MAKERERE UNIVERSITY

HLA CLASS I AND PEDIATRIC HIV DISEASE PROGRESSION IN BOTSWANA AND UGANDA

BY

SAMUEL KYOBE

MB.ChB (Mak); MMED Microbiology (Mak)
Registration No.: 2016/HD07/19061U
Student No.: 1600790619

A THESIS SUBMITTED TO THE DIRECTORATE OF RESEARCH AND GRADUATE TRAINING FOR THE AWARD OF DEGREE OF DOCTOR OF PHILOSOPHY OF MAKERERE UNIVERSITY

SEPTEMBER 2023

Declaration

I declare that the work outlined in this PhD thesis is my original contribution, and this work has not been presented anywhere else for the award of an academic qualification.

SAMUEL KYOBE

MB.ChB., MMed. Microbiology

Reg. No.: 2016/HD07/19061U

Signature:.....

Approved by Supervisors:

This thesis has been submitted for examination with the approval of the following supervisors.

Dr. Jacqueline Kyosiimire-Lugemwa, Ph.D.

MRC/UVRI & London School of Hygiene and Tropical Medicine Research Unit

……..

Prof. Graeme Mardon PhD.

Department of Pathology & Immunology

Department of Molecular and Human Genetics

Baylor College of Medicine

……..

Dr. Hakim Sendagire MB.ChB, MSc., PhD.

Department of Medical Microbiology, School of Biomedical Sciences

College of Health Sciences, Makerere University

……..

Copyright Statement

I Samuel Kyobe do declare that the work presented in this proposal is my original work and has not been presented anywhere else for any award of a degree.

Dedication

To

My parents E. G. K. Njuki (RIP) and Nassozi Regina
who started this journey of faith in education.

My wife Rachel, and sons Elisha Eric, David, Jaden Samuel and Gabriel Charles
For the overwhelming support coupled with endless love that has brought me one step closer to the verge of the next challenge.

My mentors, Prof. G. Mardon, N. Hanchard and Dr. J K. Lugemwa
For being the most encouraging mentors, any student could have,
Thank you for taking a chance on me.

"It is like a voyage of discovery into unknown lands,
seeking not new territory but for new knowledge.
It should appeal to those with a good sense of adventure" - **Fredrick Sanger**

Acknowledgement

Ultimately, this journey comes to a climax. Filled with ups and down but nonetheless a very satisfying experience surrounded by endless love, endurance and willingness to guide and mentor. My special thanks go to Profs. Graeme Mardon, Alison Elliot, and Neil Hanchard for their continued support throughout this endeavor and for being my strongest advocates. This dissertation would not have been possible without your guidance and advice. Your dedication to advancing career is an example that should shape my future endeavors. At all turns and twists in this journey you challenged my intellectual thought to find a greater understanding of my results and how they fit within a broader picture. No acknowledgment would be complete without thanking my local supervisors Drs. Jackie L. Kyosiimire and Hakim Sendagire, for their constant support throughout the journey. With Dr. Jackie, you maintained an open-door policy for discussion and exhibited a high degree of patience that allowed me to pursue my goals to a logical conclusion. Thank you to the members of my doctoral committee, in particular Drs. Andrew Kambugu, Bernard Bagaya, Henry Kajumbula and recently Prof. Benon Asiimwe for encouraging me to examine a more complete story in pediatric HIV and enabling me to cross the finish line.

Thanks to Savannah Mwesigwa and Grace Kisitu for their great contribution to this work. Savannah deserves special mention for assisting me greatly to learn bioinformatics, and access and use servers and other software like R. I can say you made my coding journey worthwhile. Thanks to Eric Katagirya, Gaone Retshabile, Marion Amujal, Lesedi Williams and Busisiwe Mlotshwa for being great PhD buddies and the excellent technical support. Special thanks to Gloria Nassuuna for emotional support and prayers. I am grateful for the positive energy and the constructive criticism that you brought during lab meetings to shape and mature my concepts. You have collectively made this journey memorable.

Thank you to the Collaborative African Genomics Network (CAfGEN) team lead by the PIs, Profs. Mogomotsi Matshaba, Adeodata Kekitiinwa, Dithan Kiragga, Sununguko M. Wata, Chester Brown, and Ishmael Kasvosve. I hold a special place in my memories for the late Prof. Gabriel Anabwani, the pioneering CAfGEN PI; at the turbulent times you steadied the ship and facilitated my journey to new PhD support. And all the members of CAfGEN including John Farirai, Angella Mirembe, John Mukisa, Gasene Sebetso, Thabo Diphoko, Edward Wampande, Masego Tsimako and Betty Nsangi. Thanks to the lab teams in the immunology and molecular

labs in Uganda led by Fred Katabazi, Edgar Kigozi, Rose Nabatanzi and in Botswana led by Koketso Morapedi for processing the samples. Thanks to the IT teams at Makerere, UVRI and Botswana, recognizing Ronald Oceng, Dafala Kezimbira, Timothy Wamala, Edward Lukyamuzi and Lesego Ketumile. I want to thank the CAfGEN study participants and the study site nurse and lab technicians that assisted with the recruitment. I want to thank the study participants that voluntarily chose to participate in the CAfGEN study, your choice to advance science is undeniably appreciated.

Thanks to the team at UVRI, Prof. Pontiano Kaleebu for recommending me to the MUIIplus PhD sponsorship and to Victoria Bukirwa, Moses Kiiza, Raymond Muganyizi, and Joshua Mandre for providing funding and a platform to present regular updates from my work. Special thanks to Gyaviira Nkurunyingi, Moses Egesa, and Joel Kabali for their support in the ELISPOT lab work.

Thank you to the team at UC Berkeley in the Coscoy Lab for the CRISPR assays and Nottingham Trent University especially Drs. David Robinson and Jody Winter for providing access to the super compute cluster. Thanks to the various funders and funding mechanisms including the NIH, Wellcome, H3Africa Consortium, Alliance for Global Health & Science and Deltas Africa.

To all my friends, past and present, thank you for the direct and indirect support.

Finally, my thanks to my greatest pillar of support in life, my family, mom Regina, wife Rachel, sons Eric, David, Jaden, and Gabriel (Angel), brothers Simon, Fredrick and Stephen, and sisters Cathy, Jenny, and Essy. My uncles the Mutumbas and Augustine for enlightening my education path. Your faith in me leaves me humbled and I love you all so much.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
ART	antiretroviral therapy
CAfGEN	Collaborative African Genomics Network
CAR	chimeric antigen receptor
CAS	computational alanine scanning
CD	Cell differentiation marker
CDC	Centers for Disease Control
CTL	$CD8^+$ cytotoxic T lymphocytes
DC	Dendritic cell
DNA	Deoxyribonucleic acid
ELISPOT	Enzyme-linked immunosorbent assay
Env	envelope
Gap	group antigen
gp	glycoprotein
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
HWE	Hardy-Weinberg equilibrium
IFN	interferon
IN	integrase
KIR	killer cell immunoglobulin-like receptors

LTNP	Long term non-progressor
LTR	Long terminal repeat
MAF	minor allele frequency
MD	molecular dynamics
MHC	Major Histocompatibility Complex
Nef	negative factor
NGS	next generation sequencing
NK	natural killer cells
PBMC	peripheral blood mononuclear cells
Pol	polymerase
Rev	regulator of expression of virion proteins
RNA	Ribonucleic acid
RT	reverse transcriptase
SFU	spot forming unit
Tat	transactivator of transcription
TCR	T cell receptor
Vif	viral infectivity factor
Vpr	viral protein R
Vpu	viral protein U
WES	whole exome sequencing
WHO	World Health Organization

ABSTRACT

Sub-Saharan Africa remains the global epicenter of the HIV epidemic; more than 70% (25 million) of all people living with HIV/AIDS are in Africa. To end the global epidemic, finding an effective HIV vaccine remains a top priority. A small fraction of people including children infected with HIV do not progress to disease (AIDS) for 10 or more years without antiretroviral therapy – the long-term non-progressors (LTNPs). HLA class I molecules are critical in the elimination of HIV infected cells through presentation of endogenously processed antigens to CD8+ cytotoxic T lymphocytes (CTL). Understanding what and how these genetic factors control viremia provides opportunities for development of new HIV vaccines and immunotherapies.

In this thesis, we first focused on estimating how many children in Uganda and Botswana infected with HIV are classified as LTNPs and the factors associated with progression. We found that one in every 16 children infected with HIV is a LTNP. Additionally, being underweight at baseline, having been enrolled after 2010 into care and being from Botswana were associated with faster progression to disease. We demonstrate that these associations remain the same when we include a baseline HIV viral load measurement in the model.

Next, we examined the HLA class I alleles associated with LTNP among children from Botswana and Uganda. We found that certain HLA class I alleles are over-represented in HIV infected children compared to the uninfected populations. In addition to the canonical alleles HLA-B*57:03 and B*58:01, we documented a novel HLA-C*03:02 allele that was associated with LTNP, and that its protective effect is additive and independent of linkage disequilibrium. Together these alleles account for 16.5% of the variation in LTNP status.

Finally, we focused on elucidating the mechanisms through which HLA-C*03:02 controls HIV infection. We documented that HLA-C*03:02 prefers to present peptides derived from structural HIV proteins, and more specifically a Gag GY9 peptide located in the p17 matrix protein. Positions E62, T142 and E151 in the HLA-C*03:02 peptide binding groove and positions p6 in GY9 are crucial in shaping the immune response and supposedly account for HIV immune escape pathways. Lastly, we demonstrated that the GY9 peptide elicits a clade-specific HLA-C*03:02-mediated response in one third of individuals carrying the allele.

In conclusion, these results support a role for HLA-C molecules in HIV control and provide prospects for development of new vaccine and immunotherapies targeting CTL responses.

CHAPTER 1: INTRODUCTION

1.1 Background

In 2022, 1.5 million children (<15 years) were estimated to be infected with HIV, of whom 1.33 million are living in Sub-Saharan Africa the epicenter of the global HIV epidemic (*UNAIDS DATA*, 2023). Globally, more than 84,000 deaths due to AIDS are reported among children aged 0-14 years, and 69,000 were registered in Sub-Saharan Africa (*UNAIDS DATA*, 2023). Although significant progress in the reduction of new infections and HIV-related deaths has been achieved, HIV/AIDS and its associated co-morbidities remain a substantial cause of mortality and morbidity. The Eastern and Southern African regions have the highest coverage (more than 90%) of the prevention of mother-to-child transmission (PMTCT) of HIV. Nonetheless, the area maintains the highest HIV transmission rates among infants (9.9% of all births) and 109,000 new infections in children in 2022 (*UNAIDS DATA*, 2023). Precisely, it is further estimated that in Botswana and Uganda, 340,000 and 1.4 million people living with HIV, while <200 and 5,900 new HIV infections and <200 and 3,800 AIDS-related deaths occur annually among children, respectively (*UNAIDS DATA*, 2023).

HIV naturally progresses to AIDS and eventually death. The frequently reported predictors of progression and death include $CD4^+$ T cell count, viral load and HAART (Eller et al., 2015; Helleberg et al., 2013; Helm et al., 2014; Iyun et al., 2020; Jiang et al., 2013; Shoko & Chikobvu, 2019). Other significant predictors are age, gender, race, year of diagnosis, geographical location, mode of infection, use of hallucinogens, and alcohol consumption (L. Chen et al., 2015; Del Amo et al., 1998; Hahn & Samet, 2010; Jiang et al., 2013; Vittinghoff et al., 2001). Clinical predictors include total lymphocyte counts, WHO stage, anaemia and all forms of malnutrition, especially in sub-Saharan Africa (Antonio George Lentoor, 2018; Del Amo et al., 1998; Helleberg et al., 2013; Matheron et al., 2003; Sunguya et al., 2011). Similarly, co-morbidities such as tuberculosis, pneumocystis pneumonia, hepatitis C, dementia, cardiovascular disease, and cancer are significant contributors to excess mortality (Del Amo et al., 1998; Greub et al., 2000; Helleberg et al., 2013; Nakku et al., 2013). The year of diagnosis involves accounting for unmeasured confounders, encompassing clinical practices such as adherence to WHO treatment guidelines, metrics reflecting enhancements in health indices such as improved healthcare accessibility and more effective health communication messages,

and progress in medical treatments and diagnostics, including advancements in early infant diagnosis. Thus, integrating the year of diagnosis as a multivariable proves instrumental in capturing the comprehensive context and temporal dynamics that wield substantial influence over the disease's trajectory (Altmann et al., 2012; Babiker et al., 2002; A. Liu et al., 2020).

The rate of clinical progression of HIV varies significantly among populations. One extreme clinical phenotype has been well described, where 1–5% of all HIV infections do not progress to AIDS for more than ten years in the absence of ART(Sabin & Lundgren, 2013). These long-term non-progressors (LTNPs) possess a high degree of control of HIV infection with viral suppression and normal-for-age $CD4^+$ T cell count (above 500 cells/mL). (Sabin & Lundgren, 2013) Nevertheless, most studies have been conducted on adults in Western and African populations, and the estimates in children remain unclear. Recently, a rare group HIV infected individuals exhibiting persistent virologic suppression over extended periods following the cessation of ART, referred to as post-treatment controllers (PTCs) (Namazi et al., 2018). They have been shown to have a distinct immunological profile with markedly reduced activation in both CD4+ and CD8+ T cells, diminished exhaustion in CD4+ T cells, and heightened Gag-specific CD4+ T cell and natural killer (NK) cell responses (Etemad et al., 2023). These populations offer opportunities to explore natural host mechanism of viral control towards an HIV cure.

Host genetic variations have been described to influence the rate of HIV disease progression significantly; however, there are scarce genetic studies in Africa, and the available literature is controversial (Fellay et al., 2007; McLaren & Carrington, 2015; Pelak et al., 2010; E. Trachtenberg & Erlich, 2001). Conversely, data concerning host genetic factors that are important to the rate of disease progression among children infected with HIV in Africa is sparse (Aouizerat et al., 2011; Chatterjee, 2010; K. K. Singh & Spector, 2009). The human major histocompatibility complex (MHC), known as the human leukocyte antigen (HLA), is a highly polymorphic locus composed of 220 genes on chromosome 6p21 (Shiina et al., 2017). The HLA class I molecules predominantly present endogenously processed antigens to cytotoxic $CD8^+$ T lymphocytes (CTL) in a cell-mediated immune response (Shiina et al., 2017). In addition, class I molecules play a significant role in the innate immune response through interactions with natural killer (NK) cell receptors via the killer cell immunoglobulin-like

receptors (KIRs) (Parham et al., 2012). This interaction puts class I molecules at the epicenter of HIV control.

Several studies corroborate the role of HLA-B*57, B*27, B*35, B*58 and some HLA-C alleles in HIV LTNP in Caucasians and African Americans (Fellay et al., 2007; Frater et al., 2007; Jin et al., 2002; Migueles et al., 2000; Ngumbela et al., 2008; Pelak et al., 2010). However, this association has not been consistently replicated in African populations. For example, Payne et al. found that in Botswana, the protective effect of HLA-B*57 and B*58:01 was absent among adult HIV LTNPs (Payne et al., 2014). In contrast, HLA B*57 and B*58:01 were reported to be consistent with the slow progression of HIV in Ugandan adults. But, this study was based on a small sample size, and the participant characteristics did not fulfil our stringent classification of LTNP (Serwanga et al., 2009). Some alleles such as HLA-B*51:01 have been reported to be both protective or susceptible in different populations (J. M. Carlson, Listgarten, et al., 2012; Kawashima et al., 2010).

HLA class I allele homozygosity is associated with faster progression to AIDS while heterozygosity confers protection (Naruto et al., 2012); however, homozygosity is very low in Africa (Cao et al., 2004). Heterozygosity reflects the broader antigen-processing repertoire needed to control a rapidly changing virus. But most of these studies have been conducted among adult Caucasian and African populations, while few studies have evaluated the role of HLA in African pediatric populations (Mehra et al., 2011; E. Trachtenberg & Erlich, 2001).

The highly polymorphic HLA genes produce a repertoire of molecules that recognize and bind specific antigens. The peptide binding characteristics of HLA class I alleles vary widely depending on the circulating HIV subtypes (P. J. R. Goulder & Walker, 2012). For example, protective alleles HLA-B*57:01, B*58:01 and B*27:05 prefer to present Gag-derived p24 peptides. In contrast, susceptible alleles HLA-B*35:01 and B*53:01 preferentially present Nef-derived viral peptides (Borghans et al., 2007). Similarly, HLA-A*2 binds a restricted set of HIV peptides derived from reverse transcriptase (RT), p17 and glycoprotein (gp) gp41 that elicit a strong CTL response and demonstrate the occurrence of immunodominance (Peter et al., 2001). However, accumulation of escape mutations in the epitopes binding to protective HLA class I alleles may abrogate the protective phenotype, presumably via defective functional antigen presentation mechanisms (Borghans et al., 2007; P. J. R. Goulder & Walker, 2012;

Montesano et al., 2014; Watanabe et al., 2011). Other HLA class I-mediated mechanisms of HIV control include HLA surface expression, peptide stabilization kinetics, T cell receptor clonotype, and peptide presentation via KIR ligands for NK cell cytotoxicity (Apps et al., 2013; Celik et al., 2016; Motozono et al., 2014; Parolini et al., 2018; Ramsuran et al., 2018). However, these efforts have not successfully translated into novel therapeutics or vaccines for HIV (Haynes & McElrath, 2013). The recent emergence of novel genomics and bioinformatics tools has enabled cheaper and more accurate mining of HLA data from next-generation sequencing (NGS) data in large populations (Bauer et al., 2018; Klasberg et al., 2019; Major et al., 2013). Harnessing these technologies redefines access to HLA typing. Therefore, this offers an opportunity to revisit these studies, especially in a population where fewer studies have been done (Luo et al., 2015).

Outstandingly, the maturing immune responses of children significantly differ from the adult populations, and this impacts the response to infections as well as design of novel HIV vaccines and therapeutics (Prendergast et al., 2012; Simon et al., 2015; Tobin & Aldrovandi, 2013, 2014). Fundamentally, there are differences in the expression and functional profiles of immune cells between children and adults (Prendergast et al., 2012; Simon et al., 2015). Therefore, we must elucidate the host genetic immune factors and mechanisms associated with pediatric HIV disease LTNP.

Therefore, this Ph.D. thesis aimed to characterize pediatric HIV disease progression (frequency of LTNP and factors associated with LTNP), determine the HLA class I alleles associated with LTNP in Botswana and Uganda, and subsequently the structural and functional mechanisms of novel alleles. The results define the protective HLA class I alleles in HIV disease progression that can be harnessed for potential novel HIV vaccines and immunotherapeutic designs (Lucas et al., 2016). These studies provide essential preliminary data to support further exploration of these models for basic and translational HIV research in sub-Saharan Africa.

1.2 Problem Statement

Despite introducing effective ART over 25 years ago, HIV remains at epidemic proportions in Sub-Saharan Africa (*World Health Statistics 2018: Monitoring Health for the SDGs, Sustainable Development Goals*, 2018). Developing an effective and safe HIV vaccine remains

a global health priority (Haynes & McElrath, 2013). There are 1.2 million children living with HIV; 95,000 AIDS-related deaths and 120,000 new infections occur in children annually. However, a small fraction of children infected with HIV naturally control the infection for more than ten years without ART; the LTNPs and elite controllers (Sabin & Lundgren, 2013). The study of these distinct sub-populations within HIV infection presents potential for uncovering novel biological mechanisms implicated in HIV control. This knowledge could be harnessed to inform the development of innovative immunotherapeutic strategies. Moreover, with the global adoption of the test-and-treat strategy, the phenomenon of LTNPs and elite controllers is anticipated to gradually diminish over the course of a few years.

Several studies have identified the HLA class I genetic locus as both protective and/or susceptible to HIV disease progression (Tshabalala et al., 2013). But, outcomes from these heavily Caucasian studies have not been consistently replicated in different populations. For example, in European populations, HLA-A*11, A*27, B*32, B*35, B*27, B*55, B*56, B*57 and C*18 alleles are frequently associated with LTNP, but these alleles are very rare or absent in Zambia, Uganda and Botswana (Kijak et al., 2009; Munderi et al., 2011; Novitsky et al., 2001; Tang et al., 2010). In a Japanese population, the known protective alleles B*27 and B*57 are virtually absent, but HIV restriction seems to be mediated by B*52:01 and the B*52:01-C*12:02 haplotype (Naruto et al., 2012). There is renewed interest in HLA-restricted epitope vaccine designs and immunotherapeutics with some promising results for toxoplasmosis, tuberculosis and cancer in recent murine experimental trials (Aspord et al., 2010; El Bissati et al., 2016; Kovjazin et al., 2013; W. Li et al., 2014; Purcell et al., 2007; Zhao et al., 2013). Nonetheless, the established AIDS-protective alleles and inherent heterogeneity contribute to only a modest fraction, ranging from 10% to 40%, of the overall spectrum of HIV resistance (Carrington et al., 1999; Pelak et al., 2010). Therefore, new genetic factors (HLA class I alleles) and epitopes that bind promiscuously or specifically to HLA molecules and induce strong CTL responses in the most affected populations are urgently needed for a broad HIV vaccine (Chakraborty et al., 2014; Gartland et al., 2014; Ogunshola et al., 2018).

In this thesis, we first characterize pediatric HIV disease progression in the African population and identify the risk factors of clinical progression. We then identify HLA class I alleles and

haplotypes that are associated with HIV resistance and susceptibility. Finally, we investigate structural and functional mechanisms of HLA-mediated HIV control.

1.3 Aim

This thesis aimed to characterize the progression of pediatric HIV disease in Uganda and Botswana, investigating the association between HLA class I alleles, haplotypes, and the mechanisms through which they influence LTNP within these populations.

1.4 Specific Objectives

1) To determine the frequency of LTNP and factors associated with pediatric HIV disease progression in Uganda and Botswana.
2) To determine the HLA class I alleles and haplotypes associated with HIV disease progression in Botswana and Ugandan African pediatric populations.
3) To describe the structural and functional mechanism of the protective HLA class I molecules.

1.5 Research Question

1) What is the frequency of LTNP, and factors associated with pediatric HIV disease progression among cohorts in Uganda and Botswana?
2) What HLA class I alleles and haplotypes are associated with pediatric HIV disease progression in two African pediatric populations?
3) What are the structural and functional mechanisms of the protective HLA class I molecules?

1.6 Study Justification

Despite an abundance of literature on the role of HLA with HIV disease progression, there are several reasons to re-evaluate these studies towards the last frontier of rolling back the HIV epidemic (P. J. R. Goulder & Walker, 2012; Kulpa & Collins, 2011; Walker et al., 2010; Zipeto & Beretta, 2012). These studies can potentially identify HLA risk and protective alleles associated with clinical progression (E. Trachtenberg & Erlich, 2001). Such genetic markers can have several roles in HIV disease, such as risk assessment for prioritization of ART and

development of novel therapeutics and vaccines (Aspord et al., 2010; El Bissati et al., 2016). The HLA genetic map has grown due to improved sequencing and genomics technologies. Since it was first discovered, the HLA class I genomic region has been fine mapped to 15,586 alleles in 2018 (Marsh et al., 2010; Robinson et al., 2015). This implies that alleles previously associated with the disease could change, or novel alleles could be discovered due to the improved understanding of the HLA genetic variability (Robinson et al., 2015). The microevolution of HIV leads to the emergency of escape mutants, especially within the HLA-restricted epitopes, which annuls the protective effect of alleles (Moore et al., 2002). In Uganda, HIV clade A/D recombinants are now responsible for an estimated 25% of HIV infections (Ssemwanga et al., 2013). Moreover, Conroy and colleagues found that the prevalence of HIV recombinant viruses changed by 1.2%, while A increased by 6.6% and D reduced by 7.8% over an 8-year period (Conroy et al., 2010). Therefore, the proposed study offers an opportunity to conduct a more extensive and comprehensive HLA class I and HIV disease progression study utilizing the latest allele database and bioinformatics tools (Robinson et al., 2015). An advantage of studying pediatric HIV cohorts is the unambiguous determination of the date of infection compared to adults (Laeyendecker et al., 2009).

A maturing pediatric immune response's expression and functional profiles differ from adults (Tobin & Aldrovandi, 2013). For example, while children experience a higher viremia they are also associated with higher levels of immune activation compared to HIV infected adults (Kourtis et al., 2000). This inverse relationship contributes to faster progression in children, presumably by increasing the number of CD4 molecules that facilitate HIV viral attachment and entry (Roider et al., 2016). However, data from Uganda shows that children infected with HIV have lower immune activation levels compared to adults (Ssewanyana et al., 2007), this would affect the balance for the required effective anti-HIV activity. Pediatric LTNPs have been demonstrated to maintain low levels of immune activation despite a high viremia and normal CD4 count (Roider et al., 2016). Also, HIV-mediated downregulation of HLA expression of mainly HLA-A and B molecules would suggest a potential role of HLA-C molecules, especially via the NK cell pathways. In fact NK cell cytotoxic activity against HIV-infected cells neonates is similar to that in adults (Jenkins et al., 1993). In addition, neonatal NK cells have been demonstrated to inhibit HIV via non-cytolytic mechanisms such as chemokine production at greater proportions compared to adults (Bernstein et al., 2004). This

data suggests a potential role of other HLA genes than the canonical B genes. It is therefore proposed that innate immune responses are more active in pediatric HIV and could account for the control of HIV. Therefore, understanding the role natural immunity plays in this population can contribute to the developing of novel therapeutics or vaccine design studies.

1.7 Theoretical Model

It is well established that HIV disease progression occurs at two clinical phenotypic extremes of LTNP and rapid progression (Sabin & Lundgren, 2013). Therefore, LTNP is associated with protective HLA class I alleles in pediatric African populations that differ from others. this is premised on three theoretical bases.

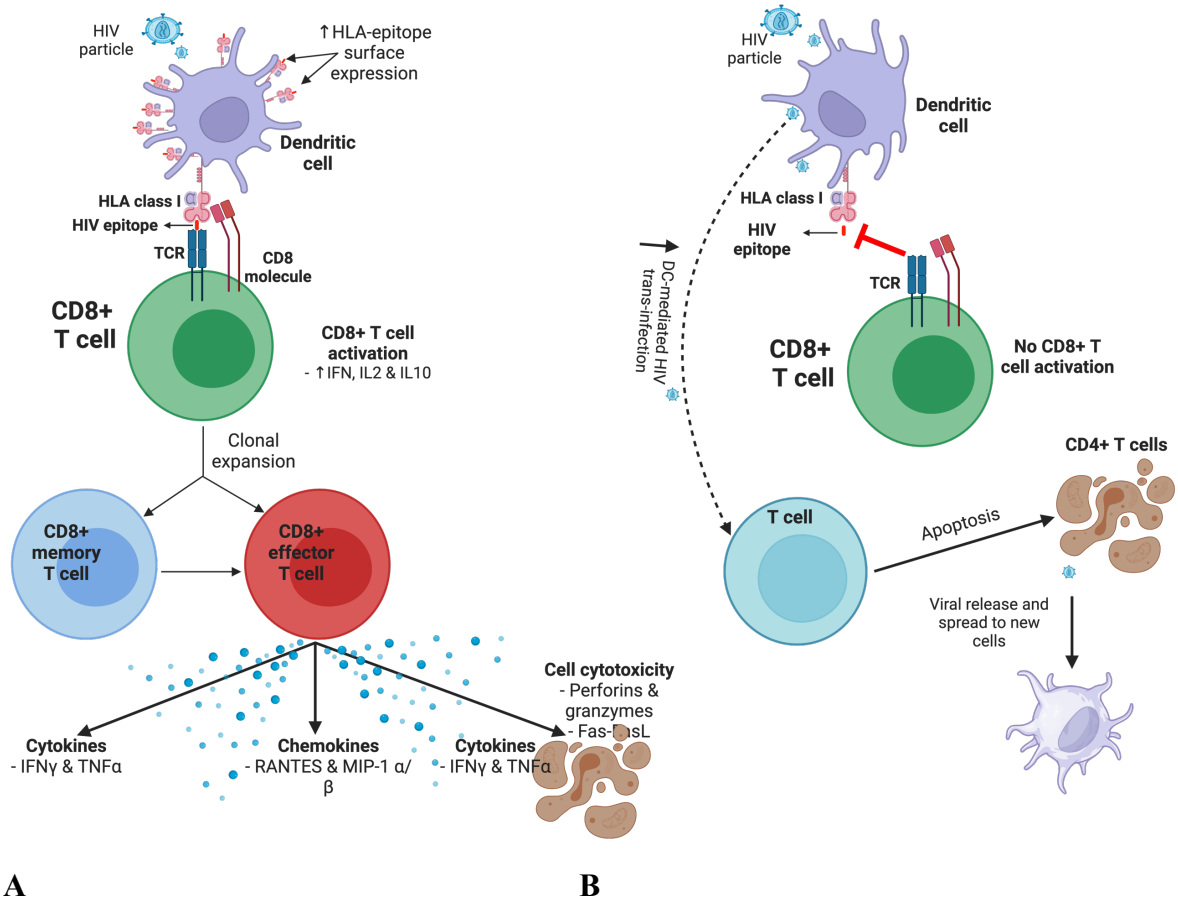
Transmission of pre-adapted virus from the maternal infection. Children inherit a copy of alleles from the mother and father. Pre-adaptation would mean they would use alternative alleles for viral control. Adland and colleagues, demonstrated in a cohort of mother-child pairs in South Africa, that the protective HLA alleles had a significant impact on the viral replicative capacity in adults but this was not present in the children (Adland et al., 2015).However, this study further showed that children expressing protective allele (B*57, B*58 & B*81) progressed faster to AIDS despite the absence of escape mutants. This would suggest that HLA protective effects observed in children may be via the non-classical mechanisms such as the engagement of HLA-C molecules with NK cells via KIRs.

Secondly, there are differences in expression of class I and II HLA molecules on the cell surface (Puri et al., 1993) between neonates and adults including expression of various other molecules such as TLR (Hikita et al., 2019). However, in adult studies the role of HLA in protective effects has been closely related to their cellular expression levels (Yarzabek et al., 2018). Therefore, in the presence of lower expression of HLA would premise that this effect may not be applicable during infant HIV infection. In fact, there is evidence to show that human extra villous trophoblasts and other placental tissue express HLA-C on their surface but not HLA-A or B (Apps et al., 2009; King et al., 2000).

Thirdly, HIV-mediated downregulation of HLA expression of mainly HLA-A and B molecules, would suggest a potential role of HLA-C molecules, especially via the NK cell pathways. In fact NK cell cytotoxic activity against HIV-infected cells neonates is similar to

that in adults (Jenkins et al., 1993). In addition, neonatal NK cells have been demonstrated to inhibit HIV via non-cytolytic mechanisms such as chemokine production at greater proportions compared to adults (Bernstein et al., 2004).

These alleles have structural and functional distinctiveness that enhances immunological viral clearance and thus control or susceptibility to HIV infection and progression (Figure1-1). Protective HLA class I alleles induce a distinct $CD8^+$ T cell immune response at both a structural and functional level that leads to clearance of virally infected cells leading to a significant delay in HIV disease progression in African children.



A **B**

Figure 1-1: Theoretical model for pediatric HIV disease and HLA Class I-mediated LTNP and RP

A) Predicted model for Protective alleles. The ltnpHLA dendritic cells (DC) expresses more cell surface HLA molecules and HLA-HIV antigen peptide complexes. The ltnp allele is able to present HIV peptides to the $CD8^+$ TCR. The $CD8^+$ T cell is activated with clonal expansion into effector and memory cells

that lead to clearance of infected cell thus, control of HIV; LTNP. B) Predicted model for Susceptible alleles. The rpHLA-DC express an unreliable HLA molecule therefore no stable HIV peptide binding and cell surface expression and thus no successful $CD8^+$ TCR engagement. This favors HIV replication and DC-mediated trans-infection. This unregulated process leads to rapid decline of T cells thus leading to rapid HIV disease progression. Created with BioRender.com

CHAPTER 2: LITERATURE REVIEW

2.1 Epidemiology of HIV

It is estimated that 39 million people live with HIV globally, of which 37.5 million are adults and 1.5 million are children under 15 years (*UNAIDS DATA*, 2023). In 2022, 1.3 million new cases of HIV were reported, of which 130,000 were children under 15 years (Figure 2-1). While 630,000 AIDS-related deaths were recorded, among whom 84,000 were reported to be children (*UNAIDS DATA*, 2023).

Sub-Saharan Africa remains the epicenter of the HIV epidemic with the highest prevalence of 25.6 million infections, accounting for more than 65% of the global burden (*UNAIDS DATA*, 2023). Regionally, Eastern and Southern Africa tend to be the most affected globally. Among children, 109,000 new cases of HIV occurred in the region in 2022, while 69,000 AIDS-related deaths were recorded in the same year (*UNAIDS DATA*, 2023).

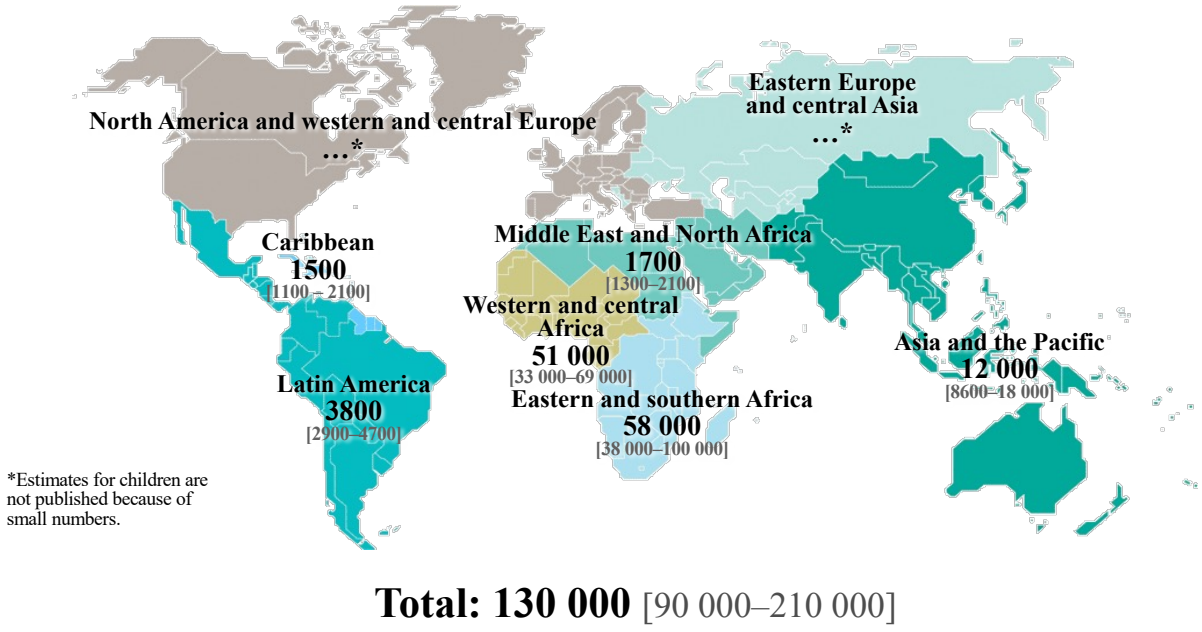


Figure 2-1: Estimated number of children (<15 years) newly infected with HIV.

Adopted from UNAIDS 2023 core epidemiology slides (UNAIDS DATA, 2023)

HIV is transmitted in humans in different ways, including sexual transmission, blood transfusions, intravenous drug use, contaminated needles and vertical transmission via the mother-child (Becerra et al., 2016; Fettig et al., 2014). The global HIV epidemic is predominated by HIV type 1, while type 2 is confined to West Africa with outbreaks in Angola, Mozambique and a few cases in Europe (GIRARD et al., 2006). The global HIV epidemic is also characterized by a large genetic variability subdivided into four groups: M (major), O (outlier), P (pending) and N (non-M, non-O) (Faria et al., 2014). Several subtypes or clades are spread across the globe, and these are typically confined to different continents(Becerra et al., 2016). The pandemic is predominantly of the M group, which is sub-divided into nine clades: A, B, C (which dominates Sub-Saharan Africa), D, E, F, G, H, J and K (Becerra et al., 2016).

2.2 HIV Structure

HIV belongs to the genus Lentivirus and family Retroviridae, with positive polarity ribonucleic acid (RNA) (Kuyl & Berkhout, 2012). They encode reverse transcriptase polymerase, an enzyme synthesizing DNA from the viral template RNA (Kuyl & Berkhout, 2012). Retroviruses can insert their genetic DNA into the genome of their host cell using its integrase enzyme (Kuyl & Berkhout, 2012; Watts et al., 2009). Retroviruses encode structural proteins Gag, Pol (polymerase) and Env (envelope) (Kuyl & Berkhout, 2012; Watts et al., 2009).

The HIV genetic material is carried as two identical single-stranded RNA strands of 9.8 Kb located inside a capsid (Watts et al., 2009). The genome is composed of 9 genes divided into three major groups. The structural genes are *gag*, *pol* and *env*, regulatory genes *tat* and *rev* and four accessory genes, *vpr*, *nef*, *vif* and *vpu,* encode up to 15 viral proteins (Kuyl & Berkhout, 2012). The RNA strands are closely associated with several viral proteins, p7, that envelop the genome (Arhel, 2010). The genome is retained in the capsid, consisting of the viral protein p24 and viral enzymes: RT, integrase (IN), and protease (PR) (Arhel, 2010; Watts et al., 2009). The nucleocapsid is surrounded by a viral matrix made of viral protein p17 (Arhel, 2010; Watts et al., 2009). The virus's envelope is formed from the lipid bilayer of the host cell's cytoplasmic membrane, acquired during the budding of viral particles on the cell surface (L. Carlson et al., 2010). The lipid envelope is studded with two viral glycoproteins, gp41 and gp120, essential for attachment to host cells (Bartha et al., 2013).

The gag gene codes for the internal structural proteins of the viral particle (Arhel, 2010; Bartha et al., 2013; Kuyl & Berkhout, 2012; Watts et al., 2009). The pol gene codes for three viral enzymes, the RT, IN and PR (Arhel, 2010; Bartha et al., 2013; Kuyl & Berkhout, 2012; Watts et al., 2009). Finally, the *env* gene produces the outer proteins of the virus envelope (Arhel, 2010; Bartha et al., 2013; Kuyl & Berkhout, 2012; Watts et al., 2009). These protein products are synthesized in the form of a polyprotein cleaved by the viral PR (Arhel, 2010; Bartha et al., 2013; Kuyl & Berkhout, 2012; Watts et al., 2009). The infectious particle of HIV also contains viral proteins categorized as auxiliary or accessory proteins. These accessory proteins include viral infectivity factor (Vif), viral protein U (Vpu), negative factor (Nef), viral protein R (Vpr), transactivator of transcription (Tat) and regulator of expression of virion proteins (Rev), which affect the virus's ability to infect, replicate and kill host cells (Arhel, 2010; Watts et al., 2009). The viral genome contains two long terminal repeat (LTR) at both ends of the genome required for viral genomic integration and control of the production of new viral particles (Arhel, 2010).

2.3 HIV Replication Cycle

The HIV infectious replication cycle starts with viral attachment to a susceptible host cell. The HIV envelope protein gp120 is the primary particle that binds the CD4 receptor expressed primarily on the surface of helper T cells, monocyte-macrophages and dendritic cells (Kijewski & Gummuluru, 2015; Z. Liu et al., 2017). Viral entry is co-facilitated by chemokine receptors such as CXCR4 and CCR5 (Adamson & Freed, 2010; Bracq et al., 2018; Z. Liu et al., 2017). Once the virus enters the cell, the viral particles are unpacked and emptied into the host cell cytoplasm. HIV has proteins which are physically associated with the viral RNA particles and have enzymatic function (Adamson & Freed, 2010). Viral entry is followed by the process of reverse transcription of single-stranded viral RNA single strand to double-stranded DNA using the RT enzyme. Reverse transcriptase is directly linked to RNA strands for transcription of double-stranded genomic DNA from RNA, also referred to as the provirus (Abdelrahman et al., 2015; Adamson & Freed, 2010). The double-stranded viral DNA is then integrated randomly into a host cell's chromosome using the viral integrase (IN) enzyme (Adamson & Freed, 2010; Anderson & Maldarelli, 2018).

Depending on the cellular circumstances, the virus may enter a lytic cycle with active production of new virions or enter a phase of latency (Anderson & Maldarelli, 2018). Latency

is characterized by a very weak transcription of viral genes and, therefore, low viral production (Anderson & Maldarelli, 2018). HIV latency is dependent on the level of activation of the host cell, as well as the integration of the provirus into the genome of the host cell (Anderson & Maldarelli, 2018). During active viral replication, the transcription of viral genes is activated by promoter sequences localized in the 5 'LTR of the viral genome (Adamson & Freed, 2010). Several types of messenger RNA (mRNA) are produced sequentially following trans-activation of the LTR (Adamson & Freed, 2010). Host cell machinery is used to translate viral mRNA into long precursor protein strands, which are later cleaved into their mature structural and enzymatic forms, by viral PR, either during or following release from the host cell (Adamson & Freed, 2010). Release of the new virions occurs by budding from the host cell and consequently acquires host membrane phospholipids and proteins (Meng & Lever, 2013; Sundquist & Krausslich, 2012). This is an important mechanism in evading the host immune response because it allows the virus to disguise itself as a host cell. Finally, the mature virion can further establish the systemic infection of the host (Meng & Lever, 2013; Sundquist & Krausslich, 2012).

The HIV genome encodes for several regulatory and accessory gene products. Tat binds to the viral LTR and sets off further viral transcription (Adamson & Freed, 2010). Rev inhibits mRNA splicing as it accumulates through interaction with alternate splicing factor/splicing factor 2 and p32 (Adamson & Freed, 2010). Vif, Nef, Vpu and Vpr regulate aspects of HIV pathogenesis such as infectivity, viral release, downregulation of MHC and viral transcription (Adamson & Freed, 2010).

2.4 Natural History and Pathogenesis of HIV

The first target cells during infection seem to be macrophages, more commonly in sexual transmission than vertical (Becerra et al., 2016). In addition, there is evidence of various cell types that could potentially play a role in the transmission, including dendritic cells (Kijewski & Gummuluru, 2015). This is followed by a 3-6 week period of primary infection, characterized by a peak in viremia and spread throughout the body (Becerra et al., 2016; Levy, 2007). About half the patients present with symptoms of mononucleosis or flu-like symptoms (WHO stage I). This peak viremia is characterized by a paralleled decline in the $CD4^+$ cells in blood (Becerra et al., 2016; Levy, 2007). Because the primary HIV receptor on target cells is

the CD4 molecule, $CD4^+$ helper T cells are the principal cells infected with HIV in vivo (Becerra et al., 2016; Levy, 2007). The virus is rapidly localized to the regional lymph nodes, which favors viral replication, partly because they possess high $CD4^+$ helper cells (Becerra et al., 2016; Levy, 2007).

Within a month following primary infection, the immune system of infected people responds by developing a specific humoral and cellular immune response against HIV (Becerra et al., 2016; McMichael et al., 2010), with the detection of antibodies against HIV in the serum of patients by the 12th week (McMichael et al., 2010). The cellular reaction, for its part, is faster than the humoral. The cellular response correlates very well over time with decreasing peak serum viremia (Becerra et al., 2016; Levy, 2007). This synchronism of the two phenomena suggests that lysis of HIV-infected cells by specific CTLs would directly cause the sudden decline in blood viremia (Becerra et al., 2016; Levy, 2007).

Following primary infection, a highly variable clinical latency phase is established. This period is characterized by an absence of symptoms and opportunistic infections. However, viral replication remains active mainly in the lymph nodes (Becerra et al., 2016; Levy, 2007). This period is characterized by a variable decline in $CD4^+$ T cells that heralds the development of AIDS. In the advanced stages of HIV infection ($CD4^+$ T cells <200 cells/mL), the virus overcomes the immune system, and blood viremia increases significantly (Becerra et al., 2016; P. J. Goulder et al., 2016; Levy, 2007). The mass loss of $CD4^+$ T cells in the patient (which can go below 50 cells/mL) and the destruction of the lymph nodes' architecture incapacitate the immune system from fighting opportunistic infections. The individual in the AIDS phase is then subject to developing many opportunistic infections that a healthy individual typically controls. The variable loss of $CD4^+$ T cells correlates with the development of various clinical entities classified in ascending severity as WHO stage II, III and IV. The patient dies from these multiple infections after a highly variable period from primary HIV infection (Becerra et al., 2016; P. J. Goulder et al., 2016; Levy, 2007).

2.5 Natural History of HIV Disease Progression in Children

Unlike adults, the natural history of children infected with HIV has some differences. Children are generally infected perinatally, and three routes have been identified including transplacental

(in utero), intrapartum and breast feeding (De Cock et al., 2000). Without breastfeeding, 30% of infections occur in utero while a large proportion of infections are accounted for by the intrapartum (De Cock et al., 2000). Breastfeeding increases the risk of mother-to-child transmission by up to 20% and accounts for 40% of vertical infections (Iliff et al., 2005). Various factors affect transmission of HIV to children including access to ART, education levels, lack of antenatal care, feeding practices, absence of infant ARV prophylaxis, maternal CD4 cell count and socio-cultural dynamics (Iwelunmor et al., 2014) and these vary from region to region (Embree et al., 2000; Jean et al., 1999; Kassa, 2018; Okoko et al., 2017).

HIV generally progresses faster in children compared to their adult counterparts with less than 36% being AIDS-free by two years following infection (Jean et al., 1999; Warszawski et al., 2007). This immediate reason is the infection occurring during a maturing immune system (Prendergast et al., 2012). Therefore, children that are infected in utero are not capable of mounting an effective immune response and progress to disease and death much faster. Compared to adults, children have a higher levels of viremia, that could contribute to their faster progression (Muenchhoff et al., 2014), and the levels take a much longer time to reach the steady state (McIntosh et al., 1996). The reasons responsible for these differences remain largely unexplained except for hypothesizing the maturing immune system.

The placenta acts as a barrier to maternal HIV infection, however, maternal infections such as malaria that affect the placental permeability increase in utero infections, and these further affect transfer of maternal IgG (de Moraes-Pinto et al., 1998; Kumar et al., 2012). However, the exact mechanism of viral infection remains unclear. HIV infection has been linked to the Hofbauer cells which are placental villous macrophages of fetal origin that are present throughout pregnancy (Backe et al., 1992). Hofbauer cells have been demonstrated to express both CD4 and HIV coreceptors DC-SIGN, CCR5, and CXCR4 (Johnson & Chakraborty, 2012) that may facilitate transmission.

It is unclear whether the route of infection has an impact on the rate of progression of HIV in perinatal infections. Nakamura and colleagues, found that there were differences in the molecular signature of transmitted viruses by either the in utero or breast feeding routes (Nakamura et al., 2017). They showed that in utero transmission favors variants with shorter, less-glycosylated V1 loops resistant to soluble CD4 neutralization, while breastfeeding

transmission selects for variants with fewer potential glycosylation sites in gp41. This would suggest that in utero transmission is adapted to transmission of more virulent strains and thus could hypothetically show faster progression of disease. Additionally, for effective production of infection, a high concentration of $CD4^+$ receptors are required, however, the oral mucosa lined by squamous epithelium has been demonstrated to have fewer $CD4^+$ receptors compared to the rectal/sigmoid and endocervical surfaces lined by columnar epithelium (Patyka et al., 2015). Therefore, in utero transmission may be associated with a high infection dose that may translate to faster HIV progression. These dynamic processes in the natural history of HIV may possess pathways that are necessary for the development of novel therapies against HIV.

Unlike adults, plasma viremia in children infected with HIV does not correlate with disease severity; Shearer and colleagues found that there was no significant difference in the viral load among rapid and non-rapid progressors in the first 6 months of life or infection (Shearer et al., 1997). This would suggest that mechanisms responsible for peadiatric HIV control may be distinct from their adult counterparts.

2.6 Clinical and Demographic Predictors of HIV disease progression

The most commonly reported predictors of clinical HIV disease progression are $CD4^+$ T cell count, viral load and HAART (Eller et al., 2015; Helleberg et al., 2013; Helm et al., 2014; Iyun et al., 2020; Jiang et al., 2013; Shoko & Chikobvu, 2019). In addition to these relatively universal factors, geographical, socio-economic, and behavioral differences provide additional factors that predict progression. Several factors have been identified in various prospective cohorts, retrospective cohorts, and cross-sectional studies reported worldwide. Age (L. Chen et al., 2015; Del Amo et al., 1998), gender, race or ethnicity (Del Amo et al., 1998), level of education (Jiang et al., 2013), marital status, year of diagnosis (L. Chen et al., 2015; Del Amo et al., 1998), country or geographical location (L. Chen et al., 2015) are frequent demographic factors identified. Particularly, being very young (<3 years), age >45 years and male are commonly associated with faster progression to AIDS and death (Del Amo et al., 1998; Little et al., 2007; Shoko & Chikobvu, 2019; Vittinghoff et al., 2001). Behavioral and modifiable risk factors include route or mode of infection (Jiang et al., 2013) such as homosexual or heterosexual transmission (L. Chen et al., 2015; Helm et al., 2014), injection drug use,

hallucinogens (Vittinghoff et al., 2001), alcohol(Hahn & Samet, 2010) and sex workers (Peterson et al., 2013).

In addition to $CD4^+$ T cell count and viral load, other clinical predictors of progression and death include total lymphocyte counts (Helleberg et al., 2013), WHO stage (Del Amo et al., 1998; Matheron et al., 2003), and hematological abnormalities such as anemia (De Santis et al., 2011). Acute and chronic malnutrition is associated with high under-five and HIV-related mortality in sub-Saharan Africa (Antonio George Lentoor, 2018; Sunguya et al., 2011). Both macro- and micronutrient deficiency have consequences such as immune dysfunction, neurodevelopmental disorders, and multiorgan dysfunction, especially in early childhood (Drotar et al., 1999; Koethe & Heimburger, 2010; Mody et al., 2014). Maternal HIV and undernutrition are associated with poor birth outcomes such as intrauterine growth restriction and failure to thrive, which persists in children infected with HIV (Antonio George Lentoor, 2018; Ezeaka et al., 2009; López et al., 2015; Young et al., 2012). Finally, co-morbidities, either HIV-related or otherwise, have been identified as significant predictors of progression. These include opportunistic infections (tuberculosis, pneumocystis pneumonia) (Del Amo et al., 1998), hepatitis C (Greub et al., 2000), dementia (Nakku et al., 2013), psychosis, cardiovascular disease, and cancer, mainly co-prevalent in adults (Helleberg et al., 2013).

However, most of these factors among adult individuals may not be directly relevant to pediatric HIV, necessitating independent analyses of pediatric HIV disease progression and death predictors. Similarly, the heterogeneity of predictors seen in these studies suggests that different populations of HIV patients have specific and or unique factors.

2.7 Long-term non-progression of HIV

All populations experience varying susceptibility to infection and disease caused by infectious pathogens. Some individuals who are exposed to a pathogen resist infection, and individuals who do become infected may experience different levels of morbidity and mortality (Hill, 1996). It is thought that the great diversity within the immune-related genes such as HLA class I and KIR has arisen throughout our evolution to ensure the survival of at least some members of a population due to selective pressure exerted by exposure to infectious agents (Dendrou et al., 2018; Goyette et al., 2015).

As observed with other infections, exposure to HIV does not necessarily result in disease with the virus (Becerra et al., 2016). Heterogeneity in HIV susceptibility has been observed in several cohorts, including commercial sex workers, injection drug users, HIV discordant couples, and HIV-negative babies of infected mothers (Kathryn et al., 2018; Martinez et al., 2011; Sampathkumar et al., 2014). Despite infection, 1-5% of individuals show a high level of viral control in the absence of ART; these are referred to as LTNPs and/or elite controllers (D. Mendoza et al., 2012; Sabin & Lundgren, 2013). This natural resistance to HIV infection and AIDS nonprogresssion has been attributed to hosts and viral factors such as MHC/HLA variability, viral subtype, and replicative capacity (Baeten et al., 2007; P. J. R. Goulder & Walker, 2012; Payne et al., 2014). Further studies on this population have the potential to determine novel viral and host genetic factors that are associated with HIV disease progression for the much-needed vaccines and therapeutics (Aspord et al., 2010; Excler et al., 2015; Novitsky et al., 2001).

Recently, a rare cohort of HIV-infected individuals, identified as post-treatment controllers (PTCs) and characterized by sustained virologic suppression following the cessation of antiretroviral therapy (ART), has been documented (Namazi et al., 2018). The PTCs were first reported in 2013 in a French VISCONTI cohort of 14 adults with chronic HIV infection (Sáez-Cirión et al., 2013). Notably, PTCs exhibit a distinctive immunological profile characterized by significantly reduced activation in both CD4+ and CD8+ T cells, diminished exhaustion in CD4+ T cells, and heightened responses in Gag-specific CD4+ T cells and natural killer (NK) cells (Etemad et al., 2023; van Paassen et al., 2023). Virologically, they exhibit a low viremia with significant reduction in viral reservoirs (Sáez-Cirión et al., 2013; van Paassen et al., 2023). The unique characteristics of these individuals present valuable opportunities for investigating natural host mechanisms underlying viral control, thereby contributing to advancements in HIV cure research.

2.8 The HIV viral reservoir

The HIV viral reservoir represents a major barrier to achieving a cure for HIV infection, as it harbors latently infected cells that persist despite effective ART. These reservoirs, primarily comprised of memory CD4+ T cells, exhibit distinct characteristics that contribute to their resilience. Key features include a state of quiescence, allowing the virus to evade immune

surveillance and remain impervious to the effects of antiretroviral drugs. The viral reservoir is heterogeneous, both spatially and temporally, with different anatomical sites and stages of infection contributing to its diversity. Clonal expansion of latently infected cells further complicates eradication efforts, as it can amplify specific viral sequences.

Transcriptional silencing, mediated in part by host epigenetic modifications, contributes to the maintenance of a latent state within these reservoirs. Additionally, the viral reservoir can persist despite prolonged treatment due to the long half-life of infected cells, which may release virus upon reactivation. Understanding the dynamics and characteristics of the HIV viral reservoir is crucial for devising strategies to target and eliminate these latent reservoirs. Ongoing research focuses on elucidating the molecular mechanisms governing reservoir establishment, maintenance, and reactivation, with the ultimate goal of developing interventions that could lead to a functional cure for HIV infection.

2.9 HLA Class I Genes and Structure

The MHC is composed of 220 genes on the short arm of chromosome 6p21 (Figure 2-2) that encodes for the highly polymorphic HLA molecules (Shiina et al., 2017). In humans, these genes are also referred to as the HLA due to the antigenic differences observed between white blood cells in different people. The HLA genes are broadly classified into two major groups, called class I and class II genes. The HLA class I locus is highly polymorphic, with over 5000 alleles in the general population (K. Y. Chen et al., 2012; L. Zhang et al., 2012). Each individual inherits a set of HLA-A, -B and -C genes from each parent, thereby possessing six HLA class I alleles. These HLA class I alleles are expressed codominant. These different alleles for each individual are collectively referred to as the haplotype (Munderi et al., 2011; L. Zhang et al., 2012).

HLA class I genes are further categorized as classical (HLA-A, -B & -C) and non-classical (HLA-E, -F, -G & -H) (Shiina et al., 2017). The classical HLA class I genes are more polymorphic compared to the other genes (Shiina et al., 2017). The HLA genes are crucial in transplant immunology and activation of the cellular immune system. Their primary role is to present endogenously processed microbial peptides to T cells during a cell-mediated immune response (Shiina et al., 2017).

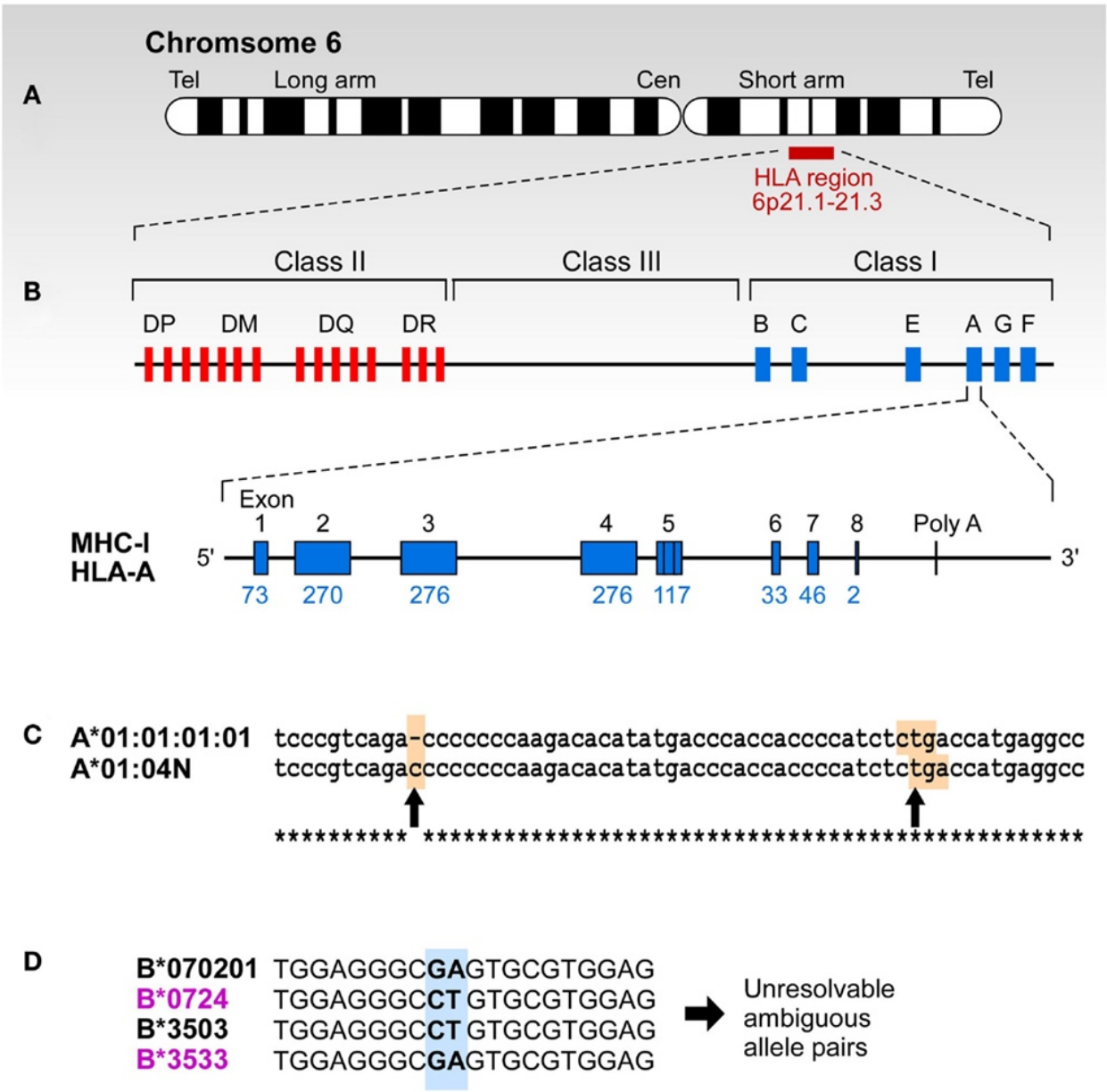


Figure 2-2: Genomic location and structure of HLA gene.

*(A) HLA resides at the short arm of human chromosome 6, spreading over approximately 4 million base pairs. (B) Class I region has three classical loci (A, B, and C), each of which has eight exons. Each individual has two copies of each locus. (C) Sometimes HLA genes are mutated to produce null alleles resulting in non-functional protein products. The example HLA-A*01:04N shows an insertion of C in codon 186 (left arrow) that caused a frameshift resulting in a premature frame stop codon (right hand). (D) HLA class I alleles are highly similar, and typically have >85% identity across A, B, and C loci. The HLA-B7/B35 example shows ambiguity – when two probes are positive, the result is*

*ambiguous. In this case, both HLA-B*07:02/B*35:03 and HLA-B*07:24/B*35:33 are possible. However, HLA-B*07:24 and B*35:33 alleles are rare; they have been observed in frequencies of >0.0001%, making it highly unlikely that this combination is present in any individual. Contrastingly, HLA-B*07:02 and B*35:03 have been observed in all of the 22 populations screened in the NMDP, and their frequencies range from 0.8-13.1% (HLA-B*07:02) and 0.04-7.2% (HLA-B*35:03). Adapted from Zhang et al., Frontiers in Immunology (2014) 5:597 (G. L. Zhang et al., 2014)*

HLA class I proteins consist of two noncovalently bound heterodimeric molecules, called α chain and β_2-microglobulin. The α chain is encoded for on the MHC locus while the β_2-microglobulin is encoded for on chromosome 15 (Natarajan et al., 2009). Class I α chain, also referred to as the heavy chain has three domains, namely α1, α2, and α3, each about 90 amino acids long (Natarajan et al., 2009). α1 and α2 are distal to the cytoplasmic membrane, while α3 is proximal and contains an intracytoplasmic anchoring tail. The α1 and α3 domains have the amino and carboxyl termini, respectively (Natarajan et al., 2009). The α1 and α2 domains form the peptide binding groove, while at the top, they provide an area for binding the T cell receptor. The amino acids for the α1 and α2 domains are highly polymorphic, while the α3 domain is less variable across individuals (Natarajan et al., 2009; Wieczorek et al., 2017a). This variability is responsible for the wide range of peptide binding and T cell recognition. The α3 domain provides an area for the attachment of T cell co-receptor, the CD8 molecule (Natarajan et al., 2009; Wieczorek et al., 2017a). The β_2-microglobulin chain is found below the α1 domain and is non-covalently bound to the α3 domain.

2.10 HLA Nomenclature

The WHO Nomenclature Committee names the HLA genes for Factors of the HLA System, Figure 2-3 (Marsh et al., 2010). The HLA nomenclature follows a sequential alphanumerical format that characterizes the genetic composition and physiological function of the molecules they encode. HLA naming starts with an HLA prefix followed by a letter such as “A”, “B”, or “C” that denotes the specific HLA gene locus. Next is the first numerical digit that indicates the allele group, followed by an integer that represents the particular allele protein and a number that denotes the presence of a silent mutation. The final digit indicates the presence of a mutation in the introns of that particular HLA gene (Marsh et al., 2010). The last alphabetical suffix denotes the expression status of the HLA molecule. Alleles that are designated as S

(secreted) are soluble; N (null) does not produce any proteins; L (low) have low surface expression, and Q (questionable) have questionable expression (Marsh et al., 2010).

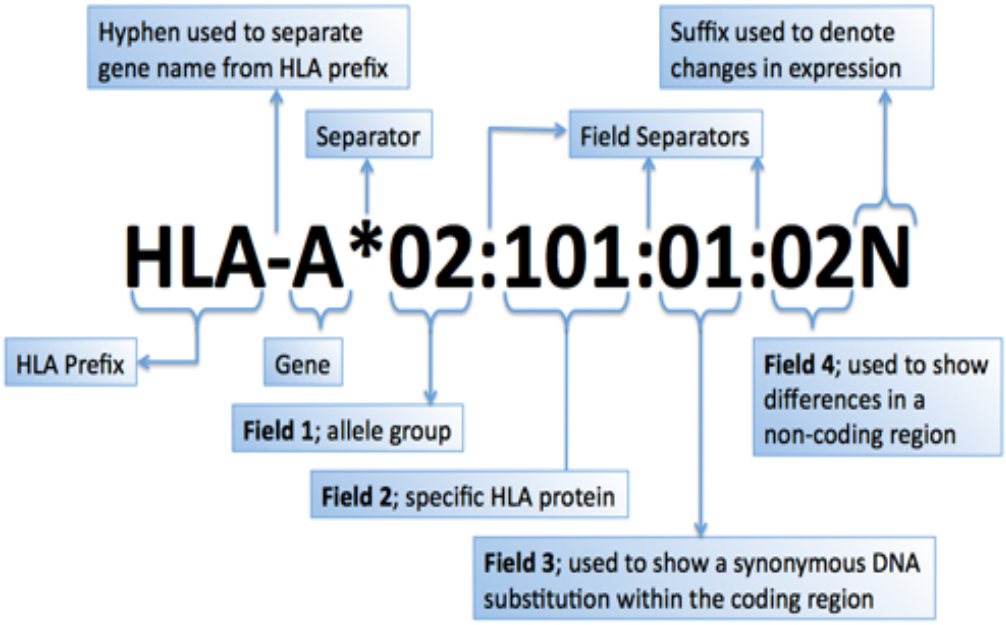


Figure 2-3: The WHO Nomenclature Committee for Factors of the HLA System.

Adapted from http://hla.alleles.org/nomenclature/naming.html (Robinson, 2001; Robinson et al., 2015)

2.11 HLA Class I Antigen Presentation Pathway

HLA class I molecules are ubiquitously expressed by all nucleated cells (K. Y. Chen et al., 2012). The HLA molecules present antigenic peptides generated in the cytosol (Figure 2-4). The peptides are derived from microbial proteins or dysfunctional cellular proteins (K. Y. Chen et al., 2012; Groettrup et al., 2010; M. Huang et al., 2016). These proteins are hydrolyzed in the cell's cytoplasm by an enzymatic complex called the proteasome (Hewitt, 2003; M. Huang et al., 2016). Ubiquitin is enzymatically added to proteins to be degraded, forming polyubiquitin chains recognized by the proteasome (Figure 2-4). Ubiquitination has an early role in antigen presentation by class I molecules as it determines the characteristics of peptides produced, which range between 3 and 22 amino acids (Hewitt, 2003; Jäger et al., 2012).

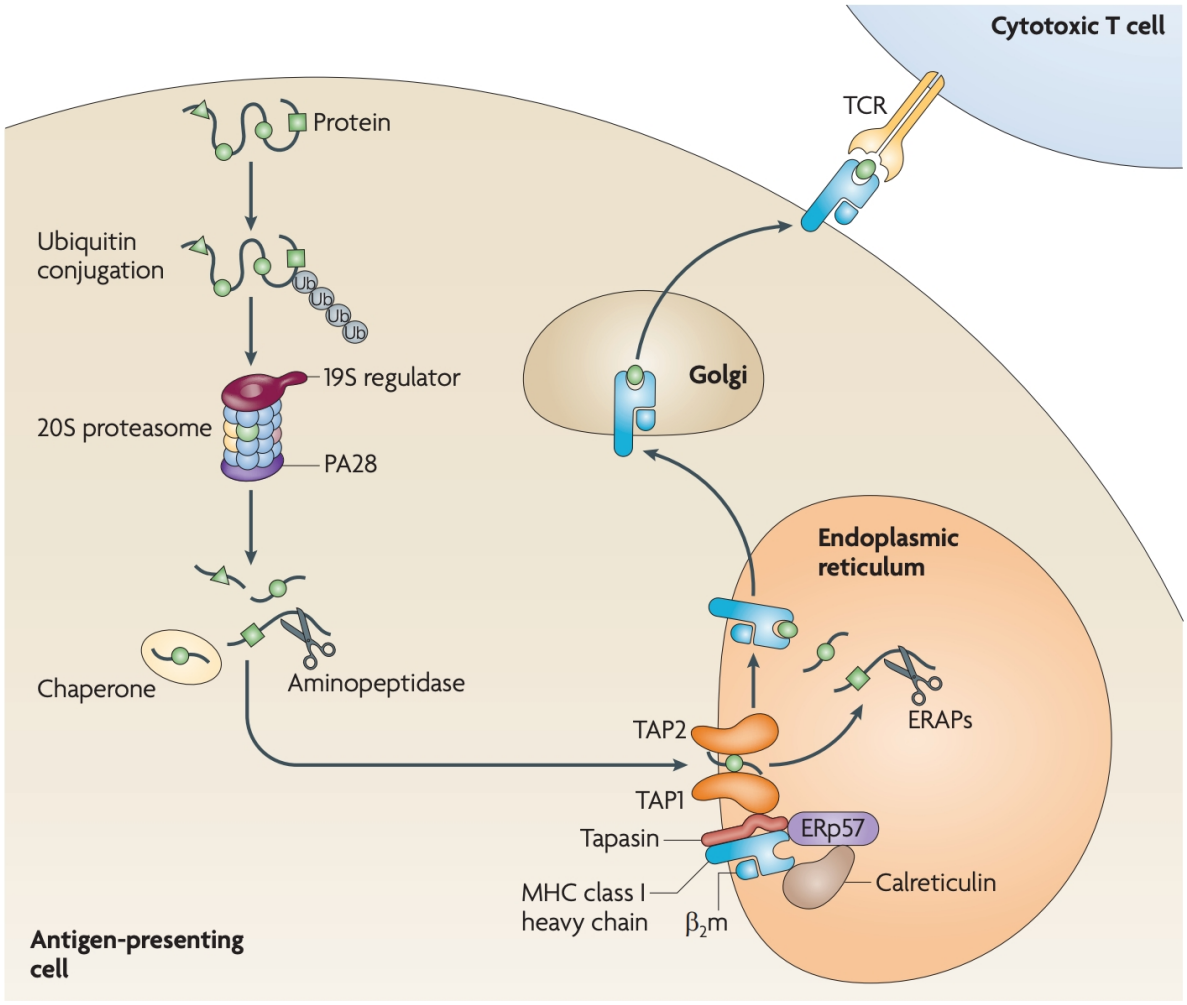


Figure 2-4: Antigen processing in the MHC class I-restricted pathway

Proteins that are synthesized in the cell (direct presentation) or are released from endosomes (cross-presentation) are polyubiquitylated in the cytoplasm and degraded by hybrid proteasomes consisting of the 20S proteasome core, the 19S regulator and PA28. The peptides that are produced are either of the ideal length for binding to MHC class I molecules (8–9 amino acids) or are amino-terminally extended precursors that can be further cleaved by aminopeptidases in the cytoplasm. Chaperones (such as heat shock protein 70 (HsP70), HsP90 α and TriC) can stabilize the peptides in the cytoplasm to prevent their rapid degradation (for example by tripeptidyl peptidase II or thimet oligopeptidase). Transporter associated with antigen processing 1 (TAP1) and TAP2, which are attached to nascent MHC class I chains through tapasin, transport the peptides into the endoplasmic reticulum (er), where they can be further trimmed at the N-terminus by er aminopeptidase 1 (erAP1) and erAP2. The oxidoreductase erp57 ensures the maintenance of disulphide bridges in the MHC class I loading complex. Note that the carboxyl terminus of a peptide ligand for MHC class I molecules is mainly

determined by proteasomal cleavage. The binding of peptides with high affinity to the MHC class I heavy chain-β2-microglobulin (β2m) complex induces a final folding and release of the MHC class I molecule from the er lumenal chaperone calreticulin to allow exit from the er and migration through the Golgi to the plasma membrane. 'Reproduced with permission from Springer Nature'.

The peptides are transported from the cytosol into the endoplasmic reticulum, where the HLA molecules are synthesized and assembled (Hewitt, 2003; M. Huang et al., 2016). Peptide transport is achieved by an endoplasmic reticulum (ER) membrane specialized molecule; the transporter associated with antigen processing (TAP) (Hewitt, 2003; M. Huang et al., 2016). Tapasin enables bridging the HLA molecule with TAP, which loads the appropriate microbial peptides onto the HLA class I molecules. The newly formed peptide-HLA (pHLA) complex is stabilized in the Golgi apparatus and then transported in an exocytic vesicle to the cell surface for expression (Hewitt, 2003; M. Huang et al., 2016; Parolini et al., 2018).

2.12 HLA Class I and HIV Disease Progression

The HLA genes are the most polymorphic genes known in humans. HLA class I genes at the HLA-A, B, and C loci encode molecules that differentially present endogenously processed viral peptides to $CD8^+$ T lymphocytes (K. Y. Chen et al., 2012). As a result of differential peptide binding, specific HLA molecules may influence susceptibility to HIV-I infection and progression (Watanabe et al., 2011). Differential peptide binding leads to activation of cellular immune responses directed towards specific viral epitopes, exerting immune pressure on HIV and resulting in viral clearance and suppression (K. Y. Chen et al., 2012).

One of the most studied aspects of susceptibility to HIV is the HLA system; both class I and class II genotypes have been implicated in protective immunity (Carrington & O'Brien, 2003). HLA B*57 & B*27 and HLA-A*23, B*37 & B*35 are associated with slow and rapid HIV disease progression in adults, respectively (Carrington & O'Brien, 2003; Migueles et al., 2000). However, due to the complexity and geographical variability of the HLA genes, the distribution of HLA class I alleles varies among populations (Munderi et al., 2011; Novitsky et al., 2001). Class I zygosity is observed to be associated with HIV disease progression (Carrington & O'Brien, 2003). Class I homozygosity at one or more loci is associated with rapid progression, while heterozygosity is associated with HIV restriction (Carrington & O'Brien, 2003).

Heterozygosity offers a broader repertoire of viral peptides to T cells for clearance, thus leading to slow progression (Carrington & O'Brien, 2003). Contrary to this, homozygosity for HLA-B*4 allele-restricted epitope was associated with slow HIV disease progression in adult Caucasians (Flores-Villanueva et al., 2001).

Studies of HLA class I association with HIV disease progression in different populations have yielded different results. Migueles (2000) observed that B*57:01 was associated with slow HIV disease progression in Caucasians and Hispanics (Migueles et al., 2000). Studies were done in South Africa, Botswana, Zimbabwe, and Zambia showed that HLA-B*13:01, B*42:01, B*44:03, B*57:01, B*58:01, B*81:01 and A*74:01 were associated with slow HIV disease progression (P. J. R. Goulder & Walker, 2012). In the same population, HLA-B*08:01, B*18:01, B*45:01, B*51:01, B*58:02 and A*36:01 were associated with rapid HIV disease progression (P. J. R. Goulder & Walker, 2012). In contrast to other African people in Kenya, Gambia and Rwanda, HLA-B*57:03 and B*35 were associated with slow progression (P. J. R. Goulder & Walker, 2012; E. Trachtenberg & Erlich, 2001). These differences are closely related to the different HIV clades that dominate the epidemics in these countries. Mismatches in HLA type between discordant couples and class I B*53 have all been associated with protection from infection. Similar findings were observed for class I B18 in Thai casual sex workers (Beyrer et al., 1999; Lockett et al., 2001; Rohowsky-Kochan et al., 1998). In addition, in the Pumwani cohort, the genotypes that have been shown to associate with HIV resistance include the HLA class I A2/6802 supertype (Sampathkumar et al., 2014).

Although previous association studies were focused on HLA typing using molecular methods such as polymerase chain reaction-sequence-specific primer (PCR-SSP) typing, similar results have shown that HLA has a more significant role in HIV disease progression with the use of newer technologies such as genome-wide association studies (GWAS) that survey other regions of the genome (Nakimuli et al., 2013; Walker et al., 2010). A recent GWAS of 1712 individuals of European ancestry showed that all the significant single-nucleotide polymorphisms (SNP) were exclusively identified in the HLA region located explicitly around the class I region (Walker et al., 2010). However, another GWAS, albeit with smaller sample size, did not find any differences in HLA class I among LTNPs and RPs (Nissen et al., 2018).

In addition, HIV is associated with viral subtype geographical diversity (Becerra et al., 2016). HIV clade D is uniquely restricted to East Africa compared to clade C in Southern Africa. Clade D is associated with rapid HIV disease progression (Barugahare, Baker, Aluoch, et al., 2005; Spira et al., 2003). Most of the studies on the role of HLA class I and HIV disease progression have been conducted in Caucasian populations, which are predominated by clade B (Junqueira & Almeida, 2016; E. Trachtenberg & Erlich, 2001). The protective role of HLA class I alleles in the context of a predominantly HIV clade C and D epidemic in Africa has not been well studied. Therefore, there is a need to diversify the existing data from previous studies with purely African populations. Furthermore, these studies are limited in number and sample size, use low-resolution allelotyping techniques and/or are a replication of Caucasian studies, which restricts their wider application (Kyosiimire-lugemwa et al., 2012; E. Trachtenberg & Erlich, 2001).

The data on the structural and functional mechanisms HLA mediates HIV disease progression remains contentious. Studies suggest that some alleles that differ from the others at single substitutions in different positions have a better peptide binding capacity in the HLA groove (van Deutekom & Keşmir, 2015). Navis et al. observed that individuals who are HLA B*57 and LTNP or RP show advantageous host factor (Navis et al., 2007). The study concluded that the viral replicative capacity was different in LTNP and RPs and thus responsible for the HIV restriction in LTNP. Similarly, Migueles (2007) and Mendoza (2012) observed no quantitative difference in HIV-specific $CD8^+$ T cell responses among LTNPs and RPs with B*57:01 (D. Mendoza et al., 2012; Migueles et al., 2000). However, the same studies reported that $CD8^+$ T cell responses were directed more to CD8 restricted gag peptides in LTNPs (Migueles et al., 2000). The peptide binding characteristics of the HLA class I alleles vary widely depending on the HIV peptide and alleles. Peter and colleagues showed that HLA*A2 binds a restricted set of HIV peptides derived from RT, p17 and gp41 that elicit a strong CTL response. The study further demonstrated the occurrence of immunodominance. They found that each peptide stimulates a strong CTL response in isolation, but this response is blunted when different peptides are administered together (Peter et al., 2001). Another study reported similar results with CTL strong responses where HIV-integrase peptides are presented by HLA*B40:2 (Watanabe et al., 2011). These studies imply that differences in the peptides presented and differences in the HLA class I alleles that recognize the peptides in the LTNP, lead to

production of a better immune response. Therefore, elucidating these mechanisms could lead to the development of candidate vaccines and immunotherapeutics (Chow et al., 2018; El Bissati et al., 2016).

2.13 Other Genetic Markers and HIV

Human genetic make-up has been implicated in HIV resistance. A well-known example is a 32 base pair deletion in the CCR5 chemokine receptor; this polymorphism is known as CCR5-Δ32, and results in a truncated CCR5 which is not expressed on the surface of $CD4^+$ T cells (Dean et al., 1996; Paxton & Kang, 1998; Smith, Dean, Carrington, Huttley, et al., 1997). In homozygotes, this prevents infection by R5 tropic viruses, which require the receptor for entry (Samson et al., 1996). Slower disease progression is seen in infected heterozygotes for CCR5-Δ32 and people with mutations in the minor CCR2 coreceptor (CCR2-64I) (Mulherin et al., 2003; Smith, Carrington, et al., 1997). While the CCR5-Δ32 mutation can be found at a heterozygote frequency of up to 20% in Eastern European populations and about 1% of Eastern Europeans are homozygous for the mutation, it is not detected in African or Asian populations to an appreciable extent (R. Liu et al., 1996; Smith, Dean, Carrington, Winkler, et al., 1997).

Critical mutations in the ligands of chemokine receptors, CCR5 and CXCR4, have also been identified. These include a promoter mutation in RANTES (ligand for CCR5) associated with delayed disease progression (H. Liu et al., 1999) and HIV individuals (Sriwanthana et al., 2001), as well as a polymorphism in the 3' untranslated region of a ligand for CXCR4, stromal-derived factor 1 (SDF-1). The SDF-1 3'A mutation results in increased SDF-1 translation, which may result in reduced T-tropic HIV binding due to competition from the more-prevalent SDF-1, and thus a lesser chance of infection (Soriano et al., 2002). Another suppressive chemokine called CCL3L1 is a ligand for the CCR5 coreceptor, and its copy number is known to differ between populations. Lower than average CCL3L1 copy numbers have been associated with increased susceptibility to HIV infection and faster disease progression in HIV-positive individuals (Gonzalez et al., 2005).

In addition, polymorphisms in the IRF-1 gene, located in the IL-4 gene cluster, have been found to associate with lower levels of IRF-1 protein expression and resistance to infection by HIV in the Pumwani cohort (Ball et al., 2007). Also, a genetic variant in the DC-SIGN repeat region,

which may cause decreased binding of HIV to DCs, was associated with a reduced risk of infection (H. Liu et al., 2004). While the genetic factors identified thus far shed light on protection against HIV, they do not explain all cases of altered susceptibility to infection. Hence, it is evident that additional factors play a significant role in influencing susceptibility to HIV infection.

2.14 Bioinformatics Approaches and Identification of HIV Epitopes

As already explained, the essentiality of HLA class I molecules in presenting peptides on the cell surface cannot be overstated, as it underpins the foundation of CD8+ T cell-mediated immune responses. Hence, the identification of peptides presented by HLA molecules holds significant implications. Diverse techniques have emerged for predicting the sequential stages within HLA peptide presentation pathways, encompassing machine learning predictive algorithms that can be experimentally validated (Zhao et al., 2013). Despite their limitations, these methodologies provide cost-effective and fast alternative methods to expedite the discovery of immunogenic peptides (R. K. Pandey et al., 2018).

Among computational tools, NetMHCpan stands out as a prominent resource for epitope prediction (M. Nielsen et al., 2007). It is trained on naturally eluted ligands and focuses on forecasting epitopes pertaining to HLA class I alleles. Nonetheless, it exclusively addresses a singular aspect of the antigen-presenting pathway, namely peptide binding prediction. This property can be further complemented by tools that examine additional pathways, such as proteasomal cleavage prediction (Keşmir et al., 2002; Saxová et al., 2003) and the anticipation of transport via transporter-associated proteins through the endoplasmic reticulum (Peters et al., 2003).

Nonetheless, these methodologies often generate a substantial pool of potential peptides, necessitating significant resources for the refinement of the most immunogenic candidate. Addressing this challenge can be achieved through the utilization of tools developed by structural biologists, encompassing techniques such as molecular docking and molecular dynamics (R. K. Pandey et al., 2018). These domains provide a unique prospect to unravel the interplay among protein complexes, thus identifying the most probable binding partners. As these algorithms progress, the search space progressively narrows down, facilitating the

precision and optimization of the selected peptides (R. K. Pandey et al., 2018; Reboul et al., 2012).

Molecular docking involves the placement of small molecules into protein receptor to analyze their interactions, enabling the prediction of binding and offering applications in drug design through the assessment of binding affinity. There are several different algorithms that can be used for flexible docking, and each has its own advantages and disadvantages. Dock 6.0 developed by Allen et al. algorithm uses a genetic algorithm to search for the lowest energy conformation of the receptor-ligand complex.(Allen et al., 2015) These algorithms are very effective at finding the lowest energy conformation of the receptor-ligand complex. However, Dock 6.0 is computationally intensive, which means that they are very slow. These were later replaced by a more popular AutoDock Vina that employs a genetic algorithm (Antunes et al., 2017). Docking programs are associated with some degree of uncertainty. Ramsbottom et al., assessed the reliability of docking programs in predicting HLA-associated drug reactions. It was revealed that the algorithms evaluated could sometimes correctly predict Abacavir's binding mode with B*57:01, but not always (Ramsbottom et al., 2018). In the instant analysis, they observed that including receptor flexibility had a negative influence on docking performance. Overall, caution is urged, as defective structures can lead to inaccurate docking predictions. In this thesis, we chose to employ incremental docking using DINC a parallelized incremental meta-docking based on Vina, and can dock large ligand with >60 degrees of freedom (flexible torsions) (Santos et al., 2020).

Docking poses are a vital component of structure-based drug design. Poses that score well against a given target protein can be used in subsequent refinement and design iterations, while poorly scoring poses can be discarded. However, the scoring process is often complex and time-consuming, requiring the manual examination of many different aspects of the docking pose. There are various opensource rescoring tools such as Convex-PL^{R}, (Kadukova & Grudinin, 2017) Δ_{vina}RF, (C. Wang & Zhang, 2017) and X-Score (R. Wang et al., 2003). Convex-PL^{R} was shown to outperform several tools in docking, scoring, ranking and screening best binders (Kadukova & Grudinin, 2017).

Unfortunately, molecular docking remains static, whereas proteins are subject to continuous motion, environmental pressures, and temperature fluctuations. Thus, to achieve precise

predictions of the most stable protein complexes, molecular dynamics simulations emerge as indispensable tools for gauging complex stability (Wan et al., 2008). The predicted binding free energy of protein-protein interactions determines the stability of complex formed. Thus, achieving a comprehensive comprehension of cellular mechanisms necessitates not only awareness of all potential protein-protein interactions but also a quantitative measure of the structural composition and stability inherent to these complexes (Ziegler et al., 2020). This consideration can be extended to determine the impact of protein mutations, which can potentially influence or even hinder the binding interactions within protein complexes (Narzi et al., 2012). Such interactions are frequently facilitated by specific subdomains or regions sometimes referred to as hotspots within proteins (Joglekar et al., 2018; Mori et al., 2014; Xia et al., 2014). For example, the previously described F pockets observed in HLA-C*01:02 and C*03:04, share near-identical characteristics, differing solely in an amino acid alteration from leucine to isoleucine at position 95, and both motifs have demonstrated an affinity for binding small hydrophobic residues (Hoof et al., 2009; Zappacosta et al., 1997). Therefore, understanding these protein-protein interactions and their impact can be important in the design of novel immunotherapies and vaccines for HIV.

2.15 Detecting Immunological Responses Through Dual-Color ELISPOT Assays

The utilization of immune surveillance techniques is crucial in understanding the factors associated with effective immune responses (Correa et al., 2007; Ndongala et al., 2009; Sandström et al., 2008). The application of the dual-color ELISPOT assay, which enables the concurrent identification of T cells that are specific to antigens and secrete either IFN-γ or IL-2 or both (Ostrowski et al., 2001; Peretz et al., 2007; Tsalimalma et al., 2011). The selection of cytokines to target in a dual-color ELISPOT assay may vary depending on the specific research question being investigated. The release of IFN-γ is a T cell activity that exhibits antiviral properties and remains unaffected by functional exhaustion during infections with high persistent antigen load (Wherry et al., 2003). This cytokine is commonly evaluated in routine ELISPOT tests, which are utilized to monitor vaccine trials (Román et al., 2013; Sandström et al., 2008). In contrast, the secretion of IL-2 is susceptible to functional exhaustion when there is a significant and sustained antigen load (Booiman et al., 2017; Wherry et al., 2003). Nevertheless, this method has a tendency to detect reactions that are inclined towards

multifunctionality, meaning they possess the capacity to multiply and efficiently regulate the reproduction of viruses (Ferrando-Martínez et al., 2012; Ndongala et al., 2010).

2.16 Conclusion

HIV continues to persist at epidemic levels in Sub-Saharan Africa. Despite numerous endeavors aimed at developing a potent HIV vaccine, the targeted protective outcomes have yet to be achieved (Haynes & McElrath, 2013). Better knowledge of the distribution, structure, and function of HLA molecules associated with HIV disease progression and their potential roles in effective peptide presentation is a valuable aid for developing an effective vaccine and immunotherapeutics against HIV (Excler et al., 2015; Haynes & McElrath, 2013; Johnston & Fauci, 2007, 2008). Identifying molecules of cellular origin that restrict HIV replication and understanding their mechanism of action could contribute to advances in strategies that would roll back the HIV epidemic. Thus, this acquired knowledge would make it possible to design better vaccines and immunotherapeutics that modify the host factors such as HLA. Note that a vaccine against HIV is recognized as one of the most appropriate ways to fight the epidemic HIV (Haynes & McElrath, 2013).

The main aim of this thesis is to define the role of the HLA class I alleles in the progression of HIV. Our work is centered on class I MHC molecules, a host cellular molecule that presents HIV peptides to $CD8^+$ T cells and is also one of the most abundantly expressed cellular molecules. Using a retrospective cohort, the first study characterizes pediatric HIV disease progression in Botswana and Uganda. We estimate the proportion of LTNP and RP and determine the factors associated with clinical progression. The second study establishes the class I alleles and haplotypes associated with HIV disease progression in African populations. These alleles are variable in different populations such as Caucasians, Asians and Africans. This suggests that the global variability of HLA necessitates an independent assessment of this region in African populations. We are also aware that the HIV epidemic is predominantly HIV-1 subtypes A, C and D in African populations, which differs from HIV-1 subtypes B in Caucasians. Finally, the third study examined the structural and functional distinctiveness of the HLA class I alleles associated with HIV disease progression. It is plausible that protective HLA class I alleles have different conformations and restrictive epitopes that aid the HLA class I antigen presentation process which can be harnessed toward novel vaccine strategies.

CHAPTER 3: MATERIALS AND METHODS

3.1 Materials

3.1.1 Study sites and cohort

This PhD study is nested within the Collaborative African Genomics Network (CAfGEN*)*, funded by the National Institutes of Health under the H3Africa program (https://www.h3africa.org). CAfGEN was initiated in 2013 as a collaboration between two pediatric centers of excellence (COEs) in Botswana and Uganda, two Universities in Africa and an American university, Makerere University, University of Botswana and Baylor College of Medicine, respectively (Figure 3-1). The overall goal of *CAfGEN* is to understand the host genetic factors associated with HIV and HIV-TB disease progression among children from Botswana and Uganda. CAfGEN offered a platform for participant recruitment used in this PhD work (Mlotshwa et al., 2017; Retshabile et al., 2018).

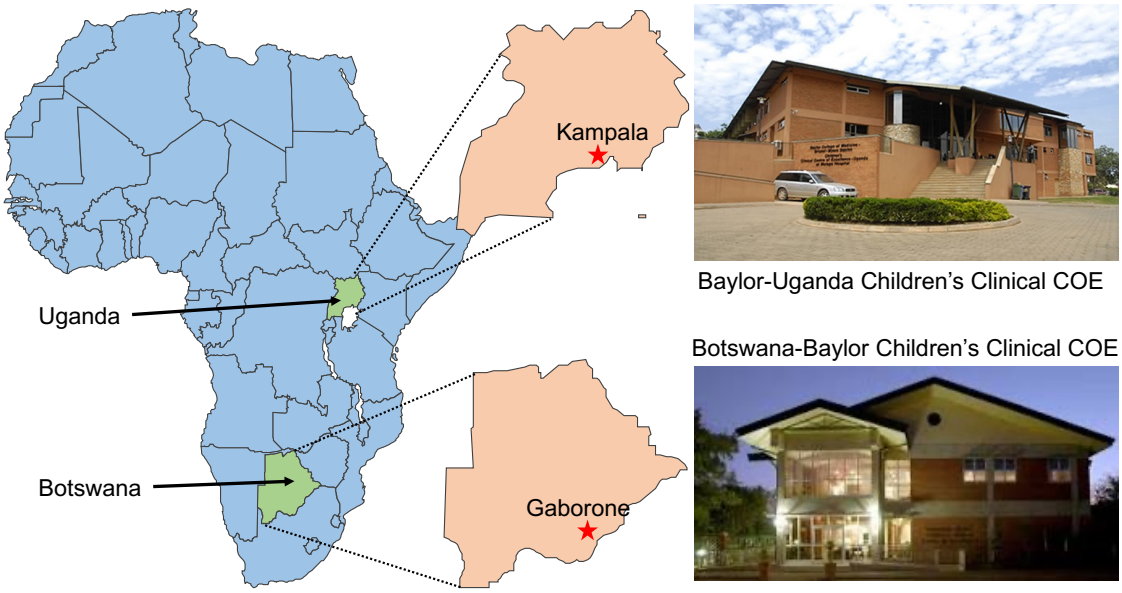


Figure 3-1: Map showing the location of the Clinical Centers of Excellence in Botswana and Uganda

Specifically, CAfGEN recruited two study populations at the extremes of the HIV disease progression spectrum as cases (LTNP) and controls (rapid progressors) as per the WHO

classification (Figure 3-2) from a cohort of 15,124 children prospectively enrolled from mid–2003. LTNPs are defined or categorized as asymptomatic HIV infection for over ten years after initial infection (the perinatal period) without ART. 'Asymptomatic' in this context refers to those individuals who did not meet the criteria for starting ART in infants and children before 2010 (Mlotshwa et al., 2017; Retshabile et al., 2018). RPs are categorized as (a) two or more $CD4^+$ T cell percentage values <15% within three years after birth, with no value >15% afterwards in the absence of ART; and/or (b) ART initiated within three years after birth, and at least one preceding CD4<15%; and/or (c) AIDS-defining illness (CDC Cat 3 or WHO Stage 3/4) (Mlotshwa et al., 2017; Retshabile et al., 2018).

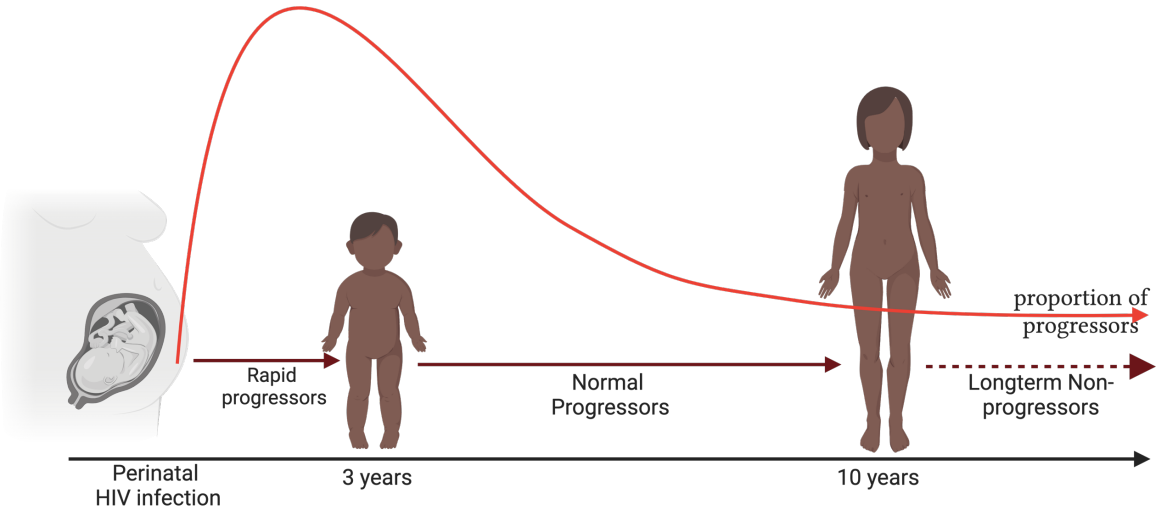


Figure 3-2: Pediatric HIV disease progression continuum

Children infected with HIV progress to disease over a variable continuum of time. Created with BioRender.com

3.1.2 Ethical considerations

CAfGEN obtained ethical approval from the School of Biomedical Sciences IRB Protocol No. SBS 112 and the National Council for Science and Technology (UNCST). Similarly, all collaborating institutions obtained ethical approval according to their national human subject research regulations. Relatedly, this study sought a waiver of consent to utilize the CAfGEN

study materials. This study was approved by the School of Biomedical Sciences HDREC and UNSCT (Appendix 9.3 and 9.5).

Privacy and confidentiality were maintained per the parent study where participant identification numbers are used for data coding and extraction. No personally identifiable information in any combination was used to collect data; therefore, we avoided the possibility of participant re-identification.

3.1.3 Participant recruitment

Clinical and demographic data on the 15,124 children with perinatal HIV infection enrolled into care in the COEs was used to achieve the first objective. For the second study, participants were consecutively recruited until the desired sample size was achieved. The participants were offered informed consent or assent to participate in the study (Appendix 9.6). The consent forms were translated into the local languages for proper understanding. In the consent forms, participants had a choice to store the samples in the biorepository and provide approval for the use of samples for future studies. Upon enrolment, the demographic data (age, sex/gender, ethnicity), CD4% and $CD4^+$ T cell counts (mean, median, minimum, date/age at minimum value), height, weight, HIV subtype (where available), viral load (mean, median, maximum, date/age of maximum), and ART status (date of commencement, regimen) were obtained from the electronic medical records to enable categorization. In addition, 8.5mL of whole blood was collected in a PAXgene DNA Blood tube which was stored at -80°C (Mlotshwa et al., 2017; Retshabile et al., 2018).

3.2 Methods

We studied a unique clinical cohort of children and applied genetics, molecular biology and immunological techniques. The study aimed to link specific differences in the human immune function of the HLA genes to HIV disease control and provide new insights into HLA-mediated mechanisms regulating HIV disease progression. To address the overall Ph.D. aim, we combined an interdisciplinary approach using cohort analysis, high-throughput HLA genotyping and cellular and molecular functional assays of HLA variants (Figure 3-3).

3.2.1 Study design

We used multiple study designs to achieve the stated objectives (Figure 3-3). Objectives 1 was conducted using a retrospective cohort study design to describe the HIV pediatric cohorts in Uganda and Botswana and the factors associated with HIV progression in these populations. While for Objective 2, we utilized an unmatched case-control study design to determine the HLA class I alleles association with HIV disease progression. Finally, for Objective 3, we use an experimental study design to determine the structural and functional uniqueness of HLA class I alleles associated with LTNP.

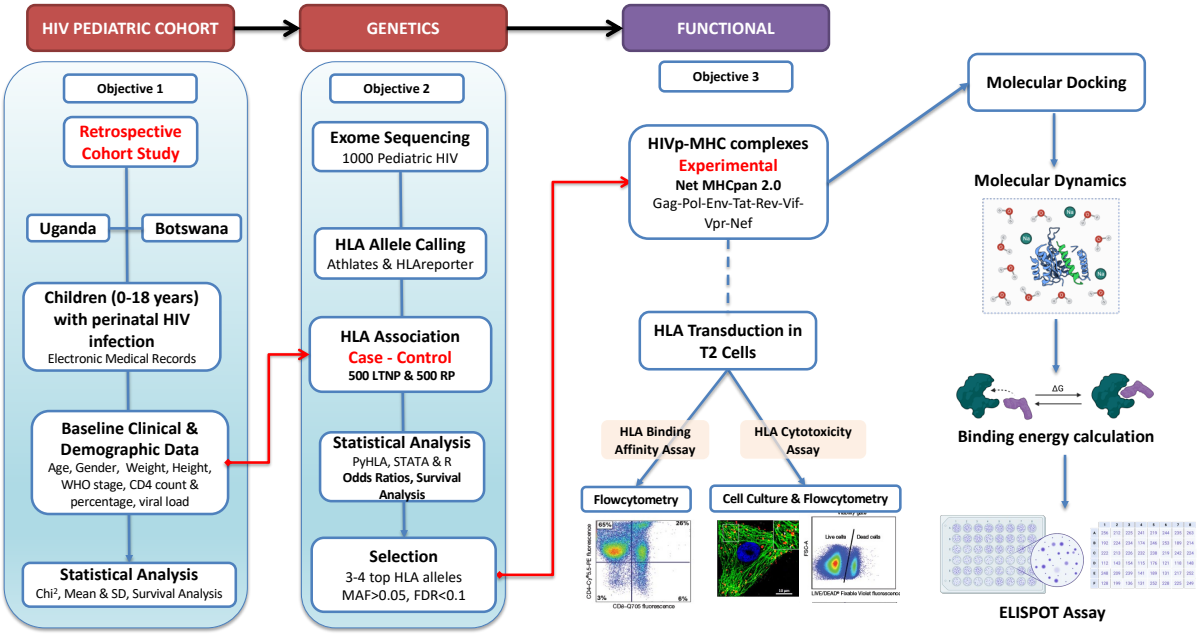


Figure 3-3: Diagrammatic flowchart of the study methods and designs

3.2.2 Sample size and power calculations

In Study I, we used clinical records from the EMR for 15,124 children enrolled at the COEs from mid-2003 when they first started in Botswana. Depending on the analysis required, children who did not meet the inclusion criteria for the study were excluded.

For Study II, one limitation of the analysis was the fixed sample size determined in *CAfGEN,* where 500 cases and 500 controls were recruited. Given this limitation, we explored the available power of the study under the following assumptions. We considered (a) class I minor allele frequency (MAF) >0.05; (b) accepted type I error (alpha, α) for statistical significance of p<0.05 and (d) a log additive disease model. When these assumptions are implemented in Quanto v1.2.4, the available sample size is powered at 98% to detect moderate to large effect sizes (genetic relative risk or odds ratio) between two and four disease-associated alleles (Figure 3-4) (Gauderman, 2002).

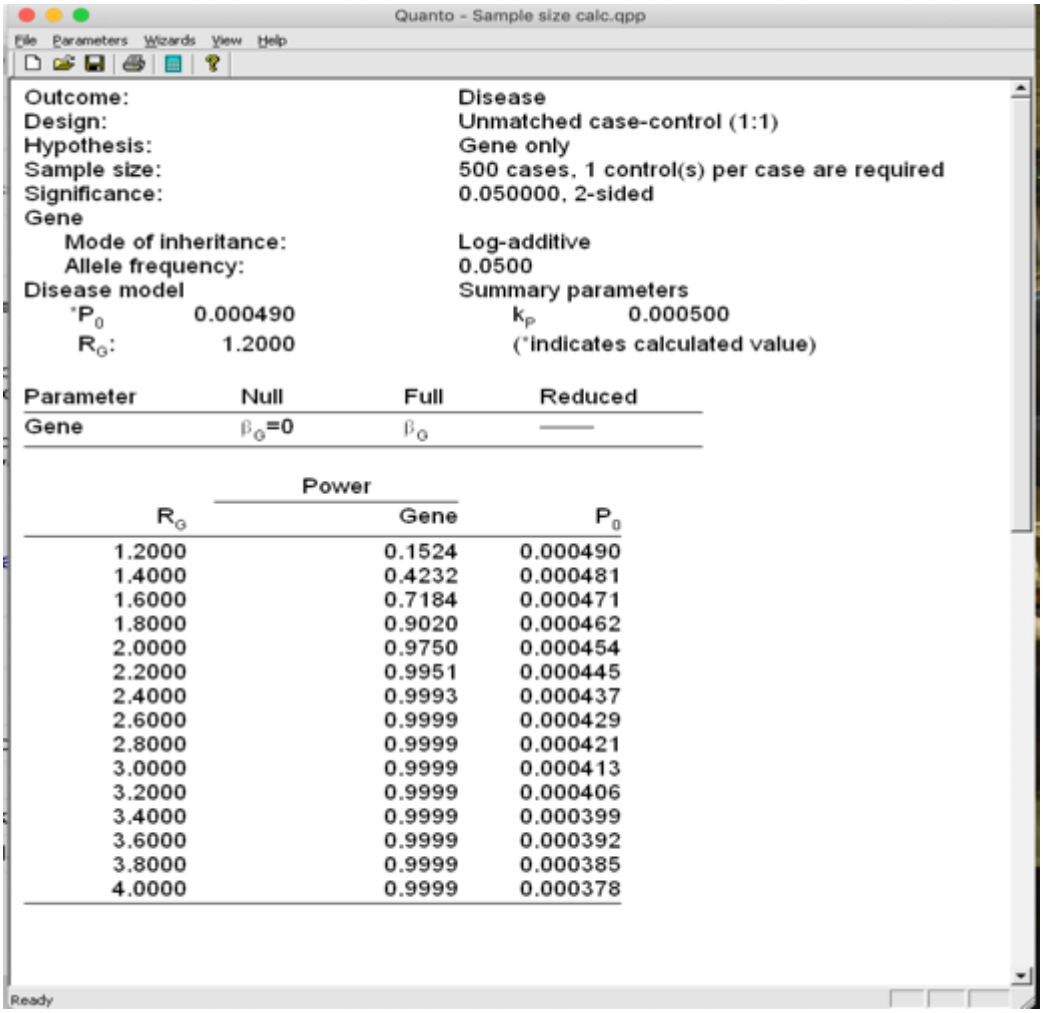


Figure 3-4: Sample size and power calculation

3.2.3 DNA extraction

The PAXgene Blood DNA (PreAnalytiX GmbH, Hombrechtikon, Switzerland) tube was used to collect 8.5ml of whole blood from all the study participants in the two recruitment centers.

The DNA used in the study was extracted from the whole blood using the PAXgene Blood DNA kit (PreAnalytiX GmbH, Hombrechtikon, Switzerland) following the manufacturer's instructions in Makerere University, Uganda and the University of Botswana, Botswana. All the blood was poured from one PAXgene Blood DNA Tube into a processing tube containing 25 ml Buffer BG1. Close the tube and mix by inverting the tube 5 to 10 times. The tube was centrifuged for 5 min at 2500g in a swing-out rotor. The supernatant was carefully discarded. Added 5 ml Buffer BG2 and washed the pellet by vortexing vigorously for 5 sec. The tube was centrifuged for 3 min at 2500g using a swing-out rotor. The supernatant was carefully discarded. Then 5 ml of Buffer BG3/PreAnalytiX protease were added and vortexed for 20 sec at high speed. The tube was placed on a heat block and incubated at 65°C for 10 min; till the protein was digested, as indicated by changing color from light red to light green. The tube was vortexed again for 5 sec at high speed. Then added 5 ml of isopropanol (100%) and mixed by inverting the tube 20 times until the white DNA strands clumped visibly precipitating the DNA. The tube is centrifuged for 3 min at 2500g. The supernatant was then discarded, and the tube was inverted to dry on absorbent paper for one min. Then, 5 ml 70% (v/v) ethanol was added and vortexed for 1 second at high speed. Then centrifuge for 3 min at 2500g. Again, the supernatant was discarded, and the tube was inverted to dry on absorbent paper for at least 5 min. Finally, 1 ml Buffer BG4 was added to dissolve the DNA and incubated for 1 hour at 65°C in a heating block. The tube was left placed at room temperature (RT) for overnight incubation. The resultant eluted DNA was quantified using a Qubit dsDNA assay kit using a Qubit 2.0 Fluorometer (Invitrogen Life Technologies, Carlsbad, CA) and NanoDrop 2000/2000C fluorometer (Thermo Fisher Scientific, Wilmington, DE USA) following the manufacturer's instructions. The DNA was then stored at -20ºC until it was shipped at RT to Baylor College of Medicine (BCM), USA, for whole exome sequencing (WES).

3.2.4 Whole Exome Sequencing (WES)

The DNA quality and quantity were reassessed at Baylor College of Medicine for integrity and quality and normalized before WES. WES was done in six batches depending on the shipment dates and funding availability. WES was done at the BCM-Human Genome Sequencing Center (BCM-HGSC, https://www.hgsc.bcm.edu)(Retshabile et al., 2018).

3.2.4.1 Pre-capture library preparation

The DNA was fragmented using 500ng of sample in 80µl volume on the Covaris M220 Focused-ultrasonicator (Covaris Inc., MA, USA). The conditions used included two cycles at 6-10°C, frequency sweeping mode, water quality testing function set to off, 10% duty cycle, 200 cycles/burst and time set at 180 sec. The quality of sheared DNA was checked to ensure that the average size is 200-300bp by gel electrophoresis. This was followed by DNA end-repair using the NEBNext End-Repair Module kit with 75µl of sheared DNA, 9µl end-repair 10X buffer and 5µl end-repair enzyme mix incubated at 25°C for 20 min. The end-repaired DNA was purified using AMPure XP beads (Beckman Coulter, CA, USA) following the manufacturer's instructions. Then 3′-adenylation was done using NEBNext dA-Tailing Module kit with 51µl of end-repaired DNA, 6µl of NEBNext dA-Tailing reaction buffer and Klenow Fragment (3'–5' exo) incubated at 37°C for 20 min. The A-tailed DNA was purified using AMPure XP beads as previously described. Illumine index adaptors ligation was done using 5µl Illumina index adaptors, 18µl quick ligase 5X buffer, 62µl A-tailed DNA and 5µl quick ligase enzyme incubated at RT for 20 min. It was then purified using AMPure XP beads as previously described. A pre-capture ligation-mediated PCR (LM-PCR) was done using 54µl of pure DNA-adaptor mix, 60µl of 2X KAPA HiFi HotStart ReadyMix and 2µl each of LM-PCR primer 1 (AATGATACGGCGACCACCGAGA) and LM-PCR primer 2 (CAAGCAGAAGACGGCATACGAG). The cycle parameters were 98°C denaturation for 45 sec, followed by six cycles of 15-sec denaturation at 98°C, 30-sec annealing at 60°C and 30-sec extension at 72°C, and then a final 1 min incubation 72°C. The quality of the PCR reaction was checked using 1µl on a 1.2% agarose gel. The LM-PCR product is purified using the AMPure XP beads kit protocol. The purified LM-PCR product is checked for size distribution and quantity on an Agilent Bioanalyzer 2100 (Agilent Technologies, Inc.). The quality of the LM-PCR product was distributed about 300bp and yielded 3-5µg.

3.2.4.2 Liquid exome sequence capture and Post-capture LM-PCR amplification

The VCRome 2.1 rebalanced probe manufactured by NimbleGen was used for exome capture according to the manufacturer's instructions. Its design covers 23,585 genes and 189,028 non-overlapping exons (July 2014 Ensembl annotation) + PKv2 covering 3643 Genes (2.5 Mbp).

The total design size was ~45 Mb capture target. A post-capture LM-PCR was done using similar conditions previously described in 3.2.4.1. The post-capture LM-PCR product is purified using the AMPure XP beads kit protocol. The purified post-capture LM-PCR product is checked for size distribution and quantity on an Agilent Bioanalyzer 2100. Good quality post-capture LM-PCR should distribute about 300bp and yield 500ng-1µg enriched library.

3.2.4.3 Whole exome sequencing

The whole exome capture library was sequenced using the HiSeq2000 for batches 1, 2 and 3, while the NovaSeq was used for batches 4, 5, and 6, both Illumina technology platforms. Whole exome data was recovered as FastQ files for each sample. Both equipment are capable of producing enough coverage to obtain ≥ 70% of the target covered at ≥ 80X sequencing.

3.2.5 HLA class I allelotyping

3.2.5.1 HLA typing from WES using HLAreporter

Recent advances in bioinformatics tools have led to many tools for high-resolution HLA typing from NGS data, including whole genome, whole exome and SNP sequence data (Bai et al., 2014; Bauer et al., 2018; Y. Huang et al., 2015; Larjo et al., 2017; C. Liu et al., 2013; Major et al., 2013; Nariai et al., 2015; Warren et al., 2012; Wittig et al., 2015). HLA class I alleles were determined from the whole exome sequence data using the HLAreporter (Y. Huang et al., 2015). HLAreporter examines exons 2 and 3 for HLA class I typing. Furthermore, to ensure accuracy, HLAreporter employs a high degree of stringency during the mapping process to the database, where only perfect matches are allowed. HLAreporter is reported to achieve 100% prediction accuracy at both the two- and four-digit typing resolution (Bauer et al., 2018; Y. Huang et al., 2015).

HLAreporter achieves accurate high-resolution HLA typing from WES data in a five (5) stepwise process (Figure 3-5): (1) mapping of WES short reads to a comprehensive HLA reference database; (2) classification of the short reads to a specific gene, e.g. HLA-A; (3) de novo assembly of the classified reads to generate contigs with high sequence read coverage (>5X); (4) the assembled contigs are queried against a major database containing known HLA class I allele sequences (exon 2 and 3) to determine the candidate alleles; and finally (5) the

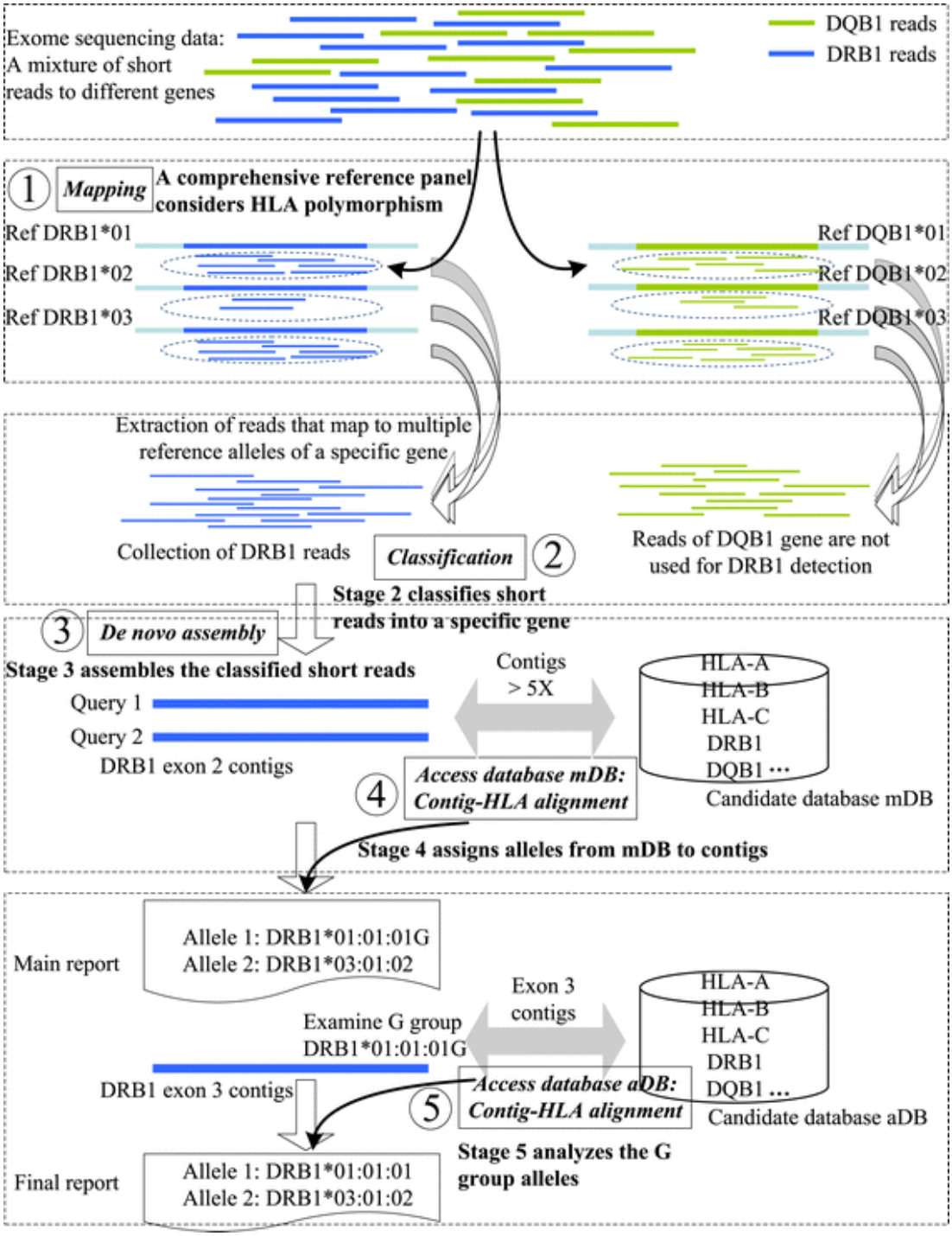


Figure 3-5: HLAreporter detection workflow.

In this example, HLA-DRB1 gene detection is demonstrated. Stages 1 and 2 show the classification of reads to a specific gene using the CRP panel-based, while stages 3, 4, and 5 show the assembly and contig-HLA matching process. Adapted from Huang et al. Genome Medicine (2015) 7:25 'Reproduced with permission from Springer Nature'.

gene-specific exon 3 contigs are queried against an additional database (with exon 3 sequences) for the determination of the G groups. HLAreporter was installed onto a high-speed large storage server cluster in BCM and/or locally. FastQ files were processed through this pipeline following the user manual. Only alleles that pass the minimum quality metrics of the platform were reported, i.e. ≥98% reads at 20X depth coverage(Y. Huang et al., 2015).

3.2.5.2 Validation of HLAreporter

We are mindful of the current gold standard for HLA typing. Therefore HLAreporter-derived data was validated using a subset of 30 samples (15 cases and 15 controls) by PCR-SSP typing with a Micro-SSP typing kit (One Lambda, CA, USA) (Marsh et al., 2010). The Micro SSP™ D-Mix and Micro SSP™ primers were thawed at RT and vortex at high speed to mix well. The negative control reaction tube was constituted by adding 1.0µL of DNA diluent to the primer set tray. Then 5 units/µl of recombinant Taq polymerase was added to the Micro SSP™ D-mix tube and vortexed for 5 sec, followed by a short pulse spin. Then 9.0 µL of the Micro SSP™ D-mix was added to the negative control reaction tube. The sample DNA was titrated according to concentration, added to the Micro SSP™ D-mix tube, vortexed for 5 sec, and pulse spun. Then 10µL of the sample-reaction mixture from the Micro SSP™ D-mix tube are added into each reaction tube, except the negative control reaction tube, of the Micro SSP™ primer set tray. The tray was sealed with parafilm seals and placed in the thermal cycler. The PCR conditions used included one cycle of 2 steps 130 sec at 96°C and 60 sec at 63°C, then nine cycles of 60 sec at 63°C followed by 20 cycles of 3 steps 10 sec at 96°C, 50 sec at 59°C and 30 sec at 72°C and final cycle at 4°C. The PCR reaction was transferred to a 2.5% agarose gel and electrophoresed for approximately 3-5 min. A gel photograph was interpreted using the provided worksheet.

3.2.6 HLA structural and functional studies

3.2.6.1 HLA structural homology modeling and validation

The HLA-C*03:02 3D structure model was predicted using SWISS-MODEL (Waterhouse et al., 2018). The 366bp HLA full protein sequence containing the three α and β_2-microglobulin chains was downloaded from the IMGT/HLA database (Robinson et al., 2019). To build the

model we selected the best template with ≥90% sequence identity, ≥70% coverage and high resolution (≤2Å) structure. The stereochemical analysis and spatial features were used to validate the model. Stereochemistry analysis of parameters such as bond length, torsion angle and rotational angle in the model was evaluated using the ERRAT, PROVE, Pro-check, and Pro-Q scores (Colovos & Yeates, 1993; *UCLA-DOE LAB — SAVES v6.0*, n.d.). ERRAT analyzes the statistics of non-bonded interactions between distinct atom types and depicts the value of the error function vs location of a 9-residue sliding window, which is generated by comparing statistics from highly refined structures. The ERRAT score is a test statistic of non-bonded interactions between atoms different types (C, O, N), and plots the value of the error function against the position of a 9-residue sliding window. The error is calculated by a comparison with statistics from highly refined structures in the PDB. PROVE predicts the quality of the model using Voronai radical planes. The volumes of atoms in the model are calculated using an approach that treats the atoms as hard spheres and computes a statistical Z-score deviation for the model from PDB-deposited structures that are well resolved (2.0 or higher) and refined (R-factor of 0.2 or better). The available error warning and pass of the modelled structure are defined by Pro-Check. It analyses residue-by-residue geometry and overall structural geometry to determine the stereochemical quality of a protein structure. Pro-Q predicts the quality of the modelled score in terms of LGscore. The stereochemical stability was confirmed with the RAMPAGE server with a Ramachandran plot (Lovell et al., 2003).

The spatial features of the model based on the 3D conformation were analyzed using the Verify 3D and ProSA scores (Eisenberg et al., 1997; Wiederstein & Sippl, 2007). VERIFY3D examines the compatibility of the predicted 3D model with its 1D protein sequence to determine whether the model is correct or misfolded. With Verify 3D, at least 80% of the amino acid residues must have a 3D score of 0.2. Verify 3D assigns a structural class based on its location and environment (i.e. alpha, beta, loop, polar, nonpolar, etc.) and compares the findings to excellent structures to determine the compatibility of an atomic model (3D) with its own amino acid sequence (1D). Pro-SA predicts the z score of the quality of the model in comparison with other experimentally determined models. ProSA uses knowledge-based energy based on the atomic coordinates to evaluate the overall model accuracy. Regions with higher-than-expected residue energy representing potential errors in the model are displayed on an energy plot. The overall quality of the model was determined from the ProTSAV score,

which is a consolidated score of model structural parameters from various model assessment tools compared to experimentally determined structures (A. Singh et al., 2016).

3.2.6.2 HIV-1 ligand prediction and preparation for docking

The HLA-C*03:02 binding motif was previously predicted based on the published motif of HLA-C*03:04 with shared sequence homology of the binding pocket residues (Yusim et al., 2003). The HIV peptides that are likely to form stable complexes with these class I alleles were determined using NetMHCpan 4.0. NetMHCpan 4.0 (www.cbs.dtu.dk/services/NetMHCpan) uses artificial neural networks to predict 8-14mer peptides that bind appropriately to given HLA molecules. Supplementary epitopes were predicted with Motif Scan (Reynisson et al., 2020; Yusim et al., 2003). We explore peptides along all known viral structural and non-structural proteins encoded by the gag, pol, vif, vpr, tat, vpu, rev, env and nef genes of HIV-1 group M. All peptides classified as strong or weak binders with rank thresholds above 2.000 and 0.500, respectively, were be selected for further analysis (Hoof et al., 2009; M. Nielsen et al., 2007).

Only peptides predicted to undergo canonical MHC class I antigen processing via proteasomal cleavage using NetChopv3.0 were considered further (Keşmir et al., 2002; Reynisson et al., 2020). The 3D structure of the selected linear peptides was predicted using the PEP-FOLD3 server and the structures with the lowest sOPEP energy value were chosen for molecular docking (Lamiable et al., 2016). The energy minimization of the ligands was done using the minimize structure module of Chimera and the charges were added using the Amber ff14SB force field. We performed 100 steps of steepest descent minimization followed by 1000 steps of conjugate gradient minimization (Pettersen et al., 2004).

3.2.6.3 Molecular docking and analysis

Model refinement was done using molecular dynamic and structural averaging with PREFMD (Heo & Feig, 2018). The prepared HLA-C*03:02 structure was used for docking HIV peptides with DINC. DINC is a parallelized meta-docking method that uses incremental docking based on AutoDock Vina. DINC is designed to dock large ligands (peptides) with high number of rotational bonds in the HLA class I molecules (Antunes et al., 2017) In our crossdocking protocol we adopted some modification to the default file. Using AutoDock Tools, a grid box

of 50 x 40 x 72 xyz points with a grid spacing of 0.375 was constructed and centred at 11.95 x 57.95 x -6.34 around the six binding pockets (El-Hachem et al., 2017). The vina exhaustiveness was set to 8 to enhance docking accuracy, and the number of binding modes created at each round of incremental docking was set to 40. To get a broader docking sampling, an extra round of docking was done employing the whole ligand with complete flexibility. Molecular docking studies were carried out in parallel on the African Centers of Excellence in Bioinformatics High Performance Computing core (320 cores, 1.25 TB RAM) (Jjingo et al., 2022) and the Uganda Virus Research Institute Compute Cluster.

All the resultant predicted ligand poses were rescored using Convex-PL, which has been found to find the best binders with greater than 80% accuracy (Kadukova & Grudinin, 2017). The top ranked ligand poses with the lowest Convex-PL score were selected as possible epitopes for future investigation. By redocking HLA-C*06:02 with the co-crystalized natural ligand, we verified our docking approach (ARTELYRSL). The top-ranking pose with an RMSD less than 2Å compared to the native peptide was judged successful docking. The intermolecular (HLA-peptide) interactions were identified and analyzed using molecular visualization software UCSF ChimeraX (Pettersen et al., 2004, 2021).

3.2.6.4 Molecular dynamics (MD) simulation and analysis

MD simulations were performed on the top eight ligands with total energy ≤10 kcal/mol from molecular docking using GROMACSv2020.3 and the CHARMM36 all-atom force field (Abraham et al., 2015; Best et al., 2012). The receptor-ligand coordinates generated during molecular docking were utilized in the re-construction of protein-ligand complexes using Chimera. All hydrogen molecules were removed from the final structure. We used the Avogadro program to add hydrogens to ligands and the CHARMM36m program to generate ligand parameters and topologies (Hanwell et al., 2012; Vanommeslaeghe et al., 2009). The resultant HLA-C*03:02-ligand complex was solvated in the center of a cubic unit cell of volume of 10000nm^3 with ~31,000 molecules of TIP3-point water. We allowed a minimum distance of 1nm between the box boundary and the complex. The system was neutralized with the addition of 10 Na+ ions. The system was subjected to energy minimization using the steepest descent method with maximum force constraint of 10kJ/mol. Position restraints were applied on both the ligand and HLA-C*03:02 receptor. The system temperature and pressure

were equilibrated at 300K using the modified Berendsen thermostat coupling method and at $4.5x10^{-5}bar^{-1}$ using the Berendsen coupling method respectively for 100ps. All MD simulations were run in duplicate for 200ns using periodic boundary conditions without ligand-protein restraints. The Parrinello-Rahman barostat and modified Berendsen thermostat were used during MD productions. The short-range Van der Waals and Coulomb interactions cutoffs were set at 1.2nm and the long-range electrostatic interactions between atoms were calculated with particle-mesh Ewald method. The bond lengths were constrained using the LINCS algorithm. All MD production runs were done using Amazon Web Services' GPU-backed EC2 (Fusaro et al., 2011) and the Nottingham Trent University Hamilton and Avicenna high performance computing cluster. The stability of the complexes was examined by analyzing hydrogen bonds, and root mean square deviation (RMSD) using GROMACS functions g_hbond, and, g_rmsd respectively(Abraham et al., 2015).

3.2.6.5 Peptide positional conservancy

In order to exhaustively evaluate the positional conservancy of the potential epitopes down to each individual amino acid residue, we used the AL2CO sequence conservation analysis server (http://prodata.swmed.edu/al2co/) (Pei & Grishin, 2001). Our approach entailed the utilization of an alignment file that originated from a compilation of HIV-1 clades A, C, D, and K representative sequences sourced from Africa. These sequences, along with their recombinant counterparts, were sourced from the LANL HIV-1 Sequence Database (https://www.hiv.lanl.gov/content/index). The objective was to derive conservancy scores as a means of quantifying the level of residue conservation within the candidate epitopes.

3.2.7 Experimental validation

3.2.7.1 Peripheral blood mononuclear cell (PBMC) isolation

Peripheral blood mononuclear cells (PBMC) were isolated by density gradient centrifugation method using Histopaque from whole blood collected in EDTA tubes (Figure 3-6). Ficoll Histopaque was brought to RT. 5ml of Ficoll was added to a 15ml centrifuge tube. Then, 4ml of blood was gently overlaid on top of the Ficoll while carefully preserving the blood and Ficoll layers. The, the tube were centrifuged for 30min at 100xg at 4°C in a swing-out rotor bucket. Then aspirated off the whitish buffy coat containing the PBMCs formed in the layer between

histopaque and the medium into a sterile 15ml tube. The cells were washed twice with 10ml of sterile phosphate buffered saline (PBS) while centrifuging at 100g for 10 minutes removing the supernatant. The cells pellet was reconstituted in 1ml of PBS.

The number of cells isolated, and their viability was established using trypan blue which stains dead cells blue while the live cells remain shinny. All cells are counted and expressed as number of cells per cm^3.

$$\text{Cell viability} = \frac{\text{total no. of live cells}}{\text{total no. of cells (live + dead)}} \times 100$$

3.2.7.2 PBMC cryopreservation

Freezing medium (80% heat-inactivated fetal bovine serum [FBS]; 20% DMSO) was brought to RT. The Ficoll-isolated PBMCs from the above procedure were adjusted to $2x10^7$ cells/ml by adding freezing medium dropwise along the tube wall. Then mixed well by reverting tube several times. We then incubated the cell/FBS/DMSO mix for 10-15 min on ice to allow DMSO to diffuse into the cells before freezing. The tubes were transferred quickly to a freezer overnight at -80°C before freezing in liquid nitrogen until required.

3.2.7.3 PBMC thawing protocol

Thawing medium R20 (80% RPMI medium; 20% FBS) was brought to RT. For thawing of cells, the PBMC stocks were removed from liquid nitrogen and transported on dry ice and quickly transferred to the water bath. When they were thawed half liquid/half solid, the cells were poured in 30ml of prewarmed thawing medium (R20) in a 50ml Falcon tube. The cells were centrifuged at 700 x g for 10 min at RT with brake (Figure 3-6). The supernatant was discarded, and the cell pellet resuspended in 10 ml fresh R20 medium using a 5ml serological pipet. The cells were counted as previously described and rested in Falcon tube (we loosened the lid of the tubes to allow gas exchange) for 18-24 h. After resting the cells, for the FluoroSpot, the cells were centrifuged at 700 x g for 10 min. The cell pellet was resuspended cells in 1 ml of fresh R20 medium. The cells were counted and dilute to achieve a concentration 200,000 cells/100µl as shown below.

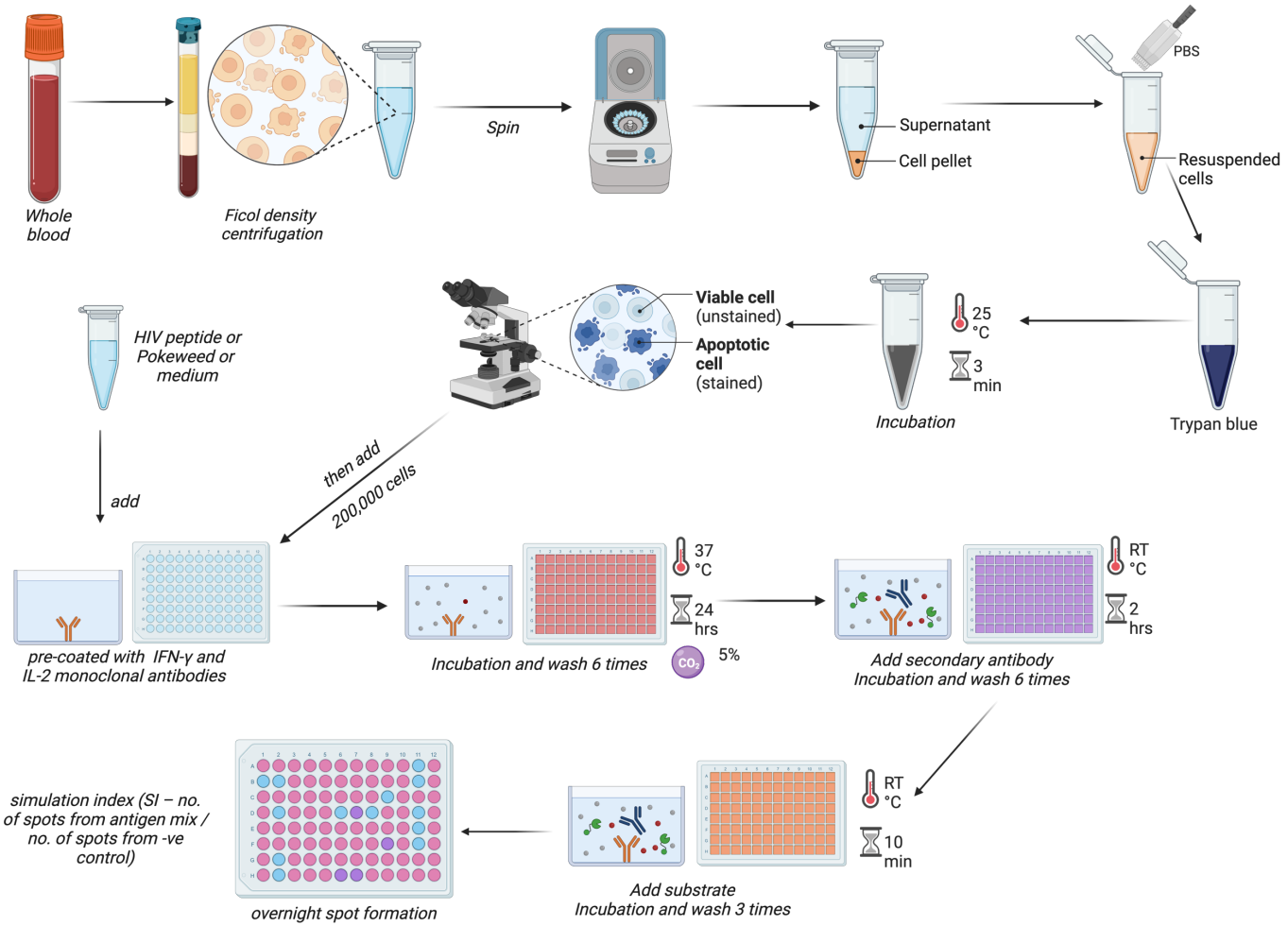


Figure 3-6: A schematic workflow for the experimental validation of the HLA-C*03:02 epitopes

$$\text{volume of R20} = \frac{counts\ in\ 1ml\ in\ millions \times \text{no. of cells/well} \times \text{no. of wells (/experiment)}}{\text{desired concentration of cells} \times \text{no. of well (/experiment)}}$$

3.2.7.4 $CD8^+$ T cell stimulation assay (IFN- γ and IL-2 Enzyme-Linked Immunosorbent Spot Assay)

HIV-1 specific HLA-C*03:02-restricted $CD8^+$ T cell responses were evaluated using an enzyme-linked immunosorbent spot (ELISPOT) assay (Figure 3-6). Specifically, the secretion of IFN-γ and IL-2 by PBMC using the dual color ELISPOT assay according to the manufacturer's instructions. HIV peptides were synthesized by Fmoc technology, and the purity confirmed using high pressure liquid chromatography. Peptides were diluted to a final concentration of 5µg/ml or 2µ/ml. PBMCs count and viability were confirmed; rested overnight before plating. All reagents were warmed to RT before use. First, 96-well plates pre-coated with both IFN-γ and IL-2 monoclonal antibodies were incubated with 100µl of 2 X 10^5 viable cells per well at 37°C in 5% CO_2 for 12hrs (or overnight). Then, 100µl of peptide solution (Figure 3-7) was added and incubated for 20h at 37°C in 5% CO_2. The wells were emptied and washed six times with 200µl of washing buffer (WP) per well. After the last washing step the plates were blotted to remove excess liquid. We added 100µl of diluted secondary antibody per well and incubated for 2 hours. The washing was repeated as stated above. We added 100µl of substrate solution to each well and incubated for 20 minutes until spots become visible. After sufficient spot development, the reaction was stopped by washing with three times with sterile water. The plates were blotted and dried properly. Media alone was used as a negative control and pokeweed as a positive control. IFN-γ and IL-2 production was directly visualized and expressed as a simulation index (SI – number of spots from antigen mixes / number of spots from negative control). An SI >3 was interpreted as a positive reaction. SI between 2 and 3 was interpreted as borderline while SI <2 was considered negative.

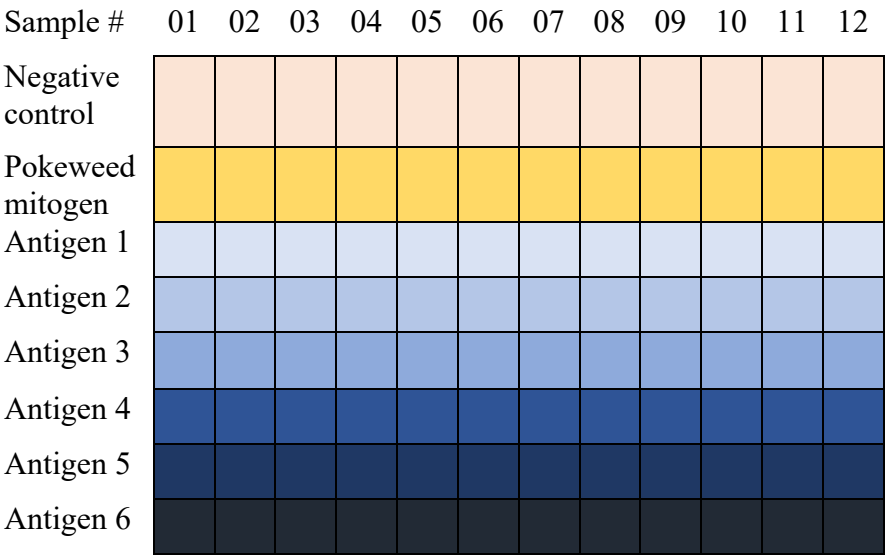


Figure 3-7: The dual-color ELISPOT plate layout

3.2.7.5 HIV-1 genotyping

Genomic DNA was extracted from whole blood with the PaxGene DNA blood kit (Qiagen, Germany) as previously described. Using HIV proviral DNA as a template, a three round nested PCR assay was performed targeting the Gag-Pol region. The primary primers were CP2F: GTTAAACAATGGCCATTGACAGAAGA and RTR2: TGTATRTCATTGACAGTC CA, the second-round primers were RTF2wd: CAGAAGARAAAATAAAAGCATTA and RT3271R: ACTGTCCATTTRTCAGGATG while the third-round nested primers containing Illumina-specific adaptor sequences were ILRT2796F: **TCGTCGGCAGCGTCAGATGTG TATAAGAGACAG**AGAACTCAAGACTTYTGGGA and ILRT3271R: **GTCTCGTGGG CTCGGAGATGTGTATAAGAGACAG**ACTGTCCATTTRTCAGGATG. All cycling conditions involved an initial 2 min denaturation at 95°C, followed by 35 cycles, each consisting of a 30 sec denaturation at 98°C, a 30 sec annealing at 60°C, and a 1 min extension at 72°C. These were followed by a 15 min extension at 72°C. The final PCR product was purified using the Agencourt AMPure XP magnetic beads (Beckman Coulter). The purified PCR was used for library preparation using the Nextera XT DNA Library Preparation Kit (indexing was done with the IDT for Illumina DNA/RNA UD Indexes Set A). Equimolar concentrations of all samples were pooled and sequenced on an Illumina **MiSeq** instrument with by paired-end (2x300bp) method using the MiSeq v3 reagent kit (Illumina Inc, CA, USA). The read quality of the generated FastQ files was assessed using FastQC v0.12.0 and the low-

quality sequences were trimmed using Trimmomatic v0.39. The resultant reads were aligned/mapped to HIV reference sequence (RefSeq: NC_001802) using the BWA-MEM (H. Li & Durbin, 2010) to generate viral contigs. HIV subtyping was done using the REGA-v3 HIV-1 Subtyping Tool (Pineda-Peña et al., 2013).

3.3 Statistical Analysis

3.3.1 Objective 1

The baseline characteristics included age at enrolment, gender, ethnicity, height/length and weight at enrolment, year of birth, year of enrolment, status in the clinic (alive or dead), date of ART and baseline pre-ART, WHO stage, viral load and $CD4^+$ T cell count, and percentage were analyzed were descriptively analyzed as overall, by country. Anthropometric data was expressed into weight-for-age, height/length-for-age, BMI-for-age and weight-for-height z-scores. Continuous data were summarized as mean (standard deviation) or median (interquartile range, IQR) depending on the normality of the distribution of data. Categorical data were summarized descriptively using the frequency and percentages. We compared the distribution of the data between the two countries using the χ^2 test for categorical variables and the Wilcoxon rank–sum test for continuous data. Due to the non-random selection of participants by disease progression, we did not compare the distribution of participants by groups of LTNP and RP.

The primary outcome of the analysis was time-to-progression to AIDS, which was defined as either ART initiation (according to prevailing WHO HIV treatment guidelines) or an AIDS-defining illness, whichever occurred first. The median and 10-year AIDS-free survival was estimated using the parametric survival model under the Weibull distribution. In contrast, we used the Kaplan-Meier method to estimate the rate of loss of LTNP status (i.e., the time-to-progression since attaining LTNP status).

Competing risk regression analysis was used to estimate the subdistribution hazard ratios for the candidate predictors of progression to AIDS, accounting for death and LTFU as competing events which preclude the observation of primary outcome. All factors with $p < 0.2$ at univariate analysis were entered into a multivariate competing risk regression model, except CD4 counts, a clinical measure of progression to AIDS. We accounted for clinically significant

interaction terms and selected the final model with the lowest Bayesian Information Criterion using the backwards-elimination method. In sensitivity analysis, we considered a parametric survival model, first allowing for the continuous predictors to vary with time using restricted cubic splines and subsequently keeping them fixed over time. We considered a two-tailed adjusted p-value <0.05 as statistically significant. We used STATA17 and R statistical software for all analyses.

3.3.2 Objective 2

Demographical data, including ethnicity, gender, age, and time to clinical progression of the 1000 study participants, was managed as described in 3.3.1 and compared between LTNPs and RPs. According to the user manual, the HLA class I allelic data obtained from HLAreporter was imported into PyHLA. The HLA class I allele frequencies were estimated using the expectation-maximization method in PyHLA (Fan & Song, 2017; A. K. Lancaster et al., 2010). We performed allele-level association, taking into consideration variants with frequencies greater than 1% in the cohort (Naruto et al., 2012). The differences in allelic frequencies were tested by the χ^2 test. Odds ratios were calculated with Haldane's correction of Woolf's method, assuming additive effects of the alleles. The p values were considered significant ($p < 0.05$) after correction for multiple testing using the false discovery rate (FDR). We calculated adjusted ORs to account for the effect of country and gender using logistic regression models. We performed zygosity analysis to determine the influence of homozygosity and heterozygosity of significant alleles on disease progression both overall and using survival analysis for time to outcome (LTNP vs RP).

The Haplo.Stats package (v1.7.9) was used to estimate the haplotype frequencies using the haplo.em function. We tested the association between haplotypes and LTNP under additive haplotype effect models. Haplotypes with frequencies greater than 1% were used during the association analysis. The haplo.score function that calculates haplotype-specific scores was used to test the association of individual haplotypes with LTNP. The haplo.glm and haplo.cc functions were used to calculate odds ratios using generalized linear regression models relative to the most frequent haplotype adjusting for non-genetic factors (gender and country). A p-value <0.05 was considered statistically significant. We estimated the population attributable

risk (PAR) due to the significant alleles and haplotypes using odds ratios from the logistic regression models. PAR was calculated according to Levin's formula:

$$PAR = \frac{p(OR-1)}{1+p(OR-1)} \; X \; 100;$$

where p denotes the frequency of alleles or haplotypes among LTNPs. Confidence intervals were calculated from the upper and lower limit of the odds ratios (ORs) of the variants. The total PAR was calculated as

$$\sum PAR = \frac{p(OR-1)}{1+p(OR-1)} \; X \; 100.$$

Finally, we evaluated the statistical power of the study using Quanto v1.2.4. We considered a two-sided significant level of $p<0.05$, minor allele (haplotype) frequency of 1%, and the population prevalence of LTNP as 0.04%. We had more than 80% power to detect genetic effect of >1.8 for alleles and haplotypes, respectively under an additive genetic model. All statistical analysis was performed using STATA v13 and R software v1.2.5033.

3.3.3 Objective 3

For this particular study objective, statistical analysis was deemed unnecessary due to the nature of the data generated. The objective focused on descriptive analysis and did not involve quantitative comparisons or hypothesis testing.

CHAPTER 4: RESULTS

4.1 Long-term non-progression and factors associated with disease progression among children living with HIV in Botswana and Uganda: a retrospective cohort study

Samuel Kyobe[1], Grace Kisitu[2], Savannah Mwesigwa[1], John Farirai[3], Eric Katagirya[4], Gaone Retshabile[5], Lesedi Williams[5], Angela Mirembe[2], Lesego Ketumile[3], Misaki Wayengera[4], John Mukisa[4], Gaseene Sebetso[5], Thabo Diphoko[5], Marion Amujal[4], Edgar Kigozi[4], Fred Katabazi[4], Ronald Oceng[2], Busisiwe Mlotshwa[5], Koketso Morapedi[5], Jacqueline Kyosiimire[6], Betty Nsangi[2], Edward Wampande[4], Masego Tsimako[7], Chester Brown[8], Ishmael Kasvosve[9], Moses Joloba[4], Gabriel Anabwani[2], Sununguko Mpoloka[5], Graeme Mardon[10,11], Dithan Kiragga[2,14], Neil Hanchard[10,12,13], Mogomotsi Matshaba[3,14], and Adeodata Kekitiinwa[2,14] on behalf of the Collaborative African Genomics Network (CAfGEN) of the H3Africa Consortium

[1]Department of Medical Microbiology, College of Health Sciences, Makerere University, Kampala, Uganda; [2]Baylor College of Medicine Children's Foundation Uganda, Kampala, Uganda; [3]Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana; [4]Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, Kampala, Uganda; [5]Department of Biological Sciences, University of Botswana, Gaborone, Botswana; [6]MRC/UVRI & LSHTM Uganda Research Unit; [7]Australian Medical Research, Hurstville, New South Wales 2220, Australia; [8]University of Tennessee Health Science Center, Memphis, Tennessee 38105, USA; [9]School of Allied Health Professionals, Faculty of Health Sciences, University of Botswana, Gaborone, Botswana; [10]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; [11]Department of Pathology and Immunology, Baylor College of Medicine, Houston, Texas, USA; [12]USDA/ARS/Children's Nutrition Research Center, Baylor College of Medicine, Houston, Texas, USA; [13]Childhood Complex Disease Genomics Section, Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; [14]Pediatric Retrovirology, Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA

Corresponding authors:

Neil A. Hanchard; Childhood Complex Disease Genomics Section, Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; neil.hanchard@nih.gov, +1 (301) 594-2151

Mogomotsi S. Matshaba; Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana; mmatshaba@baylorbotswana.org.bw, +267 319 0083

Adeodata R. Kekitiinwa; Baylor College of Medicine Children's Foundation Uganda, Kampala, Uganda; akekitiinwa@baylor-uganda.org, +256-417-119100

Highlights

The proportion of long-term non progression has not been well studied especially in African populations.

Around 7% of children infected with HIV were characterized as long-term non progressors.

The rate of HIV progression varies among childhood cohorts in Africa.

Undernutrition is an independent predictor of HIV progression at enrolment into care.

4.1.1 Abstract

Objective: We utilize a large retrospective study cohort derived from electronic medical records (EMR) to estimate the prevalence of LTNP and investigate factors associated with progression among children in Botswana and Uganda.

Methods: EMRs from large tertiary HIV clinical centers in Botswana and Uganda were queried to identify LTNP children 0–18 years enrolled between June 2003 and May 2014 and extract demographic and nutritional parameters. Multivariate subdistribution hazard analyses were used to examine demographic factors and nutritional status in progression in the pre-ART era.

Results: Between the two countries, 14,246 ART-naïve children infected with HIV were enrolled into clinical care. The overall proportion of LTNP was 6.3% (9.5% in Botswana vs. 5.9% in Uganda). The median progression-free survival for the cohort was 6.3 years, although this was lower in Botswana than in Uganda (6.6 vs 8.8 years; $p<0.001$). At baseline, the adjusted subdistribution hazard ratio (aHR_{sd}) of progression was increased among underweight children (aHR_{sd} 1.42; 95% confidence interval [CI]: 1.32–1.53), enrolled after 2010 (aHR_{sd} 1.32; 95%CI 1.22–1.42), and those from Botswana (aHR_{sd} 2; 95%CI 1.91–2.10).

Conclusion: In our study, the prevalence of pediatric LTNP was lower than that observed among adult populations, but progression-free survival was higher than expected. Stunting, wasting, year of enrolment into care, and country of origin are independent predictors of progression among children.

Keywords: LTNP, HIV progression, pediatric, Africa,

4.1.2 Introduction

Sub-Saharan Africa remains the global epicenter of the HIV epidemic; more than 70% of all people living with HIV/AIDS are in Africa. In 2021, UNAIDS estimates indicated that 150,000 new HIV infections and 99,000 deaths occurred among African children.[1] Before the initiative of universal antiretroviral therapy (ART), it was widely noted that some children would remain AIDS-free for more than ten years and could maintain normal-for-age $CD4^+$ T cell counts[2] – so-called long-term nonprogressors (LTNP). Children capable of controlling HIV infection present the opportunity for unique insights into the natural host immune responses, which is critical for the development of novel therapeutics and vaccines.[3] LTNP has been noted among adults, particularly from Western populations; however, most studied individuals were infected via horizontal transmission, necessitating an estimated date of seroconversion. In contrast to adults, vertically HIV infection offers an objective estimate for progression-free survival (i.e. majority of infections occurs [or detectable] within 45 days of birth while the remainder occur later during breast feeding or maternal seroconversion while breastfeeding);[2,4–6] added to the temporally fewer, yet measurable, exposures of childhood (e.g. less environmental, age-related or lifestyle confounders such smoking),[7] pediatric cohorts offer a unique opportunity to estimate the proportion of LTNP robustly. Further, the rapidly developing immunological system in children, forged through new exposures to infectious agents, is another crucial difference in pediatric HIV progression. For instance, children infected with HIV produce stronger de novo autologous HIV variant-specific $CD8^+$ T cell lymphocyte (CTL) responses than adults,[8,9] and LTNP children are known to have more neutralizing antibodies and immune quiescence than their adult counterparts.[10] The factors associated with progression include host-, pathogen-, and environmental factors, which vary widely between populations; thus, understanding the demographic, clinical, and laboratory factors that contribute to HIV disease progression in general among children, particularly population-specific characteristics, could help to facilitate prioritization of resources, especially in low-and-middle-income countries.

As is the norm in adult literature, the few studies on the prevalence of LTNP in children have been focused on Western populations.[11–14] A thorough examination of that literature, however, uncovers substantial clinical heterogeneity and small sample sizes among contributing studies, which limit the generalizability and applicability of the results. Nielsen et al., for example,

estimated a 27% rate of LTNP among 143 children vertically infected with HIV. In contrast, Martin et al. observed a substantially lower estimate of 2.1% among 422 newborns in the USA classified as LTNP, indicating the need for larger, more definitive studies.[11,14] Most crucially, most studies have employed non-uniform definitions of LTNP, including those among adult populations, which makes comparisons with other populations difficult.[13] As a result, the generalizability of results from LTNP studies in adults to children is limited and compounded by the perceived limitations of pediatric studies - small available sample sizes, limited follow-up time, and a lack of consensus definitions of LTNP.[13] Therefore, there is very limited data on pediatric LTNP following vertical infection, especially among African populations with the highest disease burden.[1] Here, we leverage historical electronic medical record (EMR) data from two large clinical centers in Botswana and Uganda to determine the prevalence of pediatric LTNP and identify factors associated with progression.

4.1.3 Materials and Methods

Study population

The Baylor International Pediatric AIDS Initiative (BIPAI) network is an open, prospective cohort of children infected with HIV initiated in 2001 with nine centers of excellence, six in Africa, one European and two South American. This study considered participants enrolled by the Botswana-Baylor and Baylor-Uganda Children's Clinical Centers of Excellence (COEs, Figure 3-1), the original centers of BIPAI and the founding clinical centers for the Collaborative African Genomics Network (CAfGEN). CAfGEN is a multi-disciplinary, multi-institutional, inter-and intra-country network of African and American scientists, clinicians, and researchers. The principal aim of CAfGEN is to utilize genomics and related approaches to study gene-environment interactions in HIV/AIDS and its comorbidities among diverse African pediatric populations. This study was approved by Research Ethics Committees in Uganda, Botswana, and the USA.[15]

Study design

From mid-2003, children aged 0–18 who were vertically infected with HIV were prospectively enrolled in care at the COEs. For older children, the potential for vertical HIV transmission was ascertained by history of maternal HIV infection or death due to HIV-related causes

(excluding all those with self-reported sexual activity).[16] It should be noted that in both countries, early infant diagnosis of HIV programs with PCR testing were initiated between 2005 (Botswana) and 2006 (Uganda).[17,18] Socio-demographic and clinical characteristics from case report files are documented in the BIPAI EMR, and data entry clerks independently validate the data accuracy. Follow-up visits are scheduled monthly for the first three months and quarterly after that to evaluate progression and guide ART initiation according to national ART guidelines. Additionally, community health liaisons maintain regular telephone contact with the participant families. For this analysis, we pooled data from eligible children enrolled in care up to the official censoring date of May 31, 2014. Therefore, we censored all children that had not progressed or initiated on ART. Only children who were ART-naïve at enrolment were included in the time-to-progression analysis. Childrenn coming into care in 2010 and are less than one year old would be placed on ART, this was also revised upwards to 2 years by 2012. These participants were removed since the reason given was solely based on age eligibility (part of 878 ART-experienced). This study adheres to the STROBE reporting guidelines for observational studies.

Demographic data collected included sex, dates of birth and enrolment, country, loss-to-follow-up (LTFU), transfers between clinics, and death. Death was determined following a home visit or proxied to the last known date alive if not otherwise defined as LTFU. Clinical data included weight, height (or length), and WHO clinical stage at enrolment in care. Although $CD4^+$ T cell (CD4) testing became available in 2002, it was only intermittently performed (i.e., semiannually) after 2010. Therefore, the earliest (i.e., within six months of enrolment) recorded CD4 count and percentage before ART were utilized as a baseline. Pre-ART viral load testing was not readily available for this cohort except for those enrolled in specific studies. Weight-for-age (WFA), length/height-for-age (HFA), and body mass index (BMI)-for-age (BFA) z-scores were computed according to WHO Child Growth Standard.[19] The WHO z-scores below two standard deviations of the population normal WFA, HFA, and BFA categorize undernutrition as underweight, stunted, and severely malnourished, respectively. The cohort was divided into three enrolment categories (pre-2006, 2006–2010, and post-2010 cohorts) based on the implementation of new national and WHO ART guidelines. The principal difference between the treatment guidelines was changes in the CD4 thresholds for starting ART from <200 cells/ml pre-2006 to ≤350 cells/ml 2006-2010 cohort. Also, 2010 heralded the

universal recommendation to start children under three years on ART irrespective of their CD4 status. LTNP was defined conservatively as ART-naive asymptomatic HIV infection or having a CD4 count >500 cells/ml (or CD4% ≥25%) where available for at least 10 years. To distinctly differentiate between long-term survivors and rapid progressors with reduced ambiguity, we used the age cutoff to 10 years. Rapid progression was defined as either two or more CD4% <15%, an AIDS-defining illness (WHO stage III or IV) or ART initiation within three years of birth (infection).[4] For the purpose of deriving factors associated we defined progression as ART initiation according to HIV treatment guidelines. The date of birth serves as the origin or start time, while the date of progression or censoring marks the end times in the analysis.

Statistical analyses

The primary end-point was time to progression of HIV infection from perinatl infection which is estimated to occur within the first 45 days of life. Baseline characteristics at enrolment were summarized descriptively for the entire cohort and compared between countries. Continuous data were assessed for normality (skewness and kurtosis test), summarized as median (IQR), and compared using the Wilcoxon rank–sum test. Categorical data were described using frequencies and compared by χ^2 test between countries. The analysis's primary outcome was time-to-progression, defined as either ART initiation (according to prevailing WHO HIV/AIDS treatment guidelines) or an AIDS-defining illness, whichever occurred first. The median and 10-year progression-free survival were estimated using the parametric survival model under the Weibull distribution. In contrast, we used the Kaplan-Meier method to estimate the rate of loss of LTNP status.

Competing risk regression analysis was used to estimate the subdistribution hazard ratios for the candidate predictors of progression accounting for LTFU as competing event. All factors with $p < 0.2$ at univariate analysis were entered into a multivariate regression model, except CD4 counts which is a clinical measure of progression, and HIV viral load measurements, which were only available for 847 participants. We accounted for clinically significant interaction terms and selected the final model with the lowest Bayesian Information Criterion. We considered a two-tailed p-value <0.05 as statistically significant. We used STATA17 and R statistical software for all analyses.

4.1.4 Results

Baseline characteristics

Overall, 15,124 children were enrolled in care between 2002 and 2014, 14,246 of whom were ART naïve at baseline and were included in this analysis (Figure 4-1). More than half of the children were female (51%), with a similar sex distribution in Botswana and Uganda (Table 4-1). Most participants (58%) were in care by the age of 5 years, with a median age at enrollment of 3.7 years (IQR 1.1–8.4); however, children in Botswana tended to be older at enrollment than in Uganda (5.3 years; IQR:2.0–10.8 vs 3.5 IQR: 1.0–8.4 in Uganda; P<0.0001; Table 4-1). Most children in Botswana presented with advanced disease, and as a result, the median age at initiating ART was lower in Botswana than in Uganda (Table 4-1). We found potential differences in the median CD4 count, HIV RNA viral load, LTFU, mortality and BFA between the two countries. However, the distribution of the baseline HIV RNA viral load, WFA and HFA were similar (Table 4-1).

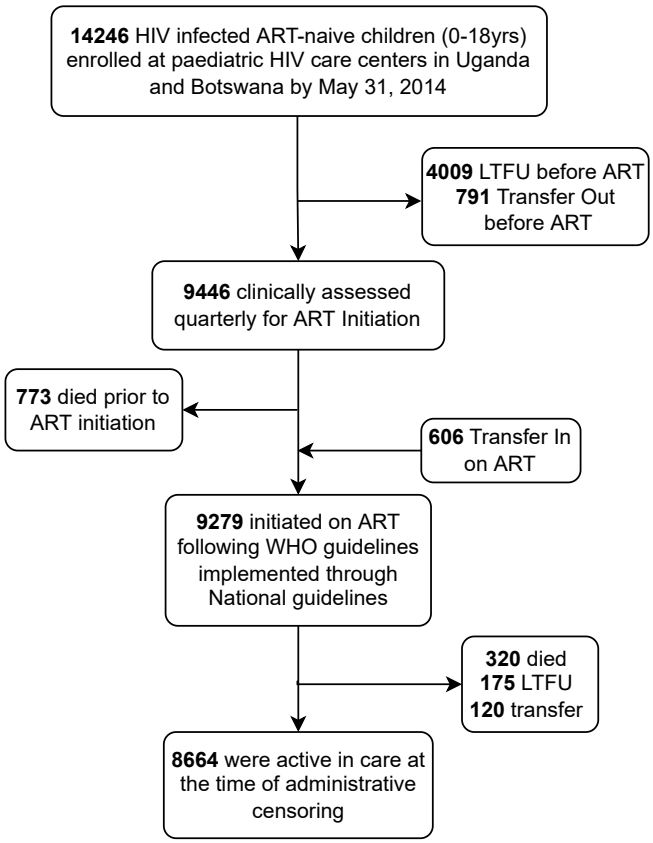


Figure 4-1: The flow of study participants into HIV care.

Proportion of LTNP among pediatric HIV populations in Africa

Selecting those children who were ART-naïve at enrolment resulted in 14,246 (94%) participants, the majority of whom (12,679; 89%) were from Uganda (Supp. Figure 1). This large difference between the countries is not unexpected, given the population sizes of Uganda (pop. 47.12 million; June 2022; World Population Prospects 2019) and Botswana (pop. 2.35 million; June 2022; World Population Prospects 2019).[20] The overall median progression-free survival was 6.3 years (95%CI: 6.1–6.4, Figure 4-2A; Uganda 6.5 years vs. Botswana 4.7 years Figure 4-2B), while the 10-year progression-free survival was 29.4% and varied between Botswana and Uganda (20.1% vs. 30.1%, Figure 4-2B). The 14,246 children studied contributed 84,475 person-years of infection. Among this group, 892 children were identified as LTNP, indicating a 6.3% prevalence. At the other end of the continuum, 28.4% (4,044) were RP. In Botswana, we found a 9.5% LTNP prevalence, which was higher than Uganda's 5.9%, which might be consistent with Botswana's older median age at enrolment in care. We then used a Kaplan-Meier estimate to examine the time progression from LTNP status. The median time to loss of LTNP status was 13 years (95%CI 12.7–13.1; Figure 4-2C). The loss of LTNP status was statistically faster among children in Botswana compared to Uganda (p = 0.0001; Figure 4-2D). Overall, the progression-free survival among LTNPs was 22.8% (95% CI 20.2–25.7) at 15 years and 1.1% (95% CI 0.6–2.1) at 20 years following vertical HIV infection.

Table 4-1: Comparison of demographic and clinical characteristics of children in Botswana and Uganda

Variable	Overall	Country		p–value
		Uganda	Botswana	
No. of Children	14246	12709 (89)	1537 (11)	
Sex, *n (%)*				0.144
Female	7295 (51)	6535 (51)	760 (49)	
Male	6951 (49)	6174 (49)	777 (51)	
Age at Enrolment, median (IQR) yrs	3.8 (1.1–8.4)	3.5 (1.0–8.1)	5.3 (1.3–9.4)	<0.001
<2	5250 (36.8)	4761 (37.4)	489 (31.8)	<0.001
2 to <5	2967 (20.8)	2722 (21.4)	245 (15.9)	
5 to <10	3375 (23.7)	2909 (22.9)	466 (30.3)	
10 to 18	2654 (18.6)	2317 (18.2)	337 (21.9)	
Year of birth, *n (%)*				<0.001

≤2004	8864 (62)	7654 (60)	1210 (79)	
>2004	5382 (38)	5055 (40)	327 (21)	
Year of enrolment, *n (%)*				<0.001
<2006	5099 (36)	4344 (34)	755 (49)	
2006–2010	6537 (46)	6027 (48)	510 (33)	
>2010	2610 (18)	2338 (18)	272 (18)	
WHO Stage, *n (%)*				<0.001
I or II	4853 (49)	4393 (51)	460 (34)	
III	3257 (33)	2746 (32)	511 (38)	
IV	1818 (18)	1457 (17)	361 (27)	
Anthropometric measurement, median (IQR)				
Weight–for–age, Z–score	-1.74 (-3.19– -0.43)	-1.7 (-3.2 – -0.3)	-1.8 (-2.97 – -0.7)	0.043
Height–for–age, Z–score	-1.76 (-2.89 – -0.55)	-1.7 (-2.9 – -0.5)	-1.6 (-2.6 – -0.6)	0.055
BMI–for–age, Z–score	-0.83 (-2.06 – 0.16)	-0.80 (-2.03 – 0.19)	-1.08 (-2.19 – -0.24)	<0.001
CD4 count (cells/ml)§	309 (101–597)	302 (90–603)	329 (159–559)	0.075
CD4%, median (IQR)†	17.0 (11.0–24.0)	17.0 (11.0–24.0)	19.2 (12.0–27.0)	<0.001
HIV RNA load (log10 copies/uL)	5.4 (4.8–5.8)	5.6 (4.9–5.8)	5.3 (4.7–5.7)	<0.001
Age at ART Initiation, median (IQR) yrs	6.3 (2.1–11.0)	6.4 (2.3–11.2)	5.7 (1.6–10.1)	<0.001
Below 5	3725 (43)	3073 (43)	652 (45)	<0.001
5 to <10	2313 (27)	1882 (26)	431 (29)	
10 to 18	2626 (30)	2248 (31)	378 (26)	
Clinical outcome, *n (%)*				
Alive	7998 (56)	6794 (53)	1204 (78)	
LTFU	4210 (30)	4126 (33)	84 (6)	
Transfer Out	911 (6)	911 (7)	0 (0)	
Death	1127 (8)	878 (7)	249 (16)	

Abbreviations: yrs – years, SD – standard deviation. p–values compare the distribution of each variable by country. *the data available for the variable is less than the total number of participants analyzed due to missing values.

§ for children above 60 months, † for children below 60 months

Note: All measurements were done at enrolment

Factors associated with HIV progression among children

Having established the proportion of LTNP among children, we then examined the predictors of progression on enrolment into care. Here we excluded 4,168 children with incomplete clinical records and examined the relationship between progression with multiple factors, including country, sex, age at enrolment, WFA, HFA, BFA, viral load, calendar year of birth, and calendar year of enrolment. In univariate analysis, male sex, country, age at enrollment, undernutrition (severe malnutrition, underweight, stunting), and being born after 2004 and enrolled after 2010 were associated with progression (Table 4-2). In multivariate analysis, underweight children at enrolment were independently associated with progression (adjusted subdistribution H.R. [aHR_{sd}][1.42, 95%CI 1.32–1.53; Table 4-2, Figure 4-3 and 4-4A) after adjusting for age, country, and enrolment cohort, while accounting for those who were LTFU. Although not statistically significant, children who were stunted at enrolment were at a higher risk of progression (aHR_{sd} 1.03, 95%CI 0.96–1.09; Figure 4-4B) than normal statured children. Children from Botswana had a 2-fold increased risk of progression compared to Uganda (aHR_{sd} 2.00, [95%CI: 1.91–2.10]). We observed a significant "cohort effect" where children enrolled in care between 2006-2010 and after 2010 progressed faster to AIDS (aHR_{sd} 1.14 and 1.13 respectively; Table 4-2). In a subset of 847 children with viral load measurements within six months of enrolment, multivariate analysis found that low BFA and enrolment after 2010 were independently associated with progression (Supp. Table 2, Figure 4-3 and 4-4C). Additionally, a high HIV RNA viral load was associated with an increased risk of progression, although not statistically significant (Figure 4-4D).

Table 4-2: Subdistribution hazard ratios of progression from competing risk models

Characteristics	Univariate HR_{sd} (95% CI)		Multivariate aHR_{sd} (95% CI)	p value
Country		<0.001		<0.001
Uganda	1		1	
Botswana	1.44 (1.36 – 1.52)		2.00 (1.91 – 2.10)	
Sex		<0.001		
Female	1			
Male	1.19 (1.14 – 1.24)			
Age at Enrolment (per 1-yr increase)	0.20 (0.18 – 0.22)	0.011	0.18 (0.16 – 0.20)	0.001

Year Birth		<0.001		
Before 2004	1			
After 2004	3.45 (3.22 – 3.70)			
Year of enrolment				<0.001
Before 2006	1		1	
2006–2009	0.96 (0.91 – 1.01)	0.140	1.14 (1.07 – 1.22)	
After 2010	1.19 (1.12 – 1.26)	<0.001	1.13 (1.02 – 1.24)	
HIV RNA load, (log10 copies/uL) §	1.89 (1.52 – 2.30)			
Weight-for-age Z–score (per Z increase)	0.68 (0.64 – 0.72)	0.031	0.77 (0.68 – 0.87)	<0.001
–1.75	1		1	
–3.75	1.24 (1.16 – 1.34)		1.42 (1.32 – 1.53)	
Height-for-age z–score (per Z increase)	0.86 (0.83 – 0.88)	<0.001	1.00 (0.92 – 1.08)	0.910
–1.75	1		1	
–3.75	1.22 (1.17 – 1.27)		1.03 (0.96–1.09)	
BMI-for-age z–score (per Z increase)†	0.79 (0.76 – 0.82)	<0.001		
–0.85	1			
–2.85	1.32 (1.28 – 1.38)			

N=10078, § n = 847, † BMI-for-age was not included in any model with weight-for-age or height-for-age, BMI - body mass index,

4.1.5 Discussion

This study presents results from one of the largest African pediatric HIV cohorts, covering approximately 8% and 12% of children living with HIV in Botswana and Uganda, respectively.[21] We aimed to estimate the frequency of LTNP and examine the factors associated with progression in children using available demographic and baseline clinical data. We applied a definition of LTNP to include only children without an AIDS-defining illness, no ART, and/or no more than one CD4 count below 500 cells/ml (or CD4% ≤25% for children) for ten years since infection meeting our criteria.[13] Most demographic and baseline clinical characteristics differed among children from Botswana and Uganda, consistent with cultural and economic differences between countries/populations.[22] However, once enrolled in care, national HIV care and treatment guidelines or programs were highly congruent, as implemented in the BIPAI network, to which both clinical centers subscribe. We, therefore, chose to consider the two populations together for outcome analyses.

Because there have been few studies on the prevalence of LTNP among children in Africa, our multiethnic study in a geographically and genetically diverse HIV background gives the first

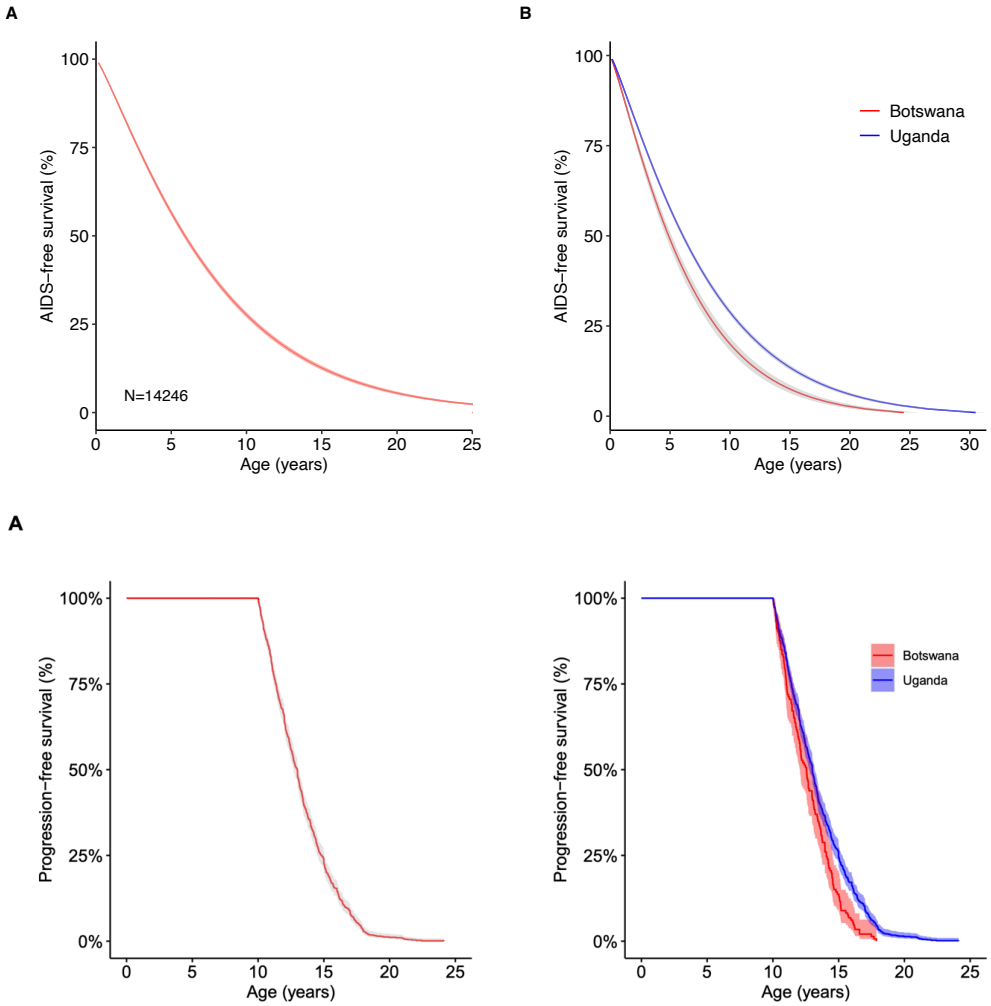


Figure 4-2: Progression-free survival curves.

Weibull parametric estimate of Progression-free survival among the total number of children (A) and children in Botswana and Uganda (B). Kaplan-Meier estimate of progression-free survival after attaining LTNP status (C) and (D) children in Botswana vs. Uganda. (The shaded area represents the 95% confidence interval).

and largest account of an estimate of LTNP prevalence in Africa. Furthermore, this study takes advantage of a large number of children infected with HIV in a location where the disease burden remains very high. Previous research, especially in Western populations, reported

conflicting rates of LTNP ranging from as low as 1.5 to 27% in substantially smaller populations and utilized varying age cut-offs (5yrs, 8yrs, or 10yrs) to define LTNP.[23,24] We observed an incidence of LTNP that was in-between that reported in these studies; however,

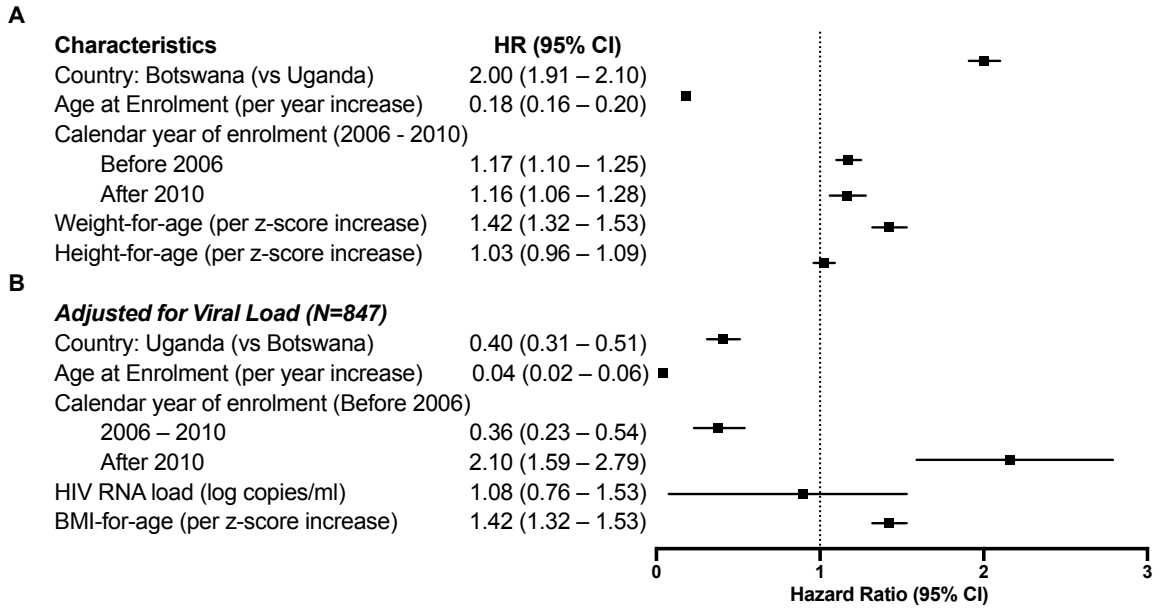


Figure 4-3: Adjusted hazard ratios of progression

(A) without HIV viral load in the model n = 10078 and (B) with HIV viral load in the model n = 847. Abbreviations: HR=adjusted sub-distribution hazard ratio, BMI=body mass index

our study is 30 times larger than the largest known study in the USA, which included 422 children and utilized a similar clinical definition.[11–13,23,24] Notably, a third of our cohort were LTFU or transferred out of the COEs and did not have recorded outcomes. Consequently, the actual proportion of LTNPs may differ from the estimates here; in Botswana, where the LTFU was significantly lower (5% vs 32% in Uganda), the proportion of LTNPs was higher (11% vs 7% in Uganda), although most of the children in Botswana were also enrolled at an older age. Given the asymptomatic nature of their presentation, it is plausible that LTNPs were present in these communities prior to the creation of pediatric HIV care clinics (survivor bias), although presenting late to clinic care creates some uncertainty around their LTNP status. Children in the COE were derived from Botswana and Uganda's urban catchment areas in the capital city. Rural-to-urban migration meant that we had all districts represented in our study. To our

knowledge, this is the first study to examine the loss of LTNP status among children. Compared with studies from adult populations, our population's median loss of LTNP status is higher than in Western adult populations.[25] In a recent study of mixed adult populations, the median time to loss of LTNP status was 12.5 years which is lower but comparable to 13 years in our study.[4] Our study further confirms that children, like adults, rapidly lose their LTNP status, with 50% progressing to AIDS within three years after ten years of stable infection.[4]

Children are expected to progress faster to AIDS than adults (e.g. <35% are progression-free by two years of infection) in part due to their developing immune system;[12] however, we unexpectedly found that half the children in our cohort remained progression-free for ≥ 6 years, and more than one-third for ≥ 10 years. A meta-analysis of 57 studies in mostly adult cohorts globally found the estimated 10-year progression-free survival to be 26%, much lower than in our cohort, especially among children from Uganda, which may be attributed to survival bias in pediatric cohorts.[26] Additionally, this study used a composite definition of progression-free survival as initiating ART regardless of the reason. We recognize that during the period under study, international and national guidelines on HIV treatment changed three times. The most significant changes affected young children, which could affect the progression-free survival estimate. However, it is anticipated that such changes should lower survival in this age group. The high progression-free survival may also be attributed to the impact of unmeasured confounding of perinatal administration of zidovudine and nevirapine monotherapy or combination therapy in PMTCT programs, which has been shown to delay HIV progression.[6,27–29] Speculatively, the disparity in progression-free survival between the two countries could be accounted for by the differences in the HIV epidemiological curves in Botswana and Uganda. Early implementation of strategies such as cotrimoxazole prophylaxis against opportunistic pathogens significantly reduces the risk of progression, especially early in infancy (in our study, we found that 60% of children in Uganda were enrolled below five years and received the prophylactic intervention).[30,31] Therefore, the estimates provided here may overestimate progression-free survival, which would be expected in an observational cohort compared to prospective cohorts.There exists a contributory relationship between HIV progression and undernutrition. HIV increases the risk of undernutrition due to higher metabolic demands and dysregulation, and conversely, undernutrition can influence immune dysfunction, which accelerates progression.[32]

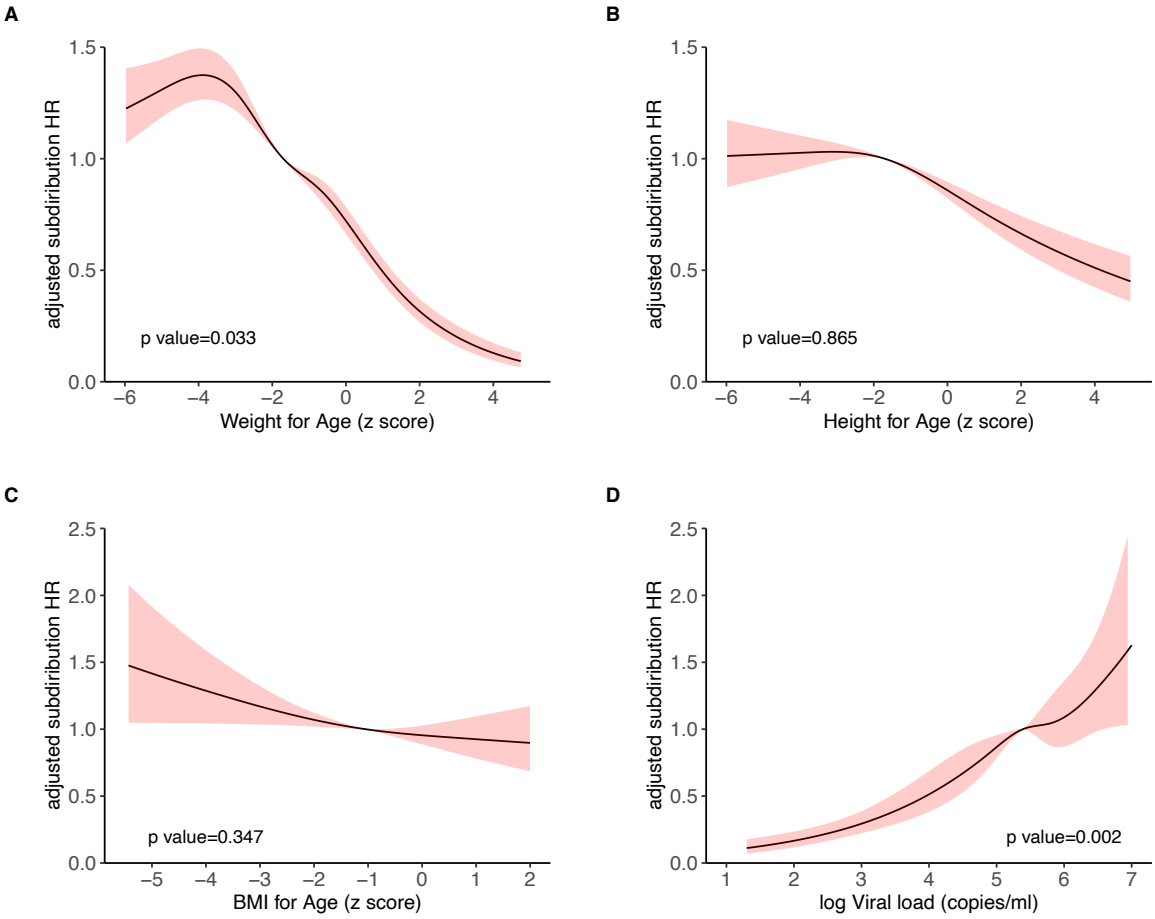


Figure 4-4: Association of weight-for-age (A), height-for-age (B), BMI-for-age (C) and HIV viral load (D) with progression.

The solid black line represents the fitted line of the association between the predictors and the estimated hazard ratio of progression. The shaded area represents the 95% confidence interval. For the upper panel the model was adjusted for country of origin, age at enrolment, year of enrolment, weight-for-age and height-for-age. For the lower panel the model was conducted in a subset of children with known viral load measurements at enrolment and was adjusted for country of origin, age at enrolment, year of enrolment, BMI-for-age and HIV RNA viral load. Cubic splines with five and three knots were used to adjust age at enrolment as well as weight-for-age and height-for-age to better fit the data. HR - hazard ratio

Predictably, the proportion of undernutrition at baseline in our cohort was higher than reported for non-HIV-infected children in African populations but similar to estimates among HIV-positive cohorts from Africa.[33] In our cohort, it is challenging to determine the direction of the

relationship; however, there is a possibility that these children experience bidirectional effects of HIV on nutrition and vice versa, leading to an excess of undernutrition.[32] Studies in pregnant women infected with HIV provide evidence of intrauterine growth retardation that could explain the persistent undernutrition even in older children infected with HIV, however, other factors such as food insecurity and frequent childhood illnesses (respiratory tract infections and diarrhea) may have a role play.[34,35] The distribution of HIV subtypes differs between Botswana and Uganda,[36] and progression has been associated with these differences, albeit with some inconsistency in direction and effect.[31,37,38]

We observed a higher proportion of rapid progressors in a predominantly HIV-1C epidemic in Botswana compared to a population with equal distribution of HIV 1A or D and their recombinant AD in Uganda.[36] In our cohort, this difference could be explained by the lower median anthropometric *z*-scores and CD4 counts at baseline among children in Botswana compared to those in Uganda (Table 4-1).[37] Furthermore, Botswana and Uganda are at different stages in the HIV epidemic transition (epidemic curve)[39]; Uganda is estimated to have reached the peak of HIV-related deaths in 1998 compared to Botswana, which reached the highest number of HIV-related deaths in 2002.[39,40] We were also able to detect a "cohort effect" with a high likelihood of progression among children enrolled after 2010, which could be a consequence of the increased number of younger children more recently enrolled at the COEs or changes in ART treatment guidelines. Incidentally, a few years earlier, the COEs started implementing programs linked to PMTCT services where HIV-exposed infants were increasingly recruited after birth to assess possible vertical infection. Furthermore, the cohort effect arises from individuals in the same recruitment cohort that will encountering similar events at comparable ages. Birth and enrolment cohorts are exposed to varying events at different life stages and experience distinct levels of economic, behavioral, policy (ART), and environmental risks of progression. Additionally, the observed cohort effect is also predicted to affect and skew LTNP estimates. The children in this study were born from 1986, the pre-ART era to 2014, when ART was widely available; however, we did not find a "calendar effect", which has not been invariably associated with progression.[4]

There are some limitations to this study. The results are potentially biased towards the Ugandan population, contributing to the vast majority (89%) of the children analyzed. However, this is

not unexpected given the population sizes in both countries; 47.12 million in Uganda compared to 2.35 million in Botswana,[20] although in a relatively random sub-study of the 847 children with HIV RNA load measurements, we found similar trends (Supp. **Table 1**). Previous studies have reported a significant mortality rate, with over 50% of deaths within the first two years if no ART is initiated.[6] As our study focuses on LTNP it is important to acknowledge that the high early mortality rate might introduce survival bias, as those who die early due to the infection would not be included in our analysis. This survival bias could potentially lead to an imprecise estimation of the prevalence of LTNP, as our study population may be enriched with individuals who have survived beyond the early critical period. Additionally, the shifting WHO ART recommendations from 2003, 2006, and 2010 can significantly impact LTNP calculations and estimation of progression-free survival. For example, in 2008, the WHO advised that all infants infected with HIV under 24 months be started on ARVs, regardless of clinical or immunological status.[41] However, the BIPAI clinics did not adopt this guideline until 2010 (consequently all these children are counted as progressors), which may have had a minor impact on our LTNP status because no children would have survived ten years by the censoring date of our study. Nonetheless, with validation from other pediatric cohorts in African populations, this study finds that our findings provide a robust estimate of LTNP and confirm the evidence for integrating nutritional programs in pediatric HIV care to reduce early mortality.[42] We intended to examine a broader range of both demographic and clinical factors previously associated with progression in children;[12,23,43] (e.g. maternal in-utero factors, maternal viral load, birth weight, prematurity, antenatal care, viral subtype, neuro psychomotor development); however, most of the desired data was either not routinely collected, not available within six months of enrolment into care, or was otherwise unavailable. This is likely the result of the piecemeal HIV care and treatment guidelines and the fractured healthcare delivery systems in Africa. Therefore, our analysis was limited to a candidate set of factors within the broader aims of CAfGEN. Nevertheless, we demonstrate a significantly increased risk of progression among children from Botswana, undernutrition (underweight and stunted) and those who were enrolled before 2006 or after 2010.

4.1.6 Conclusion

Progression-free survival in African pediatric HIV populations is more prolonged than previously recognized, and undernutrition are independently associated with progression in children. Finally, extended phenotyping of this pool of children identified as LTNP can support studies based on extremes of disease progression to understand mechanisms of natural resistance to HIV.

Funding source

The project is supported by Award No. U54AI110398-03 by the National Institute of Allergy and Infectious Disease as part of the NIH Common Fund H3Africa Initiative. This work was supported through the DELTAS Africa Initiative, an independent funding scheme of the African Academy of Sciences (AAS), Alliance for Accelerating Excellence in Science in Africa (AESA). The AESA is supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant no.107743) and the U.K. Government. All funders had no role in study design; collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Author contributions

SK, MW, SK, IK, JKL, BN, EDP, MTJ, CWB, SWM, GM, NAH, DK MSM, and ARK contributed to the conception of the study. SK, JKL, MSM, ARK, and NAH contributed to the study's design. SK performed the primary statistical analysis and drafted the manuscript. GPK, ANM, LK, and JF contributed to the data collection. SK, SM, EK, FAK, GR, BCM, LW, KM, EMW, and MTJ contributed to the manuscript's critical review. SK, GPK, NAH, MSM and ARK had access to and verified the underlying data.

Ethical Approval statement

This study was approved by the School of Biomedical Sciences Institutional Review Board (IRB), Uganda National Council for Science and Technology, University of Botswana IRB,

Botswana Health Research and Development Committee, and the Baylor College of Medicine IRB.

Conflict of Interest

The author reports no financial or non-financial conflicts of interest in this work.

Acknowledgements

We would like to acknowledge the clients and families of the Centres of Excellence, respective Ministries of Health in Botswana and Uganda, as well as the following individuals for their contributions on behalf of the *CAfGEN* Consortium: Bhekumusa Lukhele, David Kateete, Edward D. Pettitt, and Marape Marape, who were participating investigators. We acknowledge Nasinghe Emmanuel, Thembela Mavuso, Buhle Dlamini, Abhilash Sathyamoorthi, Bathusi Mathuba, Yves Mafulu, Gerald Mboowa, Harriet Nakayiza, Bheki Ntshangase, Keboletse Mokete, Kennedy Sichone, Keofentse Mathuba, LeToya Balebetse, Muambi Muyaya, Nancy Zwane, Nicholas Muriithi, Sibongile Mumanga, Thobile Jele, Olekantse Molatlhegi and Thato Regonamanye. We also acknowledge Alison Elliot's contribution, Damalie Nakanjako, Victoria Bukirwa, Joshua Mandre, and Moses Kiiza from the MUIIplus UVRI/MRC/LSHTM AIDS Research Unit.

4.1.7 References

1. United Nations Programme on HIV/AIDS (UNAIDS). UNAIDS Data 2021.; 2021.

2. Rimawi BH, Rimawi RH, Micallef M, Pinckney L, Fowler SL, Dixon TC. Pediatric HIV Long-Term Nonprogressors. Case Rep Infect Dis. 2014;2014:1-3. doi:10.1155/2014/752312

3. Berendam SJ, Nelson AN, Goswami R, et al. Pediatric HIV: the Potential of Immune Therapeutics to Achieve Viral Remission and Functional Cure. Curr HIV/AIDS Rep. 2020;17(3):237-248. doi:10.1007/s11904-020-00495-1

4. Helm JJ Van Der, Geskus R, Lodi S, Meyer L, Schuitemaker H, Gunsenheimer-bartmeyer B. Characterisation of long-term non-progression of HIV-1 infection after seroconversion : a cohort study. Lancet HIV. 2014;1:41-48.

5. Pearson DA, McGrath NM, Nozyce M, et al. Predicting HIV disease progression in children using measures of neuropsychological and neurological functioning. Pediatric AIDS clinical trials 152 study team. Pediatrics. 2000;106(6):E76. doi:10.1542/peds.106.6.e76

6. Dabis F, Elenga N, Meda N, et al. 18-Month mortality and perinatal exposure to zidovudine in West Africa. AIDS. 2001;15(6):771-779. doi:10.1097/00002030-200104130-00013

7. Goulet JL, Fultz SL, Rimland D, et al. Do Patterns of Comorbidity Vary by HIV Status, Age, and HIV Severity? Clin Infect Dis. 2007;45(12):1593–1601. doi:10.1086/523577

8. Leitman EM, Thobakgale CF, Adland E, et al. Role of HIV-specific CD8+ T cells in pediatric HIV cure strategies after widespread early viral escape. J Exp Med. 2017;214(11):3239–3261. doi:10.1084/jem.20162123

9. Chakraborty R, Morel AS, Sutton JK, et al. Correlates of Delayed Disease Progression in HIV-1-Infected Kenyan Children. J Immunol. 2005;174(12):8191-8199. doi:10.4049/jimmunol.174.12.8191

10. Card CM, McLaren PJ, Wachihi C, Kimani J, Plummer FA, Fowke KR. 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. 2009;199(9):1318-1322. doi:10.1086/597801

11. Nielsen K, Mcsherry G, Petru A, et al. A Descriptive Survey of Pediatric Human Immunodeficiency Virus–infected Long-term Survivors. Pediatrics. 1997;99(4).

12. Warszawski J, Lechenadec J, Faye A, et al. Long-Term Nonprogression of HIV Infection in Children: Evaluation of the ANRS Prospective French Pediatric Cohort. Clin Infect Dis. 2007;45(6):785-794. doi:10.1086/521165

13. Gurdasani D, Iles L, Dillon DG, et al. A systematic review of definitions of extreme phenotypes of HIV control and progression. AIDS. 2014;28(2):149-162. doi:10.1097/QAD.0000000000000049

14. Martin N, Koup R, Kaslow R, Coffin J, Ammann A, Jansen JK. Workshop on perinatally acquired human immunodeficiency virus infection in long-term surviving children: A collaborative study of factors contributing to slow disease progression. AIDS Res Hum Retroviruses. 1996;12(16):1565-1570. doi:10.1089/aid.1996.12.1565

15. Kyobe S, Mwesigwa S, Kisitu GP, et al. Exome Sequencing Reveals a Putative Role for HLA-C*03:02 in Control of HIV-1 in African Pediatric Populations. Front Genet. 2021;12(August):1-14. doi:10.3389/fgene.2021.720213

16. Biraro S, Morison LA, Nakiyingi-Miiro J, Whitworth JAG, Grosskurth H. The Role of Vertical Transmission and Health Care-Related Factors in HIV Infection of Children. JAIDS J Acquir Immune Defic Syndr. 2007;44(2):222-228. doi:10.1097/QAI.0b013e31802e2954

17. Motswere-Chirwa C, Voetsch A, Lu L, et al. Follow-up of infants diagnosed with HIV - Early infant diagnosis program, Francistown, Botswana, 2005-2012. Morb Mortal Wkly Rep. 2014;63(7):158-160.

18. Kiyaga C, Sendagire H, Joseph E, et al. Uganda's New National Laboratory Sample Transport System: A Successful Model for Improving Access to Diagnostic Services for Early Infant HIV Diagnosis and Other Programs. Pai M, ed. PLoS One. 2013;8(11):e78609. doi:10.1371/journal.pone.0078609

19. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards. Rev Chil pediatría. 2009;80(4). doi:10.4067/S0370-41062009000400012

20. United Nations. World Population Prospects 2019 Volume I: Comprehensive Tables. Vol I.; 2019. https://population.un.org/wpp/Publications/Files/WPP2019_Volume-I_Comprehensive-Tables.pdf

21. UNICEF. UNICEF Data: Monitoring the situation of children and women. HIV estimates for children dashboard. Published 2022. Accessed August 25, 2022. https://data.unicef.org/resources/hiv-estimates-for-children-dashboard/

22. Boender TS, Sigaloff KCE, Kayiwa J, et al. Barriers to Initiation of Pediatric HIV Treatment in Uganda: A Mixed-Method Study. AIDS Res Treat. 2012;2012:1-10. doi:10.1155/2012/817506

23. Paul ME, Mao C, Charurat M, et al. Predictors of immunologic long-term nonprogression in HIV-infected children: Implications for initiating therapy. J Allergy Clin Immunol. 2005;115(4):848-855. doi:10.1016/j.jaci.2004.11.054

24. Ofori-Mante JA, Kaul A, Rigaud M, et al. Natural history of HIV infected pediatric long-term or slow progressor population after the first decade of life. Pediatr Infect Dis J. 2007;26(3):217-220. doi:10.1097/01.inf.0000254413.11246.e1

25. Buchbinder S, Vittinghoff E. HIV-infected long-term nonprogressors: Epidemiology, mechanisms of delayed progression, and clinical and research implications. Microbes Infect. 1999;1(13):1113-1120. doi:10.1016/S1286-4579(99)00204-X

26. Poorolajal J, Hooshmand E, Mahjub H, Esmailnasab N, Jenabi E. Survival rate of AIDS disease and mortality in HIV-infected patients: a meta-analysis. Public Health. 2016;139:3-12. doi:10.1016/j.puhe.2016.05.004

27. Aboulker JP, Babiker A, Darbyshire JH, et al. Five year follow up of vertically HIV infected children in a randomised double blind controlled trial of immediate versus deferred zidovudine: The PENTA 1 trial. Arch Dis Child. 2001;84(3):230-236. doi:10.1136/adc.84.3.230

28. Volberding PA. The Duration of Zidovudine Benefit in Persons With Asymptomatic HIV Infection. JAMA. 1994;272(6):437. doi:10.1001/jama.1994.03520060037029

29. Faye A, Le Chenadec J, Dollfus C, et al. Early versus deferred antiretroviral multidrug therapy in infants infected with HIV type 1. Clin Infect Dis. 2004;39(11):1692-1698. doi:10.1086/425739

30. Grimwade K, Swingler GH. Cotrimoxazole prophylaxis for opportunistic infections in children with HIV infection. Cochrane Database Syst Rev. 2006;Issue 1(Art. No.: CD003508). doi:10.1002/14651858.cd003508.pub2

31. Venner CM, Nankya I, Kyeyune F, et al. Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa. EBioMedicine. 2016;13:305-314. doi:10.1016/j.ebiom.2016.10.014

32. Mody A, Bartz S, Hornik CP, et al. Effects of HIV infection on the metabolic and hormonal status of children with severe acute malnutrition. PLoS One. 2014;9(7):1-11. doi:10.1371/journal.pone.0102233

33. Kimani-Murage EW, Norris SA, Pettifor JM, et al. Nutritional status and HIV in rural South African children. BMC Pediatr. 2011;11(23):1-13. doi:10.1186/1471-2431-11-23

34. López M, Palacio M, Goncé A, et al. Risk of intrauterine growth restriction among HIV-infected pregnant women: a cohort study. Eur J Clin Microbiol Infect Dis. 2015;34(2):223-230. doi:10.1007/s10096-014-2224-6

35. Mendoza JA, Matshaba M, Makhanda J, Liu Y, Boitshwarelo M, Anabwani GM. Food Insecurity and CD4% Among HIV+ Children in Gaborone, Botswana. JAIDS J Acquir Immune Defic Syndr. 2014;66(4):414-418. doi:10.1097/QAI.0000000000000190

36. Bbosa N, Kaleebu P, Ssemwanga D. HIV subtype diversity worldwide. Curr Opin HIV AIDS. 2019;14(3):153-160. doi:10.1097/COH.0000000000000534

37. Amornkul PN, Karita E, Kamali A, et al. Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. AIDS. 2013;27(17):2775-2786. doi:10.1097/QAD.0000000000000012

38. Mlisana K, Werner L, Garrett NJ, et al. Rapid disease progression in HIV-1 subtype C-infected South African women. Clin Infect Dis. 2014;59(9):1322–1331. doi:10.1093/cid/ciu573

39. Nixon SA, Hanass-Hancock J, Whiteside A, Barnett T. The increasing chronicity of HIV in sub-Saharan Africa: Re-thinking “HIV as a long-wave event” in the era of widespread access to ART. Global Health. 2011;7(41):1-5. doi:10.1186/1744-8603-7-41

40. UNAIDS. UNAIDS 2019 Data. Available at https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf Accessed April 7, 2020. doi:978-92-9173-945-5

41. Wieczorek M, Abualrous ET, Sticht J, et al. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. Front Immunol. 2017;8(Imci). doi:10.3389/fimmu.2017.00292

42. Baum MK, Campa A, Lai S, et al. Effect of micronutrient supplementation on disease progression in asymptomatic, antiretroviral-naive, HIV-infected adults in Botswana: A randomized clinical trial. J Am Med Assoc. 2013;310(20):2154-2163. doi:10.1001/jama.2013.280923

43. Pearson DA, McGrath NM, Nozyce M, et al. Predicting HIV disease progression in children using measures of neuropsychological and neurological functioning. Pediatrics. 2000;106(6).

4.1.8 Published Article

International Journal of Infectious Diseases

Available online 28 November 2023

In Press, Journal Pre-proof What's this?

Long-term non-progression and risk factors for disease progression among children living with HIV in Botswana and Uganda: a retrospective cohort study

Samuel Kyobe [1], Grace Kisitu [2], Savannah Mwesigwa [1], John Farirai [3], Eric Katagirya [4], Gaone Retshabile [5], Lesedi Williams [5], Angela Mirembe [2], Lesego Ketumile [3], Misaki Wayengera [4], John Mukisa [4], Gaseene Sebetso [5], Thabo Diphoko [5], Marion Amujal [4], Edgar Kigozi [4], Fred Katabazi [4], Ronald Oceng [2], Busisiwe Mlotshwa [5], Koketso Morapedi [5], Betty Nsangi [2]...Dithan Kiragga [2 11]

Show more

Outline | Share Cite

https://doi.org/10.1016/j.ijid.2023.11.030

Get rights and content

Highlights

- The proportion of LTNPs has not been well studied especially in African populations.
- Around 7% of children infected with HIV were characterized as LTNPs.
- The rate of HIV progression varies among childhood cohorts in Africa.
- Undernutrition is an independent predictor of HIV progression at enrolment.

The full article is available as open access online at: https://www.sciencedirect.com/science/article/pii/S120197122 3007919

4.2 Exome Sequencing Reveals a Putative Role for HLA-C*03:02 in Control of HIV-1 in African Pediatric Populations

Samuel Kyobe[1], Savannah Mwesigwa[1,4], Grace P. Kisitu[2], John Farirai[3], Eric Katagirya[4], Angella N Mirembe[2], Lesego Ketumile[3], Misaki Wayengera[4], Fred Ashaba Katabazi[4], Edgar Kigozi[4], Edward M Wampande[4], Gaone Retshabile[5], Busisiwe C. Mlotshwa[5], Lesedi Williams[5], Koketso Morapedi[5], Ishmael Kasvosve[5] Jacqueline Kyosiimire-Lugemwa[6], Betty Nsangi[2], Masego Tsimako-Johnstone[7], Chester W. Brown[8], Moses Joloba[4], Gabriel Anabwani[3], Lukhele Bhekumusa[9], Sununguko W. Mpoloka[5], Graeme Mardon,[10,11] Mogomotsi Matshaba,[3,12] Adeodata Kekitiinwa,[2,12] and Neil A. Hanchard[10,*] on behalf of the Collaborative African Genomics Network (CAfGEN) of the H3Africa Consortium

[1]Department of Medical Microbiology, College of Health Sciences, Makerere University, Kampala, Uganda, [2]Baylor College of Medicine Children's Foundation, Kampala, Uganda, [3]Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana, [4]Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, Kampala, Uganda, [5]School of Allied Health Professions, Faculty of Health Sciences, University of Botswana, Gaborone, Botswana, [6]The Medical Research Council (MRC)/UVRI & LSHTM Uganda Research Unit, [7]Australian Medical Research, Hurstville, New South Wales 2220, Australia, [8]University of Tennessee Health Science Center, Memphis, TN 38105, USA, [9]Eswatini - Baylor College of Medicine Children's Foundation, Mbabane, Eswatini, [10]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA, [11]Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX 77030, USA, [12]Pediatric Retrovirology, Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA.

***Corresponding author:**

Current Address: National Human Genome Research Institute, National Institutes of Health, USA; Neil.hanchard@nih.gov

Keywords:

MHC, HIV disease progression, HLA, Africa, Genomes

4.2.1 Abstract

Human leucocyte antigen (HLA) class I molecules present endogenously processed antigens to T cells and have been linked to differences in HIV-1 disease progression. HLA allelotypes show considerable geographical and inter-individual variation, as does the rate of progression of HIV-1 disease, with long-term non-progression (LTNP) of disease having most evidence of an underlying genetic contribution. However, most genetic analyses of LTNP have occurred in adults of European ancestry, limiting the potential transferability of observed associations to diverse populations who carry the burden of disease. This is particularly true of HIV-1 infected children. Here, using exome sequencing (ES) to infer HLA allelotypes, we determine associations with HIV-1 LTNP in two diverse African pediatric populations. We performed a case-control association study of 394 LTNPs and 420 rapid progressors retrospectively identified from electronic medical records of pediatric HIV-1 populations in Uganda and Botswana. We utilized high-depth ES to perform high-resolution HLA allelotyping and assessed evidence of association between HLA class I alleles and LTNP. Sixteen HLA alleles and haplotypes had significantly different frequencies between Uganda and Botswana, with allelic differences being more prominent in HLA-A compared to HLA-B and C allelotypes. Three HLA allelotypes showed association with LTNP, including a novel association in HLA-C (HLA-B*57:03, aOR 3.21, Pc=0.0259; B*58:01, aOR 1.89, Pc=0.033; C*03:02, aOR 4.74, Pc=0.033). Together, these alleles convey an estimated population attributable risk (PAR) of non-progression of 16.5%. We also observed novel haplotype associations with HLA-B*57:03-C*07:01(aOR 5.40, Pc=0.025) and HLA-B*58:01-C*03:02 (aOR 4.88, Pc=0.011) with a PAR of 9.8%, as well as a previously unreported independent additive effect and heterozygote advantage of HLA-C*03:02 with B*58:01(aOR 4.15 , Pc = 0.005) that appears to limit disease progression, despite weak LD (r2=0.18) between these alleles. These associations remained irrespective of gender or country. In one of the largest studies of HIV in Africa, we find evidence of a protective effect of canonical HLA-B alleles and a novel HLA-C association that appears to augment existing HIV-1 control alleles in pediatric populations. Our findings outline the value of using multi-ethnic populations in genetic studies and offer a novel HIV-1 association of relevance to ongoing vaccine studies.

4.2.2 Background

HIV-1 remains a significant public health concern in Africa, with an estimated 36 million people living with HIV-1 in the region (UNAIDS, 2019). Despite the introduction of effective antiretroviral therapy (ART) and the roll-out of prevention of mother-to-child transmission programs, more than 70% of all new childhood infections occur in Africa, numbering close to 220,000 new HIV-1 infections and accounting for more than 110,000 deaths in children annually on the continent (UNAIDS, 2019). However, 1–5% of children infected with HIV-1 do not progress to AIDS for more than ten years even without ART (Warszawski et al., 2007) – so-called LTNPs. We have documented a cohort of childhood LTNPs from Uganda and Botswana (Kyobe et al., n.d.) having one of the highest documented frequencies of LTNPs (8%) in both countries. In general, LTNPs possess a high degree of HIV-1 infection control with significant viral suppression and high $CD4^+$ T cell count (>500 cells/mL) (Warszawski et al., 2007), and have lower levels of the immune activation markers such as HLA-DR and CD38 despite chronic infection (Hua et al., 2014); this pattern is frequently associated with the expression of broadly neutralizing antibodies (bNAbs) (González et al., 2018).

HLA class I molecules predominantly function to present endogenously processed antigens to cytotoxic $CD8^+$ T lymphocytes (CTL) in a cell-mediated immune response (Frater et al., 2007), and play a significant role in the innate immune response through interactions with natural killer (NK) cell receptors via the killer cell immunoglobulin-like receptors (KIRs) (Boudreau & Hsu, 2018). The latter interaction puts class I molecules at the epicenter of HIV-1 control and makes them a focal point for developing and designing epitope-based vaccines (Kaseke et al., 2021). As a result, genetic variation at HLA class I has been frequently associated with LTNP status. Several studies, mainly conducted among adults of European ancestry, have reported associations between LTNP and HLA-B*57, B*27, B*35, B*58, and some HLA-C alleles (P. J. R. Goulder & Walker, 2012). HLA class I genotypes, however, vary substantially across populations, and the association between specific alleles and HIV-1 disease progression in African populations remains controversial, with previous studies mostly conducted in adult populations, showing country/region-specificity of results utilizing small sample sizes and candidate allele approaches (Matthews et al., 2011; Peterson et al., 2013; Sampathkumar et al., 2014; Serwanga et al., 2009; Shepherd et al., 2015). For example, HLA-B*58:01 has been

associated with slow progression in Uganda and South Africa (Payne et al., 2014; Serwanga et al., 2009), but this observation was not seen in Kenya and Botswana (Payne et al., 2014; Peterson et al., 2013); however, the initial Ugandan study was based on only 14 slow progressors, and the definition of LTNP was less stringent than comparable studies in Europeans (Serwanga et al., 2009). Notably, the HIV-1 epidemic in East Africa is predominantly due to HIV-1 subtypes A and D, while HIV-1 subtype C is predominant in southern Africa (Bbosa et al., 2019).

The now established protective effects of HLA-B*57 in adults are also known to be population-specific - HLA-B*57:01 is associated with control in Europeans while HLA-B*57:03 is mostly strongly associated among Africans (Matthews et al., 2011). Most importantly, none of the canonical HIV-1 protective alleles (HLA-B*57, B*27, B*35, B*58) were replicated in a cohort of 123 HIV-1 clade C infected adolescents in Zimbabwe; instead, HIV-1 LTNP was attributed mainly to the HLA-C locus (HLA-C*08:02 and -C*08:04) (Shepherd et al., 2015). Adland et al. provided evidence of discordance in HLA-HIV associations between children and adults (Adland et al., 2015) in a South African cohort, among whom the putative protective effects of B*57, HLA-B*58:01, and HLA-B*81:01 in 47 HIV-1 clade C infected mothers were non-existent in 84 children (Adland et al., 2015). HLA homozygosity has also been associated with rapid progression to AIDS, while heterozygosity has been associated with protection; however, HLA homozygosity is very low in Africa (Cao et al., 2001; Carrington et al., 1999; Peterson et al., 2013). Taken together, the current body of literature fails to provide a clear picture of the role of HLA alleles in pediatric HIV-1 LTNP, particularly among large African pediatric population groups (Adland et al., 2015; Shepherd et al., 2015; Thobakgale et al., 2009).

We have demonstrated uncaptured genetic variation in Botswana and Uganda (Retshabile et al., 2018) from exome sequencing data, which, alongside other studies of genetic variation across the continent (Choudhury et al., 2020; Sherman et al., 2019a), suggest a potential role for novel genetic variants in disease susceptibility and argue for the use of multi-ethnic African populations in such studies (Sherman et al., 2019b). We hypothesized that looking for genetic association between HLA and pediatric LTNP in multi-ethnic African populations could yield novel HLA variant associations, which might be relevant for epitope-based vaccine designs. To evaluate this, we inferred HLA class I alleles from exome sequencing data derived from a

cohort of about ~800 pediatric HIV-1 positive children from Uganda and Botswana, and determined evidence for an association between HLA class I variants and LTNP status.

4.2.3 Material and Methods

Study populations and design

The Collaborative African Genomics Network (*CAfGEN*) is a collaboration between the Baylor International Pediatric AIDS Initiative (BIPAI) network sites in Uganda, Eswatini and Botswana, Makerere University, University of Botswana, and Baylor College of Medicine. The details of the primary cohort's demographic and clinical characteristics have been described (Kyobe et al., n.d.; Mwesigwa et al., 2021; Retshabile et al., 2018). This study was approved by the School of Biomedical Sciences Institutional Review Board (IRB), Uganda National Council for Science and Technology, University of Botswana IRB, Botswana Health Research and Development Committee, and the Baylor College of Medicine IRB.

We conducted a retrospective case-control study based on two extreme phenotypes of HIV-1 disease progression; LTNP and RP (Peloso et al., 2016). LTNPs were defined as children perinatally infected with HIV-1 and asymptomatic for more than ten years with a CD4 count below 500 cells/ml or CD4 above 25%. The RPs were defined as an AIDS-defining illness, two or more consecutive CD4 below 15%, and ART initiation within three years of perinatal HIV-1 infection. Using electronic medical records (EMR), we identified 1,000 participants who met the criteria and were stratified equally in both groups. All participants provided written informed consent or assent according to age.

Whole-Exome Sequencing (WES) and HLA allelotyping

DNA exome sequencing for this cohort has been described in detail elsewhere (Retshabile et al., 2018). Overall, at least 96% of the bases targeted were covered at >20X depth (Retshabile et al., 2018). 822 participants had successful WES in three batches using the Illumina HiSeq 2000, HiSeq 2500, and NovaSeq. Four-digit HLA typing was inferred from WES data using HLAreporter, whose precision ranges from 96% to 100% depending on the locus and sequencing quality (Y. Huang et al., 2015). Briefly, HLAreporter achieves accurate high-resolution typing of HLA alleles in four steps. First, WES FastQ files containing a mixture of

short reads were mapped on a comprehensive reference panel from the IMGT/HLA database (v3.33) of HLA polymorphism. Secondly, the filtered and mapped reads (≥90bp) were classified into specific HLA genes. Thirdly, the classified short reads were assembled de-novo using a TASR assembler into contigs. The final two steps involve contig size score-based contig-HLA alignment using two candidate databases to identify only perfectly matched alleles (step 4) and assign G groups (step 5). Only contigs with sequencing depth above 10X were used during alignment, this is stringent and improves the accuracy of HLAreporter (Y. Huang et al., 2015).

Validation

We conducted validation of HLAreporter using Lamda Micro-SSP kits (according to manufacturer's guidelines) on 20 randomly determined samples and observed 95% (19/20) concordance with four-digit high-resolution typing.

Statistical analysis

Categorical demographic characteristics were summarized as counts and compared using χ^2 tests, while the continuous variables were summarised as median (interquartile range, IQR) and compared using the Wilcoxon rank–sum test. HLA allele frequencies were determined by expectation-maximization methods using PyHLA and compared using the prop.test in R (Fan & Song, 2017). The gametic phase was reconstructed using the ELB algorithm and pairwise linkage disequilibrium (LD) was measured using the Pearson's squared correlation coefficient r^2 (or Hendrik's D' for global LD between loci) in PyPop or Arlequin (Excoffier & Lischer, 2010) which is less sensitive to allele frequencies. The p-value threshold for significant LD ($p < 3.18 \times 10^{-6}$) was derived from Fisher's exact test and corrected for multiple testing (https://www.hiv.lanl.gov/content/index) (A. Lancaster et al., 2003; A. K. Lancaster et al., 2007). Allelic associations were performed for alleles with frequencies >1% in the cohort using Fisher's exact test. Odds ratios were calculated with Haldane's correction of Woolf's method (F. Yang et al., 2020), assuming an additive effects model. Our preliminary study (Kyobe et al., n.d.), we found some differences between children classified as LTNP by gender and geographical origin. We performed logistic regression and included country (as proxy for the genetic differentiation by PCA and Fst (see Results) as well as differences in HIV-1 subtypes

between Uganda and Botswana) and gender as covariates to account for these differences. To establish the influence of homozygosity and heterozygosity of the significant alleles on disease progression, we performed zygosity analysis in PyPop (Svejgaard & Ryder, 1994). We constructed two-by-four tables and performed Svejgaard tests to establish the degree of inter-dependence between alleles associated with LTNP (Fan & Song, 2017).

For haplotype analyses, we used the Haplo.Stats package (v1.7.9) (Martin & Martin, 2021). Haplotype frequencies were estimated using the haplo.em function. We tested the association between haplotypes expressed at a frequency >1% and LTNP under an additive effect model using the haplo.score function. The haplo.glm and haplo.cc functions were used to calculate odds ratios using generalized linear regression models relative to the most frequent haplotype, controlling non-genetic factors (gender and country). We estimated the population attributable risk (PAR) due to the significant alleles and haplotypes using odds ratios from the logistic regression models (Bruzzi et al., 1985) according to Levin's formula: $PAR = \frac{\mathrm{P(OR-1)}}{\mathrm{P(OR-1)+1}}\ X\ 100$, where p denotes the frequency of alleles or haplotypes among LTNPs and OR, is the odds ratios from the logistic regression models (Levin, 1953). False discovery rate (FDR) was used to adjust P values for the number of alleles tested (Pcorected, Pc <0.05). All statistical analysis was performed using Stata13 and R software v1.2.5033 (Gutierrez, 2010).

4.2.4 Results

A total of 1000 participants, equally drawn from both countries, were recruited to participate in this stage of the study. As summarized in Table 4-3, we found that gender and country were similar at the geographic level, but there were significant differences in the number of LTNPs and RPs in the two countries ($p<0.001$). Consistent with clinical observations, there were more female LTNPs than male. By definition, the median time to progression among LTNPs is significantly different from RPs, and we confirmed this in our cohort (157 months vs. 17 months, $p<0.001$, Wilcoxon rank–sum test). Exome sequencing was successful for 394 LTNPs (207 Uganda, 187 Botswana) and 420 RPs (162 Uganda, 258 Botswana), and this final set of 814 individuals was utilized for downstream association analyses. Eight (~1%; 8/822) samples were excluded from further analysis because no allele was typed at any locus. A total of 68 HLA-A, 88 HLA-B, and 63 HLA-C alleles were detected in our cohort.

Table 4-3: Baseline characteristics among LTNPs and RPs

Variables	LTNPs	RPs	P-value
Country			
Botswana	187 (47.5)	258 (61.4)	<0.001
Uganda	207 (52.5)	162 (38.6)	
Gender			
Female	228 (57.9)	192 (45.7)	<0.001
Male	166 (42.1)	228 (54.3)	
Age (yr)	16.6 (13.7–19.4)	9.6 (6.6–12.1)	<0.001
Time to progression (mo)	157 (138–176.5)	17.0 (10.0–25.0)	<0.001
HLA Class I			
A	55	57	0.963
B	65	63	
C	47	48	

LTNP, long-term non-progressors, RP, rapid progressors, mo, months, yr, years

Population-specific HLA alleles and haplotypes are enriched in HIV-infected children from Uganda and Botswana

We have previously demonstrated low to moderate genetic differentiation between Uganda and Botswana populations using principal component analysis (PCA; using SNPrelate v1.10.2 in R see Supp. Figure 3) and Fst based on Weir and Cockerham's method using genome-wide data (Weir and Cockerham, 1984; Retshabile et al., 2018). The PCA plots (Supp. Figure 3) and Fst of 0.0065 show some differentiation between the two populations at the global level of PC1 and 2, however, there is strong concordance at later PCs (PC1 versus PC3) and the overall Fst differences are small. Therefore, we started by exploring differences in HLA allele distributions (Figure 4-5) and pairwise LD between allele classes in the two countries (Figure 4-6A-C, Supp. Tables 2–4). Six allelotypes - HLA-A*30:01, B*42:01, C*04:01, C*06:02, C*17:01, and C*07:01 - all had allelotype frequencies above 10% in the full dataset (Figure 4-5). These alleles collectively accounted for 10%, 11%, and 52% of the total number of allelotypes at the HLA-A, -B, and -C loci, respectively (Figure 4-5 and Supp. Table 5). Surprisingly, HLA-A*30:01, B*42:01 and C*17:01 in Ugandans and HLA-C*03:0207:01 in Botswana were observed at between two to ten times higher sample frequencies, respectively, than previous studies of healthy Ugandan (Z test for proportions: p=0.026, p=0.002 and p=0.006 respectively) and South African Black (Z test for proportions: p=0.021) populations

(Supp. Table 6) (Cao et al., 2004; Gonzalez-Galarza et al., 2020; Tshabalala et al., 2018). This trend was similar when the data were analyzed by country (Supp. Table 5). The frequencies of 16 alleles were significantly different between Uganda and Botswana (Figure 4-5, Supp. Table 7), with HLA-B*44:03 being almost four times as common in Botswana as in Uganda (9.2% vs. 2.4%, $p=1.63 \times 10^{-4}$). This allele is known to occur at low/uncommon frequencies in healthy populations worldwide, including Uganda and Zimbabwe (p=0.005) (Cao et al., 2001, 2004; Gonzalez-Galarza et al., 2020; Kijak et al., 2009; Tshabalala et al., 2018), and has been independently associated with toxic epidermal necrolysis - a severe form of Stevens-Johnson syndrome (Ueta et al., 2014). HLA-B*44:03-restricted epitopes have also been associated with asymptomatic chronic human cytomegalovirus (CMV) infection (Attaf et al., 2018).

Table 4-4: Global pairwise linkage disequilibrium

HLA loci	Wn	D'	P-value
A:B	0.430	0.606	p<0.0001*
A:C	0.372	0.552	p<0.0001*
B:C	0.631	0.868	p<0.0001*

Wn, Cramer's V Statistic, D' Hedrick's statistic
*p < 0.05 is indicative of overall significant LD

We then examined global pairwise LD between the loci as a way of understanding the underlying HLA haplotype structure in our cohort. As expected, HLA-B and C loci were in strong LD (D' = 0.867, Table 4-4), although quantitatively lower than reported in European ancestry cohorts (D' =0.928) (Cao et al., 2001; Tokić et al., 2020), it is consistent with LD in healthy African populations. In our combined cohort, highest LD was observed between HLA-B*42:01 and C*17:01 ($r^2 = 0.65$, Figure 4-6A). While in Uganda and Botswana, HLA-B*44:15 and C*04:07 and HLA-B*42:01 and C*17:01 showed high LD values $r^2 = 0.78$ and $r^2 = 0.61$ respectively. Overall, both populations exhibit low LD patterns between class I HLA loci (Figure 4-6B and 4-6C).

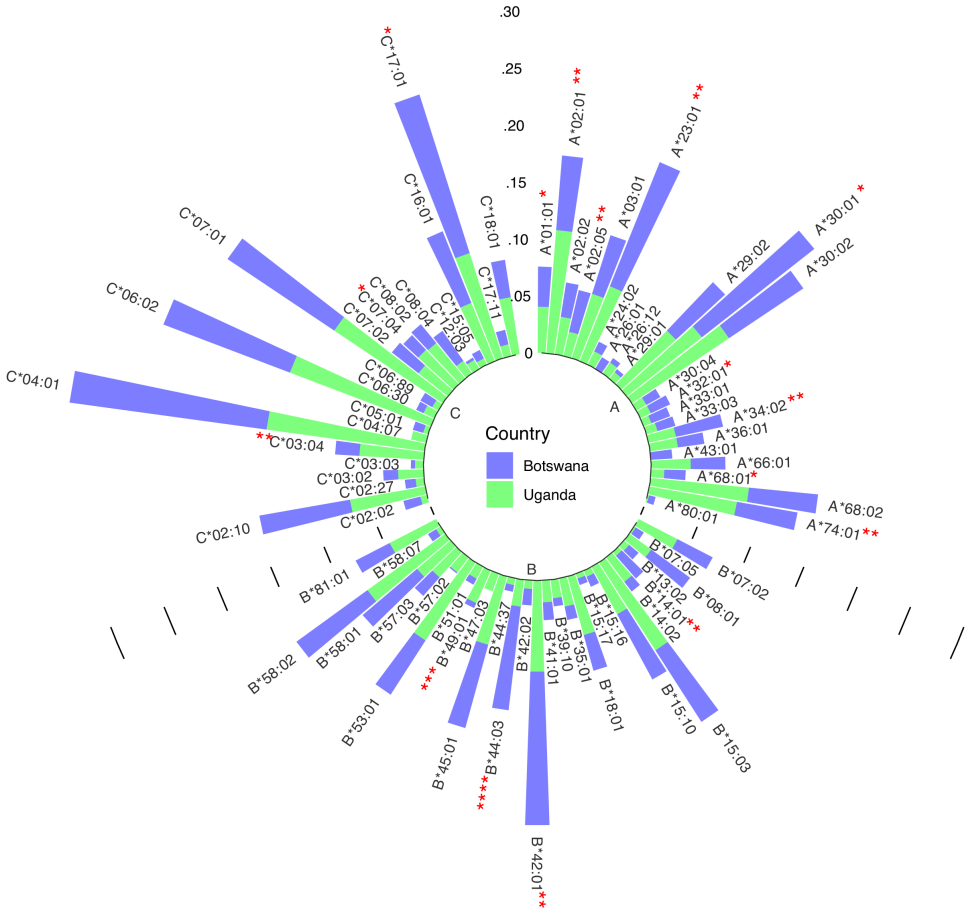


Figure 4-5: Circular bar plot of HLA class I distribution in Uganda and Botswana.

Each bar represents a class I HLA allele. The heights represent the relative frequency of each allele per country and significantly different alleles are shown with asterixis.

The number of bi-allelic haplotypes was not significantly different between Uganda and Botswana (Supp. Table 8), suggesting a similar LD structure in the two countries. When we looked at the resulting pair-wise haplotypes, HLA-A*30:01~B*42:01 (5.2%, $r^2 = 0.19$), HLA-B*42:01~C*17:01 (9.6%, $r^2 = 0.65$) and A*30:01~C*17:01 (5.6%, $r^2 = 0.19$) were the most frequent across our two sample populations (Figure 4-7, Supp. Tables 9 and 10). These four haplotypes have also been observed at high frequencies in healthy South Africans and Pumwani sex workers (Kloverpris et al., 2012; Sampathkumar et al., 2014), with HLA-B*42:01-C*17:01 being associated with faster seroconversion (rapid progression) in the Pumwani cohort (Sampathkumar et al., 2014). The frequency of these alleles and their

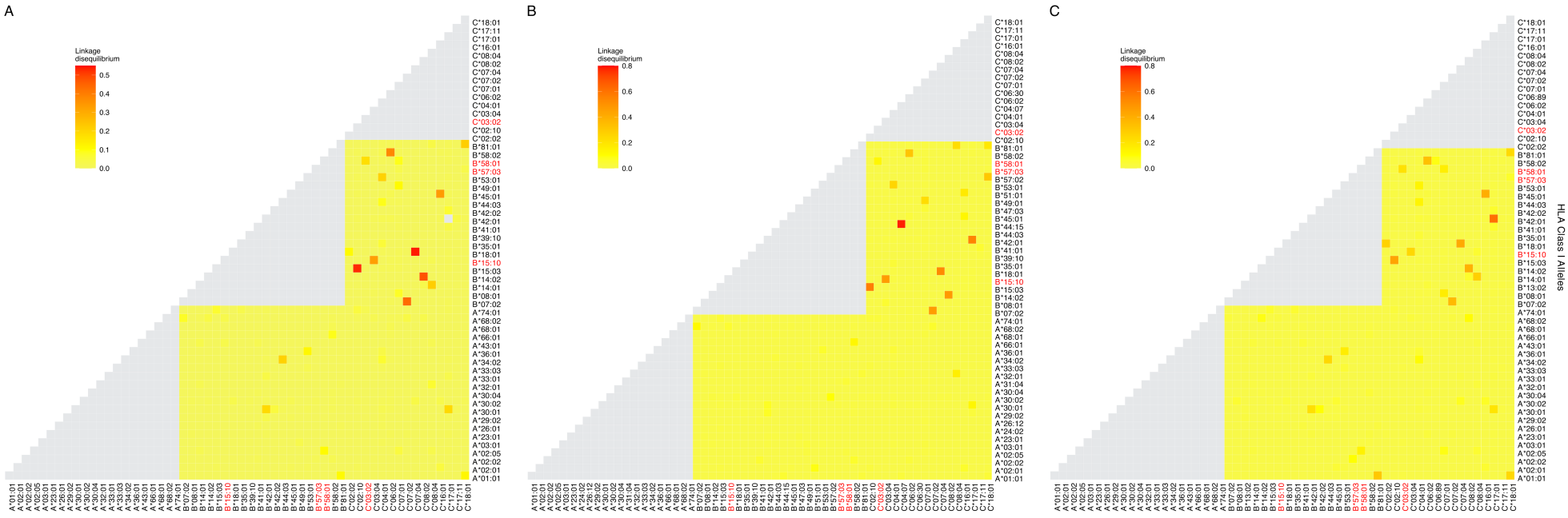


Figure 4-6: Patterns of linkage disequilibrium (LD) between the most frequent class I HLA alleles.

LD in our cohort is expressed as the R^2 values. LD plots were produced with ***(A)*** *56 alleles in the combined cohort,* ***(B)*** *59 alleles in Uganda and* ***(C)*** *56 alleles in Botswana with a frequency ≥1%. The colors indicate increasing strength of LD from yellow to red. Alleles on the same HLA locus (no pairwise LD comparisons) are indicated in grey. Alleles associated with LTNP in the cohort are highlighted in red.*

concomitant haplotypes among healthy individuals of African ancestry (Ugandans, South Africans, Kenyans and Zimbabweans) is much lower than in healthy European and Hispanic groups (Cao et al., 2001; Kijak et al., 2009; Tshabalala et al., 2018), but they are enriched in HIV-1 infected populations generally (Sampathkumar et al., 2014); this may reflect enrichment of population- and disease-specific susceptibility factors (Supp. Tables 7 and 9). This pattern of shared haplotypes could be consistent with a codominant model of protection or susceptibility that might be relevant to migration patterns across the continent and the concomitant emergence of new pathogens.

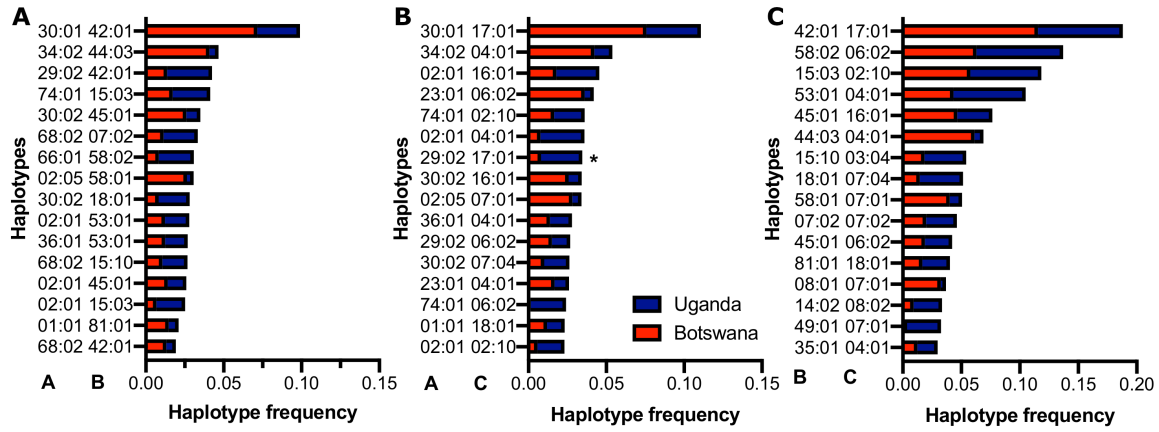


Figure 4-7: Distribution of the commonest HLA class I haplotypes in Uganda and Botswana.

**-indicates alleles that are statistically different among both countries*

Established HLA-B alleles and a novel HLA-C allele show association with LTNP in Ugandan and Batswana children

Although there was no difference in the total number of alleles detected between LTNP (cases) and RP (controls) (p=0.963, Table 4-3), among the 56 alleles with frequencies >1%, we identified three alleles with a significant positive association with LTNP status after adjusting for multiple testing and controlling for gender and country (Pc<0.05, Figure 4-8, Table 4-5) - HLA-B*57:03 (3.9% vs 1.2%, aOR 3.21 [95% CI 1.50–6.86], Pc= 0.026), HLA-B*58:01 (6.6% vs 3.9%, aOR 1.89 [95% CI 1.21–2.96], Pc=0.033) and HLA-C*03:02 (2.9% vs 0.6%, aOR 4.74 [95% CI 1.74–12.85], Pc = 0.033) (Figure 4-8). Additionally, we found HLA-

B*15:10 (3.7% vs 7.7%, OR 0.48 [95% CI 0.30–0.77], Pc=0.0259) to be enriched among RPs (Figure 4-8 and Table 4-5) suggesting a predisposition towards rapid progression. These effects were generally consistent (in both magnitude and direction) across our two populations (Figure 4-9, Supp. Table 11). Most of these associated alleles were in weak LD (Figure 4-6A), suggesting that the associations observed are more likely to be independent. Given the moderate effect sizes observed in our cohort, we also investigated the population attributable risk due to the protective alleles; together, HLA-B*57:03, B*58:01, and C*03:02 accounted for ~16.5% (95% CI 3.5–40) of LTNP among children infected with HIV-1 in Uganda and Botswana.

Next, we examined allelic zygosity, which influences susceptibility or protection from HIV-1 disease progression (Carrington et al., 1999). The overall homozygosity was consistent with previous reports, ranging from 5.3% to 8.6% at any one of the three loci; as expected, this was lower than the 13% reported in populations of European ancestry (Alter et al., 2017). Although there was no evidence for association with LTNP among participants who were homozygous for any of our putative protective alleles (Supp. Table 12), we found that participants who were heterozygous at any of our four significantly associated alleles (HLA-B*57:03, Pc=0.0043; -B*58:01, Pc=0.0258; -C*03:02, Pc=5.7 X 10^{-4}), and B*15:10 Pc=0.0082) were significantly more likely to be LTNPs (Supp. Table 12).

Next, we examined the association between common HLA haplotypes and LTNP status using multivariate regression, controlling again for the known effects of gender and country (Supp. Table 13). We found two protective B~C haplotypes and one susceptible A~C haplotype with evidence of significant association. Globally, HLA haplotypes between B-C loci had the most statistically significant association with LTNP (global p=0.006). Consistent with the single allele analyses, haplotypes HLA-B*57:03-C*07:01 (2.0% vs 0.3%, OR 5.40 [95% CI 1.40–20.79], p=0.025) and HLA-B*58:01-C*03:02 (2.2% vs 0.5%, OR 4.88 [95% CI 1.50–15.86], p=0.011) were over-represented and significantly associated with LTNP relative to the most frequent haplotype HLA-B*42:01-C*17:01 (Table 4-6 and Supp. Table 13). Participants with the HLA-A*29:02-C*17:01 (1.0% vs 2.9%, OR 0.26 [95% CI 0.08–0.79], p=0.003) haplotype were 74% less likely to be LTNPs compared to individuals with the most frequent haplotype (HLA-A*30:01-C*17:01) (Table 4-6, Supp. Table 13). Taken together, the two observed

protective haplotypes account for 9.8% (95% CI 1.0–35.2) of variation in disease progression attributed to LTNP in our cohort.

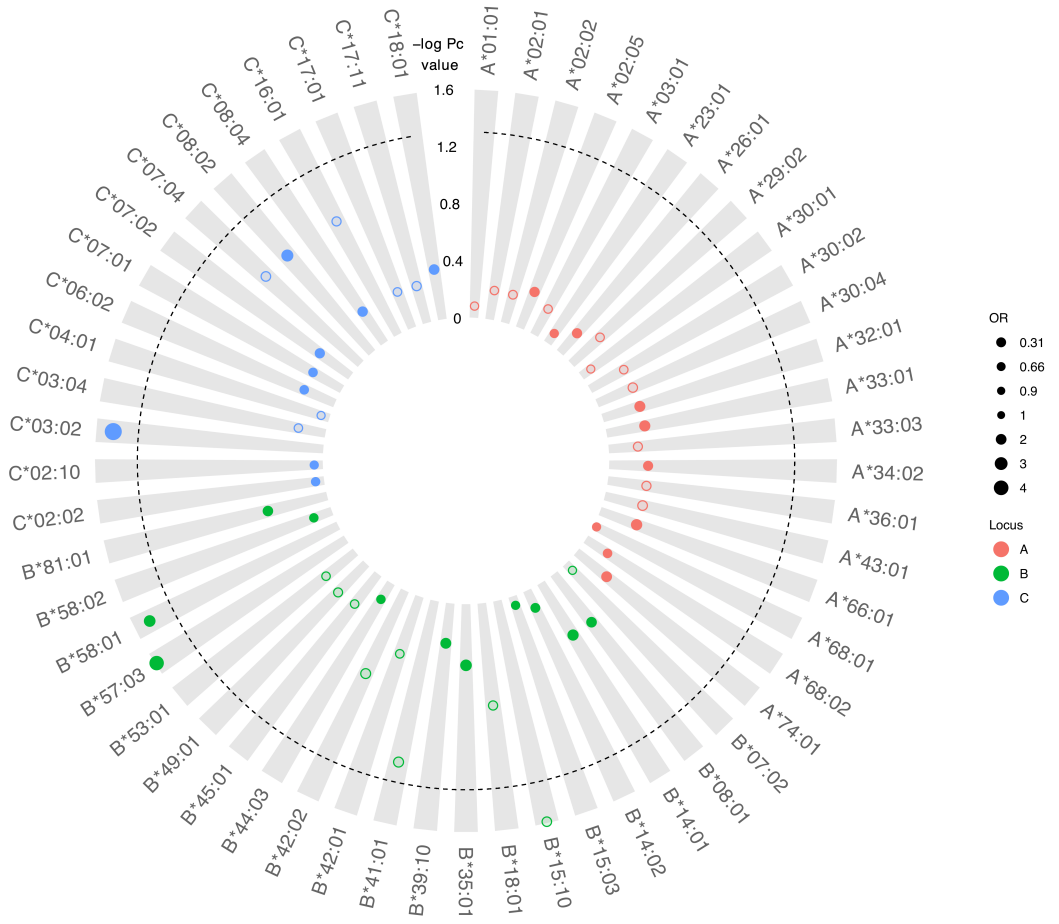


Figure 4-8: Circular bubble plot of the Class I HLA association with LTNP in the CafGEN cohort.

Each bubble represents an HLA class I allele and the distance from the center is a measure of the negative log of the correct P (Pc) value. Protective alleles are represented with solid bubbles while the susceptible alleles are shown in blank bubbles. The size of the bubbles depicts the odds ratio from unity. The broken line represents the cut off of Pc value (<0.05).

Protective effects of HLA-C*03:02 appear to be additive and independent of linkage disequilibrium

Our data suggested that the HLA-C*03:02 allele is independently associated with a longer time to progression; however, the HLA-C*03:02-HLA-B*58:01 haplotype and the HLA-B*58:01

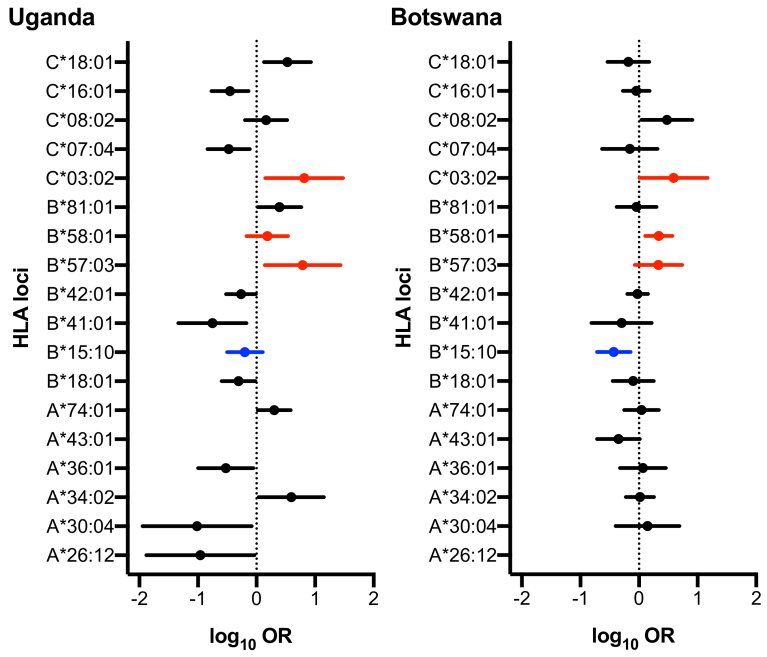


Figure 4-9: Forest plot of HLA alleles associated with LTNP in Uganda and Botswana.

Only alleles found to be statistically associated with HIV progression in either country are shown. Alleles highlighted in red and blue have protective and detrimental effects respectively in both populations. Log Odds Ratios (95% CI) are obtained from logistic regression adjusting the effect of gender on HIV disease progression in each country.

allele were also associated with a longer time to progression in our cohort. This raised the question of the primary driver of our association. In our dataset HLA-C*03:02 showed weak LD ($r^2 = 0.18$) with HLA-B*58:01 (although this varied between LTNPs: $r^2 = 0.22$, and RPs:$r^2 = 0.13$ and between countries) (Figure 4-6A, Supp. Table 10). Therefore we sought to understand whether the effects of HLA-C*03:02 were being mediated or augmented through interactions with B*58:01. For this analysis, we utilized the interaction analysis proposed by Svejgaard in which the effect of the alleles under consideration are conditional upon co-inheritance (C03+/B58+) or not (C03+/B58- and C03-/B58+) (Svejgaard & Ryder, 1994).

Surprisingly, the protective effect of HLA-C*03:02 was not statistically significant in the absence of B*58:01 (B58-/C03+; OR 4.15, p=0.106; Figure 4-10), which was also observed for children expressing HLA-B*58:01 without C*03:02 (B58+/C03-OR 1.37, Pc=0.311). We found that the protective effect of HLA-C*03:02 remained even without concomitant

Table 4-5: HLA alleles associated with LTNP in Uganda and Botswana

HLA Allele	Allele Frequency			P-value	OR (95% CI)	Pc value†
	LTNPs (n = 393)	RPs (n = 416)	Total			
Protective						
B*57:03	0.038	0.012	0.025	0.0026	3.21 (1.50–6.86)	0.025
B*58:01	0.066	0.039	0.052	0.005	1.89 (1.21–2.96)	0.033
C*03:02	0.028	0.006	0.017	0.0022	4.74 (1.74–12.85)	0.033
C*08:02	0.042	0.019	0.030	0.0284	2.05 (1.07–3.90)	0.124
Susceptible						
B*15:10	0.037	0.077	0.057	0.0025	0.48 (0.30–0.77)	0.025
B*41:01	0.008	0.026	0.017	0.0139	0.31 (0.12–0.78)	0.069
C*07:04	0.021	0.036	0.029	0.0206	0.45 (0.23–0.88)	0.124
C*16:01	0.046	0.077	0.062	0.0333	0.62 (0.41–0.96)	0.124
B*18:01	0.031	0.051	0.041	0.0475	0.60 (0.36–0.99)	0.189

NOTE: Results of all observed HLA allele association per country are provided in Supp. Table 11
Bold indicates alleles that satisfy FDR correction $p < 0.05$
† FDR adjusted P-value
n, number of participants

Table 4-6: Haplotype frequency and association with LTNP

Haplotype*	Haplotype Frequency		P-value†	Haplotype effect§	
	LTNP	RP		OR (95% CI)	Pc value
Protective					
B*57:03~C*07:01	0.020	0.003	0.007	5.40 (1.40–20.79)	0.025
B*58:01~C*03:02	0.022	0.005	0.004	4.88 (1.50–15.86)	0.011
global p = 0.006					
Susceptible					
A*29:02~C*17:01	0.010	0.029	0.008	0.26 (0.08–0.79)	0.003
global p = 0.155					

†The p-value is based on the individual haplotype association with LTNP compared to RP
§Odds ratios from generalized linear regression models comparing haplotype frequency with the most frequent haplotype and adjusted for country and gender
Note: In healthy population the haplotype frequencies of HLA-B*57:03~C*07:01 are 1.4% in

Uganda and 2.5% in Kenyan Nandi, HLA-B*58:01~C*03:02 are 2% in Uganda and 4.2% in Kenyan Nandi and HLA-A*29:02~C*17:01 are 2.5% and 0.352% in B*42:01 containing haplotypes in Kenyan Nandi and black South Africans (Gonzalez-Galarza et al., 2020).

co-inheritance of B*58:01 (C03+/B58-) although the statistical significance was reduced in accordance with the limited sample size (OR 4.15, Pc=0.106); the reverse was also true (B58+/C03-, OR 1.37, Pc=0.311). Nevertheless, the co-expression of HLA-C*03:02 and B*58:01 (B58+/C03+) was associated with an effect size that was comparable with HLA-C on its own and was statistically significant (OR 4.15, Pc=0.005 [$p < 0.05/6$ Svejgaard correction for established LD]). This suggested that the effect of HLA-C*03:02 was either dominant with respect to HLA-B*58:01 or synergistic with it. To better understand the potential for synergistic effects, we looked for additive effects with other putatively associated alleles such as HLA-A*74:01, which was marginally associated with LTNP in Uganda (Figure 4-8, Figure 4-8 and Supp. Table 11) and has been associated with viremic control (Matthews et al., 2011). Whereas HLA-A*74:01 was marginally associated with LTNP in the absence of HLA-C*03:02 (A74+/C03-; aOR 1.66, Pc=0.048), the combined protective effect of the haplotype was much stronger than either allele separately (OR 6.58, Pc=0.0149). HLA-C*03:02 and A*74:01 are in very weak LD in our cohort (r2 = 0.02, Figure 4-6A), suggesting the effect of HLA-C*03:02

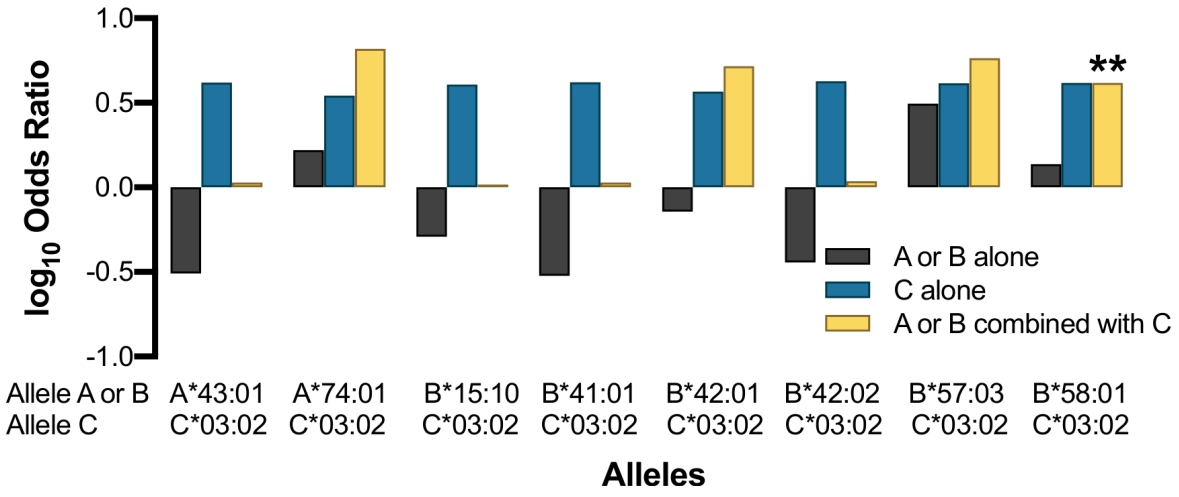


Figure 4-10: Effect of HLA-C*03:02 on other HLA class I alleles.

Asterix indicates allele combination that is statistically significant (Pc < 0.05) after correcting by a factor of 9 (Svejgaard & Ryder, 1994)).

was synergistic to the marginal effect of HLA-A*74:01. Although our study was not fully powered statistically to detect epistatic differences, we consistently observed the emergence of a protective effect in all allelic combinations that included HLA-C*03:02 with strongly protective (B*57:03) as well as putatively detrimental (HLA-B*42:01, A*43:01, B*15:10, and B*41:01) alleles (Figure 4-10), despite little LD between the alleles; this effect was not evident with overtly detrimental alleles. These data provide preliminary evidence that HLA-C*03:02 may have a cooperative role when co-inherited with other protective alleles in African pediatric HIV-1 disease progression.

4.2.5 Discussion

Using HLA typing derived from exome sequencing, we provide what to our knowledge is the largest survey of HLA allelotypes in pediatric HIV-1 infection in Africa. The sample sizes employed were more than sufficient to identify reported strong effects on disease progression. We confirmed previously known associations (HLA-B*57:03 and B*58:01) alongside a novel HLA-C*03:02 association with LTNP. The effect sizes of the protective alleles and haplotypes ranged from 1.89 to 5.4, and, consistent with previous studies, these variants explain up to 16.5% of LTNP variation in our cohort. Our results expand upon the known and anticipated alleles and effect sizes in comparable HIV-1 infected pediatric populations in Africa (Shepherd et al., 2015).

The replication of previously reported HLA-HIV-1 associations (HLA-B*57:03, B*58:01, and B*15:10) in African populations supports our methodological approach of using exome data for HLA typing in these cohorts. We observed slightly higher protective effects (OR 3.21) of HLA-B*57:03 in our gender-mixed cohort than expected from the female sex workers cohort in Pumwani Kenya (OR 2.68) (Peterson et al., 2013). In Southern African adults, HLA-B*57:03 and B*58:01 have been associated with low set-point viral load and a higher $CD4^+$ T cell count during early infection - two phenotypes that positively correlate with LTNP status in adults (Chopera et al., 2011; Tang et al., 2010). Indeed, functional studies among adult LTNPs and elite controllers expressing B*57:03 alleles have demonstrated that this allele presents epitopes that elicit superior $CD8^+$ T cell cytotoxicity activity (Migueles et al., 2015). These HLA-B associations are consistent with previous studies among adults infected with HIV-1 in Uganda, Zambia, Kenya, and South Africa (Leslie et al., 2010; Matthews et al., 2011;

Peterson et al., 2013; Serwanga et al., 2009; Tang et al., 2010), as well as other Caucasian populations (J. M. Carlson, Listgarten, et al., 2012; Chopera et al., 2011; Gijsbers et al., 2013; Klepiela et al., 2004; Lazaryan et al., 2006) and collectively, support the notion that heterozygous states of HLA protective alleles are advantageous, probably in both children and adults (Arora et al., 2020; Carrington et al., 1999), likely by providing alternative or diverse epitope presentation pathways especially in the emergence of HIV-1 escape mutants (Arora et al., 2020; J. M. Carlson, Listgarten, et al., 2012; Gijsbers et al., 2013). This is the first report of HLA-B associations in pediatric HIV-1 infected populations from Africa (Adland et al., 2015; Shepherd et al., 2015).

The HLA-C*03:02 allele was associated with a four-fold increase in LTNP status among carriers; a similar association has not been reported in LTNP studies in adults. HLA-C*03:02 has been associated with a plethora of immune and immune-mediated phenomena, including methimazole-induced hepatotoxicity in patients treated for Graves disease (X. Li et al., 2019), lower BMI (Shen et al., 2018), and the development of eclampsia, allopurinol-induced SJS, and toxic epidermal necrolysis (Cristallo et al., 2011; Ziraldo et al., 2013). Other HLA-C alleles - HLA-C*08:02 and C*08:04 - have been reported as the main drivers of HIV-1 control in adolescents from Zimbabwe (n=126) (Shepherd et al., 2015). In our cohort HLA-C*08:02 did not reach statistical significance after correction for multiple testing (Table 4-5). HLA-C*03:02 and C*08:02/04 molecules differ at positions 35 (R35Q), 114 (D114N), 116 (S116F), and 163 (L165T), located in pockets B, C, D, and E, respectively of the protein binding groove, suggesting that these alleles may have differing peptide binding affinities for different HIV-1 epitopes (Kloverpris et al., 2012). Higher extracellular surface expression of HLA-C molecules has been proposed and demonstrated as the most likely mechanism in HLA-C-mediated delayed HIV-1 disease progression (Apps et al., 2013; Thomas et al., 2009), however, the mean expression of HLA-C*03:02 is lower than C*08:02 (Apps et al., 2013) on $CD3^+$ cells; therefore there may be additional mechanisms through which HLA-C*03:02 can control HIV-1 in African children, such as microRNA regulation of HLA-C expression and KIR recognition (Kulkarni et al., 2011; H. Li et al., 2018; Vargas et al., 2020).

HLA-C*03:02 was found in weak LD with B*58:01 ($r^2 = 0.18$), and both were associated with long-term control of HIV. Interestingly, the haplotypic effect of HLA-C*03:02 in combination

with HLA-B*58:01 is similar to the individual allelic effect of HLA-C*03:02 (OR 4.88 vs. 4.78), and the allelic effect of carrying HLA-C*03:02 is twice that seen among carriers of B*58:01 (OR 4.68 vs. 1.72). However, our analysis of carriers of HLA-C*03:02 without B*58:01 showed that the associations were lost despite the strong association with protection observed when the allele was assessed agnostic to other alleles or as part of a haplotype. HLA-C*03:02 also showed an apparent additive effect when combined with other putatively protective alleles such as HLA-A*74:01, and this apparent synergistic effect of HLA-C*03:02 was consistently observed in combinations with detrimental alleles (Figure 4-10). These observations, coupled with the overall heterozygous advantage of the HLA-C*03:02 allele ($p= 5.7 \times 10^{-4}$), lead us to postulate that HLA-C*03:02 may play a synergistic role in LTNP in our population groups.

We only observed a single susceptible allele and haplotype (enriched among rapid progressors) in our cohort - we confirm a previous independent association between HLA-B*15:10 and rapid progression with a 55% reduction in the odds of LTNP being observed in our cohort. This is congruent with prior data from Kenya (OR 0.45 vs. 0.49) (Peterson et al., 2013), which may partially reflect the predominant HIV-1A and D clades seen in both Uganda and Kenya (Sampathkumar et al., 2014). This association was also seen in Botswana, where the predominant clade is HIV-1C (Payne et al., 2014; Serwanga et al., 2009).; some HLA alleles are known to bind and present epitopes from multiple clades [allelic promiscuity (Chappell et al., 2015)], which could explain the broad applicability of this observation, as such alleles could have poor specificity and so potentiate viral replication (Honeyborne et al., 2006). HLA-B*15:10 was found to be expressed at lower levels in cells defiecient of transporter associated with antigen processing 1 (TAP1) and TAP2, implying that B*15:10 might utilize TAP-dependent pathways for loading epitopes (Geng et al., 2018). TAP-dependent pathways are frequent targets of viral immune evasion (Verweij et al., 2015). Haplotype HLA-A*29:01-C*17:01 was also associated with rapid progression; consistent with Sampathkumar *et al*. who found that a HLA-C*17:01 containing haplotype was associated with faster HIV-1 seroconversion among sex-workers in Kenya (Sampathkumar et al., 2014).

Some limitations of our study merit mention. Given the long-term follow-up in our cohort, the study is limited by survival bias, therefore, the associations observed could be due to

enrichment of protective HLA variants. HLA genotypes were inferred from exome sequence data, as opposed to other sequence-based or sequence-specific primer typing techniques. Despite the enrichment for exonic regions in the genome during exome sequencing, the high polymorphism in the HLA region makes such data less reliable, potentially leading to misclassification. Current evidence, however, suggests that this may not be as significant a concern as initially feared (Duke et al., 2016; Sverchkova et al., 2019; Tokić et al., 2020), and we observed a high degree of concordance between inferred and allelotyped alleles in our validation experiments. While HWE checks are useful for validating genotypes/genotyping in disease-free controls, we did not have access to suitably matched healthy cohorts and deviations from HWE are not always sufficiently sensitive to case ascertainment status, particularly with dense datasets. Instead, we performed a limited validation of the results using an orthogonal method of allelotyping, which showed almost perfect concordance with our inferred allelotypes. Thus, we do not believe the allelotyping to be generating false positive results. Given the modest sample sizes employed, we were underpowered to stratify the haplotype analyses by country; independent replication of our findings in other larger pediatric populations is thus highly desired. Unfortunately, WHO changes on ART initiation (test and treat strategy) (Kretzschmar et al., 2013) mean that such studies will necessarily need to rely upon large retrospective datasets, which are relatively scarce within this age group.

The primary objective of our study was to evaluate enrichment of class I HLA loci on HIV disease progression (rather than HIV acquisition); necessarily, this required all participants to have the exposure of being HIV positive. A better context for our results would be comparison of allele frequencies among unaffected (normal) individuals from our population groups; however, there are no unaffected pediatric cohorts from our cohort countries, and comparisons across generations (adults vs children), particularly in the context of HIV where there is likely to be ascertainment differences between sites, are made with trepidation–. None-the-less, we provide population frequencies of the highlighted alleles from healthy available populations for completeness (Supp. Table 6). Unlike Uganda, there is no comparable HLA frequency data from Botswana in the Allele Frequency Database (AFND)(Gonzalez-Galarza et al., 2020). We therefore chose to use South African Blacks or Zimbabwe Shona from the AFND and adult data from Uganda and Kenya (Luo and Nandi) as proxy populations for our Botswana and Uganda cohorts, respectively. In Botswana, this is undoubtedly imperfect, although the black

South African populations are believed to be closely related to the Tswana people in Botswana, with some evidence of linguistic links across countries (Table 4-6 and Supp. Table 6).

Overall, the protective alleles and haplotypes reported here account for <20% of LTNP in our cohort, leaving more than two-thirds of LTNP unexplained. Therefore, we hypothesise that as yet unrecognized host and other factors could play key roles in determining HIV-1 progression in African children. For instance, we have recently documented an abundance of Anelloviridae viral species among LTNPs in our cohort (Mwesigwa et al., 2021). Anelloviridae are thought to impact NK cell activity through NF-kB, and, since HLA is known to interact with killer immunoglobulin-like receptors (KIR) on natural killer (NK) cells, the synergy between these two mechanisms could provide an alternative pathway for clearance of virally infected cells via NK cell cytotoxicity (Paximadis et al., 2011). Exploring these and other factors in our population cohorts may yield additional factors that mediate pediatric LTNP status in Africa and provide a potential path to new vaccines and therapeutics.

4.2.6 Conclusion

We provide evidence for the benefit of geographically independent multi-ethnic African populations to unravel novel HLA and HIV-1 disease associations. We confirm known and identify novel HLA associations with LTNP with quantitative interactions in two African pediatric populations. Consistent with emerging evidence, we demonstrate the role of HLA-C in the control of HIV-1 infection in children. Our results bolster a growing body of literature in support of an important role for HLA-C alleles in the control of HIV-1 among children and adolescents.

Conflict of Interest

The authors disclose no financial, commercial, or other relationships that the academic community might perceive as representing a potential conflict of interest. The funders of this research played no role in the design and conduct of the study.

Author Contributions

***Conceptualization*: SK MW MLJ GPK ARK BN MM GA IK SWM MTJ CWB GM NH; *Data curation:* SK SM EK MW FAK ANM BN JF BCM LW KM; *Formal analysis*: SK SM NH; *Funding acquisition:* SK MW MLJ GPK ARK BN MM GA IK SWM MTJ CWB GM NH JKL; *Investigation:* SK MW MLJ WEM GPK ARK BN MM GA IK SWM MTJ CWB JKL GM NH; *Methodology:* SK GPK BN MM CWB GM NH; *Project administration:* SK MW WEM MLJ GPK ARK BN MM GA IK SWM MTJ CWB GM NH; *Supervision:* GM NH MLJ ARK MM JKL; *Validation:* SK NH; *Writing – original draft:* SK NH; *Writing – review & editing:* SK MW MLJ GPK ARK BN MM GA IK SWM MTJ CWB GM NH**

Funding

The project described was supported by Award Number, U54AI110398 administered by the National Institute of Allergy and Infectious Disease (NIAID), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), and National Human Genome Research Institute (NHGRI) as part of the NIH Common Fund H3Africa Initiative. This work was also supported by the Makerere University-Uganda Virus Research Institute Centre of Excellence for Infection and Immunity Research and Training (MUII). MUII is supported through the DELTAS Africa Initiative (Grant no. 107743). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS), Alliance for Accelerating Excellence in Science in Africa (AESA), and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant no. 107743) and the UK Government.

Acknowledgments

We would like to acknowledge the following individuals for their contributions on behalf of the CAfGEN Consortium: David Kateete, Edward D. Pettitt, Marape Marape, and Bhekumusa Lukhele, who were participating investigators. We acknowledge Nasinghe Emmanuel, John Mukisa, Gaseene Sebetso, Thembela Mavuso, Bheki Ntshangase, Buhle Dlamini, Abhilash Sathyamoorthi, Bathusi Mathuba, Yves Mafulu, Gerald Mboowa, Harriet Nakayiza, Edgar Kigozi, Fred Katabazi, Keboletse Mokete, Lesego Ketumile, Kennedy Sichone, Keofentse Mathuba, LeToya Balebetse, Muambi Muyaya, Nancy Zwane, Nicholas Muriithi, Sibongile

Mumanga, Thabo Diphoko, Thobile Jele, Marion Amujal, Ronald Oceng, and Thato Regonamanye. We also acknowledge Alison Eliot's contribution, Damalie Nakanjako, Victoria Bukirwa, Joshua Mandre, and Moses Kiiza from the MUIIplus UVRI/MRC/LSHTM AIDS Research Unit.

4.2.7 References

Adland, E., Paioni, P., Thobakgale, C., Laker, L., Mori, L., Muenchhoff, M., et al. (2015). Discordant Impact of HLA on Viral Replicative Capacity and Disease Progression in Pediatric and Adult HIV Infection. *PLoS Pathog.* 11, 1–26. doi:10.1371/journal.ppat.1004954.

Alter, I., Gragert, L., Fingerson, S., Maiers, M., and Louzoun, Y. (2017). HLA class I haplotype diversity is consistent with selection for frequent existing haplotypes. *PLoS Comput. Biol.* 13, e1005693. doi:10.1371/journal.pcbi.1005693.

Apps, R., Qi, Y., Carlson, J. M., Chen, H., Gao, X., Thomas, R., et al. (2013). Influence of HLA-C expression level on HIV control. *Science (80-.).* 340, 87–91. doi:10.1126/science.1232685.

Arora, J., Pierini, F., McLaren, P. J., Carrington, M., Fellay, J., and Lenz, T. L. (2020). HLA Heterozygote Advantage against HIV-1 Is Driven by Quantitative and Qualitative Differences in HLA Allele-Specific Peptide Presentation. *Mol. Biol. Evol.* 37, 639–650. doi:10.1093/molbev/msz249.

Attaf, M., Malik, A., Severinsen, M. C., Roider, J., Ogongo, P., Buus, S., et al. (2018). Major TCR repertoire perturbation by immunodominant HLA-B*44: 03 restricted CMV-specific T cells. *Front. Immunol.* 9. doi:10.3389/fimmu.2018.02539.

Bbosa, N., Kaleebu, P., and Ssemwanga, D. (2019). HIV subtype diversity worldwide. *Curr. Opin. HIV AIDS* 14, 153–160. doi:10.1097/COH.0000000000000534.

Boudreau, J. E., and Hsu, K. C. (2018). Natural killer cell education in human health and disease. *Curr. Opin. Immunol.* 50, 102–111. doi:10.1016/j.coi.2017.11.003.

Bruzzi, P., Green, S. B., Byar, D. P., Brinton, L. A., and Schairer, C. (1985). Estimating the population attributable risk for multiple risk factors using case-control data. *Am. J. Epidemiol.* 22, 904–914. doi:10.1093/oxfordjournals.aje.a114174.

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., and Fernández-Viña, M. A. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum. Immunol.* 62, 1009–1030. doi:10.1016/S0198-8859(01)00298-1.

Cao, K., Moormann, A. M., Lyke, K. E., Masaberg, C., Sumba, O. P., Doumbo, O. K., et al. (2004). Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens* 63, 293–325. doi:10.1111/j.0001-2815.2004.00192.x.

Carlson, J. M., Listgarten, J., Pfeifer, N., Tan, V., Kadie, C., Walker, B. D., et al. (2012). Widespread Impact of HLA Restriction on Immune Control and Escape Pathways of HIV-1. *J. Virol.* 86, 5230–5243. doi:10.1128/jvi.06728-11.

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., et al. (1999). HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. *Science (80-.).* 283, 1748–1752. doi:10.1126/science.283.5408.1748.

Chappell, P., Meziane, E. K., Harrison, M., Magiera, L., Hermann, C., Mears, L., et al. (2015). Expression levels of mhc class i molecules are inversely correlated with promiscuity of peptide binding. *Elife* 2015, 1–22. doi:10.7554/eLife.05345.

Chopera, D. R., Mlotshwa, M., Woodman, Z., Mlisana, K., de Assis Rosa, D., Martin, D. P., et al. (2011). Virological and immunological factors associated with HIV-1 differential disease progression in HLA-B 58:01-positive individuals. *J. Virol.* 85, 7070–7080.

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-depth African genomes inform human migration and health. *Nature* 586, 741–748. doi:10.1038/s41586-020-2859-7.

Cristallo, A. F., Schroeder, J., Citterio, A., Santori, G., Ferrioli, G. M., Rossi, U., et al. (2011). A study of HLA class I and class II 4-digit allele level in Stevens-Johnson syndrome and toxic epidermal necrolysis. *Int. J. Immunogenet.* 38, 303–309. doi:10.1111/j.1744-313X.2011.01011.x.

Duke, J. L., Lind, C., Mackiewicz, K., Ferriola, D., Papazoglou, A., Gasiewski, A., et al. (2016). Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA* 87, 141–152. doi:10.1111/tan.12736.

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x.

Fan, Y., and Song, Y. Q. (2017). PyHLA: Tests for the association between HLA alleles and diseases. *BMC Bioinformatics* 18, 1–5. doi:10.1186/s12859-017-1496-0.

Frater, A. J., Brown, H., Oxenius, A., Gunthard, H. F., Hirschel, B., Robinson, N., et al. (2007). Effective T-Cell Responses Select Human Immunodeficiency Virus Mutants and Slow Disease Progression. *J. Virol.* 81, 6742–6751. doi:10.1128/JVI.00022-07.

Geng, J., Zaitouna, A. J., and Raghavan, M. (2018). Selected HLA-B allotypes are resistant to inhibition or deficiency of the transporter associated with antigen processing (TAP). *PLoS Pathog.* 14, e1007171. doi:10.1371/journal.ppat.1007171.

Gijsbers, E. F., Anton Feenstra, K., Van Nuenen, A. C., Navis, M., Heringa, J., Schuitemaker, H., et al. (2013). HIV-1 replication fitness of HLA-B*57/58:01 CTL Escape variants is restored by the accumulation of compensatory mutations in Gag. *PLoS One* 8, e81235. doi:10.1371/journal.pone.0081235.

Gonzalez-Galarza, F. F., McCabe, A., Santos, E. J. M. Dos, Jones, J., Takeshita, L., Ortega-Rivera, N. D., et al. (2020). Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* 48, D783–D788. doi:10.1093/nar/gkz1029.

González, N., McKee, K., Lynch, R. M., Georgiev, I. S., Jimenez, L., Grau, E., et al. (2018). Characterization of broadly neutralizing antibody responses to HIV-1 in a cohort of long term non-progressors. *PLoS One* 13, e0193773. doi:10.1371/journal.pone.0193773.

Goulder, P. J. R., and Walker, B. D. (2012). HIV and HLA Class I: an evolving relationship. *Immunity* 37, 426–440. doi:10.1016/j.immuni.2012.09.005.HIV.

Gutierrez, R. G. (2010). Stata. *Wiley Interdiscip. Rev. Comput. Stat.* doi:10.1002/wics.116.

Honeyborne, I., Rathod, A., Buchli, R., Ramduth, D., Moodley, E., Rathnavalu, P., et al. (2006). Motif Inference Reveals Optimal CTL Epitopes Presented by HLA Class I Alleles Highly Prevalent in Southern Africa 1. *J. Immunol.* 4201, 4699–4705.

Hua, S., Lécuroux, C., Sáez-Cirión, A., Pancino, G., Girault, I., Versmisse, P., et al. (2014). Potential role for HIV-specific CD38-/HLA-DR+ CD8+ T cells in viral suppression and cytotoxicity in HIV controllers. *PLoS One* 9, e101920. doi:10.1371/journal.pone.0101920.

Huang, Y., Yang, J., Ying, D., Zhang, Y., Shotelersuk, V., Hirankarn, N., et al. (2015). HLAreporter: A tool for HLA typing from next generation sequencing data. *Genome Med.* 7. doi:10.1186/s13073-015-0145-3.

Kaseke, C., Tano-Menka, R., Senjobe, F., and Gaiha, G. D. (2021). The Emerging Role for CTL Epitope Specificity in HIV Cure Efforts. *J. Infect. Dis.* 223, S32–S37. doi:10.1093/infdis/jiaa333.

Kijak, G. H., Walsh, A. M., Koehler, R. N., Moqueet, N., Eller, L. A., Eller, M., et al. (2009). HLA class i allele and haplotype diversity in Ugandans supports the presence of a major east African genetic cluster. *Tissue Antigens* 73, 262–269. doi:10.1111/j.1399-0039.2008.01192.x.

Klepiela, P., Leslie, A. J., Honeyborne, I., Ramduth, D., Thobakgale, C., Chetty, S., et al. (2004). Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432, 769–775. doi:10.1038/nature03113.

Kloverpris, H. N., Harndahl, M., Leslie, A. J., Carlson, J. M., Ismail, N., van der Stok, M., et al. (2012). HIV Control through a Single Nucleotide on the HLA-B Locus. *J. Virol.* 86, 11493–11500. doi:10.1128/jvi.01020-12.

Kretzschmar, M. E., Schim Van Der Loeff, M. F., Birrell, P. J., De Angelis, D., and Coutinho, R. A. (2013). Prospects of elimination of HIV with test-and-treat strategy. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15538–15543. doi:10.1073/pnas.1301801110.

Kulkarni, S., Savan, R., Qi, Y., Gao, X., Yuki, Y., Bass, S. E., et al. (2011). Differential microRNA regulation of HLA-C expression and its association with HIV control. *Nature* 472, 495–498. doi:10.1038/nature09914.

Kyobe, S., Kisitu, G. P., Farirai, J., Mwesigwa, S., Mirembe, A. N., Katagirya, E., et al. Characterization of HIV progression among children in Uganda and Botswana: A retrospective cohort study. Unpublished manuscript.

Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P., and Thomson, G. (2007). PyPop update - A software pipeline for large-scale multilocus population genomics. *Tissue Antigens* 69, 192–197. doi:10.1111/j.1399-0039.2006.00769.x.

Lancaster, A., Nelson, M. P., Meyer, D., Single, R. M., and Thomson, G. (2003). PyPop: A Software Framework for Population Genomics: Analyzing Large-Scale Multi-Locus Genotype Data. *Pac Symp Biocomput* 6, 790–795. doi:10.1016/j.pmrj.2014.02.014.Lumbar.

Lazaryan, A., Lobashevsky, E., Mulenga, J., Karita, E., Allen, S., Tang, J., et al. (2006). Human Leukocyte Antigen B58 Supertype and Human Immunodeficiency Virus Type 1 Infection in Native Africans. *J. Virol.* 80, 6056–6060. doi:10.1128/jvi.02119-05.

Leslie, A., Matthews, P. C., Listgarten, J., Carlson, J. M., Kadie, C., Ndung'u, T., et al. (2010). Additive Contribution of HLA Class I Alleles in the Immune Control of HIV-1 Infection. *J. Virol.* 84, 9879–9888. doi:10.1128/jvi.00320-10.

Levin, M. L. (1953). The occurrence of lung cancer in man. *Acta Unio Int. Contra Cancrum* 9, 531–541.

Li, H., Ivarsson, M. A., Walker-Sperling, V. E., Subleski, J., Johnson, J. K., Wright, P. W., et al. (2018). Identification of an elaborate NK-specific system regulating HLA-C expression. *PLoS Genet.* 14, e1007163. doi:10.1371/journal.pgen.1007163.

Li, X., Jin, S., Fan, Y., Fan, X., Tang, Z., Cai, W., et al. (2019). Association of HLA-C*03:02 with methimazole-induced liver injury in Graves' disease patients. *Biomed. Pharmacother.* 117, 109095. doi:10.1016/j.biopha.2019.109095.

Martin, G., and Martin, G. (2021). "R Studio," in *An Introduction to Programming with R* doi:10.1007/978-3-030-69664-1_1.

Matthews, P. C., Adland, E., Listgarten, J., Leslie, A., Mkhwanazi, N., Carlson, J. M., et al. (2011). HLA-A*7401–Mediated Control of HIV Viremia Is Independent of Its Linkage Disequilibrium with HLA-B*5703. *J. Immunol.* 186, 5675–5686. doi:10.4049/jimmunol.1003711.

Migueles, S. A., Mendoza, D., Zimmerman, M. G., Martins, K. M., Toulmin, S. A., Kelly, E. P., et al. (2015). CD8+ T-cell Cytotoxic Capacity Associated with Human Immunodeficiency Virus-1 Control Can Be Mediated through Various Epitopes and Human Leukocyte Antigen Types. *EBioMedicine* 2, 46–58. doi:10.1016/j.ebiom.2014.12.009.

Mwesigwa, S., Williams, L., Retshabile, G., Katagirya, E., Mboowa, G., Mlotshwa, B., et al. (2021). Unmapped exome reads implicate a role for Anelloviridae in childhood HIV-1 long-term non-progression. Unpublished manuscript. *Genomic Med.* 6. doi:10.1038/s41525-021-00185-w.

Paximadis, M., Minevich, G., Winchester, R., Schramm, D. B., Gray, G. E., Sherman, G. G., et al. (2011). KIR-HLA and maternal-infant HIV-1 transmission in sub-Saharan Africa. *PLoS One* 6, e1654. doi:10.1371/journal.pone.0016541.

Payne, R., Muenchhoff, M., Mann, J., Roberts, H. E., Matthews, P., Adland, E., et al. (2014). Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proc. Natl. Acad. Sci.* 111, E5393–E5400. doi:10.1073/pnas.1413339111.

Peloso, G. M., Rader, D. J., Gabriel, S., Kathiresan, S., Daly, M. J., and Neale, B. M. (2016). Phenotypic extremes in rare variant study designs. *Eur. J. Hum. Genet.* 24, 924–930. doi:10.1038/ejhg.2015.197.

Peterson, T. A., Kimani, J., Wachihi, C., Bielawny, T., Mendoza, L., Thavaneswaran, S., et al. (2013). HLA class I associations with rates of HIV-1 seroconversion and disease progression in the Pumwani Sex Worker Cohort. *Tissue Antigens* 81, 93–107. doi:10.1111/tan.12051.

Retshabile, G., Mlotshwa, B. C. B. C., Williams, L., Mwesigwa, S., Mboowa, G., Huang, Z., et al. (2018). Whole-Exome Sequencing Reveals Uncaptured Variation and Distinct Ancestry in the Southern African Population of Botswana. *Am. J. Hum. Genet.* 102, 731–743.

Sampathkumar, R., Peters, H. O., Mendoza, L., Bielawny, T., Ngugi, E., Kimani, J., et al. (2014). Influence of HLA class I haplotypes on HIV-1 seroconversion and disease progression in Pumwani sex worker cohort. *PLoS One* 9, 1–12. doi:10.1371/journal.pone.0101475.

Serwanga, J., Shafer, L. A., Pimego, E., Auma, B., Watera, C., Rowland, S., et al. (2009). Host HLA B allele-associated multi-clade gag T-cell recognition correlates with slow HIV-1 disease progression in antiretroviral theraphy-naïve Ugandans. *PLoS One* 4, e4188. doi:10.1371/journal.pone.0004188.

Shen, J., Guo, T., Wang, T., Zhen, Y., Ma, X., Wang, Y., et al. (2018). HLA-B∗07, HLA-DRB1∗07, HLA-DRB1∗12, and HLA-c∗03:02 strongly associate with BMI: Data from 1.3 million healthy Chinese adults. *Diabetes* 67, 861–871. doi:10.2337/db17-0852.

Shepherd, B. L., Ferrand, R., Munyati, S., Folkard, S., Boyd, K., Bandason, T., et al. (2015). HLA correlates of long-term survival in vertically infected HIV-1-positive adolescents in Harare, Zimbabwe. *AIDS Res. Hum. Retroviruses* 31, 504–507. doi:10.1089/aid.2014.0338.

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., et al. (2019a). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* 51, 30–35. doi:10.1038/s41588-018-0335-1.

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., et al. (2019b). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* 51, 30–35. doi:10.1038/s41588-018-0273-y.

Svejgaard, A., and Ryder, L. P. (1994). HLA and disease associations: Detecting the strongest association. *Tissue Antigens* 43, 18–27. doi:10.1111/j.1399-0039.1994.tb02291.x.

Sverchkova, A., Anzar, I., Stratford, R., and Clancy, T. (2019). Improved HLA typing of Class I and Class II alleles from next-generation sequencing data. *HLA* 94, 504–513. doi:10.1111/tan.13685.

Tang, J., Malhotra, R., Song, W., Brill, I., Hu, L., Farmer, P. K., et al. (2010). Human leukocyte antigens and HIV type 1 viral load in early and chronic infection: Predominance of evolving relationships. *PLoS One* 5, e9629. doi:10.1371/journal.pone.0009629.

Thobakgale, C. F., Prendergast, A., Crawford, H., Mkhwanazi, N., Ramduth, D., Reddy, S., et al. (2009). Impact of HLA in Mother and Child on Disease Progression of Pediatric Human Immunodeficiency Virus Type 1 Infection. *J. Virol.* 83, 10234–10244. doi:10.1128/jvi.00921-09.

Thomas, R., Apps, R., Qi, Y., Gao, X., Male, V., O'Huigin, C., et al. (2009). HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat. Genet.* 41, 1290–1294. doi:10.1038/ng.486.

Tokić, S., Žižkova, V., Štefanić, M., Glavaš-Obrovac, L., Marczi, S., Samardžija, M., et al. (2020). HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 allele and haplotype frequencies defined by next generation sequencing in a population of East Croatia blood donors. *Sci. Rep.* 10, Article number: 5513. doi:10.1038/s41598-020-62175-9.

Tshabalala, M., Ingram, C., Schlaphoff, T., Borrill, V., Christoffels, A., and Pepper, M. S. (2018). Human leukocyte antigen-A, B, C, DRB1, and DQB1 allele and haplotype frequencies in a subset of 237 donors in the South African bone marrow registry. *J. Immunol. Res.* 2018. doi:10.1155/2018/2031571.

Ueta, M., Kaniwa, N., Sotozono, C., Tokunaga, K., Saito, Y., Sawai, H., et al. (2014). Independent strong association of HLA-A*02:06 and HLA-B*44:03 with cold medicine-related Stevens-Johnson syndrome with severe mucosal involvement. *Sci. Rep.* 4. doi:10.1038/srep04862.

UNAIDS (2019). UNAIDS 2019 Data. Available https//www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf Accessed April 7, 2020. doi:978-92-9173-945-5.

van der Zwan, A., van der Meer-Prins, E. M. W., van Miert, P. P. M. C., van den Heuvel, H., Anholts, J. D. H., Roelen, D. L., et al. (2018). Cross-Reactivity of Virus-Specific CD8+ T Cells Against Allogeneic HLA-C: Possible Implications for Pregnancy Outcome. *Front. Immunol.* 9. doi:10.3389/fimmu.2018.02880.

Vargas, L. de B., Dourado, R. M., Amorim, L. M., Ho, B., Calonga-Solís, V., Issler, H. C., et al. (2020). Single Nucleotide Polymorphism in KIR2DL1 Is Associated With HLA-C Expression in Global Populations. *Front. Immunol.* 11. doi:10.3389/fimmu.2020.01881.

Verweij, M. C., Horst, D., Griffin, B. D., Luteijn, R. D., Davison, A. J., Ressing, M. E., et al. (2015). Viral Inhibition of the Transporter Associated with Antigen Processing (TAP): A Striking Example of Functional Convergent Evolution. *PLoS Pathog.* 11, e1004743. doi:10.1371/journal.ppat.1004743.

Warszawski, J., Lechenadec, J., Faye, A., Dollfus, C., Firtion, G., Meyer, L., et al. (2007). Long-Term Nonprogression of HIV Infection in Children: Evaluation of the ANRS Prospective French Pediatric Cohort. *Clin. Infect. Dis.* 45, 785–794. doi:10.1086/521165.

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N. Y).* 38, 1358–1370.

Yang, F., Jiang, M., Zhang, W., Qiao, Y., Chen, S. an, Wang, D., et al. (2020). HLA-A*24:02 is associated with metronidazole-induced cutaneous adverse drug reactions in Han Chinese individuals: A pilot case-control study with both HLA gene and T cell receptor repertoire analysis. *Basic Clin. Pharmacol. Toxicol.* 126, 133–143. doi:10.1111/bcpt.13315.

Data availability

Exome Sequence data used in this paper are publicly available at the European Genome and Phenome Archive (EGA) through the Data and Biospecimen Committee (DBAC) of H3Africa. The procedures for access are elaborated in the policy found at https://h3africa.org/wp-content/uploads/2018/05/App-D-H3Africa-Data-and-Biospecimen-Access-Committee-Guidelines-final-10-July-2017.pdf . Requests are subject to approval by the DBAC.

4.2.8 Published article

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fgene.2021.720213

Exome Sequencing Reveals a Putative Role for HLA-C*03:02 in Control of HIV-1 in African Pediatric Populations

Samuel Kyobe[1], Savannah Mwesigwa[1,2], Grace P. Kisitu[3], John Farirai[4], Eric Katagirya[2], Angella N. Mirembe[3], Lesego Ketumile[4], Misaki Wayengera[2], Fred Ashaba Katabazi[2], Edgar Kigozi[2], Edward M. Wampande[2], Gaone Retshabile[5], Busisiwe C. Mlotshwa[5], Lesedi Williams[5], Koketso Morapedi[5], Ishmael Kasvosve[5], Jacqueline Kyosiimire-Lugemwa[6], Betty Nsangi[3], Masego Tsimako-Johnstone[7], Chester W. Brown[8], Moses Joloba[2], Gabriel Anabwani[4], Lukhele Bhekumusa[9], Sununguko W. Mpoloka[5], Graeme Mardon[10,11], Mogomotsi Matshaba[4,12], Adeodata Kekitiinwa[3,12] and Neil A. Hanchard[10†] on behalf of the Collaborative African Genomics Network (CAfGEN) of the H3Africa Consortium*

[1] Department of Medical Microbiology, College of Health Sciences, Makerere University, Kampala, Uganda, [2] Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, Kampala, Uganda, [3] Baylor College of Medicine Children's Foundation, Kampala, Uganda, [4] Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana, [5] School of Allied Health Professions, Faculty of Health Sciences, University of Botswana, Gaborone, Botswana, [6] The Medical Research Council (MRC)/UVRI & LSHTM Uganda Research Unit, Entebbe, Uganda, [7] Australian Medical Research, Hurstville, NSW, Australia, [8] University of Tennessee Health Science Center, Memphis, TN, United States, [9] Eswatini - Baylor College of Medicine Children's Foundation, Mbabane, Eswatini, [10] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States, [11] Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, United States, [12] Pediatric Retrovirology, Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States

OPEN ACCESS

Edited by:
Dana C. Crawford, Case Western Reserve University, United States

Reviewed by:
Abeer Madbouly, Center for International Blood and Marrow Transplant Research (CIBMTR), United States
Juan Carlos Fernandez-Lopez, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico

***Correspondence:**
Neil A. Hanchard
Neil.hanchard@nih.gov

†Present address:
Neil A. Hanchard
National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, United States

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *03 June 2021*
Accepted: *05 August 2021*
Published: *26 August 2021*

Citation:
*Kyobe S, Mwesigwa S, Kisitu GP, Farirai J, Katagirya E, Mirembe AN, Ketumile L, Wayengera M, Katabazi FA, Kigozi E, Wampande EM, Retshabile G, Mlotshwa BC, Williams L, Morapedi K, Kasvosve I, Kyosiimire-Lugemwa J, Nsangi B, Tsimako-Johnstone M, Brown CW, Joloba M, Anabwani G, Bhekumusa L, Mpoloka SW, Mardon G, Matshaba M, Kekitiinwa A and Hanchard NA (2021) Exome Sequencing Reveals a Putative Role for HLA-C*03:02 in Control of HIV-1 in African Pediatric Populations. Front. Genet. 12:720213. doi: 10.3389/fgene.2021.720213*

Human leucocyte antigen (HLA) class I molecules present endogenously processed antigens to T-cells and have been linked to differences in HIV-1 disease progression. HLA allelotypes show considerable geographical and inter-individual variation, as does the rate of progression of HIV-1 disease, with long-term non-progression (LTNP) of disease having most evidence of an underlying genetic contribution. However, most genetic analyses of LTNP have occurred in adults of European ancestry, limiting the potential transferability of observed associations to diverse populations who carry the burden of disease. This is particularly true of HIV-1 infected children. Here, using exome sequencing (ES) to infer HLA allelotypes, we determine associations with HIV-1 LTNP in two diverse African pediatric populations. We performed a case-control association study of 394 LTNPs and 420 rapid progressors retrospectively identified from electronic medical records of pediatric HIV-1 populations in Uganda and Botswana. We utilized high-depth ES to perform high-resolution HLA allelotyping and assessed evidence of association between HLA class I alleles and LTNP. Sixteen HLA alleles and haplotypes had significantly different frequencies between Uganda and Botswana, with allelic differences being more prominent in HLA-A compared to HLA-B and C allelotypes. Three HLA allelotypes showed association with LTNP, including a novel association in HLA-C (HLA-B*57:03, aOR 3.21, *Pc* = 0.0259; B*58:01, aOR 1.89, *Pc* = 0.033; C*03:02, aOR 4.74, *Pc* = 0.033). Together, these alleles convey an estimated population attributable risk (PAR) of non-progression of 16.5%. We also observed novel haplotype

The full article is available as open access online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.720213/full

4.3 Identification of a clade-specific HLA-C*03:02-restricted epitope GY9 derived from the HIV-1 p17 matrix protein

Authors: Samuel Kyobe,[1,*,†] Savannah Mwesigwa,[1,2] Gyaviira Nkurunungi,[3,4] Gaone Retshabile,[5] Moses Egesa,[3,4] Joel Kabali,[3] Grace P. Kisitu,[6] Eric Katagirya,[2] Marion Amujal,[2] Misaki Wayengera,[2] Fred Katabazi,[2] Edgar Kigozi,[2] Fahim Yiga,[2] Busisiwe C. Mlotshwa,[5] Lesedi Williams,[5] Jacqueline Kyosiimire-Lugemwa,[3] Hakim Sendagire,[1] Moses Joloba,[3] Sununguko W. Mpoloka,[5] Dithan Kiragga,[6] Graeme Mardon,[7,8] Adeodata Kekitiinwa,[6,9] Mogomotsi Matshaba,[9,10] Neil A. Hanchard[11] and David Robinson,[12] on behalf of the Collaborative African Genomics Network (CAfGEN) of the H3Africa Consortium

[1]Department of Medical Microbiology, College of Health Sciences, Makerere University, Kampala, Uganda; [2]Department of Immunology and Molecular Biology, College of Health Sciences, Makerere; [3]The Medical Research Council (MRC)/UVRI & LSHTM Uganda Research Unit; [4]Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, United; [5]School of Allied Health Professions, Faculty of Health Sciences, University of Botswana, Gaborone, Botswana; [6]Baylor College of Medicine Children's Foundation, Kampala, Uganda; [7]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [8]Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX 77030, USA; [9]Pediatric Retrovirology, Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; [10]Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana; [11]National Human Genome Research Institute, National Institutes of Health, USA; [12]Department of Chemistry and Forensics, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, United Kingdom.[†]Lead contact

*Corresponding author: samuelkyobe@gmail.com

4.3.1 Abstract

Efforts towards an effective HIV-1 vaccine have remained mainly unsuccessful. There is increasing evidence for a potential role of HLA-C-restricted $CD8^+$ T cell responses in HIV-1 control, including our recent report of HLA-C*03:02 among African children. However, there are no documented optimal HIV-1 $CD8^+$ T cell epitopes restricted by HLA-C*03:02; additionally, the structural influence of HLA-C*03:02 on epitope binding is undetermined. Immunoinformatics approaches provide a fast and inexpensive method to discover HLA-restricted epitopes. Here, we employed immunopeptidomics to identify HLA-C*03:02-restricted $CD8^+$ T cell epitopes. We identified a clade-specific Gag-derived GY9 (GTEELRSLY) HIV-1 p17 matrix epitope exclusively restricted to HLA-C*03:02. Residues E62, T142, and E151 in the HLA-C*03:02 binding groove and positions p3, p6, and p9 on the GY9 epitope are crucial in shaping and stabilizing the epitope binding. Our findings support the growing evidence of the contribution of HLA-C molecules to HIV-1 control and provide a prospect for vaccine strategies.

4.3.2 Background

Host genetic factors play an important role in HIV-1 control.[1] Outside the Δ32 mutation in the chemokine receptor (CCR5) gene, genome-wide association studies (GWAS) have consistently identified variants in the major histocompatibility complex (MHC, also known as human leucocyte antigens [HLA]) class I alleles to play a significant role in the control of HIV-1 infection.[2,3] The HLA class I alleles predominantly display intracellularly processed viral antigens on cell surfaces to elicit CD8+ T lymphocytes (CTL) in the adaptive immune responses.[4,5] This cell-mediated immune response is responsible for the clearance of virally infected cells.[4,5] Therefore, the mechanisms of intracellular antigen processing, including proteasome cleavage, peptide loading, and transportation into the endoplasmic reticulum via the transporter associated with antigen processing and peptide stabilization on the MHC molecule for stable cell surface presentation are a subject of interest to understand the contribution of the HLA class I in viral control.[6–10]

We recently documented the putative role of HLA-C*03:02, HLA-B*57:03, and HLA-B*58:01 in long-term non-progression (LTNP) of HIV-1 among children in Uganda and Botswana.[11] In contrast to HLA-B*57:03 and HLA-B*58:01, the specific mechanisms underlying HLA-C*03:02-mediated HIV-1 control have not been fully elucidated. As such, the molecular and immunological basis of how HLA-C*03:02 confers its protective effects against HIV-1 infection remains unknown. Prior to our work, the HLA-C*03:02 allele was demonstrated to have a significant correlation with both reduced viral load and elevated CD4+ T cell count within the South African population, although these associations did not attain statistical significance.[12] Considering these findings, it is evident that HLA-C*03:02 holds potential significance in HIV control, stimulating further investigation into the mechanisms underlying its immune-mediated effects. HLA-B*57:03 and HLA-B*58:01 molecules display highly restricted HIV-1 epitopes in the structural and non-structural proteins that mediate HIV control.[10,13–15] Most of these peptides are derived from the Gag protein; however, high immunogenicity has been demonstrated in the non-structural proteins Nef, Vif, and Vpr.[15–18] Additional research is therefore needed to unravel the intricate interplay between HLA-C*03:02 and HIV-1 proteins, as well as the immune responses triggered by this particular HLA allele. The HLA-restricted epitopes are characterized by their ability to induce effective

qualitative and quantitative cellular immune responses.[19,20] These epitopes are crucial in driving robust and polyfunctional cellular immune responses, contributing to the recognition and targeting of HIV-infected cells.[19–22] Furthermore, HIV-1 epitopes have been demonstrated to elicit humoral immune responses, generating broadly neutralizing antibodies.[23] The selective pressure of protective HLA alleles is known to drive the emergence of escape mutants, though at the expense of viral replication fitness, which other compensatory mutations may counter.[13,24] However, accumulation of escape mutations in HLA-B*57:03/B*58:01-restricted epitopes abrogates the protective effect through various mechanisms, including qualitative binding to killer immunoglobulin-like receptors.[13,25] Nonetheless, developing epitope-based vaccines that efficiently elicit both humoral and cellular immune responses has re-emerged as a strategy to control the global HIV-1 epidemic.[5,26,27]

The success of multi-epitope HIV-1 vaccines remains generally challenging due to the rapid genetic evolution of the virus, diverse HLA genetic polymorphism, and viral-clade geographical diversity.[11,28,29] Previous research has predominantly focused on characterizing HLA-restricted epitopes specific to protective HLA-A and HLA-B alleles in the context of HIV. At the same time, comparatively limited consideration has been given to exploring the protective HLA-C alleles.[10] Therefore, identifying and prioritizing protective HLA-C-restricted epitopes from locally prevalent HIV-1 clades remains viable for designing an optimal vaccine candidate. The scientific literature presents many methodological approaches for identifying optimal HIV-1 CTL epitopes, each yielding diverse outcomes.[14,30] This diversity underscores the complexity of epitope prediction and necessitates careful consideration of the most appropriate methodologies for accurate and comprehensive epitope discovery. However, it's crucial to emphasize that these approaches consistently exhibit a strong agreement between predictive and experimental methods.[31] In this study, we employed an immunoinformatics approach and identified four potentially HLA-C*03:02-restricted $CD8^+$ T cell epitopes. Furthermore, using an ELISPOT assay, we experimentally validated that a clade-specific GY9 epitope derived from the p17 HIV-1 matrix protein is exclusively restricted to HLA-C*03:02 alleles in our population. Our observations further support the growing evidence of the contribution of HLA-C molecules to HIV-1 control and provide an opportunity for innovative vaccine strategies.

4.3.3 Methods

Ethics statement

This study was approved by the School of Biomedical Sciences Institutional Review Board (IRB), Uganda National Council for Science and Technology, University of Botswana IRB, Botswana Health Research and Development Committee, and the Baylor College of Medicine IRB. All participants provided written informed consent.

Patient recruitment

We used stored PBMC samples from 25 previously recruited participants in the parent study: the Collaborative African Genomics Network (CAfGEN). The details of participant recruitment have been described in detail elsewhere.[11,32] The clinical characteristics of patients before and after treatment are presented in Supp. Table 18. We selected all 13 participants expressing the HLA-C*03:02 allele and 12 controls that are HLA-C*03:02^{-ve}.

HLA-C*03:02 homology modeling and validation

The 3D structure model of HLA-C*03:02 was predicted using SWISS-MODEL (https://swissmodel.expasy.org/), starting with the 366 amino acids full-length protein sequence downloaded from the IMGT/HLA database (https://www.ebi.ac.uk/ipd/imgt/hla/). An optimal template to model the HLA-C*03:02 protein was selected based on PDB ID: 5w6a.2 HLA-C*06:02 with a sequence identity >90%, query coverage ≥70% and X-ray resolution at ≤2Å. The constructed model underwent comprehensive validation assessments in two distinct stereochemical and spatial analysis domains. The stereochemical analysis of parameters, including bond length, torsion angle, and rotational angle, within the model was evaluated using online tools servers SAVES (https://saves.mbi.ucla.edu/) and Pro-Q scores (https://proq.bioinfo.se/cgi-bin/ProQ/ProQ.cgi). The Ramachandran plot confirmed stereochemical quality (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/). The spatial features of the model based on the 3D conformation were analyzed using the Verify 3D (https://saves.mbi.ucla.edu/) and ProSA scores (https://prosa.services.came.sbg.ac.at/prosa.php). The model's overall quality was determined from the ProTSAV score (http://www.scfbio-iitd.res.in/software/proteomics/protsav.jsp).

HIV-1 ligand prediction and preparation for docking

HIV-1 ligands (8-14mer) predicted to bind HLA-C*03:02 were determined using NetMHCpan-4.1b server (https://services.healthtech.dtu.dk/services/NetMHCpan-4.1/) and supplementary epitopes with Motif Scan (https://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan). We used full-length HIV-1 subtype A/A1 and C consensus sequences retrieved from the HIV-1 Sequence Database (https://www.hiv.lanl.gov/content/index) representative of the predominant circulating clades in Uganda and Botswana, respectively.[33] From the available ligands, we selected a subset of those predicted as strong or weak binders by NetMHCpan-4 or Motif Scan. Additionally, these ligands were predicted to undergo proteasomal cleavage according to NetChop v3.1 (https://services.healthtech.dtu.dk/services/NetChop-3.1/). For the docking experiment, the 3D structure of the selected linear peptides was predicted using the PEP-FOLD3 server (https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3) followed by an energy minimization step using the minimize structure module in Chimera.[34] Briefly, essential hydrogens were added, and Gastiger charges were assigned to ligand residues using the Amber ff14SB force field. The 3D structures achieved convergence after 100 steps of steepest descent followed by 1000 steps of conjugate gradient.[34]

Molecular docking protocol and analysis

The model HLA-C*03:02 structure was used for docking HIV-1 ligands with DINC, a parallelized meta-docking method for the incremental docking of large ligands. Some modifications were adopted to the default DINC protocol.[35] The grid box of 50 x 40 x 72 xyz points with a grid spacing of 0.375Å was generated and centered at 11.95 x 57.95 x -6.34 around the six binding pockets using AutoDock Tools.[36] To maximize the docking accuracy, the vina exhaustiveness was set to 8, and the number of binding modes generated at each round of incremental docking was set at 40. An additional round of docking was performed using the whole ligand with full flexibility to get a larger docking sampling. The predicted ligand poses were rescored using Convex-PL, shown to achieve >80% accuracy in identifying the best binders.[37] Molecular visualization with UCSF ChimeraX was used to identify and analyze the intermolecular (pHLA) interactions.[34,38] The selection of top-ranked ligand poses was guided by several rigorous criteria, including a Convex-PL score ≥7, a DINC binding score ≤ –7.0

kcal/mol, a minimum of 6 strong hydrogen bonds formed with the pHLA complex, and an RMSD ≤1.7Å compared to the native ligand (PDB ID: 5w6a.2) of the C*06:02 allele.

Molecular dynamics simulation protocol and analysis

MD simulations were performed on the top-ranked ligands using GROMACSv2020.3 software under the CHARMM36 all-atom force field.[39,40] The receptor-ligand coordinates generated during molecular docking were utilized to reconstruct protein-ligand complexes using Chimera. All hydrogen molecules were removed from the final structure. We used the Avogadro program to add hydrogens to ligands and the CHARMM General Force Field (CGenFF) program to generate ligand parameters and topologies.[41,42] The resultant HLA-C*03:02-unbound and HLA-C*03:02-ligand complex were solvated in the center of a cubic unit cell of the volume of 10000nm3 with ~31,000 molecules of TIP3-point water. We allowed a minimum distance of 1nm between the box boundary and the complex. The system was neutralized with the addition of 10 Na+ ions. The system was subjected to energy minimization using the steepest descent method with a maximum force constraint of 10kJ/mol. Position restraints were applied on both the ligand and HLA-C*03:02 receptor. The system temperature and pressure were equilibrated at 300K using the modified Berendsen thermostats coupling method and at 4.5x10-5bar-1 using the Berendsen coupling barostats method, respectively, for 1000ps. All relaxed systems were subjected to MD simulations for 200ns using periodic boundary conditions without ligand-protein restraints. The stability of the complexes was examined by analyzing changes in the root mean square deviation (RMSD) and hydrogen bonds network using GROMACS functions hbond and rms respectively.[39]

The binding free energy (denoted as $\Delta G_{bind} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{peptide}$) was calculated using the molecular mechanics (MM) with Poisson–Boltzmann (PB) and surface area solvation method implemented in the gmx_MMPBSA program.[43] The critical residues at the interface of pHLA binding (within 5Å of the ligand) were determined by performing computational alanine scanning (CAS) experiments on the ligand and HLA-C*03:02. The resultant binding free energy due to the mutant residue was calculated by comparing the wild-type ($\Delta G_{wild\text{-}type}$) and mutant ($\Delta G_{mutant}$) complexes, as denoted by the equation: $\Delta\Delta G_{bind} = \Delta G_{wild\text{-}type} - \Delta G_{mutant}$.

HIV-1 epitope conservancy analysis

To assess the positional conservancy of the candidate epitopes at the individual residue level, we used the AL2CO sequence conservation analysis server (http://prodata.swmed.edu/al2co/). Specifically, we utilized an alignment file generated from African representative HIV-1 clades A, C, D, and K and their recombinant sequences deposited in the LANL HIV-1 Sequence Database (https://www.hiv.lanl.gov/content/index) for calculating the conservancy scores.

HIV-1 genotyping

The participants' genomic DNA was extracted from whole blood with the PaxGene DNA blood kit (Qiagen) as previously described. A three-round nested PCR assay was performed targeting the HIV-1 proviral DNA Gag-Pol region (the third round nested PCR was to add Illumina-specific adaptor sequences).[44,45] The final PCR product was purified using the Agencourt AMPure XP magnetic beads (Beckman Coulter). The purified PCR was used for library preparation using the Nextera XT DNA Library Preparation Kit (Illumina) (indexing was done with the IDT for Illumina DNA/RNA UD Indexes Set A) according to the manufacturer's protocol. Equimolar concentrations of all samples were pooled and sequenced on an Illumina MiSeq instrument (Illumina) using the paired-end (2x300bp) method with the MiSeq-v3 reagent kit (Illumina). The read quality of the generated files was determined using FastQC, and the low-quality sequences were trimmed using Trimmomatic. The resultant reads were aligned/mapped to HIV-1 reference (RefSeq: NC_001802.1) using the BWA to generate viral contigs. HIV-1 subtyping was done using the REGA-v3 HIV-1 Subtyping Tool,[46] and any refractory sequences/samples were resolved using the RIP tool (https://www.hiv.lanl.gov /content/sequence/RIP/RIP.html) or HIV-1 blast (https://www.hiv.lanl.gov/content/sequence/ BASIC_BLAST/basic_blast.html) followed by phylogenetic analysis with PhyML (https:// www.hiv.lanl.gov/content/sequence/PHYML/interface.html).

HIV-1-specific IFN-γ and IL-2 dual ELISPOT assay

HIV-1 specific HLA-C*03:02-restricted CD8+ T cell responses were evaluated using a dual ELISPOT assay. HIV-1 peptides were synthesized using the Fmoc (fluoren-9-ylmethoxy carbonyl) means of solid-phase peptide synthesis technology and the purity was confirmed using high-pressure liquid chromatography (Bio-Synthesis, Inc). Peptides were diluted to a

final concentration of 10µg/ml. PBMCs were isolated by density gradient centrifugation from EDTA whole blood and cryopreserved. Frozen PBMCs were thawed, and viability was confirmed by trypan blue, then rested overnight before plating. We evaluated the secretion of IFN-γ and/or IL-2 by PBMC using the ELSP5710/5810 AID iSpot FluoroSpot kit (AID Autoimmun Diagnostika) according to the manufacturer's instructions. Briefly, 96-well plates pre-coated with both IFN-γ and IL-2 monoclonal antibodies were incubated with 100µl of 2 X 105 viable cells and 100µl of peptide solution per well at 37°C in humidified 5% CO2 for 40 hrs. Media alone was used as a negative control (NC), and pokeweed as a positive control. Plates were washed and stained with biotinylated anti-human IL-2 and anti-human IFN-γ FITC. IFN-γ and IL-2 production was quantified using an AID iSpot EliSpot/FluoroSpot Reader (AID Autoimmun Diagnostika) and expressed as spot-forming cells (SFC) per million PBMC after subtraction of background spots from NC.

4.3.4 Results

HIV-1 clades C and A have private and shared HLA-C*03:02-epitopes and preferentially accommodate hydrophobic residues in the distal pocket

We used the NetMHCpan 4.1 and MotifScan servers to predict the epitope repertoire of HLA-C*03:02, and a total of 42,679 and 92 peptides were predicted, respectively. Expectedly the env and pol genes contribute the largest number of epitopes (Figure 4-11A). Among the NetMHCpan-predicted epitopes, 75 and 321 were predicted to meet the strong and weak binders' threshold, respectively (Supp. Table 14). The thresholds are expressed in terms of %Rank, the percentile of the predicted binding affinity compared to the distribution of binding affinities calculated on a set of random natural peptides. A similar number (and proportion) of strong and weak binders were predicted from the HIV-1 A and C proteome, and 76 (23.5%) peptides were found to be shared among the clades (Figures 4-11B and C). We then used NetChop 3.0 to determine a final set of 238 epitopes predicted to undergo proteasomal cleavage (Figure 4-11D). These epitope sequences range from 8–13mers with a predominance of 9-mers (65%, Figure 4-11E). Further analysis of their amino acid sequence

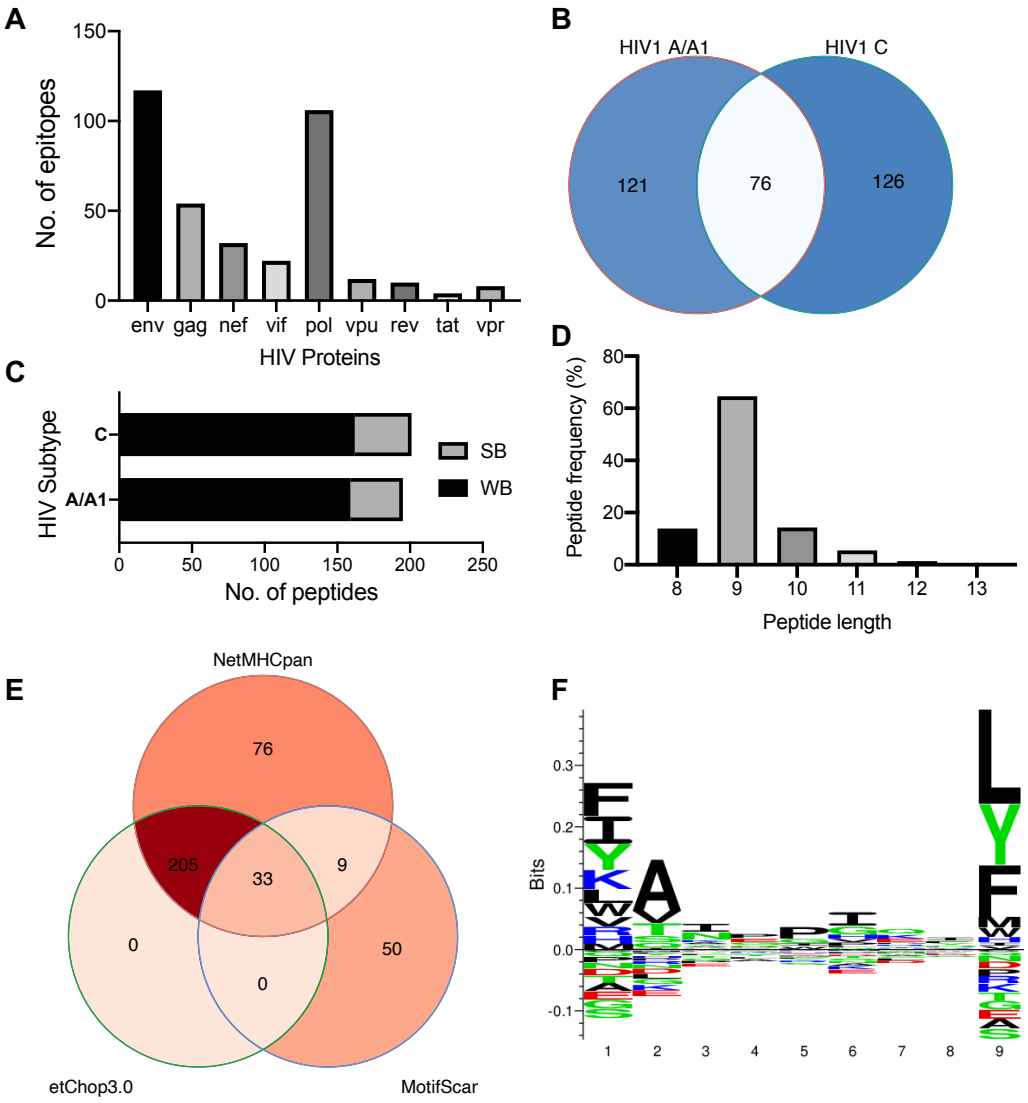


Figure 4-11: HIV-1 subtypes A/A1 and C predicted epitopes.

*(A) Distribution of predicted peptide length of HLA-C*03:02 allele in the HIV-1 A/C consensus peptidome. (B) Venn diagram showing the number of shared predicted epitopes between HIV-1 subtype A/A1 and C. (C) Proportion of HIV 1 predicted to be strong or weak binders. (D) Venn diagram showing the number of peptides predicted by NetMHCpan and Motif Scan methods. Also, the number of epitopes predicted to undergo proteasomal cleavage by TAP is shown. (E) Number of epitopes predicted by HIV protein. (E) Frequency of the length of predicted peptides.(F) Amino acid sequence logo representation of the most abundant residue at each position in the epitope. Prominent amino acid symbols indicate*

the most preferred amino acid in that epitope position. The sequence logo was calculated using clustering and pseudo counts with a weight on prior at 200, and data was handled with probability-weighted Kullback–Leibler probability distribution. Abbreviations: SB, strong binders; WB, weak binders

pattern at the HLA-C*03:02 motif using sequence logos (Figure 4-11F) found that certain amino acids are predominant or conserved at positions 1 (p1), 2 (p2), and 9 (p9, C-terminus). The P9 position of the HLA-C*03:02 motif is occupied by leucine, a large hydrophobic amino acid, but the position also accepts large hydrophobic and neutral residues phenylalanine and tyrosine, respectively (Figure 4-11F). At position p2, the small hydrophobic residue alanine is preferred, but the small hydrophilic and neutral residues valine and threonine are also accommodated. Similar to p9, position p1 equally favours large hydrophobic residues phenylalanine, isoleucine, and lysine but also accepts a large neutral residue, tyrosine (Figure 4-11F).

C*03:02-restricted stable epitopes are mainly derived from structural proteins of HIV

Next, we performed in silico docking to determine and characterize the top-ranked HLA-C*03:02 epitopes that preferentially elicit CD8+ T cell responses accounting for the putative protective effect.[11] First, we designed a 3D structural model of the HLA-C*03:02 molecule. The best template for model building was protein data bank (PDB) ID 5w6a.2 (HLA-C*06:02), with high sequence identity (94.7) and coverage, resulting in a model with high confidence scores, favorable stereochemistry, and stability, suitable for ligand binding studies (molecular docking) (Supp. Figures 4 and 5 and Supp. Table 15). According to our docking protocol, we found eight top-ranked conformations (peptides); with the best energetically favoured docking scores and extensive strong peptide-HLA (pHLA) hydrogen bonds (Figure 4-12A-D, Supp. Figure 6, Table 4-7). Four epitopes were found in structural HIV-1 proteins, including ^{71}GTEELRSLY^{79} (GY9) located on the gag gene derived from the p17 matrix protein, ^{43}GAERQGTLNF^{52} (GF10), and ^{324}AQNPEIVIY^{332} (AY9) encoded on the pol gene and ^{58}KAYETEMHN^{66} (KN9) located in the env gene derived from the gp120 protein. Other epitopes were derived from non-structural HIV-1 proteins, such as ^{84}GAFDLSFFL^{92} (GL9) and ^{114}WVYNTQGYF^{122} (WF9) from the Nef protein, while ^{128}VVSPRCEY^{135} (VY8) and ^{109}VSVESPVIL^{117} (VL9) are derived from Vif and Rev proteins respectively.

Table 4-7: Molecular docking results of top-ranked HIV epitopes docked with C*0302

HIV-1 clade	Protein and Position	Peptide Sequence		No. of H bonds	RMSD[a]	Binding energy (kcal/mol)	Convex-PL score	Amino acids involved in H bond interaction
	Structural							
C/A1	Gag^{71}*	GTEELRSLY	GY9	7	0.811	–8.6	7.23	Lys65, Tyr83, Tyr98, Thr142, Trp146 & Tyr158
A1	Env^{58}	KAYETEMHN	KN9	11	0.202	–7.4	7.36	Tyr6, Arg61, Glu62, Lys65, Tyr66, Ser76, Glu151 & Tyr158
C	Pol^{43}*	GAERQGTLNF	GF10	7	1.165	–8.7	7.22	Tyr6, Glu62, Lys65, Tyr98, Thr142, Lys145 & Gln154
C	Pol^{324}*	AQNPEIVIY	AY9	6	1.547	–7.7	7.16	Glu62, Lys65, Lys145, Tyr158 & Tyr170
	Regulatory							
C /A1	Nef^{84}	GAFDLSFFL	GL9	7	1.030	–9.1	7.24	Tyr8, Glu62, Lys65, Thr72, Tyr98, Lys145 & Tyr158
C/A1	Nef^{114}	WVYNTQGYF	WF9	7	0.873	–9.7	7.53	Tyr6, Arg68, Tyr98, Thr142, Lys145, Trp146 & Glu151
A1	Vif^{128}	VVSPRCEY	VY8	8	1.473	–8.3	7.10	Gln69, Thr72, Ser76, Tyr83, Tyr98, Thr142, Ly145 & Glu151
A1	Rev^{109}*	VSVESPVIL	VL9	7	1.638	–8.0	7.27	Tyr8, Arg61, Thr72, Asn79, Tyr83 & Tyr98

[a]RMSD Root mean square deviation (in Å) in comparison to the native ligand (PDB ID: 5w6a.2) of the C*0602 allele

*stable on molecular dynamics

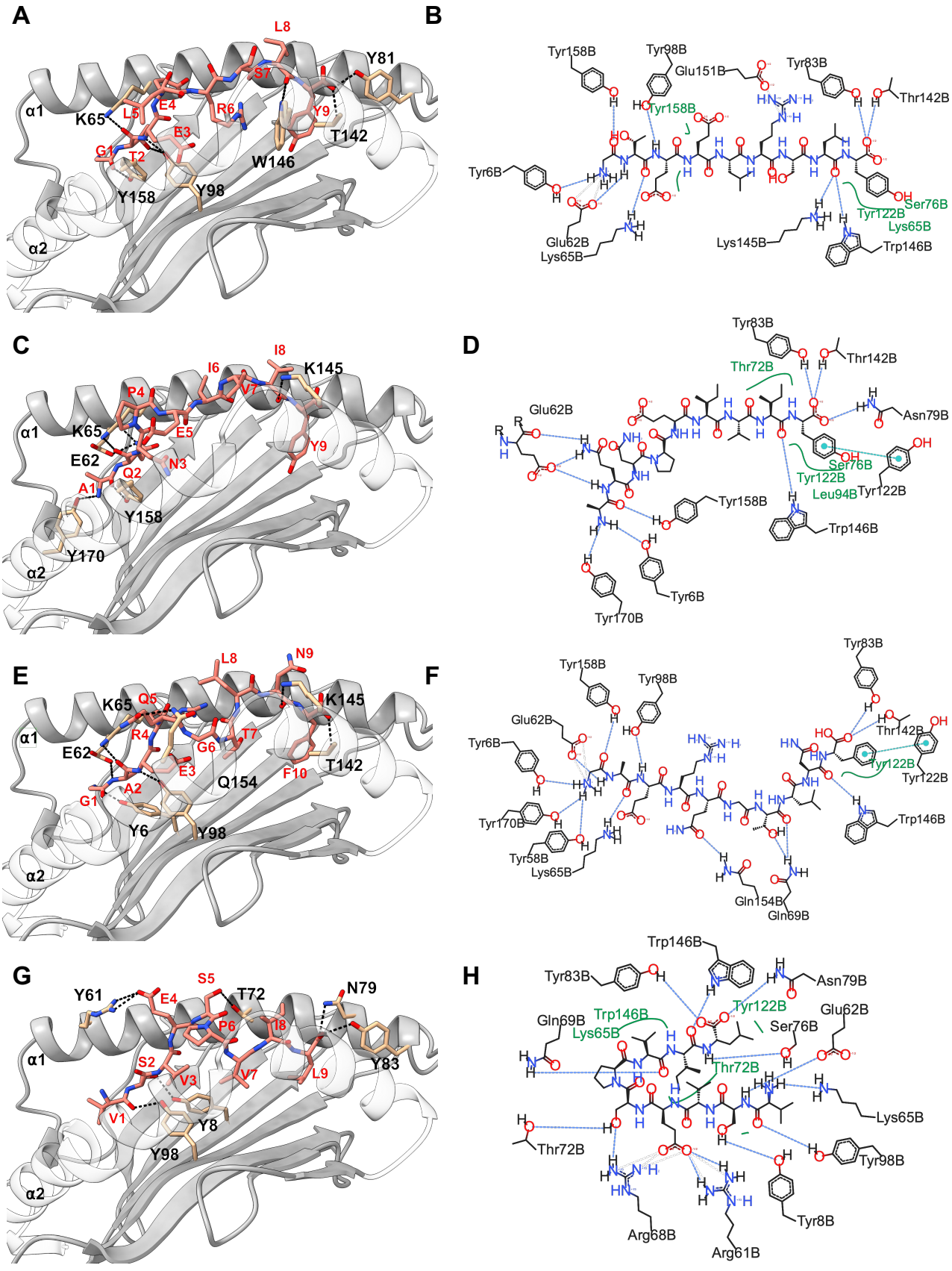

A
L8
Y81
S7
α1
K65
L5
E4
R6
Y9
E3
T142
G1
T2
W146
Y158
Y98
α2
B
Tyr158B
Tyr98B
Glu151B
Tyr83B
Thr142B
Tyr158B
Tyr6B
Ser76B
Tyr122B
Lys65B
Glu62B
Lys65B
Lys145B
Trp146B
C
I8
I6
K145
P4
V7
α1
K65
E5
Y9
E62
N3
Q2
A1
Y158
Y170
α2
D
Tyr83B
Thr142B
Thr72B
Glu62B
Asn79B
Ser76B
Tyr122B
Leu94B
Tyr122B
Tyr158B
Trp146B
Tyr6B
Tyr170B
E
L8
N9
K65
Q5
K145
α1
R4
T7
E62
G6
E3
T142
F10
Q154
A2
G1
Y6
Y98
α2
F
Tyr83B
Tyr158B
Tyr98B
Glu62B
Tyr6B
Thr142B
Tyr122B
Tyr122B
Tyr170B
Trp146B
Tyr58B
Lys65B
Gln154B
Gln69B
G
S5
Y61
E4
T72
N79
P6
I8
α1
S2
V3
V7
L9
Y83
V1
Y8
Y98
α2
H
Trp146B
Tyr83B
Asn79B
Tyr122B
Glu62B
Gln69B
Trp146B
Lys65B
Ser76B
Thr72B
Lys65B
Thr72B
Tyr98B
Tyr8B
Arg68B
Arg61B

Figure 4-12: Molecular docking analysis and representation of HLA-C*03:02 interactions with docked epitopes GY9 (A&B), AY9 (C&D), GF10 (E&F), and VL9 (G&H).

*In the left panel is a 3D depiction of the HLA-C*03:02 molecule (ribbon representation) alongside the epitope (stick representation). Dashed lines highlight the presence of hydrogen bonds. Residues within HLA-C*03:02 contributing to hydrogen bond interactions are labeled in black, and residues in the epitope are labeled in red. The α2 chain has been rendered transparent, enabling clear visualization of the epitope. In the left panel is a 2D depiction (generated using PoseView, https://proteins.plus/) of the docked epitopes, with hydrogen bonds depicted as black dashed lines and van der Waals forces shown in green. These illustrations provide an insightful view of the molecular interactions.*

To establish the structural basis of the stability of these predicted pHLA complexes, we performed an extensive conventional molecular dynamic (MD) simulation. We performed all-atom MD simulations of HLA-C*03:02 in the unbound form and on each of the eight pHLA complexes. The root means square deviation (RMSD) of protein atoms from their initial structural position over time provides an assessment of the stability of the protein-ligand complexes. We calculated and compared the average RMSD of the Cα atoms of the pHLA complexes and the unbound HLA-C*03:02 (1.05Å). Four pHLA complexes exhibited convergence, particularly evident within the final 100ns of MD, as depicted in Figure 4-13A-D. The achieved convergence is reflected in notably reduced average RMSD values: 0.66Å (GY9), 0.67Å (GF10), 0.70Å (AY9), and 0.59Å (VL9) compared to the free HLA-C*03:02 molecule (Figure 4-13A-D). The remaining pHLA complexes with KN9, GL9, WF9, and VY8 epitopes exhibited a lack of stability (Supp. Figure 6G and H). In our subsequent molecular dynamics (MD) trajectory analyses, we focused on analyzing the last 100 nanoseconds. This data indicates that the molecules GY9, GF10, AY9, and VL9 exhibit enhanced conformational stability of the HLA-C*03:02 molecule upon binding.

We then examined the formation of intermolecular interactions of the stable pHLA complexes. Hydrogen bonds play a significant role in forming and stabilizing pHLA complexes in the binding groove of HLA-C*03:02. We examined the hydrogen bond occupancy between the four epitopes and HLA-C*03:02 using the hydrogen bond module in

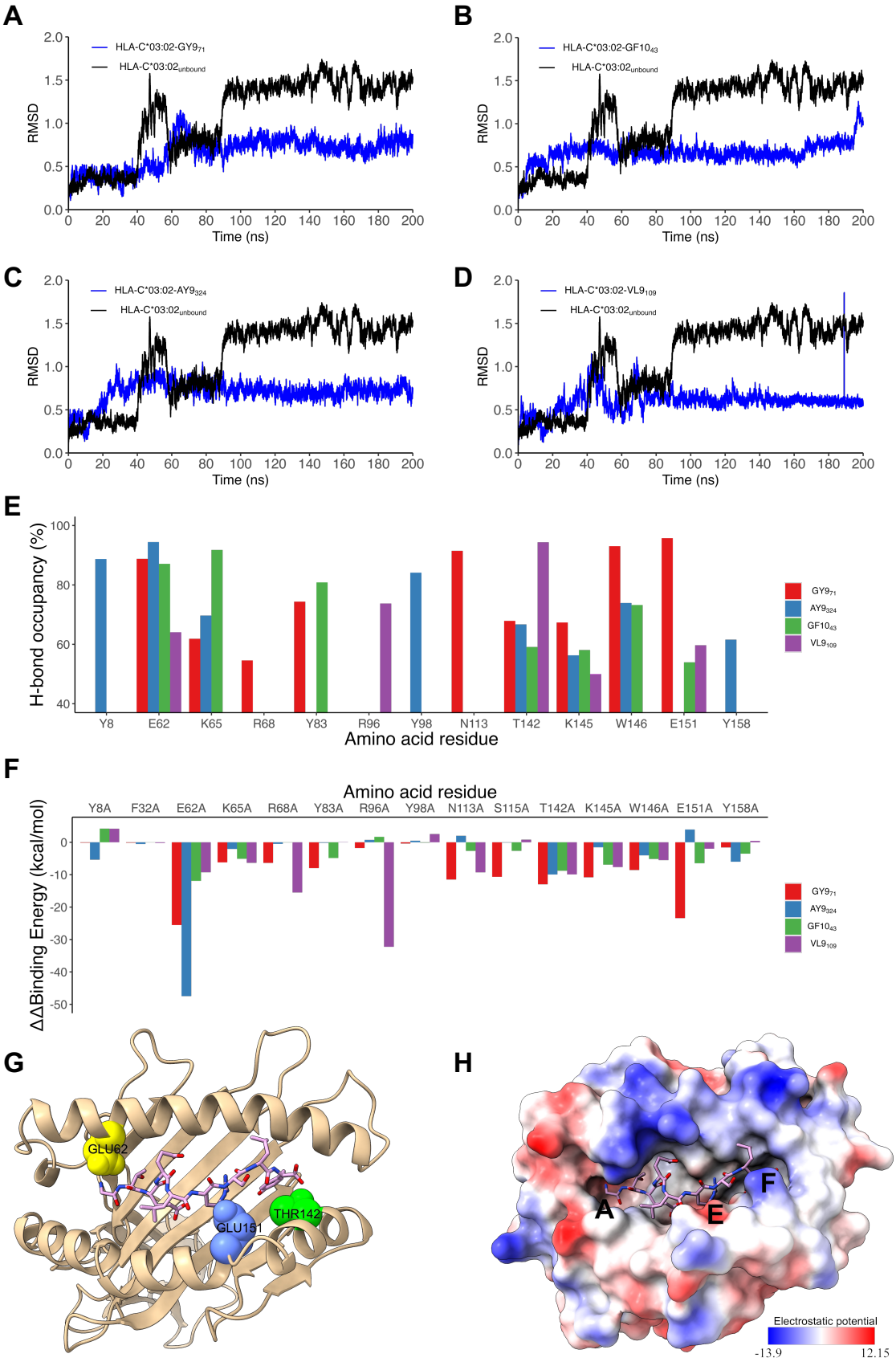

A
HLA-C*03:02-GY9$_{71}$
HLA-C*03:02$_{unbound}$
RMSD
Time (ns)
B
HLA-C*03:02-GF10$_{43}$
HLA-C*03:02$_{unbound}$
RMSD
Time (ns)
C
HLA-C*03:02-AY9$_{324}$
HLA-C*03:02$_{unbound}$
RMSD
Time (ns)
D
HLA-C*03:02-VL9$_{109}$
HLA-C*03:02$_{unbound}$
RMSD
Time (ns)
E
H-bond occupancy (%)
Y8 E62 K65 R68 Y83 R96 Y98 N113 T142 K145 W146 E151 Y158
Amino acid residue
GY9$_{71}$
AY9$_{324}$
GF10$_{43}$
VL9$_{109}$
F
Amino acid residue
Y8A F32A E62A K65A R68A Y83A R96A Y98A N113A S115A T142A K145A W146A E151A Y158A
ΔΔBinding Energy (kcal/mol)
GY9$_{71}$
AY9$_{324}$
GF10$_{43}$
VL9$_{109}$
G
GLU62
THR142
GLU151
H
A
E
F
Electrostatic potential
-13.9
12.15

Figure 4-13: Molecular dynamics simulations analysis of HLA-C*03:02/peptide complexes.

*(A) through (D) are RMSD plots of the Cα backbone of HLA-C*03:02 in complex with VL9 (A), GY9 (B), GF10 (C) and AY9 (D). (E) The percent of hydrogen-bond occupancy for interactions between HLA-C*03:02 residues (donors or acceptors) and stable epitopes (VL9, GF10, GY9, and AY9) across the final 100ns. (F) Binding free energy change of key residues involved in HLA-C*03:02 binding with epitopes. (G) 3D structure of HLA-C*03:02 displaying the key residues Glu62, Glu151, and Thr142 in pocket A (yellow), E (blue), and F (green), respectively. (H) The electrostatic surface potential of HLA-C*03:02 Electrostatic potential was calculated and visualized using ChimeraX default settings. The color scale ranges from −13 (red) to +12 (blue) kT/e*

VMD software.[47] We analyzed strong hydrogen bonds with an acceptor-donor atom distance ≤3.5Å and a hydrogen-to-donor-acceptor angle greater than 120°. We observed that residues Glu62, Thr142, and Lys170 in HLA-C*03:02 were involved in hydrogen bond formation with all four epitopes with more than 50% occupancy (Figure 4-12 and Figure 4-13E, Supp. Table 16). More than 50% hydrogen bond occupancy existed between Lys90, Trp171, and Glu151 in HLA-C*03:02 with at least three of the four epitopes (Figure 4-13E, Supp. Table 16). These results suggest that the HLA-C*03:02 binding groove favorably and stably binds three epitopes derived from structural HIV-1 proteins.

Positions 62, 142, and 151 in HLA-C*03:02 and P6 in the GY9 epitope provide the structural basis for the preferential binding of GY9

The binding free energy (ΔG) of pHLA complexes determines the stability of complex formation. Therefore, we applied the molecular mechanic/Poisson-Boltzmann surface area (MM/PBSA) method to estimate the binding free energies of GY9, AY9, GF1,0, and VL9 complexation with HLA-C*03:02. Generally, a more negative magnitude of the binding free energy corresponds to strong (high) binding affinities of pHLA complexes. Among the epitopes, GY9 showed a much stronger binding free energy of -88.41 kcal/mol, indicating a strong and favourable binding affinity to HLA-C*03:02 (Table 4-8). Notably, the van der Waals and electrostatic energies of -53.01 kcal/mol and -547.44 kcal/mol, respectively, between GY9 and HLA-C*03:02 contribute significantly to the binding (Table 4-8). We found that the electrostatic contribution of the GY9 epitope is much higher compared to other

Table 4-8: Binding free energies obtained by the MM/PBSA method of pHLA complexes with GY9, GF10, AY9, and VL9 peptides

HIV-1 subtype	Peptide Sequence		Energy Components (kcal/mol)				
			van der Waals	Electrostatics	Polar solvation	ΔG Binding energy	
	Structural						
C/A1	GTEELRSLY	GY9	-53.01	-547.44	521.66	-88.41	
C	GAERQGTLNF	GF10	-53.45	-335.96	347.44	-50.59	
C	AQNPEIVIY	AY9	-68.75	-258.85	285.79	-51.24	
	Regulatory						
A1	VSVESPVIL	VL9	-67.29	-367.42	394.36	-49.73	

epitopes. Given the significant contribution of hydrogen bonding formation to the electrostatic energy, this means that hydrogen bonds are likely to play a critical role in GY9 binding to HLA-C*03:02. Also, the van der Waals energy is an indicator of the compactness of a ligand in the receptor binding groove; we found that GY9 also had the strongest value (-53.01 kcal/mol), suggesting a more favourable packing arrangement of GY9 in the HLA-C*03:02 binding groove (Table 4-8). This compactness is essential in pHLA complexes and affects efficient T cell receptor (TCR) engagement.[48,49] Overall, these findings provide valuable insights into the importance of van der Waals interactions, electrostatic interactions, and hydrogen bonding in the binding dynamics of HIV-1 epitopes to HLA-C*03:02 as well as the preference for GY9.

To gain insight into the individual contributions of the amino acids within the HLA-C*03:02 binding groove to the binding free energy, we performed a computational alanine scanning (CAS, or mutagenesis) based on the MM/PBSA method.[43] A negative value of ΔΔG indicates a favourable contribution for the wild-type residue in that position and vice versa. We mutated 35 amino acid residues within 5Å of the epitopes to alanine and computed the binding free energy difference between wild-type and mutant pHLA complexes. Notably, mutants E62A, T142A, and E151A in HLA-C*03:02 resulted in a significant loss of binding free energy with GY9 (Figure 4-13F, Supp. Table 5). For position 62, the mutation to alanine (E62A) resulted

in a loss of binding free energy ranging from -9.25 kcal/mol to -47.47 kcal/mol across different peptide ligands (GY9, GF10, AY9, VL9). Similarly, for position 142, alanine mutation (T142A) led to a decrease in binding free energy ranging from -9.91 kcal/mol to -9.94 kcal/mol. Likewise, for position 151, alanine mutation (E151A) resulted in reduced binding free energy ranging from -1.98 kcal/mol to -23.40 kcal/mol, except for AY9, where a positive change in binding free energy of 3.93 kcal/mol was observed. Consequently, these three positions, E62, T142, and E151, found in the A, E, and F pockets of the HLA-C*03:02 peptide-binding groove (Figure 4-13G), are predicted to confer epitope specificity.

Previous studies have shown that point mutations within epitopes significantly diminish or abrogate immune responses.[49] We performed CAS on the GY9 epitope to establish the most influential positions to the binding affinity. Surprisingly, we noted a consistent trend in a decrease in the binding free energy (ΔΔG) across all the amino acid positions in GY9 except p1 with a small glycine residue (similar to alanine). However, p6 demonstrated a significant negative loss in binding free energy (ΔΔG -27.00 kcal/mol; Table 4-9), further reinforcing the importance of Arg6 at this position for binding affinity. We observed that Arg at position p6 in GY9 leads to forming of three hydrogen bonds (donor) with residues E176, W171, and N138 (Supp. Table 16). Remarkably, these hydrogen bonds exhibit a high occupancy of over 90% throughout the MD simulation, indicating their persistent and stable nature. This suggests that stabilizing p6 is vital to prevent protrusion of the epitope out of the peptide binding groove that would considerably alter the structure of the pHLA-TCR binding platform. We computed the conservancy score by aligning viral sequences from all publicly available HIV-1 subtypes A, C, D, and K and their recombinants. We also found that p3, p6, and p9 had the lowest conservancy score (Table 4-9). These results suggest that p6 contributes favorably to GY9 binding and may serve as the primary anchor residue, while positions p3 and p9 are secondary anchor residues refining epitope binding.

The GY9 epitope elicits a clade-specific HLA-C*03:02-restricted IFN-γ response

To discern the immunogenic potential of GY9 ex vivo, we assessed GY9-specific CD8+ T cells, employing a dual color enzyme-linked immunospot (ELISPOT) assay to measure the production of IFN-γ and IL-2. IFN-γ production indicates an active immune response, reflecting ongoing T cell effector functions. On the other hand, IL-2 secreted by activated T

cells or NK cells plays a crucial role in driving the proliferation and differentiation of naive T cells, B cells, and NK cells, facilitating their transition into effector (such as T_h1) and memory cells, and promoting the release of secondary cytokines. We used peripheral blood mononuclear cells (PBMC) from a study population that included 25 subjects (age range 2.3 – 20.8 years, Supp. Table 18) on antiretroviral therapy (ART) recruited from Uganda, of whom 13 were expressing the HLA-C*03:02 allele (HLA typing is described elsewhere).[11]

Table 4-9: Change in binding free energy and conservancy scores of the GY9 peptide

HIV1 Clade[a]	Amino Acid Residue and Position								
	G1	T2	E3	E4	L5	R6	S7	L8	Y9
A1	.	.	.	.	.	R/K	.	.	Y/F
C	.	.	.	.	.	K	.	.	Y/F/H
D	.	.	.	.	I	K	.	.	Y/F
K	.	.	.	.	I	K	.	.	Y/F
ΔΔG GY9[b]	NA	-8.18	-10.3	-4.82	-2.83	-27.00	-5.85	-2.82	-7.33
Conservancy score[c]	5	4	2	3	3	2	5	5	2

[a]Common clades in Botswana and Uganda populations; NA, glycine is of similar size to alanine

[b]Binding free energy change due to mutation of amino acid residue to alanine

[c]The scores are from one to nine to show the conservation level (low to high, respectively)

GY9-specific IFN-γ production ranging from 65 to 940 SFU/million PBMC, was found in 3/10 (30%) HLA-C*03:02^{+ve} subjects. Still, no response was found among any HLA-C*03:02^{-ve} subjects (Figure 4-14, Supp. Table 18). All GY9 responders were coincidentally infected with HIV-1 clade A1 (2/3) or C (1/3) (Figure 4-14, Supp. Table 18). It should be noted that GY9 originated from both the A1 and C clade consensus sequences (https://www.hiv.lanl.gov/content/index). Except for three individuals for whom HIV-1 could not be typed, all non-responders to GY9 were found to be infected with HIV-1 clades A1, C, D, or A1D recombinant strains (Supp. Table 18). Unsurprisingly, in our cohort on chronic ART (1–121 months), we detected no IL-2 production in HLA-C*03:02^{+ve} or HLA-C*03:02^{-ve} individuals (Supp. Table 18).50 We have already demonstrated above that some positions with the GY9 epitope are under selective pressure (CAS and conservancy scores). We think the lack of IL-2 response in HLA-C*03:02^{+ve} subjects infected with A1 may suggest the presence of escape mutants, especially in positions p6 > p3 > p9. These findings indicate that GY9 elicits a clade-specific immune response and exhibits non-promiscuity for HLA types.

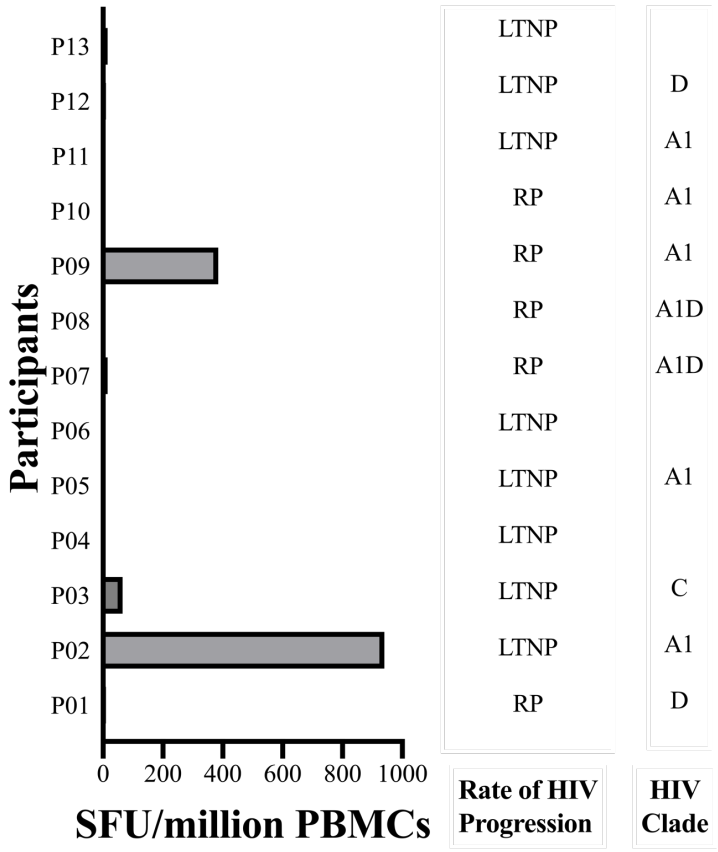


Figure 4-14: GY9-specific IFN-γ CD8^{+} T cell responses among HLA-C*03:02^{+ve} participants

*The magnitude of IFN-γ CD8^{+} T cell responses against GY9 epitope were measured among HLA-C*03:02^{+ve} and HLA-C*03:02^{-ve} (Supp. Table 18) participants by ELISPOT assay. LTNP, long-term nonprogresssion; RP, rapid progression*

4.3.5 Discussion

Several approaches are being investigated to develop novel HIV-1 vaccines. Among these approaches is the search for multi-epitope vaccine candidates that elicit effective quantitative and qualitative humoral and cellular immune responses.[26,51] Theoretically, T cell-based vaccines, utilizing peptides identified through in silico predictions, hold promise as effective vaccination strategies, particularly when focused on pinpointing the most immunogenic antigens.[26] In this study, we have utilized a synergy of computational techniques and empirical functional validation to uncover a previously unrecognized HIV-1 epitope, GY9. This epitope

displays a distinctive potential for presentation by the HLA-C*03:02 allele, associated with effective HIV-1 control among African pediatric populations. Consistent with previous reports, the role of the HIV-1 Gag protein is prominent in providing the most immunogenic peptides presented by class I HLA molecules. Additionally, we report three epitopes in the Env and Pol proteins that map to HIV-1 subtypes A and C, suggesting that some control of HIV-1 may be attributable to HLA-C*03:02.

The HIV-1 Gag protein is preferred for T cell vaccine candidates because it is highly immunogenic and conserved across HIV-1 clades.[16,52] Several T cell candidate vaccines have so far shown variable immunogenicity;[52] however, thus far, this GY9 epitope has not been reported to show immunogenicity or restriction to HLA-C*03:02 and therefore has not been considered a potential vaccine candidate.[52,53] This could be attributed to the lack of prioritization of HLA-C preferential antigens. The role of HLA-C class I molecules in delaying HIV-1 progression has been historically considered less significant, primarily attributed to their lower surface expression levels. This phenomenon might arise from the underappreciation of HLA-C-preferred antigens. The historical perception regarding the impact of HLA-C class I molecules on the control of HIV progression has been relatively suppressed, often linked to their comparatively lower levels of surface expression and high LD with HLA-A and B alleles. Consequently, their contribution to HIV-1 control has not been prominently emphasized. Notably, to date, less than 10% (22/280) of optimal HIV-1 CTL epitopes ("A list") defined in the LANL HIV-1 epitope database are HLA-C-restricted epitopes. Evidence is progressively accumulating, highlighting the potential contribution of HLA-C molecules to control HIV-1, particularly when focusing on the conserved Gag epitopes.[30,54,55] In our study, we observed a variable magnitude of IFN-γ responses and no detectable IL-2 response upon stimulation of T cells from people living with HIV with the GY9 epitope, consistent with findings reported in previous studies.[55,56] This variable magnitude of cytokine (IFN-γ and IL-2) response could be attributed to several factors, such as immune exhaustion due to chronic infection and ART (IL-2 ablation and low IFN-γ production) among this cohort and viral escape within the GY9 epitope (no IFN-γ responses).[55,57–59] Indeed, we measured the magnitude of response using study participants on highly active ART (HAART; average duration 32 months [1–121 months]).[59] A striking absence of both IFN-γ and IL-2 responses were observed in a larger number of participants with the HLA-C*03:02 allele; this aligns with the likelihood of amino

acid modifications/mutations within the GY9 epitope sequence. Indeed, when we calculated conservancy scores, p6 and p3/p9 had a very low score (2), which means that these positions are associated with a high rate of mutations/variation across the various HIV-1 clades A, C, D, and K. Similarly, our CAS studies detected significant differences in the pHLA relative binding free energy when residues in p6 and p3/p9 are mutated to alanine, suggesting a very high contribution to epitope binding. A recent report by Li et al. suggests that mutations within the epitope significantly impact pHLA binding due to conformational changes and eventually affect TCR recognition and antigen presentation.[49] Collectively, our data strongly indicate that p6 and p3/p9 within the GY9 epitope serve as primary and secondary anchor residues, respectively, crucial for robust binding within the HLA-C*03:02 antigen-binding cleft, thereby facilitating optimal T cell receptor (TCR) engagement. Indeed, Joglekar et al. demonstrated that peptide-MHC binding is essential for TCR binding and that peptide mutations play an important role in viral escape.[60] Therefore, we argue that these findings show a potential viral escape and immune evasion pathway within the GY9 epitope.[49] The absence of detectable IL-2 responses to the GY9 epitope underscores the impaired capacity to reactivate HIV-specific memory T cells elicited during chronic infection, indicating a compromised immune response.[50] This observation aligns with the known phenomenon that HIV-1 infection leads to an expansion of $CD8^{+}CD28^{-}$ T cells, characterized by their compromised ability to produce IL-2.[61] Our data shows that residues E62, T142 and E151 in the HLA-C*03:02 binding groove, along with positions p3, p6 and p9 on the GY9 epitope, are critical hot spots for binding. These residues play a crucial role in shaping and stabilizing the protein-protein interface, significantly contributing to its stability. These compelling results strongly indicate the prominent influence of the HLA-C*03:02-restricted GY9 epitope sequence in shaping HIV control.

Immunogenicity in HIV-1 is not restricted to the Gag protein since numerous studies have established the role of epitopes derived from other HIV-1 proteins.[30] Indeed, HIV-1 vaccine candidate studies have demonstrated an advantage of multi-epitope prototypes.[52] In this study, our immunoinformatic approach identified three potentially immunogenic epitopes, GF10/AY9 and VL9, derived from the Pol and Rev proteins, respectively; however, we did not find any detectable HIV-1-specific CD8+ T cell responses against these epitopes in our population. While the lack of responses could be explained by similar factors noted above, the epitopes GF10 and AY9 are derived from the HIV1-C subtype; all our HLA-C*03:02^{+ve}

participants used for the dual IFN-γ/IL-2 ELISPOT assay were infected with HIV1-A1, C, D and the A1D recombinant. When we performed a conservancy score, we found that many positions along the epitopes had a very low conservancy score (VL9>GF10>AY9, data not shown) that could explain these positions as escape mutations that abrogate responses to epitopes derived from other HIV-1 clades and the potential unsuitability of these epitopes.[49] Overall, our docking results are similar to an experimental biological study where only 6-8 HIV-1 derived peptides were identified as restricted to HLA class I alleles.[62] In that study, Ziegler et al. infected CD4+ T cells with HIV-1 and measured HLA class I (HLA-A*02:01/*02:01, B*27:05/*40:01, C*02:02/*03:04) repertoire, suggesting that these molecules present a small set of epitopes derived from the HIV-1 proteome at variable relative quantities.[62]

While previous research on HIV-1 vaccine candidates has predominantly focused on the protective HLA-B alleles, it is noteworthy that HIV-nef attachment selectively downregulates the cell surface expression of both HLA-A and B molecules.[63,64] This downregulation phenomenon facilitates immune evasion through CTL escape by virally infected cells. Consequently, HLA-C-restricted CTL responses remain intact to facilitate the recognition and destruction of HIV-infected cells. Most crucially, the HLA-C*03:02 cytoplasmic tail lacks both tyrosine and aspartate, which are the targets of Nef-dependent downregulation of HLA cellular surface expression. Instead, HLA-C*03:02 has Leu321 and Val328 in the cytoplasmic tail.[63] Therefore, compensatory mechanisms enhance HLA-C cell surface expression, favorably explaining the role of HLA-C-restricted CTL responses in HIV-1 control.[65] Furthermore, HLA-C alleles lacking a binding site for microRNA-148a in the 3′ untranslated region of their messenger RNA exhibit a compensated high surface expression, potentially influencing immune recognition and responsiveness.[66] Interestingly, the HLA-C*03:02 allele demonstrates strong linkage disequilibrium with a C variant located 35kb upstream of the HLA-C gene. The presence of the -35C allele is strongly associated with increased cell surface expression of HLA-C molecules, potentially providing a mechanistic explanation for the observed impact of HLA-C*03:02 on HIV-1 control.[65,67] In this study, we find that GY9-induced IFN-γ responses were not shared with other HLA-C, -A, or -B alleles (Supp. Table 18); this would suggest that clade-specific GY9 HLA-C*03:02-restricted responses are highly allele-specific. This restricted binding specificity of the GY9 epitope is predicted to play a crucial role in

determining immune responses following HIV-1 infection and may have implications for vaccine design and understanding the individual variation in immune recognition.

In conclusion, we have used an immunoinformatics approach to identify an HLA-C*03:02-restricted epitope, eliciting T cell-specific responses, suggesting that the GY9 epitope plays a significant role in HLA-C*03:02-mediated HIV-1 control among children. This study supports the hypothesis that an effective HIV-1 vaccine should be clade-specific; therefore, efforts for a global vaccine may not be feasible. And as such, more focus should be placed on identifying possible epitopes mapped among all clades, especially those restricted to protective HLA-C alleles. Finally, our study expands upon prior studies by providing evidence supporting the notion that the HIV-1 matrix protein p17 represents a promising epitope candidate for developing a vaccine against HIV/AIDS.

Limitations of the study

Our study has some notable limitations. Full or partial viral sequencing and CD4+ T cell counts were not performed on all participants in our study. Viral sequencing could have yielded valuable insights into the degree of conservancy within the GY9 sequence among individuals positive or negative for the HLA-C*03:02 allele, enabling empirical assessment of epitope dominance in HIV-infected individuals. Our use of consensus and primary strain sequences for epitope prediction may potentially overlook naturally occurring epitopes in the studied population. A comprehensive analysis of primary strain sequences is crucial to identify conserved epitopes capable of eliciting robust and broad immune responses. Nonetheless, a recent study by Bugembe et al. demonstrated that the same computational tools used here identify 95% of experimentally mapped HIV-1 clade A and D epitopes.[31] Furthermore, measurements of CD4+ T cell levels would have provided a baseline assessment of immunocompetence, a factor known to influence immune responses to HIV-1 epitopes.68 In the future, investigations employing well-characterized study populations, incorporating advanced immunopeptidomics techniques, intracellular cytokine flow cytometry, and tetramer staining assays will be essential to build upon our current findings and overcome the methodological limitations observed in our study.[30,69] These approaches hold the potential to deepen our understanding of the immunological responses and contribute valuable insights to the field of immunology.

AUTHOR CONTRIBUTIONS

SK designed the study, performed the research, and analyzed the data; HIV sequencing was done by FY, FK, and EK. SK, JK, ME, and GN designed and performed the ELISPOT; DR provided support for the modeling experiments together with computational resources; The manuscript was drafted by SK with additional inputs from all the authors, including SM, GR, GPK, EK, MA, MW, FY, BCM, LW, JKL, HS, MJ, SWM, DR, DK, GM, AK, MM, and NAH.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY STATEMENT

We express our support for efforts for inclusive, diverse, and equitable conduct of research.

4.3.6 References

1. McLaren, P.J., and Carrington, M. (2015). The impact of host genetic variation on infection with HIV-1. Nat. Immunol. 10.1038/ni.3147.

2. Cohen, O.J., Paolucci, S., Bende, S.M., Daucher, M., Moriuchi, H., Moriuchi, M., Cicala, C., Davey, R.T., Baird, B., and Fauci, A.S. (1998). CXCR4 and CCR5 Genetic Polymorphisms in Long-Term Nonprogressive Human Immunodeficiency Virus Infection: Lack of Association with Mutations other than CCR5-Δ32. J. Virol. *72*, 6215–6217. 10.1128/jvi.72.7.6215-6217.1998.

3. Walker, B.D., Pereyra, F., Jia, X., McLaren, P.J., Telenti, A., Bakker, P.I.W. de, De Bakker, P.I.W., and Walker, B.D. (2010). The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. Science (80-.). *330*, 1551–1557. 10.1126/science.1195271.

4. Boudreau, J.E., and Hsu, K.C. (2018). Natural killer cell education in human health and disease. Curr. Opin. Immunol. *50*, 102–111. 10.1016/j.coi.2017.11.003.

5. Kaseke, C., Tano-Menka, R., Senjobe, F., and Gaiha, G.D. (2021). The Emerging Role for CTL Epitope Specificity in HIV Cure Efforts. J. Infect. Dis. *223*, S32–S37. 10.1093/infdis/jiaa333.

6. Pymm, P., Tenzer, S., Wee, E., Weimershaus, M., Burgevin, A., Kollnberger, S., Gerstoft, J., Josephs, T.M., Ladell, K., McLaren, J.E., et al. (2022). Epitope length variants balance protective immune responses and viral escape in HIV-1 infection. Cell Rep. *38*. 10.1016/j.celrep.2022.110449.

7. Kutsch, O., Vey, T., Kerkau, T., Hünig, T., and Schimpl, A. (2002). HIV type 1 abrogates TAP-mediated transport of antigenic peptides presented by MHC class I. AIDS Res. Hum. Retroviruses *18*, 1319–1325. 10.1089/088922202320886361.

8. Li, L., and Bouvier, M. (2004). Structures of HLA-A*1101 Complexed with Immunodominant Nonamer and Decamer HIV-1 Epitopes Clearly Reveal the Presence of a Middle, Secondary Anchor Residue. J. Immunol. *172*, 6175–6184. 10.4049/jimmunol.172.10.6175.

9. Carlson, J.M., Brumme, C.J., Martin, E., Listgarten, J., Brockman, M.A., Le, A.Q., Chui, C.K.S., Cotton, L.A., Knapp, D.J.H.F., Riddler, S.A., et al. (2012). Correlates of Protective Cellular Immunity Revealed by Analysis of Population-Level Immune Escape Pathways in HIV-1. J. Virol. *86*, 13202–13216. 10.1128/jvi.01998-12.

10. Stewart-Jones, G.B.E., Gillespie, G., Overton, I.M., Kaul, R., Roche, P., McMichael, A.J., Rowland-Jones, S., and Jones, E.Y. (2005). Structures of Three HIV-1 HLA-B*5703-Peptide Complexes and Identification of Related HLAs Potentially Associated with Long-Term Nonprogression. J. Immunol. *175*, 2459–2468. 10.4049/jimmunol.175.4.2459.

11. Kyobe, S., Mwesigwa, S., Kisitu, G.P., Farirai, J., Katagirya, E., Mirembe, A.N., Ketumile, L., Wayengera, M., Katabazi, F.A., Kigozi, E., et al. (2021). Exome Sequencing Reveals a Putative Role for HLA-C*03:02 in Control of HIV-1 in African Pediatric Populations. Front. Genet. *12*, 1–14. 10.3389/fgene.2021.720213.

12. Leslie, A., Matthews, P.C., Listgarten, J., Carlson, J.M., Kadie, C., Ndung'u, T., Brander, C., Coovadia, H., Walker, B.D., Heckerman, D., et al. (2010). Additive Contribution of HLA Class I Alleles in the Immune Control of HIV-1 Infection. J. Virol. *84*, 9879–9888. 10.1128/jvi.00320-10.

13. Gijsbers, E.F., Anton Feenstra, K., Van Nuenen, A.C., Navis, M., Heringa, J., Schuitemaker, H., and Kootstra, N.A. (2013). HIV-1 replication fitness of HLA-B*57/58:01 CTL Escape variants is restored by the accumulation of compensatory mutations in Gag. PLoS One *8*, e81235. 10.1371/journal.pone.0081235.

14. Borghans, J.A.M., Mølgaard, A., de Boer, R.J., and Keşmir, C. (2007). HLA Alleles Associated with Slow Progression to AIDS Truly Prefer to Present HIV-1 p24. PLoS One *2*, e920. 10.1371/journal.pone.0000920.

15. Chen, H., Piechocka-Trocha, A., Miura, T., Brockman, M.A., Julg, B.D., Baker, B.M., Rothchild, A.C., Block, B.L., Schneidewind, A., Koibuchi, T., et al. (2009). Differential Neutralization of Human Immunodeficiency Virus (HIV) Replication in Autologous CD4 T Cells by HIV-Specific Cytotoxic T Lymphocytes. J. Virol. *83*, 3138–3149. 10.1128/jvi.02073-08.

16. Murakoshi, H., Zou, C., Kuse, N., Akahoshi, T., Chikata, T., Gatanaga, H., Oka, S., Hanke, T., and Takiguchi, M. (2018). CD8+ T cells specific for conserved, cross-reactive Gag epitopes with strong ability to suppress HIV-1 replication. Retrovirology *15*, 1–14. 10.1186/s12977-018-0429-y.

17. Adland, E., Carlson, J.M., Paioni, P., Kløverpris, H., Shapiro, R., Ogwu, A., Riddell, L., Luzzi, G., Chen, F., Balachandran, T., et al. (2013). Nef-specific CD8+ T cell responses contribute to HIV-1 immune control. PLoS One *8*, 1–10. 10.1371/journal.pone.0073117.

18. Altfeld, M., Addo, M.M., Eldridge, R.L., Yu, X.G., Thomas, S., Khatri, A., Strick, D., Phillips, M.N., Cohen, G.B., Islam, S.A., et al. (2001). Vpr Is Preferentially Targeted by CTL During HIV-1 Infection. J. Immunol. *167*, 2743–2752. 10.4049/jimmunol.167.5.2743.

19. Ferrando-Martínez, S., Casazza, J.P., Leal, M., Machmach, K., Muñoz-Fernández, M.Á., Viciana, P., Koup, R.A., and Ruiz-Mateos, E. (2012). Differential Gag-Specific Polyfunctional T Cell Maturation Patterns in HIV-1 Elite Controllers. J. Virol. *86*, 3667–3674. 10.1128/jvi.07034-11.

20. Falivene, J., Ghiglione, Y., Laufer, N., Socías, M.E., Holgado, M.P., Ruiz, M.J., Maeto, C., Figueroa, M.I., Giavedoni, L.D., Cahn, P., et al. (2015). Th17 and Th17/Treg ratio at early HIV infection associate with protective HIV-specific CD8+ T-cell responses and disease progression. Sci. Rep. *5*, 11511. 10.1038/srep11511.

21. Boulet, S., Song, R., Kamya, P., Bruneau, J., Shoukry, N.H., Tsoukas, C.M., and Bernard, N.F. (2010). HIV Protective KIR3DL1 and HLA-B Genotypes Influence NK Cell Function Following Stimulation with HLA-Devoid Cells. J. Immunol. *184*, 2057–2064. 10.4049/jimmunol.0902621.

22. Zappacosta, F., Borrego, F., Brooks, A.G., Parker, K.C., and Coligan, J.E. (1997). Peptides isolated from HLA-Cw*0304 confer different degrees of protection from natural killer cell-mediated lysis. Proc. Natl. Acad. Sci. *94*, 6313–6318. 10.1073/pnas.94.12.6313.

23. Hraber, P., Seaman, M.S., Bailer, R.T., Mascola, J.R., Montefiori, D.C., and Korber, B.T. (2014). Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. Aids *28*, 163–169. 10.1097/QAD.0000000000000106.

24. Payne, R.P., Branch, S., Kloverpris, H., Matthews, P.C., Koofhethile, C.K., Strong, T., Adland, E., Leitman, E., Frater, J., Ndung'u, T., et al. (2014). Differential Escape Patterns within the Dominant HLA-B*57:03-Restricted HIV Gag Epitope Reflect Distinct Clade-Specific Functional Constraints. J. Virol. *88*, 4668–4678. 10.1128/jvi.03303-13.

25. Brackenridge, S., Evans, E.J., Toebes, M., Goonetilleke, N., Liu, M.K.P., di Gleria, K., Schumacher, T.N., Davis, S.J., McMichael, A.J., and Gillespie, G.M. (2011). An Early HIV Mutation within an HLA-B*57-Restricted T Cell Epitope Abrogates Binding to the Killer Inhibitory Receptor 3DL1. J. Virol. *85*, 5415–5422. 10.1128/jvi.00238-11.

26. Yang, Y., Sun, W., Guo, J., Zhao, G., Sun, S., Yu, H., Guo, Y., Li, J., Jin, X., Du, L., et al. (2015). In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. Hum. Vaccines Immunother. *11*, 795–805. 10.1080/21645515.2015.1012017.

27. Haynes, B.F., and McElrath, M.J. (2013). Progress in HIV-1 vaccine development. Curr. Opin. HIV AIDS *8*, 326–332. 10.1097/COH.0b013e328361d178.

28. Song, H., Giorgi, E.E., Ganusov, V. V., Cai, F., Athreya, G., Yoon, H., Carja, O., Hora, B., Hraber, P., Romero-Severson, E., et al. (2018). Tracking HIV-1 recombination to resolve its contribution to HIV-1 evolution in natural infection. Nat. Commun. *9*, 1928. 10.1038/s41467-018-04217-5.

29. Hraber, P., Korber, B.T., Lapedes, A.S., Bailer, R.T., Seaman, M.S., Gao, H., Greene, K.M., McCutchan, F., Williamson, C., Kim, J.H., et al. (2014). Impact of Clade, Geography, and Age of the Epidemic on HIV-1 Neutralization by Antibodies. J. Virol. *88*, 12623–12643. 10.1128/jvi.01705-14.

30. Chikata, T., Paes, W., Akahoshi, T., Partridge, T., Murakoshi, H., Gatanaga, H., Ternette, N., Oka, S., Borrow, P., and Takiguchi, M. (2019). Identification of Immunodominant HIV-1 Epitopes Presented by HLA-C*12:02, a Protective Allele, Using an Immunopeptidomics Approach. J. Virol. *93*. 10.1128/JVI.00634-19.

31. Bugembe, D.L., Ekii, A.O., Ndembi, N., Serwanga, J., Kaleebu, P., and Pala, P. (2020). Computational MHC-I epitope predictor identifies 95% of experimentally mapped HIV-1 clade A and D epitopes in a Ugandan cohort. BMC Infect. Dis. *20*, 172. 10.1186/s12879-020-4876-4.

32. Retshabile, G., Mlotshwa, B.C., Williams, L., Mwesigwa, S., Mboowa, G., Huang, Z., Rustagi, N., Swaminathan, S., Katagirya, E., Kyobe, S., et al. (2018). Whole-Exome Sequencing Reveals Uncaptured Variation and Distinct Ancestry in the Southern African Population of Botswana. Am. J. Hum. Genet. *102*, 731–743. 10.1016/j.ajhg.2018.03.010.

33. Bbosa, N., Kaleebu, P., and Ssemwanga, D. (2019). HIV subtype diversity worldwide. Curr. Opin. HIV AIDS *14*, 153–160. 10.1097/COH.0000000000000534.

34. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. 10.1002/jcc.20084.

35. Antunes, D.A., Moll, M., Devaurs, D., Jackson, K.R., Lizée, G., and Kavraki, L.E. (2017). DINC 2.0: A new protein-peptide docking webserver using an incremental approach. Cancer Res. *77*, e55–e57. 10.1158/0008-5472.CAN-17-0511.

36. El-Hachem, N., Haibe-Kains, B., Khalil, A., Kobeissy, F.H., and Nemer, G. (2017). AutoDock and AutoDockTools for Protein-Ligand Docking: Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1(BACE1) as a Case Study. In Methods in Molecular Biology, pp. 391–403. 10.1007/978-1-4939-6952-4_20.

37. Kadukova, M., and Grudinin, S. (2017). Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. J. Comput. Aided. Mol. Des. *31*, 943–958. 10.1007/s10822-017-0068-8.

38. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. 10.1002/pro.3943.

39. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25. 10.1016/j.softx.2015.06.001.

40. Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., and MacKerell, A.D. (2012). Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. J. Chem. Theory Comput. *8*, 3257–3273. 10.1021/ct300400x.

41. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., et al. (2009). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem., NA-NA. 10.1002/jcc.21367.

42. Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., and Hutchison, G.R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminform. *4*, 17. 10.1186/1758-2946-4-17.

43. Valdés-Tresanco, M.S., Valdés-Tresanco, M.E., Valiente, P.A., and Moreno, E. (2021). gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. J. Chem. Theory Comput. *17*, 6281–6291. https://doi.org/10.1021/acs.jctc.1c00645.

44. Lapointe, H.R., Dong, W., Lee, G.Q., Bangsberg, D.R., Martin, J.N., Mocello, A.R., Boum, Y., Karakas, A., Kirkby, D., Poon, A.F.Y., et al. (2015). HIV Drug Resistance Testing by High-Multiplex "Wide" Sequencing on the MiSeq Instrument. Antimicrob. Agents Chemother. *59*, 6824–6833. 10.1128/AAC.01490-15.

45. Tachbele, E., Kyobe, S., Katabazi, F.A., Kigozi, E., Mwesigwa, S., Joloba, M., Messele, A., Amogne, W., Legesse, M., Pieper, R., et al. (2021). Genetic Diversity and Acquired Drug Resistance Mutations Detected by Deep Sequencing in Virologic Failures among Antiretroviral Treatment Experienced Human Immunodeficiency Virus-1 Patients in a Pastoralist Region of Ethiopia. Infect. Drug Resist. *Volume 14*, 4833–4847. 10.2147/IDR.S337485.

46. Pineda-Peña, A.-C., Faria, N.R., Imbrechts, S., Libin, P., Abecasis, A.B., Deforche, K., Gómez-López, A., Camacho, R.J., de Oliveira, T., and Vandamme, A.-M. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. Infect. Genet. Evol. *19*, 337–348. 10.1016/j.meegid.2013.04.032.

47. Hsin, J., Arkhipov, A., Yin, Y., Stone, J.E., and Schulten, K. (2008). Using VMD: An Introductory Tutorial. Curr. Protoc. Bioinforma. *24*. 10.1002/0471250953.bi0507s24.

48. Knapp, B., van der Merwe, P.A., Dushek, O., and Deane, C.M. (2019). MHC binding affects the dynamics of different T-cell receptors in different ways. PLOS Comput. Biol. *15*, e1007338. 10.1371/journal.pcbi.1007338.

49. Li, X., Singh, N.K., Collins, D.R., Ng, R., Zhang, A., Lamothe-Molina, P.A., Shahinian, P., Xu, S., Tan, K., Piechocka-Trocha, A., et al. (2023). Molecular basis of differential HLA class I-restricted T cell recognition of a highly networked HIV peptide. Nat. Commun. *14*, 2929. 10.1038/s41467-023-38573-8.

50. Ndongala, M.L., Kamya, P., Boulet, S., Peretz, Y., Rouleau, D., Tremblay, C., Leblanc, R., Côté, P., Baril, J., Thomas, R., et al. (2010). Changes in Function of HIV-Specific T-Cell Responses with Increasing Time from Infection. Viral Immunol. *23*, 159–168. 10.1089/vim.2009.0084.

51. Excler, J.L., Robb, M.L., and Kim, J.H. (2015). Prospects for a globally effective HIV-1 vaccine. Vaccine *33*, D4–D12. 10.1016/j.vaccine.2015.03.059.

52. Korber, B., and Fischer, W. (2020). T cell-based strategies for HIV-1 vaccines. Hum. Vaccin. Immunother. *16*, 713–722.

53. Shin, S.Y. (2016). Recent update in HIV vaccine development. Clin. Exp. Vaccine Res. *5*, 6. 10.7774/cevr.2016.5.1.6.

54. Zipeto, D., and Beretta, A. (2012). HLA-C and HIV-1: friends or foes ? Retrovirology *9*, 1–9.

55. Buranapraditkun, S., Hempel, U., Pitakpolrat, P., Allgaier, R.L., Thantivorasit, P., Lorenzen, S.-I., Sirivichayakul, S., Hildebrand, W.H., Altfeld, M., Brander, C., et al. (2011). A Novel Immunodominant CD8+ T Cell Response Restricted by a Common HLA-C Allele Targets a Conserved Region of Gag HIV-1 Clade CRF01_AE Infected Thais. PLoS One *6*, e23603. 10.1371/journal.pone.0023603.

56. Hashimoto, M., Akahoshi, T., Murakoshi, H., Ishizuka, N., Oka, S., and Takiguchi, M. (2012). CTL recognition of HIV-1-infected cells via cross-recognition of multiple overlapping peptides from a single 11-mer Pol sequence. Eur. J. Immunol. *42*, 2621–2631. 10.1002/eji.201242483.

57. Perdomo-Celis, F., Taborda, N.A., and Rugeles, M.T. (2019). CD8+ T-Cell Response to HIV Infection in the Era of Antiretroviral Therapy. Front. Immunol. *10*. 10.3389/fimmu.2019.01896.

58. Addo, M.M., Yu, X.G., Rathod, A., Cohen, D., Eldridge, R.L., Strick, D., Johnston, M.N., Corcoran, C., Wurcel, A.G., Fitzpatrick, C.A., et al. (2003). Comprehensive Epitope Analysis of Human Immunodeficiency Virus Type 1 (HIV-1)-Specific T-Cell Responses Directed against the Entire Expressed HIV-1 Genome Demonstrate Broadly Directed Responses, but No Correlation to Viral Load. J. Virol. *77*, 2081–2092. 10.1128/JVI.77.3.2081-2092.2003.

59. Rosás-Umbert, M., Gunst, J.D., Pahus, M.H., Olesen, R., Schleimann, M., Denton, P.W., Ramos, V., Ward, A., Kinloch, N.N., Copertino, D.C., et al. (2022). Administration of broadly neutralizing anti-HIV-1 antibodies at ART initiation maintains long-term CD8+ T cell immunity. Nat. Commun. *13*, 6473. 10.1038/s41467-022-34171-2.

60. Joglekar, A. V., Liu, Z., Weber, J.K., Ouyang, Y., Jeppson, J.D., Noh, W.J., Lamothe-Molina, P.A., Chen, H., Kang, S. gu, Bethune, M.T., et al. (2018). T cell receptors for the HIV KK10 epitope from patients with differential immunologic control are functionally indistinguishable. Proc. Natl. Acad. Sci. U. S. A. *115*, 1877–1882. 10.1073/pnas.1718659115.

61. Tassiopoulos, K., Landay, A., Collier, A.C., Connick, E., Deeks, S.G., Hunt, P., Lewis, D.E., Wilson, C., and Bosch, R. (2012). CD28-Negative CD4+ and CD8+ T Cells in Antiretroviral Therapy–Naive HIV-Infected Adults Enrolled in Adult Clinical Trials Group Studies. J. Infect. Dis. *205*, 1730–1738. 10.1093/infdis/jis260.

62. Ziegler, M.C., Nelde, A., Weber, J.K., Schreitmüller, C.M., Martrus, G., Huynh, T., Bunders, M.J., Lunemann, S., Stevanovic, S., Zhou, R., et al. (2020). HIV-1 induced changes in HLA-C*03:04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. AIDS *34*, 1713–1723. 10.1097/QAD.0000000000002596.

63. Cohen, G.B., Gandhi, R.T., Davis, D.M., Mandelboim, O., Chen, B.K., Strominger, J.L., and Baltimore, D. (1999). The selective downregulation of class I major

histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. Immunity *10*, 661–671. 10.1016/S1074-7613(00)80065-5.

64. Mann, J.K., Byakwaga, H., Kuang, X.T., Le, A.Q., Brumme, C.J., Mwimanzi, P., Omarjee, S., Martin, E., Lee, G.Q., Baraki, B., et al. (2013). Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. Retrovirology *10*, 100. 10.1186/1742-4690-10-100.

65. Fellay, J., Shianna, K. V, Ge, D., Colombo, S., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cozzi-lepri, A., Luca, A. De, et al. (2007). A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1. Science (80-.). *317*, 944–947.

66. Kulkarni, S., Savan, R., Qi, Y., Gao, X., Yuki, Y., Bass, S.E., Martin, M.P., Hunt, P., Deeks, S.G., Telenti, A., et al. (2011). Differential microRNA regulation of HLA-C expression and its association with HIV control. Nature *472*, 495–498. 10.1038/nature09914.

67. Thomas, R., Apps, R., Qi, Y., Gao, X., Male, V., O'hUigin, C., O'Connor, G., Ge, D., Fellay, J., Martin, J.N., et al. (2009). HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. Nat. Genet. *41*, 1290–1294. 10.1038/ng.486.

68. Koita, O.A., Dabitao, D., Mahamadou, I., Tall, M., Dao, S., Tounkara, A., Guiteye, H., Noumsi, C., Thiero, O., Kone, M., et al. (2006). Confirmation of Immunogenic Consensus Sequence HIV-1 T-cell Epitopes in Bamako, Mali and Providence, Rhode Island. Hum. Vaccin. *2*, 119–128. 10.4161/hv.2869.

69. Goepfert, P.A., Bansal, A., Edwards, B.H., Ritter, G.D., Tellez, I., McPherson, S.A., Sabbaj, S., and Mulligan, M.J. (2000). A Significant Number of Human Immunodeficiency Virus Epitope-Specific Cytotoxic T Lymphocytes Detected by Tetramer Binding Do Not Produce Gamma Interferon. J. Virol. *74*, 10249–10255. 10.1128/JVI.74.21.10249-10255.2000.

ACKNOWLEDGEMENT

The project described was supported by Award Number, U54AI110398 administered by the National Institute of Allergy and Infectious Disease (NIAID), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), and National Human

Genome Research Institute (NHGRI) as part of the NIH Common Fund H3Africa Initiative. This work was also supported by the Makerere University-Uganda Virus Research Institute Centre of Excellence for Infection and Immunity Research and Training (MUII). MUII is supported through the DELTAS Africa Initiative (Grant no. 107743). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS), Alliance for Accelerating Excellence in Science in Africa (AESA), and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant no. 107743) and the UK Government. The government of Uganda funded this work through Makerere University Research and Innovations Fund (grant number: MAK-RIF). We thank Nottingham Trent University (NTU) for providing the Hamilton and Avicenna high-performance computing clusters on which this work was performed.

4.3.7 Published Article

International Journal of
Molecular Sciences

MDPI

Article

Identification of a Clade-Specific HLA-C*03:02 CTL Epitope GY9 Derived from the HIV-1 p17 Matrix Protein

Samuel Kyobe [1,*,†], Savannah Mwesigwa [1,2], Gyaviira Nkurununngi [3,4], Gaone Retshabile [5], Moses Egesa [3,4], Eric Katagirya [2], Marion Amujal [2], Busisiwe C. Mlotshwa [5], Lesedi Williams [5], Hakim Sendagire [1,‡] on behalf of the CAfGEN Consortium, Dithan Kiragga [6], Graeme Mardon [7,8], Mogomotsi Matshaba [9,10], Neil A. Hanchard [11], Jacqueline Kyosiimire-Lugemwa [3] and David Robinson [12]

[1] Department of Medical Microbiology, College of Health Sciences, Makerere University, P. O Box 7072 Kampala, Uganda; savannah.mwesigwa@gmail.com (S.M.); hakimsendagire@gmail.com (H.S.)
[2] Department of Immunology and Molecular Biology, College of Health Sciences, Makerere University, P. O Box 7072 Kampala, Uganda; katagiryaeric@gmail.com (E.K.); amujalcrystal@gmail.com (M.A.)
[3] The Medical Research Council /Uganda Virus Research Institute & London School Hygine Tropical Medicine Uganda Research Unit, PO Box 49 Entebbe, Uganda; gyaviira.nkurununngi@mrcuganda.org (G.N.); m.c.egesa@gmail.com (M.E.); jlugemwa@gmail.com (J.K.-L.)
[4] Department of Infection Biology, London School of Hygiene & Tropical Medicine, Keppel Street London WC1E 7HT London, UK
[5] Department of Biological Sciences, University of Botswana, Private Bag UB 0022 Gaborone, Botswana; retshabileg@ub.ac.bw (G.R.); b.mlotshwa@hotmail.co.uk (B.C.M.); williamsl@ub.ac.bw (L.W.)
[6] Baylor College of Medicine Children's Foundation, P.O Box 72052 Kampala, Uganda; dkiragga@baylor-uganda.org
[7] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; gmardon@bcm.edu
[8] Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX 77030, USA
[9] Pediatric Retrovirology, Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; matshaba@bcm.edu
[10] Botswana-Baylor Children's Clinical Centre of Excellence, Private Bag BR 129 Gaborone, Botswana
[11] National Human Genome Research Institute, National Institutes of Health, 50 South Drive 20892 Bethesda, MD, USA; neil.hanchard@nih.gov
[12] Department of Chemistry and Forensics, School of Science and Technology, Nottingham Trent University Clifton Lane, Nottingham NG11 8NS, UK; david.robinson@ntu.ac.uk
* Correspondence: samuelkyobe@gmail.com
† Current address: Affiliation 1.
‡ Membership of the CAfGEN Consortium is provided in the Acknowledgments.

Citation: Kyobe, S.; Mwesigwa, S.; Nkurununngi, G.; Retshabile, G.; Egesa, M.; Katagirya, E.; Amujal, M.; Mlotshwa, B.C.; Williams, L.; Sendagire, H., on behalf of the CAfGEN Consortium; et al. Identification of a Clade-Specific HLA-C*03:02 CTL Epitope GY9 Derived from the HIV-1 p17 Matrix Protein. *Int. J. Mol. Sci.* **2023**, *24*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Received: 7 October 2023
Revised: 20 November 2023
Accepted: date
Published: date

Abstract: Efforts towards an effective HIV-1 vaccine have remained mainly unsuccessful. There is increasing evidence for a potential role of HLA-C-restricted $CD8^+$ T cell responses in HIV-1 control, including our recent report of HLA-C*03:02 among African children. However, there are no documented optimal HIV-1 $CD8^+$ T cell epitopes restricted by HLA-C*03:02; additionally, the structural influence of HLA-C*03:02 on epitope binding is undetermined. Immunoinformatics approaches provide a fast and inexpensive method to discover HLA-restricted epitopes. Here, we employed immunopeptidomics to identify HLA-C*03:02 $CD8^+$ T cell epitopes. We identified a clade-specific Gag-derived GY9 (GTEELRSLY) HIV-1 p17 matrix epitope potentially restricted to HLA-C*03:02. Residues E62, T142, and E151 in the HLA-C*03:02 binding groove and positions p3, p6, and p9 on the GY9 epitope are crucial in shaping and stabilizing the epitope binding. Our findings support the growing evidence of the contribution of HLA-C molecules to HIV-1 control and provide a prospect for vaccine strategies.

Keywords: HIV control; HLA-C*03:02; immunoinformatics

The full article is available as open access online at: *Int. J. Mol. Sci.* **2023**, *24*, x. https://doi.org/10.3390/xxxxx

CHAPTER 5: GENERAL DISCUSSION

The HIV epidemic is multifaceted, with a considerable reduction in prevalence reported in many African areas. This reduction is attributable to successful public health campaigns with tailored messages and an emphasis on targeted interventions in high-risk populations. However, a substantial number of people are infected with HIV annually. Unfortunately, to date, most preventive strategies, including vaginal microbicides and candidate vaccines, have failed to achieve an appreciable level of protection. A small proportion of people infected with HIV possess the ability to naturally control HIV for a long period, including children. These cohorts offer hope for developing novel vaccines; however, since the Test-and-Treat Strategy, we can no longer explore them. We thus remain at the last frontiers to explore and discover unknown natural mechanisms that enhance HIV control.

The hypervariable HLA class I molecules widely existent on nucleated cells present intracellularly processed viral antigens to $CD8^+$ T cells. This immune synapse plays a fundamental role in eliciting the cell-mediated immune responses, often called the *"conductor of the immune orchestra."* Some HLA class I molecules, especially the HLA-C, also present viral antigens to NK cells via the KIRs mediating various effector functions. However, most current evidence was derived from Western adult populations, in whom different HIV-1 clades are responsible for HIV infections. These observations prompted the formulation of the study hypothesis in this thesis that there is a variable distribution of LTNP in Africa, and certain population-specific antigen-HLA complexes are associated with LTNP through their role in mediating $CD8^+$ T cell activity.

5.1 LTNP and Factors Associated with Progression

First, this thesis aimed to determine the frequency of LTNPs and factors associated with progression among children infected with HIV in Uganda and Botswana. Perinatal HIV infection remains a significant public health concern, and understanding the factors contributing to HIV progression among infected children is crucial for developing and implementing targeted interventions (De Pee & Semba, 2010; J. A. Mendoza et al., 2014). Our retrospective study enrolled an ethnically diverse cohort of children with perinatal HIV infection that was followed longitudinally over time to identify individuals who develop AIDS

and are initiated on ART according to prevailing national ART guidelines. LTNP was defined as maintaining normal $CD4^+$ T cell counts or, conservatively, including children who exhibited asymptomatic or mildly symptomatic infection without encountering AIDS-defining conditions for ten years or longer (K. Nielsen et al., 1997). The results of our study revealed a notable prevalence of LTNP among children with perinatal HIV infection. Approximately 6.3% of the cohort demonstrated LTNP status, and the proportion varies from 9.5% in Botswana to 5.9% in Uganda. Nielsen et al. reported a 27% proportion of LTNP among children in the USA, while a French cohort study (K. Nielsen et al., 1997; Warszawski et al., 2007) found a much lower proportion of 2.4% of children infected with HIV as LTNP. These findings demonstrate the existence of geographical variations in the rate of HIV disease progression, which aligns with the observations presented in this thesis. The difference in the rate of progression between children from Uganda and Botswana may be attributable to the baseline $CD4^+$ T cell thresholds (seen in this cohort) and the various circulating HIV-1 clades in both countries (Bbosa et al., 2019); however, the later contribution remains controversial in the literature (Amornkul et al., 2013; Ssemwanga et al., 2013; Venner et al., 2016). Despite using Nielson's similar definition of LTNP (K. Nielsen et al., 1997), we report a substantially lower proportion among African children. Nevertheless, identifying a substantial number of LTNPs in our study cohort suggests that certain immunological or genetic factors might contribute to the ability to control viral replication and delay disease progression in perinatally infected children. However, LTNP status is not a lifelong status as more than 50% of children classified as LTNPs in our cohort experience disease progression by the 13^{th} year of life, slightly six months later compared to adult Western populations (Buchbinder & Vittinghoff, 1999; Helm et al., 2014; K. Nielsen et al., 1997).

Upon further comparison of our data with previous studies, several difficulties are brought to the fore, mainly due to the inconsistent definition of LTNP (Gurdasani et al., 2014; Paul et al., 2005). Despite the paucity of data, previous studies have reported LTNP frequencies ranging from 1.5% to 27% (Paul et al., 2005). Our study found an LTNP proportion that falls within this range, providing additional insights into the prevalence of LTNPs in this population. Notwithstanding certain limitations, such as survival bias and LTFU in our cohorts, this thesis substantially contributes to the current knowledge by conducting an extended longitudinal analysis of a significantly larger and diverse cohort. The significant proportion of children with

LTFU adversely impacts our ability to observe them prospectively, influencing our estimate of LTNP. The observed high 10-year AIDS-free survival can be attributed to survival bias commonly encountered in observational studies, particularly in conditions where the risk of death during the early periods is notably very high (Poorolajal et al., 2016; Tchetgen Tchetgen et al., 2015). While LTNP individuals demonstrate a distinctive ability to control HIV infection, the impact of the viral reservoir remains a critical aspect of their long-term viral dynamics (Leitman et al., 2017). It is believed that there is a critical balance control of infection in LTNPs and reduction of viral reservoirs, presumably via CD8 T cells (Chun et al., 2001). Understanding the interplay between the reservoir and the host immune response in LTNP individuals is crucial for advancing our knowledge of HIV pathogenesis and informing strategies aimed at achieving sustained viral control and potentially a functional cure (Berendam et al., 2020; Dangeti, 2014; Leitman et al., 2017; Tobin & Aldrovandi, 2014).

Furthermore, this study demonstrates significant associations between stunting, wasting, year of enrolment into care, and country of origin with HIV disease progression among children with perinatal HIV infection. Stunting, characterized by chronic malnutrition and impaired linear growth, strongly predicted progression in this population. Similarly, wasting, indicative of acute malnutrition and severe weight loss, emerged as a significant risk factor for faster HIV disease progression. The results of our study corroborate and extend previous research on the high prevalence (Jesson et al., 2015; Sewale et al., 2018) and impact (Muenchhoff et al., 2018) of malnutrition on HIV disease progression in African children. Previous studies have shown that malnutrition negatively impacts immune function, increasing T cell activation and exhaustion (Bourke et al., 2019; Ibrahim et al., 2017; Nájera et al., 2004; Rytter et al., 2014). Consequently, there is an escalation in viral replication and an accelerated progression of HIV disease with poor CD4 recovery (Muenchhoff et al., 2018). Equally, HIV is a state of heightened metabolic demands and dysregulation (high microbial translocation) that elevates the risk of undernutrition, thereby accelerating the risk of progression (Kimani-Murage et al., 2011; Muenchhoff et al., 2018). The risk of malnutrition is further worsened by the high risk of food insecurity among their households (J. A. Mendoza et al., 2014; Rose et al., 2014). Identifying stunting and wasting as risk factors for HIV disease progression in our cohort reaffirms the need to integrate nutritional support into the clinical care of children with perinatal HIV infection (Baum et al., 2013).

Additionally, the year of enrolment into care was identified as a relevant factor ("cohort effect"), with children enrolled after 2006 paradoxically exhibiting a faster rate of progression compared to those enrolled later. Our finding is inconsistent with other studies reporting improvements in pediatric HIV progression over time. However, this is mainly attributable to early initiation of ART among children infected with HIV and not necessarily improvements in socioeconomic factors and access to healthcare (Boender et al., 2012; Iyun et al., 2020). Also, survival bias is another aspect that could account for the increased rate of progression observed in the later cohort, with the earlier cohort (<2006) being enriched for older children who have survived. Also, the cohort effect could be explained by the increased recruitment of children below five years, who are known to progress much faster to AIDS (Warszawski et al., 2007). Similarly, the changes in national ART policies and guidelines over time likely contribute to these observed differences. For example, in 2008 the WHO recommended initiating ART for all infants infected with HIV under 24 months of age, regardless of their clinical or immunological condition (Wieczorek et al., 2017b).

Furthermore, the country of origin also influenced HIV disease progression, with children from Botswana experiencing more rapid disease advancement compared to Uganda. Socioeconomic factors, access to healthcare, HIV epidemiologic curves, and regional variations in HIV subtypes may contribute to this disparity (Bbosa et al., 2019; Boender et al., 2012; Nixon et al., 2011). When establishing the HIV care centers, Uganda and Botswana were at distinct stages in the transitions of their respective HIV epidemic curves (Nixon et al., 2011). The country-specific variations in HIV disease progression among children with perinatal HIV infection underscores the need for region-specific approaches to manage the epidemic effectively. Various studies have indicated that the prevalence of different HIV-1 subtypes and genetic diversity across countries can influence disease progression and treatment outcomes (Bbosa et al., 2019; Touloumi et al., 2013). Our study provides additional evidence for the critical role of early and comprehensive HIV care, focusing on addressing malnutrition and region-specific challenges to promote long-term non-progression and improved quality of life for children living with perinatal HIV infection worldwide. We firmly believe our findings can be directly applied and integrated into the current clinical HIV care practice, emphasizing the Test-and-Treat Strategy.

5.2 HLA class I and HIV LTNP

The host immune response to HIV-1 infection exhibits considerable variability among individuals, including children. Not all children exposed to HIV-1 become infected; even among those infected, the disease's progression varies significantly (Paximadis et al., 2011; Warszawski et al., 2007). Numerous studies have demonstrated the influence of host genetic factors, such as the 32-bp deletion within the CCR5 coding sequence (CCR5-Δ32), in determining susceptibility to HIV-1 infection and progression (R. Liu et al., 1996; Paximadis et al., 2011). Specifically, the role of specific HLA allotypes and haplotypes in HIV-1 disease progression has been extensively studied across Western populations. Despite the efforts to investigate this association, inconsistent results among studies have contributed to a complex and ambiguous understanding of the interplay between HLA class I and HIV-1 non-progression, highlighting the importance of replicating genetic study findings in diverse populations.

This thesis investigated the distribution of HLA class I alleles and haplotypes among children infected with HIV in Uganda and Botswana. Our findings reveal a striking enrichment of population-specific HLA class I alleles and haplotypes in the respective cohorts. In Uganda, certain HLA class I alleles and haplotypes, such as the HLA-A*30:01, HLA-B*42:01, and HLA-C*17:01 (Supp. Table 5), previously associated with enhanced viral control in adult HIV populations, were found to be overrepresented among infected children (Kloverpris et al., 2012; Peterson et al., 2013). Similarly, in Botswana, distinct HLA class I alleles and haplotypes like the HLA-C*06:02 (Supp. Table 5) allele, known for their potential role in modulating immune responses to control HIV, showed higher prevalence in children infected with HIV (Peterson et al., 2013). The observed enrichment of population-specific HLA variants suggests a unique immunogenetic landscape that may influence HIV disease progression in these regions (Kloverpris et al., 2012; Mori et al., 2014). Furthermore, the enrichment of population-specific HLA class I alleles and haplotypes in children infected with HIV in Uganda and Botswana aligns with previous studies in adult populations, indicating that HLA class I alleles play a crucial role in HIV pathogenesis across age groups (Paximadis et al., 2011). These observed differences in HLA class I profiles between the two countries may also reflect the distinct genetic ancestries and unique HIV-1 subtypes prevalent in each region (Bbosa et al.,

2019). Indeed, we have also demonstrated some genetic differentiation (Fst = 0.0065) between Uganda and Botswana (Retshabile et al., 2018). Enriching population-specific HLA class I alleles in children infected with HIV from Uganda and Botswana provides additional evidence of geographically determined genetic factors that influence HIV disease outcomes. These findings emphasize the importance of accounting for genetic diversity in HIV research and underscore the need for population-specific studies to enhance our understanding of HIV immunogenetics. These findings may also explain the differences in the rates of disease progression we observed between Uganda and Botswana.

Notably, we observed a low LD among HLA class I alleles in children infected with HIV in Uganda and Botswana. The low LD suggests a high diversity of HLA class I allele combinations, which may contribute to variations in HIV disease progression. Moreover, the low LD observed in our study indicates a diverse repertoire of HLA class I alleles within the children infected with HIV, which may offer tremendous potential for presenting a wide range of viral antigens, impacting the host's ability to mount an effective immune response (Carrington et al., 1999). Additionally, the low LD within populations suggests that each HLA class I allele might play a relatively independent role in shaping HIV immune control (Matthews et al., 2011). As the field of HIV immunogenetics continues to evolve with cheaper HLA genotyping from genomic data such as WES (Karnes et al., 2017; Kyobe et al., 2021; C. Wang et al., 2012), our findings add to the growing body of evidence supporting the potential integration of genetic information into clinical decision-making, ultimately leading to improved outcomes for children infected with HIV worldwide.

We further identified a significant association between LTNP and HLA-B alleles (B*57:03 and B*58:01), known for their ability to present conserved HIV epitopes, allowing for efficient recognition and activation of potent antiviral immune responses. Our results are consistent with previous research that has demonstrated the significance of HLA-B*57:03 and HLA-B*58:01 alleles in LTNP and/or control HIV (low viremia, high CD4 counts and low T cell activation) infection among adult and pediatric African and Western populations (P. J. R. Goulder & Walker, 2012; Kovacs et al., 2020; Leslie et al., 2010; McLaren et al., 2012; Peterson et al., 2013; E. A. Trachtenberg & Erlich, 2001). However, our findings are in discordance with a Botswana-based study that reported the absence of a protective effect associated with HLA-

B*57 and HLA-B*58:01; this phenomenon was primarily attributed to viral adaptation within their cohort (Payne et al., 2014). The association of HLA-B*57:03 and HLA-B*58:01 with LTNP in both adult and pediatric populations underscores the importance of these alleles in immune control across different age groups. However, these findings contrast with those of Naruto et al., who found HLA-B*67:01 and B*52:01 in higher frequency among Japanese adults infected with HIV with a low viral load and high CD4 count (Naruto et al., 2012). They also reported HLA-C*12:02 as a possible marker for slow HIV disease progression. Surprisingly, in an earlier investigation among Japanese hemophiliacs, Kawashima et al. observed that participants with HLA-B*51:01 alleles showed significantly slow HIV disease progression (Kawashima et al., 2010). These findings indicate that the rarity of protective HLA-B*57 and B*27 alleles in the Japanese population means that other HLA class I alleles play a prominent role in influencing HIV progression in Japan. The contrasting findings of Naruto and Kawashima underscore the notion that even within populations exhibiting relative genetic homogeneity, and there can be diverse HLA class I alleles that play a role in HIV disease progression. They also emphasize the need for region-specific studies to inform region-specific vaccine strategies.

Similarly, we also observed an association between the HLA-C*03:02 allele and LTNP, a previously unreported finding in African populations. Our results support those from Naruto et al., who, in addition to two HLA-B alleles, found an additional single HLA-C allele associated with low viral load and CD4 counts (Naruto et al., 2012). Notably, various HLA-C alleles (such as C*12:02, C*12:03, and C*02) have been implicated in reduced viral load set-point and slower disease progression in multiple populations (Chikata et al., 2018; Leslie et al., 2010; Naruto et al., 2012; K. K. Singh et al., 2011). It is important to highlight that while in the study conducted by Leslie and colleagues in South Africa, the statistical significance of the HLA-C*03:02 allele was not maintained after adjusting for multiple testing, it was observed to be correlated with a lower median viral load (15,000 copies/ml) (Leslie et al., 2010). Further still, they observed that participants with HLA-C*03:02 alleles had a high median CD4 count of 428 cells/ml (Leslie et al., 2010). This suggests that our approach of extreme phenotyping proved superior in revealing the protective effects of HLA-C*03:02 in African populations. Additional evidence substantiating the role of the HLA-C*03:02 allele in LTNP comes from a study by Wright and colleagues (Wright et al., 2010). Their study found that individuals

expressing this allele had a lower viral replicative capacity. This implies that the allele exerts an HLA-mediated immune pressure of HIV-1 that encourages the emergence of virus escape mutants that are not very good at replicating, resulting in fitness costs. Identifying the novel HLA-C*03:02 allele as a contributor to LTNP in our study adds to the growing body of evidence supporting the diverse role of HLA-C alleles in shaping the immune response against HIV (Blais et al., 2012).

HLA class I heterozygosity confers significant advantages in HIV control (Carrington et al., 1999). The diversity of HLA molecules resulting from heterozygosity enhances the presentation of a broader range of viral epitopes to the immune system, increasing the likelihood of eliciting robust and effective T cell responses (Arora et al., 2020; Matthews et al., 2012). This diversity of presented epitopes reduces the chances of viral escape mutations, where the virus mutates to evade immune recognition. Studies have shown that certain HLA class I haplotypes (such as HLA-B*52:01-C*12:02 in Japanese, HLA-A*30-B*13-C*06 in Chinese, HLA-A*74:01-B*57:03 and B*39:10-Cw*12:03 in South Africans) exhibit more efficient control of HIV, resulting in lower viral loads, higher CD4 counts and therefore delayed disease progression (Leslie et al., 2010; Naruto et al., 2012; H. Zhang et al., 2013). In our study, we demonstrated an association between LTNP and HLA class I haplotypes B*57:03-C*07:01 and B*58:01-C*03:02. As expected, our findings do not align with previous studies, which can be attributed to the distinct populations and the specific distribution patterns of HLA class I alleles in our study cohort. Because HLA class I heterozygosity enables the recognition of a broader spectrum of HIV variants, understanding and leveraging their benefits in HIV control could hold promising implications for designing future immunotherapies and vaccine strategies.

Although not the primary focus of this thesis, we also made noteworthy observations regarding HLA class I allelic and haplotypic associations with rapid HIV progression (Sampathkumar et al., 2014). The HLA-B*15:10 allele and the HLA-A*29:02-C*17:01 haplotype were associated with a more rapid course of HIV disease progression. However, it is important to note that these findings do not replicate those observed in the study conducted by Sampathkumar and colleagues among sex workers in Kenya (Sampathkumar et al., 2014) but are well corroborated

with the allelic association of HLA-B*15:10 among adults in Durban, South Africa (Leslie et al., 2010).

Previous research has shown that the co-inheritance of certain HLA class I alleles within haplotypes makes it challenging to disentangle the specific contributions of individual alleles to HIV outcomes (Matthews et al., 2011). After noting a significant association between HLA-C*03:02 in haplotype with HLA-B*58:01 and LTNP status, we analyzed the relationship between these alleles and established their significance in HIV control. First, we investigated whether the protective effects of HLA-C*03:02 against HIV disease progression were attributed to LD. Our findings in the examined populations revealed that HLA-C*03:02 exhibited weak LD with B*58:01 (r^2= 0.18). Consequently, we deduce that the protective impact of HLA-C*03:02 is unlikely solely influenced by LD. We then employed a basic statistical analysis based on Svejgaard's method to determine the relative strength of the association between HLA-C*03:02 and HLA-B*58:01 with LTNP status (Svejgaard & Ryder, 1994). We also compared HLA-C*03:02 with other alleles, including those predicted to be related to rapid disease progression. We found that the protective effects of HLA-C03:02 are not dependent on the presence of the HLA-B*58:01 allele. Moreover, we noted that the protective effects of HLA-C*03:02 appear to be additive (Figure 4-10), indicating that this allele independently contributes to reducing the susceptibility to HIV disease progression within our study cohort. The log odds of protection are increased when HLA-C*03:02 is in existence with other protective HLA alleles, such as HLA-A*74:01 and B*58:01. Moreover, in the presence of HLA-C*03:02 together with alleles shown to be associated with rapid HIV disease progression (HLA-B*15:10), we putatively observed a reversal of the phenotype. Our findings corroborated with the study by Mathew et al., who found that the protective effect of HLA-A*74:01 was not due to its high LD with HLA-B*57:03 (Matthews et al., 2011). This HLA allele was also shown to be associated with protection among adults with HIV in Mbeya, Tanzania (Koehler et al., 2010). Leslie provided additional evidence of the additive contributions of HLA class I protective effects in a 2010 study among adults in Durban, South Africa. Their findings demonstrated that individuals expressing both HLA-B*57:03 and A*7401 alleles exhibited a significantly lower median viral load (1,650 copies/ml) compared to those with HLA-B*57:03 alone (10,400 copies/ml) (Leslie et al., 2010). The subsequent investigation in a mixed Southern African population provided corroborating evidence for their

study, demonstrating cooperative additive effects of six HLA class I pairs (HLA-A*02-B*81, HLA-B*44-C*04, HLA-A*74-B*57, HLA-A*74-B*81, HLA-B*58:01-B*81, and HLA-B*58:01-C*04) with favorable implications for HIV-1 disease control and strengthening the validity and strength of our findings (Matthews et al., 2012). Our distinct study design found that HLA-C*03:02 exerts its protective effects independently of the HLA-B*58:01 allele. We also observed that the protective effects of HLA-C*03:02 appear to be additive, signifying that this allele independently contributes to reduced susceptibility to HIV disease progression in our study cohort. This observation reinforces that certain combinations of HLA class I alleles can confer enhanced control of viral replication in HIV-infected individuals, underscoring the importance of considering the synergistic effects of specific HLA genotypes in the context of HIV immunity. Additionally, the independence of HLA-C*03:02 from HLA-B*58:01 highlights the importance of considering specific HLA class I allelic effects in dissecting the complex interplay between host genetics and HIV outcomes and we further assert that these observations can extended in the current era of Test-and-Treat Strategy (Kretzschmar et al., 2013). Identifying HLA-C*03:02 as a standalone contributor to protection against HIV infection is a significant finding that expands our understanding of the role of HLA-C alleles in HIV immune control. This thesis provides novel insights into the independent nature of HLA-C*03:02's protective effects, even in the presence of weak LD with HLA-B*58:01. This supports the notion that multiple HLA class I alleles can independently influence the immune control of HIV and their combined effects may enhance the overall protection against viral replication and disease progression.

In our study, a novel finding emerged regarding the protective effect of HLA-C*03:02 in LTNPs among children infected with HIV. However, upon further analysis, we found that the combined protective effect of HLA-C*03:02, B*57:03, and B*58:01 accounted for 16.5% of the overall LTNP status in our study cohort. Furthermore, the observation that HLA-C*03:02, B*57:03, and B*58:01 account for less than ~17% of LTNP highlights the heterogeneity within LTNP individuals. In an comparable study involving adults, the contributory role of protective genes in determining HIV control exhibits a comparable magnitude of variation. Fellay conducted a GWAS focusing on differences in viral set points during the asymptomatic stage, revealing that overall the genetic effects contribute to 15% of the observed variation, while individual SNPs in the HLA-C and B genes collectively account for 6.5–9.5% (Fellay et al.,

2007). Another study examining HIV controllers versus progressors demonstrated that four SNPs in the MHC region explained approximately 19% of HIV control variation in European participants (Walker et al., 2010). Consequently, our findings align with previous cohorts, indicating similarities in the genetic determinants of HIV control among adults.

This observation has significant implications for our understanding of the complex genetic determinants that underlie LTNP and the multifactorial nature of HIV immune control. The limited contribution of protective HLA class I alleles to the LTNP phenotype suggests that additional host and viral genetic (Magierowska et al., 1999; Pant Pai et al., 2012; Soriano et al., 2002), environmental (De Pee & Semba, 2010; Mwesigwa et al., 2021), or immunological (Okoye & Picker, 2013) factors play a substantial role in determining the ability of children to control HIV infection without progressing to AIDS. While the presence of protective HLA class I alleles is undoubtedly beneficial, it is not the sole factor responsible for achieving LTNP status. This highlights the importance of considering a comprehensive set of genetic and non-genetic factors when studying HIV disease progression and immune control. Also, investigating the interactions between protective HLA class I alleles and other genetic or immune-related determinants could yield valuable insights into the underlying mechanisms of HIV control (Mwesigwa et al., 2021) in children on ART.

Overall, the observed associations with LTNP in Ugandan and Batswana children support the generalizability of these findings across these diverse populations. The genetic diversity among various ethnic groups can influence HLA class I allele distributions and may contribute to population-specific differences in HIV disease progression. As such, our study contributes to the growing understanding that novel HLA-C alleles impact HIV immune control, including pediatric populations in Africa.

5.3 Mechanisms of HLA class I mediated HIV control

Different alleles of the same HLA gene show variations in their effects on HIV progression, and various mechanisms of action have been presented to explain these differences (P. J. R. Goulder & Walker, 2012). Numerous studies have shown that distinct alleles exhibit varying epitope specificity, and these epitopes are linked to diverse CTL functions (Kaseke et al., 2021). The CTLs induced by these alleles demonstrate enhanced proliferative capacity and

display polyfunctional responses (Ferrando-Martínez et al., 2012). Additional evidence supports the role of specific residues within the peptide binding groove and the epitope in determining epitope specificity (Ferrando-Martínez et al., 2012). Furthermore, previous studies have shown that protective alleles induce compensatory mutations within the virus, often at the cost of viral fitness (Payne et al., 2014). In this study, we investigated these mechanisms to understand the role of HLA-C*03:02 in HIV control.

Understanding the distinct peptides and peptide-binding properties of novel population-specific protective HLA class I alleles may provide valuable insights for vaccine development and immunotherapies targeting specific HIV-1 variants prevalent in Africa (Ferrando-Martínez et al., 2012). The findings from our study confirm that predictive epitope mapping of HIV-1 clades C and A/A1 reveals both private and shared HLA-C*03:02-epitopes. The presence of private epitopes suggests that certain regions of the virus are unique to specific clades, potentially influencing viral pathogenesis and immune responses in a clade-specific manner. On the other hand, identifying shared HLA-C*03:02-epitopes implies common immunogenic regions that can be targeted across different clades. This observation has important implications for the development of effective HIV vaccines, as it highlights the necessity of considering clade-specific epitopes alongside conserved ones to achieve broader immune protection against diverse HIV strains, especially in the design of peptide or epitope-based vaccine candidates (Hansen et al., 2019; Korber & Fischer, 2020; Y. Yang et al., 2015; Zhao et al., 2013).

Furthermore, using Logos plots, we found that HLA-C*03:02 molecules prefer large hydrophobic residues in the F pocket when accommodating epitopes. This structural feature has not been demonstrated in the context of HLA-C*03:02 antigen presentation of HIV. The preferential accommodation of hydrophobic residues may play a crucial role in shaping the immune response to HIV by affecting epitopes' binding affinity and subsequent recognition by T cells. Indeed, the preferential accommodation of hydrophobic residues in the F pocket by HLA-C*03:02 is shared with other HLA class I protective alleles such as HLA-B*57:03 (Stewart-Jones et al., 2005). This may suggest that the HLA-C*03:02 allele may have the potential for cross-presentation of HLA-B*57:03-restricted epitopes, given that both molecules have a large surface area in the F pocket (Barugahare, Baker, K'Aluoch, et al., 2005; Gnjatic et al., 2003; S. Zhang et al., 2011). The significance of this finding lies in its potential impact

on the diversity and specificity of T cell responses mounted against HIV. Further investigations into the functional implications of hydrophobic residue preference in HLA-C*03:02 presentation could provide valuable insights into the dynamics of HLA-peptide interactions and their role in shaping the immune response to HIV.

Furthermore, the utilization of molecular docking and molecular dynamics simulations enabled us to identify four stable complexes formed between HLA-C*03:02 molecules and nano peptides (GY9, GF10, AY9, and VL9) mainly derived from HIV-1 structural proteins (1/4) Gag and (2/4) Pol. Molecular docking enables the virtual screening of thousands of potential peptides and reveals the best docked pHLA complexes (R. K. Pandey et al., 2018; Usman Mirza et al., 2016). The MD simulations can accurately represent the time-dependent behavior of the biological complexes. It has been further reported that MD simulations and binding free energy calculations can accurately predict the binding affinities of pHLA complexes and effectively rank peptides (Kongkaew et al., 2015). The docking interactions between GY9, GF10, AY9, and VL9 epitopes and HLA-C*03:02 demonstrated a high binding affinity (predicted binding energy > –7.5kcal/mol), exhibiting a robust and stable complex formation. We re-evaluated our docking results from DINC with the Convex-PL scoring function, and in each case, these pHLA complexes had a good score.

The docking conformations of these four peptides were closely related (exhibited small RMSD values <1.65Å) to the native peptide co-crystalized with HLA-C*06:02 that was used as a template for the structural homology modeling of HLA-C*03:02. Also, these complexes are stabilized through multiple strong hydrogen bonds (≥10) and van der Waals interactions between the peptide residues and the specific residues and pockets of the HLA-C*03:02 peptide binding groove. Additionally, MD analyses reveal that GY9, GF10, AY9, and VL9 epitopes display minimal conformational changes and stabilization of HLA-C*03:02 upon complex formation; this stability over the simulation time indicates a highly compatible fit within the peptide binding groove. In our extended MD, the simulated complexes are further characterized by forming strong hydrogen bonds that exhibit long bond occupancy (>50%), revealing insights into their stability and interaction dynamics. These peptides also showed the most negative in-silico binding free energy calculated from the MM/PBSA method (Kongkaew et al., 2015; Wan et al., 2008). These methods combine molecular mechanics calculations with

Poisson-Boltzmann continuum solvation models to estimate the free energy of binding in molecular complexes. The resultant binding free energy is the difference between the free energy of the pHLA complex and the peptide and HLA protein. HLA-C*03:02-GY9 complex showed the best results with the lowest binding free energy (–88.41 kcal/mol) as well as the lowest electrostatic interaction energy (-547.44 kcal/mol) as compared to GF10, AY9, and VL9 peptides. Broadly, the electrostatic interactions contribute the most to the binding free energy. This strong interaction would signify that the GY9 epitope is probably preferentially bound and presented by HLA-C*03:02 molecules for immune recognition and control of HIV.

Despite finding a large number of epitopes to be predicted as strong or weak binders, only four were predicted to be structurally stable in complex with HLA-C*03:02. Chikata and colleagues demonstrated that the HIV proteome is efficiently processed into thousands of short peptides (8-12mer) for HLA class I presentation (Chikata et al., 2019), and Ziegler and colleagues further showed that only a limited number (0.2% ~ 8 peptides) are stably presented on the $CD4^+$ T cell (cell expressing HLA-A*02:01/02:01, B*27:05/40:01, C*02:02/03:04) surfaces for immune recognition (Ziegler et al., 2020). Our finding of four epitopes predicted to be structurally stabilizing agrees with these previous findings. Furthermore, the high prioritization of epitopes from structural proteins conforms with the fact that structural proteins are known to contain conserved regions essential for virus survival, making them attractive targets for immune recognition.

Understanding the molecular interactions between HLA-C*03:02 and specific HIV-1 GY9 epitope derived from structural protein could offer critical insights into the structural basis of HLA-C*03:02 peptide binding (Kloverpris et al., 2012). Computational alanine mutagenesis (CAS) has been widely applied and/or experimentally verified in various complexes and diseases (van Deutekom & Keşmir, 2015). The energetic information explains the essential targets for peptide-based vaccine design, as these regions could elicit robust and broadly reactive T cell responses against HIV. The findings from our study, utilizing MD simulation analysis and CAS, provide valuable insights into the structural basis for the preferential binding of the GY9 epitope to HLA-C*03:02. Our results revealed that specific positions, namely positions 62, 142, and 151 in the HLA-C*03:02 molecule, play a critical role in mediating the preferential binding interaction. Notably, Glu62 is also associated with a significant loss of

binding energy across the other three peptides, with the highest loss (ΔΔG -47.47 kcal/mol) detected with the AY9 epitope. Mutating other residues within the peptide binding groove revealed no significant changes in the binding free energy. This would suggest that Glu62, located in the A pocket, contributes substantially to the binding of epitopes in the peptide binding groove. These three residues are found in the A, F, and E pockets of the HLA-C*03:02 binding groove, respectively. Our data agrees with the observation that a limited number of residues in the peptide binding groove contribute significantly to the binding energy of pHLA complexes (van Deutekom & Keşmir, 2015) and these could be responsible for shaping the immune response. For example, despite their strong relationship with HIV progression, only positions 67, 70, and 97 in HLA-B alleles are significantly associated with differences in HIV viral load (Walker et al., 2010). Additionally, these key positions sometimes referred to as hotspots, have been demonstrated to exhibit frequent variations among HLA class I alleles compared to other positions (Mori et al., 2014; van Deutekom & Keşmir, 2015). However, a recent study (Chikata et al., 2022) has shown that residues found outside the peptide binding groove may influence pHLA complexation, although this study was a technical assessment closely related to HLA-C*14:02 (non-protective) and HLA-C*14:03 (protective) that are differently associated with HIV progression. Our data suggests that the residues Glu62, Thr142, and Glu151 are the major determinants of the binding strength in the HLA-C*03:02 peptide binding groove.

Peptides are anchored within the grooves of distinct class I molecules through either a pair of preferred amino acids at specific positions along their length or, less frequently, by a single amino acid (L. Li & Bouvier, 2004; Mobbs et al., 2017; Stewart-Jones et al., 2005; Toh et al., 2000; Watanabe et al., 2011). Evidence suggests that modifications in viral proteins, which impact these specific epitopes over the course of an infection, serve as a mechanism for viral evasion of the immune system (J. M. Carlson, Brumme, et al., 2012; Frater et al., 2007; Navis et al., 2007; Pymm et al., 2022). These alterations might influence how well the epitopes bind to class I molecules, either by changing primary or secondary anchor residues or by maintaining the binding pattern while altering the residues the TCR recognizes (Joglekar et al., 2018; X. Li et al., 2023). The presence of both primary and secondary anchor positions in epitopes is of significant interest in understanding pHLA binding and, therefore, the design of peptide-based vaccines (Insaidoo et al., 2011; L. Li & Bouvier, 2004; Toh et al., 2000;

Zappacosta et al., 1997). Primary anchor positions are well-established as critical residues within an epitope, directly interacting with specific pockets/residues in the peptide binding groove, ensuring stable binding and presentation (L. Li & Bouvier, 2004). Although not as predominant in their influence on MHC binding, secondary anchor positions play a crucial role in modulating the overall pHLA interaction (L. Li & Bouvier, 2004; Ruppert et al., 1993). These secondary anchors seem to fine-tune the binding affinity and stability, contributing to the specificity and immunogenicity of the epitope. Our study supports these findings by demonstrating a substantial loss in binding energy at position 6 in the GY9 epitope. The observed change in binding energy was more than 2.5-fold compared to other positions, strongly indicating that Arg at P6 serves as a primary anchor residue. Notably, the positions p3 and p9 within the GY9 epitope were identified as secondary anchor positions, further reinforcing the importance of these residues in facilitating the stable interaction with HLA-C*03:02. The presence of anchor residues in HIV epitopes, as predicted in silico, can be validated through experimental methods (L. Li & Bouvier, 2004). Understanding the interplay between primary and secondary anchor positions is pivotal for unraveling the complexities of antigen presentation and designing more effective vaccines and immunotherapies.

Through these analyses, we have identified critical structural constraints or determinants responsible for the stability and specificity of the HLA-C*03:02-GY9 complex. These structural insights provide a deeper understanding of the molecular mechanisms governing GY9 presentation by HLA-C*03:02 and contribute to the broader knowledge of HLA-peptide interactions in the context of LTNP. Furthermore, our research underscores the value of MD analysis and CAS in unraveling novel pHLA interactions, conformational changes, and stability. These computational approaches have become indispensable tools in modern theoretical immunology, enabling detailed investigations into biomolecular complexes' dynamic behavior and energetics, leading to design of novel vaccine candidates (Krishnan et al., 2020; Nezafat et al., 2016; R. K. Pandey et al., 2018; Usman Mirza et al., 2016).

Finally, this thesis explored whether the epitope GY9, distinguished by its unique predicted stability and strong binding affinity, is associated with the induction of IFN-γ and/or IL-2 responses. The IFN-γ response is of particular significance, as this pro-inflammatory cytokine is a central mediator of cellular immune responses against viral infections, promoting antiviral

effector functions and immune activation (A. Rakityanskaya et al., 2022; Whitmire et al., 2005). Dual ELISPOT assays have been used to demonstrate the presence of distinct T cell functional subsets in children infected with HIV and their variation over time (Ndongala et al., 2010). The results of our study, utilizing a dual ELISPOT assay to measure IL-2 and IFN-γ responses (Boulet et al., 2007), revealed an interesting finding regarding the GY9 epitope in children infected with HIV on ART. We observed that the GY9 epitope elicited a clade-specific HLA-C*03:02-restricted IFN-γ response in one-third of the study participants expressing the HLA-C*03:02 allele but no response among HLA-C*03:02^{-ve} individuals. This suggests that the GY9 epitope is recognized by specific CD8^{+} T cells in an HLA-C*03:02-dependent manner, leading to the production of IFN-γ (H. Chen et al., 2009). Furthermore, our findings indicate the presence of heterogeneity in responses (30% positivity) to HIV-1 epitopes among individuals expressing the same HLA alleles, which is supported by results of several previous studies (Day et al., 2001; Jin et al., 2002; Mkhwanazi et al., 2010; Schmitt-Haendle et al., 2005).

Additionally, all GY9 responses were found among individuals infected with HIV-1 A1 or C, but none were detected from individuals with D or A1D recombinants. This response's variable and clade-specific nature suggests that the GY9 epitope may be subject to sequence variation, particularly in clades present in the study population, leading to clade-specific CD8^{+} T cell responses. Our study also revealed significant variability in the magnitude of IFN-γ responses to GY9 epitope among the three individuals, aligning with the observations reported in previous studies (Day et al., 2001), which could be explained by difference in the absolute CD8^{+} T cell counts (Schmitt-Haendle et al., 2005), persistent immune activation and exhaustion and HAART (Addo et al., 2003; Booiman et al., 2017; Buggert et al., 2014). Although our participant with the smallest response magnitude had been on ART longer, other factors like CD8^{+} T cell counts may be a more important confounder in this study. Immune exhaustion is a functional impairment in T cells that occurs during chronic HIV infection (Booiman et al., 2017; Buggert et al., 2014). Exhausted T cells gradually lose their ability to produce effector cytokines like IL-2 and IFN-γ, leading to impaired immune control, persistence of viral replication, and HIV progression despite ART (Alrubayyi et al., 2022).

Furthermore, our study's absolute absence of IL-2 response is intriguing and may suggest a specific functional profile of the GY9-specific T cells. IL-2 is a crucial cytokine secreted by activated T cells that promotes T cell proliferation, differentiation, survival, cytolytic function and maintenance of memory T cell pools (Janas et al., 2005; Rochman et al., 2009). The absence of IL-2 responses could be attributed to several factors. First, the lack of IL-2 response may indicate a distinct functional phenotype of the responding T cells, potentially involved in effector functions rather than expansion (Ross & Cantrell, 2018). Second, is that the GY9-specific T cells in the study population may have reached a terminally differentiated state, where they have limited capacity for IL-2 production despite their effector functions (Booiman et al., 2017). This can result from prolonged HIV infection and antigen exposure, which was shown to recover upon successful virologic control (Behrens et al., 2018). Third, it could be the influence of ART on T cell function, which has been shown to modulate T cell responses (Rosás-Umbert et al., 2022). Long-term ART use impacts the functionality of HIV-specific T cells. Moreover, the absolute reduction in IL-2 production has detrimental effects on the proliferation of HIV-specific $CD8^+$ T cells (Zimmerli et al., 2005); this event may also account for the variable IFN-γ response. Besides, Mkhwanazi and colleagues showed that during HIV progression, there is a loss of polyfunctional $CD8^+$ T cells and an increase in the monofunctional cells that could explain the absence of dual-secreting cells (Mkhwanazi et al., 2010). Therefore, a further investigation into the functional characteristics of the GY9-specific T cells and the factors influencing IL-2 production could provide deeper insights into the dynamics of HLA-C-induced T cell-mediated immune responses against HIV.

Overall, our findings are consistent with previous studies showing clade-specific T cell responses against HIV-1 epitopes (Frahm et al., 2008; Hraber et al., 2014). HIV is known for its genetic diversity, with various clades circulating in different regions. This genetic diversity can influence epitope presentation and T cell recognition, leading to clade-specific immune responses. Several studies have demonstrated that HLA-C alleles, such as HLA-C*12:02, significantly shape T cell responses against HIV by presenting immunodominant epitopes (Chikata et al., 2019). Identifying a clade-specific HLA-C*03:02-restricted IFN-γ response to the GY9 epitope adds to the growing body of evidence on the clade-specificity of T cell responses in HIV infection. Further, it highlights the importance of considering clade-specific HLA-C-restricted epitopes in vaccine design and immunotherapeutic strategies.

5.4 HLA-C alleles: A complementary mechanism for HIV control

HLA-C molecules have emerged as attractive targets for vaccine development because of distinctive attributes that distinguish them from HLA-A and HLA-B molecules. Here, we discuss why HLA-C molecules generally and particularly HLA-C*03:02, may offer some advantages as targets compared to HLA-A and HLA-B molecules for HIV control.

A significant difference between HLA-C and HLA-A/B molecules is their interaction with the HIV-1 Nef protein (Figure 5-1). HIV-1 Nef has evolved to selectively downregulate the surface expression of HLA-A and HLA-B (Cohen et al., 1999), leading to immune evasion via reduced recognition of infected cells by $CD8^+$ T cells. On the other hand, HLA-C molecules, including HLA-C*03 alleles, are less efficiently targeted by Nef, resulting in relatively higher cell surface expression levels (Cohen et al., 1999; Mahiti et al., 2016; Mann et al., 2013) which have been demonstrated to correlate with HIV progression (Apps et al., 2013). Furthermore, Cohen and colleagues demonstrated that Nef interacts explicitly with the HLA-A/B cytoplasmic tail, with a specific binding mediated by Tyr321 and Asp328 residues. As a result, the relatively compensatory high surface expression of HLA-C alleles allows them to predominantly present HLA-C-restricted epitopes in the absence of other HLA molecules (Papúchová et al., 2019; Parolini et al., 2018), which forms the basis for their significant role in HIV control. Moreover, Makadzange and colleagues showed that HLA-C-restricted CTL responses account for more than 50% of the total CTL response, further underlining the role of HLA-C in HIV control (Blais et al., 2012; Makadzange et al., 2010).

Additionally, regulating HLA-C cell surface expression has been linked to a distinct genetic variant located 35kb upstream of HLA-C alleles (Blais et al., 2012; Corrah et al., 2011). Thomas and colleagues established that the -35C (rs9264942) allele was linked to elevated HLA-C cell surface expression and a more gradual course of HIV progression (Thomas et al., 2009). They also demonstrated that the C allele at -35kb is in strong LD with HLA-C*03:02, which may individually explain its significance in HIV control, although the underlying mechanism has not been established (Corrah et al., 2011). Although, this LD is strongest with HLA-C*07, there is a higher inter-individual variation with other HLA-C alleles however this remains unknown in African populations (Bettens et al., 2014). However, Kulkarni et al. later characterized a variant (indel) situated downstream of the HLA-C allele's stop codon in the 3'

untranslated region (UTR) at position 263 (rs67384697) as a target for microRNA(miR)-148a binding (Figure 5-1). This binding regulates HLA-C post-transcriptional regulation and protein expression in individuals of European American descent and occurs in high LD with -35 SNP (Kulkarni et al., 2011, 2013). Specifically, an insertion at position 263 (263ins) was correlated with decreased HLA-C protein surface expression due to miR-148a binding (Corrah et al., 2011). While not within the scope of this thesis, investigating the correlation of these variants in HLA-C*03:02 individuals, an allele with a low frequency (0.002) in Western populations, could yield exciting insights. Although the 263 indel and -35 SNP exhibit strong LD within some European populations, their LD appears to be relatively weak in African and Swiss populations (Bettens et al., 2014; Gentle et al., 2013; Kulkarni et al., 2011). This observation implies the possibility of distinct regulatory mechanisms governing HLA-C expression in our population. Given the genetic diversity in African populations, it remains unknown if other miRNAs (e.g., miR-148b, miR-152, miR-657) or genetic variants may have an impact (positive or negative) on HLA-C expression in African populations (Kulkarni et al., 2011). We suggest that similar mechanisms may contribute to HIV control mediated by HLA-C*03:02 in our populations. These observations prompt inquiry into potential strategies for augmenting HLA-C surface expression. A comprehensive study by Schaefer et al. demonstrated that HLA-C surface expression is also influenced by a dihydrophobic target signal (^{333}DXSLI^{337}) in the cytoplasmic tail that promotes intracellular localization (Schaefer et al., 2008). Most importantly, they showed that targeting this region by mutating isoleucine at position 337 to threonine (I337T) increases surface expression. This data offers hypothetical insights into prospective interventions that could be utilized to enhance the surface expression of HLA-C molecules.

HLA-C molecules engage with specific KIRs on NK cells, modulating NK cell activity and influencing a broader immune response (Kulpa & Collins, 2011; Pende et al., 2019; Vivier et al., 2008). These HLA-C molecules are categorized as C1 or C2 based on the presence of asparagine or lysine at position 80, respectively (Kulpa & Collins, 2011; Pende et al., 2019). The C1 subgroup functions exclusively as cognate ligands presenting self or endogenous (viral) peptides to inhibitory KIR2DL2/3 receptors, triggering NK cells through the ITAM-based signaling pathway, which leads to inhibition of NK cell cytotoxicity (Pende et al., 2019; van Teijlingen et al., 2014). HLA-C*03:02 belongs within the C1 group; however, it has been

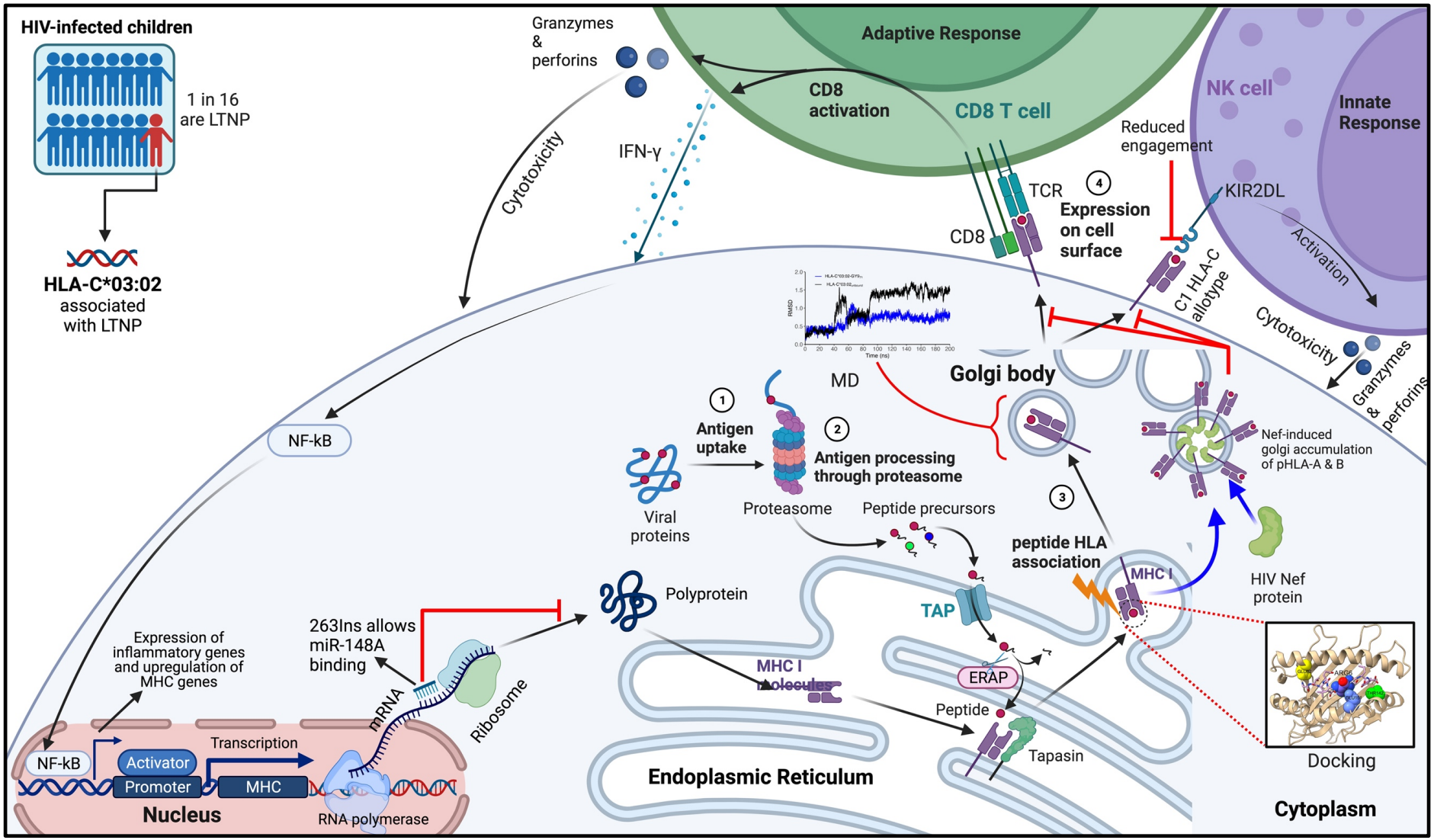

HIV-infected children
1 in 16 are LTNP
HLA-C*03:02 associated with LTNP
Granzymes & perforins
Cytotoxicity
IFN-γ
Adaptive Response
CD8 activation
CD8 T cell
NK cell
Innate Response
Reduced engagement
TCR
CD8
4
Expression on cell surface
KIR2DL
C1 HLA-C allotype
Activation
Cytotoxicity
Granzymes & perforins
MD
Golgi body
NF-kB
1
Antigen uptake
2
Antigen processing through proteasome
Viral proteins
Proteasome
Peptide precursors
3
peptide HLA association
MHC I
Nef-induced golgi accumulation of pHLA-A & B
HIV Nef protein
Polyprotein
TAP
ERAP
MHC I molecules
Peptide
Tapasin
Endoplasmic Reticulum
Docking
Cytoplasm
263Ins allows miR-148A binding
mRNA
Ribosome
Expression of inflammatory genes and upregulation of MHC genes
Transcription
NF-kB
Activator
Promoter
MHC
Nucleus
RNA polymerase

Figure 5-1: A model of the HLA-C antigen presentation, it's regulatory mechanisms and immune responses

This figure illustrates the complex process of HLA-C antigen presentation, highlighting key steps and regulatory mechanisms that influence intracellular transcription and surface expression. The figure shows that the proposed immune responses responsible for HIV control. Inset are images of some of the areas that were assessed in this study. Created with BioRender.com

demonstrated that engagement of inhibitory KIR in chronic viral infections rather enhances CTL responses (Boelen et al., 2018) and, therefore provides a plausible mechanism to HLA-C*03:02-mediated control of HIV. Multiple studies have established that C1 alleles such as HLA-C*14:03, C*03:04, and C*12:02 presumably in complex with HIV-1 epitopes serve as KIR2DL ligands, and they have been linked to the delayed progression of HIV (Lin et al., 2016; Tiemessen et al., 2011; Zappacosta et al., 1997; Ziegler et al., 2020). Moreover, Ziegler and co-authors presented compelling evidence that HIV-derived peptides presented by a homologous C1 allele (HLA-C*03:04) resulted in reduced interaction with inhibitory KIR2DL3 on NK cells (Ziegler et al., 2020). The downstream effects of attenuated inhibitory KIR signaling subsequently lead to reduced KIR-mediated inhibition of NK functionality, enhancing direct cytotoxicity against virally infected cells. Indeed, upregulation of NK cell receptors has been documented in individuals with HIV infection, particularly in the pre-HAART period, further substantiating their pivotal role in the immune response (Mavilio et al., 2003). Körner and colleagues have demonstrated that NK cells possess the capacity to detect Vpu-mediated HLA-C downmodulation, characterized by reduced binding of inhibitory KIR receptors to HLA-C, thereby responding by enhancing their antiviral activity (Körner et al., 2017). Therefore, the NK cell-mediated lysis offers an alternative mechanism to explain the HIV control associated with HLA-C*03:02 alleles within our populations. Consequently, focusing research efforts on HLA-C molecules within the context of HIV cure and vaccine development could harness these regulatory interactions, potentially refining and enhancing immune responses and ultimately enhancing vaccine efficacy.

Lastly, upon examining the electrostatic potentials and closely examining the peptide binding groove of the HLA-C*03:02 molecule, we observe similar characteristics shared with HLA-B*57:03 and HLA-B*58:01 molecules (Figure 5-2). Like these molecules, HLA-C*03:02 also

possesses many tyrosine residues within the peptide binding groove (Stewart-Jones et al., 2005). Tyrosine is a large non-polar hydrophobic amino acid with an aromatic R group. The HLA-C*03:02 molecule harbors four tyrosine residues at positions 6, 8, 98, and 158, constituting a comparable tyrosine bed akin to that found in HLA-B*57:03. We hypothesize that this cluster might perform a similar function to its counterpart in HLA-B*57:03, where it was experimentally observed to create hydrogen bond networks with bound peptides and

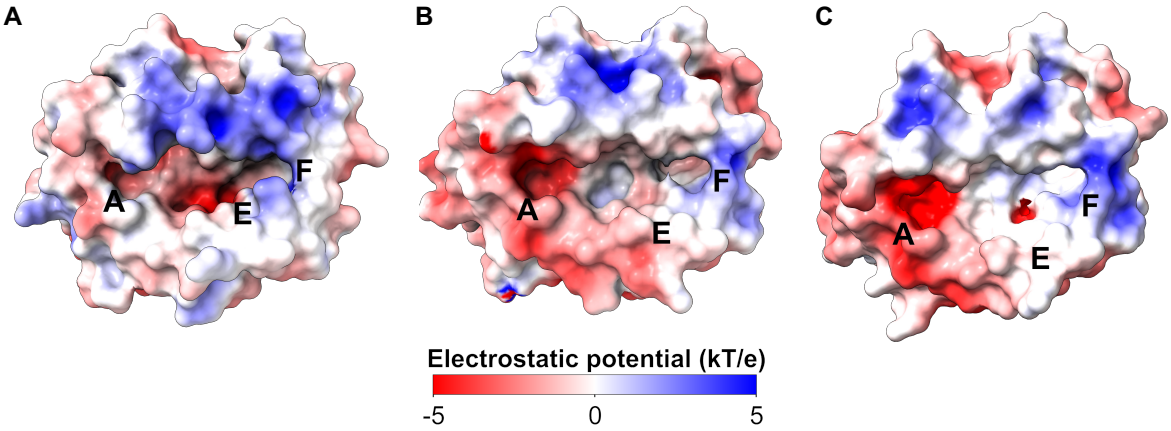


Figure 5-2: Structure and surface electrostatic potentials of HLA molecules

*The electrostatic potential of (A) HLA-C*03:02 (B) HLA-B*57:03 [PDB 5vvp], and (C) HLA-B*58:01 [PDB 5v5l] molecules were calculated using the APBS (Adaptive Poisson-Boltzmann Solver) server using the Amber forcefield to add charges in each system* (Jurrus et al., 2018). *Color scales range from red (-5kT/e) for electronegative potentials to blue (5kT/e) for electropositive potentials and neutral potentials (0 kT/e) are colored white (k is the Boltzmann constant and T is the temperature).*

sequester water molecules (Stewart-Jones et al., 2005). Indeed, this may be validated because when we performed a CAS on these tyrosine residues, we did not find a significant change in binding free energy, which could be associated with the lack of modeling of explicit water molecules in our study, which has been shown to improve the prediction of good binders (Lu et al., 2019)

Furthermore, our investigation reveals that HLA-C*03:02 exhibits a similar electrostatic potential to HLA-B*57:03 and B*58:01 in both the peptide binding groove and the TCR binding platform (Figure 5-2). These three alleles share a common feature of an electronegative A pocket and a neutral F pocket. While both HLA-B*58:01 and C*03:02 possess an electronegative E pocket, pocket HLA-B*57:03 appears neutral. The electronegative and

electropositive properties of the distal, proximal, and middle platforms on the α2 chain relevant to TCR binding are consistent across all three alleles. Analogous electrostatic potentials are also present on the α1 chain engaged in TCR binding. Given these structural equivalences, HLA-C*03:02 might share functional attributes with HLA-B*57:03 and B*58:01 in the context of LTNP outcomes.

In summary, HLA-C molecules exhibit distinct attributes that render them appealing candidates for vaccine design compared to HLA-A and HLA-B molecules. Their expression profiles, regulatory interactions, and potential capacity for cross-presentation of peptides offer benefits for formulating vaccines with broader epitope inclusivity, augmented immune reactions, and the potential to address diverse viral strains and escape mutants (Apps et al., 2013; Kulkarni et al., 2011; Kulpa & Collins, 2011; Parolini et al., 2018; Thomas et al., 2009). Exploiting the features of HLA-C in vaccine design may contribute to developing more effective vaccines against infectious diseases, including challenging pathogens like HIV.

5.5 Study Limitations

Some limitations of our study warrant acknowledgment. We conducted a retrospective study, underscoring that our findings warrant cautious interpretation; regrettably, analogous studies have been precluded by the Test-and-Treat Strategy. Nevertheless, the heterogeneous response to ART has been extensively recorded (Fabbiani et al., 2009), underscoring the necessity for endeavors to determine who and why in order to enhance clinical outcomes. Although our analysis of time-to-progression factors was limited in scope, we successfully validated previous research findings and offered insights into observed distinctions among populations in two African countries.

Survival bias within this cohort might theoretically enhance the prevalence of protective genetic variants. Nevertheless, using a case-control study design of extreme phenotyping and leveraging large sample sizes, we showed HLA class I alleles with a notable impact on HIV control.

Water molecules have been shown to stabilizes epitope binding within the middle region of the peptide binding groove (Lu et al., 2019); therefore, molecular docking protocols are encourage to model water molecules. While our docking protocol did not account for flexible docking

involving explicit water molecules, the dominant epitopes identified through this protocol exhibited $CD8^+$ T cell responses restricted by HLA-C.

Our functional study focused solely on the effector and proliferative functions of $CD8^+$ T cells, as indicated by the measurement of dual IFN-γ and IL-2 production. However, it is worth noting some limitation of targeting IL-2, it has been previously demonstrated that the production of IL-2 by memory CD8 T cells is highly variable and selects for only a small subset of antigen-specific CD8 T cells (Kahan et al., 2022). Also, it's important to note that HLA-C alleles also interact with NK cells via KIR, which results in the cytotoxic targeting of virally infected cells. In a recent study conducted by Ziegler et al., it was shown that within HIV-infected $CD4^+$ T cells, the binding of pHLA-C*03:04 with the inhibitory KIR2DL3 is weaker compared to non-infected $CD4^+$ T cells (Ziegler et al., 2020). This data suggests that the activation pathway of NK cells in the immune response becomes significantly more active in HIV infection upon interaction with stable peptides within HLA-C molecules linked to delayed disease progression. It's plausible that a significant portion of responses remains unaccounted for through cytotoxic pathways in our study.

In summary, we elucidate the potential role of the recently identified HLA-C*03:02 in HIV control and detail their structural and functional adaptations. Although the role of HLA-C alleles in controlling HIV infection is not well documented, we provide empirical evidence for HLA-C*03:02 that supports its prior association with a low viral load set point and a high CD4 count (Leslie et al., 2010), which could biologically underlie its favorable function in HIV LTNP.

5.6 Concluding Remarks

HIV-1 continues to pose a substantial public health challenge, particularly in Africa, the region most heavily impacted. Despite the implementation of programs targeting the prevention of mother-to-child transmission of HIV, over a quarter of a million children are infected with HIV through perinatal transmission each year. The development of a successful vaccine remains a pivotal objective in the global endeavor to combat the HIV pandemic. LTNPs offer a certain prospect to uncover essential immune mechanisms that could be exploited in the advance of new vaccines and immunotherapies. Amidst the renewed interest in peptide-based vaccine

development, the emphasis remains on harnessing the classical function of HLA class I alleles, which present intracellularly processed antigens on cell surfaces to induce CTL responses. In this thesis, a potential association between the HLA-C*03:02 allele and long-term HIV control among African children was identified. Furthermore, our study revealed the preferential presentation of the GY9 nanomer peptide by HLA-C*03:02. Despite the moderate IFN responses to the GY9 epitope, we established the specificity of GY9 responses to clade and HLA-C*03:02. This work adds to the growing understanding of HLA-C molecules' role in HIV control, holding implications for the design of global vaccines and immunotherapies aimed at curbing the HIV epidemic.

CHAPTER 6: CONCLUSION

In this chapter, we present a conclusion from the principal research findings of the Ph.D. research objectives and questions, highlighting their significance and contribution to the field. In this thesis, we tackled the issue of limited studies estimating the frequency of LTNP in children infected with HIV, particularly in African populations. We investigated the protective impact of HLA class I molecules and conducted functional characterization of HLA-C in controlling HIV-1 infection. The unique strength of this thesis lies in its multidisciplinary approach, combining epidemiology, biostatistics, bioinformatics, genetics, computational chemistry, and immunological methodologies to study HIV disease progression in children. Consequently, the insights gained from this study can serve as a valuable model for addressing similar disease conditions in Africa.

Firstly, this thesis sought to ascertain the frequency of LTNP and identify the risk factors linked to HIV disease progression in perinatal infection. We noted a relatively small percentage of children meeting the LNTP status, with this proportion displaying variability between Uganda and Botswana. The thesis further highlights the significant associations between stunting, wasting, year of enrolment into care, and country of origin with HIV disease progression among children with perinatal HIV infection. These findings are significant because they reinforce the importance of addressing malnutrition and timely enrollment into care to enhance clinical outcomes in this vulnerable population. Moreover, the study emphasizes the need for tailored interventions that consider regional variations such as HIV subtypes and healthcare access, providing valuable insights for developing effective strategies to mitigate HIV outcomes among children.

Furthermore, we aimed to investigate the association between HLA class I alleles and LTNP. Our study reveals a notable enrichment of population-specific HLA class I alleles and haplotypes in children infected with HIV in Uganda and Botswana. We provide evidence for the association of known HLA-B alleles (B*57:03 and B*58:01) and a novel HLA-C*03:02 allele with LTNP in Ugandan and Batswana children infected with HIV. These results are consistent with prior investigations concerning the role of HLA-B alleles and introduce a novel component to our knowledge of HLA-C in HIV immune control. Observing an HLA-C allele is important because previous studies have been at odds with the role of these molecules in

HIV control. However, these genetic factors collectively explain only ~17% of the variability in LTNP. Therefore, understanding the independent and additive protective effects of HLA-C*03:02 may inform the development of targeted therapeutic strategies, such as multi-epitope HLA-based vaccines, to improve immune responses in HIV-vulnerable populations. But most importantly, the modest contribution of the identified HLA alleles to LTNP underscores the complexity of HIV immune control and the multifactorial nature of LTNP. These findings provide a basis for further investigations into the interplay of various genetic and immunological factors contributing to HIV immune control. Such knowledge is crucial for developing targeted therapeutic strategies and vaccines for HIV.

Finally, we investigated the structural and functional characteristics of the HLA-C*03:02 molecule within a population predominantly infected with HIV-1 clades A and C. As expected, we found that HIV-1 clades A and C have shared and private HLA-C*03:02 epitopes. Based on immunoinformatics studies, HLA-C*03:02 molecules prefer to present peptides derived from structural HIV proteins, where the GY9 peptide derived from HIV-1 matrix protein forms the most stable pHLA complex. We show that specific residues in HLA-C*03:02 and GY9 determine the epitope specificity. Furthermore, our study reveals that the GY9 epitope elicits a clade-specific HLA-C*03:02-restricted IFN-γ response in one in three children infected with HIV on ART and expressing the molecules. The differential response to GY9 epitope shows a critical dimension to the functional profile of GY9-specific $CD8^+$ T cells. This indicates potential differences in T cell functionality and potential role for effector functions, including via interaction with HLA-C ligands the KIRs on NK cells.

Generally, population-specific studies like ours advance our knowledge of the complex host-virus interactions and provide valuable insights into the immunogenetic mechanisms governing HIV infection, including the differential response to HIV therapeutics and vaccines, paving the way for novel therapeutic interventions. The variability in GY9-specific responses among study participants underscores the complex interplay of genetic, immunological, and viral factors that shape the immune response to HIV epitopes. These findings have implications for designing immunotherapies and vaccines targeting clade-specific epitopes to enhance T cell-mediated immunity and control viral replication in individuals living with HIV.

CHAPTER 7: RECOMMENDATIONS AND FUTURE DIRECTIONS

7.1 Recommendations

When the doctoral studies began in 2016, the decision to start ART was based on specific immunologic and clinical criteria. This led us to examine other factors that might further improve how we optimize the decision when to begin ART. Starting from 2009, the WHO revised its approach to treating HIV, moving to the Test-and-Treat Strategy. The goal was to provide more than 90% of people with HIV access to ART. However, even with this change, there are substantial differences in how children respond to ART include drug hypersensitivity reactions. We suggest that our methods could be used to investigate the reasons or factors why children respond differently to ART. These studies might assist in making ART better for children and enhancing their treatment outcomes while also considering any unique variations within the population.

The results of our research indicate that specific genetic markers might offer natural protection against HIV disease progression. Exploring the development of precise treatments that mimic these protective influences holds the potential for discovering novel therapies and vaccines thus improving treatment outcomes and reducing the impact of HIV-associated morbidity. Including genetic components in public health communication can give individuals insights into how they might respond to ART, which could enhance commitment to ART and inspire active participation in testing and developing novel treatment strategies.

7.2 Future Directions

The first urgent requirement is to perform additional functional studies to validate our genetic associations. We propose cloning the HLA-C*03:02 gene into a lentiviral vector incorporating a green fluorescent protein (GFP) tag, followed by HEK293 T cell transfection. Transfected 721 cells can be employed to facilitate the evaluation of the following aspects subsequently: (i) the expression of HLA-C*03:02 molecules on the cellular surface and their localization in the endoplasmic reticulum, (ii) conducting cytotoxicity assays, and (iii) performing binding affinity assays (Kim et al., 2007; Y. Liu et al., 2023; Sidney et al., 2018). An alternative strategy involves generating HLA-C monomers within bacterial cell systems (Vaurs et al., 2021),

followed by their refolding with the corresponding GY9 epitope to generate pHLA tetramers. These tetramers can be employed to achieve similar research objectives.

The implications of the findings in this thesis extend beyond the initial observations here. Exploring the entire surfaceome/peptidome of $CD4^+$ T cells expressing protective HLA-C alleles with immunogenic ligands could reveal potential targets for immunotherapeutic interventions (K. Pandey et al., 2023; Ziegler et al., 2020). Moreover, investigating distinctive HLA-C*03:02 epitope-specific CTL responses, including aspects such as proliferation and polyfunctionality (such as cytotoxic degranulation and multiple cytokine production), is recommended. The breakthroughs observed in chimeric antigen receptor (CAR) T cell therapy within cancer treatment have transformative implications. Advancing CAR T cell therapeutics for HIV necessitates a comprehensive description of the protective HLA-HIV peptide repertoire.

A vital aspect involves the direct exploration of HLA-C*03:02's role in promoting cytotoxicity within HIV-infected cells. Understanding the specific mechanisms by which this HLA allele influences cytotoxic responses against the virus could unveil critical insights into immune-mediated control of HIV (Lo Nigro et al., 2019; Vollmers et al., 2021). Additionally, an essential avenue for future studies revolves around elucidating the interactions between HLA-C*03:02 and NK cells. Investigating these interactions comprehensively would shed light on how this particular HLA variant modulates NK cell responses, potentially uncovering novel avenues for therapeutic interventions aimed at bolstering immune defenses against HIV.

A challenge in advancing HLA-C class I molecules as lead targets for vaccine design lies in their low surface expression, particularly in non-APCs. Schaefer and colleagues have revealed that introducing di-hydrophobic signals within the cytoplasmic tail induces HLA-C surface expression by triggering intracellular degradation (Schaefer et al., 2008). CRISPR-Cas9 investigations focused on these mechanisms could potentially enhance HLA-C surface expression, unveiling its potential benefits in the context of HIV control.

Finally, we propose that our understanding of selective immunogenic HLA-based surface presentation could open new avenues for research. In this context, we suggest the term "*schooling the proteasome*," signifying the potential to direct the proteasomal machinery to

prioritize or target highly immunogenic regions within the HIV proteome. Progress in these innovative research domains may be relevant for similar viral infections and cancer studies.

CHAPTER 8: REFERENCES

A. Rakityanskaya, I., S. Ryabova, T., A. Kalashnikova, A., S. Balasaniants, G., D. Kaprin, A., I. Ershov, F., V. Kir'janova, V., B. Korzhenevskaja, T., V. Barbinov, D., V. Ignatovskij, A., Y. Grivtsova, L., G. Isaeva, V., A. Falaleeva, N., I. Gil', A., A. Berns, S., V. Vasil'eva, N., V. Dolgo-Saburova, J., V. Shagdileeva, E., V. Frolova, E., & S. Astanina, N. (2022). Perspective Chapter: The Role of Interferon Gamma in Clinical Medicine. In *Basic and Clinical Aspects of Interferon Gamma*. IntechOpen. https://doi.org/10.5772/intechopen.105476

Abdelrahman, T., Hughes, J., Main, J., Mclauchlan, J., Thursz, M., & Thomson, E. (2015). Next-generation sequencing sheds light on the natural history of hepatitis C infection in patients who fail treatment. *Hepatology*, *61*(1), 88–97. https://doi.org/10.1002/hep.27192

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Adamson, C. S., & Freed, E. O. (2010). Novel Approaches to Inhibiting HIV-1 Replication. *Antiviral Res*, *85*(1), 1–55. https://doi.org/10.1016/j.antiviral.2009.09.009.Novel

Addo, M. M., Yu, X. G., Rathod, A., Cohen, D., Eldridge, R. L., Strick, D., Johnston, M. N., Corcoran, C., Wurcel, A. G., Fitzpatrick, C. A., Feeney, M. E., Rodriguez, W. R., Basgoz, N., Draenert, R., Stone, D. R., Brander, C., Goulder, P. J. R., Rosenberg, E. S., Altfeld, M., & Walker, B. D. (2003). Comprehensive Epitope Analysis of Human Immunodeficiency Virus Type 1 (HIV-1)-Specific T-Cell Responses Directed against the Entire Expressed HIV-1 Genome Demonstrate Broadly Directed Responses, but No Correlation to Viral Load. *Journal of Virology*, *77*(3), 2081–2092. https://doi.org/10.1128/JVI.77.3.2081-2092.2003

Adland, E., Paioni, P., Thobakgale, C., Laker, L., Mori, L., Muenchhoff, M., Csala, A., Clapson, M., Flynn, J., Novelli, V., Hurst, J., Naidoo, V., Shapiro, R., Huang, K. H. G., Frater, J., Prendergast, A., Prado, J. G., Ndung'u, T., Walker, B. D., … Goulder, P. J. R.

(2015). Discordant Impact of HLA on Viral Replicative Capacity and Disease Progression in Pediatric and Adult HIV Infection. *PLoS Pathogens*, *11*(6), 1–26. https://doi.org/10.1371/journal.ppat.1004954

Allen, W. J., Balius, T. E., Mukherjee, S., Brozell, S. R., Moustakas, D. T., Lang, P. T., Case, D. A., Kuntz, I. D., & Rizzo, R. C. (2015). DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, *36*(15), 1132–1156. https://doi.org/10.1002/jcc.23905

Alrubayyi, A., Moreno-Cubero, E., Hameiri-Bowen, D., Matthews, R., Rowland-Jones, S., Schurich, A., & Peppa, D. (2022). Functional Restoration of Exhausted CD8 T Cells in Chronic HIV-1 Infection by Targeting Mitochondrial Dysfunction. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/fimmu.2022.908697

Alter, I., Gragert, L., Fingerson, S., Maiers, M., & Louzoun, Y. (2017). HLA class I haplotype diversity is consistent with selection for frequent existing haplotypes. *PLoS Computational Biology*, *13*(8), e1005693. https://doi.org/10.1371/journal.pcbi.1005693

Altmann, M., an der Heiden, M., Scheufele, R., Hartmann, K., Houareau, C., Bartmeyer, B., & Hamouda, O. (2012). The risk of AIDS-defining events is decreasing over time in the German HIV-1 Seroconverter Cohort. *BMC Infectious Diseases*, *12*(1), 94. https://doi.org/10.1186/1471-2334-12-94

Amornkul, P. N., Karita, E., Kamali, A., Rida, W. N., Sanders, E. J., Lakhi, S., Price, M. A., Kilembe, W., Cormier, E., Anzala, O., Latka, M. H., Bekker, L. G., Allen, S. A., Gilmour, J., & Fast, P. E. (2013). Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. *AIDS*, *27*(17), 2775–2786. https://doi.org/10.1097/QAD.0000000000000012

Anderson, E. M., & Maldarelli, F. (2018). The role of integration and clonal expansion in HIV infection: live long and prosper. *Retrovirology*, *15*(1), 71. https://doi.org/10.1186/s12977-018-0448-8

Antonio George Lentoor. (2018). Nutritional Status of Perinatally HIV-infected Children on

Antiretroviral Therapy from a Resource-poor Rural South African Community. *African Journal of Medical and Health Sciences*, *17*, 1–6. https://doi.org/10.4103/ajmhs.ajmhs

Antunes, D. A., Moll, M., Devaurs, D., Jackson, K. R., Lizée, G., & Kavraki, L. E. (2017). DINC 2.0: A new protein-peptide docking webserver using an incremental approach. *Cancer Research*, *77*(21), e55–e57. https://doi.org/10.1158/0008-5472.CAN-17-0511

Aouizerat, B. E., Pearce, C. L., & Miaskowski, C. (2011). The Search for Host Genetic Factors of HIV/AIDS Pathogenesis in the Post-Genome Era: Progress to Date and New Avenues for Discovery. *Current HIV/AIDS Reports*, *8*(1), 38–44. https://doi.org/10.1007/s11904-010-0065-1

Apps, R., Qi, Y., Carlson, J. M., Chen, H., Gao, X., Thomas, R., Yuki, Y., Del Prete, G. Q., Goulder, P., Brumme, Z. L., Brumme, C. J., John, M., Mallal, S., Nelson, G., Bosch, R., Heckerman, D., Stein, J. L., Soderberg, K. A., Moody, M. A., … Carrington, M. (2013). Influence of HLA-C expression level on HIV control. *Science*, *340*(6128), 87–91. https://doi.org/10.1126/science.1232685

Arhel, N. (2010). Revisiting HIV-1 uncoating. *Retrovirology*, *7*(96), 1–10.

Arora, J., Pierini, F., McLaren, P. J., Carrington, M., Fellay, J., & Lenz, T. L. (2020). HLA Heterozygote Advantage against HIV-1 Is Driven by Quantitative and Qualitative Differences in HLA Allele-Specific Peptide Presentation. *Molecular Biology and Evolution*, *37*(3), 639–650. https://doi.org/10.1093/molbev/msz249

Aspord, C., Charles, J., Leccia, M.-T., Laurin, D., Richard, M.-J., Chaperot, L., & Plumas, J. (2010). A Novel Cancer Vaccine Strategy Based on HLA-A*0201 Matched Allogeneic Plasmacytoid Dendritic Cells. *PLoS ONE*, *5*(5), e10458. https://doi.org/10.1371/journal.pone.0010458

Attaf, M., Malik, A., Severinsen, M. C., Roider, J., Ogongo, P., Buus, S., Ndung'u, T., Leslie, A., Kløverpris, H. N., Matthews, P. C., Sewell, A. K., & Goulder, P. (2018). Major TCR repertoire perturbation by immunodominant HLA-B*44: 03 restricted CMV-specific T cells. *Frontiers in Immunology*, *9*(2539). https://doi.org/10.3389/fimmu.2018.02539

Babiker, A., Darbyshire, J., Pezzotti, P., Porter, K., Rezza, G., Walker, S. A., Beral, V., Coutinho, R., Amo, J. Del, Gill, N., Lee, C., Meyer, L., Tyrer, F., Dabis, F., Thiebaut, R., F, S. L.-A., Angelis, D. De, & CASCADE collaboration. (2002). Changes over calendar time in the risk of specific first AIDS-defining events following HIV seroconversion, adjusting for competing risks. *International Journal of Epidemiology*, *31*(5), 951–958. https://doi.org/10.1093/ije/31.5.951

Backe, E., Jimenez, E., Unger, M., Schafer, A., Jauniaux, E., & Vogel, M. (1992). Demonstration of HIV-1 infected cells in human placenta by in situ hybridisation and immunostaining. *Journal of Clinical Pathology*, *45*(10), 871–874. https://doi.org/10.1136/jcp.45.10.871

Baeten, J. M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R. S., Certain, L., Mandaliya, K., Jaoko, W., & Overbaugh, J. (2007). HIV-1 Subtype D Infection Is Associated with Faster Disease Progression than Subtype A in Spite of Similar Plasma HIV-1 Loads. *The Journal of Infectious Diseases*, *195*(8), 1177–1180. https://doi.org/10.1086/512682

Bai, Y., Ni, M., Cooper, B., Wei, Y., & Fury, W. (2014). Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics*, *15*(1), 1–16. https://doi.org/10.1186/1471-2164-15-325

Ball, T. B., Ji, H., Kimani, J., McLaren, P., Marlin, C., Hill, A. V., & Plummer, F. A. (2007). Polymorphisms in IRF-1 associated with resistance to HIV-1 infection in highly exposed uninfected Kenyan sex workers. *AIDS*, *21*(9), 1091–1101. https://doi.org/10.1097/QAD.0b013e3280ef6ae1

Bartha, I., Carlson, J. M., Brumme, C. J., McLaren, P. J., Brumme, Z. L., John, M., Haas, D. W., Martinez-Picado, J., Dalmau, J., López-Galíndez, C., Casado, C., Rauch, A., Günthard, H. F., Bernasconi, E., Vernazza, P., Klimkait, T., Yerly, S., O'Brien, S. J., Listgarten, J., … Fellay, J. (2013). A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *ELife*, *2*(2), 1–16. https://doi.org/10.7554/eLife.01123

Barugahare, B., Baker, C., Aluoch, O. K., Donovan, R., Elrefaei, M., Eggena, M., Jones, N.,

Mutalya, S., Kityo, C., Mugyenyi, P., & Cao, H. (2005). Human Immunodeficiency Virus-Specific Responses in Adult Ugandans: Patterns of Cross-Clade Recognition. *Journal of Virology*, *79*(7), 4132–4139. https://doi.org/10.1128/JVI.79.7.4132

Barugahare, B., Baker, C., K'Aluoch, O., Donovan, R., Elrefaei, M., Eggena, M., Jones, N., Mutalya, S., Kityo, C., Mugyenyi, P., Cao, H., Aluoch, O. K., Donovan, R., Elrefaei, M., Eggena, M., Jones, N., Mutalya, S., Kityo, C., Mugyenyi, P., & Cao, H. (2005). Human Immunodeficiency Virus-Specific Responses in Adult Ugandans: Patterns of Cross-Clade Recognition. *Journal of Virology*, *79*(7), 4132–4139. https://doi.org/10.1128/JVI.79.7.4132-4139.2005

Bauer, D. C., Zadoorian, A., Wilson, L. O. W., & Thorne, N. P. (2018). Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Briefings in Bioinformatics*, *19*(2), 179–187. https://doi.org/10.1093/bib/bbw097

Baum, M. K., Campa, A., Lai, S., Sales Martinez, S., Tsalaile, L., Burns, P., Farahani, M., Li, Y., Van Widenfelt, E., Page, J. B., Bussmann, H., Fawzi, W. W., Moyo, S., Makhema, J., Thior, I., Essex, M., & Marlink, R. (2013). Effect of micronutrient supplementation on disease progression in asymptomatic, antiretroviral-naive, HIV-infected adults in Botswana: A randomized clinical trial. *Journal of the American Medical Association*, *310*(20), 2154–2163. https://doi.org/10.1001/jama.2013.280923

Bbosa, N., Kaleebu, P., & Ssemwanga, D. (2019). HIV subtype diversity worldwide. *Current Opinion in HIV and AIDS*, *14*(3), 153–160. https://doi.org/10.1097/COH.0000000000000534

Becerra, J. C., Bildstein, L. S., & Gach, J. S. (2016). Recent Insights into the HIV/AIDS Pandemic. *Microbial Cell*, *3*(9), 450–474. https://doi.org/10.15698/mic2016.09.529

Behrens, N. E., Wertheimer, A., Klotz, S. A., & Ahmad, N. (2018). Reduction in terminally differentiated T cells in virologically controlled HIV-infected aging patients on long-term antiretroviral therapy. *PLOS ONE*, *13*(6), e0199101. https://doi.org/10.1371/journal.pone.0199101

Berendam, S. J., Nelson, A. N., Goswami, R., Persaud, D., Haigwood, N. L., Chahroudi, A., Fouda, G. G., & Permar, S. R. (2020). Pediatric HIV: the Potential of Immune Therapeutics to Achieve Viral Remission and Functional Cure. *Current HIV/AIDS Reports*, *17*(3), 237–248. https://doi.org/10.1007/s11904-020-00495-1

Bernstein, H. B., Kinter, A. L., Jackson, R., & Fauci, A. S. (2004). Neonatal Natural Killer Cells Produce Chemokines and Suppress HIV Replication in Vitro. *AIDS Research and Human Retroviruses*, *20*(11), 1189–1195. https://doi.org/10.1089/aid.2004.20.1189

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., & MacKerell, A. D. (2012). Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. *Journal of Chemical Theory and Computation*, *8*(9), 3257–3273. https://doi.org/10.1021/ct300400x

Bettens, F., Brunet, L., & Tiercy, J.-M. (2014). High-allelic variability in HLA-C mRNA expression: association with HLA-extended haplotypes. *Genes & Immunity*, *15*(3), 176–181. https://doi.org/10.1038/gene.2014.1

Beyrer, C., Artenstein, A. W., Rugpao, S., Stephens, H., VanCott, T. C., Robb, M. L., Rinkaew, M., Birx, D. L., Khamboonruang, C., Zimmerman, P. A., Nelson, K. E., & Natpratan, C. (1999). Epidemiologic and Biologic Characterization of a Cohort of Human Immunodeficiency Virus Type 1 Highly Exposed, Persistently Seronegative Female Sex Workers in Northern Thailand. *The Journal of Infectious Diseases*, *179*(1), 59–67. https://doi.org/10.1086/314556

Blais, M.-E., Zhang, Y., Rostron, T., Griffin, H., Taylor, S., Xu, K., Yan, H., Wu, H., James, I., John, M., Dong, T., & Rowland-Jones, S. L. (2012). High Frequency of HIV Mutations Associated with HLA-C Suggests Enhanced HLA-C–Restricted CTL Selective Pressure Associated with an AIDS-Protective Polymorphism. *The Journal of Immunology*, *188*(9), 4663–4670. https://doi.org/10.4049/jimmunol.1103472

Boelen, L., Debebe, B., Silveira, M., Salam, A., Makinde, J., Roberts, C. h., Wang, E. C. Y., Frater, J., Gilmour, J., Twigger, K., Ladell, K., Miners, K. L., Jayaraman, J., Traherne, J.

A., Price, D. A., Qi, Y., Martin, M. P., Macallan, D. C., Thio, C. L., … Asquith, B. (2018). Inhibitory killer cell immunoglobulin-like receptors strengthen CD8 + T cell–mediated control of HIV-1, HCV, and HTLV-1. *Science Immunology*, *3*(29). https://doi.org/10.1126/sciimmunol.aao2892

Boender, T. S., Sigaloff, K. C. E., Kayiwa, J., Musiime, V., Calis, J. C. J., Hamers, R. L., Nakatudde, L. K., Khauda, E., Mukuye, A., Ditai, J., Geelen, S. P., Mugyenyi, P., Rinke de Wit, T. F., & Kityo, C. (2012). Barriers to Initiation of Pediatric HIV Treatment in Uganda: A Mixed-Method Study. *AIDS Research and Treatment*, *2012*, 1–10. https://doi.org/10.1155/2012/817506

Booiman, T., Wit, F. W., Girigorie, A. F., Maurer, I., De Francesco, D., Sabin, C. A., Harskamp, A. M., Prins, M., Franceschi, C., Deeks, S. G., Winston, A., Reiss, P., & Kootstra, N. A. (2017). Terminal differentiation of T cells is strongly associated with CMV infection and increased in HIV-positive individuals on ART and lifestyle matched controls. *PLOS ONE*, *12*(8), e0183357. https://doi.org/10.1371/journal.pone.0183357

Borghans, J. A. M., Mølgaard, A., de Boer, R. J., & Keşmir, C. (2007). HLA Alleles Associated with Slow Progression to AIDS Truly Prefer to Present HIV-1 p24. *PLoS ONE*, *2*(9), e920. https://doi.org/10.1371/journal.pone.0000920

Boudreau, J. E., & Hsu, K. C. (2018). Natural killer cell education in human health and disease. *Current Opinion in Immunology*, *50*, 102–111. https://doi.org/10.1016/j.coi.2017.11.003

Boulet, S., Ndongala, M. L., Peretz, Y., Boisvert, M.-P., Boulassel, M.-R., Tremblay, C., Routy, J.-P., Sekaly, R.-P., & Bernard, N. F. (2007). A dual color ELISPOT method for the simultaneous detection of IL-2 and IFN-γ HIV-specific immune responses. *Journal of Immunological Methods*, *320*(1–2), 18–29. https://doi.org/10.1016/j.jim.2006.11.010

Bourke, C. D., Jones, K. D. J., & Prendergast, A. J. (2019). Current Understanding of Innate Immune Cell Dysfunction in Childhood Undernutrition. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.01728

Bracq, L., Xie, M., Benichou, S., & Bouchet, J. (2018). Mechanisms for Cell-to-Cell

Transmission of HIV-1. *Frontiers in Immunology*, *9*(February), 1–14. https://doi.org/10.3389/fimmu.2018.00260

Bruzzi, P., Green, S. B., Byar, D. P., Brinton, L. A., & Schairer, C. (1985). Estimating the population attributable risk for multiple risk factors using case-control data. *American Journal of Epidemiology*, *22*(5), 904–914. https://doi.org/10.1093/oxfordjournals.aje.a114174

Buchbinder, S., & Vittinghoff, E. (1999). HIV-infected long-term nonprogressors: Epidemiology, mechanisms of delayed progression, and clinical and research implications. *Microbes and Infection*, *1*(13), 1113–1120. https://doi.org/10.1016/S1286-4579(99)00204-X

Buggert, M., Tauriainen, J., Yamamoto, T., Frederiksen, J., Ivarsson, M. A., Michaëlsson, J., Lund, O., Hejdeman, B., Jansson, M., Sönnerborg, A., Koup, R. A., Betts, M. R., & Karlsson, A. C. (2014). T-bet and Eomes Are Differentially Linked to the Exhausted Phenotype of CD8+ T Cells in HIV Infection. *PLoS Pathogens*, *10*(7), e1004251. https://doi.org/10.1371/journal.ppat.1004251

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., & Fernández-Viña, M. A. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Human Immunology*, *62*(9), 1009–1030. https://doi.org/10.1016/S0198-8859(01)00298-1

Cao, K., Moormann, A. M., Lyke, K. E., Masaberg, C., Sumba, O. P., Doumbo, O. K., Koech, D., Lancaster, A., Nelson, M., Meyer, D., Single, R., Hartzman, R. J., Plowe, C. V., Kazura, J., Mann, D. L., Sztein, M. B., Thomson, G., Fernández-Viná, M. A., & Fernández-Vina, M. A. (2004). Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens*, *63*(4), 293–325. https://doi.org/10.1111/j.0001-2815.2004.00192.x

Carlson, J. M., Brumme, C. J., Martin, E., Listgarten, J., Brockman, M. A., Le, A. Q., Chui, C. K. S., Cotton, L. A., Knapp, D. J. H. F., Riddler, S. A., Haubrich, R., Nelson, G., Pfeifer,

N., DeZiel, C. E., Heckerman, D., Apps, R., Carrington, M., Mallal, S., Harrigan, P. R., … Brumme, Z. L. (2012). Correlates of Protective Cellular Immunity Revealed by Analysis of Population-Level Immune Escape Pathways in HIV-1. *Journal of Virology*, *86*(24), 13202–13216. https://doi.org/10.1128/jvi.01998-12

Carlson, J. M., Listgarten, J., Pfeifer, N., Tan, V., Kadie, C., Walker, B. D., Ndung'u, T., Shapiro, R., Frater, J., Brumme, Z. L., Goulder, P. J. R., & Heckerman, D. (2012). Widespread Impact of HLA Restriction on Immune Control and Escape Pathways of HIV-1. *Journal of Virology*, *86*(9), 5230–5243. https://doi.org/10.1128/jvi.06728-11

Carlson, L., de Marco, A., Oberwinkler, H., Habermann, A., Briggs, J. A. G., Kräusslich, H.-G., & Grünewald, K. (2010). Cryo Electron Tomography of Native HIV-1 Budding Sites. *PLoS Pathogens*, *6*(11), e1001173. https://doi.org/10.1371/journal.ppat.1001173

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K., & O'Brien, S. J. (1999). HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. *Science*, *283*(5408), 1748–1752. https://doi.org/10.1126/science.283.5408.1748

Carrington, M., & O'Brien, S. J. (2003). The Influence of HLA Genotype on AIDS. *Annual Review of Medicine*, *54*(1), 535–551. https://doi.org/10.1146/annurev.med.54.101601.152346

Celik, A. A., Kraemer, T., Huyton, T., Blasczyk, R., & Bade-Döding, C. (2016). The diversity of the HLA-E-restricted peptide repertoire explains the immunological impact of the Arg107Gly mismatch. *Immunogenetics*, *68*(1), 29–41. https://doi.org/10.1007/s00251-015-0880-z

Chakraborty, S., Rahman, T., & Chakravorty, R. (2014). Characterization of the Protective HIV-1 CTL Epitopes and the Corresponding HLA Class I Alleles: A Step towards Designing CTL Based HIV-1 Vaccine. *Advances in Virology*, *2014*, 1–17. https://doi.org/10.1155/2014/321974

Chappell, P. E., Meziane, E. K., Harrison, M., Magiera, Ł., Hermann, C., Mears, L., Wrobel,

A. G., Durant, C., Nielsen, L. L., Buus, S., Ternette, N., Mwangi, W., Butter, C., Nair, V., Ahyee, T., Duggleby, R., Madrigal, A., Roversi, P., Lea, S. M., & Kaufman, J. (2015). Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *ELife*, *4*(4), 1–22. https://doi.org/10.7554/eLife.05345

Chatterjee, K. (2010). Host genetic factors in susceptibility to HIV-1 infection and progression to AIDS. *Journal of Genetics*, *89*(1), 109–116.

Chen, H., Piechocka-Trocha, A., Miura, T., Brockman, M. A., Julg, B. D., Baker, B. M., Rothchild, A. C., Block, B. L., Schneidewind, A., Koibuchi, T., Pereyra, F., Allen, T. M., & Walker, B. D. (2009). Differential Neutralization of Human Immunodeficiency Virus (HIV) Replication in Autologous CD4 T Cells by HIV-Specific Cytotoxic T Lymphocytes. *Journal of Virology*, *83*(7), 3138–3149. https://doi.org/10.1128/jvi.02073-08

Chen, K. Y., Liu, J., & Ren, E. C. (2012). Structural and functional distinctiveness of HLA-A2 allelic variants. *Immunologic Research*, *53*(1–3), 182–190. https://doi.org/10.1007/s12026-012-8295-5

Chen, L., Yang, J., Zhang, R., Xu, Y., Zheng, J., Jiang, J., Jiang, J., He, L., Wang, N., Yeung, P. C., & Pan, X. (2015). Rates and risk factors associated with the progression of HIV to AIDS among HIV patients from Zhejiang, China between 2008 and 2012. *AIDS Research and Therapy*, *12*(1), 32. https://doi.org/10.1186/s12981-015-0074-7

Chikata, T., Paes, W., Akahoshi, T., Partridge, T., Murakoshi, H., Gatanaga, H., Ternette, N., Oka, S., Borrow, P., & Takiguchi, M. (2019). Identification of Immunodominant HIV-1 Epitopes Presented by HLA-C*12:02, a Protective Allele, Using an Immunopeptidomics Approach. *Journal of Virology*, *93*(17). https://doi.org/10.1128/JVI.00634-19

Chikata, T., Paes, W., Kuse, N., Partridge, T., Gatanaga, H., Zhang, Y., Kuroki, K., Maenaka, K., Ternette, N., Oka, S., Borrow, P., & Takiguchi, M. (2022). Impact of Micropolymorphism Outside the Peptide Binding Groove in the Clinically Relevant Allele HLA-C*14 on T Cell Responses in HIV-1 Infection. *Journal of Virology*, *96*(10). https://doi.org/10.1128/jvi.00432-22

Chikata, T., Tran, G. Van, Murakoshi, H., Akahoshi, T., Qi, Y., Naranbhai, V., Kuse, N., Tamura, Y., Koyanagi, M., Sakai, S., Nguyen, D. H., Nguyen, D. T., Nguyen, H. T., Nguyen, T. V., Oka, S., Martin, M. P., Carrington, M., Sakai, K., Nguyen, K. Van, & Takiguchi, M. (2018). HLA Class I-Mediated HIV-1 Control in Vietnamese Infected with HIV-1 Subtype A/E. *Journal of Virology*, *92*(5). https://doi.org/10.1128/JVI.01749-17

Chopera, D. R., Mlotshwa, M., Woodman, Z., Mlisana, K., de Assis Rosa, D., Martin, D. P., Abdool Karim, S., Gray, C. M., & Williamson, C. (2011). Virological and immunological factors associated with HIV-1 differential disease progression in HLA-B 58:01-positive individuals. *Journal of Virology*, *85*(14), 7070–7080.

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y. J., Ghoorah, A. W., Dareng, E., Odia, T., Falola, O., Adebiyi, E., Hazelhurst, S., Mazandu, G., Nyangiri, O. A., Mbiyavanga, M., … Hanchard, N. A. (2020). High-depth African genomes inform human migration and health. *Nature*, *586*(7831), 741–748. https://doi.org/10.1038/s41586-020-2859-7

Chow, V. A., Shadman, M., & Gopal, A. K. (2018). Translating anti-CD19 CAR T-cell therapy into clinical practice for relapsed/refractory diffuse large B-cell lymphoma. *Blood*, *132*(8), 777–781. https://doi.org/10.1182/blood-2018-04-839217

Chun, T.-W., Justement, J. S., Moir, S., Hallahan, C. W., Ehler, L. A., Liu, S., McLaughlin, M., Dybul, M., Mican, J. M., & Fauci, A. S. (2001). Suppression of HIV replication in the resting CD4 + T cell reservoir by autologous CD8 + T cells: Implications for the development of therapeutic strategies. *Proceedings of the National Academy of Sciences*, *98*(1), 253–258. https://doi.org/10.1073/pnas.98.1.253

Cohen, G. B., Gandhi, R. T., Davis, D. M., Mandelboim, O., Chen, B. K., Strominger, J. L., & Baltimore, D. (1999). The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. *Immunity*, *10*(6), 661–671. https://doi.org/10.1016/S1074-7613(00)80065-5

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), 1511–1519.

https://doi.org/10.1002/pro.5560020916

Conroy, S. A., Laeyendecker, O., Redd, A. D., Collinson-Streng, A., Kong, X., Makumbi, F., Lutalo, T., Sewankambo, N., Kiwanuka, N., Gray, R. H., Wawer, M. J., Serwadda, D., & Quinn, T. C. (2010). Changes in the Distribution of HIV Type 1 Subtypes D and A in Rakai District, Uganda Between 1994 and 2002. *AIDS Research and Human Retroviruses*, *26*(10), 1087–1091. https://doi.org/10.1089/aid.2010.0054

Corrah, T. W., Goonetilleke, N., Kopycinski, J., Deeks, S. G., Cohen, M. S., Borrow, P., McMichael, A., & Brackenridge, S. (2011). Reappraisal of the Relationship between the HIV-1-Protective Single-Nucleotide Polymorphism 35 Kilobases Upstream of the HLA-C Gene and Surface HLA-C Expression. *Journal of Virology*, *85*(7), 3367–3374. https://doi.org/10.1128/JVI.02276-10

Correa, R., Harari, A., Vallelian, F., Resino, S., Munoz-Fernandez, M. A., & Pantaleo, G. (2007). Functional patterns of HIV-1-specific CD4 T-cell responses in children are influenced by the extent of virus suppression and exposure. *AIDS*, *21*(1), 23–30. https://doi.org/10.1097/QAD.0b013e32801120bc

Cristallo, A. F., Schroeder, J., Citterio, A., Santori, G., Ferrioli, G. M., Rossi, U., Bertani, G., Cassano, S., Gottardi, P., Ceschini, N., Barocci, F., Ribizzi, G., Cutrupi, V., Cairoli, R., Rapisarda, V., Pastorello, E. A., & Barocci, S. (2011). A study of HLA class I and class II 4-digit allele level in Stevens-Johnson syndrome and toxic epidermal necrolysis. *International Journal of Immunogenetics*, *38*(4), 303–309. https://doi.org/10.1111/j.1744-313X.2011.01011.x

Dangeti, S. R. (2014). Distinct advancements and challenges in HIV 1 vaccine development and cure - A review. *HIV and AIDS Review*, *13*(1), 1–5. https://doi.org/10.1016/j.hivar.2013.10.001

Day, C. L., Shea, A. K., Altfeld, M. A., Olson, D. P., Buchbinder, S. P., Hecht, F. M., Rosenberg, E. S., Walker, B. D., & Kalams, S. A. (2001). Relative Dominance of Epitope-Specific Cytotoxic T-Lymphocyte Responses in Human Immunodeficiency Virus Type 1-Infected Persons with Shared HLA Alleles. *Journal of Virology*, *75*(14), 6279–6291.

https://doi.org/10.1128/JVI.75.14.6279-6291.2001

De Cock, K. M., Fowler, M. G., Mercier, E., de Vincenzi, I., Saba, J., Hoff, E., Alnwick, D. J., Rogers, M., & Shaffer, N. (2000). Prevention of Mother-to-Child HIV Transmission in Resource-Poor Countries. *JAMA*, *283*(9), 1175. https://doi.org/10.1001/jama.283.9.1175

de Moraes-Pinto, M. I., Verhoeff, F., Chimsuku, L., Milligan, P. J. M., Wesumperuma, L., Broadhead, R. L., Brabin, B. J., Johnson, P. M., & Hart, C. A. (1998). Placental antibody transfer: influence of maternal HIV infection and placental malaria. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *79*(3), F202–F205. https://doi.org/10.1136/fn.79.3.F202

De Pee, S., & Semba, R. D. (2010). Role of nutrition in HIV infection: Review of evidence for more effective programming in resource-limited settings. *Food and Nutrition Bulletin*, *13*(4), S313-344. https://doi.org/10.1177/15648265100314s403

De Santis, G. C., Brunetta, D. M., Vilar, F. C., Brandão, R. A., de Albernaz Muniz, R. Z., de Lima, G. M. N., Amorelli-Chacel, M. E., Covas, D. T., & Machado, A. A. (2011). Hematological abnormalities in HIV-infected patients. *International Journal of Infectious Diseases*, *15*(12), e808–e811. https://doi.org/10.1016/j.ijid.2011.08.001

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., & O'Brien, S. J. (1996). Genetic Restriction of HIV-1 Infection and Progression to AIDS by a Deletion Allele of the CKR5 Structural Gene. *Science*, *273*(5283), 1856–1862. https://doi.org/10.1126/science.273.5283.1856

Del Amo, J., Petruckevitch, A., Phillips, A., Johnson, A. M., Stephenson, J., Desmond, N., Hanscheid, T., Low, N., Newell, A., Obasi, A., Paine, K., Pym, A., Theodore, C. M., & De Cock, K. M. (1998). Disease progression and survival in HIV-1-infected Africans in London. *AIDS*, *12*(10), 1203–1209. https://doi.org/10.1097/00002030-199810000-00013

Dendrou, C. A., Petersen, J., Rossjohn, J., & Fugger, L. (2018). HLA variation and disease. *Nature Reviews Immunology*, *18*(5), 325–339. https://doi.org/10.1038/nri.2017.143

Drotar, D., Olness, K., Wiznitzer, M., Schatschneider, C., Marum, L., Guay, L., Fagan, J., Hom, D., Svilar, G., Ndugwa, C., & Mayengo, R. K. (1999). Neurodevelopmental outcomes of Ugandan infants with HIV infection: An application of growth curve analysis. *Health Psychology*, *18*(2), 114–121. https://doi.org/10.1037/0278-6133.18.2.114

Duke, J. L., Lind, C., Mackiewicz, K., Ferriola, D., Papazoglou, A., Gasiewski, A., Heron, S., Huynh, A., McLaughlin, L., Rogers, M., Slavich, L., Walker, R., & Monos, D. S. (2016). Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA*, *87*(3), 141–152. https://doi.org/10.1111/tan.12736

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. In *Methods in Enzymology* (pp. 396–404). https://doi.org/10.1016/S0076-6879(97)77022-8

El-Hachem, N., Haibe-Kains, B., Khalil, A., Kobeissy, F. H., & Nemer, G. (2017). AutoDock and AutoDockTools for Protein-Ligand Docking: Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1(BACE1) as a Case Study. In *Methods in Molecular Biology* (pp. 391–403). https://doi.org/10.1007/978-1-4939-6952-4_20

El Bissati, K., Chentoufi, A. A., Krishack, P. A., Zhou, Y., Woods, S., Dubey, J. P., Vang, L., Lykins, J., Broderick, K. E., Mui, E., Suzuki, Y., Sa, Q., Bi, S., Cardona, N., Verma, S. K., Frazeck, L., Reardon, C. A., Sidney, J., Alexander, J., … McLeod, R. (2016). Adjuvanted multi-epitope vaccines protect HLA-A*11:01 transgenic mice against Toxoplasma gondii. *JCI Insight*, *1*(15). https://doi.org/10.1172/jci.insight.85955

Eller, M. A., Opollo, M. S., Liu, M., Redd, A. D., Eller, L. A., Kityo, C., Kayiwa, J., Laeyendecker, O., Wawer, M. J., Milazzo, M., Kiwanuka, N., Gray, R. H., Serwadda, D., Sewankambo, N. K., Quinn, T. C., Michael, N. L., Wabwire-Mangen, F., Sandberg, J. K., & Robb, M. L. (2015). HIV Type 1 Disease Progression to AIDS and Death in a Rural Ugandan Cohort Is Primarily Dependent on Viral Load Despite Variable Subtype and T-Cell Immune Activation Levels. *Journal of Infectious Diseases*, *211*(10), 1574–1584. https://doi.org/10.1093/infdis/jiu646

Embree, J. E., Njenga, S., Datta, P., Nagelkerke, N. J. D., Ndinya-Achola, J. O., Mohammed, Z., Ramdahin, S., Bwayo, J. J., & Plummer, F. A. (2000). Risk factors for postnatal mother–child transmission of HIV-1. *AIDS*, *14*(16), 2535–2541. https://doi.org/10.1097/00002030-200011100-00016

Etemad, B., Sun, X., Li, Y., Melberg, M., Moisi, D., Gottlieb, R., Ahmed, H., Aga, E., Bosch, R. J., Acosta, E. P., Yuki, Y., Martin, M. P., Carrington, M., Gandhi, R. T., Jacobson, J. M., Volberding, P., Connick, E., Mitsuyasu, R., Frank, I., … Li, J. Z. (2023). HIV post-treatment controllers have distinct immunological and virological features. *Proceedings of the National Academy of Sciences*, *120*(11). https://doi.org/10.1073/pnas.2218960120

Excler, J. L., Robb, M. L., & Kim, J. H. (2015). Prospects for a globally effective HIV-1 vaccine. *Vaccine*, *33*(6), D4–D12. https://doi.org/10.1016/j.vaccine.2015.03.059

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Ezeaka, V. C., Iroha, E. O., Akinsulie, A. O., Temiye, E. O., & Adetifa, I. M. O. (2009). Anthropometric indices of infants born to HIV-1-infected mothers: a prospective cohort study in Lagos, Nigeria. *International Journal of STD & AIDS*, *20*(8), 545–548. https://doi.org/10.1258/ijsa.2008.008446

Fabbiani, M., Di Giambenedetto, S., Bracciale, L., Bacarelli, A., Ragazzoni, E., Cauda, R., Navarra, P., & De Luca, A. (2009). Pharmacokinetic variability of antiretroviral drugs and correlation with virological outcome: 2 years of experience in routine clinical practice. *Journal of Antimicrobial Chemotherapy*, *64*(1), 109–117. https://doi.org/10.1093/jac/dkp132

Fan, Y., & Song, Y. Q. (2017). PyHLA: Tests for the association between HLA alleles and diseases. *BMC Bioinformatics*, *18*(1), 1–5. https://doi.org/10.1186/s12859-017-1496-0

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., & Pybus, O. G. (2014).

The early spread and epidemic ignition of HIV-1 in human populations. *Science*, *346*(6205), 51–61. https://doi.org/10.1126/science.1256739

Fellay, J., Shianna, K. V, Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J. M., Obel, N., Smith, J. P., … Goldstein, D. B. (2007). A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1. *Science*, *317*(5840), 944–947. https://doi.org/10.1126/science.1143767

Ferrando-Martínez, S., Casazza, J. P., Leal, M., Machmach, K., Muñoz-Fernández, M. Á., Viciana, P., Koup, R. A., & Ruiz-Mateos, E. (2012). Differential Gag-Specific Polyfunctional T Cell Maturation Patterns in HIV-1 Elite Controllers. *Journal of Virology*, *86*(7), 3667–3674. https://doi.org/10.1128/jvi.07034-11

Fettig, J., Swaminathan, M., Murrill, C. S., & Kaplan, J. E. (2014). Global Epidemiology of HIV. *Infect Dis Clin North Am*, *28*(3), 323–337. https://doi.org/10.1586/14737175.2015.1028369.Focused

Flores-Villanueva, P. O., Yunis, E. J., Delgado, J. C., Vittinghoff, E., Buchbinder, S., Leung, J. Y., Uglialoro, A. M., Clavijo, O. P., Rosenberg, E. S., Kalams, S. A., Braun, J. D., Boswell, S. L., Walker, B. D., & Goldfeld, A. E. (2001). Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proceedings of the National Academy of Sciences*, *98*(9), 5140–5145. https://doi.org/10.1073/pnas.071548198

Frahm, N., Nickle, D. C., Linde, C. H., Cohen, D. E., Zuñiga, R., Lucchetti, A., Roach, T., Walker, B. D., Allen, T. M., Korber, B. T., Mullins, J. I., & Brander, C. (2008). Increased detection of HIV-specific T cell responses by combination of central sequences with comparable immunogenicity. *AIDS*, *22*(4), 447–456. https://doi.org/10.1097/QAD.0b013e3282f42412

Frater, A. J., Brown, H., Oxenius, A., Gunthard, H. F., Hirschel, B., Robinson, N., Leslie, A. J., Payne, R., Crawford, H., Prendergast, A., Brander, C., Kiepiela, P., Walker, B. D., Goulder, P. J. R., McLean, A., & Phillips, R. E. (2007). Effective T-Cell Responses Select

Human Immunodeficiency Virus Mutants and Slow Disease Progression. *Journal of Virology*, *81*(12), 6742–6751. https://doi.org/10.1128/JVI.00022-07

Fusaro, V. A., Patil, P., Gafni, E., Wall, D. P., & Tonellato, P. J. (2011). Biomedical Cloud Computing With Amazon Web Services. *PLoS Computational Biology*, *7*(8), e1002147. https://doi.org/10.1371/journal.pcbi.1002147

Gartland, A. J., Li, S., McNevin, J., Tomaras, G. D., Gottardo, R., Janes, H., Fong, Y., Morris, D., Geraghty, D. E., Kijak, G. H., Edlefsen, P. T., Frahm, N., Larsen, B. B., Tovanabutra, S., Sanders-Buell, E., deCamp, A. C., Magaret, C. A., Ahmed, H., Goodridge, J. P., … Hertz, T. (2014). Analysis of HLA A*02 Association with Vaccine Efficacy in the RV144 HIV-1 Vaccine Trial. *Journal of Virology*, *88*(15), 8242–8255. https://doi.org/10.1128/JVI.01164-14

Gauderman, W. J. (2002). Sample size requirements for matched case-control studies of gene-environment interaction. *Statistics in Medicine*, *21*(1), 35–50. https://doi.org/10.1002/sim.973

Geng, J., Zaitouna, A. J., & Raghavan, M. (2018). Selected HLA-B allotypes are resistant to inhibition or deficiency of the transporter associated with antigen processing (TAP). *PLoS Pathogens*, *14*(7), e1007171. https://doi.org/10.1371/journal.ppat.1007171

Gentle, N. L., Paximadis, M., Puren, A., & Tiemessen, C. T. (2013). Genetic Variability in Markers of HLA-C Expression in Two Diverse South African Populations. *PLoS ONE*, *8*(7), e67780. https://doi.org/10.1371/journal.pone.0067780

Gijsbers, E. F., Anton Feenstra, K., Van Nuenen, A. C., Navis, M., Heringa, J., Schuitemaker, H., & Kootstra, N. A. (2013). HIV-1 replication fitness of HLA-B*57/58:01 CTL Escape variants is restored by the accumulation of compensatory mutations in Gag. *PLoS ONE*, *8*(12), e81235. https://doi.org/10.1371/journal.pone.0081235

GIRARD, M., OSMANOV, S., & KIENY, M. (2006). A review of vaccine research and development: The human immunodeficiency virus (HIV). *Vaccine*, *24*(19), 4062–4081. https://doi.org/10.1016/j.vaccine.2006.02.031

Gnjatic, S., Atanackovic, D., Matsuo, M., Jäger, E., Lee, S. Y., Valmori, D., Chen, Y.-T., Ritter, G., Knuth, A., & Old, L. J. (2003). Cross-Presentation of HLA Class I Epitopes from Exogenous NY-ESO-1 Polypeptides by Nonprofessional APCs. *The Journal of Immunology*, *170*(3), 1191–1196. https://doi.org/10.4049/jimmunol.170.3.1191

Gonzalez-Galarza, F. F., McCabe, A., Santos, E. J. M. Dos, Jones, J., Takeshita, L., Ortega-Rivera, N. D., Cid-Pavon, G. M. D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., Middleton, D., & Jones, A. R. (2020). Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Research*, *48*(D1), D783–D788. https://doi.org/10.1093/nar/gkz1029

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R. J., Freedman, B. I., Quinones, M. P., Bamshad, M. J., Murthy, K. K., Rovin, B. H., Bradley, W., Clark, R. A., Anderson, S. A., O'Connell, R. J., Agan, B. K., Ahuja, S. S., Bologna, R., … Ahuja, S. K. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. https://doi.org/10.1126/science.1101160

González, N., McKee, K., Lynch, R. M., Georgiev, I. S., Jimenez, L., Grau, E., Yuste, E., Kwong, P. D., Mascola, J. R., & Alcamí, J. (2018). Characterization of broadly neutralizing antibody responses to HIV-1 in a cohort of long term non-progressors. *PLoS ONE*, *13*(3), e0193773. https://doi.org/10.1371/journal.pone.0193773

Goulder, P. J., Lewin, S. R., & Leitman, E. M. (2016). Paediatric HIV infection: The potential for cure. *Nature Reviews Immunology*, *16*(4), 259–271. https://doi.org/10.1038/nri.2016.19

Goulder, P. J. R., & Walker, B. D. (2012). HIV and HLA Class I: an evolving relationship. *Immunity*, *37*(3), 426–440. https://doi.org/10.1016/j.immuni.2012.09.005.HIV

Goyette, P., Boucher, G., Mallon, D., Ellinghaus, E., Jostins, L., Huang, H., Ripke, S., Gusareva, E. S., Annese, V., Hauser, S. L., Oksenberg, J. R., Thomsen, I., Leslie, S., Daly, M. J., Van Steen, K., Duerr, R. H., Barrett, J. C., McGovern, D. P. B., Schumm, L. P., … Zhao, Z. Z. (2015). High-density mapping of the MHC identifies a shared role for HLA-

DRB1∗01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nature Genetics*, *47*(2), 172–179. https://doi.org/10.1038/ng.3176

Greub, G., Ledergerber, B., Battegay, M., Grob, P., Perrin, L., Furrer, H., Burgisser, P., Erb, P., Boggian, K., Piffaretti, J. C., Hirschel, B., Janin, P., Francioli, P., Flepp, M., & Telenti, A. (2000). Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus coinfection: The swiss HIV cohort study. *Lancet*, *356*(9244), 1800–1805. https://doi.org/10.1016/S0140-6736(00)03232-3

Groettrup, M., Kirk, C. J., & Basler, M. (2010). Proteasomes in immune cells: more than peptide producers? *Nature Reviews Immunology*, *10*(1), 73–78. https://doi.org/10.1038/nri2687

Gurdasani, D., Iles, L., Dillon, D. G., Young, E. H., Olson, A. D., Naranbhai, V., Fidler, S., Gkrania-Klotsas, E., Post, F. A., Kellam, P., Porter, K., & Sandhu, M. S. (2014). A systematic review of definitions of extreme phenotypes of HIV control and progression. *AIDS*, *28*(2), 149–162. https://doi.org/10.1097/QAD.0000000000000049

Gutierrez, R. G. (2010). Stata. *WIREs Computational Statistics*, *2*(6), 728–733. https://doi.org/10.1002/wics.116

Hahn, J. A., & Samet, J. H. (2010). Alcohol and HIV Disease Progression: Weighing the Evidence. *Current HIV/AIDS Reports*, *7*(4), 226–233. https://doi.org/10.1007/s11904-010-0060-6

Hansen, S. G., Marshall, E. E., Malouli, D., Ventura, A. B., Hughes, C. M., Ainslie, E., Ford, J. C., Morrow, D., Gilbride, R. M., Bae, J. Y., Legasse, A. W., Oswald, K., Shoemaker, R., Berkemeier, B., Bosche, W. J., Hull, M., Womack, J., Shao, J., Edlefsen, P. T., … Picker, L. J. (2019). A live-attenuated RhCMV/SIV vaccine shows long-term efficacy against heterologous SIV challenge. *Science Translational Medicine*, *11*(501). https://doi.org/10.1126/scitranslmed.aaw2607

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis

platform. *Journal of Cheminformatics*, *4*(1), 17. https://doi.org/10.1186/1758-2946-4-17

Haynes, B. F., & McElrath, M. J. (2013). Progress in HIV-1 vaccine development. *Current Opinion in HIV and AIDS*, *8*(4), 326–332. https://doi.org/10.1097/COH.0b013e328361d178

Helleberg, M., Kronborg, G., Larsen, C. S., Pedersen, G., Pedersen, C., Obel, N., & Gerstoft, J. (2013). CD4 Decline Is Associated With Increased Risk of Cardiovascular Disease, Cancer, and Death in Virally Suppressed Patients With HIV. *Clinical Infectious Diseases*, *57*(2), 314–321. https://doi.org/10.1093/cid/cit232

Helm, J. J. Van Der, Geskus, R., Lodi, S., Meyer, L., Schuitemaker, H., & Gunsenheimer-bartmeyer, B. (2014). Characterisation of long-term non-progression of HIV-1 infection after seroconversion : a cohort study. *Lancet HIV*, *1*, 41–48.

Heo, L., & Feig, M. (2018). PREFMD: A web server for protein structure refinement via molecular dynamics simulations. *Bioinformatics*, *34*(6), 1063–1065. https://doi.org/10.1093/bioinformatics/btx726

Hewitt, E. W. (2003). The MHC class I antigen presentation pathway: Strategies for viral immune evasion. *Immunology*, *110*(2), 163–169. https://doi.org/10.1046/j.1365-2567.2003.01738.x

Hill, A. V. (1996). Genetics of infectious disease resistance. *Current Opinion in Genetics & Development*, *6*(3), 348–353.

Honeyborne, I., Rathod, A., Buchli, R., Ramduth, D., Moodley, E., Rathnavalu, P., Chetty, S., Day, C., Brander, C., Hildebrand, W., Walker, B. D., Kiepiela, P., & Goulder, P. J. R. (2006). Motif Inference Reveals Optimal CTL Epitopes Presented by HLA Class I Alleles Highly Prevalent in Southern Africa 1. *Journal of Immunology*, *4201*(4), 4699–4705.

Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., Buus, S., & Nielsen, M. (2009). NetMHCpan, a method for MHC class i binding prediction beyond humans. *Immunogenetics*, *61*(1), 1–13. https://doi.org/10.1007/s00251-008-0341-z

Hraber, P., Korber, B. T., Lapedes, A. S., Bailer, R. T., Seaman, M. S., Gao, H., Greene, K. M., McCutchan, F., Williamson, C., Kim, J. H., Tovanabutra, S., Hahn, B. H., Swanstrom, R., Thomson, M. M., Gao, F., Harris, L., Giorgi, E., Hengartner, N., Bhattacharya, T., … Montefiori, D. C. (2014). Impact of Clade, Geography, and Age of the Epidemic on HIV-1 Neutralization by Antibodies. *Journal of Virology*, *88*(21), 12623–12643. https://doi.org/10.1128/jvi.01705-14

Hua, S., Lécuroux, C., Sáez-Cirión, A., Pancino, G., Girault, I., Versmisse, P., Boufassa, F., Taulera, O., Sinet, M., Lambotte, O., & Venet, A. (2014). Potential role for HIV-specific CD38-/HLA-DR+ CD8+ T cells in viral suppression and cytotoxicity in HIV controllers. *PLoS ONE*, *9*(7), e101920. https://doi.org/10.1371/journal.pone.0101920

Huang, M., Zhang, W., Guo, J., Wei, X., Phiwpan, K., Zhang, J., & Zhou, X. (2016). Improved Transgenic Mouse Model for Studying HLA Class I Antigen Presentation. *Scientific Reports*, *6*(1), 33612. https://doi.org/10.1038/srep33612

Huang, Y., Yang, J., Ying, D., Zhang, Y., Shotelersuk, V., Hirankarn, N., Sham, P. C., Lau, Y. L., & Yang, W. (2015). HLAreporter: a tool for HLA typing from next generation sequencing data. *Genome Medicine*, *7*(1), 25. https://doi.org/10.1186/s13073-015-0145-3

Ibrahim, M. K., Zambruni, M., Melby, C. L., & Melby, P. C. (2017). Impact of Childhood Malnutrition on Host Defense and Infection. *Clinical Microbiology Reviews*, *30*(4), 919–971. https://doi.org/10.1128/CMR.00119-16

Iliff, P. J., Piwoz, E. G., Tavengwa, N. V, Zunguza, C. D., Marinda, E. T., Nathoo, K. J., Moulton, L. H., Ward, B. J., & Humphrey, J. H. (2005). Early exclusive breastfeeding reduces the risk of postnatal HIV-1 transmission and increases HIV-free survival. *AIDS*, *19*(7), 699–708. https://doi.org/10.1097/01.aids.0000166093.16446.c9

Insaidoo, F. K., Borbulevych, O. Y., Hossain, M., Santhanagopolan, S. M., Baxter, T. K., & Baker, B. M. (2011). Loss of T Cell Antigen Recognition Arising from Changes in Peptide and Major Histocompatibility Complex Protein Flexibility. *Journal of Biological Chemistry*, *286*(46), 40163–40173. https://doi.org/10.1074/jbc.M111.283564

Iwelunmor, J., Ezeanolue, E. E., Airhihenbuwa, C. O., Obiefune, M. C., Ezeanolue, C. O., & Ogedegbe, G. G. (2014). Socio-cultural factors influencing the prevention of mother-to-child transmission of HIV in Nigeria: a synthesis of the literature. *BMC Public Health*, *14*(1), 771. https://doi.org/10.1186/1471-2458-14-771

Iyun, V., Technau, K. G., Eley, B., Rabie, H., Boulle, A., Fatti, G., Egger, M., Tanser, F., Wood, R., Fairlie, L., Cotton, M. F., & Davies, M. A. (2020). Earlier Antiretroviral Therapy Initiation and Decreasing Mortality among HIV-infected Infants Initiating Antiretroviral Therapy within 3 Months of Age in South Africa, 2006-2017. *Pediatric Infectious Disease Journal*, *39*(2), 127–133. https://doi.org/10.1097/INF.0000000000002516

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., Shales, M., Mercenne, G., Pache, L., Li, K., Hernandez, H., Jang, G. M., Roth, S. L., Akiva, E., Marlett, J., Stephens, M., D'Orso, I., Fernandes, J., Fahey, M., … Krogan, N. J. (2012). Global landscape of HIV–human protein complexes. *Nature*, *481*(7381), 365–370. https://doi.org/10.1038/nature10719

Janas, M. L., Groves, P., Kienzle, N., & Kelso, A. (2005). IL-2 Regulates Perforin and Granzyme Gene Expression in CD8+ T Cells Independently of Its Effects on Survival and Proliferation. *The Journal of Immunology*, *175*(12), 8003–8010. https://doi.org/10.4049/jimmunol.175.12.8003

Jean, S. S., PAPE, J. W., VERDIER, R.-I., REED, G. W., HUTTO, C., JOHNSON, W. D., & WRIGHT, P. F. (1999). The natural history of human immunodeficiency virus 1 infection in Haitian infants. *The Pediatric Infectious Disease Journal*, *18*(1), 58–63. https://doi.org/10.1097/00006454-199901000-00014

Jenkins, M., Mills, J., & Kohl, S. (1993). Natural Killer Cytotoxicity and Antibody-Dependent Cellular Cytotoxicity of Human Immunodeficiency Virus–Infected Cells by Leukocytes from Human Neonates and Adults. *Pediatric Research*, *33*(5), 469–474. https://doi.org/10.1203/00006450-199305000-00010

Jesson, J., Masson, D., Adonon, A., Tran, C., Habarugira, C., Zio, R., Nicimpaye, L.,

Desmonde, S., Serurakuba, G., Kwayep, R., Sare, E., Konate, T., Nimaga, A., Saina, P., Kpade, A., Bassuka, A., Gougouyor, G., & Leroy, V. (2015). Prevalence of malnutrition among HIV-infected children in Central and West-African HIV-care programmes supported by the Growing Up Programme in 2011: a cross-sectional study. *BMC Infectious Diseases*, *15*(1), 216. https://doi.org/10.1186/s12879-015-0952-6

Jiang, H., Xie, N., Cao, B., Tan, L., Fan, Y., Zhang, F., Yao, Z., Liu, L., & Nie, S. (2013). Determinants of progression to AIDS and death following HIV diagnosis: A retrospective cohort study in Wuhan, China. *PLoS ONE*, *8*(12), 1–11. https://doi.org/10.1371/journal.pone.0083078

Jin, X., Gao, X., Murugappan Ramanathan, J., Deschenes, G. R., Nelson, G. W., O'Brien, S. J., Goedert, J. J., Ho, D. D., O'Brien, T. R., & Carrington, M. (2002). Human immunodeficiency virus type 1 (HIV-1)-specific CD8+-T-cell responses for groups of HIV-1-infected individuals with different HLA-B*35 genotypes. *Journal of Virology*, *76*(24), 12603–12610. https://doi.org/10.1128/JVI.76.24.12603

Jjingo, D., Mboowa, G., Sserwadda, I., Kakaire, R., Kiberu, D., Amujal, M., Galiwango, R., Kateete, D., Joloba, M., & Whalen, C. C. (2022). Bioinformatics mentorship in a resource limited setting. *Briefings in Bioinformatics*, *23*(1), 1–8. https://doi.org/10.1093/bib/bbab399

Joglekar, A. V., Liu, Z., Weber, J. K., Ouyang, Y., Jeppson, J. D., Noh, W. J., Lamothe-Molina, P. A., Chen, H., Kang, S. gu, Bethune, M. T., Zhou, R., Walker, B. D., & Baltimore, D. (2018). T cell receptors for the HIV KK10 epitope from patients with differential immunologic control are functionally indistinguishable. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), 1877–1882. https://doi.org/10.1073/pnas.1718659115

Johnson, E. L., & Chakraborty, R. (2012). Placental Hofbauer cells limit HIV-1 replication and potentially offset mother to child transmission (MTCT) by induction of immunoregulatory cytokines. *Retrovirology*, *9*(1), 101. https://doi.org/10.1186/1742-4690-9-101

Johnston, M. I., & Fauci, A. S. (2007). An HIV Vaccine — Evolving Concepts. *New England Journal of Medicine*, *356*(20), 2073–2081. https://doi.org/10.1056/NEJMra066267

Johnston, M. I., & Fauci, A. S. (2008). An HIV Vaccine — Challenges and Prospects. *New England Journal of Medicine*, *359*(9), 888–890. https://doi.org/10.1056/NEJMp0806162

Junqueira, D. M., & Almeida, S. E. de M. (2016). HIV-1 subtype B: Traces of a pandemic. *Virology*, *495*, 173–184. https://doi.org/10.1016/j.virol.2016.05.003

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., … Baker, N. A. (2018). Improvements to the <scp>APBS</scp> biomolecular solvation software suite. *Protein Science*, *27*(1), 112–128. https://doi.org/10.1002/pro.3280

Kadukova, M., & Grudinin, S. (2017). Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization. *Journal of Computer-Aided Molecular Design*, *31*(10), 943–958. https://doi.org/10.1007/s10822-017-0068-8

Kahan, S. M., Bakshi, R. K., Ingram, J. T., Hendrickson, R. C., Lefkowitz, E. J., Crossman, D. K., Harrington, L. E., Weaver, C. T., & Zajac, A. J. (2022). Intrinsic IL-2 production by effector CD8 T cells affects IL-2 signaling and promotes fate decisions, stemness, and protection. *Science Immunology*, *7*(68). https://doi.org/10.1126/sciimmunol.abl6322

Karnes, J. H., Shaffer, C. M., Bastarache, L., Gaudieri, S., Glazer, A. M., Steiner, H. E., Mosley, J. D., Mallal, S., Denny, J. C., Phillips, E. J., & Roden, D. M. (2017). Comparison of HLA allelic imputation programs. *PLoS ONE*, *12*(2), 1–12. https://doi.org/10.1371/journal.pone.0172444

Kaseke, C., Tano-Menka, R., Senjobe, F., & Gaiha, G. D. (2021). The Emerging Role for CTL Epitope Specificity in HIV Cure Efforts. *The Journal of Infectious Diseases*, *223*(Supplement_1), S32–S37. https://doi.org/10.1093/infdis/jiaa333

Kassa, G. M. (2018). Mother-to-child transmission of HIV infection and its associated factors

in Ethiopia: a systematic review and meta-analysis. *BMC Infectious Diseases*, *18*(1), 216. https://doi.org/10.1186/s12879-018-3126-5

Kathryn, P., Mcclelland, R. S., Overbaugh, J., Richardson, B. A., Bosire, R., Kiarie, J. N., Farquhar, C., & Guthrie, B. L. (2018). Higher prevalence of viral control in HIV-1-infected women in serodiscordant relationships. *PLoS ONE*, *13*(12), 1–11.

Kawashima, Y., Kuse, N., Gatanaga, H., Naruto, T., Fujiwara, M., Dohki, S., Akahoshi, T., Maenaka, K., Goulder, P., Oka, S., & Takiguchi, M. (2010). Long-Term Control of HIV-1 in Hemophiliacs Carrying Slow-Progressing Allele HLA-B*5101. *Journal of Virology*, *84*(14), 7151–7160. https://doi.org/10.1128/JVI.00171-10

Keşmir, C., Nussbaum, A. K., Schild, H., Detours, V., & Brunak, S. (2002). Prediction of proteasome cleavage motifs by neural networks. *Protein Engineering, Design and Selection*, *15*(4), 287–296. https://doi.org/10.1093/protein/15.4.287

Kijak, G. H., Walsh, A. M., Koehler, R. N., Moqueet, N., Eller, L. A., Eller, M., Currier, J. R., Wang, Z., Wabwire-Mangen, F., Kibuuka, H. N., Michael, N. L., Robb, M. L., & McCutchan, F. E. (2009). HLA class i allele and haplotype diversity in Ugandans supports the presence of a major east African genetic cluster. *Tissue Antigens*, *73*(3), 262–269. https://doi.org/10.1111/j.1399-0039.2008.01192.x

Kijewski, S. D., & Gummuluru, S. (2015). A mechanistic overview of dendritic cell-mediated HIV-1 trans infection: the story so far. *Future Virol*, *10*(3), 257–269. https://doi.org/0.2217/fvl.15.2.

Kim, G. G., Donnenberg, V. S., Donnenberg, A. D., Gooding, W., & Whiteside, T. L. (2007). A novel multiparametric flow cytometry-based cytotoxicity assay simultaneously immunophenotypes effector cells: Comparisons to a 4 h51Cr-release assay. *Journal of Immunological Methods*, *325*(1–2), 51–66. https://doi.org/10.1016/j.jim.2007.05.013

Kimani-Murage, E. W., Norris, S. A., Pettifor, J. M., Tollman, S. M., Klipstein-Grobusch, K., Gómez-Olivé, X. F., Dunger, D. B., & Kahn, K. (2011). Nutritional status and HIV in rural South African children. *BMC Pediatrics*, *11*(23), 1–13.

https://doi.org/10.1186/1471-2431-11-23

Klasberg, S., Surendranath, V., Lange, V., & Schöfl, G. (2019). Bioinformatics Strategies, Challenges, and Opportunities for Next Generation Sequencing-Based HLA Genotyping. *Transfusion Medicine and Hemotherapy*, *46*(5), 312–325. https://doi.org/10.1159/000502487

Klepiela, P., Leslie, A. J., Honeyborne, I., Ramduth, D., Thobakgale, C., Chetty, S., Rathnavalu, P., Moore, C., Pfafferott, K. J., Hilton, L., Zimbwa, P., Moore, S., Allen, T., Brander, C., Addo, M. M., Altfeld, M., James, I., Mallal, S., Bunce, M., … Goulder, P. J. R. (2004). Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature*, *432*(7018), 769–775. https://doi.org/10.1038/nature03113

Kloverpris, H. N., Harndahl, M., Leslie, A. J., Carlson, J. M., Ismail, N., van der Stok, M., Huang, K.-H. G., Chen, F., Riddell, L., Steyn, D., Goedhals, D., van Vuuren, C., Frater, J., Walker, B. D., Carrington, M., Ndung'u, T., Buus, S., & Goulder, P. (2012). HIV Control through a Single Nucleotide on the HLA-B Locus. *Journal of Virology*, *86*(21), 11493–11500. https://doi.org/10.1128/jvi.01020-12

Koehler, R. N., Walsh, A. M., Saathoff, E., Tovanabutra, S., Arroyo, M. A., Currier, J. R., Maboko, L., Hoelsher, M., Robb, M. L., Michael, N. L., McCutchan, F. E., Kim, J. H., & Kijak, G. H. (2010). Class I HLA-A*7401 Is Associated with Protection from HIV-1 Acquisition and Disease Progression in Mbeya, Tanzania. *The Journal of Infectious Diseases*, *202*(10), 1562–1566. https://doi.org/10.1086/656913

Koethe, J. R., & Heimburger, D. C. (2010). Nutritional aspects of HIV-associated wasting in sub-Saharan Africa. *American Journal of Clinical Nutrition*, *91*(4), 1138S–1142S. https://doi.org/10.3945/ajcn.2010.28608D

Kongkaew, S., Yotmanee, P., Rungrotmongkol, T., Kaiyawet, N., Meeprasert, A., Kaburaki, T., Noguchi, H., Takeuchi, F., Kungwan, N., & Hannongbua, S. (2015). Molecular Dynamics Simulation Reveals the Selective Binding of Human Leukocyte Antigen Alleles Associated with Behçet's Disease. *PLOS ONE*, *10*(9), e0135575. https://doi.org/10.1371/journal.pone.0135575

Korber, B., & Fischer, W. (2020). T cell-based strategies for HIV-1 vaccines. *Human Vaccines & Immunotherapeutics*, *16*(3), 713–722. https://doi.org/10.1080/21645515.2019.1666957

Körner, C., Simoneau, C. R., Schommers, P., Granoff, M., Ziegler, M., Hölzemer, A., Lunemann, S., Chukwukelu, J., Corleis, B., Naranbhai, V., Kwon, D. S., Scully, E. P., Jost, S., Kirchhoff, F., Carrington, M., & Altfeld, M. (2017). HIV-1-Mediated Downmodulation of HLA-C Impacts Target Cell Recognition and Antiviral Activity of NK Cells. *Cell Host & Microbe*, *22*(1), 111-119.e4. https://doi.org/10.1016/j.chom.2017.06.008

Kourtis, A. P., Nesheim, S. R., Thea, D., Ibegbu, C., Nahmias, A. J., & Lee, F. K. (2000). Correlation of virus load and soluble L-selectin, a marker of immune activation, in pediatric HIV-1 infection. *AIDS*, *14*(16), 2429–2436. https://doi.org/10.1097/00002030-200011100-00003

Kovacs, A. A. Z., Kono, N., Wang, C.-H., Wang, D., Frederick, T., Operskalski, E., Tien, P. C., French, A. L., Minkoff, H., Kassaye, S., T. Golub, E., Aouizerat, B. E., Kuniholm, M. H., & Millstein, J. (2020). Association of HLA Genotype With T-Cell Activation in Human Immunodeficiency Virus (HIV) and HIV/Hepatitis C Virus–Coinfected Women. *The Journal of Infectious Diseases*, *221*(7), 1156–1166. https://doi.org/10.1093/infdis/jiz589

Kovjazin, R., Shitrit, D., Preiss, R., Haim, I., Triezer, L., Fuks, L., Nader, R., Raz, M., Bardenstein, R., Horn, G., Smorodinsky, N. I., & Carmon, L. (2013). Characterization of Novel Multiantigenic Vaccine Candidates with Pan-HLA Coverage against Mycobacterium tuberculosis. *Clinical and Vaccine Immunology*, *20*(3), 328–340. https://doi.org/10.1128/CVI.00586-12

Kretzschmar, M. E., Schim Van Der Loeff, M. F., Birrell, P. J., De Angelis, D., & Coutinho, R. A. (2013). Prospects of elimination of HIV with test-and-treat strategy. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15538–15543. https://doi.org/10.1073/pnas.1301801110

Krishnan, S. G., Joshi, A., & Kaushik, V. (2020). T cell epitope designing for dengue peptide vaccine using docking and molecular simulation studies. *Molecular Simulation*, *46*(10), 787–795. https://doi.org/10.1080/08927022.2020.1772970

Kulkarni, S., Qi, Y., O'hUigin, C., Pereyra, F., Ramsuran, V., McLaren, P., Fellay, J., Nelson, G., Chen, H., Liao, W., Bass, S., Apps, R., Gao, X., Yuki, Y., Lied, A., Ganesan, A., Hunt, P. W., Deeks, S. G., Wolinsky, S., … Carrington, M. (2013). Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. *Proceedings of the National Academy of Sciences*, *110*(51), 20705–20710. https://doi.org/10.1073/pnas.1312237110

Kulkarni, S., Savan, R., Qi, Y., Gao, X., Yuki, Y., Bass, S. E., Martin, M. P., Hunt, P., Deeks, S. G., Telenti, A., Pereyra, F., Goldstein, D., Wolinsky, S., Walker, B., Young, H. A., & Carrington, M. (2011). Differential microRNA regulation of HLA-C expression and its association with HIV control. *Nature*, *472*(7344), 495–498. https://doi.org/10.1038/nature09914

Kulpa, D. A., & Collins, K. L. (2011). The emerging role of HLA-C in HIV-1 infection. *Immunology*, *134*(2), 116–122. https://doi.org/10.1111/j.1365-2567.2011.03474.x

Kumar, S. B., Rice, C. E., Milner, D. A., Ramirez, N. C., Ackerman, W. E., Mwapasa, V., Turner, A. N., & Kwiek, J. J. (2012). Elevated cytokine and chemokine levels in the placenta are associated with in-utero HIV-1 mother-to-child transmission. *AIDS*, *26*(6), 685–694. https://doi.org/10.1097/QAD.0b013e3283519b00

Kuyl, A. C. Van Der, & Berkhout, B. (2012). The biased nucleotide composition of the HIV genome : a constant factor in a highly variable virus. *Retrovirology*, *9*(92), 1–14.

Kyobe, S., Kisitu, G. P., Farirai, J., Mwesigwa, S., Mirembe, A. N., Katagirya, E., Ketumile, L., Wayengera, M., Retshabile, G., Mlotshwa, B. C., Williams, L., Morapedi, K., Kyosiimire-Lugemwa, J., Nsangi, B., Wampande, E. M., Masego, T.-J., Brown, C. W., Joloba, M., Anabwani, G., … Kekitiinwa, A. R. (n.d.). *Characterization of HIV progression among children in Uganda and Botswana: A retrospective cohort study. Unpublished manuscript*.

Kyobe, S., Mwesigwa, S., Kisitu, G. P., Farirai, J., Katagirya, E., Mirembe, A. N., Ketumile, L., Wayengera, M., Katabazi, F. A., Kigozi, E., Wampande, E. M., Retshabile, G., Mlotshwa, B. C., Williams, L., Morapedi, K., Kasvosve, I., Kyosiimire-Lugemwa, J., Nsangi, B., Tsimako-Johnstone, M., … Hanchard, N. A. (2021). Exome Sequencing Reveals a Putative Role for HLA-C*03:02 in Control of HIV-1 in African Pediatric Populations. *Frontiers in Genetics*, *12*(August), 1–14. https://doi.org/10.3389/fgene.2021.720213

Kyosiimire-lugemwa, J., Ssemaganda, A., Musinguzi, K., Nsubuga, R., Imami, N., Gotch, F., & Pala, P. (2012). *Programmed Death-1 (PD-1), a Correlate of Protection Against Disease Progression in HIV-1 Infected Long-Term Non-Progressors*. 2012.

Laeyendecker, O., Redd, A. D., Lutalo, T., Gray, R. H., Wawer, M., Ssempijja, V., Gamiel, J., Bwanika, J. B., Makumbi, F., Nalugoda, F., Opendi, P., Kigozi, G., Ndyanabo, A., Iga, B., Kiwanuka, N., Sewankambo, N., Reynolds, S. J., Serwadda, D., & Quinn, T. C. (2009). Frequency of long-term nonprogressors in HIV-1 seroconverters from rakai uganda. *Journal of Acquired Immune Deficiency Syndromes*, *52*(3), 316–319. https://doi.org/10.1097/QAI.0b013e3181bc08f5

Lamiable, A., Thevenet, P., Rey, J., Vavrusa, M., Derreumaux, P., & Tuffery, P. (2016). PEP-FOLD3: faster denovo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*, *44*(1), W449–W454. https://doi.org/10.1093/nar/gkw329

Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P., & Thomson, G. (2007). PyPop update - A software pipeline for large-scale multilocus population genomics. *Tissue Antigens*, *69*(SUPPL. 1), 192–197. https://doi.org/10.1111/j.1399-0039.2006.00769.x

Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P., & Thomson, G. (2010). Lightning talk: PyPop – a software pipeline for large-scale multilocus population genomics. *Nature Precedings*, *713*, 1–6. https://doi.org/10.1038/npre.2010.4700.1

Lancaster, A., Nelson, M. P., Meyer, D., Single, R. M., & Thomson, G. (2003). PyPop: A Software Framework for Population Genomics: Analyzing Large-Scale Multi-Locus Genotype Data. *Pac Symp Biocomput*, *6*(9), 790–795.

https://doi.org/10.1016/j.pmrj.2014.02.014.Lumbar

Larjo, A., Eveleigh, R., Kilpeläinen, E., Kwan, T., Pastinen, T., Koskela, S., & Partanen, J. (2017). Accuracy of programs for the determination of human leukocyte antigen alleles from next-generation sequencing data. *Frontiers in Immunology*, *8*(DEC), 1–9. https://doi.org/10.3389/fimmu.2017.01815

Lazaryan, A., Lobashevsky, E., Mulenga, J., Karita, E., Allen, S., Tang, J., & Kaslow, R. A. (2006). Human Leukocyte Antigen B58 Supertype and Human Immunodeficiency Virus Type 1 Infection in Native Africans. *Journal of Virology*, *80*(12), 6056–6060. https://doi.org/10.1128/jvi.02119-05

Leitman, E. M., Thobakgale, C. F., Adland, E., Ansari, M. A., Raghwani, J., Prendergast, A. J., Tudor-Williams, G., Kiepiela, P., Hemelaar, J., Brener, J., Tsai, M. H., Mori, M., Riddell, L., Luzzi, G., Jooste, P., Ndung'u, T., Walker, B. D., Pybus, O. G., Kellam, P., … Goulder, P. J. R. (2017). Role of HIV-specific CD8+ T cells in pediatric HIV cure strategies after widespread early viral escape. *Journal of Experimental Medicine*, *214*(11), 3239–3261. https://doi.org/10.1084/jem.20162123

Leslie, A., Matthews, P. C., Listgarten, J., Carlson, J. M., Kadie, C., Ndung'u, T., Brander, C., Coovadia, H., Walker, B. D., Heckerman, D., & Goulder, P. J. R. (2010). Additive Contribution of HLA Class I Alleles in the Immune Control of HIV-1 Infection. *Journal of Virology*, *84*(19), 9879–9888. https://doi.org/10.1128/jvi.00320-10

Levin, M. L. (1953). The occurrence of lung cancer in man. *Acta - Unio Internationalis Contra Cancrum*, *9*(3), 531–541.

Levy, J. A. (2007). *HIV and the Pathogenesis of AIDS* (Second Edi). ASM Press.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Ivarsson, M. A., Walker-Sperling, V. E., Subleski, J., Johnson, J. K., Wright, P. W., Carrington, M., Björkström, N. K., McVicar, D. W., & Anderson, S. K. (2018). Identification of an elaborate NK-specific system regulating HLA-C expression. *PLoS*

Genetics, *14*(1), e1007163. https://doi.org/10.1371/journal.pgen.1007163

Li, L., & Bouvier, M. (2004). Structures of HLA-A*1101 Complexed with Immunodominant Nonamer and Decamer HIV-1 Epitopes Clearly Reveal the Presence of a Middle, Secondary Anchor Residue. *The Journal of Immunology*, *172*(10), 6175–6184. https://doi.org/10.4049/jimmunol.172.10.6175

Li, W., Joshi, M., Singhania, S., Ramsey, K., & Murthy, A. (2014). Peptide Vaccine: Progress and Challenges. *Vaccines*, *2*(3), 515–536. https://doi.org/10.3390/vaccines2030515

Li, X., Jin, S., Fan, Y., Fan, X., Tang, Z., Cai, W., Yang, J., & Xiang, X. (2019). Association of HLA-C*03:02 with methimazole-induced liver injury in Graves' disease patients. *Biomedicine & Pharmacotherapy*, *117*, 109095. https://doi.org/10.1016/j.biopha.2019.109095

Li, X., Singh, N. K., Collins, D. R., Ng, R., Zhang, A., Lamothe-Molina, P. A., Shahinian, P., Xu, S., Tan, K., Piechocka-Trocha, A., Urbach, J. M., Weber, J. K., Gaiha, G. D., Takou Mbah, O. C., Huynh, T., Cheever, S., Chen, J., Birnbaum, M., Zhou, R., … Wang, J. (2023). Molecular basis of differential HLA class I-restricted T cell recognition of a highly networked HIV peptide. *Nature Communications*, *14*(1), 2929. https://doi.org/10.1038/s41467-023-38573-8

Lin, Z., Kuroki, K., Kuse, N., Sun, X., Akahoshi, T., Qi, Y., Chikata, T., Naruto, T., Koyanagi, M., Murakoshi, H., Gatanaga, H., Oka, S., Carrington, M., Maenaka, K., & Takiguchi, M. (2016). HIV-1 Control by NK Cells via Reduced Interaction between KIR2DL2 and HLA-C∗12:02/C∗14:03. *Cell Reports*, *17*(9), 2210–2220. https://doi.org/10.1016/j.celrep.2016.10.075

Little, K., Thorne, C., Luo, C., Bunders, M., Ngongo, N., McDermott, P., & Newell, M.-L. (2007). Disease Progression in Children with Vertically-Acquired HIV Infection in Sub-Saharan Africa: Reviewing the Need for HIV Treatment. *Current HIV Research*, *5*(2), 139–153. https://doi.org/10.2174/157016207780077002

Liu, A., Gong, B., Liu, X., Shen, Y., Wu, Y., Zhang, W., & Cao, J. (2020). A retrospective

epidemiological analysis of human Cryptosporidium infection in China during the past three decades (1987-2018). *PLOS Neglected Tropical Diseases*, *14*(3), e0008146. https://doi.org/10.1371/journal.pntd.0008146

Liu, C., Yang, X., Duffy, B., Mohanakumar, T., Mitra, R. D., Zody, M. C., & Pfeifer, J. D. (2013). ATHLATES: Accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Research*, *41*(14), 1–8. https://doi.org/10.1093/nar/gkt481

Liu, H., Chao, D., Nakayama, E. E., Taguchi, H., Goto, M., Xin, X., Takamatsu, J., Saito, H., Ishikawa, Y., Akaza, T., Juji, T., Takebe, Y., Ohishi, T., Fukutake, K., Maruyama, Y., Yashiki, S., Sonoda, S., Nakamura, T., Nagai, Y., … Shioda, T. (1999). Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. *Proceedings of the National Academy of Sciences*, *96*(8), 4581–4585. https://doi.org/10.1073/pnas.96.8.4581

Liu, H., Hwangbo, Y., Holte, S., Lee, J., Wang, C., Kaupp, N., Zhu, H., Celum, C., Corey, L., McElrath, M. J., & Zhu, T. (2004). Analysis of Genetic Polymorphisms in CCR5, CCR2, Stromal Cell–Derived Factor–1, RANTES, and Dendritic Cell–Specific Intercellular Adhesion Molecule–3–Grabbing Nonintegrin in Seronegative Individuals Repeatedly Exposed to HIV-1. *The Journal of Infectious Diseases*. https://doi.org/10.1086/423209

Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A., & Landau, N. R. (1996). Homozygous Defect in HIV-1 Coreceptor Accounts for Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection. *Cell*, *86*(3), 367–377. https://doi.org/10.1016/S0092-8674(00)80110-5

Liu, Y., Keib, A., Neuber, B., Wang, L., Riemer, A. B., Bonsack, M., Hückelhoven-Krauss, A., Schmitt, A., Müller-Tidow, C., & Schmitt, M. (2023). Definition and Characterization of SOX11-Derived T Cell Epitopes towards Immunotherapy of Glioma. *International Journal of Molecular Sciences*, *24*(3), 1943. https://doi.org/10.3390/ijms24031943

Liu, Z., Chen, S., Jin, X., Wang, Q., Yang, K., Li, C., Xiao, Q., Hou, P., Liu, S., Wu, S., Hou, W., Xiong, Y., Kong, C., Zhao, X., Wu, L., Li, C., Sun, G., & Guo, D. (2017). Genome editing of the HIV co-receptors CCR5 and CXCR4 by CRISPR-Cas9 protects CD4+T cells from HIV-1 infection. *Cell and Bioscience*, *7*(1), 1–15.

https://doi.org/10.1186/s13578-017-0174-2

Lo Nigro, C., Macagno, M., Sangiolo, D., Bertolaccini, L., Aglietta, M., & Merlano, M. C. (2019). NK-mediated antibody-dependent cell-mediated cytotoxicity in solid tumors: biological evidence and clinical perspectives. *Annals of Translational Medicine*, *7*(5), 105–105. https://doi.org/10.21037/atm.2019.01.42

Lockett, S. F., Robertson, J. R., Brettle, R. P., Yap, P. L., Middleton, D., & Brown, A. J. L. (2001). Mismatched human leukocyte antigen alleles protect against heterosexual HIV transmission. *J Acquir Immune Defic Syndr*, *27*, 277–280. https://doi.org/10.1021/acs.chemmater.5b02331

López, M., Palacio, M., Goncé, A., Hernàndez, S., Barranco, F. J., García, L., Loncà, M., Coll, J. O., Gratacós, E., & Figueras, F. (2015). Risk of intrauterine growth restriction among HIV-infected pregnant women: a cohort study. *European Journal of Clinical Microbiology & Infectious Diseases*, *34*(2), 223–230. https://doi.org/10.1007/s10096-014-2224-6

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, *50*(3), 437–450. https://doi.org/10.1002/prot.10286

Lu, J., Hou, X., Wang, C., & Zhang, Y. (2019). Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *Journal of Chemical Information and Modeling*, *59*(11), 4540–4549. https://doi.org/10.1021/acs.jcim.9b00645

Lucas, C. G. D. O., Matassoli, F. L., Peçanha, L. M. T., Santillo, B. T., Oliveira, L. M. da S., Oshiro, T. M., Marques, E. T. D. A., Oxenius, A., & Arruda, L. B. (2016). Dendritic cells primed with a chimeric plasmid containing HIV-1- gag associated with lysosomal-associated protein-1 (LAMP/ gag) is a potential therapeutic vaccine against HIV. *The FASEB Journal*, *30*(8), 2970–2984. https://doi.org/10.1096/fj.201500059

Luo, H., Ye, H., Ng, H. W., Shi, L., Tong, W., Mendrick, D. L., & Hong, H. (2015). Machine Learning Methods for Predicting HLA-Peptide Binding Activity. *Bioinformatics and Biology Insights*, *9s3*, BBI.S29466. https://doi.org/10.4137/BBI.S29466

Magierowska, M., Theodorou, I., Debré, P., Sanson, F., Autran, B., Rivière, Y., Charron, D., & Costagliola, D. (1999). Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. *Blood*, *93*(3), 936–941. http://www.ncbi.nlm.nih.gov/pubmed/9920843

Mahiti, M., Toyoda, M., Jia, X., Kuang, X. T., Mwimanzi, F., Mwimanzi, P., Walker, B. D., Xiong, Y., Brumme, Z. L., Brockman, M. A., & Ueno, T. (2016). Relative Resistance of HLA-B to Downregulation by Naturally Occurring HIV-1 Nef Sequences. *MBio*, *7*(1). https://doi.org/10.1128/mBio.01516-15

Major, E., Rigó, K., Hague, T., Bérces, A., & Juhos, S. (2013). HLA Typing from 1000 Genomes Whole Genome and Whole Exome Illumina Data. *PLoS ONE*, *8*(11), e78410. https://doi.org/10.1371/journal.pone.0078410

Makadzange, A. T., Gillespie, G., Dong, T., Kiama, P., Bwayo, J., Kimani, J., Plummer, F., Easterbrook, P., & Rowland-Jones, S. L. (2010). Characterization of an HLA-C-restricted CTL response in chronic HIV infection. *European Journal of Immunology*, *40*(4), 1036–1041. https://doi.org/10.1002/eji.200939634

Mann, J. K., Byakwaga, H., Kuang, X. T., Le, A. Q., Brumme, C. J., Mwimanzi, P., Omarjee, S., Martin, E., Lee, G. Q., Baraki, B., Danroth, R., McCloskey, R., Muzoora, C., Bangsberg, D. R., Hunt, P. W., Goulder, P. J., Walker, B. D., Harrigan, P. R., Martin, J. N., … Brumme, Z. L. (2013). Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. *Retrovirology*, *10*(1), 100. https://doi.org/10.1186/1742-4690-10-100

Marsh, S. G. E., Albert, E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., Fernández-Viña, M., Geraghty, D. E., Holdsworth, R., Hurley, C. K., Lau, M., Lee, K. W., Mach, B., Maiers, M., Mayr, W. R., Müller, C. R., Parham, P., Petersdorf, E. W.,

Sasazuki, T., … Trowsdale, J. (2010). Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, *75*(4), 291–455. https://doi.org/10.1111/j.1399-0039.2010.01466.x

Martin, G., & Martin, G. (2021). R Studio. In *An Introduction to Programming with R*. https://doi.org/10.1007/978-3-030-69664-1_1

Martinez, A. N., Bluthenthal, R. N., Flynn, N. M., Anderson, R. L., & Kral, A. H. (2011). HIV Risks and Seroprevalence Among Mexican American Injection Drug Users in California. *AIDS and Behavior*, *15*(1), 95–102. https://doi.org/10.1007/s10461-009-9614-2

Matheron, S., Pueyo, S., Damond, F., Simon, F., Leprêtre, A., Campa, P., Salamon, R., Chêne, G., & Brun-Vezinet, F. (2003). Factors associated with clinical progression in HIV-2 infected-patients. *AIDS*, *17*(18), 2593–2601. https://doi.org/10.1097/00002030-200312050-00006

Matthews, P. C., Adland, E., Listgarten, J., Leslie, A., Mkhwanazi, N., Carlson, J. M., Harndahl, M., Stryhn, A., Payne, R. P., Ogwu, A., Huang, K.-H. G., Frater, J., Paioni, P., Kloverpris, H., Jooste, P., Goedhals, D., van Vuuren, C., Steyn, D., Riddell, L., … Goulder, P. J. R. (2011). HLA-A*7401–Mediated Control of HIV Viremia Is Independent of Its Linkage Disequilibrium with HLA-B*5703. *The Journal of Immunology*, *186*(10), 5675–5686. https://doi.org/10.4049/jimmunol.1003711

Matthews, P. C., Listgarten, J., Carlson, J. M., Payne, R., Huang, K.-H. G., Frater, J., Goedhals, D., Steyn, D., van Vuuren, C., Paioni, P., Jooste, P., Ogwu, A., Shapiro, R., Mncube, Z., Ndung'u, T., Walker, B. D., Heckerman, D., & Goulder, P. J. R. (2012). Co-Operative Additive Effects between HLA Alleles in Control of HIV-1. *PLoS ONE*, *7*(10), e47799. https://doi.org/10.1371/journal.pone.0047799

Mavilio, D., Benjamin, J., Daucher, M., Lombardo, G., Kottilil, S., Planta, M. A., Marcenaro, E., Bottino, C., Moretta, L., Moretta, A., & Fauci, A. S. (2003). Natural killer cells in HIV-1 infection: Dichotomous effects of viremia on inhibitory and activating receptors and their functional correlates. *Proceedings of the National Academy of Sciences*, *100*(25), 15011–15016. https://doi.org/10.1073/pnas.2336091100

McIntosh, K., SHEVITZ, A., ZAKNUN, D., KORNEGAY, J., CHATIS, P., KARTHAS, N., & BURCHETT, S. K. (1996). Age- and Time-related Changes in Extracellular Viral Load in Children Vertically Infected by Human Immunodeficiency Virus. *The Pediatric Infectious Disease Journal*, *15*(12), 1087–1091. https://doi.org/10.1097/00006454-199612000-00006

McLaren, P. J., & Carrington, M. (2015). The impact of host genetic variation on infection with HIV-1. *Nature Immunology*, *16*(6), 577–583. https://doi.org/10.1038/ni.3147

McLaren, P. J., Ripke, S., Pelak, K., Weintrob, A. C., Patsopoulos, N. A., Jia, X., Erlich, R. L., Lennon, N. J., Kadie, C. M., Heckerman, D., Gupta, N., Haas, D. W., Deeks, S. G., Pereyra, F., Walker, B. D., & de Bakker, P. I. W. (2012). Fine-mapping classical HLA variation associated with durable host control of HIV-1 infection in African Americans. *Human Molecular Genetics*, *21*(19), 4334–4347. https://doi.org/10.1093/hmg/dds226

McMichael, A. J., Borrow, P., Tomaras, G. D., Goonetilleke, N., & Haynes, B. F. (2010). The immune response during acute HIV-1 infection: Clues for vaccine development. *Nature Reviews Immunology*, *10*(1), 11–23. https://doi.org/10.1038/nri2674

Mehra, N., Sharma, G., & Kaur, G. (2011). Genetic correlates influencing immunopathogenesis of HIV infection. *The Indian Journal of Medical Research*, *134*(6), 749. https://doi.org/10.4103/0971-5916.92623

Mendoza, D., Royce, C., Ruff, L. E., Ambrozak, D. R., Quigley, M. F., Dang, T., Venturi, V., Price, D. A., Douek, D. C., Migueles, S. A., & Connors, M. (2012). HLA B*5701-Positive Long-Term Nonprogressors/Elite Controllers Are Not Distinguished from Progressors by the Clonal Composition of HIV-Specific CD8+ T Cells. *Journal of Virology*, *86*(7), 4014–4018. https://doi.org/10.1128/JVI.06982-11

Mendoza, J. A., Matshaba, M., Makhanda, J., Liu, Y., Boitshwarelo, M., & Anabwani, G. M. (2014). Food Insecurity and CD4% Among HIV+ Children in Gaborone, Botswana. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *66*(4), 414–418. https://doi.org/10.1097/QAI.0000000000000190

Meng, B., & Lever, A. M. L. (2013). Wrapping up the bad news – HIV assembly and release. *Retrovirology*, *10*(5), 1–12.

Migueles, S. A., Mendoza, D., Zimmerman, M. G., Martins, K. M., Toulmin, S. A., Kelly, E. P., Peterson, B. A., Johnson, S. A., Galson, E., Poropatich, K. O., Patamawenu, A., Imamichi, H., Ober, A., Rehm, C. A., Jones, S., Hallahan, C. W., Follmann, D. A., & Connors, M. (2015). CD8+ T-cell Cytotoxic Capacity Associated with Human Immunodeficiency Virus-1 Control Can Be Mediated through Various Epitopes and Human Leukocyte Antigen Types. *EBioMedicine*, *2*(1), 46–58. https://doi.org/10.1016/j.ebiom.2014.12.009

Migueles, S. A., Sabbaghian, M. S., Shupert, W. L., Bettinotti, M. P., Marincola, F. M., Martino, L., Hallahan, C. W., Selig, S. M., Schwartz, D., Sullivan, J., & Connors, M. (2000). HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proceedings of the National Academy of Sciences*, *97*(6), 2709–2714. https://doi.org/10.1073/pnas.050567397

Mkhwanazi, N., Thobakgale, C. F., van der Stok, M., Reddy, S., Mncube, Z., Chonco, F., Walker, B. D., Altfeld, M., Goulder, P. J. R., & Ndung'u, T. (2010). Immunodominant HIV-1-specific HLA-B- and HLA-C-restricted CD8+ T cells do not differ in polyfunctionality. *Virology*, *405*(2), 483–491. https://doi.org/10.1016/j.virol.2010.06.002

Mlotshwa, B. C. B. C., Mwesigwa, S., Mboowa, G., Williams, L., Retshabile, G., Kekitiinwa, A., Wayengera, M., Kyobe, S., Brown, C. W. C. W., Hanchard, N. A. N. A., Mardon, G., Joloba, M., Anabwani, G., & Mpoloka, S. W. S. W. (2017). The collaborative African genomics network training program: A trainee perspective on training the next generation of African scientists. *Genetics in Medicine*, *19*(7), 826–833. https://doi.org/10.1038/gim.2016.177

Mobbs, J. I., Illing, P. T., Dudek, N. L., Brooks, A. G., Baker, D. G., Purcell, A. W., Rossjohn, J., & Vivian, J. P. (2017). The molecular basis for peptide repertoire selection in the human leucocyte antigen (HLA) C∗06:02 molecule. *Journal of Biological Chemistry*, *292*(42), 17203–17215. https://doi.org/10.1074/jbc.M117.806976

Mody, A., Bartz, S., Hornik, C. P., Kiyimba, T., Bain, J., Muehlbauer, M., Kiboneka, E., Stevens, R., St. Peter, J. V., Newgard, C. B., Bartlett, J., & Freemark, M. (2014). Effects of HIV infection on the metabolic and hormonal status of children with severe acute malnutrition. *PLoS ONE*, *9*(7), 1–11. https://doi.org/10.1371/journal.pone.0102233

Montesano, C., Bonanno, C. T., Grifoni, A., Sano, C. Di, Palma, P., Castelli-Gattinara, G., Mattei, M., Mancino, G., Salerno, A., Colizzi, V., & Amicosante, M. (2014). Impact of Human Leukocyte Antigen Polymorphisms in Human Immunodeficiency Virus Progression in a Paediatric Cohort Infected with a Mono-phyletic Human Immunodeficiency Virus-1 Strain. *Journal of AIDS & Clinical Research*, *5*(3), 1–6. https://doi.org/10.4172/2155-6113.

Moore, C. B., John, M., James, I. R., Christiansen, F. T., Witt, C. S., & Mallal, S. A. (2002). Evidence of HIV-1 Adaptation to HLA-Restricted Immune Responses at a Population Level. *Science*, *296*(May), 1439–1444.

Mori, M., Wichukchinda, N., Miyahara, R., Rojanawiwat, A., Pathipvanich, P., Maekawa, T., Miura, T., Goulder, P., Yasunami, M., Ariyoshi, K., & Sawanpanyalert, P. (2014). HLA-B*35:05 is a protective allele with a unique structure among HIV-1 CRF01-AE-infected Thais, in whom the B*57 frequency is low. *AIDS*, *28*(7), 959–967. https://doi.org/10.1097/QAD.0000000000000206

Motozono, C., Kuse, N., Sun, X., Rizkallah, P. J., Fuller, A., Oka, S., Cole, D. K., Sewell, A. K., & Takiguchi, M. (2014). Molecular Basis of a Dominant T Cell Response to an HIV Reverse Transcriptase 8-mer Epitope Presented by the Protective Allele HLA-B*51:01. *The Journal of Immunology*, *192*(7), 3428–3434. https://doi.org/10.4049/jimmunol.1302667

Muenchhoff, M., Healy, M., Singh, R., Roider, J., Groll, A., Kindra, C., Sibaya, T., Moonsamy, A., McGregor, C., Phan, M. Q., Palma, A., Kloverpris, H., Leslie, A., Bobat, R., LaRussa, P., Ndung'u, T., Goulder, P., Sobieszczyk, M. E., & Archary, M. (2018). Malnutrition in HIV-Infected Children Is an Indicator of Severe Disease with an Impaired Response to Antiretroviral Therapy. *AIDS Research and Human Retroviruses*, *34*(1), 46–55. https://doi.org/10.1089/aid.2016.0261

Muenchhoff, M., Prendergast, A. J., & Goulder, P. J. R. (2014). Immunity to HIV in Early Life. *Frontiers in Immunology*, *5*. https://doi.org/10.3389/fimmu.2014.00391

Mulherin, S. A., O'Brien, T. R., Ioannidis, J. P. A., Goedert, J. J., Buchbinder, S. P., Coutinho, R. A., Jamieson, B. D., Meyer, L., Michael, N. L., Pantaleo, G., Rizzardi, G. P., Schuitemaker, H., Sheppard, H. W., Theodorou, I. D., Vlahov, D., & Rosenberg, P. S. (2003). Effects of CCR5-Δ32 and CCR2-64I alleles on HIV-1 disease progression. *AIDS*, *17*(3), 377–387. https://doi.org/10.1097/00002030-200302140-00012

Munderi, P., Snowden, W. B., Walker, A. S., Kityo, C., Mosteller, M., Kabuye, G., Thoofer, N. K., Ssali, F., Gilks, C. F., & Hughes, A. R. (2011). Distribution of HLA-B alleles in a Ugandan HIV-infected adult population: NORA pharmacogenetic substudy of DART. *Tropical Medicine and International Health*, *16*(2), 200–204. https://doi.org/10.1111/j.1365-3156.2010.02688.x

Mwesigwa, S., Williams, L., Retshabile, G., Katagirya, E., Mboowa, G., Mlotshwa, B., Kyobe, S., Kateete, D. P., Wampande, E. M., Wayengera, M., Mpoloka, S. W., Mirembe, A. N., Kasvosve, I., Morapedi, K., Kisitu, G. P., Kekitiinwa, A. R., Anabwani, G., Joloba, M. L., Matovu, E., … Hanchard, N. A. (2021). Unmapped exome reads implicate a role for Anelloviridae in childhood HIV-1 long-term non-progression. *Npj Genomic Medicine*, *6*(1), 24. https://doi.org/10.1038/s41525-021-00185-w

Nájera, O., González, C., Toledo, G., López, L., & Ortiz, R. (2004). Flow Cytometry Study of Lymphocyte Subsets in Malnourished and Well-Nourished Children with Bacterial Infections. *Clinical and Vaccine Immunology*, *11*(3), 577–580. https://doi.org/10.1128/CDLI.11.3.577-580.2004

Nakamura, K. J., Heath, L., Sobrera, E. R., Wilkinson, T. A., Semrau, K., Kankasa, C., Tobin, N. H., Webb, N. E., Lee, B., Thea, D. M., Kuhn, L., Mullins, J. I., & Aldrovandi, G. M. (2017). Breast milk and in utero transmission of HIV-1 select for envelope variants with unique molecular signatures. *Retrovirology*, *14*(1), 6. https://doi.org/10.1186/s12977-017-0331-z

Nakimuli, A., Chazara, O., Farrell, L., Hiby, S. E., Tukwasibwe, S., Knee, O., Jayaraman, J.,

Traherne, J. A., Elliott, A. M., Kaleebu, P., Mirembe, F., & Moffett, A. (2013). Killer cell immunoglobulin-like receptor (KIR) genes and their HLA-C ligands in a Ugandan population. *Immunogenetics*, *65*(11), 765–775. https://doi.org/10.1007/s00251-013-0724-7

Nakku, J., Kinyanda, E., & Hoskins, S. (2013). Prevalence and factors associated with probable HIV dementia in an African population: A cross-sectional study of an HIV/AIDS clinic population. *BMC Psychiatry*, *13*(1), 126. https://doi.org/10.1186/1471-244X-13-126

Namazi, G., Fajnzylber, J. M., Aga, E., Bosch, R. J., Acosta, E. P., Sharaf, R., Hartogensis, W., Jacobson, J. M., Connick, E., Volberding, P., Skiest, D., Margolis, D., Sneller, M. C., Little, S. J., Gianella, S., Smith, D. M., Kuritzkes, D. R., Gulick, R. M., Mellors, J. W., … Li, J. Z. (2018). The Control of HIV After Antiretroviral Medication Pause (CHAMP) Study: Posttreatment Controllers Identified From 14 Clinical Studies. *The Journal of Infectious Diseases*, *218*(12), 1954–1963. https://doi.org/10.1093/infdis/jiy479

Nariai, N., Kojima, K., Saito, S., Mimori, T., Sato, Y., Kawai, Y., Yamaguchi-Kabata, Y., Yasuda, J., & Nagasaki, M. (2015). HLA-VBSeq: Accurate HLA typing at full resolution from whole-genome sequencing data. *BMC Genomics*, *16*(Suppl 2), 1–6. https://doi.org/10.1186/1471-2164-16-S2-S7

Naruto, T., Gatanaga, H., Nelson, G., Sakai, K., Carrington, M., Oka, S., & Takiguchi, M. (2012). HLA Class I-Mediated Control of HIV-1 in the Japanese Population, in Which the Protective HLA-B*57 and HLA-B*27 Alleles Are Absent. *Journal of Virology*, *86*(19), 10870–10872. https://doi.org/10.1128/JVI.00689-12

Narzi, D., Becker, C. M., Fiorillo, M. T., Uchanska-Ziegler, B., Ziegler, A., & Böckmann, R. A. (2012). Dynamical characterization of two differentially disease associated MHC class i proteins in complex with viral and self-peptides. *Journal of Molecular Biology*, *415*(2), 429–442. https://doi.org/10.1016/j.jmb.2011.11.021

Natarajan, K., Li, H., Mariuzza, R. A., & Margulies, D. H. (2009). MHC class I molecules, structure and function. In *Seikagaku The Journal Of Japanese Biochemical Society* (Vol. 1, Issue 3, pp. 189–199). http://www.ncbi.nlm.nih.gov/pubmed/19378785

Navis, M., Schellens, I., van Baarle, D., Borghans, J., van Swieten, P., Miedema, F., Kootstra, N., & Schuitemaker, H. (2007). Viral Replication Capacity as a Correlate of HLA B57/B5801-Associated Nonprogressive HIV-1 Infection. *The Journal of Immunology*, *179*(5), 3133–3143. https://doi.org/10.4049/jimmunol.179.5.3133

Ndongala, M. L., Kamya, P., Boulet, S., Peretz, Y., Rouleau, D., Tremblay, C., Leblanc, R., Côté, P., Baril, J., Thomas, R., Vézina, S., Boulassel, M. R., Routy, J.-P., Sékaly, R. P., & Bernard, N. F. (2010). Changes in Function of HIV-Specific T-Cell Responses with Increasing Time from Infection. *Viral Immunology*, *23*(2), 159–168. https://doi.org/10.1089/vim.2009.0084

Ndongala, M. L., Peretz, Y., Boulet, S., Doroudchi, M., Yassine-Diab, B., Boulassel, M.-R., Rouleau, D., Tremblay, C., LeBlanc, R., Routy, J.-P., Sékaly, R.-P., & Bernard, N. F. (2009). HIV Gag p24 specific responses secreting IFN-γ and/or IL-2 in treatment-naïve individuals in acute infection early disease (AIED) are associated with low viral load. *Clinical Immunology*, *131*(2), 277–287. https://doi.org/10.1016/j.clim.2008.11.006

Nezafat, N., Karimi, Z., Eslami, M., Mohkam, M., Zandian, S., & Ghasemi, Y. (2016). Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. *Computational Biology and Chemistry*, *62*, 82–95. https://doi.org/10.1016/j.compbiolchem.2016.04.006

Ngumbela, K. C., Day, C. L., Mncube, Z., Nair, K., Ramduth, D., Thobakgale, C., Moodley, E., Reddy, S., de Pierres, C., Mkhwanazi, N., Bishop, K., van der Stok, M., Ismail, N., Honeyborne, I., Crawford, H., Kavanagh, D. G., Rousseau, C., Nickle, D., Mullins, J., … Walker, B. D. (2008). Targeting of a CD8 T Cell Env Epitope Presented by HLA-B*5802 Is Associated with Markers of HIV Disease Progression and Lack of Selection Pressure. *AIDS Research and Human Retroviruses*, *24*(1), 72–82. https://doi.org/10.1089/aid.2007.0124

Nielsen, K., Mcsherry, G., Petru, A., Frederick, T., Wara, D., Bryson, Y., Martin, N., Hutto, C., Ammann, A. J., Grubman, S., Oleske, J., & Scott, G. B. (1997). A Descriptive Survey of Pediatric Human Immunodeficiency Virus–infected Long-term Survivors. *Pediatrics*, *99*(4).

Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., & Buus, S. (2007). NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence. *PLoS ONE*, *2*(8), e796. https://doi.org/10.1371/journal.pone.0000796

Nissen, S. K., Christiansen, M., Helleberg, M., Kjær, K., Jørgensen, S. E., Ge, J., Katzenstein, T. L., Benfield, T., Kronborg, G., Larsen, C. S., Laursen, A., Pedersen, G., Jakobsen, M. R., Tolstrup, M., & Mogensen, T. H. (2018). Whole Exome Sequencing of identifies rare variants in genes encoding innate immune sensors and signaling molecules. *Scientific Reports*, *8*(15253), 1–15. https://doi.org/10.1038/s41598-018-33481-0

Nixon, S. A., Hanass-Hancock, J., Whiteside, A., & Barnett, T. (2011). The increasing chronicity of HIV in sub-Saharan Africa: Re-thinking "HIV as a long-wave event" in the era of widespread access to ART. *Globalization and Health*, *7*(41), 1–5. https://doi.org/10.1186/1744-8603-7-41

Novitsky, V., Flores-Villanueva, P. O., Chigwedere, P., Gaolekwe, S., Bussman, H., Sebetso, G., Marlink, R., Yunis, E. J., & Essex, M. (2001). Identification of most frequent HLA class I antigen specificities in Botswana: Relevance for HIV vaccine design. *Human Immunology*, *62*(2), 146–156. https://doi.org/10.1016/S0198-8859(00)00236-6

Ogunshola, F., Anmole, G., Miller, R. L., Goering, E., Nkosi, T., Muema, D., Mann, J., Ismail, N., Chopera, D., Ndung'u, T., Brockman, M. A., & Ndhlovu, Z. M. (2018). Dual HLA B*42 and B*81-reactive T cell receptors recognize more diverse HIV-1 Gag escape variants. *Nature Communications*, *9*(1), 5023. https://doi.org/10.1038/s41467-018-07209-7

Okoko, N. A., Owuor, K. O., Kulzer, J. L., Owino, G. O., Ogolla, I. A., Wandera, R. W., Bukusi, E. A., Cohen, C. R., & Abuogi, L. L. (2017). Factors associated with mother to child transmission of HIV despite overall low transmission rates in HIV-exposed infants in rural Kenya. *International Journal of STD & AIDS*, *28*(12), 1215–1223. https://doi.org/10.1177/0956462417693735

Okoye, A. A., & Picker, L. J. (2013). CD4+ T cell depletion in HIV infection: mechanisms of

immunological failure. *Immunol Rev*, *254*(1), 54–64. https://doi.org/10.1111/imr.12066

Ostrowski, M. A., Gu, J. X., Kovacs, C., Freedman, J., Luscher, M. A., & MacDonald, K. S. (2001). Quantitative and Qualitative Assessment of Human Immunodeficiency Virus Type 1 (HIV-1)–Specific CD4 + T Cell Immunity to gag in HIV-1–Infected Individuals with Differential Disease Progression: Reciprocal Interferon-γ and Interleukin-10 Responses. *The Journal of Infectious Diseases*, *184*(10), 1268–1278. https://doi.org/10.1086/324005

Pandey, K., Wang, S. S., Mifsud, N. A., Faridi, P., Davenport, A. J., Webb, A. I., Sandow, J. J., Ayala, R., Monje, M., Cross, R. S., Ramarathinam, S. H., Jenkins, M. R., & Purcell, A. W. (2023). A combined immunopeptidomics, proteomics, and cell surface proteomics approach to identify immunotherapy targets for diffuse intrinsic pontine glioma. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1192448

Pandey, R. K., Ojha, R., Aathmanathan, V. S., Krishnan, M., & Prajapati, V. K. (2018). Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. *Vaccine*, *36*(17), 2262–2272. https://doi.org/10.1016/j.vaccine.2018.03.042

Pant Pai, N., Shivkumar, S., & Cajas, J. M. (2012). Does Genetic Diversity of HIV-1 Non-B Subtypes Differentially Impact Disease Progression in Treatment-Naive HIV-1–Infected Individuals? A Systematic Review of Evidence. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *59*(4), 382–388. https://doi.org/10.1097/QAI.0b013e31824a0628

Papúchová, H., Meissner, T. B., Li, Q., Strominger, J. L., & Tilburgs, T. (2019). The Dual Role of HLA-C in Tolerance and Immunity at the Maternal-Fetal Interface. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.02730

Parham, P., Norman, P. J., Abi-Rached, L., & Guethlein, L. A. (2012). Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 800–811. https://doi.org/10.1098/rstb.2011.0266

Parolini, F., Biswas, P., Serena, M., Sironi, F., Muraro, V., Guizzardi, E., Cazzoletti, L., Scupoli, M. T., Gibellini, D., Ugolotti, E., Biassoni, R., Beretta, A., Malnati, M., Romanelli, M. G., & Zipeto, D. (2018). Stability and Expression Levels of HLA-C on the Cell Membrane Modulate HIV-1 Infectivity. *Journal of Virology*, *92*(1). https://doi.org/10.1128/JVI.01711-17

Patyka, M., Malamud, D., Weissman, D., Abrams, W. R., & Kurago, Z. (2015). Periluminal Distribution of HIV-Binding Target Cells and Gp340 in the Oral, Cervical and Sigmoid/Rectal Mucosae: A Mapping Study. *PLOS ONE*, *10*(7), e0132942. https://doi.org/10.1371/journal.pone.0132942

Paul, M. E., Mao, C., Charurat, M., Serchuck, L., Foca, M., Hayani, K., Handelsman, E. L., Diaz, C., McIntosh, K., & Shearer, W. T. (2005). Predictors of immunologic long-term nonprogression in HIV-infected children: Implications for initiating therapy. *Journal of Allergy and Clinical Immunology*, *115*(4), 848–855. https://doi.org/10.1016/j.jaci.2004.11.054

Paximadis, M., Minevich, G., Winchester, R., Schramm, D. B., Gray, G. E., Sherman, G. G., Coovadia, A. H., Kuhn, L., & Tiemessen, C. T. (2011). KIR-HLA and maternal-infant HIV-1 transmission in sub-Saharan Africa. *PLoS ONE*, *6*(2), e1654. https://doi.org/10.1371/journal.pone.0016541

Paxton, W. A., & Kang, S. (1998). Chemokine receptor allelic polymorphisms: Relationships to HIV resistance and disease progression. *Seminars in Immunology*, *10*(3), 187–194. https://doi.org/10.1006/smim.1998.0132

Payne, R., Muenchhoff, M., Mann, J., Roberts, H. E., Matthews, P., Adland, E., Hempenstall, A., Huang, K.-H., Brockman, M., Brumme, Z., Sinclair, M., Miura, T., Frater, J., Essex, M., Shapiro, R., Walker, B. D., Ndung'u, T., McLean, A. R., Carlson, J. M., & Goulder, P. J. R. (2014). Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proceedings of the National Academy of Sciences*, *111*(50), E5393–E5400. https://doi.org/10.1073/pnas.1413339111

Pei, J., & Grishin, N. V. (2001). AL2CO: calculation of positional conservation in a protein

sequence alignment. *Bioinformatics*, *17*(8), 700–712.

Pelak, K., Goldstein, D. B., Walley, N. M., Fellay, J., Ge, D., Shianna, K. V., Gumbs, C., Gao, X., Maia, J. M., Cronin, K. D., Hussain, S. K., Carrington, M., Michael, N. L., & Weintrob, A. C. (2010). Host Determinants of HIV-1 Control in African Americans. *The Journal of Infectious Diseases*, *201*(8), 1141–1149. https://doi.org/10.1086/651382

Peloso, G. M., Rader, D. J., Gabriel, S., Kathiresan, S., Daly, M. J., & Neale, B. M. (2016). Phenotypic extremes in rare variant study designs. *European Journal of Human Genetics*, *24*(6), 924–930. https://doi.org/10.1038/ejhg.2015.197

Pende, D., Falco, M., Vitale, M., Cantoni, C., Vitale, C., Munari, E., Bertaina, A., Moretta, F., Del Zotto, G., Pietra, G., Mingari, M. C., Locatelli, F., & Moretta, L. (2019). Killer Ig-Like Receptors (KIRs): Their Role in NK Cell Modulation and Developments Leading to Their Clinical Exploitation. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.01179

Peretz, Y., Ndongala, M. L., Boulet, S., Boulassel, M. R., Rouleau, D., Côté, P., Longpré, D., Routy, J.-P., Falutz, J., Tremblay, C., Tsoukas, C. M., Sékaly, R. P., & Bernard, N. F. (2007). Functional T cell subsets contribute differentially to HIV peptide-specific responses within infected individuals: Correlation of these functional T cell subsets with markers of disease progression. *Clinical Immunology*, *124*(1), 57–68. https://doi.org/10.1016/j.clim.2007.04.004

Peter, K., Men, Y., Pantaleo, G., Gander, B., & Corradin, G. (2001). Induction of a cytotoxic T-cell response to HIV-1 proteins with short synthetic peptides and human compatible adjuvants. *Vaccine*, *19*(30), 4121–4129. https://doi.org/10.1016/S0264-410X(01)00179-7

Peters, B., Bulik, S., Tampe, R., van Endert, P. M., & Holzhütter, H.-G. (2003). Identifying MHC Class I Epitopes by Predicting the TAP Transport Efficiency of Epitope Precursors. *The Journal of Immunology*, *171*(4), 1741–1749. https://doi.org/10.4049/jimmunol.171.4.1741

Peterson, T. A., Kimani, J., Wachihi, C., Bielawny, T., Mendoza, L., Thavaneswaran, S., Narayansingh, M. J., Kariri, T., Liang, B., Ball, T. B., Ngugi, E. N., Plummer, F. A., & Luo, M. (2013). HLA class I associations with rates of HIV-1 seroconversion and disease progression in the Pumwani Sex Worker Cohort. *Tissue Antigens*, *81*(2), 93–107. https://doi.org/10.1111/tan.12051

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Pineda-Peña, A.-C., Faria, N. R., Imbrechts, S., Libin, P., Abecasis, A. B., Deforche, K., Gómez-López, A., Camacho, R. J., de Oliveira, T., & Vandamme, A.-M. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infection, Genetics and Evolution*, *19*, 337–348. https://doi.org/10.1016/j.meegid.2013.04.032

Poorolajal, J., Hooshmand, E., Mahjub, H., Esmailnasab, N., & Jenabi, E. (2016). Survival rate of AIDS disease and mortality in HIV-infected patients: a meta-analysis. *Public Health*, *139*, 3–12. https://doi.org/10.1016/j.puhe.2016.05.004

Prendergast, A. J., Klenerman, P., & Goulder, P. J. R. (2012). The impact of differential antiviral immunity in children and adults. *Nature Reviews Immunology*, *12*(9), 636–648. https://doi.org/10.1038/nri3277

Purcell, A. W., McCluskey, J., & Rossjohn, J. (2007). More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews Drug Discovery*, *6*(5), 404–414. https://doi.org/10.1038/nrd2224

Pymm, P., Tenzer, S., Wee, E., Weimershaus, M., Burgevin, A., Kollnberger, S., Gerstoft, J., Josephs, T. M., Ladell, K., McLaren, J. E., Appay, V., Price, D. A., Fugger, L., Bell, J. I., Schild, H., van Endert, P., Harkiolaki, M., & Iversen, A. K. N. (2022). Epitope length variants balance protective immune responses and viral escape in HIV-1 infection. *Cell Reports*, *38*(9). https://doi.org/10.1016/j.celrep.2022.110449

Ramsbottom, K. A., Carr, D. F., Jones, A. R., & Rigden, D. J. (2018). Critical assessment of approaches for molecular docking to elucidate associations of HLA alleles with adverse drug reactions. *Molecular Immunology*, *101*(August), 488–499. https://doi.org/10.1016/j.molimm.2018.08.003

Ramsuran, V., Naranbhai, V., Horowitz, A., Qi, Y., Martin, M. P., Yuki, Y., Gao, X., Walker-sperling, V., Prete, G. Q. Del, Schneider, D. K., Lifson, J. D., Fellay, J., Deeks, S. G., Martin, J. N., Goedert, J. J., Wolinsky, S. M., Michael, N. L., Kirk, G. D., Buchbinder, S., … Carrington, M. (2018). Elevated HLA-A expression impairs HIV control through inhibition of NKG2A-expressing cells. *Science*, *359*(January), 86–90.

Reboul, C. F., Meyer, G. R., Porebski, B. T., Borg, N. A., & Buckle, A. M. (2012). Epitope Flexibility and Dynamic Footprint Revealed by Molecular Dynamics of a pMHC-TCR Complex. *PLoS Computational Biology*, *8*(3), e1002404. https://doi.org/10.1371/journal.pcbi.1002404

Retshabile, G., Mlotshwa, B. C., Williams, L., Mwesigwa, S., Mboowa, G., Huang, Z., Rustagi, N., Swaminathan, S., Katagirya, E., Kyobe, S., Wayengera, M., Kisitu, G. P., Kateete, D. P., Wampande, E. M., Maplanka, K., Kasvosve, I., Pettitt, E. D., Matshaba, M., Nsangi, B., … Hanchard, N. A. (2018). Whole-Exome Sequencing Reveals Uncaptured Variation and Distinct Ancestry in the Southern African Population of Botswana. *The American Journal of Human Genetics*, *102*(5), 731–743. https://doi.org/10.1016/j.ajhg.2018.03.010

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*(W1), W449–W454. https://doi.org/10.1093/nar/gkaa379

Robinson, J. (2001). IMGT/HLA Database – a sequence database for the human major histocompatibility complex. *Nucleic Acids Research*, *29*(1), 210–213. https://doi.org/10.1093/nar/29.1.210

Robinson, J., Barker, D. J., Georgiou, X., Cooper, M. A., Flicek, P., & Marsh, S. G. E. (2019). IPD-IMGT/HLA Database. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz950

Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P., & Marsh, S. G. E. (2015). The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Research*, *43*(D1), D423–D431. https://doi.org/10.1093/nar/gku1161

Rochman, Y., Spolski, R., & Leonard, W. J. (2009). New insights into the regulation of T cells by γc family cytokines. *Nature Reviews Immunology*, *9*(7), 480–490. https://doi.org/10.1038/nri2580

Rohowsky-Kochan, C., Skurnick, J., Molinaro, D., & Louria, D. (1998). HLA antigens associated with susceptibility/resistance to HIV-1 infection. *Human Immunology*, *59*(12), 802–815. https://doi.org/10.1016/S0198-8859(98)00086-X

Roider, J. M., Muenchhoff, M., & Goulder, P. J. R. (2016). Immune activation and paediatric HIV-1 disease outcome. *Current Opinion in HIV and AIDS*, *11*(2), 146–155. https://doi.org/10.1097/COH.0000000000000231

Román, V. R. G., Jensen, K. J., Jensen, S. S., Leo-Hansen, C., Jespersen, S., Té, D. da S., Rodrigues, C. M., Janitzek, C. M., Vinner, L., Katzenstein, T. L., Andersen, P., Kromann, I., Andreasen, L. V., Karlsson, I., & Fomsgaard, A. (2013). Therapeutic Vaccination Using Cationic Liposome-Adjuvanted HIV Type 1 Peptides Representing HLA-Supertype-Restricted Subdominant T Cell Epitopes: Safety, Immunogenicity, and Feasibility in Guinea-Bissau. *AIDS Research and Human Retroviruses*, *29*(11), 1504–1512. https://doi.org/10.1089/aid.2013.0076

Rosás-Umbert, M., Gunst, J. D., Pahus, M. H., Olesen, R., Schleimann, M., Denton, P. W., Ramos, V., Ward, A., Kinloch, N. N., Copertino, D. C., Escribà, T., Llano, A., Brumme, Z. L., Brad Jones, R., Mothe, B., Brander, C., Fox, J., Nussenzweig, M. C., Fidler, S., …

Søgaard, O. S. (2022). Administration of broadly neutralizing anti-HIV-1 antibodies at ART initiation maintains long-term CD8+ T cell immunity. *Nature Communications*, *13*(1), 6473. https://doi.org/10.1038/s41467-022-34171-2

Rose, A. M., Hall, C. S., & Martinez-Alier, N. (2014). Aetiology and management of malnutrition in HIV-positive children. *Archives of Disease in Childhood*, *99*(6), 546–551. https://doi.org/10.1136/archdischild-2012-303348

Ross, S. H., & Cantrell, D. A. (2018). Signaling and Function of Interleukin-2 in T Lymphocytes. *Annual Review of Immunology*, *36*(1), 411–433. https://doi.org/10.1146/annurev-immunol-042617-053352

Ruppert, J., Sidney, J., Celis, E., Kubo, R. T., Grey, H. M., & Sette, A. (1993). Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell*, *74*(5), 929–937. https://doi.org/10.1016/0092-8674(93)90472-3

Rytter, M. J. H., Kolte, L., Briend, A., Friis, H., & Christensen, V. B. (2014). The Immune System in Children with Malnutrition—A Systematic Review. *PLoS ONE*, *9*(8), e105017. https://doi.org/10.1371/journal.pone.0105017

Sabin, C. A., & Lundgren, J. D. (2013). The natural history of HIV infection. *Current Opinion in HIV and AIDS*, *8*(4), 311–317. https://doi.org/10.1097/COH.0b013e328361fa66

Sáez-Cirión, A., Bacchus, C., Hocqueloux, L., Avettand-Fenoel, V., Girault, I., Lecuroux, C., Potard, V., Versmisse, P., Melard, A., Prazuck, T., Descours, B., Guergnon, J., Viard, J.-P., Boufassa, F., Lambotte, O., Goujard, C., Meyer, L., Costagliola, D., Venet, A., … Rouzioux, C. (2013). Post-Treatment HIV-1 Controllers with a Long-Term Virological Remission after the Interruption of Early Initiated Antiretroviral Therapy ANRS VISCONTI Study. *PLoS Pathogens*, *9*(3), e1003211. https://doi.org/10.1371/journal.ppat.1003211

Sampathkumar, R., Peters, H. O., Mendoza, L., Bielawny, T., Ngugi, E., Kimani, J., Wachihi, C., Plummer, F. A., & Luo, M. (2014). Influence of HLA class I haplotypes on HIV-1 seroconversion and disease progression in Pumwani sex worker cohort. *PLoS ONE*, *9*(7),

1–12. https://doi.org/10.1371/journal.pone.0101475

Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C.-M., Saragosti, S., Lapouméroulie, C., Cognaux, J., Forceille, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R. J., Collman, R. G., … Parmentier, M. (1996). Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*, *382*(6593), 722–725. https://doi.org/10.1038/382722a0

Sandström, E., Nilsson, C., Hejdeman, B., Bråve, A., Bratt, G., Robb, M., Cox, J., VanCott, T., Marovich, M., Stout, R., Aboud, S., Bakari, M., Pallangyo, K., Ljungberg, K., Moss, B., Earl, P., Michael, N., Birx, D., Mhalu, F., … Biberfeld, G. (2008). Broad Immunogenicity of a Multigene, Multiclade HIV-1 DNA Vaccine Boosted with Heterologous HIV-1 Recombinant Modified Vaccinia Virus Ankara. *The Journal of Infectious Diseases*, *198*(10), 1482–1490. https://doi.org/10.1086/592507

Santos, K. B., Guedes, I. A., Karl, A. L. M., & Dardenne, L. E. (2020). Highly Flexible Ligand Docking: Benchmarking of the DockThor Program on the LEADS-PEP Protein–Peptide Data Set. *Journal of Chemical Information and Modeling*, *60*(2), 667–683. https://doi.org/10.1021/acs.jcim.9b00905

Saxová, P., Buus, S., Brunak, S., & Keşmir, C. (2003). Predicting proteasomal cleavage sites: A comparison of available methods. *International Immunology*, *15*(7), 781–787. https://doi.org/10.1093/intimm/dxg084

Schaefer, M. R., Williams, M., Kulpa, D. A., Blakely, P. K., Yaffee, A. Q., & Collins, K. L. (2008). A Novel Trafficking Signal within the HLA-C Cytoplasmic Tail Allows Regulated Expression upon Differentiation of Macrophages. *The Journal of Immunology*, *180*(12), 7804–7817. https://doi.org/10.4049/jimmunol.180.12.7804

Schmitt-Haendle, M., Bachmann, O., Harrer, E., Schmidt, B., Bäuerle, M., & Harrer, T. (2005). Recognition Patterns of HLA-A2–Restricted Human Immunodeficiency Virus-1–Specific Cytotoxic T-Lymphocytes in a Cohort of HIV-1–Infected Individuals. *Viral Immunology*, *18*(4), 627–636. https://doi.org/10.1089/vim.2005.18.627

Serwanga, J., Shafer, L. A., Pimego, E., Auma, B., Watera, C., Rowland, S., Yirrell, D., Pala, P., Grosskurth, H., Whitworth, J., Gotch, F., & Kaleebu, P. (2009). Host HLA B allele-associated multi-clade gag T-cell recognition correlates with slow HIV-1 disease progression in antiretroviral theraphy-naïve Ugandans. *PLoS ONE*, *4*(1), e4188. https://doi.org/10.1371/journal.pone.0004188

Sewale, Y., Hailu, G., Sintayehu, M., Moges, N. A., & Alebel, A. (2018). Magnitude of malnutrition and associated factors among HIV infected children attending HIV-care in three public hospitals in East and West Gojjam Zones, Amhara, Northwest, Ethiopia, 2017: a cross-sectional study. *BMC Research Notes*, *11*(1), 788. https://doi.org/10.1186/s13104-018-3882-8

Shearer, W. T., Quinn, T. C., LaRussa, P., Lew, J. F., Mofenson, L., Almy, S., Rich, K., Handelsman, E., Diaz, C., Pagano, M., Smeriglio, V., & Kalish, L. A. (1997). Viral Load and Disease Progression in Infants Infected with Human Immunodeficiency Virus Type 1. *New England Journal of Medicine*, *336*(19), 1337–1342. https://doi.org/10.1056/NEJM199705083361901

Shen, J., Guo, T., Wang, T., Zhen, Y., Ma, X., Wang, Y., Zhang, Z. X., Cai, J. P., Mao, W., Zhu, F. M., Li, J. P., Wang, Z. L., Zhang, D. M., Liu, M. L., Shan, X. Y., Zhang, B. W., Zhu, C. F., Deng, Z. H., Yu, W. J., … Baier, L. J. (2018). HLA-B∗07, HLA-DRB1∗07, HLA-DRB1∗12, and HLA-c∗03:02 strongly associate with BMI: Data from 1.3 million healthy Chinese adults. *Diabetes*, *67*(5), 861–871. https://doi.org/10.2337/db17-0852

Shepherd, B. L., Ferrand, R., Munyati, S., Folkard, S., Boyd, K., Bandason, T., Jallow, S., Rowland-Jones, S. L., & Yindom, L. M. (2015). HLA correlates of long-term survival in vertically infected HIV-1-positive adolescents in Harare, Zimbabwe. *AIDS Research and Human Retroviruses*, *31*(5), 504–507. https://doi.org/10.1089/aid.2014.0338

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M. P., Chavan, S., Vergara, C., Ortega, V. E., Levin, A. M., Eng, C., Yazdanbakhsh, M., Wilson, J. G., Marrugo, J., Lange, L. A., Williams, L. K., Watson, H., Ware, L. B., … Salzberg, S. L. (2019a). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nature Genetics*, *51*(2), 30–35. https://doi.org/10.1038/s41588-018-0335-1

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M. P., Chavan, S., Vergara, C., Ortega, V. E., Levin, A. M., Eng, C., Yazdanbakhsh, M., Wilson, J. G., Marrugo, J., Lange, L. A., Williams, L. K., Watson, H., Ware, L. B., … Salzberg, S. L. (2019b). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nature Genetics*, *51*(2), 30–35. https://doi.org/10.1038/s41588-018-0273-y

Shiina, T., Blancher, A., Inoko, H., & Kulski, J. K. (2017). Comparative genomics of the human, macaque and mouse major histocompatibility complex. *Immunology*, *150*(2), 127–138. https://doi.org/10.1111/imm.12624

Shoko, C., & Chikobvu, D. (2019). A superiority of viral load over CD4 cell count when predicting mortality in HIV patients on therapy. *BMC Infectious Diseases*, *19*(1), 169. https://doi.org/10.1186/s12879-019-3781-1

Sidney, J., Vela, J. L., Friedrich, D., Kolla, R., von Herrath, M., Wesley, J. D., & Sette, A. (2018). Low HLA binding of diabetes-associated CD8+ T-cell epitopes is increased by post translational modifications. *BMC Immunology*, *19*(1), 12. https://doi.org/10.1186/s12865-018-0250-3

Simon, A. K., Hollander, G. A., & McMichael, A. (2015). Evolution of the immune system in humans from infancy to old age. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1821), 20143085. https://doi.org/10.1098/rspb.2014.3085

Singh, A., Kaushik, R., Mishra, A., Shanker, A., & Jayaram, B. (2016). ProTSAV: A protein tertiary structure analysis and validation server. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1864*(1), 11–19. https://doi.org/10.1016/j.bbapap.2015.10.004

Singh, K. K., Gray, P. K., Wang, Y., Fenton, T., Trout, R. N., & Spector, S. A. (2011). HLA Alleles are Associated With Altered Risk for Disease Progression and Central Nervous System Impairment of HIV-Infected Children. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *57*(1), 32–39. https://doi.org/10.1097/QAI.0b013e3182119244

Singh, K. K., & Spector, \Stephen A. (2009). Host Genetic Determinants of Human

Immunodeficiency Virus Infection and Disease Progression in Children. *Pediatric Research*, *65*(5 Part 2), 55R-63R. https://doi.org/10.1203/PDR.0b013e31819dca03

Smith, M. W., Carrington, M., Winkler, C., Lomb, D., Dean, M., Huttley, G., & O'brien, S. (1997). CCR2 chemokine receptor and AIDS progression. *Nature Medicine*, *3*(10), 1052–1053. https://doi.org/10.1038/nm1097-1052c

Smith, M. W., Dean, M., Carrington, M., Huttley, G. A., & O'Brien, S. J. (1997). CCR5-Δ32 gene deletion in HIV-1 infected patients. *The Lancet*, *350*(9079), 741. https://doi.org/10.1016/S0140-6736(05)63551-9

Smith, M. W., Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Lomb, D. A., Goedert, J. J., O'Brien, T. R., Jacobson, L. P., Kaslow, R., Buchbinder, S., Vittinghoff, E., Vlahov, D., Hoots, K., Hilgartner, M. W., Study, Multicenter AIDS Cohort Study (MACS), M. H., & O'Brien, S. J. (1997). Contrasting Genetic Influence of CCR2 and CCR5 Variants on HIV-1 Infection and Disease Progression. *Science*, *277*(5328), 959–965. https://doi.org/10.1126/science.277.5328.959

Soriano, A., Martínez, C., García, F., Plana, M., Palou, E., Lejeune, M., Aróstegui, J. I., De Lazzari, E., Rodriguez, C., Barrasa, A., Lorenzo, J. I., Alcamí, J., del Romero, J., Miró, J. M., Gatell, J. M., & Gallart, T. (2002). Plasma Stromal Cell–Derived Factor (SDF)-1 Levels, SDF1-3′A Genotype, and Expression of CXCR4 on T Lymphocytes: Their Impact on Resistance to Human Immunodeficiency Virus Type 1 Infection and Its Progression. *The Journal of Infectious Diseases*, *186*(7), 922–931. https://doi.org/10.1086/343741

Spira, S., Wainberg, M. A., Loemba, H., Turner, D., & Brenner, B. G. (2003). Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *Journal of Antimicrobial Chemotherapy*, *51*(2), 229–240. https://doi.org/10.1093/jac/dkg079

Sriwanthana, B., Hodge, T., Mastro, T. D., Dezzutti, C. S., Bond, K., Stephens, H. A. F., Kostrikis, L. G., Limpakarnjanarat, K., Young, N. L., Qari, S. H., Lal, R. B., Chandanayingyong, D., & McNicholl, J. M. (2001). HIV-Specific Cytotoxic T Lymphocytes, HLA-A11, and Chemokine-Related Factors May Act Synergistically to Determine HIV Resistance in CCR5 Δ32-Negative Female Sex Workers in Chiang Rai,

Northern Thailand. *AIDS Research and Human Retroviruses*, *17*(8), 719–734. https://doi.org/10.1089/088922201750236997

Ssemwanga, D., Nsubuga, R. N., Mayanja, B. N., Lyagoba, F., Magambo, B., Yirrell, D., van der Paal, L., Grosskurth, H., & Kaleebu, P. (2013). Effect of HIV-1 Subtypes on Disease Progression in Rural Uganda: A Prospective Clinical Cohort Study. *PLoS ONE*, *8*(8), 1–7. https://doi.org/10.1371/journal.pone.0071768

Ssewanyana, I., Elrefaei, M., Dorsey, G., Ruel, T., Jones, N. G., Gasasira, A., Kamya, M., Nakiwala, J., Achan, J., Charlebois, E., Havlir, D., & Cao, H. (2007). Profile of T Cell Immune Responses in HIV-Infected Children from Uganda. *The Journal of Infectious Diseases*, *196*(11), 1667–1670. https://doi.org/10.1086/522013

Stewart-Jones, G. B. E., Gillespie, G., Overton, I. M., Kaul, R., Roche, P., McMichael, A. J., Rowland-Jones, S., & Jones, E. Y. (2005). Structures of Three HIV-1 HLA-B*5703-Peptide Complexes and Identification of Related HLAs Potentially Associated with Long-Term Nonprogression. *The Journal of Immunology*, *175*(4), 2459–2468. https://doi.org/10.4049/jimmunol.175.4.2459

Sundquist, W. I., & Krausslich, H. (2012). HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine*, *2*.

Sunguya, B. F., Poudel, K. C., Otsuka, K., Yasuoka, J., Mlunde, L. B., Urassa, D. P., Mkopi, N. P., & Jimba, M. (2011). Undernutrition among HIV-positive children in Dar es Salaam, Tanzania: Antiretroviral therapy alone is not enough. *BMC Public Health*, *11*(869), 1–11. https://doi.org/10.1186/1471-2458-11-869

Svejgaard, A., & Ryder, L. P. (1994). HLA and disease associations: Detecting the strongest association. *Tissue Antigens*, *43*(1), 18–27. https://doi.org/10.1111/j.1399-0039.1994.tb02291.x

Sverchkova, A., Anzar, I., Stratford, R., & Clancy, T. (2019). Improved HLA typing of Class I and Class II alleles from next-generation sequencing data. *HLA*, *94*(6), 504–513. https://doi.org/10.1111/tan.13685

Tang, J., Malhotra, R., Song, W., Brill, I., Hu, L., Farmer, P. K., Mulenga, J., Allen, S., Hunter, E., & Kaslow, R. A. (2010). Human leukocyte antigens and HIV type 1 viral load in early and chronic infection: Predominance of evolving relationships. *PLoS ONE*, *5*(3), e9629. https://doi.org/10.1371/journal.pone.0009629

Tchetgen Tchetgen, E. J., Phiri, K., & Shapiro, R. (2015). A Simple Regression-based Approach to Account for Survival Bias in Birth Outcomes Research. *Epidemiology*, *26*(4), 473–480. https://doi.org/10.1097/EDE.0000000000000317

Thobakgale, C. F., Prendergast, A., Crawford, H., Mkhwanazi, N., Ramduth, D., Reddy, S., Molina, C., Mncube, Z., Leslie, A., Prado, J., Chonco, F., Mphatshwe, W., Tudor-Williams, G., Jeena, P., Blanckenberg, N., Dong, K., Kiepiela, P., Coovadia, H., Ndung'u, T., … Goulder, P. J. R. (2009). Impact of HLA in Mother and Child on Disease Progression of Pediatric Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology*, *83*(19), 10234–10244. https://doi.org/10.1128/jvi.00921-09

Thomas, R., Apps, R., Qi, Y., Gao, X., Male, V., O'hUigin, C., O'Connor, G., Ge, D., Fellay, J., Martin, J. N., Margolick, J., Goedert, J. J., Buchbinder, S., Kirk, G. D., Martin, M. P., Telenti, A., Deeks, S. G., Walker, B. D., Goldstein, D., … Carrington, M. (2009). HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nature Genetics*, *41*(12), 1290–1294. https://doi.org/10.1038/ng.486

Tiemessen, C. T., Paximadis, M., Minevich, G., Winchester, R., Shalekoff, S., Gray, G. E., Sherman, G. G., Coovadia, A. H., & Kuhn, L. (2011). Natural Killer Cell Responses to HIV-1 Peptides are Associated With More Activating KIR Genes and HLA-C Genes of the C1 Allotype. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *57*(3), 181–189. https://doi.org/10.1097/QAI.0b013e3182174a76

Tobin, N. H., & Aldrovandi, G. M. (2013). Immunology of pediatric HIV infection. *Immunological Reviews*, *254*(1), 143–169. https://doi.org/10.1111/imr.12074

Tobin, N. H., & Aldrovandi, G. M. (2014). Are Infants Unique in Their Ability to be "Functionally Cured" of HIV-1? *Current HIV/AIDS Reports*, *11*(1), 1–10. https://doi.org/10.1007/s11904-013-0189-1

Toh, H., Savoie, C. J., Kamikawaji, N., Muta, S., Sasazuki, T., & Kuhara, S. (2000). Changes at the floor of the peptide-binding groove induce a strong preference for proline at position 3 of the bound peptide: Molecular dynamics simulations of HLA-A*0217. *Biopolymers*, *54*(5), 318–327. https://doi.org/10.1002/1097-0282(20001015)54:5<318::AID-BIP30>3.0.CO;2-T

Tokić, S., Žižkova, V., Štefanić, M., Glavaš-Obrovac, L., Marczi, S., Samardžija, M., Sikorova, K., & Petrek, M. (2020). HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 allele and haplotype frequencies defined by next generation sequencing in a population of East Croatia blood donors. *Scientific Reports*, *10*(5513), Article number: 5513. https://doi.org/10.1038/s41598-020-62175-9

Touloumi, G., Pantazis, N., Pillay, D., Paraskevis, D., Chaix, M.-L., Bucher, H. C., Kucherer, C., Zangerle, R., Kran, A.-M. B., Porter, K., Zangerle, R., Kelleher, A. D., Cooper, D. A., Grey, P., Finlayson, R., Bloch, M., Kelleher, T., Ramacciotti, T., Gelgor, L., … Pillay, D. (2013). Impact of HIV-1 Subtype on CD4 Count at HIV Seroconversion, Rate of Decline, and Viral Load Set Point in European Seroconverter Cohorts. *Clinical Infectious Diseases*, *56*(6), 888–897. https://doi.org/10.1093/cid/cis1000

Trachtenberg, E. A., & Erlich, H. a. (2001). A Review of the Role of the Human Leukocyte Antigen (HLA) System as a Host Immunogenetic Factor Influencing HIV Transmission and Progression to AIDS. In *HIV Molecular Immunology* (Vol. 43, Issue December). www.cdc.gov/mmwr/preview/mmwrhtml/00018871.htm

Trachtenberg, E., & Erlich, H. (2001). A review of the role of HLA system as a host immunogenetic factor influencing HIV transmission and progression to AIDS. *HIV Molecular Immunology*, *43*(December), 43–60.

Tsalimalma, K., Kordossis, T., & Choremi-Papadopoulou, E. (2011). Antiretroviral Therapy-Induced Dominant Interleukin-2 HIV-1 Gag CD4+ T Cell Response: Evidence of Functional Recovery of HIV-1-Specific CD4+ T Cells. *Scandinavian Journal of Immunology*, *73*(3), 256–265. https://doi.org/10.1111/j.1365-3083.2010.02502.x

Tshabalala, M., Ingram, C., Schlaphoff, T., Borrill, V., Christoffels, A., & Pepper, M. S.

(2018). Human Leukocyte Antigen-A, B, C, DRB1, and DQB1 Allele and Haplotype Frequencies in a Subset of 237 Donors in the South African Bone Marrow Registry. *Journal of Immunology Research*, *2018*, 1–8. https://doi.org/10.1155/2018/2031571

Tshabalala, Morse, E., & Zijenah, L. (2013). HLA Genetic Polymorphisms: Role in HIV-1 Susceptibility, Disease Progression and Treatment Outcomes. *Retrovirology: Research and Treatment*, 1. https://doi.org/10.4137/RRT.S11589

UCLA-DOE LAB — SAVES v6.0. (n.d.). Retrieved September 11, 2021, from https://saves.mbi.ucla.edu/

Ueta, M., Kaniwa, N., Sotozono, C., Tokunaga, K., Saito, Y., Sawai, H., Miyadera, H., Sugiyama, E., Maekawa, K., Nakamura, R., Nagato, M., Aihara, M., Matsunaga, K., Takahashi, Y., Furuya, H., Muramatsu, M., Ikezawa, Z., & Kinoshita, S. (2014). Independent strong association of HLA-A*02:06 and HLA-B*44:03 with cold medicine-related Stevens-Johnson syndrome with severe mucosal involvement. *Scientific Reports*, *4*(1), 4862. https://doi.org/10.1038/srep04862

UNAIDS. (2019). *UNAIDS 2019 Data*. Available at Https://Www.Unaids.Org/Sites/Default/Files/Media_asset/2019-UNAIDS-Data_en.Pdf Accessed April 7, 2020. https://doi.org/978-92-9173-945-5

UNAIDS DATA. (2023). https://www.unaids.org/sites/default/files/media_asset/data-book-2023_en.pdf

Usman Mirza, M., Rafique, S., Ali, A., Munir, M., Ikram, N., Manan, A., Salo-Ahen, O. M. H., & Idrees, M. (2016). Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific Reports*, *6*(1), 37313. https://doi.org/10.1038/srep37313

van Deutekom, H. W. M., & Keşmir, C. (2015). Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? *Immunogenetics*, *67*(8), 425–436. https://doi.org/10.1007/s00251-015-0849-y

van Paassen, P. M., van Pul, L., van der Straten, K., Buchholtz, N. V. J. E., Grobben, M., van

Nuenen, A. C., van Dort, K. A., Boeser-Nunnink, B. D., van den Essenburg, M. D., Burger, J. A., van Luin, M., Jurriaans, S., Sanders, R. W., Swelsen, W. T., Symons, J., Klouwens, M. J., Nijhuis, M., van Gils, M. J., Prins, J. M., … Kootstra, N. A. (2023). Virological and immunological correlates of HIV posttreatment control after temporal antiretroviral therapy during acute HIV infection. *AIDS*, *37*(15), 2297–2304. https://doi.org/10.1097/QAD.0000000000003722

van Teijlingen, N. H., Hölzemer, A., Körner, C., García-Beltrán, W. F., Schafer, J. L., Fadda, L., Suscovich, T. J., Brander, C., Carrington, M., Evans, D. T., van Baarle, D., & Altfeld, M. (2014). Sequence variations in HIV-1 p24 Gag-derived epitopes can alter binding of KIR2DL2 to HLA-C*03. *AIDS*, *28*(10), 1399–1408. https://doi.org/10.1097/QAD.0000000000000284

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., & Mackerell, A. D. (2009). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, NA-NA. https://doi.org/10.1002/jcc.21367

Vargas, L. de B., Dourado, R. M., Amorim, L. M., Ho, B., Calonga-Solís, V., Issler, H. C., Marin, W. M., Beltrame, M. H., Petzl-Erler, M. L., Hollenbach, J. A., & Augusto, D. G. (2020). Single Nucleotide Polymorphism in KIR2DL1 Is Associated With HLA-C Expression in Global Populations. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.01881

Vaurs, J., Douchin, G., Echasserieau, K., Oger, R., Jouand, N., Fortun, A., Hesnard, L., Croyal, M., Pecorari, F., Gervois, N., & Bernardeau, K. (2021). A novel and efficient approach to high-throughput production of HLA-E/peptide monomer for T-cell epitope screening. *Scientific Reports*, *11*(1), 17234. https://doi.org/10.1038/s41598-021-96560-9

Venner, C. M., Nankya, I., Kyeyune, F., Demers, K., Kwok, C., Chen, P.-L., Rwambuya, S., Munjoma, M., Chipato, T., Byamugisha, J., Van Der Pol, B., Mugyenyi, P., Salata, R. A., Morrison, C. S., & Arts, E. J. (2016). Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa. *EBioMedicine*, *13*, 305–314.

https://doi.org/10.1016/j.ebiom.2016.10.014

Verweij, M. C., Horst, D., Griffin, B. D., Luteijn, R. D., Davison, A. J., Ressing, M. E., & Wiertz, E. J. H. J. (2015). Viral Inhibition of the Transporter Associated with Antigen Processing (TAP): A Striking Example of Functional Convergent Evolution. *PLoS Pathogens*, *11*(4), e1004743. https://doi.org/10.1371/journal.ppat.1004743

Vittinghoff, E., Hessol, N. A., Bacchetti, P., Fusaro, R. E., Holmberg, S. D., & Buchbinder, S. P. (2001). Cofactors for HIV Disease Progression in a Cohort of Homosexual and Bisexual Men. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *27*(3), 308–314. https://doi.org/10.1097/00126334-200107010-00015

Vivier, E., Tomasello, E., Baratin, M., Walzer, T., & Ugolini, S. (2008). Functions of natural killer cells. *Nature Immunology*, *9*(5), 503–510. https://doi.org/10.1038/ni1582

Vollmers, S., Lobermeyer, A., & Körner, C. (2021). The New Kid on the Block: HLA-C, a Key Regulator of Natural Killer Cells in Viral Immunity. *Cells*, *10*(11), 3108. https://doi.org/10.3390/cells10113108

Walker, B. D., Pereyra, F., Jia, X., McLaren, P. J., Telenti, A., Bakker, P. I. W. de, De Bakker, P. I. W., & Walker, B. D. (2010). The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. *Science*, *330*(6010), 1551–1557. https://doi.org/10.1126/science.1195271

Wan, S., Flower, D. R., & Coveney, P. V. (2008). Toward an atomistic understanding of the immune synapse: Large-scale molecular dynamics simulation of a membrane-embedded TCR-pMHC-CD4 complex. *Molecular Immunology*, *45*(5), 1221–1230. https://doi.org/10.1016/j.molimm.2007.09.022

Wang, C., Krishnakumar, S., Wilhelmy, J., Babrzadeh, F., Stepanyan, L., Su, L. F., Levinson, D., Fernandez-Vina, M. A., Davis, R. W., Davis, M. M., & Mindrinos, M. (2012). High-throughput, high-fidelity HLA genotyping with deep sequencing. *Proceedings of the National Academy of Sciences*, *109*(22), 8676–8681. https://doi.org/10.1073/pnas.1206614109

Wang, C., & Zhang, Y. (2017). Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest. *Journal of Computational Chemistry*, *38*(3), 169–177. https://doi.org/10.1002/jcc.24667

Wang, R., Lu, Y., & Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *Journal of Medicinal Chemistry*, *46*(12), 2287–2303. https://doi.org/10.1021/jm0203783

Warren, R. L., Choe, G., Freeman, D. J., Castellarin, M., Munro, S., Moore, R., & Holt, R. A. (2012). Derivation of HLA types from shotgun sequence datasets. *Genome Medicine*, *4*(12). https://doi.org/10.1186/gm396

Warszawski, J., Lechenadec, J., Faye, A., Dollfus, C., Firtion, G., Meyer, L., Douard, D., Monpoux, F., Tricoire, J., Benmebarek, Y., Rouzioux, C., & Blanche, S. (2007). Long-Term Nonprogression of HIV Infection in Children: Evaluation of the ANRS Prospective French Pediatric Cohort. *Clinical Infectious Diseases*, *45*(6), 785–794. https://doi.org/10.1086/521165

Watanabe, T., Murakoshi, H., Gatanaga, H., Koyanagi, M., Oka, S., & Takiguchi, M. (2011). Effective recognition of HIV-1-infected cells by HIV-1 integrase-specific HLA-B*4002-restricted T cells. *Microbes and Infection*, *13*(2), 160–166. https://doi.org/10.1016/j.micinf.2010.10.006

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess Jr, J. W., Swanstrom, R., Burch, C. L., & Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, *460*(7256), 711–716. https://doi.org/10.1038/nature08237

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370.

Wherry, E. J., Blattman, J. N., Murali-Krishna, K., van der Most, R., & Ahmed, R. (2003). Viral Persistence Alters CD8 T-Cell Immunodominance and Tissue Distribution and Results in Distinct Stages of Functional Impairment. *Journal of Virology*, *77*(8), 4911–4927. https://doi.org/10.1128/JVI.77.8.4911-4927.2003

Whitmire, J. K., Tan, J. T., & Whitton, J. L. (2005). Interferon-γ acts directly on CD8+ T cells to increase their abundance during virus infection. *The Journal of Experimental Medicine*, *201*(7), 1053–1059. https://doi.org/10.1084/jem.20041463

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017a). Major histocompatibility complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation. *Frontiers in Immunology*, *8*(MAR), 1–16. https://doi.org/10.3389/fimmu.2017.00292

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., & Freund, C. (2017b). Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Frontiers in Immunology*, *8*(Imci). https://doi.org/10.3389/fimmu.2017.00292

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(SUPPL.2), 407–410. https://doi.org/10.1093/nar/gkm290

Wittig, M., Anmarkrud, J. A., Kässens, J. C., Koch, S., Forster, M., Ellinghaus, E., Hov, J. R., Sauer, S., Schimmler, M., Ziemann, M., Görg, S., Jacob, F., Karlsen, T. H., & Franke, A. (2015). Development of a high-resolution NGS-based HLA-typing and analysis pipeline. *Nucleic Acids Research*, *43*(11), 1–8. https://doi.org/10.1093/nar/gkv184

World Health Statistics 2018: Monitoring Health for the SDGs, Sustainable Development Goals. (2018). https://doi.org/Licence: CC BY-NC-SA 3.0 IGO

Wright, J. K., Brumme, Z. L., Carlson, J. M., Heckerman, D., Kadie, C. M., Brumme, C. J., Wang, B., Losina, E., Miura, T., Chonco, F., van der Stok, M., Mncube, Z., Bishop, K., Goulder, P. J. R., Walker, B. D., Brockman, M. A., & Ndung'u, T. (2010). Gag-Protease-

Mediated Replication Capacity in HIV-1 Subtype C Chronic Infection: Associations with HLA Type and Clinical Parameters. *Journal of Virology*, *84*(20), 10820–10831. https://doi.org/10.1128/JVI.01084-10

Xia, Z., Chen, H., Kang, S., Huynh, T., Fang, J. W., Lamothe, P. A., Walker, B. D., & Zhou, R. (2014). The complex and specific pMHC interactions with diverse HIV-1 TCR clonotypes reveal a structural basis for alterations in CTL function. *Scientific Reports*, *4*(1), 4087. https://doi.org/10.1038/srep04087

Yang, F., Jiang, M., Zhang, W., Qiao, Y., Chen, S. an, Wang, D., Zhu, H., Zhang, J., Qin, S., Zhu, Q., Lv, Y., Xing, Q., & Luo, X. (2020). HLA-A*24:02 is associated with metronidazole-induced cutaneous adverse drug reactions in Han Chinese individuals: A pilot case-control study with both HLA gene and T cell receptor repertoire analysis. *Basic and Clinical Pharmacology and Toxicology*, *126*(2), 133–143. https://doi.org/10.1111/bcpt.13315

Yang, Y., Sun, W., Guo, J., Zhao, G., Sun, S., Yu, H., Guo, Y., Li, J., Jin, X., Du, L., Jiang, S., Kou, Z., & Zhou, Y. (2015). In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Human Vaccines and Immunotherapeutics*, *11*(3), 795–805. https://doi.org/10.1080/21645515.2015.1012017

Young, S., Murray, K., Mwesigwa, J., Natureeba, P., Osterbauer, B., Achan, J., Arinaitwe, E., Clark, T., Ades, V., Plenty, A., Charlebois, E., Ruel, T., Kamya, M., Havlir, D., & Cohan, D. (2012). Maternal nutritional status predicts adverse birth outcomes among HIV-infected rural Ugandan women receiving combination antiretroviral therapy. *PLoS ONE*, *7*(8), 1–7. https://doi.org/10.1371/journal.pone.0041934

Yusim, K., Szinger, J. J., & Honeyborne, I. (2003). Enhanced Motif Scan: A Tool to Scan for HLA Anchor Residues in Proteins. *HIV Immunology and HIV/SIV Vaccine Databases 2003*, 25–36.

Zappacosta, F., Borrego, F., Brooks, A. G., Parker, K. C., & Coligan, J. E. (1997). Peptides isolated from HLA-Cw*0304 confer different degrees of protection from natural killer cell-mediated lysis. *Proceedings of the National Academy of Sciences*, *94*(12), 6313–

6318. https://doi.org/10.1073/pnas.94.12.6313

Zhang, G. L., Keskin, D. B., Lin, H.-N., Lin, H. H., DeLuca, D. S., Leppanen, S., Milford, E. L., Reinherz, E. L., & Brusic, V. (2014). Human Leukocyte Antigen Typing Using a Knowledge Base Coupled with a High-Throughput Oligonucleotide Probe Array Analysis. *Frontiers in Immunology*, *5*. https://doi.org/10.3389/fimmu.2014.00597

Zhang, H., Zhao, B., Han, X., Wang, Z., Liu, B., Lu, C., Zhang, M., Liu, J., Chen, O., Hu, Q., Jiang, F., & Shang, H. (2013). Associations of HLA Class I antigen specificities and haplotypes with disease progression in HIV-1-infected Hans in Northern China. *Human Immunology*, *74*(12), 1636–1642. https://doi.org/10.1016/j.humimm.2013.08.287

Zhang, L., Cheng, D., Tao, N., Zhao, M., Zhang, F., Yuan, Y., & Qiu, X. (2012). Distribution of HLA-A, -B and -DRB1 Genes and Haplotypes in the Tujia Population Living in the Wufeng Region of Hubei Province, China. *PLoS ONE*, *7*(6), e38774. https://doi.org/10.1371/journal.pone.0038774

Zhang, S., Liu, J., Cheng, H., Tan, S., Qi, J., Yan, J., & Gao, G. F. (2011). Structural basis of cross-allele presentation by HLA-A*0301 and HLA-A*1101 revealed by two HIV-derived peptide complexes. *Molecular Immunology*, *49*(1–2), 395–401. https://doi.org/10.1016/j.molimm.2011.08.015

Zhao, L., Zhang, M., & Cong, H. (2013). Advances in the study of HLA-restricted epitope vaccines. *Human Vaccines & Immunotherapeutics*, *9*(12), 2566–2577.

Ziegler, M. C., Nelde, A., Weber, J. K., Schreitmüller, C. M., Martrus, G., Huynh, T., Bunders, M. J., Lunemann, S., Stevanovic, S., Zhou, R., & Altfeld, M. (2020). HIV-1 induced changes in HLA-C*03:04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. *AIDS (London, England)*, *34*(12), 1713–1723. https://doi.org/10.1097/QAD.0000000000002596

Zimmerli, S. C., Harari, A., Cellerai, C., Vallelian, F., Bart, P.-A., & Pantaleo, G. (2005). HIV-1-specific IFN-γ/IL-2-secreting CD8 T cells support CD4-independent proliferation of HIV-1-specific CD8 T cells. *Proceedings of the National Academy of Sciences*, *102*(20),

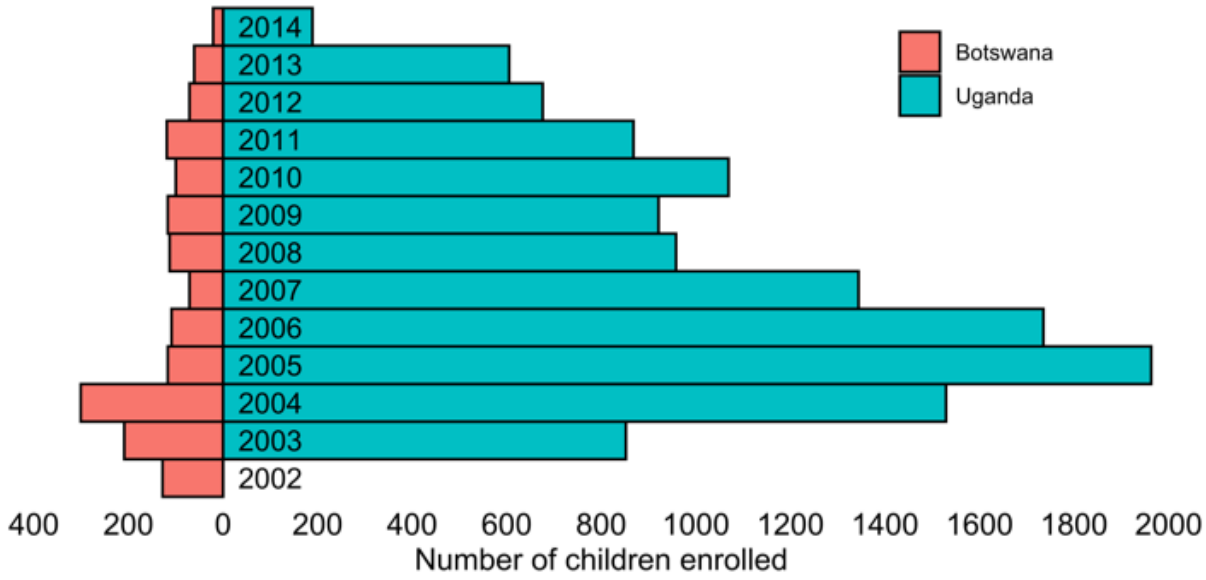
7239–7244. https://doi.org/10.1073/pnas.0502393102

Zipeto, D., & Beretta, A. (2012). HLA-C and HIV-1: friends or foes? *Retrovirology*, *9*(1), 39. https://doi.org/10.1186/1742-4690-9-39

Ziraldo, C., Vodovotz, Y., Namas, R. A., Almahmoud, K., Tapias, V., Mi, Q., Barclay, D., Jefferson, B. S., Chen, G., Billiar, T. R., & Zamora, R. (2013). Central role for MCP-1/CCL2 in injury-induced inflammation revealed by in vitro, in silico, and clinical studies. *PloS One*, *8*(12), e79804. https://doi.org/10.1371/journal.pone.0079804

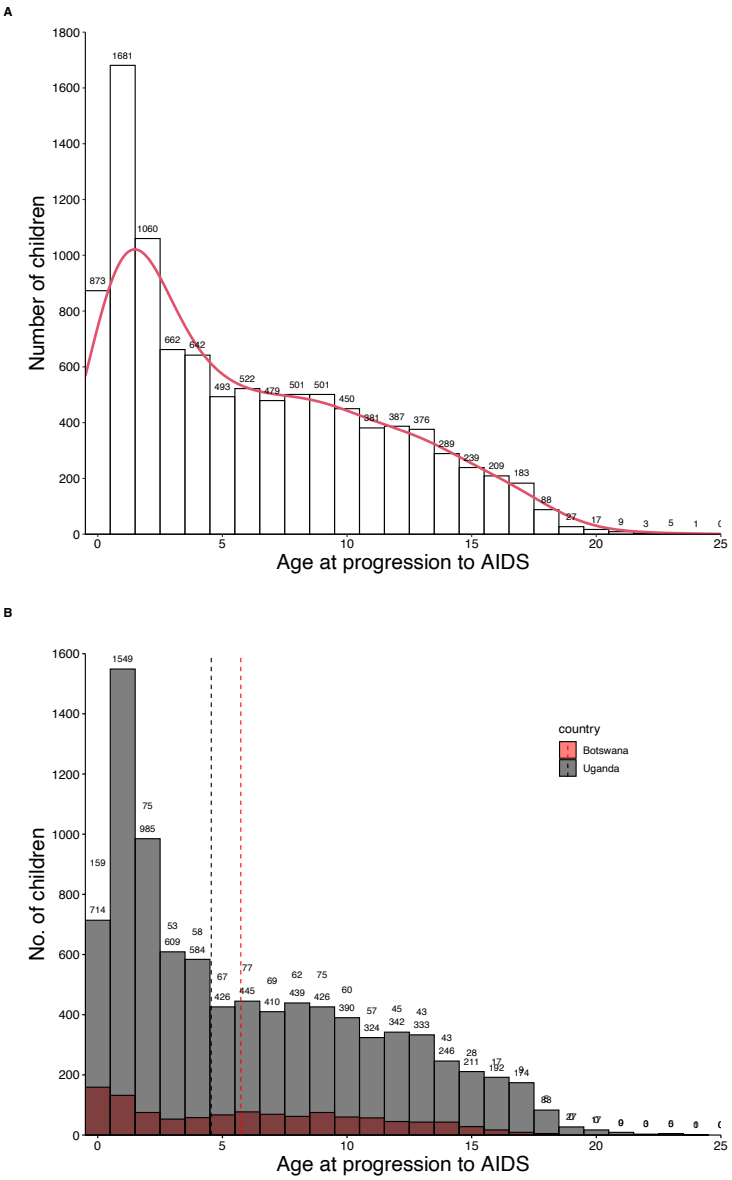
CHAPTER 9: APPENDIX

9.1 Supplementary Figures



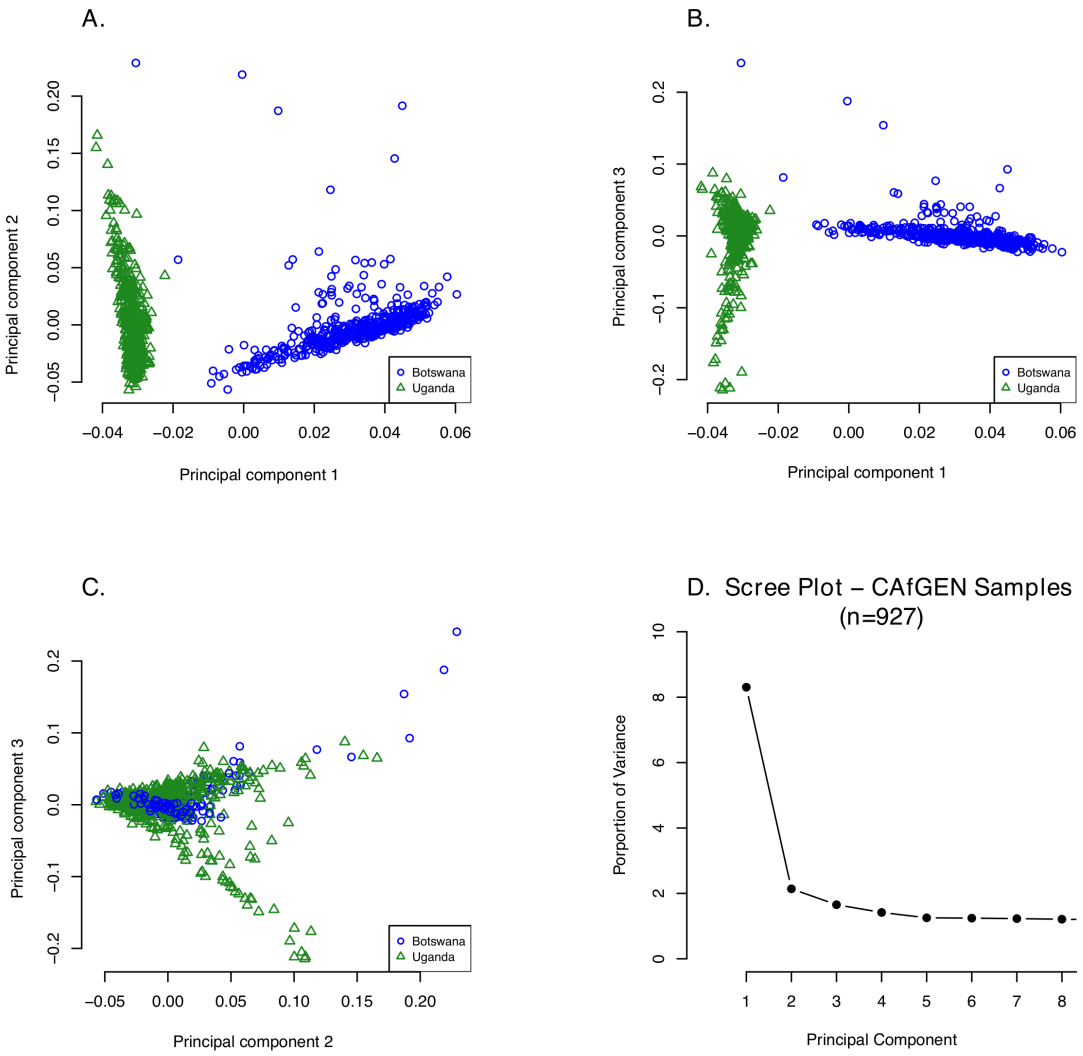
Supp. Figure 1: Number of children enrolled per year in Botswana and Uganda.

Enrolment into care at the Baylor Pediatric HIV Clinical Center of Excellence started in 2002 and 2003 in Botswana and Uganda, respectively.



Supp. Figure 2: Distribution of age of progression in the entire cohort (A) and in Uganda vs. Botswana (B).

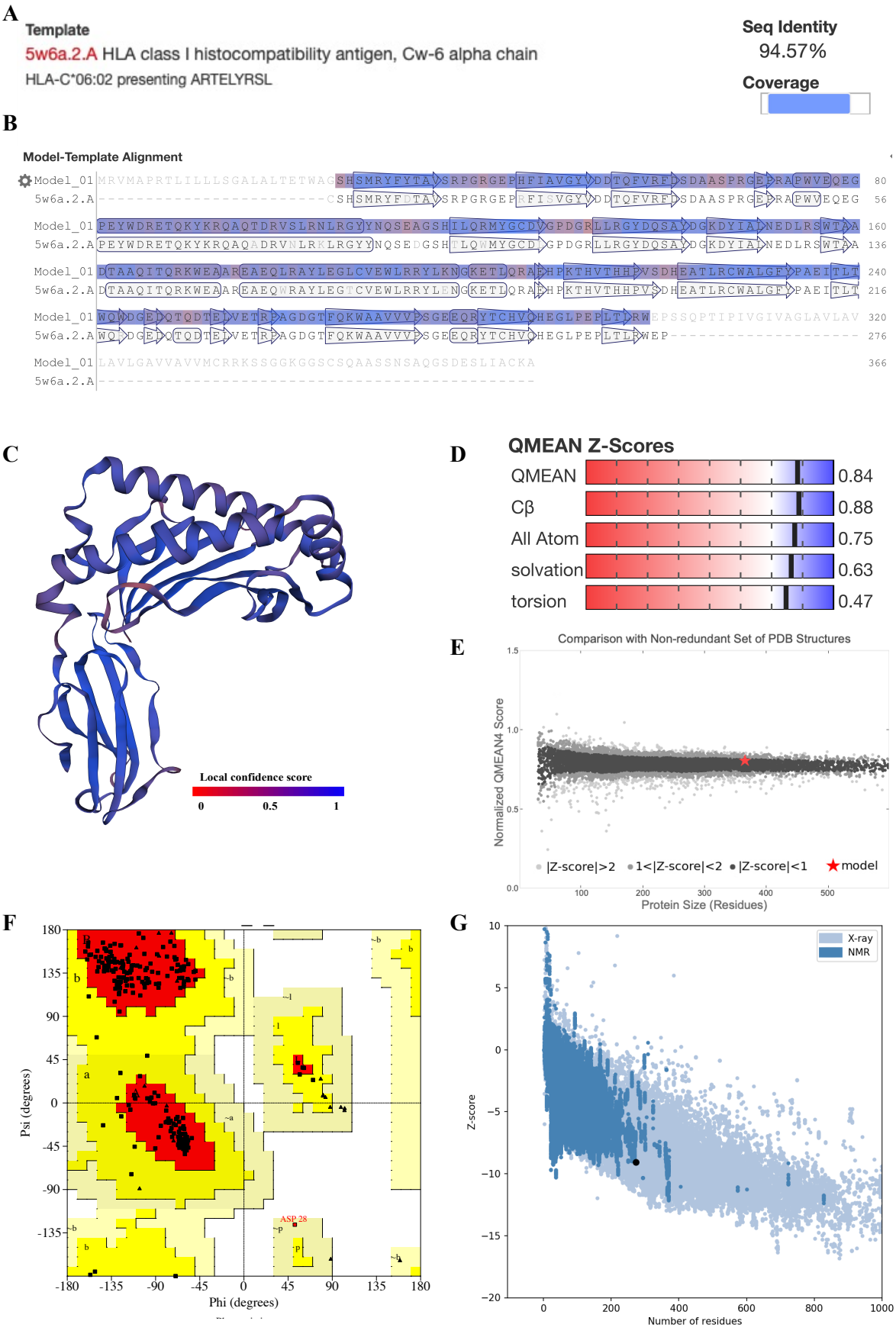
The dashed lines in Fig. S1B indicate the median age at progression in the two countries.



Supp. Figure 3: Principal component plots.

A and B show structure between the two CAfGEN populations along PC1. PC2 and PC3 show clines suggestive of gene flow within the respective populations. C shows the shared sub-Saharan ancestry between the two population. Most of the variation is from the structure between the two populations as seen in D.

Note: The Fst = 0.0065 was calculated using the SNPRelate package v1.10.2 in R v3.4.3. the working space was 198,319 autosomal biallelic SNPs, from 922 samples which were LD pruned with an r^2 = 0.2, in windows of 50 SNPs and steps of 5 SNPs for each population in addition to QC that was also run separatelyfor the tow populations (sex check, heterozygosity, HWE, genotype missingness, SNP missingness and relatedness)

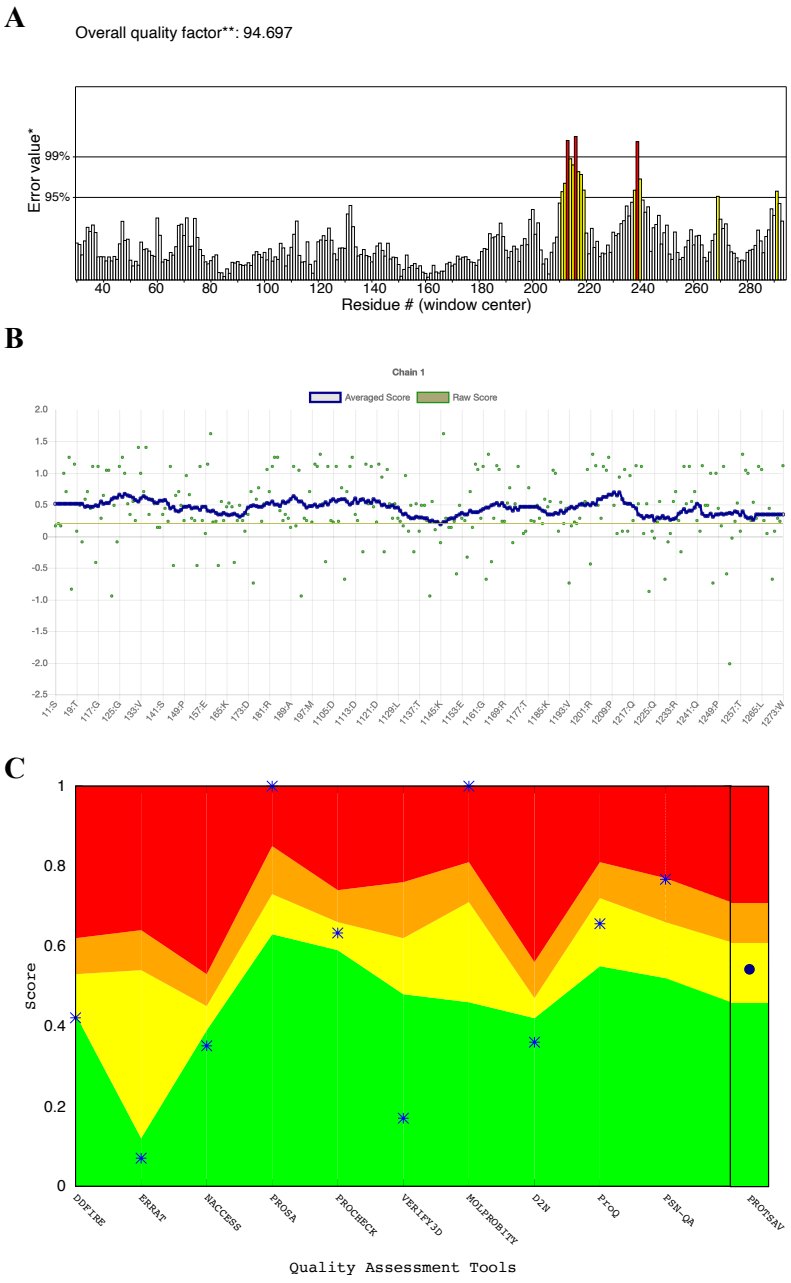

A
Template
5w6a.2.A HLA class I histocompatibility antigen, Cw-6 alpha chain
HLA-C*06:02 presenting ARTELYRSL
Seq Identity
94.57%
Coverage
B
Model-Template Alignment
Model_01
5w6a.2.A
C
Local confidence score
0
0.5
1
D
QMEAN Z-Scores
QMEAN
0.84
Cβ
0.88
All Atom
0.75
solvation
0.63
torsion
0.47
E
Comparison with Non-redundant Set of PDB Structures
Normalized QMEAN4 Score
|Z-score|>2
1<|Z-score|<2
|Z-score|<1
model
Protein Size (Residues)
F
Psi (degrees)
Phi (degrees)
ASP 28
G
X-ray
NMR
Z-score
Number of residues

Supp. Figure 4: Model building and validation

*(A) Parameters of 5w6a.2.A the template selected for model building. (B) Model-template alignment. The predicted model is colored by a QMEANDisCo score similar to Figure 1C. (C) 3D structure of the model colored by local residue confidence score, red being low and blue very high score estimated with the QMEANDisCo scoring function (range 0–1). An oval ring shows the peptide binding groove. (D) QMEAN z-scores of the model. All five scores are a comparison of the model with the expected experimentally determined structures of similar size. (E) Plot of the normalized QMEAN scores of the reference set used in constructing the protein model of C*0302. (F) Ramachandran plot shows the values of dihedral angles Φ and Ψ of each residue. Red indicates favored region, brown shows the allowed region, while yellow and light yellow are outlier regions. No residue lies outside of the favored region. (G) ProSA – Protein structural analysis plot of the overall 3D quality z-score of the model compared to X-ray and NMR solved structures. Abbreviations: QMEAN, Qualitative Model Energy Analysis; QMEANDisCo, Qualitative Model Energy Analysis Distance constraints.*

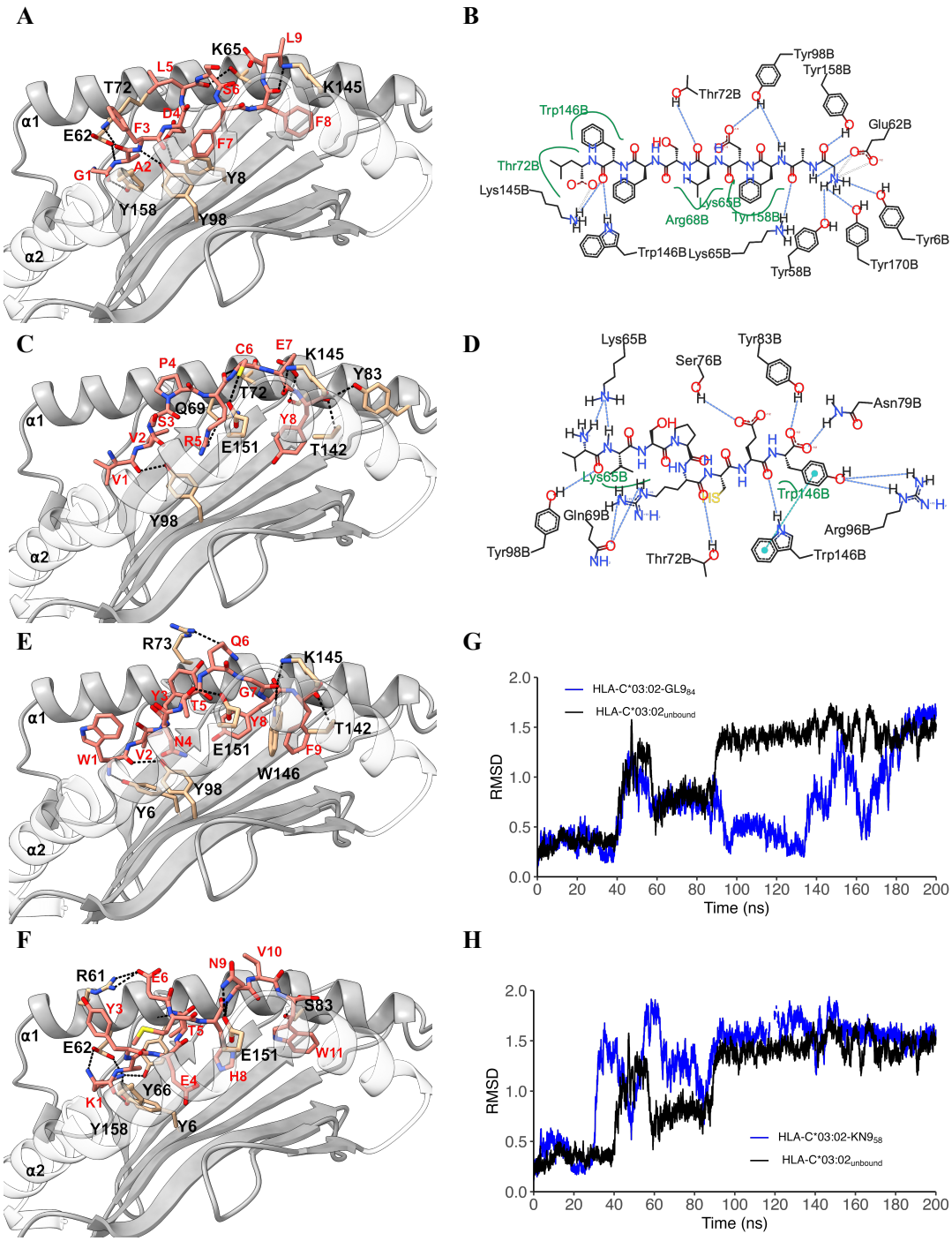
Accompanying Notes:

From the top 50 suitable templates generated by SWISS-MODEL, the best template for model building was 5w6a.2. (HLA-C*06:02) with 95% and 75% sequence identity and coverage, respectively (Fig. Sxx and Sxx). Overall (QMEAN z-score 0.84) and individual amino acid residues have very high confidence scores across the length of the model (Fig. Sxx, B and C). The stereochemical analysis found that 98.9% of the residues were located in the Ramachandran-favoured regions (Fig. SxxE), with an overall average G-factor of -0.11. No residue was found in the outlier regions demonstrating the model is of good quality. ProSA analysis identifies errors in the 3D structure of the model. The overall model ProSA z-score is -9.08, which is typical of native proteins resolved experimentally by X-ray or NMR (Fig. Sxx 1F). At least 96% of the amino acid in the predicted model have an averaged 3D-1D score ≥0.2 (Fig. S1A). Other stereochemical and spatial features are summarized in Table 1 (Fig. S1). The predicted model is stable with a ProTSAV score of 0.54 and root-mean-square distance (RMSD) within the range of 2-5Å (Fig. S1B). Overall the predicted model is of good quality and acceptable for ligand binding studies.



Supp. Figure 5: Modeling validation.

(A) ERRAT score for the selected model. Yellow color shows the error region between 95 and 99%, Red color shows the misfolded region, and white color shows the lower error rate of protein folding. (B) Verify-3D plot showed that 99.63% of the residues have averaged a 3D-1D score of ≥ 0.2. (C) ProTSAV score of 0.54 and root-mean-square distance (RMSD) within the range of 2-5Å, suggesting moderate accuracy of the model.



Supp. Figure 6: The molecular docking representation and RMSD analysis.

*HLA-C*03:02 with docked epitopes GL9 (A, B and G), VY8 (C&D), WF9 (E) and KN9 (F and H). (G) and (H) show the RMSD plot of GL9 and KN9 showing the lack of stability.*

9.2 Supplementary Tables

Supp. Table 1: Subdistribution hazard ratios of progression from competing risk models among children with viral load measurements

Characteristics §	Univariate sHR (95% CI)	p value	Multivariate aHR_{sd} (95% CI)	p value
Country		0.852		< .001
Botswana	1		1	
Uganda	1.01 (0.87 – 1.18)		0.40 (0.31 – 0.51)	
Gender, *n (%)*		0.194		
Male	1			
Female	0.91 (0.80 – 1.05)			
Age at Enrolment (per 1-yr increase)	0.11 (0.06 – 0.19)	< .001	0.04 (0.02 – 0.06)	< .001
Year Birth		< .001		
Before 2004	1			
After 2004	4.51 (3.25 – 6.26)			
Year of enrolment		0.038		< .001
Before 2006	1		1	
2006 – 2010	1.12 (0.91 – 1.38)		0.36 (0.23 – 0.54)	
After 2010	1.29 (1.05 – 1.59)		2.10 (1.59 – 2.79)	
HIV RNA load, (log10 copies/uL)	1.87 (1.52 – 2.30)	< .001	1.08 (0.76 – 1.53)	0.001
Weight-for-age z–score		0.004		
–1.75	1			
–3.75	1.06 (0.96 – 1.17)			
Height-for-age z–score		0.022		
–1.75	1			
–-3.75	1.02 (0.89 – 1.17)			
BMI-for-age z–score		< .001		0.347
–0.85	1		1	
–2.85	1.18 (1.06 – 1.30)		1.17 (1.03 – 1.33)	

§ N = 847, † BMI-for-age was not included in any model with weight-for-age or height-for-age, BMI - body mass index

Supp. Table 2: LD values between class I HLA loci in the cohort

Supp. Table 3: LD values between class I HLA loci in Uganda

Supp. Table 4: LD values between class I HLA loci in Botswana

Supp. Table 5: Class I HLA allele count and frequency among pediatric HIV populations in Uganda and Botswana (N=814)

Supp. Table 6: HLA allele frequencies in different populations

Population	Allele Frequencies									
	A*30:01	B*15:10	B*42:01	B*57:03	B*58:01	C*03:02	C*04:01	C*06:02	C*07:01	C*17:01
HIV Infected										
Overall CAfGEN	0.1069	0.0571	0.1102	0.0252	0.0518	0.0172	0.1592	0.1248	0.1136	0.1242
Ugandan CAfGEN	0.0838	0.0473	0.0799	0.0296	0.0385	0.0223	0.1399	0.1324	0.1146	0.0967
Botswana CAfGEN	0.1256	0.0651	0.1349	0.0217	0.0627	0.0131	0.1746	0.1188	0.1128	0.1461
Healthy populations										
Ugandans	0.0307	0.0250	0.0124	0.0120	0.0400	0.0180	0.1411	0.1166	0.1595	0.0276
Kenyan Nandi	0.0477	0.0170	0.0688	0.0290	0.1000	0.0460	0.1146	0.2167	0.1521	0.1021
Kenyan Luo	0.0642	0.0360	0.0774	0.0090	0.0700	0.0320	0.1321	0.1868	0.1170	0.0868
Zimbabwe Shona	0.0910	0.0840	0.0690	0.0440	0.0440	0.0110	0.1480	0.1280	0.1060	0.0820
South African Blacks	0.1021	0.0775	0.0739	0.0247	0.0704	0.0106	0.1409	0.1690	0.0563	0.1056
African Americans	0.0691	0.0303	0.0546	0.0338	0.0351	0.0137	0.1846	0.0886	0.1240	0.0736

Data was extracted from Cao et al., Tshabalala et al., Kijak et al., and IMGT/HLA database

Supp. Table 7: Differences in HLA alleles found in pediatric HIV populations in Uganda and Botswana

HLA Allele	Allele frequency			OR (95% CI)	P value†
	Uganda (n = 370)	Botswana (n = 451)	Total		
HLA A					
A*02:01	0.108	0.059	0.081	1.94 (1.34–2.8)	0.005
A*74:01	0.078	0.038	0.056	2.16 (1.38–3.39)	0.005
A*34:02	0.025	0.061	0.045	0.38 (0.2–0.67)	0.005
A*02:05	0.022	0.054	0.040	0.38 (0.21–0.70)	0.008
A*23:01	0.071	0.117	0.097	0.57 (0.40–0.81)	0.010
A*01:01	0.041	0.016	0.027	2.52 (1.31–4.8)	0.020
A*32:01	0.010	0.031	0.021	0.32 (0.14–0.75)	0.024
A*30:01	0.084	0.126	0.107	0.6 (0.45–0.89)	0.024
A*68:01	0.012	0.031	0.022	0.37 (0.16–0.8)	0.039
HLA B					
B*44:03	0.024	0.092	0.061	0.24 (0.13–0.41)	1.63×10^{-4}
B*49:01	0.031	0.005	0.017	6.62 (2.26–19.38)	0.001
B*42:01	0.080	0.135	0.110	0.55 (0.39–0.78)	0.006
B*14:01	0.004	0.027	0.017	0.16 (0.04–0.54)	0.008
HLA C					
C*03:04	0.057	0.021	0.037	2.74 (1.55–4.85)	0.007
C*17:01	0.097	0.146	0.124	0.6 (0.45–0.86)	0.036
C*07:04	0.042	0.018	0.028	2.39 (1.26–4.52)	0.04

† FDR adjusted P value

n, number of participants

Supp. Table 8: Global pairwise linkage disequilibrium

HLA loci	Wn	D'	ln(L_1)	ln(L_0)	p-value
Uganda					
A:B	0.45271	0.65307	-3053.95	-3716.03	-
A:C	0.30553	0.59446	-2918.19	-3379.71	-
B:C	0.54272	0.87593	-2406.65	-3530.9	-
Botswana					
A:B	0.42587	0.64506	-3612.17	-4401.32	-
A:C	0.38323	0.60042	-3584.52	-4231.64	-
B:C	0.59406	0.86833	-2879.95	-4158.78	-

Wn Cramer's V Statistic, D' Hedrick's statistic, In(L_1) log-likelihood of obtaining the observed data given the inferred haplotype

* $p < 0.05$ is indicative of overall significant LD

Supp. Table 9: The distribution HLA haplotypes in Uganda and Botswana

Haplotype	Overall	Botswana	Uganda	Haplotype	Overall	Botswana	Uganda	Haplotype	Overall	Botswana	Uganda	Haplotype	Overall	Botswa	Uganda
A*30:01~B*42:01	0.053	0.072	0.027	A*30:01~C*17:01	0.057	0.076	0.035	B*4201~C*1701	0.096	0.115	0.073	A*30:01~B*42:01~C*17:01	0.047	0.062	0.027
A*34:02~B*44:03	0.025	0.041	0.006	A*34:02~C*04:01	0.028	0.042	0.012	B*5802~C*0602	0.068	0.063	0.074	A*30:02~B*45:01~C*16:01	0.018	0.023	0.009
A*74:01~B*15:03	0.020	0.017	0.006	A*02:01~C*16:01	0.022	0.018	0.028	B*1503~C*0210	0.059	0.058	0.061	A*29:02~B*42:01~C*17:01	0.017	0.011	0.024
A*30:02~B*45:01	0.020	0.026	0.006	A*23:01~C*06:02	0.022	0.036	0.006	B*5301~C*0401	0.052	0.043	0.062	A*02:05~B*58:01~C*07:01	0.016	0.027	0.001
A*29:02~B*42:01	0.019	0.014	0.006	A*02:05~C*07:01	0.018	0.028	0.006	B*4501~C*1601	0.040	0.046	0.030	A*02:01~B*45:01~C*16:01	0.012	0.013	0.010
A*02:05~B*58:01	0.016	0.026	0.006	A*30:02~C*16:01	0.018	0.026	0.008	B*4403~C*0401	0.037	0.061	0.008	A*01:01~B*81:01~C*18:01	0.012	0.012	0.011
A*68:02~B*07:02	0.016	0.011	0.006	A*74:01~C*02:10	0.018	0.017	0.020	B*5801~C*0701	0.027	0.040	0.011	A*30:02~B*18:01~C*07:04	0.011	0.009	0.017
A*66:01~B*58:02	0.015	0.008	0.006	A*29:02~C*17:01	0.017	0.008	0.027	B*1510~C*0304	0.026	0.019	0.036	A*02:01~B*53:01~C*04:01	0.011	0.011	0.014
A*36:01~B*53:01	0.014	0.012	0.006	A*02:01~C*04:01	0.017	0.008	0.028	B*18:01~C*07:04	0.024	0.014	0.037	A*02:01~B*15:03~C*02:10	0.010	0.004	0.017
A*02:01~B*45:01	0.014	0.014	0.006	A*36:01~C*04:01	0.015	0.014	0.014	B*07:02~C*07:02	0.023	0.020	0.026	A*30:01~B*42:02~C*17:01	0.009	0.015	
A*68:02~B*15:10	0.013	0.011	0.006	A*29:02~C*06:02	0.014	0.015	0.012	B*45:01~C*06:02	0.021	0.019	0.023	A*23:01~B*58:02~C*06:02	0.009	0.016	
A*30:02~B*18:01	0.013	0.008	0.006	A*23:01~C*04:01	0.014	0.017	0.009	B*08:01~C*07:01	0.020	0.033	0.004	A*23:01~B*15:03~C*02:10	0.009	0.011	0.007
A*02:01~B*53:01	0.012	0.012	0.006	A*30:02~C*07:01	0.013	0.014	0.009	B*81:01~C*18:01	0.020	0.017	0.023	A*23:01~B*45:01~C*06:02	0.008	0.010	0.005

A*02:01~B*15:03	0.012	0.007	0.006	A*30:02~C*07:04	0.012	0.010	0.016	B*14:02~C*08:02	0.016	0.009	0.024	A*30:02~B*08:01~C*07:01	0.007	0.010	
A*01:01~B*81:01	0.012	0.015	0.006	A*74:01~C*06:02	0.012	0.001	0.023	B*49:01~C*07:01	0.015	0.004	0.029	A*29:02~B*45:01~C*06:02	0.007	0.007	0.006
A*68:02~B*42:01	0.010	0.013	0.006	A*01:01~C*18:01	0.012	0.012	0.011	B*35:01~C*04:01	0.014	0.013	0.017	A*02:02~B*58:02~C*06:02	0.006	0.001	0.013
A*23:01~B*58:02	0.010	0.018		A*23:01~C*02:10	0.012	0.014	0.008	B*15:10~C*04:01	0.014	0.024		A*23:01~B*08:01~C*07:01	0.006	0.008	0.003
A*23:01~B*45:01	0.009	0.012	0.007	A*68:02~C*07:02	0.011	0.008	0.014	B*58:01~C*03:02	0.013	0.013	0.014	A*23:01~B*44:03~C*04:01	0.006	0.010	
A*23:01~B*44:03	0.009	0.014	0.002	A*68:01~C*06:02	0.011	0.016	0.005	B*44:03~C*07:01	0.013	0.020	0.007	A*02:01~B*51:01~C*16:01	0.006	0.002	0.011
A*29:02~B*45:01	0.009	0.008	0.012	A*68:02~C*08:02	0.011	0.016	0.005	B*81:01~C*04:01	0.012	0.013	0.011	A*03:01~B*49:01~C*07:01	0.006	0.002	0.010
A*23:01~B*15:03	0.009	0.008	0.007	A*03:01~C*06:02	0.011	0.009	0.012	B*57:03~C*07:01	0.012	0.009	0.014	A*02:01~B*58:02~C*06:02	0.005	0.004	0.008
A*30:01~B*42:02	0.009	0.014		A*02:01~C*02:10	0.011	0.006	0.017	B*42:02~C*17:01	0.011	0.014	0.006	A*30:02~B*57:02~C*18:01	0.005		0.013
A*23:01~B*15:10	0.009	0.011	0.005	A*74:01~C*04:01	0.011	0.010	0.013	B*41:01~C*17:01	0.010	0.011	0.007	A*30:01~B*42:01~C*17:11	0.005	0.007	0.003
A*30:02~B*08:01	0.008	0.014		A*23:01~C*07:01	0.010	0.009	0.012	B*14:01~C*08:04	0.009	0.017					
A*74:01~B*58:02	0.008	0.003	0.014	A*68:02~C*17:01	0.010	0.011	0.009	B*58:01~C*06:02	0.009	0.011	0.006				
A*33:01~B*42:01	0.008	0.014	0.001	A*30:01~C*07:01	0.010	0.005	0.014	B*42:01~C*17:11	0.009	0.011	0.006				
A*30:02~B*57:03	0.008	0.006	0.010	A*66:01~C*06:02	0.010	0.005	0.017	B*53:01~C*06:02	0.008	0.007	0.011				
A*68:02~B*14:02	0.008	0.008	0.007	A*68:02~C*03:04	0.009	0.006	0.013	B*57:02~C*18:01	0.008	0.001	0.017				
A*23:01~B*58:01	0.008	0.010	0.003	A*33:01~C*17:01	0.009	0.013	0.014	B*51:01~C*16:01	0.008	0.005	0.011				

A*43:01~B*15:10	0.007	0.015		A*03:01~C*70:1	0.009	0.004	0.020	B*13:02~C*06:02	0.008	0.009	0.006
A*23:01~B*53:01	0.007	0.006	0.006	A*30:02~C*18:01	0.009	0.005	0.018	B*18:01~C*02:02	0.007	0.011	0.002
A*30:01~B*57:03	0.007	0.003	0.011	A*23:01~C*07:02	0.009	0.014	0.003	B*58:02~C*06:89	0.007	0.010	0.003
A*02:01~B*58:02	0.007	0.004	0.008	A*68:02~C*04:01	0.008	0.004	0.017	B*81:01~C*08:04	0.007	0.002	0.012
A*23:01~B*08:01	0.007	0.007	0.006	A*43:01~C*04:01	0.008	0.013		B*15:03~C*02:27	0.006	0.001	0.007
A*02:02~B*58:02	0.007	0.001	0.013	A*02:02~C*07:01	0.007	0.009	0.004	B*57:03~C*18:01	0.006	0.007	0.005
A*68:01~B*58:02	0.007	0.011		A*32:01~C*08:04	0.007	0.008	0.005	B*15:10~C*08:04	0.006	0.010	
A*68:02~B*53:01	0.007	0.001	0.014	A*33:03~C*04:01	0.007	0.009	0.004	B*15:10~C*16:01	0.006	0.005	0.007
A*02:01~B*51:01	0.006	0.001	0.012	A*03:01~C*17:01	0.007	0.008	0.002	B*39:10~C*12:03	0.005	0.002	0.009
A*03:01~B*49:01	0.006	0.002	0.011	A*29:02~C*02:10	0.006	0.007	0.003	B*47:03~C*07:01	0.005		0.012
A*29:02~B*44:03	0.006	0.010		A*30:01~C*17:11	0.006	0.007	0.004				
A*03:01~B*15:10	0.006	0.005	0.006	A*80:01~C*02:02	0.006	0.009					
A*30:02~B*57:02	0.006		0.014	A*68:02~C*07:01	0.006	0.006	0.005				
A*23:01~B*70:02	0.006	0.007	0.003	A*29:02~C*04:01	0.006	0.006	0.006				
A*74:01~B*35:01	0.005	0.003	0.009	A*02:02~C*06:02	0.006	0.000	0.013				
A*03:01~B*58:02	0.005	0.008	0.004	A*30:01~C*03:04	0.005	0.003	0.008				

A*74:01~B*58:01	0.005	0.003	0.008	A*34:02~C*02:10	0.005	0.009	0.002
A*30:01~B*15:03	0.005	0.011		A*30:02~C*17:01	0.005	0.009	0.003
A*33:03~B*53:01	0.005	0.007		A*02:01~C*06:02	0.005	0.006	0.006
A*34:02~B*15:03	0.005	0.007	0.002	A*02:05~C*08:04	0.005	0.009	
A*30:02~B*14:02	0.005		0.010	A*02:01~C*07:01	0.005	0.008	
				A*02:05~C*04:01	0.005	0.006	0.001
				A*30:01~C*02:10	0.005	0.009	0.002

Supp. Table 10: LD associations between class I HLA loci

Haplotypes		Frequency	D′	R2	p-value
A0101	C1801	0.0231	0.42	0.12	1.49E-12
A0101	B8101	0.0195	0.37	0.11	1.74E-09
A0201	C1601	0.0402	0.25	0.05	6.92E-09
A0205	B5801	0.028	0.39	0.11	5.93E-12
A0205	C0701	0.0353	0.4	0.05	4.08E-08
A0214	B1803	0.0037	0.6	0.36	2.16E-06
A2902	B4201	0.0426	0.25	0.03	2.24E-06
A3001	C1701	0.1011	0.48	0.19	2.50E-24
A3001	B4201	0.095	0.44	0.19	3.09E-24
A3001	B4202	0.0158	0.74	0.05	2.73E-07
A3002	C0704	0.0244	0.37	0.04	9.85E-07
A3002	C1601	0.0365	0.22	0.03	3.15E-06
A3201	C0804	0.0122	0.28	0.07	1.66E-07
A3402	B4403	0.0426	0.55	0.21	2.65E-20
A3402	C0401	0.0499	0.53	0.07	7.56E-10
A3601	B5301	0.028	0.58	0.12	4.73E-14
A3601	C0401	0.0305	0.56	0.04	1.57E-07
A6601	B5802	0.0268	0.51	0.09	5.94E-10
A6802	B0702	0.0317	0.42	0.08	1.45E-10
A6802	C0702	0.0231	0.31	0.04	1.45E-06
A7401	C0210	0.0317	0.26	0.05	5.19E-07
A7401	B1503	0.0341	0.33	0.07	6.00E-07
A8001	C0202	0.0085	0.87	0.41	8.31E-12
A8001	B1801	0.0085	0.77	0.09	2.25E-07
B0702	C0702	0.0402	0.69	0.44	3.37E-33
B0702	C0719	0.0061	1	0.09	1.13E-06
B0705	C0702	0.0085	0.87	0.13	9.90E-09
B0801	C0701	0.0378	0.55	0.08	1.23E-11
B1401	C0804	0.0171	0.57	0.22	8.86E-15
B1402	C0802	0.0292	0.92	0.48	4.74E-31
B1503	C0210	0.0938	0.78	0.53	2.22E-58
B1510	C0304	0.0463	0.71	0.33	2.05E-30
B1516	C1402	0.0085	1	0.58	1.63E-14
B1516	C1412	0.0037	1	0.25	2.39E-06
B1517	C0501	0.0073	0.75	0.41	3.07E-11
B1801	C0704	0.0438	0.85	0.54	3.73E-40
B1801	C0202	0.0122	0.65	0.11	9.87E-09

B3501	C0401	0.0256	0.76	0.06	9.60E-09
B3910	C1203	0.0085	0.87	0.43	1.05E-12
B4201	C1701	0.1608	0.86	0.65	1.50E-88
B4201	C0602	0.0134			3.35E-07
B4201	C1711	0.0158	0.67	0.04	9.67E-07
B4202	C1701	0.0183	0.93	0.07	4.32E-09
B4403	C0401	0.0646	0.49	0.09	1.18E-12
B4415	C0407	0.0085	1	0.78	7.40E-16
B4501	C1601	0.0658	0.63	0.34	5.34E-35
B4901	C0701	0.028	0.91	0.11	3.39E-15
B5101	C1601	0.0134	0.84	0.1	5.97E-10
B5301	C0401	0.0938	0.73	0.21	2.00E-31
B5702	C1801	0.0134	1	0.18	1.77E-12
B5801	C0302	0.0244	0.75	0.18	5.65E-18
B5801	C0701	0.0536	0.46	0.09	1.77E-16
B5802	C0602	0.1096	0.79	0.38	2.62E-46
B5802	C0689	0.0122	0.69	0.06	1.60E-06
B8101	C1801	0.0341	0.48	0.22	3.53E-19
B8202	C0302	0.0061	1	0.2	1.26E-07

Number of distinct two-way tests performed = N2 = 15681.

p-value for two-way comparisons = 0.05/N2 = 3.189e-06 for a 95% confidence level.

Displaying results with p-values less than 3.189e-06

Supp. Table 11: HLA alleles associated with LTNP by Country of Origin

HLA Allele	Allele frequency in cohort			P value	OR (95% CI)	P value†
	LTNPs	RPs	Total			
Protective						
Uganda						
A*34:02	0.036	0.010	0.025	0.037	3.93 (1.09–14.17)	0.1665
A*74:01	0.099	0.048	0.077	0.0351	2.01 (1.05–3.85)	0.1665
B*57:03	0.044	0.007	0.028	0.0171	6.14 (1.38–27.29)	0.1812
B*58:01	0.042	0.035	0.039	0.3097	1.53 (0.67–3.45)	0.6814
B*81:01	0.060	0.028	0.046	0.0392	2.47 (1.04–5.82)	0.1812
C*03:02	0.032	0.007	0.021	0.0162	6.52 (1.41–30.10)	0.065
C*08:02	0.048	0.030	0.040	0.3719	1.46 (0.64–3.35)	0.7438
C*18:01	0.075	0.020	0.051	0.0099	3.37 (1.34–8.48)	0.0539
Botswana						
A*30:04	0.014	0.010	0.012	0.6018	1.40 (0.40–4.92)	0.935
A*34:02	0.063	0.060	0.061	0.8911	1.04 (0.60–1.81)	0.935
A*36:01	0.026	0.022	0.024	0.734	1.17 (0.48–2.89)	0.935
A*74:01	0.040	0.036	0.038	0.7811	1.10 (0.56–2.18)	0.935
B*57:03	0.032	0.015	0.022	0.1081	2.15 (0.85–5.45)	0.4107
B*58:01	0.093	0.042	0.063	0.0044	2.18 (1.28–3.72)	0.0417
C*03:02	0.023	0.006	0.013	0.0472	3.89 (1.02–14.89)	0.378
C*08:02	0.035	0.012	0.021	0.0311	3.01 (1.10–8.19)	0.378
Susceptible						
Uganda						
A*26:12	0.003	0.024	0.012	0.0416	0.11 (0.01–0.92)	0.1665
A*30:04	0.003	0.020	0.010	0.0328	0.09 (0.01–0.83)	0.1665
A*36:01	0.013	0.041	0.025	0.029	0.30 (0.10–0.88)	0.1665
A*43:01						
B*18:01	0.034	0.077	0.052	0.0377	0.49 (0.25–0.96)	0.1812
B*15:10	0.039	0.059	0.048	0.2002	0.63 (0.31–1.28)	0.533
B*41:01	0.008	0.035	0.019	0.0112	0.18 (0.05–0.68)	0.1812
B*42:01	0.062	0.105	0.080	0.0412	0.54 (0.30–0.98)	0.1812
C*07:04	0.027	0.061	0.042	0.0101	0.33 (0.15–0.77)	0.0539
C*16:01	0.030	0.088	0.055	0.0053	0.35 (0.17–0.73)	0.0539
Botswana						
A*26:12						
A*43:01	0.020	0.046	0.035	0.0587	0.45 (0.19–1.03)	0.8227
B*18:01	0.029	0.035	0.033	0.5775	0.80 (0.35–1.78)	0.7726

B*15:10	0.035	0.087	0.065	0.0033	0.37 (0.19–0.72)	0.0417
B*41:01	0.009	0.021	0.016	0.2565	0.50 (0.16–1.65)	0.6093
B*42:01	0.130	0.139	0.135	0.7873	0.95 (0.63–1.42)	0.8567
B*81:01	0.032	0.035	0.034	0.8116	0.91 (0.41–1.99)	0.8567
C*07:04	0.014	0.020	0.018	0.515	0.69 (0.23–2.07)	0.8107
C*16:01	0.063	0.071	0.068	0.6964	0.90 (0.53–1.52)	0.8567
C*18:01	0.026	0.039	0.033	0.315	0.66 (0.29–1.49)	0.8107

NOTE: Results of all observed HLA alleles are provided in Supplementary Table Sxxx

Bold indicates alleles that satisfy FDR correction $p < 0.05$

† FDR adjusted P value

Number of participants in Uganda was 370 (208 LTNPs vs 162 RPs) while Botswana was 451 (190 LTNPs vs 261 RPs)

Supp. Table 12: Zygosity tests for HLA alleles associated with LTNP and RP

Allele	Homozygosity		Heterozygosity		Zygosity	
	OR	p value	OR	p value	OR	p value
Protective						
A*74:01	1.46	1	0.55	0.0165	2.65	0.3464
B*57:03	1.86	1	0.33	0.0043	5.51	0.2272
B*58:01	0.77	1	0.54	0.0258	1.42	0.7141
C*03:02	0.25	1	0.22	5.7 X 10-4	1.14	1
C*08:02	0.14	0.3182	0.38	0.0042	0.38	1
Susceptible						
A*43:01	3.9	0.3889	3.82	0.0024	1.02	1
B*15:10	0.2	0.297	1.93	0.0082	0.1	0.0594
B*41:01	4.11	0.3804	4.07	0.004	1.01	1
C*16:01	0.64	0.7053	1.63	0.0549	0.39	0.2869

Supp. Table 13: Haplotype frequency and association with LTNP

HLA Haplotype*	Haplotype Frequency			Haplotype effect§	
	LTNP	RP	p value†	OR (95% CI)	p value†
A*01:01~B*81:01	0.012	0.012	0.976	0.97 (0.31–3.04)	0.845
A*02:01~B*15:03	0.008	0.018	0.145	0.51 (0.15–1.73)	0.482
A*02:01~B*45:01	0.012	0.014	0.957	1.04 (0.35–3.11)	0.18
A*02:01~B*53:01	0.013	0.010	0.585	1.44 (0.42–4.90)	0.861
A*02:05~B*5801	0.021	0.013	0.298	1.52 (0.53–4.29)	0.358
A*29:02~B*42:01	0.013	0.031	0.032	0.43 (0.15–1.24)	0.02
A*30:01~B*42:01	0.051	0.051	0.755	1	
A*30:02~B*18:01	0.006	0.020	0.037	0.35 (0.10–1.22)	0.028
A*30:02~B*45:01	0.016	0.023	0.392	0.75 (0.28–1.96)	0.394
A*34:02~B*44:03	0.030	0.021	0.355	1.42 (0.62–3.24)	0.39
A*36:01~B*53:01	0.008	0.020	0.064	0.41 (0.13–1.27)	0.042
A*66:01~B*58:02	0.018	0.010	0.305	1.63 (0.59–4.54)	0.823
A*68:02~B*07:02	0.010	0.010	0.831	0.89 (0.24–3.26)	0.435
A*68:02~B*15:10	0.013	0.012	0.847	0.88 (0.30–2.58)	0.383
A*68:02~B*42:01	0.016	0.016	0.976	0.98 (0.34–2.77)	0.769
A*74:01~B*15:03	0.022	0.019	0.549	1.33 (0.55–3.18)	0.912
global p-val = 0.26515					
A*01:01~C*18:01	0.012	0.011	0.739	1.10 (0.36–3.35)	0.931
A*02:01~C*04:01	0.025	0.008	0.206	1.89 (0.53–6.63)	0.597
A*02:01~C*16:01	0.017	0.028	0.129	0.56 (0.22–1.41)	0.089
A*02:05~C*07:01	0.024	0.013	0.112	1.88 (0.70–5.03)	0.182
A*23:01~C*02:10	0.014	0.008	0.490	1.40 (0.46–4.23)	0.357
A*23:01~C*04:01	0.012	0.017	0.262	0.44 (0.12–1.61)	0.157
A*23:01~C*06:02	0.021	0.024	0.680	0.79 (0.33–1.92)	0.713
A*29:02~C*06:02	0.019	0.011	0.318	1.72 (0.54–5.43)	0.556
A*29:02~C*17:01	0.010	0.029	0.008	0.26 (0.08–0.79)	0.003
A*30:01~C*17:01	0.056	0.055	0.879	1	
A*30:02~C*07:01	0.008	0.010	0.342	0.66 (0.19–2.28)	0.708
A*30:02~C*07:04	0.008	0.017	0.119	0.43 (0.13–1.45)	0.059
A*30:02~C*16:01	0.018	0.020	0.775	0.97 (0.36–2.60)	0.92
A*34:02~C*040:1	0.029	0.026	0.606	1.20 (0.52–2.73)	0.609
A*36:01~C*04:01	0.012	0.018	0.397	0.63 (0.21–1.84)	0.202
A*68:01~C*06:02	0.011	0.011	0.946	0.89 (0.27–2.87)	0.85
A*68:02~C*07:02	0.012	0.013	0.897	1.05 (0.28–3.85)	0.635
A*68:02~C*08:02	0.014	0.006	0.154	2.72 (0.76–9.68)	0.106
A*74:01~C*02:10	0.021	0.015	0.405	1.35 (0.55–3.27)	0.835

A*74:01~C*06:02	0.018	0.006	0.026	3.70 (0.66–20.77)	0.372
global p-val = 0.15544					
B*07:02~C*07:02	0.030	0.016	0.089	2.14 (0.96–4.75)	0.154
B*08:01~C*07:01	0.026	0.016	0.229	2.05 (0.86–4.90)	0.068
B*14:02~C*08:02	0.020	0.013	0.265	1.92 (0.77–4.74)	0.509
B*15:03~C*02:10	0.056	0.062	0.595	1.15 (0.66–2.00)	0.923
B*15:10~C*03:04	0.023	0.028	0.532	0.95 (0.45–1.99)	0.472
B*15:10~C*04:01	0.006	0.021	0.012	0.35 (0.11–1.16)	0.173
B*18:01~C*07:04	0.019	0.030	0.157	0.74 (0.31–1.72)	0.11
B*35:01~C*04:01	0.020	0.009	0.085	3.12 (1.01–9.55)	0.076
B*42:01~C*17:01	0.084	0.108	0.112	1	
B*42:02~C*17:01	0.006	0.015	0.061	0.41 (0.11–1.50)	0.163
B*44:03~C*04:01	0.036	0.038	0.878	1.24 (0.64–2.39)	0.358
B*44:03~C*07:01	0.010	0.018	0.337	0.80 (0.27–2.37)	0.732
B*45:01~C*06:02	0.024	0.018	0.465	1.50 (0.64–3.50)	0.571
B*45:01~C*16:01	0.032	0.046	0.154	0.84 (0.44–1.60)	0.552
B*49:01~C*07:01	0.014	0.016	0.724	0.90 (0.33–2.43)	0.401
B*53:01~C*04:01	0.049	0.055	0.580	1.10 (0.60–2.01)	0.705
B*57:03~C*07:01	0.020	0.003	0.007	5.40 (1.40–20.79)	0.025
B*58:01~C*03:02	0.022	0.005	0.004	4.88 (1.50–15.86)	0.011
B*58:01~C*07:01	0.033	0.022	0.174	1.65 (0.77–3.52)	0.106
B*58:02~C*06:02	0.069	0.067	0.809	1.28 (0.75–2.18)	0.633
B*81:01~C*04:01	0.013	0.011	0.658	1.47 (0.48–4.47)	0.633
B*81:01~C*18:01	0.020	0.019	0.759	1.44 (0.64–3.24)	0.573
global p-val = 0.006047					
A*01:01~B*81:01~C*18:01	0.012	0.011	0.898	0.99 (0.30–3.24)	0.819
A*02:01~B*45:01~C*16:01	0.009	0.014	0.359	0.68 (0.18–2.47)	0.346
A*02:01~B*53:01~C*04:01	0.015	0.009	0.227	2.04 (0.46–8.92)	0.553
A*02:05~B*58:01~C*07:01	0.018	0.013	0.313	1.46 (0.49–4.37)	0.378
A*29:02~B*42:01~C*17:01	0.009	0.027	0.016	0.36 (0.11–1.16)	0.022
A*30:01~B*42:01~C*17:01	0.047	0.045	0.967	1	
A*30:02~B*18:01~C*07:04	0.008	0.016	0.185	0.47 (0.13–1.69)	0.1
A*30:02~B*45:01~C*16:01	0.017	0.019	0.629	0.82 (0.27–2.47)	0.729
A*34:02~B*44:03~C*04:01	0.025	0.019	0.536	1.20 (0.47–3.05)	0.605
A*36:01~B*53~01~C*04:01	0.008	0.018	0.106	0.43 (0.12–1.48)	0.086
A*66:01~B*58:02~C*06:02	0.012	0.009	0.619	1.23 (0.36–4.21)	0.855
A*68:02~B*07:02~C*07:02	0.010	0.011	0.807	0.87 (0.26–2.93)	0.713
A*68:02~B*15:10~C*03:04	0.012	0.011	0.710	1.13 (0.30–4.15)	0.438
A*74:01~B*15:03~C*02:10	0.018	0.013	0.381	1.44 (0.54–3.87)	0.81
global p-val = 0.38837					

*Haplotypes with frequency >1% in the pooled population

†The p value is based on the individual haplotype association with LTNP compared to RP
§Odds ratios are adjusted for country and gender

Supp. Table 14: List of peptides classified as strong or weak binders

Peptide	Protein	Position	Subtype	Score_EL	%Rank_EL	Score_BA	%Rank_BA	Aff(nM)	MotifScan	Binding Level
RAVGIGAVF	Env	493/4	A1 & C	0.882943	0.029	0.830774	0.023	6.24	Yes	SB
RAVGIGAVF	Env	494	C	0.882943	0.029	0.830774	0.023	6.24		SB
ISNYTHIIY	Env	617	A1	0.761367	0.089	0.757922	0.072	13.73		SB
LAWDDLRSL	Env	737	A1 & C	0.742696	0.098	0.694673	0.166	27.21		SB
RAIEAQQHL	Env	539	A1	0.719846	0.124	0.621234	0.337	60.23		SB
FSYHRLRDF	Env	748	A1 & C	0.715692	0.127	0.785577	0.048	10.18		SB
VSFEPIPIHY	Env	203	A1	0.572694	0.24	0.513237	0.791	193.77		SB
HSFNCGGEF	Env	367	A1	0.531917	0.286	0.765019	0.064	12.71		SB
FCASDAKAY	Env	52	A1 & C	0.47309	0.366	0.582521	0.478	91.56		SB
NAKTIIVQL	Env	275	A1	0.451979	0.395	0.355564	2.382	1067.05		SB
SFEPIPIHY	Env	204	A1	0.436763	0.418	0.251438	4.882	3292.08		SB
SAAENLWVTVY	Env	28	A1	0.398212	0.478	0.507083	0.826	207.11	Yes	SB
RAIEAQQHM	Env	529	C	0.83873	0.05	0.72826	0.116	18.92		SB
VSFDPIPIHY	Env	196	C	0.67235	0.159	0.526122	0.725	168.55		SB
TSKLFNSTY	Env	373	C	0.660196	0.172	0.607834	0.378	69.63		SB
SFDPIPIHY	Env	197	C	0.653036	0.178	0.296926	3.526	2012.44		SB
YSPLSFQTL	Env	684	C	0.523339	0.297	0.592497	0.435	82.19		SB
TAVPWNSSW	Env	578	C	0.498843	0.33	0.553949	0.6	124.73		SB
ISNYTDTIY	Env	607	C	0.49404	0.337	0.680644	0.192	31.67		SB
HSFNCRGEF	Env	360	C	0.447834	0.4	0.722646	0.124	20.1		SB

NATNATNTM	Env	132	C	0.434379	0.422	0.622937	0.332	59.13		SB
YTDTIYRLL	Env	610	C	0.428991	0.43	0.517677	0.768	184.68		SB
SIRIGPGQTF	Env	294	C	0.395192	0.483	0.43147	1.402	469.36		SB
VGNLWVTVY	Env	30	C	0.386815	0.496	0.521671	0.748	176.87		SB
RALGPGATL	Gag	335	A1 & C	0.769621	0.085	0.698606	0.157	26.08		SB
FSPEVIPMF	Gag	164	A1 & C	0.601813	0.216	0.573088	0.519	101.4		SB
YVDRFFKTL	Gag	296	A1 & C	0.512706	0.311	0.532807	0.692	156.79		SB
NSSKVSQNY	Gag	124	A1	0.463877	0.378	0.471721	1.06	303.65		SB
FALNPSLL	Gag	44	A1	0.414924	0.452	0.532447	0.694	157.4		SB
IALEMHPEF	Nef	195	A	0.911434	0.016	0.779453	0.051	10.87		SB
YTPGPGTRF	Nef	128	A	0.731763	0.109	0.550209	0.616	129.88		SB
MARELHPEY	Nef	195	C	0.927552	0.013	0.802619	0.041	8.46		SB
YTPGPGVRY	Nef	128	C	0.799714	0.067	0.562132	0.565	114.16		SB
AAFDLSFFL	Nef	84	C	0.447516	0.401	0.601996	0.396	74.17		SB
QVPLRPMTY	Nef	74	C	0.422209	0.441	0.346374	2.524	1178.61		SB
KAAFDLSFF	Nef	83	C	0.401476	0.473	0.634637	0.297	52.1		SB
FSVPLDESF	Pol	271	A1	0.801681	0.066	0.774497	0.055	11.47		SB
VAVHVASGY	Pol	790	A1 & C	0.734988	0.105	0.733173	0.109	17.94		SB
FSFPQITL	Pol	54	A1	0.709584	0.131	0.611932	0.366	66.61		SB
FSFPQITLW	Pol	54	A1	0.611108	0.209	0.624421	0.328	58.19		SB
KAQEEHERY	Pol	722	A1	0.567706	0.244	0.422455	1.503	517.45		SB
TVLDVGDAY	Pol	262	A1	0.508985	0.316	0.536598	0.676	150.49		SB
WANIQQEF	Pol	847	A1	0.456545	0.388	0.488867	0.943	252.23		SB

KALTDIVTL	Pol	442	A1	0.434042	0.422	0.48895	0.942	252		SB
IQQEFGIPY	Pol	850	A1 & C	0.422751	0.44	0.561249	0.569	115.26		SB
IVTDSQYAL	Pol	650	A1 & C	0.422403	0.44	0.525228	0.73	170.19		SB
VIWGKTPKF	Pol	536	A1 & C	0.417946	0.447	0.396999	1.785	681.53		SB
YAGIKVKQL	Pol	426	A1	0.416351	0.45	0.433945	1.379	456.96		SB
FQQGEARKF	Pol	8	A1	0.399945	0.476	0.413004	1.597	573.16		SB
IIRDYGKQM	Pol	982	A1	0.385613	0.498	0.447044	1.256	396.57		SB
AQNPEIVIY	Pol	324	C	0.633581	0.192	0.405893	1.675	619		SB
KAQEEHEKY	Pol	718	C	0.629062	0.195	0.416035	1.567	554.67		SB
LAFPQGEAREF	Pol	6	C	0.503965	0.323	0.55705	0.587	120.62		SB
IAMESIVIW	Pol	526	C	0.453247	0.393	0.584738	0.469	89.39		SB
FSVPLDEGF	Pol	267	C	0.420922	0.443	0.643744	0.274	47.21		SB
IIKDYGKQM	Pol	978	C	0.411548	0.457	0.411849	1.609	580.37		SB
RAQNPEIVIY	Pol	323	C	0.395124	0.483	0.440535	1.319	425.51		SB
SAEPVPLQL	Rev	67	A1	0.826617	0.055	0.504657	0.841	212.62		SB
VSVESPVIL	Rev	109	A1	0.560202	0.25	0.582496	0.478	91.59		SB
QTKGLGISY	Tat	39	C	0.5222	0.298	0.410085	1.629	591.55		SB
VSSEVHIPL	Vif	51	A1 & C	0.527779	0.291	0.697677	0.159	26.34		SB
QVVSPRCEY	Vif	127	A1	0.517982	0.304	0.510408	0.806	199.79		SB
FPRPWLHSL	Vpr	34	C	0.39766	0.479	0.453593	1.193	369.45		SB
SIRIGPGQAF	Env	301	A1	0.382582	0.504	0.475587	1.035	291.21		WB
IAARTVEL	Env	759	A1	0.366525	0.54	0.401091	1.732	652.01		WB
KSIRIGPGQAF	Env	300	A1	0.355927	0.564	0.431358	1.404	469.93		WB

VAKQLRKYF	Env	339	A1	0.353734	0.569	0.399043	1.759	666.62	Yes	WB
VSFEPIPIH	Env	203	A1	0.307562	0.672	0.387403	1.905	756.09		WB
KVAKQLRKY	Env	338	A1	0.294791	0.701	0.309951	3.231	1747.91		WB
IAARTVELL	Env	759	A1	0.291271	0.711	0.512399	0.795	195.53		WB
LIAARTVEL	Env	758	A1	0.284241	0.733	0.528858	0.711	163.64		WB
YAPPIQGVI	Env	418	A1	0.270802	0.774	0.515124	0.781	189.85		WB
LALDKWANL	Env	643	A1	0.267282	0.785	0.491717	0.922	244.57		WB
KRAVGIGAVF	Env	492	A1 & C	0.250841	0.84	0.56029	0.573	116.46		WB
FLGAAGSTM	Env	504	A1 & C	0.248595	0.847	0.637507	0.29	50.51		WB
HSSLKGLRL	Env	769	A1	0.235975	0.891	0.385263	1.937	773.8		WB
TQACPKVSF	Env	197	A1 & C	0.230732	0.911	0.381144	1.998	809.07		WB
TVYYGVPVW	Env	36	A1 & C	0.230436	0.912	0.448681	1.24	389.61		WB
LTVQARQLL	Env	519	A1 & C	0.229157	0.918	0.483657	0.98	266.86		WB
CSFNMTTEL	Env	150	A1	0.227933	0.923	0.658451	0.238	40.27		WB
GTMKNTITL	Env	391	A1	0.223283	0.943	0.456227	1.173	359.06		WB
IRIGPGQAF	Env	302	A1	0.221586	0.95	0.394757	1.813	698.26		WB
RIGPGQAFY	Env	303	A1	0.22056	0.954	0.363576	2.262	978.45		WB
YFKNKTIIF	Env	346	A1	0.215331	0.976	0.371941	2.136	893.78		WB
RSIRLVSGF	Env	726	A1 & C	0.203926	1.032	0.527622	0.717	165.84		WB
LTVWGIKQL	Env	550	A1 & C	0.187713	1.125	0.408074	1.651	604.56		WB
RLVSGFLAL	Env	729	A1 & C	0.180556	1.172	0.509643	0.811	201.45		WB
YSPLSFQTH	Env	694	A1	0.179217	1.18	0.403213	1.705	637.21		WB
EAQQHLLKL	Env	542	A1	0.178821	1.183	0.262888	4.51	2908.49		WB

YCAPAGFAI	Env	212	A1	0.17839	1.186	0.513188	0.791	193.87		WB
FTNSSGGDL	Env	354	A1	0.177427	1.192	0.581559	0.482	92.52		WB
VINRVRQGY	Env	686	A1	0.173406	1.219	0.333906	2.752	1348.83		WB
GAASITLTV	Env	513	A1 & C	0.153775	1.361	0.442535	1.301	416.4		WB
WQRAGQAMY	Env	410	A1	0.141279	1.472	0.44034	1.321	426.41		WB
RQGYSPLSF	Env	691	A1 & C	0.13327	1.546	0.335467	2.723	1326.24		WB
IVQQQSNLL	Env	530	A1 & C	0.127759	1.598	0.308892	3.255	1768.05		WB
RIGPGQAF	Env	303	A1	0.127464	1.601	0.261377	4.559	2956.43		WB
RVIEIGQRI	Env	817	A1	0.126817	1.608	0.325536	2.905	1476.68		WB
SAAENLWVTVYY	Env	28	A1	0.123482	1.643	0.43549	1.365	449.38		WB
VTNNTTNTH	Env	135	A1	0.122096	1.658	0.32029	3	1562.92		WB
HLENVTEEF	Env	84	A1	0.119104	1.69	0.281899	3.926	2367.75		WB
RVRQGYSPL	Env	689	A1 & C	0.118692	1.694	0.498367	0.879	227.59		WB
KVSFEPIPIHY	Env	202	A1	0.116024	1.728	0.201961	7.138	5622.91		WB
IISLWDQSL	Env	107	A1 & C	0.110283	1.804	0.425397	1.471	501.24		WB
QGYSPLSF	Env	692	A1 & C	0.105649	1.866	0.22955	5.738	4171.78		WB
VAKQLRKY	Env	339	A1	0.105324	1.87	0.160003	10.071	8853.62	Yes	WB
HSFNCGGEFF	Env	367	A1	0.104954	1.875	0.508729	0.817	203.45		WB
KQKVYSLFY	Env	162	A1	0.104182	1.885	0.31117	3.204	1725.01		WB
QACPKVSF	Env	198	A1 & C	0.101887	1.916	0.172265	9.063	7753.57	Yes	WB
QMHTDIISL	Env	102	A1	0.099229	1.953	0.342537	2.594	1228.57		WB
CASDAKAY	Env	53	A1 & C	0.09624	1.995	0.308427	3.265	1776.97	Yes	WB
MIVGGLIGL	Env	669	A1 & C	0.09606	1.998	0.437634	1.346	439.08		WB

VSFDPIPIH	Env	196	C	0.34104	0.598	0.393241	1.832	709.81		WB
CSFNITTEL	Env	146	C	0.337204	0.607	0.663407	0.227	38.16		WB
KSIRIGPGQTF	Env	293	C	0.335109	0.611	0.377459	2.053	841.98		WB
IVNRVRQGY	Env	676	C	0.319602	0.646	0.423979	1.486	508.99		WB
NAKTIIVHL	Env	268	C	0.318164	0.649	0.323603	2.94	1507.89		WB
RIGPGQTFY	Env	296	C	0.293707	0.704	0.347757	2.499	1161.1		WB
IAARAVEL	Env	749	C	0.281786	0.74	0.358909	2.332	1029.12		WB
EAQQHMLQL	Env	532	C	0.267757	0.783	0.335862	2.716	1320.58		WB
LALDSWKNL	Env	633	C	0.263671	0.796	0.464013	1.113	330.06		WB
IRIGPGQTF	Env	295	C	0.253702	0.83	0.361972	2.286	995.58		WB
NSTNSTITL	Env	382	C	0.25187	0.836	0.475012	1.039	293.02		WB
IAARAVELL	Env	749	C	0.24898	0.846	0.509625	0.811	201.49		WB
QMHEDIISL	Env	102	C	0.226914	0.927	0.402093	1.719	644.98		WB
FDPIPIHY	Env	198	C	0.215602	0.975	0.167134	9.491	8196.2		WB
YAILKCNNKTF	Env	211	C	0.203654	1.033	0.536423	0.676	150.78		WB
LVQYWGLEL	Env	778	C	0.195315	1.08	0.552698	0.605	126.43		WB
LIAARAVEL	Env	748	C	0.192955	1.093	0.475596	1.035	291.18		WB
YCAPAGYAI	Env	205	C	0.190789	1.105	0.509632	0.811	201.48		WB
KVSFDPIPIHY	Env	195	C	0.174861	1.209	0.21699	6.32	4779.04		WB
KSNITGLLL	Env	420	C	0.148587	1.404	0.420552	1.522	528.21		WB
YAPPIAGNI	Env	409	C	0.145736	1.43	0.428847	1.432	482.87		WB
FFYCNTSKL	Env	368	C	0.13947	1.489	0.543822	0.644	139.18		WB
YTDTIYRL	Env	610	C	0.132907	1.55	0.242296	5.213	3634.36		WB

LRAIEAQQHM	Env	528	C	0.130522	1.572	0.447056	1.256	396.52		WB
RIGPGQTF	Env	296	C	0.127788	1.598	0.232341	5.613	4047.68		WB
KYLGSLVQY	Env	773	C	0.126143	1.615	0.208671	6.756	5229.15		WB
IVLENVTENF	Env	83	C	0.119627	1.684	0.371536	2.142	897.7		WB
VSKKLKEHF	Env	332	C	0.11672	1.719	0.21887	6.23	4682.81		WB
ILKCNNKTF	Env	213	C	0.107861	1.836	0.362187	2.283	993.26		WB
FPNKTIKF	Env	340	C	0.104282	1.884	0.185568	8.115	6714.16		WB
MWQEVGRAM	Env	400	C	0.103597	1.893	0.384677	1.945	778.73		WB
STNSTITL	Env	383	C	0.10245	1.909	0.23879	5.349	3774.88		WB
YLGSLVQY	Env	774	C	0.097504	1.978	0.237775	5.388	3816.56		WB
SQVQHTNIM	Gag	368	A1	0.335296	0.611	0.540339	0.659	144.52		WB
IVGGHQAAM	Gag	190	A1	0.332089	0.618	0.531842	0.696	158.44		WB
RMYSPVSIL	Gag	275	A1 & C	0.300269	0.688	0.568383	0.539	106.7		WB
HQSLSPRTL	Gag	144	A1	0.27481	0.762	0.343764	2.572	1212.36		WB
EQDPPLVSL	Gag	479	A1	0.265043	0.792	0.168239	9.399	8098.79		WB
EVIPMFSAL	Gag	167	A1	0.230496	0.912	0.448962	1.237	388.43		WB
KARVLAEAM	Gag	359	A1 & C	0.22823	0.922	0.534478	0.685	153.98		WB
SLYNTVATL	Gag	77	A1 & C	0.227893	0.923	0.477865	1.02	284.12		WB
SSKGRPGNF	Gag	438	A1	0.207597	1.012	0.332815	2.772	1364.84		WB
WASRELERF	Gag	36	A1 & C	0.196382	1.074	0.457215	1.165	355.25		WB
HLVWASREL	Gag	33	A1 & C	0.184639	1.145	0.46826	1.083	315.23		WB
FSPEVIPM	Gag	164	A1 & C	0.184534	1.146	0.511259	0.801	197.96		WB
GATPQDLNM	Gag	178	A1	0.178777	1.183	0.354918	2.392	1074.54		WB

LRALGPGATL	Gag	334	A1 & C	0.176845	1.196	0.468908	1.079	313.03		WB
KAFSPEVIPM	Gag	162	A1 & C	0.162729	1.291	0.570562	0.529	104.21		WB
LYNTVATLY	Gag	78	A1 & C	0.158278	1.325	0.341834	2.607	1237.95		WB
GTEELRSLY	Gag	71	A1 & C	0.157273	1.333	0.290496	3.697	2157.44		WB
KAFSPEVIPMF	Gag	162	A1 & C	0.138827	1.495	0.325694	2.902	1474.16		WB
IVQNAQGQM	Gag	134	A1	0.136572	1.516	0.386398	1.92	764.36		WB
SSKVSQNY	Gag	125	A1	0.130482	1.572	0.205323	6.943	5422.05		WB
ILRALGPGATL	Gag	333	A1 & C	0.127028	1.605	0.327163	2.875	1450.91		WB
HAGPIPPGQM	Gag	219	A1	0.119532	1.685	0.264835	4.449	2847.86		WB
KVIEEKAF	Gag	157	A1 & C	0.105035	1.874	0.190356	7.822	6375.19		WB
AFSPEVIPM	Gag	163	A1 & C	0.097858	1.973	0.318572	3.039	1592.24		WB
RAEQATQEV	Gag	305	A1	0.097632	1.976	0.33118	2.802	1389.2		WB
KAADGKVSQNY	Gag	119	C	0.381886	0.506	0.328432	2.852	1431.13		WB
FALNPGLL	Gag	44	C	0.327012	0.629	0.500559	0.866	222.26		WB
QANNTNIMM	Gag	366	C	0.316923	0.651	0.534963	0.683	153.18		WB
QAISPRTL	Gag	142	C	0.301281	0.686	0.286115	3.814	2262.16		WB
HQAISPRTL	Gag	141	C	0.299385	0.69	0.376814	2.063	847.88		WB
EVIPMFTAL	Gag	164	C	0.268549	0.781	0.434537	1.374	454.04		WB
AADGKVSQNY	Gag	120	C	0.246056	0.856	0.269077	4.315	2720.1		WB
TAPPAESF	Gag	452	C	0.210741	0.996	0.249181	4.956	3373.46		WB
TVGGHQAAM	Gag	187	C	0.208381	1.007	0.434397	1.375	454.73		WB
IIKQLQPAL	Gag	60	C	0.177103	1.194	0.335437	2.724	1326.67		WB
IAGTTSTL	Gag	233	C	0.147722	1.412	0.319726	3.013	1572.49		WB

SQANNTNIM	Gag	365	C	0.139661	1.487	0.463489	1.117	331.93		WB
IVQNLQGQM	Gag	131	C	0.119719	1.683	0.353613	2.411	1089.82		WB
ATPQDLNTM	Gag	176	C	0.116598	1.721	0.283877	3.873	2317.61		WB
MTSNPPIPV	Gag	247	C	0.115201	1.739	0.549135	0.621	131.4		WB
PTAPPAESF	Gag	451	C	0.113002	1.768	0.21742	6.3	4756.86		WB
WVYNTQGYF	Nef	114	A	0.357859	0.56	0.68489	0.184	30.25		WB
IALEMHPEFY	Nef	195	A	0.339334	0.602	0.512904	0.793	194.47		WB
QVPLRPMTF	Nef	74	A	0.301437	0.686	0.316784	3.079	1623.35		WB
YSQKRQEIL	Nef	103	A	0.241465	0.872	0.446334	1.263	399.63		WB
GAFDLSFFL	Nef	84	A	0.228536	0.92	0.491318	0.925	245.63		WB
HIALEMHPEF	Nef	194	A	0.16498	1.275	0.460682	1.139	342.17		WB
MTFKGAFDL	Nef	80	A	0.143811	1.448	0.586978	0.459	87.25		WB
WVYNTQGY	Nef	114	A	0.133409	1.545	0.385191	1.938	774.41		WB
FPVRPQVPL	Nef	69	A & C	0.120275	1.677	0.441135	1.314	422.76		WB
GAFDLSFF	Nef	84	A	0.117278	1.712	0.318159	3.048	1599.38		WB
AATQASCAW	Nef	49	A	0.109135	1.82	0.431006	1.408	471.72		WB
LARRHIAL	Nef	190	A	0.09746	1.978	0.326682	2.884	1458.48		WB
MARELHPEYY	Nef	195	C	0.361584	0.552	0.590551	0.444	83.94		WB
WVYHTQGYF	Nef	114	C	0.28466	0.732	0.642211	0.277	48		WB
AAFDLSFF	Nef	84	C	0.28226	0.739	0.421807	1.51	521.09		WB
HMARELHPEY	Nef	194	C	0.26171	0.802	0.467187	1.09	318.91		WB
YSKKRQEIL	Nef	103	C	0.209644	1	0.380848	2.003	811.66		WB
YKAAFDLSF	Nef	82	C	0.199563	1.056	0.571632	0.525	103.01		WB

KAAFDLSF	Nef	83	C	0.176032	1.201	0.398621	1.764	669.67		WB
WVYHTQGY	Nef	114	C	0.164097	1.281	0.372584	2.127	887.58		WB
VSSGIRKVL	Pol	707	A1 & C	0.36556	0.543	0.451144	1.215	379.37		WB
ESFRKYTAF	Pol	277	A1	0.3559	0.564	0.510061	0.809	200.54		WB
MAVFIHNF	Pol	893	A1 & C	0.312	0.662	0.497293	0.885	230.25		WB
KTAVQMAVF	Pol	888	A1 & C	0.286168	0.727	0.557407	0.585	120.15		WB
SKNPEIIIY	Pol	328	A1	0.282558	0.738	0.27689	4.07	2499.61		WB
LKDPVHGVY	Pol	465	A1	0.275662	0.759	0.371671	2.14	896.39		WB
KIQNFRVYY	Pol	934	A1 & C	0.26776	0.783	0.401027	1.733	652.46		WB
QSQGVVESM	Pol	861	A1 & C	0.252004	0.836	0.412605	1.601	575.64		WB
FVNTPPLVKL	Pol	571	A1 & C	0.239215	0.88	0.459389	1.149	346.99		WB
KQITKIQNF	Pol	930	A1	0.224701	0.937	0.303688	3.371	1870.46		WB
LTEEKIKAL	Pol	181	A1 & C	0.224105	0.939	0.326478	2.887	1461.7		WB
ILKDPVHGVY	Pol	464	A1	0.223912	0.94	0.352937	2.421	1097.82		WB
FQSSMTKIL	Pol	315	A1 & C	0.218146	0.964	0.504703	0.841	212.51		WB
TAVQMAVF	Pol	889	A1 & C	0.206477	1.018	0.431294	1.404	470.25		WB
KLNWASQIY	Pol	418	A1 & C	0.204547	1.028	0.449963	1.227	384.24		WB
LGIPHPAGL	Pol	247	A1 & C	0.200432	1.051	0.413018	1.597	573.07		WB
QGTGPTFSF	Pol	48	A1	0.192154	1.097	0.329447	2.833	1415.49		WB
VVMESIVIW	Pol	530	A1	0.19088	1.105	0.357055	2.36	1049.98		WB
TKIQNFRVY	Pol	933	A1	0.188567	1.12	0.383669	1.96	787.26		WB
ISKIGPENPY	Pol	202	A1	0.187844	1.124	0.489952	0.935	249.29		WB
KLIGKDKVY	Pol	679	A1	0.1755	1.205	0.29419	3.599	2072.91		WB

RAHLLSWGF	Pol	361	A1	0.174719	1.21	0.469707	1.074	310.34		WB
LAFQQGEARKF	Pol	6	A1	0.164966	1.275	0.428127	1.44	486.65		WB
WTVNDIQKL	Pol	407	A1 & C	0.157896	1.328	0.43906	1.333	432.35		WB
IATDIQTKEL	Pol	919	A1 & C	0.157511	1.331	0.356511	2.368	1056.17		WB
FVNTPPLVKLWY	Pol	571	A1 & C	0.155907	1.344	0.422651	1.501	516.35		WB
TQIGCTLNF	Pol	147	A1	0.152978	1.368	0.461041	1.136	340.84		WB
LVNQIIEKL	Pol	672	A1	0.149606	1.395	0.287707	3.771	2223.53		WB
YFSVPLDESF	Pol	270	A1	0.137998	1.502	0.467101	1.091	319.21		WB
ATWIPEWEF	Pol	563	A1 & C	0.13467	1.533	0.282179	3.918	2360.58		WB
ATDIQTKEL	Pol	920	A1 & C	0.133302	1.546	0.198066	7.368	5864.94		WB
ITKIQNFRVY	Pol	932	A1	0.121022	1.67	0.477253	1.024	286		WB
ETPGIRYQY	Pol	293	A1 & C	0.116612	1.721	0.178555	8.582	7243.45		WB
VVHTDNGSNF	Pol	827	A1	0.115483	1.736	0.324699	2.92	1490.11		WB
RSKNPEIIIY	Pol	327	A1	0.115474	1.736	0.305731	3.325	1829.57		WB
TAYFLLKL	Pol	812	A1	0.113739	1.759	0.277299	4.057	2488.57		WB
FSVPLDESFRKY	Pol	271	A1	0.112764	1.772	0.457089	1.166	355.73		WB
VYYDPSKDL	Pol	472	A1 & C	0.112341	1.777	0.245583	5.086	3507.38		WB
FPISPIETV	Pol	155	A1 & C	0.112292	1.778	0.328478	2.851	1430.41		WB
FFRENLAF	Pol	1	A1 & C	0.109798	1.811	0.36058	2.307	1010.68		WB
GGFIKVKQY	Pol	107	A1	0.109739	1.812	0.180556	8.441	7088.31		WB
HTDNGSNF	Pol	829	A1 & C	0.107207	1.845	0.230935	5.676	4109.73		WB
VTVLDVGDAY	Pol	261	A1 & C	0.106708	1.852	0.402489	1.714	642.22		WB
ILKDPVHGVYY	Pol	464	A1	0.106233	1.858	0.203807	7.03	5511.72		WB

SKIGPENPY	Pol	203	A1	0.104838	1.877	0.289602	3.721	2178.41		WB
KMIGGIGGF	Pol	101	A1 & C	0.104399	1.882	0.451857	1.208	376.45		WB
IQNFRVYY	Pol	935	CnA1	0.103473	1.895	0.230427	5.699	4132.38		WB
HQKEPPFLW	Pol	376	A1 & C	0.102662	1.906	0.158378	10.229	9010.66		WB
YALGIIQAQ	Pol	656	A1 & C	0.101253	1.925	0.408886	1.642	599.28		WB
TFSFPQITL	Pol	53	A1	0.097074	1.984	0.249974	4.93	3344.64		WB
FPQGEAREF	Pol	8	C	0.34726	0.584	0.406087	1.672	617.7		WB
IKIQNFRVY	Pol	929	C	0.21752	0.967	0.427914	1.442	487.77		WB
LVNQIIEQL	Pol	668	C	0.213079	0.986	0.343043	2.585	1221.86		WB
YTAFTIPSI	Pol	278	C	0.211829	0.991	0.620301	0.34	60.84		WB
TAYYILKL	Pol	808	C	0.205678	1.022	0.316606	3.083	1626.48		WB
KALTDIVPL	Pol	438	C	0.204854	1.027	0.546496	0.632	135.21		WB
WAGIQQEF	Pol	843	C	0.196719	1.072	0.356579	2.367	1055.4		WB
RANSPTSREL	Pol	22	C	0.185245	1.141	0.389929	1.873	735.71		WB
FSVPLDEGFRKY	Pol	267	C	0.16102	1.302	0.457413	1.164	354.49		WB
TLNFPQITL	Pol	49	C	0.159456	1.315	0.312476	3.175	1700.8		WB
GGFIKVRQY	Pol	103	C	0.1474	1.415	0.218402	6.253	4706.58		WB
ILKEPVHGVY	Pol	460	C	0.139875	1.485	0.314137	3.138	1670.51		WB
QLIKKERVY	Pol	675	C	0.133957	1.54	0.248639	4.973	3393.3		WB
KQIIKIQNF	Pol	926	C	0.131944	1.559	0.252557	4.846	3252.46		WB
ITKIGPENPY	Pol	198	C	0.1308	1.57	0.421384	1.514	523.48		WB
RAMASEFNL	Pol	731	C	0.121006	1.67	0.492241	0.919	243.19		WB
LKEPVHGVY	Pol	461	C	0.116511	1.722	0.312324	3.178	1703.61		WB

FRAQNPEIVIY	Pol	322	C	0.114352	1.75	0.276719	4.075	2504.24		WB
RAQNPEIVI	Pol	323	C	0.101652	1.92	0.337334	2.689	1299.71		WB
VPLDEGFRKY	Pol	269	C	0.100791	1.932	0.187678	7.98	6562.61		WB
KAVRIIKIL	Rev	14	A1	0.153297	1.365	0.371364	2.145	899.38		WB
RSAEPVPLQL	Rev	66	A1	0.142886	1.457	0.267964	4.35	2753.06		WB
AVRIIKILY	Rev	15	A1 & C	0.101483	1.922	0.204064	7.014	5496.41		WB
LQLPPIERL	Rev	73	C	0.14727	1.416	0.226179	5.889	4326.75		WB
QAVRIIKIL	Rev	14	C	0.132609	1.553	0.334777	2.736	1336.17		WB
RPAEPVPLQL	Rev	66	C	0.108441	1.829	0.204054	7.015	5497.01		WB
FLNKGLGISY	Tat	38	A1	0.229185	0.918	0.45994	1.144	344.93		WB
LNKGLGISY	Tat	39	A1	0.150828	1.385	0.288096	3.761	2214.19		WB
FQTKGLGISY	Tat	38	C	0.331945	0.618	0.50803	0.821	205		WB
NSLVKHHMY	Vif	22	A1 & C	0.315654	0.654	0.380199	2.012	817.38		WB
DARLVVRTY	Vif	61	A1	0.239778	0.878	0.29818	3.494	1985.32		WB
LADQLIHLHY	Vif	102	A1	0.211333	0.993	0.422226	1.505	518.73		WB
VVSPRCEY	Vif	128	A1	0.205318	1.024	0.294344	3.595	2069.46		WB
KTKPPLPSV	Vif	158	A1	0.195951	1.076	0.267004	4.381	2781.8		WB
LADQLIHLH	Vif	102	A1	0.183138	1.155	0.301829	3.412	1908.47		WB
YSTQIDPDL	Vif	94	A1	0.12009	1.679	0.465731	1.1	323.98		WB
HNKVGSLQY	Vif	139	A1 & C	0.115281	1.738	0.240839	5.269	3692.11		WB
LGHGVSIEW	Vif	81	A1 & C	0.110124	1.806	0.295045	3.576	2053.82		WB
LQYLALKAL	Vif	145	A1	0.108206	1.832	0.431922	1.398	467.07		WB
KAILGHIVI	Vif	122	C	0.227424	0.925	0.511344	0.801	197.78		WB

LADQLIHMHY	Vif	102	C	0.210579	0.996	0.442515	1.301	416.49		WB
FADSAIRKA	Vif	115	C	0.174186	1.213	0.39092	1.861	727.86		WB
YSTQVDPGL	Vif	94	C	0.171802	1.23	0.506153	0.832	209.21		WB
DARLVIKTY	Vif	61	C	0.168372	1.253	0.238014	5.379	3806.71		WB
LADQLIHMH	Vif	102	C	0.164259	1.28	0.281472	3.937	2378.71		WB
LQYLALTAL	Vif	145	C	0.158573	1.322	0.511877	0.798	196.64		WB
KIKPPLPSV	Vif	158	C	0.142395	1.461	0.222732	6.046	4491.16		WB
LADQLIHM	Vif	102	C	0.134856	1.532	0.265169	4.438	2837.58		WB
HIVIPRCDY	Vif	127	C	0.099586	1.949	0.365113	2.239	962.31		WB
RANGWFYRHHY	Vif	34	C	0.097793	1.973	0.331941	2.788	1377.81		WB
LGQHIYNTY	Vpr	42	A	0.243689	0.864	0.424712	1.478	504.96		WB
FPRPWLHGL	Vpr	34	A	0.182312	1.16	0.338313	2.671	1286.02		WB
AIIRTLQQL	Vpr	59	A	0.117168	1.713	0.283589	3.881	2324.84		WB
HSLGQYIY	Vpr	40	C	0.11724	1.712	0.270843	4.26	2668.62		WB
WTLELLEEL	Vpr	18	C	0.106043	1.861	0.445803	1.269	401.93		WB
LGQYIYETY	Vpr	42	C	0.103312	1.897	0.400359	1.742	657.2		WB
WTIVGIEY	Vpu	23	A	0.196678	1.072	0.360771	2.304	1008.6		WB
YAIVALVVAF	Vpu	7	A	0.183715	1.151	0.660056	0.234	39.57		WB
AIVALVVAF	Vpu	8	A	0.131763	1.561	0.356484	2.368	1056.48		WB
IVALVVAF	Vpu	9	A	0.125815	1.618	0.274184	4.155	2573.88		WB
YAIVALVV	Vpu	7	A	0.105075	1.873	0.438317	1.34	435.84		WB
MVDMGHLRL	Vpu	72	C	0.319918	0.645	0.497757	0.883	229.1		WB
AIVVWTIVY	Vpu	24	C	0.17373	1.217	0.378767	2.034	830.15		WB

IVYIEYRKL	Vpu	30	C	0.168865	1.249	0.345592	2.539	1188.62		WB
IVVWTIVY	Vpu	25	C	0.136583	1.516	0.3267	2.883	1458.2		WB
IAIVVWTIVY	Vpu	23	C	0.125981	1.617	0.542971	0.648	140.46		WB
YTPGPGTRFPLTF	Nef	128	A	f 0.097320	0 1.98	0 0.35952	8 2.32	3 1022.25		WB

Supp. Table 15: Stereochemical and spatial analysis of the HLA-C*03:02 model by different computational tools

Spatial analysis		Steriochemical analysis							Overall
Verify 3D	ProSA	ERRAT	PROVE	Pro-Check			ProQ	RAMPAGE	ProTSAV
				E	W	P			
99.63	-9.08	94.667	1.8	1	6	2	8.708	99.578	~0.525

Supp. Table 16: Hydrogen bond occupancy of C*03:02 residues making contact with peptides

Amino Acid Residue	Peptide			
	$GY9_{71}$	$AY9_{324}$	$GF10_{43}$	$VL9_{109}$
Y8	0.00%	88.73%	0.00%	0.00%
E62-1	88.79%	94.43%	87.10%	64.06%
E62-2	84.76%	92.40%	0.00%	0.00%
E62-3	0.00%	91.09%	0.00%	0.00%
K65	61.87%	69.68%	91.77%	0.00%
R68	54.58%	0.00%	0.00%	0.00%
Y83	74.37%	0.00%	80.85%	0.00%
R96	0.00%	0.00%	0.00%	73.74%
Y98	0.00%	84.12%	0.00%	0.00%
N113	91.48%	0.00%	0.00%	0.00%
T142	67.90%	66.70%	59.10%	94.36%
K145-1	67.35%	56.30%	58.07%	50.00%
K145-2	54.80%	52.26%	0.00%	0.00%
W146-1	93.04%	73.91%	73.25%	0.00%
W146-2	65.79%	0.00%	0.00%	0.00%
E151-1	95.74%	0.00%	53.94%	59.68%
E151-2	92.87%	0.00%	0.00%	0.00%
E151-3	80.76%	0.00%	0.00%	0.00%
Y158	0.00%	61.57%	0.00%	0.00%

Supp. Table 17: Calculated MM-PBSA binding energy difference

Amino acid mutation	Binding Free Energy Difference			
	$GY9_{71}$	$AY9_{324}$	$GF10_{43}$	$VL9_{109}$
E62A	-25.54	-47.47	-11.91	-9.25
E151A	-23.4	3.93	-6.47	-1.98
T142A	-12.97	-9.94	-8.77	-9.91
N113A	-11.49	2	-2.64	-9.27
K145A	-10.8	-1.54	-6.9	-7.65
S115A	-10.66	-0.09	-2.63	0.8
W146A	-8.55	-4.03	-5.13	-5.51
Y83A	-7.98	-0.16	-4.84	0.07
R68A	-6.36	-0.49	-0.06	-15.49
K65A	-6.2	-2.05	-5.08	-6.33
N79A	-3.12	1.37	-0.57	1.43
S76A	-2.57	1.03	-1.35	0.62
L155A	-2.11	1.62	0.21	3.28
R96A	-1.78	0.73	1.68	-32.25
L80A	-1.64	0.33	-1.62	0.17
I123A	0.53	-0.91	-1.62	0.16
Y158A	0.43	-3.51	-1.59	-5.99
Y122A	-1.44	-0.87	-1.13	-1.16
V75A	-0.11	0.14	-0.65	0.25
Y98A	2.53	-0.13	-0.38	0.48
R61A	-0.41	0.01	-0.26	0.23
F32A	-0.22	-0.05	-0.14	-0.53
Y8A	4.19	4.21	-0.13	-5.36
L94A	1.22	-0.16	-0.13	3.05
M4A	-0.49	-0.35	-0.1	-0.56
Y170A	2.91	0.75	-0.06	0.94
Q154A	0.02	0.3	-0.04	0.16
Q95A	0.03	-0.05	-0.01	0.03
Y58A	2.81	0.74	0.11	-1.04
I141A	0.31	0.12	0.3	0.3
T72A	-1.45	-0.37	0.56	-1.4
Y6A	1.7	-0.82	0.64	-2.69
W166A	1.21	2.39	1.16	0.8
Y66A	2.93	0.53	1.25	-3.58
Q69A	0.73	3.13	1.53	0.78

All values are given in kcal/mol

Supp. Table 18: Participant characteristics and IFN-γ response

No.	Age (Yrs)	Sex	ART Duration (Mo)	Progression status	HLA locus						IFN-γ response (SFU/10^6 cells)	HIV-1 Clade
					A		B		C			
HLA-C*03:02^{+ve}												
P1	2.3	Female	18	RP	A*31:01	A*23:01	B*82:02	B*08:01	C*07:01	C*03:02	10	D
P2	13.9	Male	30	LTNP	ND	ND	B*81:01	B*45:01	C*03:02	C*18:01	940	A1
P3	20.8	Female	35	LTNP	A*30:01	A*02:01	B*42:01	B*58:01	C*03:02	C*17:01	65	C
P4	15.5	Male	4	LTNP	A*74:01	A*74:01	B*35:01	B*58:01	C*04:01	C*03:02	0	No DNA
P5	14.3	Male	45	LTNP	A*30:02	A*31:04	B*15:64	B*58:01	C*07:01	C*03:02	5	A1
P6	12.0	Female	1	LTNP	A*03:01	A*68:02	B*82:02	B*45:01	C*16:01	C*03:02		No DNA
P7	6.4	Female	68	RP	A*02:02	A*02:01	B*58:01	B*49:01	C*07:01	C*03:02	15	A1D
P8	3.0	Female	15	RP	A*74:01	A*30:02	B*58:01	B*45:01	C*06:02	C*03:02	0	A1D
P9	2.5	Female	8	RP	A*74:01	A*23:01	B*07:02	B*58:01	C*07:02	C*03:02	385	A1
P10	9.6	Male	104	RP	A*30:01	A*74:01	B*82:02	B*57:03	C*03:02	C*07:01	0	A1
P11	14.1	Female	26	LTNP	A*30:02	A*23:01	B*45:01	B*58:01	C*03:02	C*06:02	5	A1
P12	14.3	Female	35	LTNP	A*30:02	A*29:02	B*39:10	B*82:02	C*12:03	C*03:02	10	D
P13	11.3	Male	2	LTNP	A*74:01	A*23:01	B*07:02	B*58:01	C*07:02	C*03:02	15	No DNA
HLA-C*03:02^{-ve}												
C1	12.5	Female	121	RP	A*29:02	A*23:01	B*42:01	B*45:01	C*06:30	C*17:11	0	
C2	11.5	Female	1	LTNP	A*26:12	A*02:05	B*58:02	B*57:03	C*07:01	C*06:02	5	
C3	13.6	Female	1	LTNP	A*03:01	A*68:02	B*15:10	B*15:03	C*02:27	C*03:07	5	
C4	18.6	Female	8	LTNP	A*74:01	A*02:01	B*53:01	B*44:15	C*04:01	C*04:07	0	
C5	3.2	Male	35	RP	A*66:01	A*29:02	B*42:01	B*58:02	C*06:30	C*17:11	0	
C6	17.5	Female	1	LTNP	A*34:02	A*30:01	B*42:01	B*47:03	C*17:01	C*07:01	0	

C7	12.9	Male	1	LTNP	A*68:02	A*32:01	B*81:01	B*15:10	C*03:04	C*08:04	5
C8	10.9	Female	1	LTNP	A*74:01	A*23:01	B*15:03	B*15:03	C*02:10	C*02:10	5
C9	10.6	Male	101	RP	A*30:02	A*26:12	B*15:03	B*57:02	C*02:10	C*18:01	15
C10	6.3	Male	59	RP	A*30:02	A*68:02	B*42:01	B*49:01	C*17:01	C*07:01	15
C11	16.4	Male	1	LTNP	A*29:02	A*02:10	B*15:16	B*45:01	C*14:02	C*06:02	20
C12	6.7	Male	53	RP	A*34:02	A*68:01	B*35:71	B*44:37	C*07:01	C*07:05	5

Abbreviations: ND, not detected; SFU, spot forming units

9.3 UNCST PhD research approval

Uganda National Council for Science and Technology

(Established by Act of Parliament of the Republic of Uganda)

Our Ref: HS421ES **10th July 2019**

Dr. Samuel Kyobe
Principal Investigator
Makerere University Kampala
Kampala

Dear Dr. Kyobe,

I am pleased to inform you that on **10/07/2019**, the Uganda National Council for Science and Technology (UNCST) approved your study titled, **Role of Class I Human Leukocyte Antigen (HLA) Alleles in HIV Disease Progression in Botswana and Uganda".** The Approval is valid for the period of **10/07/2019** to **10/07/2022.**

Your study reference number is **HS421ES**. Please, cite this number in all your future correspondences with UNCST in respect of the above study.

Please, note that as Principal Investigator, you are responsible for:

1. Keeping all co-investigators informed about the status of the study.
2. Submitting any changes, amendments, and addenda to the study protocol or the consent form, where applicable, to the designated local Research Ethics Committee (REC) or Lead Agency, where applicable, for re-review and approval prior to the activation of the changes.
3. Notifying UNCST about the REC or lead agency approved changes, where applicable, within five working days.
4. For clinical trials, reporting all serious adverse events promptly to the designated local REC for review with copies to the National Drug Authority.
5. Promptly reporting any unanticipated problems involving risks to study subjects/participants to the UNCST.
6. Providing any new information which could change the risk/benefit ratio of the study to the UNCST for final registration and clearance.
7. Submitting annual progress reports electronically to UNCST. Failure to do so may result in termination of the research project.

Please, note that this approval includes all study related tools submitted as part of the application.

Yours sincerely,

Hellen Opolot
For: Executive Secretary
UGANDA NATIONAL COUNCIL FOR SCIENCE AND TECHNOLOGY

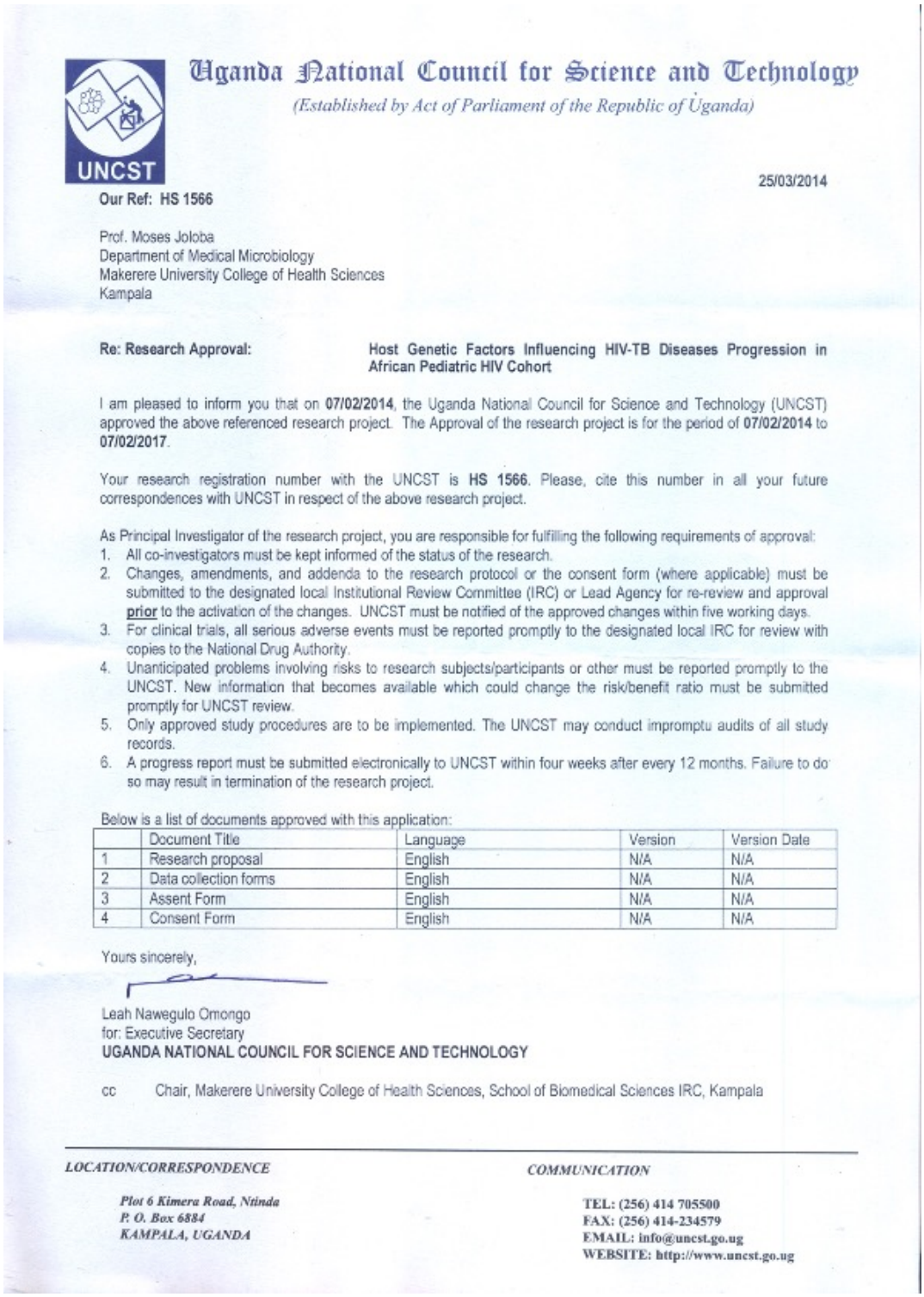
LOCATION/CORRESPONDENCE

Plot 6 Kimera Road, Ntinda
P. O. Box 6884
KAMPALA, UGANDA

COMMUNICATION

TEL: (256) 414 705500
FAX: (256) 414-234579
EMAIL: info@uncst.go.ug
WEBSITE: http://www.uncst.go.ug

9.4 UNCST CAfGEN research approval



Uganda National Council for Science and Technology

(Established by Act of Parliament of the Republic of Uganda)

UNCST

25/03/2014

Our Ref: HS 1566

Prof. Moses Joloba
Department of Medical Microbiology
Makerere University College of Health Sciences
Kampala

Re: Research Approval: **Host Genetic Factors Influencing HIV-TB Diseases Progression in African Pediatric HIV Cohort**

I am pleased to inform you that on **07/02/2014**, the Uganda National Council for Science and Technology (UNCST) approved the above referenced research project. The Approval of the research project is for the period of **07/02/2014** to **07/02/2017**.

Your research registration number with the UNCST is **HS 1566**. Please, cite this number in all your future correspondences with UNCST in respect of the above research project.

As Principal Investigator of the research project, you are responsible for fulfilling the following requirements of approval:

1. All co-investigators must be kept informed of the status of the research.
2. Changes, amendments, and addenda to the research protocol or the consent form (where applicable) must be submitted to the designated local Institutional Review Committee (IRC) or Lead Agency for re-review and approval **prior** to the activation of the changes. UNCST must be notified of the approved changes within five working days.
3. For clinical trials, all serious adverse events must be reported promptly to the designated local IRC for review with copies to the National Drug Authority.
4. Unanticipated problems involving risks to research subjects/participants or other must be reported promptly to the UNCST. New information that becomes available which could change the risk/benefit ratio must be submitted promptly for UNCST review.
5. Only approved study procedures are to be implemented. The UNCST may conduct impromptu audits of all study records.
6. A progress report must be submitted electronically to UNCST within four weeks after every 12 months. Failure to do so may result in termination of the research project.

Below is a list of documents approved with this application:

	Document Title	Language	Version	Version Date
1	Research proposal	English	N/A	N/A
2	Data collection forms	English	N/A	N/A
3	Assent Form	English	N/A	N/A
4	Consent Form	English	N/A	N/A

Yours sincerely,

Leah Nawegulo Omongo
for: Executive Secretary
UGANDA NATIONAL COUNCIL FOR SCIENCE AND TECHNOLOGY

cc Chair, Makerere University College of Health Sciences, School of Biomedical Sciences IRC, Kampala

LOCATION/CORRESPONDENCE

Plot 6 Kimera Road, Ntinda
P. O. Box 6884
KAMPALA, UGANDA

COMMUNICATION

TEL: (256) 414 705500
FAX: (256) 414-234579
EMAIL: info@uncst.go.ug
WEBSITE: http://www.uncst.go.ug

9.5 Makerere School of Biomedical Sciences IRB approval

MAKERERE UNIVERSITY
P.O. Box 7072 Kampala, Uganda
E-mail: biomedicalresearch62@gmail.com
Phone:256752575050
Fax: 256 414 532204
COLLEGE OF HEALTH SCIENCES
SCHOOL OF BIOMEDICAL SCIENCES
HIGHER DEGREES RESEARCH AND ETHICS COMMITTEE

24TH June 2019

SBS-677

To Mr. Kyobe Samuel
Principal Investigator
Department of Microbiology

Category of review
[X] Initial review
[ ] Continuing review
[ ] Amendment
[ ] Termination of study
[ ] SAEs

Decision of the School of Biomedical Sciences Higher Degrees Research and Ethics Committee (SBS-HDREC) at its 92nd REC meeting held on 18th April 2019.

In the matter concerning the review of a PhD proposal entitled, **"Role of class 1 HLA alleles in HIV disease progression in Botswana and Uganda" SBS-HDREC - 677**

The investigators have met all the requirements as stated by **SBS-HDREC** and therefore, the protocol is **APPROVED.**

The approval granted includes all materials submitted by the investigators for **SBS-HDREC** review;

1. Protocol version 1 of June 2019

and is valid until **17th April 2020**

Please note that the annual report and the request for renewal where applicable, should be submitted six weeks before expiry date of approval.

Any problems of a serious nature related to the execution of the research protocol should be promptly reported to the **SBS-HDREC**, and any changes to the research protocol should not be implemented without approval from **SBS-HDREC**, except when necessary to eliminate apparent immediate hazards to the research participant(s).

You are required to register the research protocol with the Uganda National Council for Science and Technology **(UNCST)** for final clearance to undertake the study in Uganda.

MAKERERE UNIVERSITY
SCHOOL OF BIOMEDICAL SCIENCES
APPROVED
VALID UNTIL
17 APR 2020
RESEARCH ETHICS COMMITTEE
P. O. BOX 7072, KAMPALA

Signed.....

Dr. Erisa Mwaka
Chairperson, School of Biomedical Sciences Higher Degrees Research and Ethics Committee.

9.6 CAfGEN Consent Forms

Baylor College of Medicine
Children's Foundation-Uganda
Mulago Hospital- Block 5
P.O. Box 72052, Clock Tower
Kampala – Uganda

Tel +256 -0417119100
Fax +256 -0417119166
Email: admin@baylor-uganda.org

CONSENT FORM

IRB#0000 7568 : **HOST GENETIC FACTORS INFLUENCING HIV DISEASE PROGRESSION IN AFRICAN PEADIATRIC HIV**

1.0 Background
Your child is invited to take part in a research study to help doctors and scientists learn more about genes and their effect on how quickly children with HIV develop AIDS and how they affect development of TB disease. Your child is invited to be in the study because of his/her HIV infection.

Genes are the chapters of information in our bodies that determine things like what we look like and how we behave; they tell the body what to do. Genes also have the information that determines how many of the soldiers of the body are made and how well they do their jobs; this complex information may also be passed on from parent to child. The goal of the study is to look at the genes of some children to try to discover new genes, or new patterns in the way genes are used, that may help understand reasons for how quickly HIV progresses. Please read this information and feel free to ask any questions before you agree to your child taking part in the study. For this study, we will collect, store, and use blood samples, and health information for research. If you agree, your child's blood sample and some of his or her health information will be collected and put into our safe storage place.

The storage place (also known as a biorepository or biobank) is a collection of samples and health information from different people which are stored for study. Researchers will use the samples and information in biobanks now and in the future to learn more about health, including HIV and several other diseases. Studying genes along with health information can help us to better understand what genes are responsible for causing certain diseases and how these diseases manifest.

2.0 Purpose
The purpose of this study is to examine genes in children with HIV to help understand why some children develop AIDS faster than others. We will use blood cells to look for genes that may be connected with this process.

3.0 Procedures
The research will be conducted at the following location(s): The Botswana-Baylor Children's Clinical Center of Excellence; University Of Botswana; Baylor College of Medicine Children's Foundation Uganda (Baylor Uganda); Makerere University and Baylor College Of Medicine Houston, USA.

3.1. This study will collect information from the questions you and your child answer and from the routine examination of your child, done to take care of your child's HIV related problems. The information will be kept in a safe place and the names of your child will not be used.

3.2. Two (2) or three (3) teaspoons (10-15 ml) of blood will be drawn and sent for genetic testing, one time only. Some of the blood sample drawn will be stored. We may treat the cells from your child's sample to allow us to grow them in a laboratory so as create an unlimited supply of cells for study for a long time. The cells from the cell line would be stored and used only for research. Once the blood sample is drawn, it will be used for research purposes only and it will not be available to you for any other kind of non-research or medical testing. We will store your child's samples and information in locked freezers in locked buildings. We will remove your child's name and any other information that could directly identify your child from his or her samples.

Baylor College of Medicine
Children's Foundation-Uganda
Mulago Hospital- Block 5
P.O. Box 72052, Clock Tower
Kampala – Uganda

Tel +256 -0417119100
Fax +256 -0417119166
Email: admin@baylor-uganda.org

3.3. You will bring your child to the Baylor –Uganda clinic for approximately 1-3 hours. You will be given the opportunity to ask additional questions about the study. Experienced blood collectors will be used to obtain your child's sample to keep him/her as comfortable as possible. .

Your child will continue to receive his/her regular medical care at the clinic. No new treatment will be introduced as a result of your child's participation in the study, and standard clinical procedures will not be affected by your decision to either participate or not.

3.4. If you accept, we will put your child's blood samples in a biorepository/biobank that other approved researchers can use. Here is some information about bio repositories/biobanks that people in research studies should know:

Researchers can ask to study the samples stored in the biorepository/biobank. This includes researchers from University of Botswana, Makerere University and the Baylor College of Medicine. In addition, researchers from other universities, the government, and drug or health companies in other countries may ask to study the samples. A special committee will look at each request to study the samples to find out what the researchers want to do and how they will protect your child's rights. If the committee approves the research request, we might give the researchers samples and information from many people, including your child. All of the samples and information will be labeled with barcodes only. We will not share information that could directly identify your child (like your child's name, government identification number and address) without your permission.

4.0 Access to your genetic information

Researchers can do more powerful genetic studies when they share information from human samples. They share this information by putting it into scientific databases. These databases store information from many studies conducted in many different places. Researchers can then study the combined information to learn even more about health and many different diseases.

Although your child's name and other information that could directly identify him or her will NOT be placed into any scientific database, your child's genetic information is unique to him or her and therefore there is a small chance that someone could trace it back to him or her. The risk of this happening is very small, but may grow in the future. **Researchers will always have a duty to protect your privacy and to keep your child's information confidential.**

How long will the samples and information be stored?

There is no limit on how long we will store your child's samples and information. We may keep using them for research for many years unless you decide to withdraw your child from the project.

5.0 Will you find out the results of the research?

You should not expect to get results about your child from research done through the biorepository/biobank. Researchers will study samples and information from many people; it may take many years before they know if the results have any meaning. However, in the future it may be possible for researchers to give you your child's genetic research results and we will make efforts to explain to you the meaning of the results of the research. There is also a small chance that researchers could find something soon that might be important to your child's

Baylor College of Medicine
Children's Foundation-Uganda
Mulago Hospital- Block 5
P.O. Box 72052, Clock Tower
Kampala – Uganda

Tel +256 -0417119100
Fax +256 -0417119166
Email: admin@baylor-uganda.org

health. If this happens, we may contact you to find out if you would like to learn more. However, even if we find something important to your child's health, we cannot guarantee that you will be contacted.

6.0 Will you be contacted in the future about this or other research?

We may want to contact you in the future. You can decide now whether or not you want to be contacted. You can also change your mind later during the course of the study.

If you agree, we may contact you for several reasons. For example, we may ask for more samples or more information about your child's health. We may want to ask if you want to participate in other research. We will not notify you every time your child's samples and information are used. However, some researchers might do a study and need to contact you. For example, they might ask you and/or your child to do a phone interview or come in to be seen by a researcher or doctor. If a study like this is approved, someone from this project will tell you about the study so you can decide if you want to participate. There will be a new consent process just for that study. If at any time you decide you no longer want to be contacted about future studies or you have any questions, you can call Dr. Angella Mirembe at (+256 703559328).

7.0 Potential Risks and Discomforts

The blood draw will cause minor pain but it will be brief and your child may get a bruise but this is not common. However, if this happens, it will clear very quickly. There may be a very small risk of infection due to the blood draw but this rarely happens. This blood draw will not differ from the routine blood draws the child gets. The gene testing may uncover other information about the risk of certain diseases. There is a risk of getting unexpected and/or unwanted information if the results of the genetics testing are shared with you.

There is a small risk of loss of confidentiality. The staff will protect confidentiality by assigning a study identification number and storing information and blood samples with this number and without information like name and address that could identify your child. The information will be locked away in filing cabinets and kept in a locked office and in a password-protected computer file available only to study staff. The confidentiality of the data will be maintained within legal limits. The blood samples will be kept in a laboratory in a freezer designed for these samples, with a study number but no name or other identifier. The data from the study may be published; however, no names or other ways of identifying the subjects will be published.

Research records are separate from medical records. We will not place any information from this project in your child's medical records.

Researchers who study your child's sample and information will not know who your child is. We will give them only barcode numbers; we will not give them any information that directly identifies your child. The researchers must sign an agreement that they will not try to find out who your child is.

Study staff will update you in a timely way on any new information that may affect your decision to stay in the study.

8.0 Potential Benefits

Baylor College of Medicine
Children's Foundation-Uganda
Mulago Hospital- Block 5
P.O. Box 72052, Clock Tower
Kampala – Uganda

Tel +256 -0417119100
Fax +256 -0417119166
Email: admin@baylor-uganda.org

The benefits of participating in this study may be: finding one or a few genes that your child or another family member might have that is/are medically important. However, you may receive no direct benefit from participating in the study.

9.0 Alternatives
You may choose to not participate in this study or agree to only parts of it, and whatever your decision is; it will not affect your routine medical care.

10.0 Subject Costs and Payments
You will not be asked to pay any costs related to this research.

You will not be paid for taking part in this study; however, you will receive fifteen thousand shillings (15000/-) to cater for your travel expenses to and from the clinic.

11.0 Subject's Rights

Your signature on this consent form means that you have received and understood the information about this study and that you agree on behalf of your child to take part in this study.

You will be given a copy of this signed form to keep. You are not giving up any of your rights by signing this form. Even after you have signed this form, you may change your mind at any time. Please contact the study staff if you decide to stop taking part in this study.

If you choose not to take part in the research or if you decide to stop taking part later, your benefits and services will stay the same as before this study was discussed with you. You will not lose these benefits, services, or rights.

12.0 Your Health Information

We may be collecting health information that could be linked to you (protected health information). This protected health information might have your name, date of birth, address, or something else that identifies you attached to it. We need your permission to collect this information for this study. Your signature on this form means that you give us permission to use your protected health information for this research study.

If you decide to take part in the study, your protected health information will not be given out except as allowed by law or as described in this form. Everyone working with your protected health information will work to keep this information private. The results of the data collected from this study may be published. However, your child will not be identified by name.

If you decide to stop taking part in the study or if you are removed from the study, you may decide that you no longer allow protected health information that identifies you to be used in this research study. Contact the study Investigator at (+256 772 462686) so that you can inform her of your decision. The investigator will honor your decision.

The principal investigator, DR. ADEODATA KEKITIINWA and/or someone he/she appoints in her place will try to answer all of your questions. If you have questions or concerns at any time, or if you need to report an injury related to the research, you may speak with a member of the study staff:

Baylor College of Medicine
Children's Foundation-Uganda
Mulago Hospital- Block 5
P.O. Box 72052, Clock Tower
Kampala – Uganda

Tel +256 -0417119100
Fax +256 -0417119166
Email: admin@baylor-uganda.org

Participants Rights:
In case you want to know your rights as a study participant contact Makerere University School of Biomedical Sciences Research and Ethics Committee chair, Dr. Erisa Mwaka on 0752 575050 he can also answer your questions and concerns about your rights as a research subject. Call the IRB office if you would like to speak to a person independent of the investigator and research staff for complaints about the research, if you cannot reach the research staff, or if you wish to talk to someone other than the research staff.

1. If you agree to your child's participating in the study now, please sign here below:

Child's name _____ Age _____ (years) Date _____

_____ _____ _____
Parent/Caregiver's Name (dd/mmm/yyyy) Signature/Thumb print Date

_____
Relationship to Subject

If unable to read and write,

_____ _____ _____
Witness' Name Signature Date (dd/mmm/yyyy)

2. If you agree to be contacted in future for this or other studies, please sign here below:

Child's name _____ Age _____ (years) Date _____

_____ _____ _____
Parent/Caregiver's Name Signature/Thumb print Date (dd/mmm/yyyy)

_____
Relationship to Subject

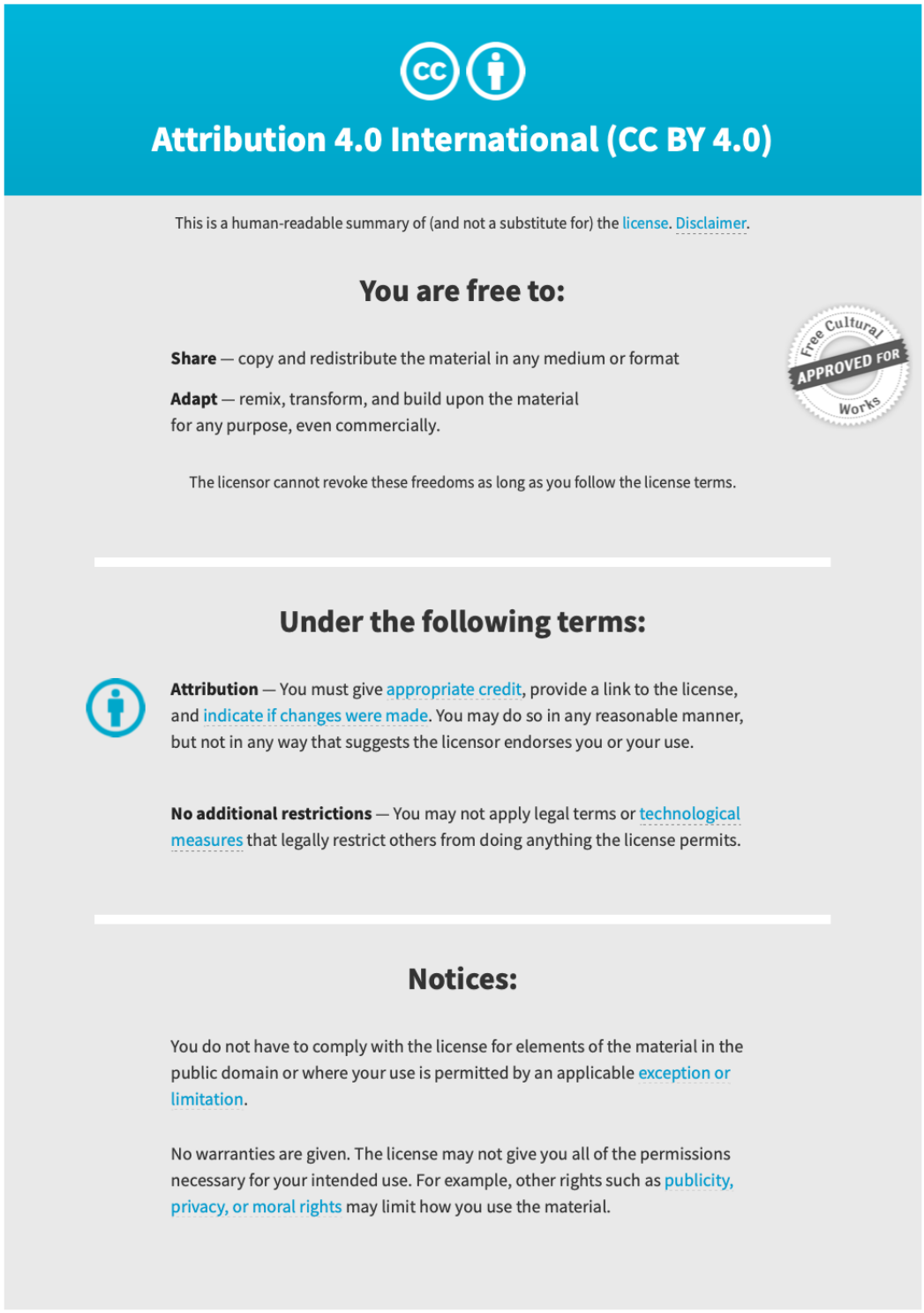
If unable to read and write,

_____ _____ _____
Witness' Name Signature Date (dd/mmm/yyyy)

_____ _____ _____
Study staff obtaining consent Name (print) Staff Signature Date (dd/mmm/yyyy)

9.7 Permissions to Use Copyrighted Material

9.7.1 Permissions to use Figure 2-2 and to reproduce Kyobe et al. Front. Genet., 26 August 2021



Attribution 4.0 International (CC BY 4.0)

This is a human-readable summary of (and not a substitute for) the license. Disclaimer.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

9.7.2 Permissions to use Figure 2-4

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Aug 24, 2023

This Agreement between Makerere University -- Samuel Kyobe ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5615330262679
License date	Aug 24, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Immunology
Licensed Content Title	Proteasomes in immune cells: more than peptide producers?
Licensed Content Author	Marcus Groettrup et al
Licensed Content Date	Dec 11, 2009
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of	1

9.7.3 Permissions to use Figure 3-4

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Aug 24, 2023

This Agreement between Makerere University -- Samuel Kyobe ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5607251496578
License date	Aug 13, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Genome Medicine
Licensed Content Title	HLAreporter: a tool for HLA typing from next generation sequencing data
Licensed Content Author	Yazhi Huang et al
Licensed Content Date	Mar 16, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations