



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

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, Katagiryra 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 SV,  
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

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

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

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,  $P_c = 0.0259$ ; B\*58:01, aOR 1.89,  $P_c = 0.033$ ; C\*03:02, aOR 4.74,  $P_c = 0.033$ ). 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,  $P_c = 0.025$ ) and HLA-B\*58:01-C\*03:02 (aOR 4.88,  $P_c = 0.011$ ) 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,  $P_c = 0.005$ ) that appears to limit disease progression, despite weak LD ( $r^2 = 0.18$ ) 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.

**Keywords: AIDS, childhood HIV, long-term non-progression, genetics, genomics**

## INTRODUCTION

HIV-1 remains a significant public health concern in Africa, with an estimated 26 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 10 years even without ART (Warszawski et al., 2007) – so-called long-term non-progressors (LTNPs). We have documented a cohort of childhood LTNPs from Uganda and Botswana (Kyobe et al., under review) 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 and 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 (Goulder and 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 (Serwanga et al., 2009; Matthews et al., 2011; Peterson et al., 2013; Sampathkumar et al., 2014; Shepherd et al., 2015). For example, HLA-B\*58:01 has been associated with slow progression in Uganda and South Africa (Serwanga et al., 2009; Payne et al., 2014), but this observation was not seen in Kenya and Botswana (Peterson et al., 2013; Payne et al., 2014); 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 and colleagues 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 (Carrington et al., 1999; Cao et al., 2001; 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 (Thobakgale et al., 2009; Adland et al., 2015; Shepherd et al., 2015).

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 (Sherman et al., 2019a; Choudhury et al., 2020), 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.

## MATERIALS 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 (Retshabile et al., 2018; Mwesigwa et al., 2021; Kyobe et al., under review). 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 10 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 3 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 >20× 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 (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 ( $\geq 90$  bp) 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 10× were used during alignment; this is stringent and improves the accuracy of HLAreporter (Huang et al., 2015).

### Validation

We conducted validation of HLAreporter using Lamda micro-SSP kits (according to manufacturers 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  $\chi^2$  tests, while continuous variables were summarized 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 and 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 and 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<sup>1</sup> (Lancaster et al. 2003, 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 (Yang et al., 2020), assuming an additive effects model. Our preliminary study (Kyobe et al., under review), 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  $F_{st}$  (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 and 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 and Song, 2017).

For haplotype analyses, we used the Haplo.Stats package (v1.7.9) (Martin and 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:  $PAR = \frac{P(OR-1)}{P(OR-1)+1} \times 100$ , where  $p$  denotes the frequency of alleles or haplotypes among LTNPs and OR, is the odds ratios from the logistic regression models (Levin,

<sup>1</sup><https://www.hiv.lanl.gov/content/index>

1953). False discovery rate (FDR) was used to adjust  $P$  values for the number of alleles tested ( $P_{\text{corrected}}$ ,  $P_c < 0.05$ ). All statistical analysis was performed using Stata13 and R software v1.2.5033 (Gutierrez, 2010).

## 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 1**, 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.

### 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 **Supplementary Figure 1**) and  $F_{\text{st}}$  based on Weir and Cockerham's method using genome-wide data (Weir and Cockerham, 1984; Retshabile et al., 2018). The PCA plots (**Supplementary Figure 1**) and  $F_{\text{st}}$  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 vs. PC3) and the overall  $F_{\text{st}}$  differences are small. We started by exploring differences in HLA allele distributions (**Figure 1**)

**TABLE 1** | Baseline characteristics among LTNPs and RPs.

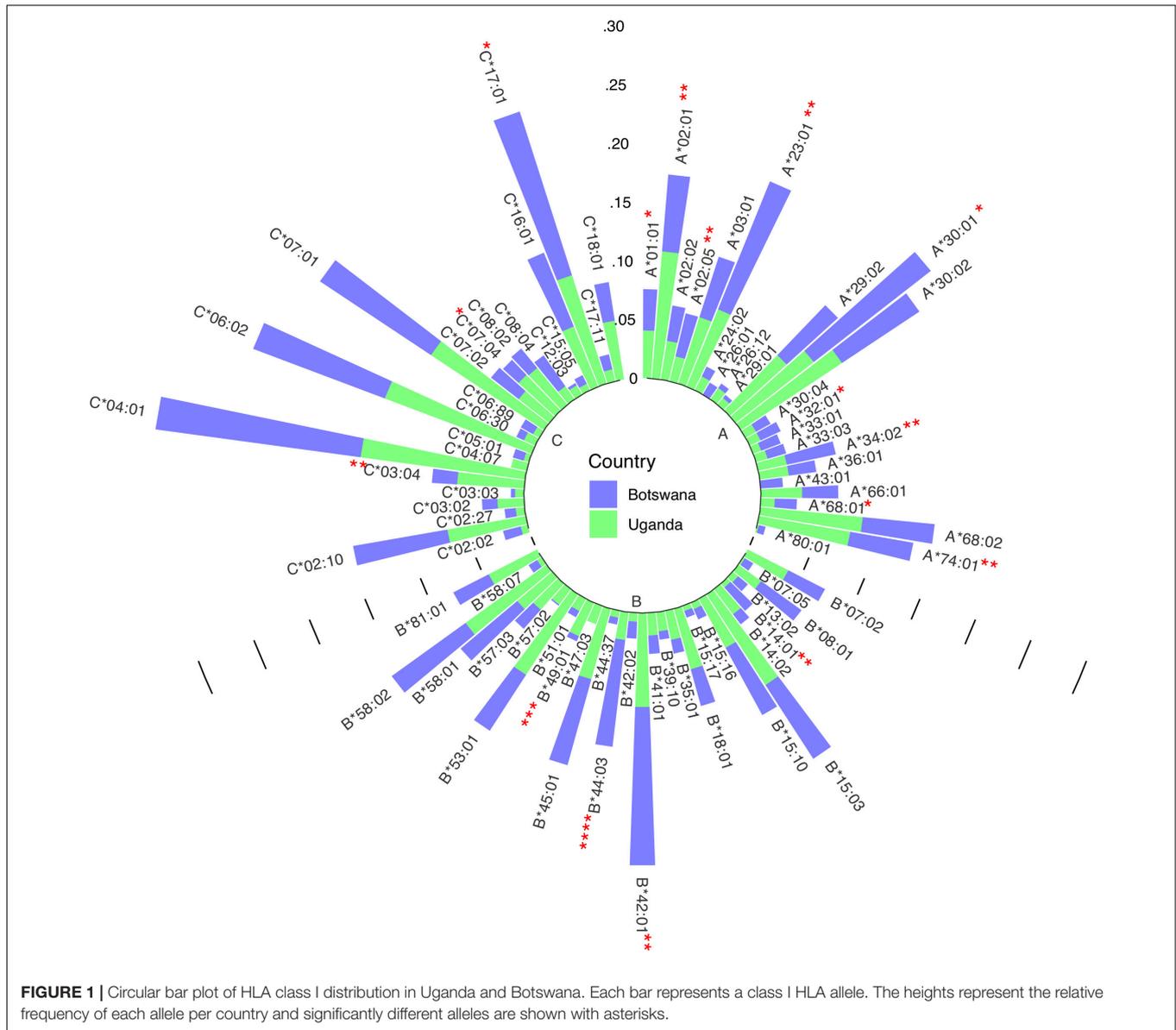
Variables	LTNPs	RPs	$P$ -value
<b>Country</b>			
Botswana	187 (47.5)	258 (61.4)	<b>&lt;0.001</b>
Uganda	207 (52.5)	162 (38.6)	
<b>Gender</b>			
Female	228 (57.9)	192 (45.7)	<b>&lt;0.001</b>
Male	166 (42.1)	228 (54.3)	
Age (yr)	16.6 (13.7–19.4)	9.6 (6.6–12.1)	<b>&lt;0.001</b>
Time to progression (mo)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

*Bold indicates variables that are different.*

and pairwise LD between allele classes in the two countries (**Figures 2A–C** and **Supplementary Tables 1–3**). 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 1**). 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 1** and **Supplementary Table 4**). Surprisingly, HLA-A\*30:01, B\*42:01, and C\*17:01 in Ugandans and C\*07:01 in Botswana were observed at between two and 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 (**Supplementary Table 5**) (Cao et al., 2004; Tshabalala et al., 2018; Gonzalez-Galarza et al., 2020). This trend was similar when the data were analyzed by country (**Supplementary Table 4**). The frequencies of 16 alleles were significantly different between Uganda and Botswana (**Figure 1** and **Supplementary Table 6**), 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; Kijak et al., 2009; Tshabalala et al., 2018; Gonzalez-Galarza et al., 2020), 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).

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 2**), although quantitatively lower than reported in European ancestry cohorts ( $D' = 0.928$ ) (Cao et al., 2001; Tokić et al., 2020), this is consistent with LD in healthy African populations. In our combined cohort, the strongest LD was observed between HLA-B\*42:01 and C\*17:01 ( $r^2 = 0.65$ , **Figure 2A**). 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 exhibited low LD patterns between class I HLA loci (**Figures 2B,C**).

The number of bi-allelic haplotypes was not significantly different between Uganda and Botswana (**Supplementary Table 7**), 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 3** and **Supplementary Tables 8, 9**). 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 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 (Supplementary Tables 6, 8). 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.

### Established HLA-B Alleles and a Novel HLA-C Allele Show Association With LTNP in Ugandan and Botswana Children

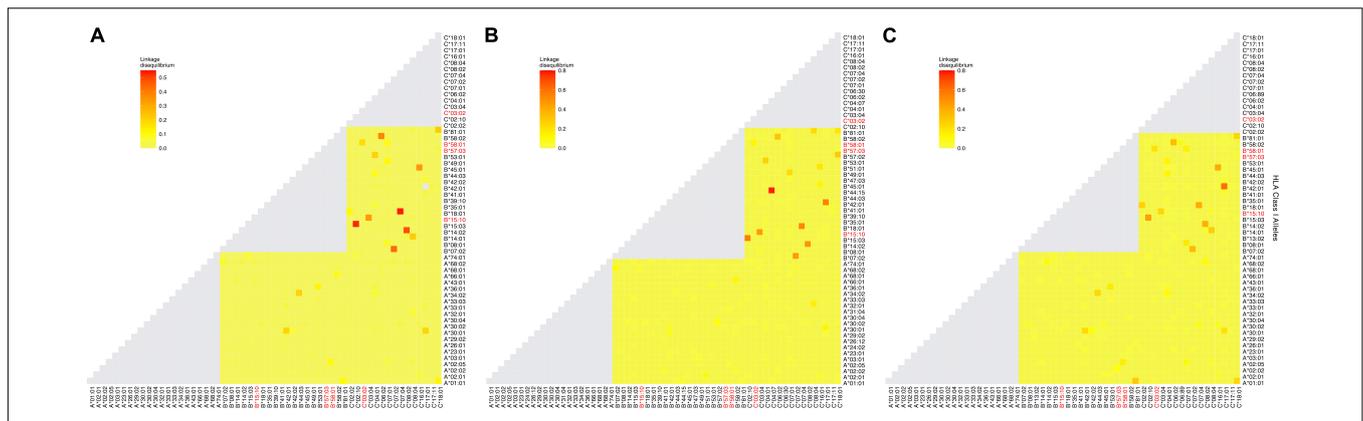
Although there was no difference detected between LTNP (cases) and RP (controls) in the total number of alleles ( $p = 0.963$ ,

Table 1), 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 ( $P_c < 0.05$ , Figure 4 and Table 3) - HLA-B\*57:03 [3.9 vs. 1.2%, aOR 3.21 (95% CI 1.50–6.86),  $P_c = 0.026$ ], HLA-B\*58:01 [6.6 vs. 3.9%, aOR 1.89 (95% CI 1.21–2.96),

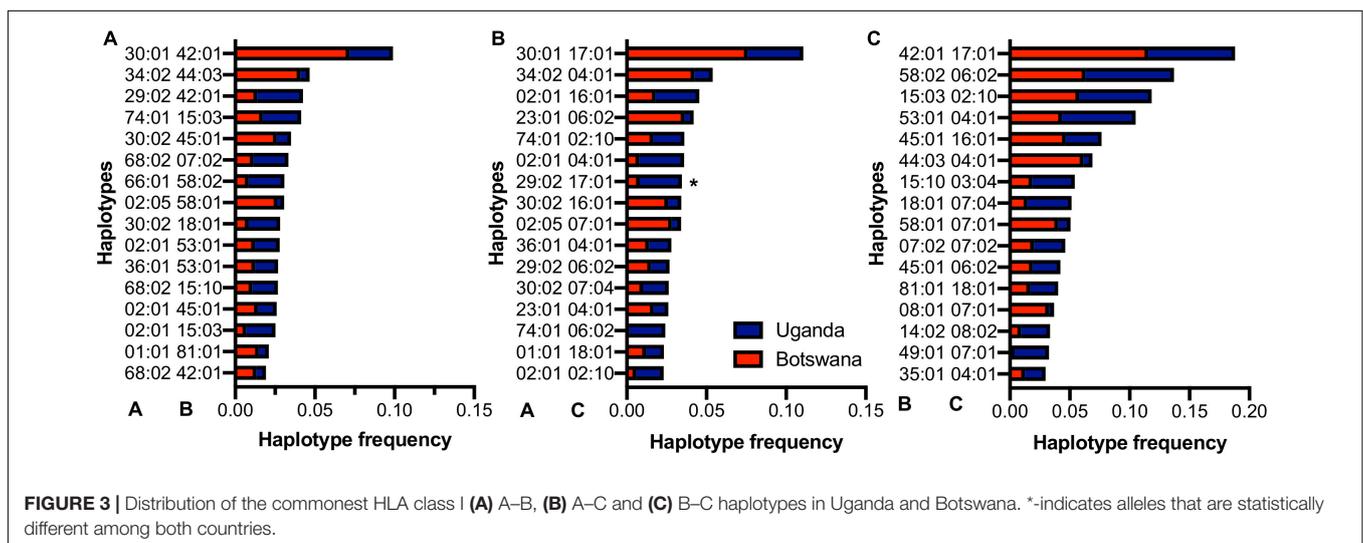
**TABLE 2 |** Global pairwise linkage disequilibrium.

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.



**FIGURE 2** | 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  $\geq 1\%$ . The colors indicate increasing strength of LD from yellow to red. Alleles on the same HLA locus (no pairwise LD comparisons) are indicated in gray. Alleles associated with LTNP in the cohort are highlighted in red.



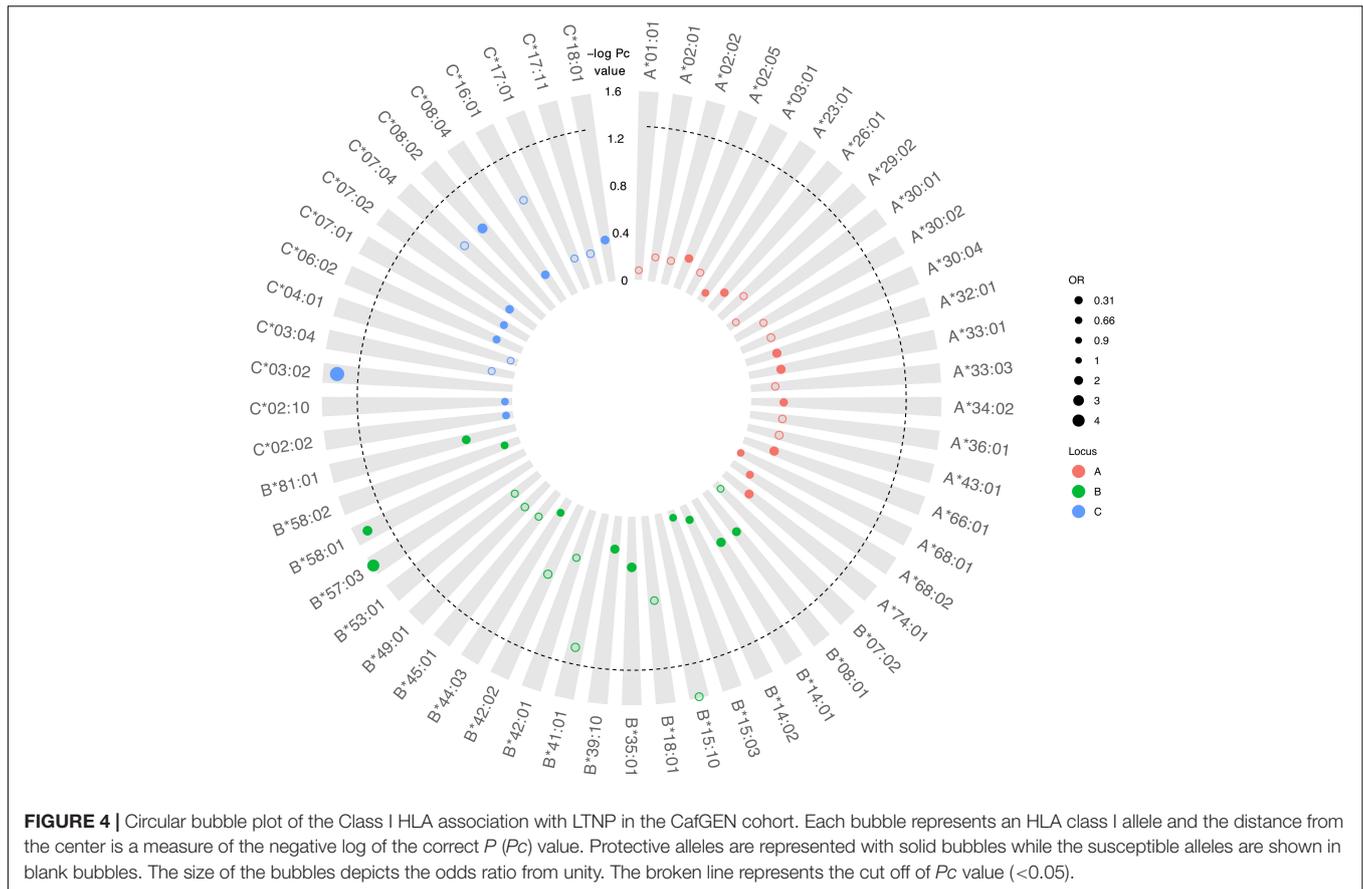
**FIGURE 3** | Distribution of the commonest HLA class I (A) A–B, (B) A–C and (C) B–C haplotypes in Uganda and Botswana. \*-indicates alleles that are statistically different among both countries.

$P_c = 0.033$ ] and HLA-C\*03:02 [2.9 vs. 0.6%, aOR 4.74 (95% CI 1.74–12.85),  $P_c = 0.033$ ] (Figure 4). Additionally, we found HLA-B\*15:10 [3.7 vs. 7.7%, OR 0.48 (95% CI 0.30–0.77),  $P_c = 0.0259$ ] to be enriched among RPs (Figure 4 and Table 3) suggesting a predisposition toward rapid progression. These effects were generally consistent (in both magnitude and direction) across our two populations (Figure 5 and Supplementary Table 10). Most of these associated alleles were in weak LD (Figure 2A), 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  $\sim 16.5\%$  (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 (Supplementary Table 11), we found that participants who were heterozygous at any of our four significantly associated alleles (HLA-B\*57:03,  $P_c = 0.0043$ ; –B\*58:01,  $P_c = 0.0258$ ; –C\*03:02,  $P_c = 5.7 \times 10^{-4}$ ), and (B\*15:10  $P_c = 0.0082$ ) were significantly more likely to be LTNPs (Supplementary Table 11).

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 (Supplementary Table 12). 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 B\*57:03-C\*07:01 [2.0 vs. 0.3%, OR 5.40 (95% CI 1.40–20.79),  $p = 0.025$ ] and 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 and Supplementary Table 12). Participants with the 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 and Supplementary Table 12). 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.

**TABLE 3 |** HLA alleles associated with LTNP in Uganda and Botswana.

HLA allele	Allele frequency in cohort			$P$ -value	OR (95% CI)	$P_c$ value <sup>†</sup>
	LTNPs ( $n = 393$ )	RPs ( $n = 416$ )	Total			
<b>Protective</b>						
B*57:03	0.038	0.012	0.025	0.0026	3.21 (1.50–6.86)	<b>0.025</b>
B*58:01	0.066	0.039	0.052	0.005	1.89 (1.21–2.96)	<b>0.033</b>
C*03:02	0.028	0.006	0.017	0.0022	4.74 (1.74–12.85)	<b>0.033</b>
C*08:02	0.042	0.019	0.030	0.0284	2.05 (1.07–3.90)	0.124
<b>Susceptible</b>						
B*15:10	0.037	0.077	0.057	0.0025	0.48 (0.30–0.77)	<b>0.025</b>
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

Results of all observed HLA allele association per country are provided in Supplementary Table 10.

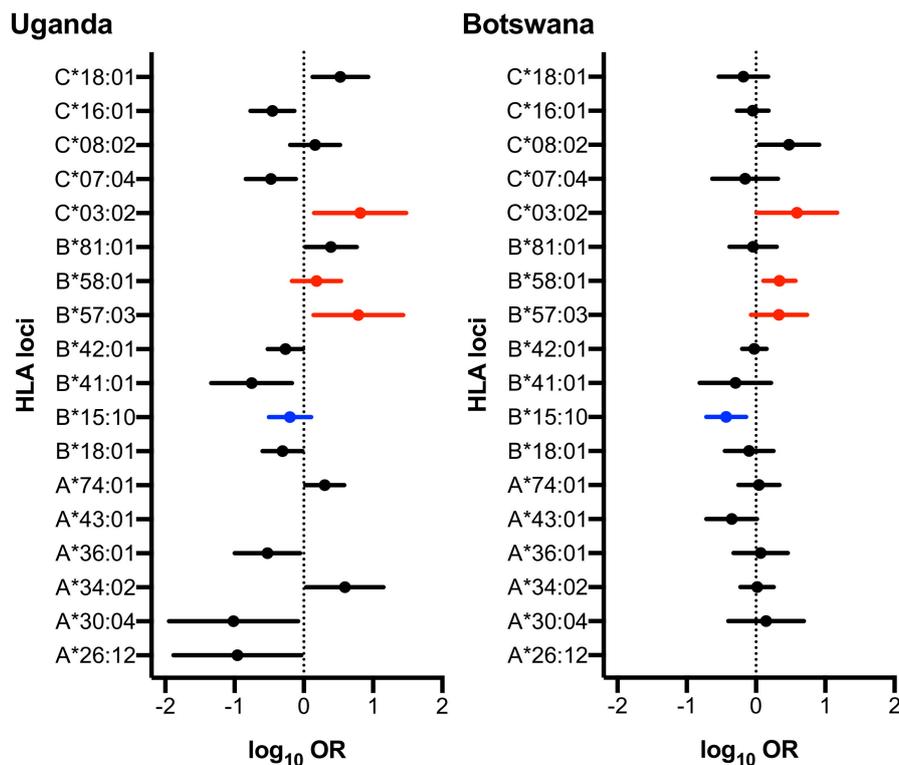
Bold indicates alleles that satisfy FDR correction  $p < 0.05$ .

<sup>†</sup>FDR adjusted  $P$ -value.

$n$ , number of participants.

### 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 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 2A and Supplementary Table 9). 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



**FIGURE 5 |** 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.

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 and 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 6**), which was also observed for children expressing HLA-B\*58:01 without C\*03:02 (B58+/C03- OR 1.37,  $P_c = 0.311$ ). We found that the protective effect of HLA-C\*03:02 remained even without concomitant co-inheritance of B\*58:01 (C03 + /B58-) although the statistical significance was reduced in accordance with the limited sample size (OR 4.15,  $P_c = 0.106$ ); the reverse was also true (B58+/ C03-, OR 1.37,  $P_c = 0.311$ ). 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,  $P_c = 0.005$  ( $p < 0.05/6$  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 (**Figures 4, 5 and Supplementary Table 10**) 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 C\*03:02 (A74+/C03- aOR 1.66,  $P_c = 0.048$ ), the combined protective

effect of the haplotype was much stronger than either allele separately (OR 6.58,  $P_c = 0.0149$ ). HLA-C\*03:02 and A\*74:01 are in very weak LD in our cohort ( $r^2 = 0.02$ , **Figure 2A**), suggesting the effect of HLA-C\*03:02 was synergistic to the marginal effect of 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 (B\*42:01, A\*43:01, B\*15:10, B\*41:01) alleles (**Figure 6**), 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.

## 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 C\*03:02 association with LTNP. The effect sizes of the protective alleles and haplotypes ranged from 1.89 to 5.4, and, consistent

**TABLE 4** | Haplotype frequency and association with LTNP.

Haplotype*	Haplotype frequency in cohort			Haplotype effect <sup>§</sup>	
	LTNP	RP	P-value <sup>†</sup>	OR (95% CI)	Pc value
<b>Protective</b>					
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 $p = 0.006$			
<b>Susceptible</b>					
A*29:02~C*17:01	0.010	0.029	0.008	0.26 (0.08–0.79)	0.003
		global $p = 0.155$			

<sup>†</sup>The  $p$ -value is based on the individual haplotype association with LTNP compared to RP.

<sup>§</sup>Odds ratios from generalized linear regression models comparing haplotype frequency with the most frequent haplotype and adjusted for country and gender.

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).

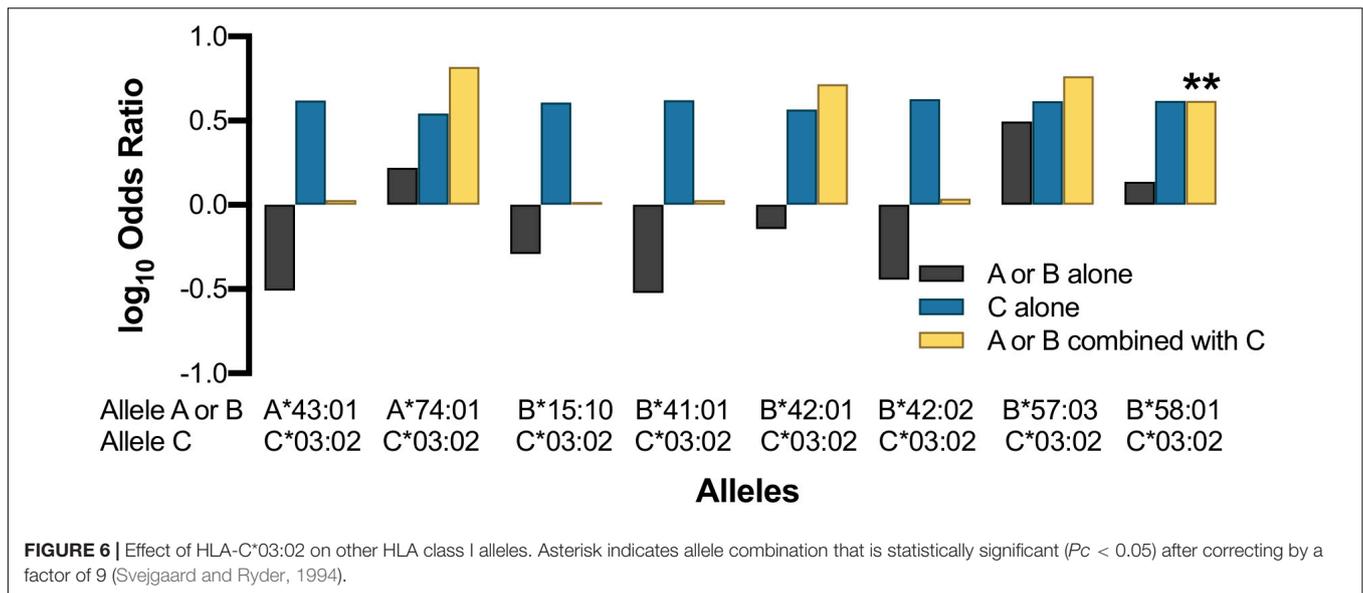
with previous studies, these variants explain up to 16.5% of LTNP variation in our cohort. Our results expand upon the known and anticipated alleles and effect sizes in comparable HIV-1 infected pediatric populations in Africa (Shepherd et al., 2015).

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<sup>+</sup> T-cell count during early infection - two phenotypes that positively correlate with LTNP status in adults (Tang et al., 2010; Chopera et al., 2011). 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<sup>+</sup> 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 (Serwanga et al., 2009; Leslie et al., 2010; Tang et al., 2010; Matthews et al., 2011; Peterson et al., 2013), as well as other Caucasian populations (Klepiela et al., 2004; Lazaryan et al., 2006; Chopera et al., 2011; Carlson et al., 2012; Gijbsbers et al., 2013) and collectively, support the notion that heterozygous states of HLA protective alleles are advantageous, probably in both children and adults (Carrington et al., 1999; Arora et al., 2020), likely by providing alternative or diverse epitope presentation pathways especially in the emergence of HIV-1 escape mutants (Carlson et al., 2012; Gijbsbers et al., 2013; Arora et al., 2020). 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 fourfold 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 (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; van der Zwan et al., 2018). Other HLA-C alleles - C\*08:02 and C\*08:02 - 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 3). 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 (Thomas et al., 2009; Apps et al., 2013), however, the mean expression of HLA-C\*03:02 is lower than C\*08:02 (Apps et al., 2013) on CD3<sup>+</sup> 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; 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 B\*58:01 is similar to the individual allelic effect of C\*03:02 (OR 4.88 vs. 4.78), and the allelic effect of carrying 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 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 C\*03:02 was consistently observed in combinations with detrimental alleles (Figure 6). 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 (Serwanga et al., 2009; Payne et al., 2014); 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 deficient 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 A\*29:01-C\*17:01 was also associated with rapid progression; consistent with Sampathkumar et al. who found that a 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 (**Supplementary Table 5**). 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** and **Supplementary Table 5**).

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 believe 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- $\kappa$ B, 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.

## 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 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.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://h3africa.org/wp-content/uploads/2018/05/App-D-H3Africa-Data-and-Biospecimen-Access-Committee-Guidelines-final-10-July-2017.pdf>, NA.x.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Makerere University School of Biomedical Sciences IRB, Uganda National Council for Science and

Technology (UNCST), University of Botswana IRB, Botswana Health Research and Development Committee (HRDC), and Baylor College of Medicine IRB. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

SK, MW, MJ, GK, AK, BN, MM, GA, IK, SWM, MT-J, CB, GM, and NH: conceptualization, writing – review and editing, funding acquisition, investigation, and project administration. SK, SM, ErK, MW, FK, AM, BN, JF, GR, BM, LW, and KM: data curation. SK, SM, and NH: formal analysis. JK-L: funding acquisition. EW and JK-L: investigation. SK, GK, GR, BN, MM, CB, GM, and NH: methodology. EW: project administration. GM, NH, MJ, AK, MM, and JK-L: supervision. SK and NH: validation and writing – original draft. All authors contributed to the article and approved the submitted version.

## 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. The result and their interpretations do not reflect the views of the National Institutes of Health. 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 was supported through the DELTAS Africa Initiative (Grant No. 107743). The DELTAS Africa Initiative was 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 United Kingdom 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, Thembele 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. Finally, we acknowledge the families and children who participated in the study.

## 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:e1004954. 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* 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., Roeder, 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.* 9:2539. 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* 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:e05345. 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

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.720213/full#supplementary-material>

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* 18:90. 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

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

Gonzalez-Galarza, F. F., McCabe, A., Dos Santos, E. J. M., 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

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.* 2, 728–733. 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.* 7:25. 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
- 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* Switzerland: Springer. 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. *NPJ Genom. Med.* 6:24. 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.
- 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. U.S.A.* 111, E5393–E5400. doi: 10.1073/pnas.1413391111
- 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., Wachihhi, 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:e101475. 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 therapy-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.
- 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., Marčić, 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.* 10:5513. 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.* 2018:2031571. 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.* 4:4862. doi: 10.1038/srep04862
- UNAIDS (2019). *UNAIDS 2019 Data*. Available online at: [https://www.unaids.org/sites/default/files/media\\_asset/2019-UNAIDS-data\\_en.pdf](https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf) (accessed April 7, 2020).
- 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.* 9:2880. doi: 10.3389/fimmu.2018.02880
- Vargas, L., de, B., Dourado, R. M., Amorim, L. M., Ho, B., Calonga-Solis, V., et al. (2020). Single nucleotide polymorphism in KIR2DL1 is associated with HLA-C expression in global populations. *Front. Immunol.* 11:1881. doi: 10.3389/fimmu.2020.01881
- Verweij, M. C., Horst, D., Griffin, B. D., Luteijn, R. D., Davison, A. J., Rensing, 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., Firion, 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. A., 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

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kyobe, Mwesigwa, Kisitu, Farirai, Katagirya, Mirembe, Ketumile, Wayengera, Katabazi, Kigozi, Wampande, Retshabile, Mlotshwa, Williams, Morapedi, Kasvosve, Kyosiimire-Lugemwa, Nsangi, Tsimako-Johnstone, Brown, Joloba, Anabwani, Bhekumusa, Mpoloka, Mardon, Matshaba, Kekitiinwa and Hanchard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.