, 995-1002 (1994).

140 Sheng, Y. H. *et al.* MUC1 and MUC13 differentially regulate epithelial inflammation in response to inflammatory and infectious stimuli. *Mucosal Immunol* **6**, 557-568, doi:10.1038/mi.2012.98 (2013).

141 Steinstraesser, L. *et al.* Inhibition of early steps in the lentiviral replication cycle by cathelicidin host defense peptides. *Retrovirology* **2**, 2, doi:10.1186/1742-4690-2-2 (2005).

142 Wong, J. H. *et al.* Effects of cathelicidin and its fragments on three key enzymes of HIV-1. *Peptides* **32**, 1117-1122, doi:10.1016/j.peptides.2011.04.017 (2011).

143 Ogawa, Y. *et al.* Antimicrobial Peptide LL-37 Produced by HSV-2-Infected Keratinocytes Enhances HIV Infection of Langerhans Cells. *Cell host & microbe* **13**, 77-86, doi:10.1016/j.chom.2012.12.002 (2013).

144 Zanetti, M. Cathelicidins, multifunctional peptides of the innate immunity. *Journal of leukocyte biology* **75**, 39-48, doi:10.1189/jlb.0403147 (2004).

145 Crombie, R. Mechanism of thrombospondin-1 anti-HIV-1 activity. *AIDS Patient Care STDS* **14**, 211-214, doi:10.1089/108729100317821 (2000).

146 Rusnati, M. *et al.* Thrombospondin-1/HIV-1 tat protein interaction: modulation of the biological activity of extracellular Tat. *FASEB J* **14**, 1917-1930, doi:10.1096/fj.99-0902com (2000).

147 Ellegard, R. *et al.* Complement opsonization of HIV-1 results in decreased antiviral and inflammatory responses in immature dendritic cells via CR3. *J Immunol* **193**, 4590-4601, doi:10.4049/jimmunol.1401781 (2014).

148 Tjomsland, V. *et al.* Complement opsonization of HIV-1 results in a different intracellular processing pattern and enhanced MHC class I presentation by dendritic cells. *Eur J Immunol* **43**, 1470-1483, doi:10.1002/eji.201242935 (2013).

149 Verani, A. & Lusso, P. Chemokines as natural HIV antagonists. *Current molecular medicine* **2**, 691-702 (2002).

150 Ma, G. *et al.* Secretory leukocyte protease inhibitor binds to annexin II, a cofactor for macrophage HIV-1 infection. *The Journal of experimental medicine* **200**, 1337-1346, doi:10.1084/jem.20041115 (2004).

151 Bingle, L. & Tetley, T. D. Secretory leukoprotease inhibitor: partnering alpha 1-proteinase inhibitor to combat pulmonary inflammation. *Thorax* **51**, 1273-1274 (1996).

152 Wingens, M. *et al.* Induction of SLPI (ALP/HUSI-I) in epidermal keratinocytes. *J Invest Dermatol* **111**, 996-1002, doi:10.1046/j.1523-1747.1998.00425.x (1998).

153 Williams, S. E., Brown, T. I., Roghanian, A. & Sallenave, J. M. SLPI and elafin: one glove, many fingers. *Clinical science* **110**, 21-35, doi:10.1042/CS20050115 (2006).

154 Devito, C. *et al.* Mucosal and plasma IgA from HIV-1-exposed uninfected individuals inhibit HIV-1 transcytosis across human epithelial cells. *Journal of immunology* **165**, 5170-5176 (2000).

155 Lamm, M. E. Interaction of antigens and antibodies at mucosal surfaces. *Annual review of microbiology* **51**, 311-340, doi:10.1146/annurev.micro.51.1.311 (1997).

156 Robinovitch, M. R. *et al.* Parotid salivary basic proline-rich proteins inhibit HIV-I infectivity. *Oral Dis* **7**, 86-93 (2001).

157 Amado, F., Lobo, M. J., Domingues, P., Duarte, J. A. & Vitorino, R. Salivary peptidomics. *Expert review of proteomics* **7**, 709-721, doi:10.1586/epr.10.48 (2010).

158 Gu, M., Haraszthy, G. G., Collins, A. R. & Bergey, E. J. Identification of salivary proteins inhibiting herpes simplex virus 1 replication. *Oral microbiology and immunology* **10**, 54-59 (1995).

159 Newman, F., Beeley, J. A. & MacFarlane, T. W. Adherence of oral microorganisms to human parotid salivary proteins. *Electrophoresis* **17**, 266-270, doi:10.1002/elps.1150170146 (1996).

160 White, S. H., Wimley, W. C. & Selsted, M. E. Structure, function, and membrane integration of defensins. *Current opinion in structural biology* **5**, 521-527 (1995).

161 Weinberg, A., Krisanaprakornkit, S. & Dale, B. A. Epithelial antimicrobial peptides: review and significance for oral applications. *Crit Rev Oral Biol Med* **9**, 399-414 (1998).

162 Furci, L., Sironi, F., Tolazzi, M., Vassena, L. & Lusso, P. Alpha-defensins block the early steps of HIV-1 infection: interference with the binding of gp120 to CD4. *Blood* **109**, 2928-2935, doi:10.1182/blood-2006-05-024489 (2007).

163 Furci, L., Tolazzi, M., Sironi, F., Vassena, L. & Lusso, P. Inhibition of HIV-1 infection by human alpha-defensin-5, a natural antimicrobial peptide expressed in the genital and intestinal mucosae. *PloS one* **7**, e45208, doi:10.1371/journal.pone.0045208 (2012).

164 Levinson, P. *et al.* Levels of innate immune factors in genital fluids: association of alpha defensins and LL-37 with genital infections and increased HIV acquisition. *AIDS* **23**, 309-317, doi:10.1097/QAD.0b013e328321809c (2009).

165 Rapista, A. *et al.* Human defensins 5 and 6 enhance HIV-1 infectivity through promoting HIV attachment.

Retrovirology **8**, 45, doi:10.1186/1742-4690-8-45 (2011).

166 Yang, D. *et al.* Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* **286**, 525-528 (1999).

167 Ellison, R. T., 3rd, LaForce, F. M., Giehl, T. J., Boose, D. S. & Dunn, B. E. Lactoferrin and transferrin damage of the gram-negative outer membrane is modulated by Ca2+ and Mg2+. *Journal of general microbiology* **136**, 1437-1446 (1990).

168 Harmsen, M. C. *et al.* Antiviral effects of plasma and milk proteins: lactoferrin shows potent activity against both human immunodeficiency virus and human cytomegalovirus replication in vitro. *The Journal of infectious diseases* **172**, 380-388 (1995).

169 Viani, R. M., Gutteberg, T. J., Lathey, J. L. & Spector, S. A. Lactoferrin inhibits HIV-1 replication in vitro and exhibits synergy when combined with zidovudine. *Aids* **13**, 1273-1274 (1999).

170 Goldman, A. S. & Smith, C. W. Host resistance factors in human milk. *The Journal of pediatrics* **82**, 1082-1090 (1973).

171 Heo, S. M. *et al.* Host Defense Proteins Derived from Human Saliva Bind to Staphylococcus aureus. *Infection and immunity*, doi:10.1128/IAI.00825-12 (2013).

172 Simpson, A. J., Maxwell, A. I., Govan, J. R., Haslett, C. & Sallenave, J. M. Elafin (elastase-specific inhibitor) has anti-microbial activity against gram-positive and gram-negative respiratory pathogens. *FEBS letters* **452**, 309-313 (1999).

173 Henriksen, P. A. *et al.* Adenoviral gene delivery of elafin and secretory leukocyte protease inhibitor attenuates NF-kappa B-dependent inflammatory responses of human endothelial cells and macrophages to atherogenic stimuli. *Journal of immunology* **172**, 4535-4544 (2004).

174 Butler, M. W. *et al.* Elafin prevents lipopolysaccharide-induced AP-1 and NF-kappaB activation via an effect on the ubiquitin-proteasome pathway. *The Journal of biological chemistry* **281**, 34730-34735, doi:10.1074/jbc.M604844200 (2006).

175 Gettins, P. G. Serpin structure, mechanism, and function. *Chem Rev* **102**, 4751-4804 (2002).

176 Fahey, J. V. *et al.* New approaches to making the microenvironment of the female reproductive tract hostile to HIV. *American journal of reproductive immunology* **65**, 334-343, doi:10.1111/j.1600-0897.2010.00949.x (2011).

177 Mangan, M. S., Kaiserman, D. & Bird, P. I. The role of serpins in vertebrate immunity. *Tissue Antigens* **72**, 1-10, doi:10.1111/j.1399-0039.2008.01059.x (2008).

178 Heutinck, K. M., ten Berge, I. J., Hack, C. E., Hamann, J. & Rowshani, A. T. Serine proteases of the human immune system in health and disease. *Molecular immunology* **47**, 1943-1955, doi:10.1016/j.molimm.2010.04.020 (2010).

179 Congote, L. F. The C-terminal 26-residue peptide of serpin A1 is an inhibitor of HIV-1. *Biochemical and biophysical research communications* **343**, 617-622, doi:10.1016/j.bbrc.2006.02.190 (2006).

180 Shapiro, L., Pott, G. B. & Ralston, A. H. Alpha-1-antitrypsin inhibits human immunodeficiency virus type 1. *FASEB J* **15**, 115-122, doi:10.1096/fj.00-0311com (2001).

181 Zhou, X., Shapiro, L., Fellingham, G., Willardson, B. M. & Burton, G. F. HIV replication in CD4+ T lymphocytes in the presence and absence of follicular dendritic cells: inhibition of replication mediated by alpha-1-antitrypsin through altered IkappaBalpha

ubiquitination. *Journal of immunology* **186**, 3148-3155, doi:10.4049/jimmunol.1001358 (2011).

182 Munch, J. *et al.* Discovery and optimization of a natural HIV-1 entry inhibitor targeting the gp41 fusion peptide. *Cell* **129**, 263-275, doi:10.1016/j.cell.2007.02.042 (2007).

183 Moriuchi, H., Moriuchi, M. & Fauci, A. S. Cathepsin G, a neutrophil-derived serine protease, increases susceptibility of macrophages to acute human immunodeficiency virus type 1 infection. *Journal of virology* **74**, 6849-6855 (2000).

184 Wittek, N. & Majewska, E. [Cystatin C--modulator of immune processes]. *Przegl Lek* **67**, 484-487 (2010).

185 Kopitar-Jerala, N. The role of cystatins in cells of the immune system. *FEBS letters* **580**, 6295-6301, doi:10.1016/j.febslet.2006.10.055 (2006).

186 Luciano-Montalvo, C. & Melendez, L. M. Cystatin B associates with signal transducer and activator of transcription 1 in monocyte-derived and placental macrophages. *Placenta* **30**, 464-467, doi:10.1016/j.placenta.2009.03.003 (2009).

187 Rivera-Rivera, L., Perez-Laspiur, J., Colon, K. & Melendez, L. M. Inhibition of interferon response by cystatin B: implication in HIV replication of macrophage reservoirs. *J Neurovirol* **18**, 20-29, doi:10.1007/s13365-011-0061-2 (2012).

188 Aboud, L., Ball, T. B., Tjernlund, A. & Burgener, A. The role of serpin and cystatin antiproteases in mucosal innate immunity and their defense against HIV. *Am J Reprod Immunol* **71**, 12-23, doi:10.1111/aji.12166 (2014).

189 Dezzutti, C. S. *et al.* Mucosal Escherichia coli bactericidal activity and immune mediators are associated with HIV-1 seroconversion in women participating in the HPTN 035 trial. *J Infect Dis* **206**, 1931-1935, doi:10.1093/infdis/jis555 (2012).

190 Muhlia-Almazan, A., Sanchez-Paz, A. & Garcia-Carreno, F. L. Invertebrate trypsins: a review. *J Comp Physiol B* **178**, 655-672, doi:10.1007/s00360-008-0263-y (2008).

191 Berti, P. J. & Storer, A. C. Alignment/phylogeny of the papain superfamily of cysteine proteases. *J Mol Biol* **246**, 273-283, doi:10.1006/jmbi.1994.0083 (1995).

192 Buhling, F. *et al.* Review: novel cysteine proteases of the papain family. *Adv Exp Med Biol* **477**, 241-254, doi:10.1007/0-306-46826-3_26 (2000).

193 Wensink, A. C., Hack, C. E. & Bovenschen, N. Granzymes regulate proinflammatory cytokine responses. *J Immunol* **194**, 491-497, doi:10.4049/jimmunol.1401214 (2015).

194 Stapels, D. A., Geisbrecht, B. V. & Rooijakkers, S. H. Neutrophil serine proteases in antibacterial defense. *Curr Opin Microbiol* **23**, 42-48, doi:10.1016/j.mib.2014.11.002 (2015).

195 Chapman, H. A., Riese, R. J. & Shi, G. P. Emerging roles for cysteine proteases in human biology. *Annu Rev Physiol* **59**, 63-88, doi:10.1146/annurev.physiol.59.1.63 (1997).

196 Sim, R. B. & Tsiftsoglou, S. A. Proteases of the complement system. *Biochem Soc Trans* **32**, 21-27, doi:10.1042/ (2004).

197 Reeves, E. P. *et al.* Killing activity of neutrophils is mediated through activation of proteases by K+ flux. *Nature* **416**, 291-297, doi:10.1038/416291a (2002).

198 Segal, A. W. How neutrophils kill microbes. *Annu Rev Immunol* **23**, 197-223, doi:10.1146/annurev.immunol.23.021704.115653 (2005).

199 Honey, K. & Rudensky, A. Y. Lysosomal cysteine proteases regulate antigen presentation. *Nat Rev Immunol* **3**, 472-482, doi:10.1038/nri1110 (2003).

200 Lin, L. & Hu, K. Tissue plasminogen activator and inflammation: from phenotype to signaling mechanisms. *Am J Clin Exp Immunol* **3**, 30-36 (2014).

201 Wilgus, T. A., Roy, S. & McDaniel, J. C. Neutrophils and Wound Repair: Positive Actions and Negative Reactions. *Adv Wound Care (New Rochelle)* **2**, 379-388, doi:10.1089/wound.2012.0383 (2013).

202 Rath, T., Roderfeld, M., Graf, J. & Roeb, E. [Matrix metalloproteinases in inflammatory bowel disease - from basic research to clinical significance]. *Z Gastroenterol* **47**, 758-769, doi:10.1055/s-0028-1109520 (2009).

203 Li, C. & Kuemmerle, J. F. Mechanisms that mediate the development of fibrosis in patients with Crohn's disease. *Inflamm Bowel Dis* **20**, 1250-1258, doi:10.1097/MIB.0000000000000043 (2014).

204 Ashton-Rickardt, P. G. An emerging role for Serine Protease Inhibitors in T lymphocyte immunity and beyond. *Immunol Lett* **152**, 65-76, doi:10.1016/j.imlet.2013.04.004 (2013).

205 Silverman, G. A. *et al.* The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem* **276**, 33293-33296, doi:10.1074/jbc.R100016200 (2001).

206 Irving, J. A., Pike, R. N., Lesk, A. M. & Whisstock, J. C. Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res* **10**, 1845-1864 (2000).

207 Mulgrew, A. T., Taggart, C. C. & McElvaney, N. G. Alpha-1-antitrypsin deficiency: current concepts. *Lung* **185**, 191-201, doi:10.1007/s00408-007-9009-y (2007).

208 Nita, I. M., Serapinas, D. & Janciauskiene, S. M. alpha1-Antitrypsin regulates CD14 expression and soluble CD14 levels in human monocytes in vitro. *Int J Biochem Cell Biol* **39**, 1165-1176, doi:10.1016/j.biocel.2007.02.017 (2007).

209 Hadzic, R. *et al.* Alpha1-antitrypsin inhibits Moraxella catarrhalis MID protein-induced tonsillar B cell proliferation and IL-6 release. *Immunol Lett* **102**, 141-147, doi:10.1016/j.imlet.2005.08.006 (2006).

210 Travis, J., Bowen, J. & Baugh, R. Human alpha-1-antichymotrypsin: interaction with chymotrypsin-like proteinases. *Biochemistry* **17**, 5651-5656 (1978).

211 Cicardi, M., Zingale, L., Zanichelli, A., Pappalardo, E. & Cicardi, B. C1 inhibitor: molecular and clinical aspects. *Springer Semin Immunopathol* **27**, 286-298, doi:10.1007/s00281-005-0001-4 (2005).

212 Liu, D. *et al.* C1 inhibitor prevents Gram-negative bacterial lipopolysaccharide-induced vascular permeability. *Blood* **105**, 2350-2355, doi:10.1182/blood-2004-05-1963 (2005).

213 Rees, D. D. *et al.* Recombinant human Monocyte/Neutrophil elastase inhibitor protects rat lungs against injury from cystic fibrosis airway secretions. *Am J Respir Cell Mol Biol* **20**, 69-78, doi:10.1165/ajrcmb.20.1.3306 (1999).

214 Bladergroen, B. A. *et al.* The granzyme B inhibitor, protease inhibitor 9, is mainly expressed by dendritic cells and at immune-privileged sites. *J Immunol* **166**, 3218-3225 (2001).

215 Barrett, A. J. *et al.* Nomenclature and classification of the proteins homologous with the cysteine-proteinase inhibitor chicken cystatin. *Biochem J* **236**, 312 (1986).

216 Turk, V., Stoka, V. & Turk, D. Cystatins: biochemical and structural properties,

and medical relevance. *Front Biosci* **13**, 5406-5420 (2008).

217 Abrahamson, M., Alvarez-Fernandez, M. & Nathanson, C. M. Cystatins. *Biochem Soc Symp*, 179-199 (2003).

218 Turk, V. & Bode, W. The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett* **285**, 213-219 (1991).

219 Bode, W. *et al.* The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J* **7**, 2593-2599 (1988).

220 Zavasnik-Bergant, T. Cystatin protease inhibitors and immune functions. *Front Biosci* **13**, 4625-4637 (2008).

221 Turk, B. *et al.* Regulation of the activity of lysosomal cysteine proteinases by pH-induced inactivation and/or endogenous protein inhibitors, cystatins. *Biol Chem Hoppe Seyler* **376**, 225-230 (1995).

222 Maizels, R. M., Gomez-Escobar, N., Gregory, W. F., Murray, J. & Zang, X. Immune evasion genes from filarial nematodes. *Int J Parasitol* **31**, 889-898 (2001).

223 Scharfstein, J. Parasite cysteine proteinase interactions with alpha 2-macroglobulin or kininogens: differential pathways modulating inflammation and innate immunity in infection by pathogenic trypanosomatids. *Immunobiology* **211**, 117-125, doi:10.1016/j.imbio.2005.10.014 (2006).

224 Cordelier, P. & Strayer, D. S. Mechanisms of alpha1-antitrypsin inhibition of cellular serine proteases and HIV-1 protease that are essential for HIV-1 morphogenesis. *Biochim Biophys Acta* **1638**, 197-207 (2003).

225 Jia, Q. *et al.* Short cyclic peptides derived from the C-terminal sequence of alpha1-antitrypsin exhibit significant anti-HIV-1 activity. *Bioorg Med Chem Lett* **22**, 2393-2395, doi:10.1016/j.bmcl.2012.02.037 (2012).

226 Forssmann, W. G. *et al.* Short-term monotherapy in HIV-infected patients with a virus entry inhibitor against the gp41 fusion peptide. *Sci Transl Med* **2**, 63re63, doi:10.1126/scitranslmed.3001697 (2010).

227 Asmal, M. *et al.* Inhibition of HCV by the serpin antithrombin III. *Virol J* **9**, 226, doi:10.1186/1743-422X-9-226 (2012).

228 Quenelle, D. C., Hartman, T. L., Buckheit, R. W., Prichard, M. N. & Lynn, R. G. Anti-HSV activity of serpin antithrombin III. *Int Trends Immun* **2**, 87-92 (2014).

229 Smee, D. F., Hurst, B. L., Day, C. W. & Geiben-Lynn, R. Influenza Virus H1N1 inhibition by serine protease inhibitor (serpin) antithrombin III. *Int Trends Immun* **2**, 83-86 (2014).

230 Whitney, J. B., Asmal, M. & Geiben-Lynn, R. Serpin induced antiviral activity of prostaglandin synthetase-2 against HIV-1 replication. *PLoS One* **6**, e18589, doi:10.1371/journal.pone.0018589 (2011).

231 Asmal, M. *et al.* In vivo anti-HIV activity of the heparin-activated serine protease inhibitor antithrombin III encapsulated in lymph-targeting immunoliposomes. *PLoS One* **7**, e48234, doi:10.1371/journal.pone.0048234 (2012).

232 Dittmann, M. *et al.* A serpin shapes the extracellular environment to prevent influenza A virus maturation. *Cell* **160**, 631-643, doi:10.1016/j.cell.2015.01.040 (2015).

233 Challacombe, S. J. & Sweet, S. P. Oral mucosal immunity and HIV infection: current status. *Oral Dis* **8 Suppl 2**, 55-62 (2002).

234 Guy, B. *et al.* A specific inhibitor of cysteine proteases impairs a Vif-dependent modification of human

immunodeficiency virus type 1 Env protein. *J Virol* **65**, 1325-1331 (1991).

235 Cantres-Rosario, Y. *et al.* Cathepsin B and cystatin B in HIV-seropositive women are associated with infection and HIV-1-associated neurocognitive disorders. *AIDS* **27**, 347-356, doi:10.1097/QAD.0b013e32835b3e47 (2013).

236 Rivera, L. E., Colon, K., Cantres-Rosario, Y. M., Zenon, F. M. & Melendez, L. M. Macrophage derived cystatin B/cathepsin B in HIV replication and neuropathogenesis. *Curr HIV Res* **12**, 111-120 (2014).

237 Izuhara, K., Ohta, S., Kanaji, S., Shiraishi, H. & Arima, K. Recent progress in understanding the diversity of the human ov-serpin/clade B serpin family. *Cell Mol Life Sci* **65**, 2541-2553, doi:10.1007/s00018-008-8049-7 (2008).

238 Gatto, M. *et al.* Serpins, immunity and autoimmunity: old molecules, new functions. *Clin Rev Allergy Immunol* **45**, 267-280, doi:10.1007/s12016-013-8353-3 (2013).

239 Palmiter, R. D., Gagnon, J. & Walsh, K. A. Ovalbumin: a secreted protein without a transient hydrophobic leader sequence. *Proc Natl Acad Sci U S A* **75**, 94-98 (1978).

240 Moosbrugger-Martinz, V. *et al.* Epidermal barrier abnormalities in exfoliative ichthyosis with a novel homozygous loss-of-function mutation in CSTA. *Br J Dermatol*, doi:10.1111/bjd.13545 (2014).

241 Brooke, M. A., Nitoiu, D. & Kelsell, D. P. Cell-cell connectivity: desmosomes and disease. *J Pathol* **226**, 158-171, doi:10.1002/path.3027 (2012).

242 Butler, M. W. *et al.* Modulation of cystatin A expression in human airway epithelium related to genotype, smoking, COPD, and lung cancer. *Cancer Res* **71**, 2572-2581, doi:10.1158/0008-5472.CAN-10-2046 (2011).

243 Vasilopoulos, Y., Sagoo, G. S., Cork, M. J., Walters, K. & Tazi-Ahnini, R. HLA-C, CSTA and DS12346 susceptibility alleles confer over 100-fold increased risk of developing psoriasis: evidence of gene interaction. *J Hum Genet* **56**, 423-427, doi:10.1038/jhg.2011.33 (2011).

244 Jorgensen, I. *et al.* Serum cysteine proteases and their inhibitors in rheumatoid arthritis: relation to disease activity and radiographic progression. *Clin Rheumatol* **30**, 633-638, doi:10.1007/s10067-010-1585-1 (2011).

245 Kato, T., Takai, T., Mitsuishi, K., Okumura, K. & Ogawa, H. Cystatin A inhibits IL-8 production by keratinocytes stimulated with Der p 1 and Der f 1: biochemical skin barrier against mite cysteine proteases. *J Allergy Clin Immunol* **116**, 169-176, doi:10.1016/j.jaci.2005.03.044 (2005).

246 Kan, Y. *et al.* Imiquimod suppresses propagation of herpes simplex virus 1 by upregulation of cystatin A via the adenosine receptor A1 pathway. *J Virol* **86**, 10338-10346, doi:10.1128/JVI.01196-12 (2012).

247 Hasselrot, K. *et al.* HIV-1 exposed uninfected men who have sex with men have increased levels of salivary CC-chemokines associated with sexual behavior. *AIDS* **24**, 1569-1575 (2010).

248 Hasselrot, K. *et al.* Oral HIV-exposure elicits mucosal HIV-neutralizing antibodies in uninfected men who have sex with men. *AIDS* **23**, 329-333, doi:10.1097/QAD.0b013e32831f924c (2009).

249 Perez, C. L., Hasselrot, K., Bratt, G., Broliden, K. & Karlsson, A. C. Induction of systemic HIV-1-specific cellular immune responses by oral exposure in the uninfected partner of discordant couples. *AIDS* **24**, 969-974 (2010).

250 Hasselrot, K. *et al.* Orally exposed uninfected individuals have systemic anti-HIV responses associating with partners' viral load. *AIDS* **24**, 35-43,

doi:10.1097/QAD.0b013e3283329853 (2010).

251 Hasselrot, K. *Mucosal and systemic immune responses in HIV-1 exposed uninfected individuals* PhD thesis, Karolinska Institutet, (2009).

252 Detels, R. *et al.* Resistance to HIV-1 infection. Multicenter AIDS Cohort Study. *J Acquir Immune Defic Syndr* **7**, 1263-1269 (1994).

253 Kaslow, R. A. *et al.* The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol* **126**, 310-318 (1987).

254 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362, doi:10.1038/nmeth.1322 (2009).

255 van 't Wout, A. B., Schuitemaker, H. & Kootstra, N. A. Isolation and propagation of HIV-1 on peripheral blood mononuclear cells. *Nat Protoc* **3**, 363-370, doi:10.1038/nprot.2008.3 (2008).

256 Fuss, I. J., Kanof, M. E., Smith, P. D. & Zola, H. Isolation of whole mononuclear cells from peripheral blood and cord blood. *Curr Protoc Immunol* **Chapter 7**, Unit7 1, doi:10.1002/0471142735.im0701s85 (2009).

257 Kim, N. M. a. J. H. *HIV protocols* 317-23 (Human Press Inc. , 1999).

258 Fung, K. P. & Wong, T. W. Estimation of quantal dose response of drugs by the Spearman-Karber method: a computer program written in BASIC. *J Singapore Paediatr Soc* **31**, 129-132 (1989).

259 Nyambi, P. N., Burda, S., Bastiani, L. & Williams, C. A virus binding assay for studying the antigenic landscape on intact, native, primary human immunodeficiency virus-type 1. *J Immunol Methods* **253**, 253-262 (2001).

260 Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* **105**, 7552-7557, doi:10.1073/pnas.0802203105 (2008).

261 Patel, P. *et al.* Estimating per-act HIV transmission risk: a systematic review. *AIDS* **28**, 1509-1519, doi:10.1097/QAD.0000000000000298 (2014).

262 Fultz, P. N. Components of saliva inactivate human immunodeficiency virus. *Lancet* **2**, 1215 (1986).

263 Baggaley, R. F. *et al.* Heterosexual anal intercourse: a neglected risk factor for HIV? *Am J Reprod Immunol* **69 Suppl 1**, 95-105, doi:10.1111/aji.12064 (2013).

264 Hasselrot, K. *The HESN men have demonstrated unique oral and systemic responses to HIV in the form HIV-neutralizing IgA1 antibodies against their partner's virus in both the saliva and sera* PhD thesis, Karolinska Institutet, (2009).

265 Pudney, J., Oneta, M., Mayer, K., Seage, G., 3rd & Anderson, D. Pre-ejaculatory fluid as potential vector for sexual transmission of HIV-1. *Lancet* **340**, 1470 (1992).

266 Baggaley, R. F., White, R. G. & Boily, M. C. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *Int J Epidemiol* **39**, 1048-1063, doi:10.1093/ije/dyq057 (2010).

267 Burgener, A. *et al.* A systems biology examination of the human female genital tract shows compartmentalization of immune factor expression. *J Virol* **87**, 5141-5150, doi:10.1128/JVI.03347-12 (2013).

268 Birse, K. M. *et al.* Unbiased proteomics analysis demonstrates significant variability in mucosal immune factor expression depending on the site and method of collection. *PLoS One* **8**, e79505, doi:10.1371/journal.pone.0079505 (2013).

269 Presland, R. Function of Filaggrin and Caspase-14 in Formation and Maintenance of the Epithelial Barrier. *Dermatol Sinica* (2009).

270 Presland, R. B. & Jurevic, R. J. Making sense of the epithelial barrier: what molecular biology and genetics tell us about the functions of oral mucosal and epidermal tissues. *J Dent Educ* **66**, 564-574 (2002).

271 Kalinin, A., Marekov, L. N. & Steinert, P. M. Assembly of the epidermal cornified cell envelope. *J Cell Sci* **114**, 3069-3070 (2001).

272 West, C. E., Jenmalm, M. C. & Prescott, S. L. The gut microbiota and its role in the development of allergic disease: a wider perspective. *Clin Exp Allergy* **45**, 43-53, doi:10.1111/cea.12332 (2015).

273 Marchetti, G., Tincati, C. & Silvestri, G. Microbial translocation in the pathogenesis of HIV infection and AIDS. *Clin Microbiol Rev* **26**, 2-18, doi:10.1128/CMR.00050-12 (2013).

274 Salas, J. T. & Chang, T. L. Microbiome in human immunodeficiency virus infection. *Clin Lab Med* **34**, 733-745, doi:10.1016/j.cll.2014.08.005 (2014).

275 Griffin, P. M. *et al.* Illnesses associated with Escherichia coli O157:H7 infections. A broad clinical spectrum. *Ann Intern Med* **109**, 705-712 (1988).

276 Hansen, R., Thomson, J. M., Fox, J. G., El-Omar, E. M. & Hold, G. L. Could Helicobacter organisms cause inflammatory bowel disease? *FEMS Immunol Med Microbiol* **61**, 1-14, doi:10.1111/j.1574-695X.2010.00744.x (2011).

277 Doncel, G. F., Anderson, S. & Zalenskaya, I. Role of semen in modulating the female genital tract microenvironment--implications for HIV transmission. *Am J Reprod Immunol* **71**, 564-574, doi:10.1111/aji.12231 (2014).

278 Brito, M. O. *et al.* Risk behaviours and prevalence of sexually transmitted infections and HIV in a group of Dominican gay men, other men who have sex with men and transgender women. *BMJ Open* **5**, e007747, doi:10.1136/bmjopen-2015-007747 (2015).

279 Palmer, C. S. *et al.* Increased glucose metabolic activity is associated with CD4+ T-cell activation and depletion during chronic HIV infection. *AIDS* **28**, 297-309, doi:10.1097/QAD.0000000000000128 (2014).

280 Gommerman, J. L., Rojas, O. L. & Fritz, J. H. Re-thinking the functions of IgA(+) plasma cells. *Gut Microbes* **5**, 652-662, doi:10.4161/19490976.2014.969977 (2014).

281 Burgener, A. *et al.* Identification of differentially expressed proteins in the cervical mucosa of HIV-1-resistant sex workers. *J Proteome Res* **7**, 4446-4454, doi:10.1021/pr800406r (2008).

282 Lajoie, J. *et al.* A distinct cytokine and chemokine profile at the genital mucosa is associated with HIV-1 protection among HIV-exposed seronegative commercial sex workers. *Mucosal Immunol* **5**, 277-287, doi:10.1038/mi.2012.7 (2012).

283 Levin, Y. The role of statistical power analysis in quantitative proteomics. *Proteomics* **11**, 2565-2567, doi:10.1002/pmic.201100033 (2011).

284 de Koning, P. J. *et al.* Intracellular serine protease inhibitor SERPINB4 inhibits granzyme M-induced cell death. *PLoS One* **6**, e22645, doi:10.1371/journal.pone.0022645 (2011).

285 Song, K. J., Ahn, H. J. & Nam, H. W. Anti-apoptotic effects of SERPIN B3 and B4 via STAT6 activation in macrophages after infection with Toxoplasma gondii. *Korean J Parasitol* **50**, 1-6, doi:10.3347/kjp.2012.50.1.1 (2012).

286 Gudjonsson, J. E. *et al.* Assessment of the psoriatic transcriptome in a large sample: additional regulated genes and

comparisons with in vitro models. *J Invest Dermatol* **130**, 1829-1840, doi:10.1038/jid.2010.36 (2010).

287 Baral, S., Sifakis, F., Cleghorn, F. & Beyrer, C. Elevated risk for HIV infection among men who have sex with men in low- and middle-income countries 2000-2006: a systematic review. *PLoS Med* **4**, e339, doi:10.1371/journal.pmed.0040339 (2007).

288 Mayer KH, M. M. Past as prologue: the refractory and evolving HIV epidemic among men who have sex with men. *Clin Infect Dis* **52**, 1371-1373 (2011).

289 Birrell, P. J. *et al.* HIV incidence in men who have sex with men in England and Wales 2001-10: a nationwide population study. *Lancet Infect Dis* **13**, 313-318, doi:10.1016/S1473-3099(12)70341-9 (2013).

290 Berbee, J. F., Havekes, L. M. & Rensen, P. C. Apolipoproteins modulate the inflammatory response to lipopolysaccharide. *J Endotoxin Res* **11**, 97-103, doi:10.1179/096805105X35215 (2005).

291 Kang, S. S. *et al.* IFN-gamma renders human intestinal epithelial cells responsive to lipopolysaccharide of Vibrio cholerae by down-regulation of DMBT1. *Comp Immunol Microbiol Infect Dis* **35**, 345-354, doi:10.1016/j.cimid.2012.02.003 (2012).

292 Zhuang, Y. & Lyga, J. Inflammaging in skin and other tissues - the roles of complement system and macrophage. *Inflamm Allergy Drug Targets* **13**, 153-161 (2014).

293 Hammad, H. & Lambrecht, B. N. Barrier Epithelial Cells and the Control of Type 2 Immunity. *Immunity* **43**, 29-40, doi:10.1016/j.immuni.2015.07.007 (2015).

294 Vakili, L. *et al.* Systemic inflammation, intestine, and paraoxonase-1. *Adv Exp Med Biol* **824**, 83-88, doi:10.1007/978-3-319-07320-0_8 (2014).

295 Ligtenberg, A. J., Karlsson, N. G. & Veerman, E. C. Deleted in malignant brain tumors-1 protein (DMBT1): a pattern recognition receptor with multiple binding sites. *Int J Mol Sci* **11**, 5212-5233, doi:10.3390/ijms1112521 (2010).

296 Sorensen, O. E., Borregaard, N. & Cole, A. M. Antimicrobial peptides in innate immune responses. *Contrib Microbiol* **15**, 61-77, doi:10.1159/000136315 (2008).

297 Moutsopoulos, N. M., Greenwell-Wild, T. & Wahl, S. M. Differential mucosal susceptibility in HIV-1 transmission and infection. *Adv Dent Res* **19**, 52-56 (2006).

298 Purchiaroni, F. *et al.* The role of intestinal microbiota and the immune system. *Eur Rev Med Pharmacol Sci* **17**, 323-333 (2013).

299 Guarner, F. & Malagelada, J. R. Gut flora in health and disease. *Lancet* **361**, 512-519, doi:10.1016/S0140-6736(03)12489-0 (2003).

300 Cebra, J. J., Periwal, S. B., Lee, G., Lee, F. & Shroff, K. E. Development and maintenance of the gut-associated lymphoid tissue (GALT): the roles of enteric bacteria and viruses. *Dev Immunol* **6**, 13-18 (1998).

301 Campion, S. L. *et al.* Proteome-wide analysis of HIV-specific naive and memory CD4(+) T cells in unexposed blood donors. *J Exp Med* **211**, 1273-1280, doi:10.1084/jem.20130555 (2014).

302 Phalipon, A. *et al.* Secretory component: a new role in secretory IgA-mediated immune exclusion in vivo. *Immunity* **17**, 107-115 (2002).

303 Denison, F. C., Grant, V. E., Calder, A. A. & Kelly, R. W. Seminal plasma components stimulate interleukin-8 and interleukin-10 release. *Mol Hum Reprod* **5**, 220-226 (1999).

304 Prakash, M., Patterson, S., Gotch, F. & Kapembwa, M. S. Recruitment of CD4 T

lymphocytes and macrophages into the cervical epithelium of women after coitus. *Am J Obstet Gynecol* **188**, 376-381 (2003).

305 Sharkey, D. J., Macpherson, A. M., Tremellen, K. P. & Robertson, S. A. Seminal plasma differentially regulates inflammatory cytokine gene expression in human cervical and vaginal epithelial cells. *Mol Hum Reprod* **13**, 491-501, doi:10.1093/molehr/gam028 (2007).

306 O'Donovan, M. R., Johns, S. & Wilcox, P. The effect of PHA stimulation on lymphocyte sub-populations in whole-blood cultures. *Mutagenesis* **10**, 371-374 (1995).

307 Conti, P. *et al.* Activation of human natural killer cells by lipopolysaccharide and generation of interleukin-1 alpha, beta, tumour necrosis factor and interleukin-6. Effect of IL-1 receptor antagonist. *Immunology* **73**, 450-456 (1991).

308 Ryckman, C., Vandal, K., Rouleau, P., Talbot, M. & Tessier, P. A. Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. *J Immunol* **170**, 3233-3242 (2003).

309 Fernandez-Romero, J. A. *et al.* Multipurpose prevention technologies: the future of HIV and STI protection. *Trends Microbiol*, doi:10.1016/j.tim.2015.02.006 (2015).

310 Brahn, E., Lee, S., Lucas, A., McFadden, G. & Macaulay, C. Suppression of collagen-induced arthritis with a serine proteinase inhibitor (serpin) derived from myxoma virus. *Clin Immunol* **153**, 254-263, doi:10.1016/j.clim.2014.05.003 (2014).

311 Rahman, S. *et al.* Mucosal serpin A1 and A3 levels in HIV highly exposed sero-negative women are affected by the menstrual cycle and hormonal contraceptives but are independent of epidemiological confounders. *Am J Reprod Immunol* **69**, 64-72, doi:10.1111/aji.12014 (2013).

312 Rodriguez-Garcia, M., Patel, M. V. & Wira, C. R. Innate and adaptive anti-HIV immune responses in the female reproductive tract. *J Reprod Immunol* **97**, 74-84, doi:10.1016/j.jri.2012.10.010 (2013).

313 Wang, L. H., Zhang, W., Gao, Q. X. & Wang, F. Expression of the luteinizing hormone receptor (LHR) gene in ovine non-gonadal tissues during estrous cycle. *Genet Mol Res* **11**, 3766-3780, doi:10.4238/2012.October.15.8 (2012).

314 Boily, M. C., Baggaley, R. F. & Masse, B. The role of heterosexual anal intercourse for HIV transmission in developing countries: are we ready to draw conclusions? *Sex Transm Infect* **85**, 408-410, doi:10.1136/sti.2009.037499 (2009).

315 Brody, S. Patients misrepresenting their risk factors for AIDS. *Int J STD AIDS* **6**, 392-398 (1995).