

# Distinct genetic architectures and environmental factors associate with host response to the $\gamma$ 2-herpesvirus infections

Neneh Sallah <sup>1,2,14</sup>✉, Wendell Miley<sup>3</sup>, Nazzarena Labo<sup>3</sup>, Tommy Carstensen<sup>1,4</sup>, Segun Fatumo<sup>1,2,5</sup>, Deepti Gurdasani<sup>1,6</sup>, Martin O. Pollard <sup>1,4</sup>, Alexander T. Dilthey <sup>7</sup>, Alexander J. Mentzer <sup>8,9</sup>, Vickie Marshall<sup>3</sup>, Elena M. Cornejo Castro <sup>3</sup>, Cristina Pomilla<sup>1,4</sup>, Elizabeth H. Young<sup>1,4</sup>, Gershim Asiki <sup>10</sup>, Martin L. Hibberd <sup>2</sup>, Manjinder Sandhu <sup>4</sup>, Paul Kellam <sup>11,12</sup>, Robert Newton<sup>5,16</sup>, Denise Whitby<sup>3,16</sup> & Inês Barroso <sup>1,13,15,16</sup>✉

Kaposi's sarcoma-associated herpesvirus (KSHV) and Epstein-Barr Virus (EBV) establish life-long infections and are associated with malignancies. Striking geographic variation in incidence and the fact that virus alone is insufficient to cause disease, suggests other co-factors are involved. Here we present epidemiological analysis and genome-wide association study (GWAS) in 4365 individuals from an African population cohort, to assess the influence of host genetic and non-genetic factors on virus antibody responses. EBV/KSHV co-infection (OR = 5.71(1.58–7.12)), HIV positivity (OR = 2.22(1.32–3.73)) and living in a more rural area (OR = 1.38(1.01–1.89)) are strongly associated with immunogenicity. GWAS reveals associations with KSHV antibody response in the *HLA-B/C* region ( $p = 6.64 \times 10^{-09}$ ). For EBV, associations are identified for VCA (rs71542439,  $p = 1.15 \times 10^{-12}$ ). Human leucocyte antigen (HLA) and trans-ancestry fine-mapping substantiate that distinct variants in *HLA-DQA1* ( $p = 5.24 \times 10^{-44}$ ) are driving associations for EBNA-1 in Africa. This study highlights complex interactions between KSHV and EBV, in addition to distinct genetic architectures resulting in important differences in pathogenesis and transmission.

<sup>1</sup>The Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK. <sup>2</sup>London School of Hygiene & Tropical Medicine, London, UK. <sup>3</sup>Viral Oncology Section, AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc., Frederick, MD, USA. <sup>4</sup>Department of Medicine, University of Cambridge, Cambridge, UK. <sup>5</sup>MRC/UVRI at the London School of Hygiene & Tropical Medicine, Entebbe, Uganda. <sup>6</sup>Queen Mary University London, London, UK. <sup>7</sup>Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany. <sup>8</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>9</sup>Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. <sup>10</sup>African Population and Health Research Center, Nairobi, Kenya. <sup>11</sup>Department of Infectious Diseases, Imperial College London, London, UK. <sup>12</sup>Kymab Ltd, Babraham Research Complex, Cambridge, UK. <sup>13</sup>MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. <sup>14</sup>Present address: London School of Hygiene & Tropical Medicine, London, UK. <sup>15</sup>Present address: Exeter Centre of Excellence in Diabetes (ExCEED), University of Exeter Medical School, Exeter, UK. <sup>16</sup>These authors contributed equally: Robert Newton, Denise Whitby, Inês Barroso. ✉email: [Neneh.Sallah@lshtm.ac.uk](mailto:Neneh.Sallah@lshtm.ac.uk); [ines.barroso@exeter.ac.uk](mailto:ines.barroso@exeter.ac.uk)

The gamma-herpesviruses, Epstein–Barr Virus (EBV) and Kaposi sarcoma-associated herpesvirus (KSHV) are oncogenic viruses associated with several malignancies, including >95% of Burkitt’s Lymphoma (BL) in parts of sub-Saharan Africa and 100% of cases of Kaposi’s Sarcoma (KS), respectively<sup>1,2</sup>. While EBV is ubiquitous globally, the prevalence of, and deaths caused by several associated malignancies, display extensive geographic variation. In contrast, KSHV displays striking geographic variation that parallels associated KS incidence, with highest prevalence reported in sub-Saharan Africa<sup>1,3</sup>. The inter and intra-continental heterogeneity in the prevalence of KSHV and incidence of associated malignancies, while attributable to a number of factors including differences in pathogen prevalence, may also be influenced by many environmental determinants, host genetics and gene–environment interactions in different populations. Synergistic and antagonistic interactions have been reported for KSHV–EBV co-infection; B-lymphocytes act as a reservoir of latent infection, in dually infected cells, both viruses subvert the host immune response and inhibit lytic reactivation of each other<sup>4–8</sup>. While viral genetic factors have been extensively studied<sup>9,10</sup>, host genetic factors and their interactions with the environment, leading to potential disease outcomes, are largely unknown and require investigation, particularly in Africa.

We and others have reported associations in the HLA class II region with EBV EBNA-I antibody response<sup>11–14</sup>. On the other hand, host genetic studies for KSHV are lacking, with no genome-wide association study (GWAS) performed for traits associated with infection or diseases. The strong correlation of serological status between siblings (not explained by known risk factors), and the familial clustering of KSHV disease<sup>15–19</sup> are all highly suggestive of host genetic influence in addition to shared environmental risk factors. Candidate gene, whole-exome and next-generation sequencing methods have suggested Mendelian causes of rare paediatric KS cases as a result of inborn errors in immunity. However, the role of genetic variants in adult KS and in the viral control in asymptomatic individuals is less clear<sup>20</sup>. To date, most studies provided marginal evidence using statistically lenient *p* value thresholds of between 0.01 and 0.05, which failed to correct for multiple testing to declare “association” for variants in immunomodulatory genes with virus biological function, pathogenesis and tumorigenesis<sup>21–24</sup>. In addition, they had other limitations including, very small sample sizes (between 1 and 350 cases), failure to adjust for environmental factors such as co-infection with other pathogens, or confounding by strong *HLA* associations with HIV and AIDS, and did not stratify controls by KSHV serostatus or include KSHV seronegative controls for comparison, and all lack replication in independent samples. Lastly, only two studies<sup>23,25</sup> have been conducted in African populations.

To overcome the limitations of previous studies and attempt to identify convincing associations with KSHV and EBV immune response traits, we assess systematic differences in >4000 individuals from an African population cohort, where both viruses are endemic, using socio-demographic and clinical data to assess intrinsic and environmental determinants of infection and then perform a GWAS using antibody responses as markers of infection. We use whole-genome sequence data, dense genotyping array data and imputation to a panel with African sequence data to identify genetic loci associated with both infections and attempt to replicate previously identified genetic loci in the context of the environment.

## Results

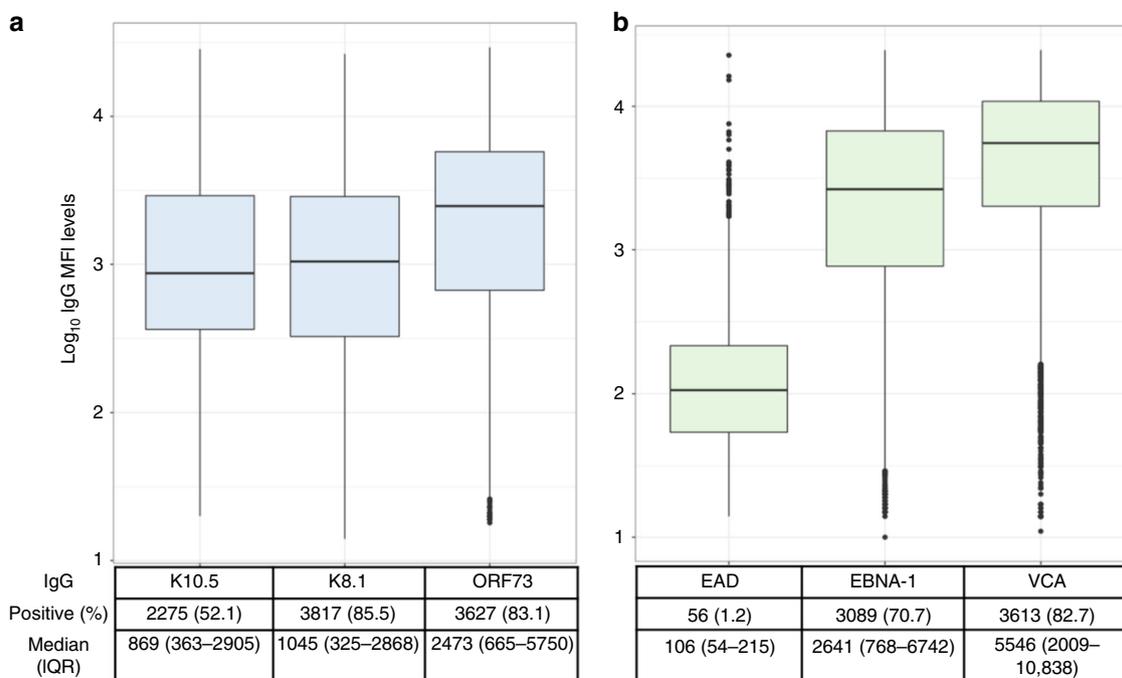
**Characteristics of samples in the Uganda General Population Cohort (GPC).** To investigate the seroprevalence of infections, we

tested serum samples from 4365 individuals in the General Population Cohort (GPC) collected during medical survey round 22 (in 2011). The GPC is a population-based cohort in Kyamulibwa, rural south-west Uganda, comprising inhabitants of 25 neighbouring villages<sup>26</sup>. Participants were over the age of 13 years and belonged mainly (>70%) to the Baganda ethnolinguistic group. Villages were categorised according to urbanicity quartiles reflecting shared ‘urban’ characteristics based on differences in economic activity, civil infrastructure, and availability of educational and healthcare services as previously described<sup>27</sup>, with 28% living in quartile 1 (very rural i.e. no educational facilities and no households with electricity) (Table 1). In this study, 91% of individuals were categorised as seropositive for EBV based on detectable IgG levels against either EBNA-1 or VCA<sup>28</sup> and 91% categorised as seropositive for KSHV based on detectable IgG levels against either ORF73, K10.5 or K8.1<sup>29,30</sup> (Table 1). HIV infection seroprevalence in this study was 6.5%, Hepatitis C virus (HCV) seroprevalence was 3.7% and Hepatitis B virus (HBV) infection had the lowest seroprevalence among the pathogens examined at 2.9% (Table 1). Nearly 95% of participants, 4134 individuals, were infected with at least one of these viruses. The majority of participants, 2743 (63%) were seropositive to at least two of the viruses tested, 23 (0.5%) participants were seropositive for four viruses and 231 (5.3%) participants were seronegative for all viruses (Supplementary Fig. 1). Co-infection with KSHV and EBV was the most common with >95% of dually infected individuals (Supplementary Table 1). Co-infections of other pathogens with EBV or KSHV was similarly frequent and mirrored the seroprevalence estimates seen in the cohort (Supplementary Table 1).

**Table 1 Characteristics of individuals in the GPC.**

Characteristic	N = 4365	(%)
Sex		
Female	2537	58.1
Male	1828	41.9
Age group (year)		
13–18	1078	24.7
19–39	1828	41.9
40+	1459	33.4
Ethnic group		
Baganda	3328	76.2
Other	1037	23.8
Education level attained		
Primary	1538	35.2
Junior secondary	1460	33.4
Senior secondary	905	20.7
Tertiary	223	5.1
Urbanicity quartile		
1	1222	28.0
2	982	22.5
3	953	21.8
4	963	22.1
KSHV		
Positive	3988	91.3
EBV		
Positive	3956	90.6
HIV		
Positive	284	6.5
Hepatitis B		
Positive	126	2.9
Hepatitis C		
Positive	161	3.7

N the number of unique individuals.



**Fig. 1** Inter-individual variability in IgG antibody responses to KSHV and EBV ( $n = 4365$  individuals). **a** Distribution of anti-KSHV antibody levels ( $\text{Log}_{10}$  MFI) for K10.5 (min = 0, max = 28,514), K8.1 (min = 0, max = 26,468) and ORF73 (min = 0, max = 29,308). **b** Distribution of anti-EBV antibody levels ( $\text{Log}_{10}$  MFI) for EAD (min = 0, max = 22,747), EBNA-1 (min = 0, max = 24,696), VCA (min = 11, max = 24,754). Seropositivity to all antigens and median (centre) plus IQR displayed in the text box.

To investigate the inter-individual variation of IgG antibody responses marking different stages of the viral life cycles (latent vs lytic) to KSHV and EBV infections, distributions of antibody levels were determined to the latent (ORF73, K10.5, EBNA-1) and lytic antigens (K8.1, VCA and EAD) in 4365 individuals. All antibody responses were highly variable across individuals as shown by the wide range of median fluorescent intensity (MFI) (Fig. 1a). For KSHV, seropositivity to the K8.1 and ORF73 were both high >80% with similar distributions observed for all anti-KSHV IgG (Fig. 1a). For EBV, seropositivity to EBNA-1 was highest at 83% (Fig. 1b) with the lowest antibody responses observed for EAD (Fig. 1b). The IgG responses were modestly phenotypically correlated ( $r^2$ : 0.1–0.43) (Supplementary Fig. 2).

**Environmental determinants of KSHV IgG antibody response.**

To assess the contribution of environmental determinants on KSHV serostatus (i.e. seropositive vs seronegative) and specific IgG antibody response levels [MFI], we investigated 4365 individuals from the GPC<sup>26</sup>. We selected age, sex, socio-demographic and infection status factors (Table 1) that could potentially influence serostatus or the levels of seroreactivity to KSHV antigens. Following a multivariate logistic regression analysis, being EBV positive had the largest effect on KSHV serostatus (OR = 5.68, 95% CI = 4.55–7.06), and being male (OR = 1.35, 95% CI = 1.17–1.57) was also significantly associated with seropositivity (Fig. 2a). Living in a lower urbanicity quartile, lower educational attainment and HIV negativity were also significantly associated with seropositivity in individuals albeit with modest effect sizes (Fig. 2a and Supplementary Data 1). We further assessed the contribution of the same factors on IgG antibody levels to all antigens using a multivariate linear regression analysis. Similar to serostatus, anti-KSHV IgG levels were significantly higher in EBV seropositive individuals (OR = 2.19, 95% CI = 1.88–2.22) and higher in males than in females, particularly, for anti-ORF73 (OR = 1.17, 95% CI = 0.08–0.148) (Fig. 2b and Supplementary Data 1). HIV seropositivity was significantly

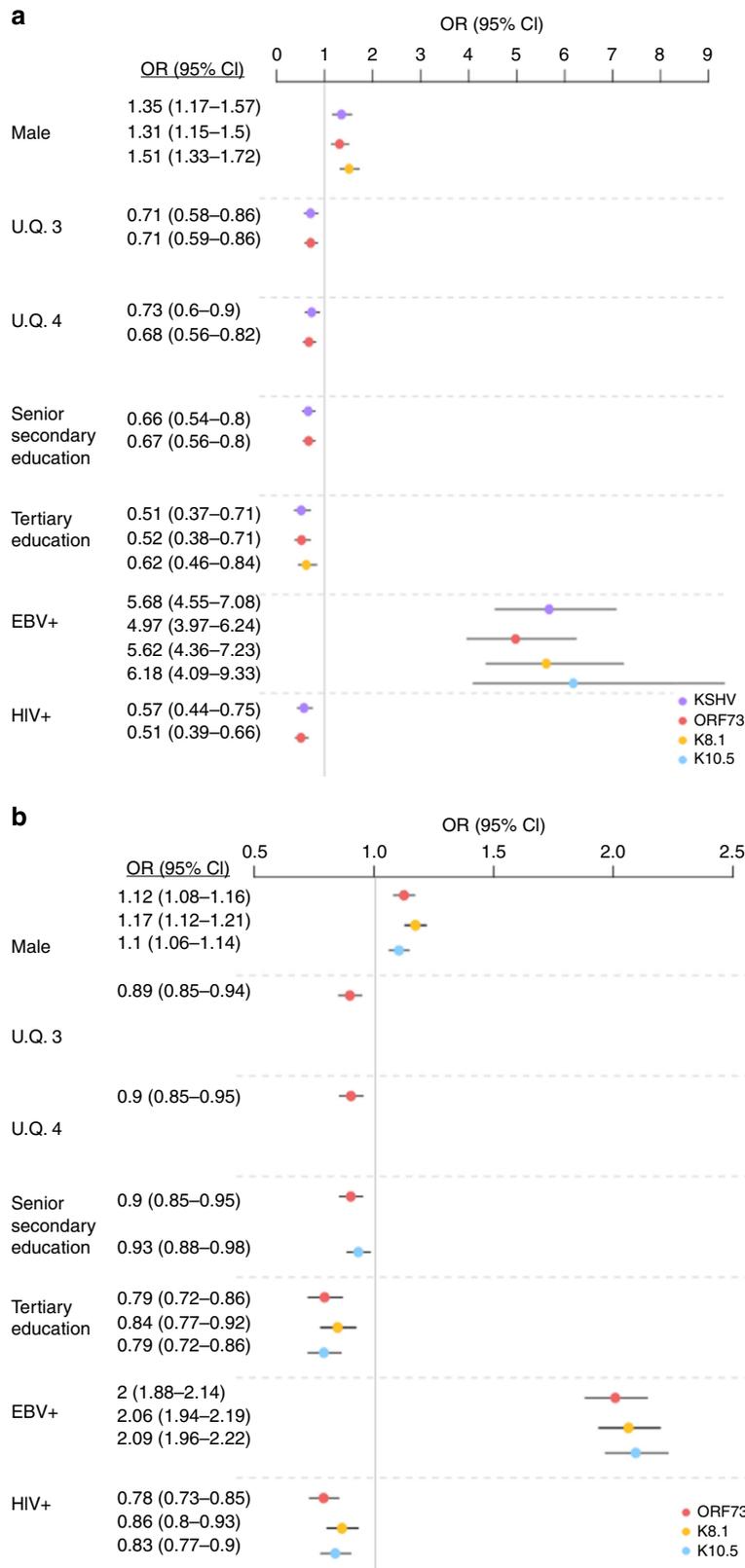
associated with decreased IgG antibody levels (OR = 0.86, 95% CI = 0.73–0.93), as were higher education levels, particularly tertiary education attainment (OR = 0.84, 95% CI = 0.72–0.92) (Fig. 2b). Age, ethnolinguistic group, HBV and HCV co-infections had no significant associations with KSHV serostatus or antibody levels. Full association results can be found in Supplementary Data 1.

**Environmental determinants of EBV IgG antibody response.**

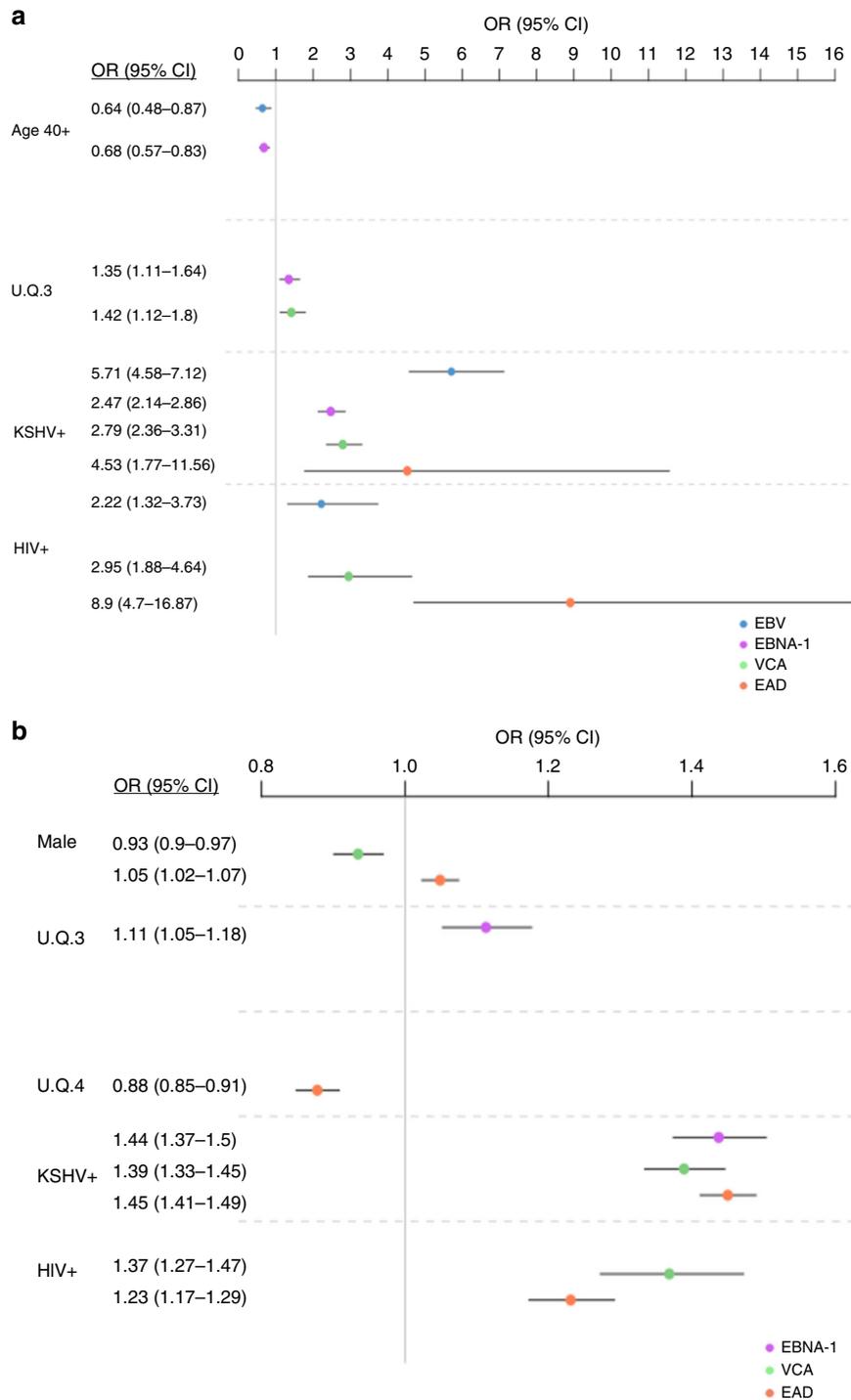
We then assessed the contribution of environmental determinants (Table 1) on EBV serostatus (i.e. seropositive vs seronegative) and the levels of EBV antigen-specific IgG antibodies. KSHV seropositivity (OR = 5.71, 95% CI = 4.58–7.12) had the largest effect on EBV serostatus (Fig. 3a). Living in an urbanicity quartile of 3 (OR = 1.38, 95% CI = 1.01–1.89) and HIV infection (OR = 2.22, 95% CI = 1.32–3.73) were also significantly associated with seropositivity to EBV (Fig. 3a and Supplementary Data 2). Unlike KSHV, sex, or education level were not significantly associated with EBV serostatus (Supplementary Data 2). KSHV seropositivity was also significantly associated with lower IgG antibody levels for all EBV antigens (Fig. 3b and Supplementary Data 2). HIV seropositivity was associated with higher antibody levels to lytic antigens VCA and EAD (OR = 1.37, 95% CI = 1.27–1.47) but not the latent antigen EBNA-1. Living in a higher urbanicity quartile of 3 was also associated with increased antibody levels to EBNA-1 (OR = 1.11, 95% CI = 1.05–1.18) but not the lytic antigens (Fig. 3b and Supplementary Data 2). Anti-EAD was the only antibody that decreased with living in an urbanicity quartile of 4 (OR = 0.88, 95% CI = 0.85–0.91) (Fig. 3b and Supplementary Data 2). Ethnicity, education level and HCV co-infection had no significant influence on IgG levels (Supplementary Data 2).

**Genetic determinants of anti-KSHV and anti-EBV antibody responses.**

To assess the contribution of human genetic variation



**Fig. 2 Predictors of KSHV IgG antibody response.** Odds ratios (ORs) and 95% confidence intervals (CI) of significant associations ( $p < 0.006$ ) between factors and IgG variable as determined based on serologies in the 4365 individuals from the GPC. **a** Predictors of IgG serostatus. ORs and 95% CI were estimated following a multivariate logistic regression analysis with serostatus (seropositive vs seronegative) as response variable. **b** Predictors of IgG response level. ORs and 95% CI were estimated following a multivariate linear regression analysis with  $\log_{10}$ -transformed MFI as response variable. UQ urbanicity quartile. Dots represent the ORs and are coloured by IgG variable: KSHV (seropositivity to any antigen)—purple, ORF73—red, K8.1—yellow, K10.5—light blue; lines represent the 95% confidence intervals. Full results from this analysis are provided in Supplementary Data 1.



**Fig. 3 Predictors of EBV IgG antibody response.** Odds ratios (ORs) and 95% confidence intervals (CI) of significant associations ( $p < 0.006$ ) between factors and IgG variable as determined based on serologies in the 4365 individuals from the GPC. **a** Predictors of IgG serostatus. ORs and 95% CI were estimated following a multivariable logistic regression analysis with serostatus (seropositive vs seronegative) as response variable. **b** Predictors of IgG response level. ORs and 95% CI were estimated following a multivariable linear regression analysis with  $\log_{10}$ -transformed MFI levels as response variable. UQ urbanicity quartile. Dots represent the ORs and are coloured by IgG variable (EBV (seropositivity to EBNA-1 or VCA)—blue, EBNA-1—magenta, VCA—light green, EAD—orange), lines represent the 95% confidence intervals. Full results from this analysis are provided in Supplementary Data 2.

on antibody responses to KSHV ( $N = 4365$ ) and EBV ( $N = 4365$ ), we first explored the heritability of anti-KSHV and anti-EBV IgG serological traits in the GPC. As most individuals in the cohort are seropositive ( $>90\%$ ) presumably from childhood<sup>30</sup>, those who present as seronegative are also likely to have been infected and thus, produced antibodies at very low/undetectable levels; as such, they are also informative and useful to include in

analysis. In this study, heritability estimates were high for anti-ORF73 ( $h^2 = 33\%$ , 95% CI = 24–42%), anti-K8.1 ( $h^2 = 25\%$ , 95% CI = 16–35%) and anti-K10.5 ( $h^2 = 32\%$ , 95% CI = 23–41%), and for anti-EBNA-1 ( $h^2 = 36\%$ , 95% CI = 28–45%) and anti-VCA IgG ( $h^2 = 32\%$ , 95% CI = 23–41%). After correction for unmeasured shared environmental correlation, accounted for by GPS coordinates<sup>11,31</sup>, our estimates were reduced yet still

**Table 2** Lead association signals with anti-KSHV IgG antibody levels.

Trait	Chr:Pos(b37)	Variant	Mapped gene(s) <sup>a</sup>	Consequence <sup>a</sup>	EA	EAf (%)	<i>p</i>	$\beta$ (95% CI)
ORF73	5:37832167	rs114429578	<i>GDNF</i>	Intergenic	T	3.1	$4.60 \times 10^{-08}$	0.36 (0.23–0.49)
	6:32584693	rs510205	<i>HLA-DQA1</i>	Intergenic	G	9.4	$5.53 \times 10^{-08}$	0.21 (0.14–0.29)
K10.5	6:31262619	rs111664408	<i>HLA-B/ HLA-C</i>	Intergenic	G	10.6	$6.64 \times 10^{-09}$	−0.22 (−0.29 to −0.15)
	10:124833806	rs10794590	<i>ACADSB</i>	Intergenic	T	25.7	$4.59 \times 10^{-08}$	−0.14 (−0.19 to −0.09)

EA effect allele, EAF effect allele frequency.

<sup>a</sup>Annotated using Ensembl VEP.

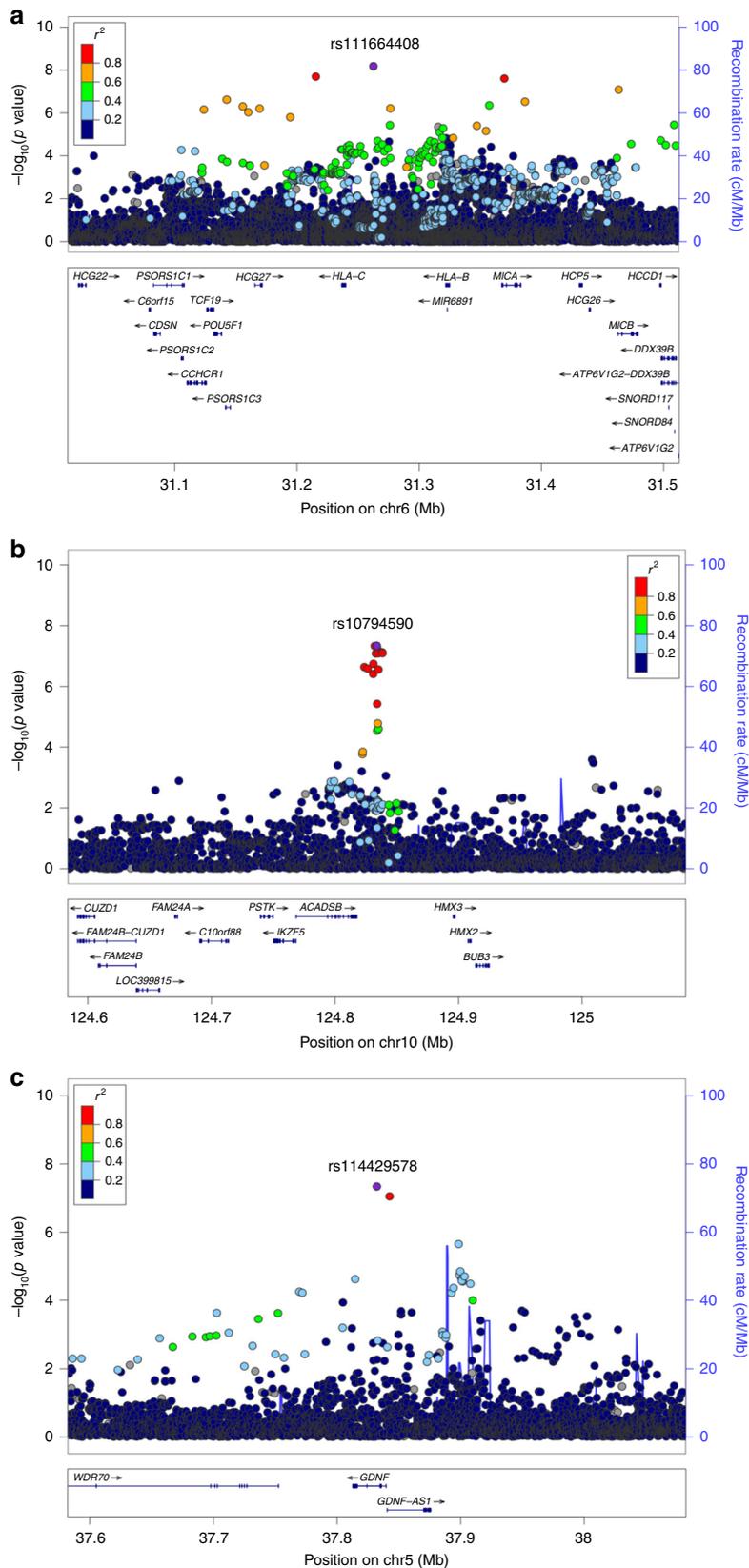
significant for all traits as previously found in our pilot study<sup>11</sup>, anti-ORF73 ( $h^2_{\text{Adjusted}} = 20\%$ , 95% CI = 13–26%), anti-K8.1 ( $h^2_{\text{Adjusted}} = 13\%$ , 95% CI = 7–19%) and anti-K10.5 ( $h^2_{\text{Adjusted}} = 18\%$ , 95% CI = 11–24%), and for anti-EBNA-1 ( $h^2_{\text{Adjusted}} = 20\%$ , 95% CI = 14–27%) and anti-VCA IgG ( $h^2_{\text{Adjusted}} = 18\%$ , 95% CI = 13–25%).

To further investigate the genetic determinants of response to infection in this population, we conducted GWAS for each continuous antibody trait combining whole-genome sequencing and dense genotyping data with imputation to a merged 1000 Genomes phase 3, AGVP<sup>32</sup> and UG2G reference panel<sup>33</sup>. Stringent sample and variant QC left 4365 individuals with all phenotypes and ~17 M SNPs across autosomal markers and X-Chromosome for analyses (see ‘Methods’). Homogeneity in the study population was previously ascertained by principal components analyses with AGVP populations as a reference panel<sup>11</sup>. This model accounted well for population structure and cryptic relatedness as shown by genomic inflation factor ( $\lambda$ ) ~1.0 for all traits (Supplementary Fig. 3). Correcting for multiple testing, accounting for the lower linkage disequilibrium (LD) in African populations and using an FDR of 5%, the genome-wide significance threshold was adjusted to  $p < 1 \times 10^{-8}$ <sup>11,33</sup>. Following GWAS of quantitative anti-ORF73, anti-K8.1 and anti-K10.5 IgG levels for 4365 individuals, one locus in association with K10.5, rs111664408-G ( $p = 6.64 \times 10^{-09}$ ,  $\beta$ , 95% CI = −0.2, −0.29 to −0.15) on chromosome 6, mapped as intergenic between *HLA-B* and *HLA-C*, reached genome-wide significance threshold (Table 2 and Fig. 4a). In addition, the following variants reached the standard genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ) used in European studies which we highlight as they merit being prioritised for replication in other studies; for K10.5, rs10794590-T ( $p = 4.59 \times 10^{-08}$ ,  $\beta$ , 95% CI = −0.14, −0.19 to −0.09), an intergenic variant on chromosome 10 (Table 2 and Fig. 4b) and for ORF73, rs114429578-T ( $p = 4.60 \times 10^{-08}$ ,  $\beta$ , 95% CI = 0.36, 0.23–0.49) on chromosome 5 mapped in *GDNF* (Table 2 and Fig. 4c). Another plausible signal rs510205-G ( $p = 5.53 \times 10^{-08}$ ,  $\beta$ , 95% CI = 0.21, 0.14–0.29) on chromosome 6 mapped in *HLA-DQA1* (Table 2) was also associated with anti-ORF73 IgG. In the GTEx database, rs10794590 (associated with K10.5 IgG levels) was associated with the differential expression of *ACADSB* ( $p = 2.6 \times 10^{-9}$ ) and *IKZF5* ( $p = 1.5 \times 10^{-3}$ ) in oesophagus mucosa and colon tissues. We also found rs510205 (associated with ORF73 IgG) associated with the differential expression of the following eight genes, *HLA-DQA2*, *HLA-DQB2*, *HLA-DQB1*, *HLA-DRB1*, *HLA-DRB6*, *HLA-DQA1*, *CYP21A1P* and *LY6G5B* in whole blood ( $p < 3.8 \times 10^{-6}$ ). Expression data for our other association signals were unavailable in the GTEx database. We were unable to refine signals linked to *HLA* alleles with no significant associations identified with ORF73 or K10.5 IgG levels (Supplementary Data 3). We also assessed whether variants within 23 genetic loci that been previously identified in other studies with marginal significance as associated with classic KS, KSHV seropositivity, KSHV viral load, antibody response or primary effusion lymphoma<sup>21,23–25,34–36</sup> were present in this

study and had plausible signals. In this study, 21 out of 31 variants were typed/imputed and were in *IL6*, *IL8RB*, *IL13*, *IL4*, *IL12A*, *FCYRIIIA*, *HLA* and *IRAK1*<sup>21,23,24,34–36</sup>, data for KIR types<sup>22</sup> were unavailable in this study. Despite adequate power (80%), we did not detect significant associations ( $p > 0.1$ ) of any of the previous suggested variants for any of the traits in this study (Supplementary Table 2).

For EBV infection, GWAS of 3289 individuals, replicated our previous finding<sup>11</sup> rs9272371 in *HLA-DQA1* ( $p = 5.24 \times 10^{-33}$ ,  $\beta = -0.37$ ) and identified a distinct secondary signal, rs3129867 in *HLA-DRA* ( $p_{\text{cond}} = 6.01 \times 10^{-11}$ ,  $\beta = -0.17$ ) (Supplementary Table 3). To maximise power, we combined our dataset with 1076 individuals from our pilot study<sup>11</sup> and performed GWAS for EBV response in 4365 individuals (Supplementary Fig. 5). We found stronger statistical associations for anti-EBNA-1 IgG antibody responses in the *HLA* class II region (Table 3 and Fig. 5). The C-allele at our lead SNP rs9272371 in *HLA-DQA1* ( $p = 3.63 \times 10^{-44}$ ,  $\beta = -0.37$ ) was associated with lower antibody levels (Fig. 5a and Table 3), with the presence of a distinct secondary signal rs3129867 in *HLA-DRA* (Fig. 5b and Table 3) with the G allele associated with lower antibody levels ( $p_{\text{cond}} = 2.78 \times 10^{-11}$ ,  $\beta = -0.17$ ). We also identified significant associations for anti-VCA IgG antibody levels also in the *HLA*-class II region (Fig. 5c and Supplementary Fig. 5). The C-allele of the lead SNP rs71542439, also in *HLA-DQA1*, associated with lower antibody levels ( $p = 1.15 \times 10^{-12}$ ,  $\beta = -0.21$ ) (Table 3). To fine map the associations observed in the *HLA* region, we tested the presence or absence of 117 four-digit imputed classical *HLA* alleles. *HLA-DRB1*\*15:01 showed the lowest *p* value for association with EBV EBNA-1 ( $p = 4 \times 10^{-4}$ ) (Supplementary Data 4) and had very weak LD with other alleles ( $r^2 < 0.02$ ) (Supplementary Data 5). We did not replicate *HLA-DRB1*\*0701 ( $p = 0.86$ ) or *HLA-DQB1*\*0301 ( $p = 0.84$ ) alleles (Supplementary Data 4) previously associated with EBNA-1 IgG in individuals of European descent<sup>13,14</sup>. We also did not identify any significant associations between any classical *HLA* alleles and VCA IgG levels (Supplementary Data 4).

**Trans-ancestry meta-analysis and fine mapping of anti-EBNA-1 IgG response.** To investigate the portability of findings across ancestry, we used MANTRA<sup>37</sup> to perform a genome-wide trans-ancestry meta-analysis for anti-EBNA IgG responses, with association summary statistics of 4365 individuals from our Ugandan GWAS combined with 2162 seropositive individuals from the 1000 Genomes-imputed European ancestry GWAS<sup>13</sup>, giving a total of 6527 individuals with ~4.9 million shared SNPs for analysis. Using a threshold of  $\log_{10}$  Bayes Factor (BF)  $> 6$ <sup>38</sup> we replicated strong evidence of association in the *HLA* class II region with lead SNP rs6927022 ( $\log_{10}\text{BF} = 44.3$ ) (Supplementary Fig. 6) previously identified as the lead association SNP in the European ancestry study<sup>13</sup>. Our Ugandan lead SNP rs9272371 ( $\log_{10}\text{BF} = 43.2$ ) displayed heterogeneity in effect sizes in the two studies ( $p_Q = 3.56 \times 10^{-8}$ ) as identified in our pilot study. To



**Fig. 4 Regional association plots for lead SNPs associated with anti-KSHV IgG levels  $p < 5 \times 10^{-8}$ ,  $n = 4365$  individuals. a** Association on chromosome 6 in *HLA* class I region with anti-K10.5 IgG (rs111664408,  $p = 6.64 \times 10^{-9}$ ). **b** Association on chromosome 10 in the *ACADSB* region with anti-K10.5 IgG (rs10794590,  $p = 4.59 \times 10^{-8}$ ). **c** Association on chromosome 5 in *GDNF* region with anti-ORF73 IgG (rs114429578,  $p = 4.60 \times 10^{-8}$ ). GWAS performed using linear mixed model accounting for kinship in GEMMA. The lead SNPs are labelled and coloured in purple. LD ( $r^2$ ) was calculated based on Ugandan SNP genotypes.

**Table 3 Loci with significant evidence of association with anti-EBV IgG levels (N = 4365).**

Trait	Chr:Pos(b37)	Variant	Nearest gene <sup>a</sup>	Consequence <sup>a</sup>	EA	EAF (%)	p	β (95% CI)
EBNA-1	6:32604654	rs9272371	<i>HLA-DQA1</i>	Intron	C	26.9	$3.63 \times 10^{-44}$	-0.37 (-0.43 to -0.31)
EBNA-1	6:32404220	rs3129867	<i>HLA-DRA</i>	Upstream	G	49.2	$2.10 \times 10^{-24*}$	-0.25 (-0.30 to -0.20)
VCA	6: 32630257	rs71542439	<i>HLA-DQB1</i>	Intron	T	21.1	$1.15 \times 10^{-12}$	-0.21 (-0.27 to -0.15)

EA effect allele, EAF effect allele frequency.

\* $p_{\text{cond}}$  on lead SNP, β (95% CI) =  $2.78 \times 10^{-11}$ , -0.15 (-0.18 to -0.11).<sup>a</sup>Annotated using Ensembl VEP.

further, refine genetic association signals, using MANTRA results we generated 99% credible sets most likely to drive association signals and contain causal variants (or tagging unobserved causal variants) and compared fine-mapping intervals for each associated lead SNP by analysing the variants 500 kb up- and downstream of the lead SNP in the combined dataset as described previously<sup>39,40</sup>. This resulted in three distinct SNPs in the credible set, rs6927022, rs9272371 and rs9274247 (Supplementary Table 4) further confirming that rs6927022 does not fully drive associations in the Ugandan population. No other locus was found to be in association with anti-EBNA-1 IgG response.

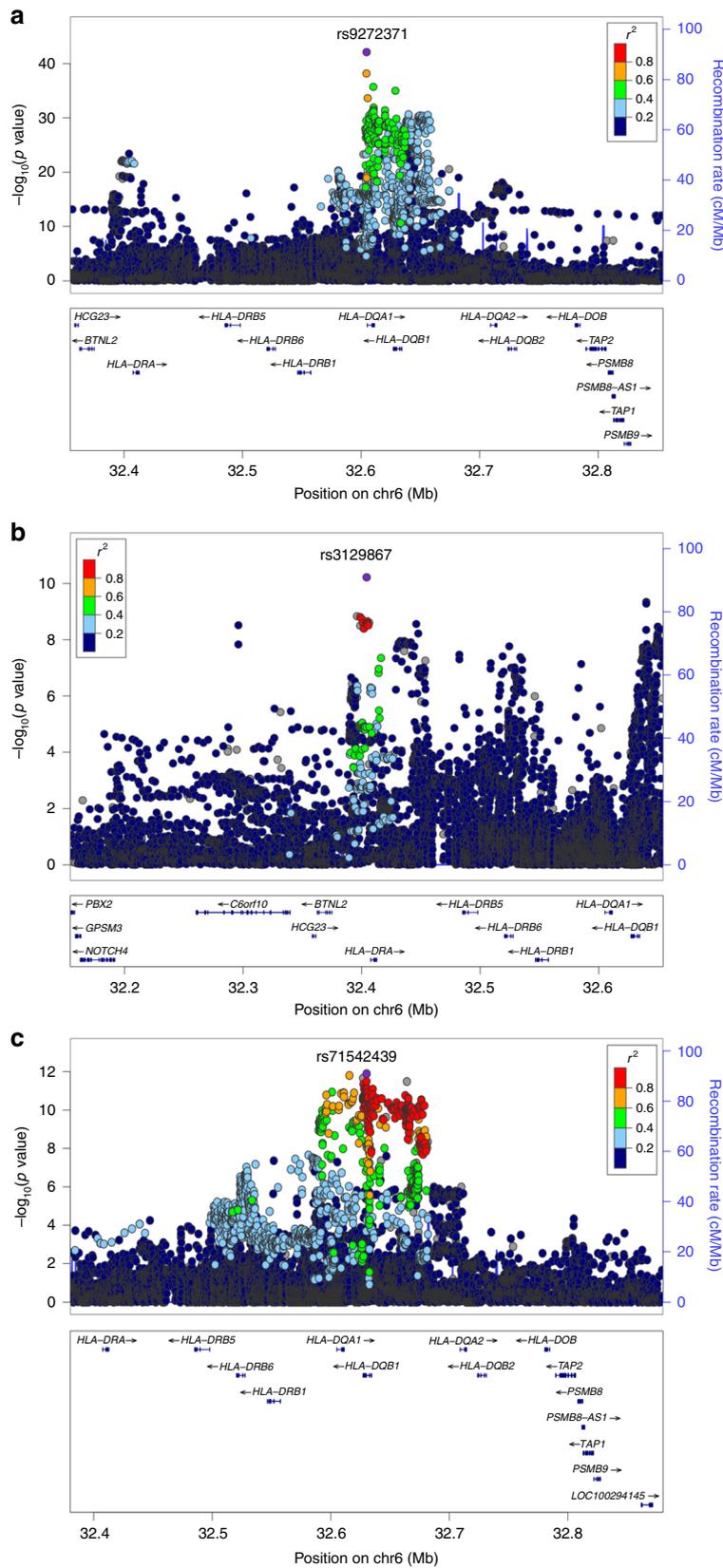
## Discussion

In this study, we assessed the seroprevalence of KSHV and EBV infections along with the burden of co-infections and the influence of intrinsic, environmental and genetic factors on IgG antibody responses to infection in a rural African population cohort of >4000 individuals. Serological assays are useful in assisting diagnosis, and in improving our understanding of prevalence and transmission of infection, in addition to understanding virology and host immunity. The presence of antibodies representing host immune response are commonly used as diagnostic markers for stage of infection and thus, seroprevalence is widely used as a measure of the frequency of infections in a population<sup>41</sup>. In the GPC, prevalence estimates of KSHV and EBV are high (>90%) and consistent with previous findings in Uganda<sup>42–46</sup>. For KSHV, we found increased antibody seroprevalence and IgG levels in males compared to females as previously reported<sup>30</sup>. For EBV, while higher age (>40 years) was associated with modest decrease in antibody response (OR = 0.64), the background antigenic milieu in this population is different to western populations; it is likely that while antibody levels increase in childhood, they could decay with time (in adulthood) as a result of immunomodulation by other factors as observed in other studies<sup>47</sup> or as a result of exposure to other pathogens. Environmental factors such as co-infection with other pathogens have been studied and found to influence variation in seroprevalence estimates, and antibody responses associated with infection<sup>45,48</sup>. This is consistent with the results presented in this study, where we found HIV infection to be associated with IgG antibody response levels to KSHV and EBV infection. While HIV infection was associated with lower antibody responses to KSHV, the inverse was observed for EBV antibodies. Similarly, lower urbanicity and educational attainment were associated with a decrease in antibody levels to KSHV and increase to EBV (Figs. 2 and 3). This could be due to differential effect of shared epidemiological risk factors for HIV, EBV and KSHV, and/or as a result of the mutual inhibition of KSHV and EBV that has been observed in other studies<sup>5,6</sup>. Here, we also found that KSHV and EBV co-infections influence antibody responses to each other, highlighting the direct or indirect interaction between the two viruses, which has been previously observed for viral shedding among Ugandan children and their mothers and also in a Cameroon KS case control study<sup>28,48</sup>. As Uganda is also a malaria

endemic area and studies have reported co-infection with *P. falciparum* and other parasites as having an influence on antibody titre<sup>45,46,49</sup>, it would be useful to also incorporate this in future study design. While the cohort is indeed rural, a marked variation in levels of urbanicity across the villages, largely attributable to differences in economic activity, civil infrastructure, and availability of educational and healthcare services has been previously described<sup>27</sup>. It is interesting that even within a rural setting different levels of ‘urbanicity’ are associated with antibodies supporting broader observations contrasting higher seroprevalence in rural vs urban settings<sup>46,48,50</sup>.

In addition to environmental factors described above, host genetic factors (which might also be influenced by the environment) can also contribute to phenotypic variation in infectious disease response traits. To explore the proportion of variance in antibody responses attributable to the host genetics, we investigated the heritable component of the IgG response to KSHV and EBV antigens in individuals with genotype data, adjusting for unmeasured environmental correlation using spatial distances and similar methodology used in previous studies<sup>51</sup>. These analyses show that while KSHV and EBV antibody response traits were heritable there is overestimation in heritability prior to accounting for shared environment, which is consistent with our previous findings in Uganda<sup>11</sup>. The heritability estimates for KSHV ( $h^2 = 13–19\%$ ) are also lower than the reported estimate of 37% in a Mexican-American ancestry cohort<sup>52</sup>; this deflation in estimates is similar to what we had previously observed for our pilot study of EBV in Uganda ( $h^2_{\text{EBNA-1}} = 11\%$ ) compared to Mexican-Americans or Europeans ( $h^2_{\text{EBNA-1}} = 37–43\%$ )<sup>11,12,53,54</sup>. These differences in the GPC compared to other populations could be attributed to a number of factors including differences in study/assay design, locus or allelic heterogeneity influencing traits, not fully adjusting for environmental effects and differences in gene–environment interactions.

Conducting GWASs for KSHV-associated diseases is challenging as sample sizes are limiting to achieve power for reliable detection of loci. Prior to this study, GWAS had not been performed for any KSHV phenotype. As previous studies have shown correlation of IgG levels with the development of disease<sup>55,56</sup>, here we used IgG antibody responses as a proxy for disease progression. While anti-KSHV IgG traits are partly heritable after accounting for shared environment, despite having >80% power to detect signals of moderate to large effect sizes ( $\beta > 0.15$ , for MAFs >25%) for all traits, only one genome-wide significant locus ( $p < 1 \times 10^{-8}$ ) was identified, the variant rs111664408 in *HLA-B/HLA-C* ( $p = 6.64 \times 10^{-09}$ ) on chromosome 6 contributing to inter-individual variability in K10.5 IgG responses. Nonetheless, using the standard genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ), two candidate loci associated with anti-ORF73 and anti-K10.5 IgG response levels were identified, reflecting the latent stage/history of infection. For anti-ORF73 IgG, relatively strong associations were in *GDNF* on chromosome 5 (rs114429578,  $p = 4.60 \times 10^{-08}$ ) and in *HLA-DQA1* (rs510205,  $p = 5.53 \times 10^{-08}$ ) (Table 2 and Fig. 4). While



**Fig. 5 Regional association plots for lead SNPs associated with anti-EBV IgG levels ( $p < 1 \times 10^{-8}$ ),  $n = 4365$  individuals. **a** Lead association on chromosome 6 in *HLA-DQA1* with anti-EBNA-1 IgG (rs9272371,  $p = 3.60 \times 10^{-44}$ ). **b** Secondary association on chromosome 6 in *HLA-DRA* with anti-EBNA-1 IgG (rs3129867,  $p_{\text{cond}} = 2.78 \times 10^{-11}$ ). **c** Lead association on chromosome 6 in the *HLA-DQB1* with anti-VCA IgG (rs71542439,  $p = 1.15 \times 10^{-12}$ ). GWAS performed using linear mixed model accounting for kinship in GEMMA. The lead SNPs are labelled and coloured in purple. LD ( $r^2$ ) was calculated based on SNP genotypes in PLINK.**

*GDNF* (encoding glial cell line-derived neurotrophic factor) has not been previously implicated in KSHV pathogenesis, it has been reported to be critical in the maintenance of latency for Herpes-simplex viruses infections<sup>57</sup>. The HLA class I and II regions, however, have been reported to play a role in the pathogenesis of KSHV and KS; nevertheless, previous genetic association studies have failed to identify convincing and reproducible associations at these loci<sup>23,25</sup>. Activation of CD4+ T cells is particularly important for anti-KSHV immunity and *in vitro* studies have shown the CD4+ T cells can inhibit viral replication in KSHV-infected tonsillar B cells<sup>58,59</sup>. Like EBV, KSHV has evolved strategies to evade immune detection, including negatively regulating the process by which HLA class II molecules present antigens to CD4+ T cells, thereby promoting its survival. Recently, ORF73 and v-IRF3 (K10.5) have been reported to inhibit MHC class II peptide presentation by blocking the transcription of the class II transactivator (CIITA), a master regulator of class II expression<sup>60–62</sup>. These findings suggest that variation in genes that play a role in modulating the immune response, particularly, T-cell immunity could contribute to inter-individual differences in anti-ORF73 and anti-K10.5 IgG levels to control KSHV viral infection, and thus, merit prioritisation in further replication studies in other populations.

For EBV infection, we and others have previously confirmed the presence of variants in the *HLA* Class II region driving associations for EBV EBNA-1 antibody response in different populations<sup>11–14</sup>; here we replicated these findings in *HLA-DQA1* (rs9272371,  $p = 3.63 \times 10^{-44}$ ) (Fig. 5a) and identified an additional distinct variant (rs3129867,  $p_{\text{cond}} = 2.78 \times 10^{-11}$ ) in *HLA-DRA* (Fig. 5b) contributing to IgG levels in Uganda that was not identified as distinct in our initial pilot study<sup>11</sup>. We also identified an association at rs28394498 ( $p = 1.15 \times 10^{-12}$ ) in *HLA-DQA1* with anti-VCA IgG levels (Fig. 5c). We attempted to fine map the *HLA* region; *HLA-DRB1\*1501* present in 1.1% of individuals in this study was the most significant association signal ( $p = 4.7 \times 10^{-4}$ ,  $\beta = 0.42$ ) for EBNA-1 IgG. Interestingly, however, it did not explain our lead GWAS signal in *HLA-DQA1*. We also did not replicate *HLA* alleles *DRB1\*0701* and *DQB1\*0301* previously associated with EBNA-1 IgG levels in two GWASs consisting individuals of European descent<sup>13,14</sup>. Both studies, however, retained the significance of their lead GWAS SNPs in the *HLA* region after adjusting for the effect of the lead *HLA* alleles (residual  $p < 2.6 \times 10^{-11}$ )<sup>13,14</sup>, and thus, could not fully explain the lead GWAS signal, which is consistent with our findings. This suggests the relevance of additional alleles or a non-*HLA* effect that did not reach significance or is absent in our study. It is also likely that allelic/haplotypic heterogeneity in Africans vs Europeans, an expression modulating effect or *HLA* imputation errors play a role particularly, given the limited representation of Africans in the multi-ethnic reference panel<sup>63</sup>. Trans-ancestry meta-analysis of European and African individuals in the *HLA* class II region confirmed distinct association signals in the two populations. As with our previous findings, rs9272371 and rs6927022 are likely to be distinct variants in the *HLA* class II region, with a single signal in Europeans (rs6927022) and a signal mostly driven by rs9272371 in Uganda (Supplementary Table 4). Thus, future studies should aim to further refine signals in larger sample sizes in individuals of African descent and using *HLA* panels based on African data which are currently in development.

In summary, we characterise systematic differences in >4000 individuals with evidence of infection by KSHV and EBV in Uganda and identify determinants of response to infection. The availability of data on environmental covariates such as co-infection with other pathogens allowed us to capture genetic variation independently of the environment. Co-infection was indeed a strong predictor of inter-individual variability in

antibody responses and highlighted significant and complex interactions between EBV and KSHV. Despite similar heritability and power for KSHV and EBV traits, it was interesting to observe less genome-wide statistically significant associations for the former; this suggests differences in underlying genetic architectures between the traits. It is possible that variants with very small effect sizes or low-frequency variants (i.e. EAF <0.5%) influence inter-individual variability in KSHV immune responses; however, this study is under-powered to detect low-frequency variants with small effect sizes and neither the genetic data nor the tools used here are optimal for accurate rare variant detection. As sub-Saharan Africa sustains such high transmission compared to the rest of the world, these insights are useful to better understand fundamental biological processes of gamma-herpesvirus infection and their epidemiological characteristics to inform appropriate therapeutic interventions and to control their transmission. Lastly, given there are no equivalent datasets from African populations, summary statistics derived from this study are a useful resource for others to use.

## Methods

**Ethics approval and consent.** Informed consent was obtained from all participants either in conjunction with parental/guardian consent for under 18-year-olds with signature, or a thumb print if the individual was unable to write. The study was approved by the Uganda Virus Research Institute, Research Ethics committee (UVRI-REC) (Ref. GC/127/10/10/25), the Uganda National Council for Science and Technology (UNCST) (Ref. HS 870), and the UK National Research Ethics Service, Research Ethics Committee (UK NRES REC) (Ref. 11/H0305/5).

**Sample selection and collection.** The GPC is a population-based cohort in Kyamulibwa, rural south-west Uganda, consisting of 25 neighbouring villages inhabited mainly by farmers who grow subsistence crop, cultivate coffee for trade and also raise livestock<sup>26,64</sup>. Households are scattered with some concentrated in the trading centres. Villages were categorised by urbanicity quartiles to reflect shared socio-demographic and ‘urban’ characteristics based on seven components: population size, economic activity, built environment, communication services, educational facilities, health services, and diversity as previously described elsewhere<sup>27</sup>. Blood samples from 7000 GPC study participants, representing 11 self-reported ethnolinguistic groups, were collected during medical survey sampling conducted in the study area between 2010 and 2011<sup>26</sup>. Details of sexual behaviour, medical, socio-demographic and geographic factors were also recorded. Serum was tested for HIV-1, Hepatitis B (HBV) and Hepatitis C (HCV) infections and the remainder was stored at  $-80^{\circ}\text{C}$  in freezers in Entebbe prior to further serological testing.

**Serology and quality control of phenotypic data.** We quantified MFI to measure the levels of serum IgG antibodies to the KSHV encoded antigens ORF73, K8.1 and K10.5; and to the EBV encoded EBNA-1 VCA and EAD covalently bound to fluorescent magnetic beads and read on a multiplexed platform<sup>28</sup>. Pairwise-correlation between phenotypes was tested for using Pearson’s correlation in R. For each virus, seropositivity was defined as having antibody level to any antigen tested above a prespecified cut-off, as previously described<sup>29</sup>. Of the original ~7000 people genotyped, we were able to link phenotype results from 4365 people (mean age  $\pm$  SD =  $34 \pm 18$  years, 54% female) from samples collected from the GPC in 2011.

**Statistical analysis of quantitative antibody levels.** To investigate factors influencing IgG antibody response to KSHV and EBV infections, categorical outcome variables for antibody serostatus (i.e. positive/negative)<sup>29</sup> and continuous variables for antibody levels ( $\log_{10}$  transformed MFI) were generated and analysed with potential risk factors. The following variables were considered to be possible risk factors or confounders: sex, age (years) or age group (categorised as 13–18, 19–35 or  $\geq 40$  years), ethnic group (Baganda or other), education level attained (none/primary, junior secondary, senior secondary or tertiary), village urbanicity quartile (1–4 with 4 being most urban) based on shared socio-demographic or ‘urban’ characteristics as previously described<sup>27</sup>. HIV infection, HBV infection, HCV infection, and anti-KSHV serostatus and  $\log_{10}$  antibody levels (anti-K8.1, anti-ORF73 and anti-K10.5), or anti-EBV antibody serostatus and antibody levels (anti-EBNA-1, anti-VCA and anti-EAD  $\log_{10}$  MFI). Multivariate logistic or linear regression models were fitted to examine variables predictive of antibody levels. All variables in the model were independent of each other except for urbanicity quartile and education ( $\chi^2$ ,  $p < 0.001$ ) and age and education ( $\chi^2$ ,  $p < 0.001$ ); nonetheless, we do not identify significant collinearity in our models as variance inflation factors were all close to 1. A multiple testing  $p$  value of  $< 0.006$  was used to determine statistical significance.

**Genetic data curation and quality control.** Of the 7000 GPC samples, 5000 were densely genotyped on the Illumina HumanOmni 2.5M BeadChip array and we then imputed additional variants into the genotype chip dataset using a merged 1000 Genomes phase 3<sup>65</sup>, African genome variation project<sup>32</sup> and UG2G (Uganda 2000 Genomes) reference panel in IMPUTE2<sup>66</sup>. Whole-genome sequencing was performed on 2000 samples with 100 base paired end sequencing at 4× coverage on the Illumina HiSeq 2000 platform following the manufacturer's protocol. Stringent variant and sample quality control (QC) filtering was performed. Low-quality variants that mapped to multiple regions within the human genome or did not map to any region, and duplicate variants genotyped on the chip were removed. We excluded samples with a call rate <97% and heterozygosity >3 SD from the mean, discordant genetic sex and reported sex, and sites deviating from Hardy–Weinberg equilibrium ( $p < 10^{-8}$ ). Following imputation, we only included high-quality sites (info score > 0.3 and  $r^2 > 0.6$ ) with minor allele frequency (MAF)  $\geq 0.5\%$ . We also removed samples without matching phenotype and genotype or sequence data. Of the merged datasets, 343 samples had overlapping genotype and sequence variant calls for which a final concordance of 98% was achieved for all SNPs. The merged datasets post QC filtering resulted in 4365 samples with KSHV and EBV phenotypes and ~17 M SNPs across the autosomes and X-chromosome for analyses. For EBV analyses we first performed primary GWAS of 3289 unique individuals and then in 4365 individuals i.e. in combination with 1076 overlapping samples present in our pilot study<sup>11</sup>.

**HLA region imputation.** Human leucocyte antigen alleles were imputed using genotype data available from across the extended MHC region only (chromosome 6 base pair positions 25,500,000–34,000,000 in genome build 37) and SNP strandedness was checked with the HLA\*IMP front-end software<sup>67</sup>. HLA genotypes of 117 alleles (AF > 1%), at 11 loci were imputed with HLA\*IMP:02<sup>63</sup> using the standard HLA\*IMP:02 multi-ethnic reference graphs, representing ~800 reference individuals for HLA-DRB3/4/5 and >6000 reference individuals for all other loci.

**Heritability of antibody response traits in the GPC.** Narrow-sense heritability ( $h^2$ ) for anti-EBNA-1 IgG, anti-VCA IgG, anti-ORF73 IgG, anti-K8.1 IgG and anti-K10.5 IgG traits were estimated using a linear mixed model (LMM) in FaST-LMM with two random effects, one based on genetic effects and the other on environmental effects using spatial location<sup>31</sup> recorded as Global Position System (GPS) coordinates as a proxy for environmental effects.

**Genome-wide association analyses.** The statistical power to identify genetic variants of genome-wide significance and with different effect sizes given the sample size was estimated using QUANTO software (<http://biostats.usc.edu/software>). We conducted analyses for all quantitative antibody traits for KSHV ( $N = 4365$ ) and EBV ( $N = 3289$  and  $N = 4365$  (including overlapping samples from pilot study) by first applying a multivariable linear regression model adjusting for significant covariates in R. Residuals of MFI values used for analyses were then transformed using inverse, rank-based normalisation in R to ensure a standard normal distribution for the phenotypes and confirmed by visualisation. To control for cryptic relatedness and population structure within the GPC, the GWAS of transformed phenotypes was performed using the standard mixed model approach in GEMMA<sup>68</sup>. A kinship matrix to define pairwise genetic relatedness among individuals was generated using pooled imputed genotypes and sequence data for all autosomes and X-chromosome using the  $k = 1$  option in GEMMA. The data were LD-pruned ( $r^2 = 0.2$ ) using dosages and an MAF threshold of 1% was applied. Environmental correlation using GPS data as a proxy, genotyping or sequencing method were also adjusted for as additional covariates during analysis in GEMMA. To identify distinct SNPs, conditional analysis was performed in GEMMA. Each SNP within 1 MB of the lead association SNP was conditioned. If any SNP was statistically significant, it was added stepwise onto the mixed model and analysed jointly; this was done until no SNPs with  $p < 5 \times 10^{-9}$ <sup>11,33</sup> remained. All SNPs remaining statistically significant were considered distinct association signals. To functionally annotate our most significant associations, we used the Ensembl Variant Effect Predictor (VEP)<sup>69</sup> and the gene/tissue expression database (GTEx: <https://gtexportal.org>) V8 release to access data on expression quantitative trait loci (eQTLs) from tissues. To test associations between imputed HLA allele and amino acid dosages and phenotypes, we performed linear regression of transformed phenotypes in GEMMA as performed in the SNP-based GWAS for that phenotype.

**Trans-ancestry meta-analysis and fine mapping.** MANTRA was used to perform a genome-wide trans-ethnic meta-analysis for anti-EBNA IgG responses with association summary statistics of 4365 individuals from our Ugandan GWAS combined with publicly available data of 1000 Genomes-imputed European ancestry GWAS from 2162 seropositive individuals, giving a total of 6527 individuals with ~4.6 million shared SNPs for analysis. The MANTRA approach leverages differences in LD structures across populations to account for differences in genetic architecture and accommodates heterogeneity of allelic effects between distantly related populations within a Bayesian partition framework<sup>37</sup>. To determine statistical significance, we use a threshold of  $\log_{10}$  Bayes Factor (BF) > 6 which is comparable to a  $p < 5 \times 10^{-8}$ , previously determined by Wang et al.<sup>38</sup>. Heterogeneity of allelic effect sizes was calculated using Cochran's Q-test for heterogeneity in METAL<sup>70</sup>. Using MANTRA results we generated 99% credible sets most likely

to drive association signals and contain causal variants (or tagging unobserved causal variants) and compared fine-mapping intervals for each associated lead SNP by analysing the variants 500 kb up- and downstream of the lead SNP in the Ugandan and combined Ugandan + European datasets. For this, posterior probabilities were calculated for SNPs and then ranked in decreasing order according to BF, proceeding down the rank until the cumulative posterior probability exceeded 99%<sup>39,40</sup>. All SNPs  $\geq 0.99$  were included in the credible sets.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Summary statistics are available on the NHGRI-EBI GWAS Catalogue <https://www.ebi.ac.uk/gwas/downloads/summary-statistics> under accession codes: GCST90000522, GCST90000523, GCST90000524, GCST90000525, GCST90000526.

Received: 2 November 2019; Accepted: 13 July 2020;

Published online: 31 July 2020

## References

- Plummer, M. et al. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob. Health* **4**, e609–e616 (2016).
- De Flora, S. & La Maestra, S. Epidemiology of cancers of infectious origin and prevention strategies. *J. Prev. Med. Hyg.* **56**, E15–E20 (2015).
- de Sanjose, S. et al. Geographic variation in the prevalence of Kaposi sarcoma-associated herpesvirus and risk factors for transmission. *J. Infect. Dis.* **199**, 1449–1456 (2009).
- Krithivas, A., Young, D. B., Liao, G., Greene, D. & Hayward, S. D. Human herpesvirus 8 LANA interacts with proteins of the mSin3 corepressor complex and negatively regulates Epstein-Barr virus gene expression in dually infected PEL cells. *J. Virol.* **74**, 9637–9645 (2000).
- Xu, D. et al. Epstein–Barr virus inhibits Kaposi's sarcoma-associated herpesvirus lytic replication in primary effusion lymphomas. *J. Virol.* **81**, 6068–6078 (2007).
- Jiang, Y., Xu, D., Zhao, Y. & Zhang, L. Mutual inhibition between Kaposi's sarcoma-associated herpesvirus and Epstein–Barr virus lytic replication initiators in dually-infected primary effusion lymphoma. *PLoS ONE* **3**, e1569 (2008).
- Spadavecchia, S., Gonzalez-Lopez, O., Carroll, K. D., Palmeri, D. & Lukac, D. M. Convergence of Kaposi's sarcoma-associated herpesvirus reactivation with Epstein–Barr virus latency and cellular growth mediated by the notch signaling pathway in coinfecting cells. *J. Virol.* **84**, 10488–10500 (2010).
- Horenstein, M. G. et al. Epstein–Barr virus latent gene expression in primary effusion lymphomas containing Kaposi's sarcoma-associated herpesvirus/human herpesvirus-8. *Blood* **90**, 1186–1191 (1997).
- Sallah, N. et al. Genome-wide sequence analysis of Kaposi sarcoma-associated herpesvirus shows diversification driven by recombination. *J. Infect. Dis.* **218**, 1700–1710 (2018).
- Correia, S. et al. Sequence variation of Epstein–Barr virus: viral types, geography, codon usage, and diseases. *J. Virol.* <https://doi.org/10.1128/JVI.01132-18> (2018).
- Sallah, N. et al. Whole-genome association study of antibody response to Epstein–Barr virus in an African population: a pilot. *Glob. Health Epidemiol. Genom.* **2**, e18 (2017).
- Rubicz, R. et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein–Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet.* **9**, e1003147 (2013).
- Hammer, C. et al. Amino acid variation in HLA class II proteins is a major determinant of humoral response to common viruses. *Am. J. Hum. Genet.* **97**, 738–743 (2015).
- Scepanovic, P. et al. Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med.* **10**, 59 (2018).
- Mbulaiteye, S. M. et al. Human herpesvirus 8 infection within families in rural Tanzania. *J. Infect. Dis.* **187**, 1780–1785 (2003).
- Guttman-Yassky, E. et al. Familial clustering of classic Kaposi sarcoma. *J. Infect. Dis.* **189**, 2023–2026 (2004).
- Plancoulaine, S. et al. Respective roles of serological status and blood specific antihuman herpesvirus 8 antibody levels in human herpesvirus 8 intrafamilial transmission in a highly endemic area. *Cancer Res.* **64**, 8782–8787 (2004).
- Plancoulaine, S. et al. Human herpesvirus 8 transmission from mother to child and between siblings in an endemic population. *Lancet* **356**, 1062–1065 (2000).
- Kaasinen, E. et al. Nationwide registry-based analysis of cancer clustering detects strong familial occurrence of Kaposi sarcoma. *PLoS ONE* **8**, e55209 (2013).

20. Jackson, C. C. et al. Kaposi sarcoma of childhood: inborn or acquired immunodeficiency to oncogenic HHV-8. *Pediatr. Blood Cancer* **63**, 392–397 (2016).
21. Brown, E. E. et al. Host immunogenetics and control of human herpesvirus-8 infection. *J. Infect. Dis.* **193**, 1054–1062 (2006).
22. Goedert, J. J. et al. Risk of classic Kaposi sarcoma with combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci: a population-based case-control study. *J. Infect. Dis.* **213**, 432–438 (2016).
23. Alkharsah, K. R., Dedicoat, M., Blaszczak, R., Newton, R. & Schulz, T. F. Influence of HLA alleles on shedding of Kaposi sarcoma-associated herpesvirus in saliva in an African population. *J. Infect. Dis.* **195**, 809–816, (2007).
24. Yang, D. et al. Interleukin 1 receptor-associated kinase 1 (IRAK1) mutation is a common, essential driver for Kaposi sarcoma herpesvirus lymphoma. *Proc. Natl. Acad. Sci. USA* **111**, E4762–E4768 (2014).
25. Cornejo Castro, E. M. et al. Relationship between human leukocyte antigen alleles and risk of Kaposi's sarcoma in Cameroon. *Genes Immun.* <https://doi.org/10.1038/s41435-019-0077-9> (2019).
26. Asiki, G. et al. The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. *Int. J. Epidemiol.* **42**, 129–141 (2013).
27. Riha, J. et al. Urbanicity and lifestyle risk factors for cardiometabolic diseases in rural Uganda: a cross-sectional study. *PLoS Med.* **11**, e1001683 (2014).
28. Labo, N. et al. Mutual detection of Kaposi's sarcoma-associated herpesvirus and Epstein-Barr virus in blood and saliva of Cameroonians with and without Kaposi's sarcoma. *Int. J. Cancer* <https://doi.org/10.1002/ijc.32546> (2019).
29. Labo, N. et al. Heterogeneity and breadth of host antibody response to KSHV infection demonstrated by systematic analysis of the KSHV proteome. *PLoS Pathog.* **10**, e1004046 (2014).
30. Newton, R. et al. Kaposi sarcoma-associated herpesvirus in a rural Ugandan cohort, 1992–2008. *J. Infect. Dis.* **217**, 263–269 (2018).
31. Heckerman, D. et al. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci. USA* **113**, 7377–7382 (2016).
32. Gurdasani, D. et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* <https://doi.org/10.1038/nature13997> (2014).
33. Gurdasani, D. et al. Uganda genome resource enables insights into population history and genomic discovery in Africa. *Cell* **179**, 984–1002 e1036 (2019).
34. Masala, M. V. et al. Classic Kaposi's sarcoma in Sardinia: HLA positive and negative associations. *Int. J. Dermatol.* **44**, 743–745 (2005).
35. van der Kuyl, A. C. et al. An IL-8 gene promoter polymorphism is associated with the risk of the development of AIDS-related Kaposi's sarcoma: a case-control study. *AIDS* **18**, 1206–1208 (2004).
36. Brown, E. E. et al. A common genetic variant in FCGR3A-V158F and risk of Kaposi sarcoma herpesvirus infection and classic Kaposi sarcoma. *Cancer Epidemiol. Biomark. Prev.* **14**, 633–637 (2005).
37. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
38. Wang, X. et al. Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **22**, 2303–2311 (2013).
39. Charles, B. A. et al. A genome-wide association study of serum uric acid in African Americans. *BMC Med. Genomics* **4**, 17 (2011).
40. Maller, J. B. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
41. Morrison, B. J., Labo, N., Miley, W. J. & Whitby, D. Serodiagnosis for tumor viruses. *Semin. Oncol.* **42**, 191–206 (2015).
42. Newton, R. et al. The sero-epidemiology of Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in adults with cancer in Uganda. *International journal of cancer.* *Int. J. Cancer* **103**, 226–232 (2003).
43. Pfeiffer, R. M. et al. Geographic heterogeneity of prevalence of the human herpesvirus 8 in sub-Saharan Africa: clues about etiology. *Ann. Epidemiol.* **20**, 958–963 (2010).
44. Shebl, F. M. et al. Population-based assessment of kaposi sarcoma-associated herpesvirus DNA in plasma among Ugandans. *J. Med. Virol.* **85**, 1602–1610 (2013).
45. Wakeham, K. et al. Parasite infection is associated with Kaposi's sarcoma associated herpesvirus (KSHV) in Ugandan women. *Infect. Agent Cancer* **6**, 15 (2011).
46. Nalwoga, A. et al. Association between malaria exposure and Kaposi's sarcoma-associated herpes virus seropositivity in Uganda. *Trop. Med. Int. Health* **20**, 665–672 (2015).
47. Stockdale, L. et al. Human cytomegalovirus epidemiology and relationship to tuberculosis and cardiovascular disease risk factors in a rural Ugandan cohort. *PLoS ONE* **13**, e0192086 (2018).
48. Newton, R. et al. Determinants of Gammaherpesvirus shedding in saliva among Ugandan children and their mothers. *J. Infect. Dis.* **218**, 892–900 (2018).
49. Wakeham, K. et al. Trends in Kaposi's sarcoma-associated Herpesvirus antibodies prior to the development of HIV-associated Kaposi's sarcoma: a nested case-control study. *Int. J. Cancer* **136**, 2822–2830 (2015).
50. Nalwoga, A. et al. Kaposi's sarcoma-associated herpesvirus seropositivity is associated with parasite infections in Ugandan fishing communities on Lake Victoria islands. *PLoS Negl. Trop. Dis.* **13**, e0007776 (2019).
51. Zaitlen, N. et al. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet.* **9**, e1003520 (2013).
52. Rubicz, R. et al. Genome-wide genetic investigation of serological measures of common infections. *Eur. J. Hum. Genet.* **23**, 1544–1548 (2015).
53. Zhou, Y. et al. Genetic loci for Epstein–Barr virus nuclear antigen-1 are associated with risk of multiple sclerosis. *Mult. Scler. J.* <https://doi.org/10.1177/1352458515626598> (2016).
54. Besson, C. et al. Strong correlations of anti-viral capsid antigen antibody levels in first-degree relatives from families with Epstein–Barr virus-related lymphomas. *J. Infect. Dis.* **199**, 1121–1127 (2009).
55. Sitas, F. et al. Antibodies against human herpesvirus 8 in black South African patients with cancer. *N. Engl. J. Med.* **340**, 1863–1871 (1999).
56. Cannon, M. J. et al. Risk factors for Kaposi's sarcoma in men seropositive for both human herpesvirus 8 and human immunodeficiency virus. *AIDS* **17**, 215–222 (2003).
57. Yanez, A. A., Harrell, T., Sriranganathan, H. J., Ives, A. M. & Bertke, A. S. Neurotrophic factors NGF, GDNF and NTN selectively modulate HSV1 and HSV2 lytic infection and reactivation in primary adult sensory and autonomic neurons. *Pathogens* <https://doi.org/10.3390/pathogens6010005> (2017).
58. Myoung, J. & Ganem, D. Active lytic infection of human primary tonsillar B cells by KSHV and its noncytolytic control by activated CD4+ T cells. *J. Clin. Invest.* **121**, 1130–1140 (2011).
59. Roshan, R. et al. T-cell responses to KSHV infection: a systematic approach. *Oncotarget* **8**, 109402–109416 (2017).
60. Butler, L. M. et al. Kaposi's sarcoma-associated herpesvirus inhibits expression and function of endothelial cell major histocompatibility complex class II via suppressor of cytokine signaling 3. *J. Virol.* **86**, 7158–7166 (2012).
61. Sabbah, S. et al. T-cell immunity to Kaposi sarcoma-associated herpesvirus: recognition of primary effusion lymphoma by LANA-specific CD4+ T cells. *Blood* **119**, 2083–2092 (2012).
62. Thakker, S. et al. Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen inhibits major histocompatibility complex class II expression by disrupting enhanceosome assembly through binding with the regulatory factor X complex. *J. Virol.* **89**, 5536–5556 (2015).
63. Dilthey, A. et al. Multi-population classical HLA type imputation. *PLoS Comput Biol.* **9**, e1002877 (2013).
64. Nunn, A. J. et al. Mortality associated with HIV-1 infection over five years in a rural Ugandan population: cohort study. *BMJ* **315**, 767–771 (1997).
65. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
66. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
67. Dilthey, A. T., Moutsianas, L., Leslie, S. & McVean, G. HLA\*IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* **27**, 968–972 (2011).
68. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
69. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
70. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

## Acknowledgements

We thank all study participants who contributed to this study. We acknowledge the African Genome Variation Project (AGVP) for sharing data resources to contextualise our results. The GPC is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement. Further funding was obtained from the Wellcome Trust (WT098051 and WT090132), the UK Medical Research Council and with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E and Contract 75N91019D00024. D.G. received a UKRI/Rutherford HDR-UK fellowship, serial number: MR/S003711/1.

## Author contributions

N.S., I.B., P.K., M.S., D.W. and R.N. designed the study, interpreted the results and wrote the manuscript. W.M., N.L. and D.W. performed phenotype assay development and validation. T.C., S.F., M.O.P. and D.G. carried out curation of sequence and genotype data, including establishment of the reference panel including the African Genome Variation Project data.

N.S. carried out all statistical analyses: regression, heritability and genetic association analyses. A.T.D. and A.J.M. performed HLA imputation of genotypes and assisted with interpretation of findings. V.M., E.M.C.C. and M.L.H. contributed to revising of the manuscript. G.A. was the programme leader for the GPC. C.P. and E.H.Y. were the scientific coordinators of the project and coordinated regulatory approvals. All authors commented on the interpretation of the results, reviewed and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-17696-2>.

**Correspondence** and requests for materials should be addressed to N.S. or I.B.

**Peer review information** *Nature Communications* thanks Paul McLaren, Jacques Fellay, and the other, anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020