

RESEARCH ARTICLE

Antibody epitope profiling of the KSHV LANA protein using VirScan

Sydney J. Bennett^{1,2}*, Dicle Yalcin²*, Sara R. Privatt^{1,2}, Owen Ngalamika³, Salum J. Lidenge^{4,5}, John T. West², Charles Wood^{1,2*}

1 School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, Nebraska, United States of America, **2** Department of Interdisciplinary Oncology, Stanley S. Scott Cancer Center, Louisiana State University Health Sciences Center, New Orleans, Louisiana, United States of America, **3** Dermatology and Venereology Section, University Teaching Hospital, University of Zambia School of Medicine, Lusaka, Zambia, **4** Ocean Road Cancer Institute, Dar es Salaam, Tanzania, **5** Department of Clinical Oncology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

✉ These authors contributed equally to this work.

* cwoo12@lsuhsc.edu



OPEN ACCESS

Citation: Bennett SJ, Yalcin D, Privatt SR, Ngalamika O, Lidenge SJ, West JT, et al. (2022) Antibody epitope profiling of the KSHV LANA protein using VirScan. PLoS Pathog 18(12): e1011033. <https://doi.org/10.1371/journal.ppat.1011033>

Editor: Ashlee V. Moses, Oregon Health and Science University, UNITED STATES

Received: August 8, 2022

Accepted: November 30, 2022

Published: December 19, 2022

Copyright: © 2022 Bennett et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This research was supported in part by the National Institute of Health grants from the National Cancer Institute R01 CA228178 (JTW and CW), R01 CA239591 (CW), U54 CA221204 (CW), U54 CA190155 (CW), R01 DAO47823 (CW), and K43 TW011418 (SJJ), Fogarty International D43 TW010354 (CW) and D43 TW012277 (CW), and Ruth L. Kirschstein National Research Service

Abstract

The humoral antibody response against Kaposi sarcoma-associated herpesvirus (KSHV) in infected individuals has been characterized demonstrating the latency-associated nuclear antigen (LANA) as the most antigenic KSHV protein. Despite the antigenicity of the protein, specific LANA epitopes have not been systematically characterized. Here, we utilized a bacteriophage T7 library, which displays 56-amino acid KSHV LANA peptides with 28-amino acid overlap (VirScan), to define those epitopes in LANA targeted by antibodies from a cohort of 62 sub-Saharan African Kaposi sarcoma (KS) patients and 22 KSHV-infected asymptomatic controls. Intra- and inter-patient breadth and magnitude of the anti-LANA responses were quantified at the peptide and amino acid levels. From these data, we derived a detailed epitope annotation of the entire LANA protein, with a high-resolution focus on the N- and C-termini. Overall, the central repeat region was highly antigenic, but the responses to this region could not be confidently mapped due to its high variability. The highly conserved N-terminus was targeted with low breadth and magnitude. In a minority of individuals, antibodies specific to the nuclear localization sequence and a portion of the proline-rich regions of the N-terminus were evident. In contrast, the first half of the conserved C-terminal domain was consistently targeted with high magnitude. Unfortunately, this region was not included in LANA partial C-terminal crystal structures, however, it was predicted to adopt predominantly random-coil structure. Coupled with functional and secondary structure domain predictions, VirScan revealed fine resolution epitope mapping of the N- and C-terminal domains of LANA that is consistent with previous antigenicity studies and may prove useful to correlate KSHV humoral immunity with pathogenesis.

Author summary

The incidence of Kaposi sarcoma (KS) coincides with high prevalence of Kaposi sarcoma-associated herpesvirus (KSHV) infection, particularly in sub-Saharan Africa, Peru, and

Award 1 T32 AI125207 (CW) from the National Institute of Allergy and Infectious Diseases. SJB and SRP were Ruth L. Kirschstein fellows; SJL and ON were Fogarty fellows. The University of Nebraska DNA Sequencing Core receives partial support from the National Institute for General Medical Science INBRE (P20 GM103427-19) grant as well as The Fred & Pamela Buffett Cancer Center Support Grant (P30 CA036727). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

the Mediterranean. KSHV infection typically occurs during childhood and enters latency typified by the expression of latency-associated nuclear antigen (LANA). Despite recognition that LANA is highly antigenic, no previous study has defined the targeted epitopes. Our study presents the first high-resolution epitope map of LANA in KS patients and asymptomatic KSHV-infected individuals. High-throughput phage display, VirScan, was leveraged to elucidate epitopes across the LANA protein and compare them between KS and asymptomatic groups. A 24-amino acid peptide was found to be highly antigenic in both groups, regardless of co-infection with human immunodeficiency virus (HIV). In KS patients, HIV co-infection was associated with a lower response to this epitope, but HIV co-infection had little to no effect in the absence of cancer. Further studies of the identified epitope will determine whether the epitope has therapeutic, diagnostic, or viral decoy potential. The application of our novel findings will be instrumental when applied to the entire KSHV proteome in designing effective vaccines for KSHV infection and therapeutics to prevent progression to KS.

Introduction

Kaposi sarcoma-associated herpesvirus (KSHV) is the causative agent of Kaposi sarcoma (KS), which occurs at high incidence in sub-Saharan Africa (SSA), and a number of other geographic locations [1]. KS incidence is higher among individuals co-infected with Human Immunodeficiency Virus-1 (HIV; Epidemic KS) compared to HIV⁻ individuals (Endemic KS), due to acquired immunodeficiency syndrome (AIDS)-induced immunosuppression. Latency-associated nuclear antigen (LANA) is a 1,162 amino acid (230kDa) protein encoded by the KSHV ORF73 gene (NC_009333.1). LANA is consistently and abundantly detected in KSHV-infected KS tumor cells, KSHV primary effusion lymphoma cells, multicentric Castleman's disease B cells, and latently infected cells *in vitro*. LANA is a multifunctional oncoprotein that plays a role in KSHV-mediated tumorigenesis through its manipulation of cell cycle machinery and deregulation of tumor suppressor pathways. Further, LANA is highly antigenic, and along with the lytic glycoprotein K8.1, its serological detection is considered the gold standard for diagnosis of KSHV infection in latency and lytic replication, respectively [2]. When detected, LANA exhibits a punctate nuclear staining pattern in infected or tumor cells that is diagnostic of KSHV etiology.

Structurally, LANA has conserved amino (N-) and carboxyl (C-) terminal regions, interspersed by an extensive, highly variable, and acidic series of repeats. The N-terminus, also known as the chromosome-binding motif (CBM), has been shown to tether the viral episome to host chromatin by interacting with histones H2A and H2B. This region plays a central role in promoting KSHV episomal persistence by coordinating episome duplication and partitioning with cellular replication [3]. This coordination also helps keep KSHV genome burden and antigen low in most infected cells. On the other hand, the C-terminus is commonly referred to as the KSHV DNA-binding domain (DBD) since it binds directly to the conserved terminal repeats (TR) in the KSHV episome to maintain viral genomic stability. Therefore, LANA constitutes a link between cellular chromatin and the KSHV episome. The C-terminal domain has a predicted hydrophobic 3-dimensional core structure. Mutagenesis of the hydrophobic residues in this tetramer interface reveals that they are essential for higher order oligomerization of LANA, which imparts the cooperativity necessary for DNA-protein complex formation and contributes to genome stability, persistence, and replication [4].

Seroepidemiological studies using protein- and infected cell-based assays to detect and quantify KSHV seroprevalence and humoral antibody (Ab) response have employed immuno-fluorescence assays (IFA), luciferase immunoprecipitation systems (LIPS), enzyme-linked immunosorbent assays (ELISA), and protein bead-arrays, each varying in specificity and sensitivity, yet overall contributing major advancements to Ab response evaluation and diagnosis of KSHV infection. Phage display technologies have recently been developed to present large libraries of potential peptide epitopes in tiled arrays to facilitate mapping of all potential Ab recognition sites that other immunoassays cannot do readily or systematically. For example, it is possible to profile Ab repertoires against entire proteomes without having to purify the target proteins or analyze each peptide individually. Moreover, individual- and group-specific Ab responses can be defined comparatively between those with and without disease or with and without an infection.

Although extensively studied and demonstrated to be highly antigenic, LANA epitopes have rarely been investigated and never using a high-resolution tiled phage display system. In this study, we mapped epitopes of the KSHV LANA protein using phage display and immunoprecipitation followed by barcoding and high-throughput sequencing (PhIP-Seq). LANA peptides were displayed on T7-bacteriophage as a library of 56-amino acid tiles, each with a 28-amino acid overlap (VirScan) [5,6]. We quantified the recognition of the peptide epitopes in LANA and comparatively assessed the breadth and magnitude of the anti-LANA response within and between 16 HIV⁻ Endemic KS (EnKS) patients, 46 HIV⁺ Epidemic KS (EpKS) patients, and 22 KSHV-infected individuals without KS (asymptomatic) from SSA.

Results

Phage library and cohort characteristics

The antibody (Ab) response against KSHV-LANA was mapped using VirScan and samples from sub-Saharan African (SSA) study cohorts in Zambia and Tanzania (Fig 1). The cohort included 16 HIV⁻ Kaposi Sarcoma patients (Endemic KS, EnKS) and 46 HIV⁺ Kaposi Sarcoma patients (Epidemic KS, EpKS). The EnKS cohort was significantly older (*Mann-Whitney*, $p = 0.0003$) and had a significantly greater proportion of males compared to EpKS cohort (Table 1, *Fisher's Exact Test*, $p = 0.012$) reflecting the epidemiology observed in SSA where EpKS is more prominent, while EnKS occurs more frequently in older males [1]. In addition to the KS individuals, 22 KSHV⁺ asymptomatic (KS⁻) individuals were included. To control for both the EnKS and EpKS groups, 11/22 of the asymptomatic individuals were HIV⁻ and 11/22 were HIV⁺. The median age (*Mann-Whitney*, $p > 0.05$) and proportion of males (*Fisher's Exact*, $p > 0.05$) were not significantly different between the KS and asymptomatic groups. Furthermore, all HIV⁺ patients included in this study were on antiretroviral therapy (ART), with self-reported adherence in all but two individuals. The extent of immunosuppression, as reflected in HIV viral load, reduced CD4 counts, or longer duration of HIV infection might have impacted LANA (and other) Ab responses; however, no significant difference in CD4 count, HIV plasma viral load, or duration of HIV infection was evident (Table 1, S4 Fig, *Mann-Whitney*, $p > 0.05$).

To normalize total immunoglobulin G (IgG) between the significantly elevated levels in EpKS compared to EnKS subjects (*Mann-Whitney*, $p = 0.0144$), a total IgG of 2 μ g was input into each PhIP-Seq assay. Each precipitation resulted in >850,000 reads/replicate, and the number of reads was not significantly different between KS and asymptomatic groups (Table 1, *Mann-Whitney*, $p > 0.05$). Greater than 95% of reads aligned with the reference sequences suggesting there were no contamination or amplification issues. Additionally, sequencing of 32 mock immunoprecipitations (PBS-only, mock IPs) revealed uniform

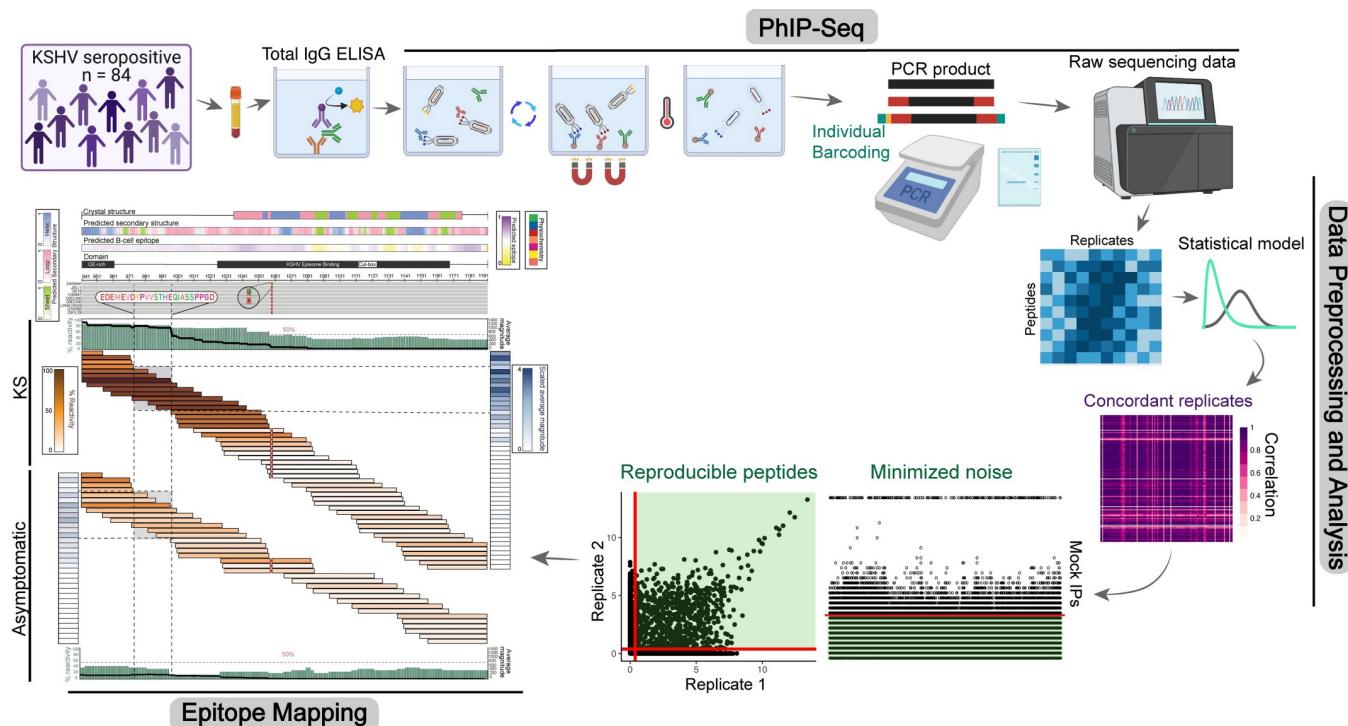


Fig 1. VirScan experiment and data analysis workflow. Phage immunoprecipitation and sequencing (PhIP-Seq) is carried out by first quantifying the levels of total IgG by ELISA, incubating the phage and plasma while rotating to allow complex formation, pulling down the phage-Ab complexes using magnetic beads and washing away the unbound phage, and finally, PCR amplification of the phage DNA followed by sequencing. The sequencing data is aligned to the reference sequences and a peptide by replicate count matrix is generated. The count data was fit to a Gamma-Poisson distribution to obtain residual p-values that represented the enrichment of a peptide above a significance-level cutoff, termed reactive peptides. Peptides were further filtered by removing peptides that had high non-specific binding in mock IPs. Samples were also filtered based on replicate correlations. Finally, statistically significant and reproducible peptides with highly concordant sample replicates were selected for downstream analyses and viral epitope mapping of KSHV-LANA. Created with [BioRender.com](#).

<https://doi.org/10.1371/journal.ppat.1011033.g001>

distribution of the phage library where >99.6% of the expected peptides were represented. From these data, we demonstrated that the phage library population was not bottlenecked during amplification.

There were 224 KSHV-LANA peptides in the VirScan library. Of these 224 peptides, 59 were non-unique amino acid sequences, meaning that Ab precipitation would be non-discriminatory between them. Thus, the resulting counts were consolidated with each matching peptide of the 165 remaining for downstream analyses. One of the 165 peptides was missing in all 32 mock-IPs, and thus was likely lost during library amplification. Finally, 9 peptides were filtered out due to high non-specific binding in the mock-IPs (S1 Fig) leaving 155 KSHV-LANA peptides for downstream analyses. At least one KS patient reacted to 151/155 potential target peptides, compared to recognition of 130/155 potential peptides in the asymptomatic individuals.

As previously shown, KSHV-specific Ab titer, determined using a BC3 monoclonal-enhanced immunofluorescence assay (mIFA) [7–10], was significantly higher in KS compared to asymptomatic individuals (Table 1, Mann-Whitney, $p < 0.0001$). While the number of reactive peptides per patient, i.e., *breadth*, significantly differed between low, medium, and high total KSHV Ab titer individuals (S2 Fig, Kruskal-Wallis, $p = 0.0006$), titer only moderately correlated with breadth (S2 Fig, Spearman $r = 0.3752$, $p = 0.0004$) and weakly correlated with log-scaled magnitude (S2 Fig, Spearman $r = 0.1124$, $p = 0.3087$) indicating that anti-LANA repertoire differences between KS and asymptomatic groups were not solely attributable to titer differentials. Importantly, KSHV Ab titer encompasses the Ab response against all KSHV components, whereas in this study, only anti-LANA responses were quantified. While the

Table 1. Cohort demographics (N = 84). Significant comparisons are denoted: ^aEnKS vs. EpKS, ^bAsymptomatic HIV⁻ vs. HIV⁺, ^cKS vs. Asymptomatic. P-values are denoted: ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05

Variable	Kaposi Sarcoma (KS, n = 62)		Asymptomatic KSHV ⁺ (n = 22)	
	HIV ⁻ (EnKS, n = 16)	HIV ⁺ (EpKS, n = 46)	HIV ⁻ (n = 11)	HIV ⁺ (n = 11)
	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)
Demographic				
Age (years) ^{a***}	54 (33.75)	32.5 (11)	33 (5)	43 (12)
Gender ^{a*}				
Female	1 (6%)	19 (41%)	6 (55%)	4 (36%)
Male	15 (94%)	27 (59%)	5 (45%)	7 (64%)
HIV				
Duration (days)	-	180 (761.3)	-	545 (1615)
ART Adherence				
Always	-	42 (92%)	-	11 (100%)
Mostly	-	2 (4%)	-	0 (0%)
Rarely	-	2 (4%)	-	0 (0%)
CD4 Count	-	210 (276.5)	-	360 (191.8)
< 200	-	22 (48%)	-	1 (9%)
200–500	-	17 (37%)	-	6 (55%)
> 500	-	6 (13%)	-	1 (9%)
Unknown	-	1 (2%)	-	3 (27%)
Viral load (cps/mL)	-	1458 (17532.7)	-	103.5 (43)
Undetectable	-	19 (41%)	-	6 (55%)
Unknown	-	13 (28%)	-	3 (27%)
Antibody				
Total IgG (mg/mL) ^{a*}	6.279 (5.103)	11.044 (8.362)	8.388 (1.63)	8.77 (1.433)
KSHV Ab Titer ^{c****}	1280 (2240)	1280 (1280)	160 (280)	40 (280)
Low	1 (6%)	9 (20%)	6 (55%)	8 (73%)
Medium	9 (56%)	26 (56%)	5 (45%)	2 (18%)
High	6 (38%)	11 (24%)	0 (0%)	1 (9%)
Sequencing Data				
Reads/Replicate ^{b*}	1,765,070 (385,470)	1,736,288 (452,445)	1,821,327 (243,137)	1,500,691 (635,943)

<https://doi.org/10.1371/journal.ppat.1011033.t001>

breadth and magnitude of response against LANA have an association with KSHV Ab titer, they cannot completely explain the titer, suggesting that other factors (i.e., the response to other KSHV proteins) are also contributing factors to KSHV Ab titer.

KSHV-LANA can be subdivided into N-terminal, Central Repeat, and C-terminal regions

Since LANA has been partially sequenced many times, the phage library incorporated six different isolates of LANA (UniProtKB accession: O40947, Q9DUM3, Q9DUN0, Q76SB0, Q91LX9, and Q9QR71). When these isolates were compared with common KSHV reference sequences, (GK18 and JSC-1), coverage of the entire LANA protein was evident in the phage library (Fig 2). Additionally, LANA is conserved at the N- and C-termini but varied in length and sequence in the central repeat region (Fig 2). Due to the nature of the repeat sequences, the central repeat region has been difficult to sequence and align with confidence. This resulted in ambiguous mapping of peptides in this region and hindered analysis of epitopes. For downstream analyses, we divided LANA into three domains based on multiple sequence alignment

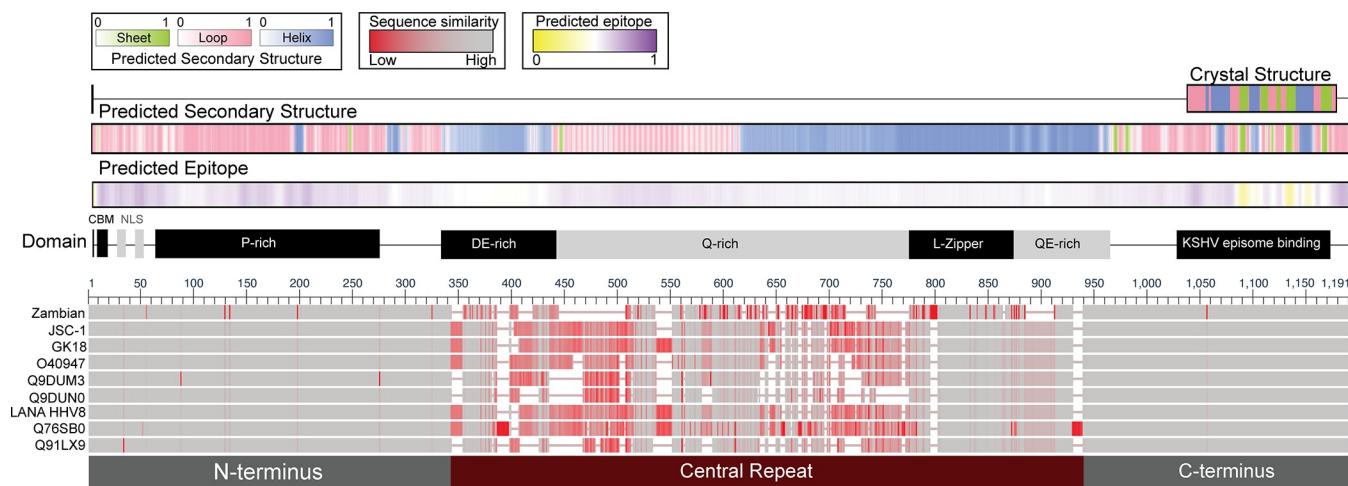


Fig 2. KSHV-LANA protein multiple sequence alignments. The reference LANA sequences (70% Zambian consensus, JSC-1, and GK18) and input library sequences (indicated by their UniProtKB IDs) were aligned, and residues were colored based on their sequence similarity; gray indicates 100% sequence identity, red represents no similarity at a given position. Computationally predicted and crystal secondary structures are annotated and color-coded by type: sheet (green), loop (pink), or helix (blue). Predicted secondary structures are represented in color gradients representing the probabilities (minimum [0; white] and maximum [1; sheet (green), loop (pink), or helix (blue)]). Additionally, the BepiPred-2.0 tool was used to map probabilities of a predicted B-cell epitope. Lastly, functional motifs/domains are annotated directly above the aligned sequences. CBM: Chromosome binding motif, NLS: Nuclear localization sequence, P-rich: Proline-rich region, DE-rich: Aspartic/Glutamic acid-rich region, L-Zipper: Leucine zipper motif, QE-rich: Glutamine/Glutamic acid-rich region.

<https://doi.org/10.1371/journal.ppat.1011033.g002>

(MSA) results: N-terminus [MSA 1–342], central repeat region [MSA 343–940], and C-terminus [MSA 941–1191] (Fig 2). The Zambian consensus sequence contained six point mutations in the conserved N- and C-termini that were not represented by the isolates in the library, indicating the peptides presented by the phage library were representative of LANA in the virus to which the study cohort would most likely have been infected (see Zambian, Fig 2).

The LANA C-terminus contains the KSHV episome binding site and is the only portion of the LANA protein for which a crystal structure has been determined (PDB: 4UZB) [11]. That model shows the LANA C-terminus is mostly random coil, with interspersed helices and beta-sheets. The secondary structure derived from the crystal structure largely agrees with secondary structure predictions, (72% of amino acids accurately predicted) (Fig 2). For this reason, the predicted secondary structure of the remaining regions of LANA that have not been crystallized was likewise inferred. The LANA N-terminus contains the chromosome binding site, nuclear localization sequences, and a proline (P)-rich region. The N-terminal region is predicted to be predominantly random coil, especially in the proline-rich region. In contrast, the central repeat region of LANA while harboring regions of random coil, also contains large helical regions, likely resulting from the various repeats and leucine zipper motifs (Fig 2). Interestingly, while LANA is known to be a highly antigenic protein [12,13], an advanced B cell epitope prediction algorithm, BepiPred 2.0, did not identify any particular region of LANA to be highly antigenic (see epitope prediction score, Fig 2). To demonstrate that VirScan can be used to map B cell epitopes of KSHV, an epitope map of the Ab response against KSHV-LANA was built to investigate which sequences of the protein contribute to its antigenicity and how well such repertoires correspond to predicted B cell epitopes, and to previous empirical Ab-response data.

The KSHV-LANA central repeat and C-terminal regions show high antigenicity

The 151 validated LANA-derived peptides in the VirScan library were partitioned into N-terminus, C-terminus, and central repeat region resulting in 19/151 peptides in the N-terminus,

48/151 peptides in the C-terminus, and 84/151 peptides in the central repeat region. The Ab response to each region was compared in multiple ways. First, the percentage of patients reactive to at least one of the peptides derived from each of the three regions was calculated and compared between KS and asymptomatic controls: 52% of KS patients and 46% of asymptomatic controls were reactive to at least one peptide within the N-terminus, 69% and 68% in the central repeat region, and 89% and 73% in the C-terminus, respectively. No significant difference in percent patient reactivity between KS patients and asymptomatic controls was observed in any of the regions (*Fisher's Exact*, $p>0.05$), and the C-terminus was the most consistently targeted region among both KS patients and asymptomatic controls (*Chi-square*, $p<0.0001$).

Second, the sum of reactive peptides per sample (*breadth*) for each region was calculated. For KS patients, breadth was significantly higher in the C-terminal and central repeat regions when compared to the N-terminus (Fig 3A, $p<0.0001$). In asymptomatic controls, only the central repeat region had significantly higher breadth than the N-terminus (Fig 3A, $p = 0.0185$). The loss of significance was likely due to an overall significantly lower total LANA breadth in asymptomatic controls compared to KS patients (S3 Fig, $p = 0.0452$). Furthermore, the breadth was moderately able to linearly separate KS and asymptomatic by principal component analysis (PCA), explaining ~43% of the total variance (S3 Fig). Lastly, the average magnitude of reactive peptides per sample for each region was calculated. The average magnitude represents how many times each reactive peptide was targeted by an Ab. Similar to breadth, the magnitude was significantly different across all three regions in KS patients, where the C-terminus was the most frequently targeted region (Fig 3B). On the contrary, there was no significant difference between the three regions of LANA in asymptomatic controls (Fig 3B). This data demonstrates that the central repeat region of LANA was nearly as antigenic as the C-terminus. However, due to the lack of confidence in amino acid position assignments, responses to the central repeat region could not be confidently mapped (Fig 2) [12]. Notably, no single peptide from the repeat region was consistently hit across the cohort and no shared

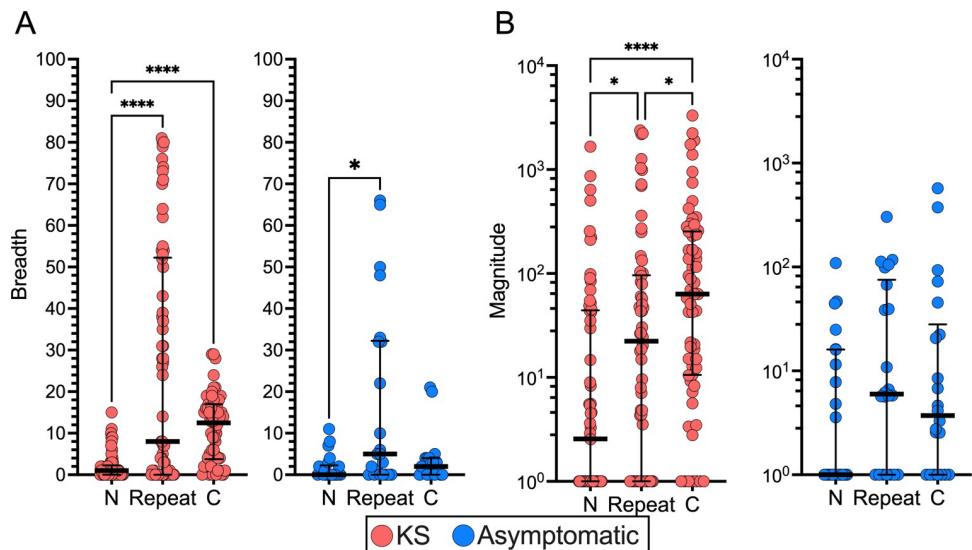


Fig 3. Breadth and magnitude of KSHV-LANA responses across regions. Comparison between KSHV-LANA responses at different regions in KS and asymptomatic individuals based on (A) breadth of response representing the number of reactive peptides, and (B) magnitude of response representing the frequency of which a reactive peptide was targeted (*Kruskal-Wallis post-hoc tests with Dunn's correction*). Significance levels are indicated with asterisks, where $p<0.05$ *, $p<0.01$ **, $p<0.001$ ***, $p<0.0001$ ****.

<https://doi.org/10.1371/journal.ppat.1011033.g003>

converging repeat pattern was observed amongst the top reactive peptides. Nevertheless, epitope mapping of the N- and C-terminal domains readily identified the peptides and amino acids responsible for LANA's antigenicity.

The KSHV-LANA N-terminus exhibits weak antigenicity

The LANA N-terminus had lower overall breadth and magnitude when compared to the C-terminus and central repeat regions (Fig 3), suggesting it was the least antigenic region of the LANA protein. To determine whether Ab recognition of the N-terminus was uniformly low or if there was a focal point of antigenicity, the peptides contained in the N-terminal domain were mapped to the reference sequences and the percent patient reactivity and average magnitude were overlayed (Fig 4). The maxima of patients reactive to each peptide in the N-terminus never reached 25%, demonstrating that there was no epitope consistently targeted in either the KS or asymptomatic groups. To confirm this was not an effect of having overlapping peptides, the percent patient reactivity per residue was calculated. For each amino acid residue, if a patient was reactive to at least one peptide encompassing that residue, they were considered

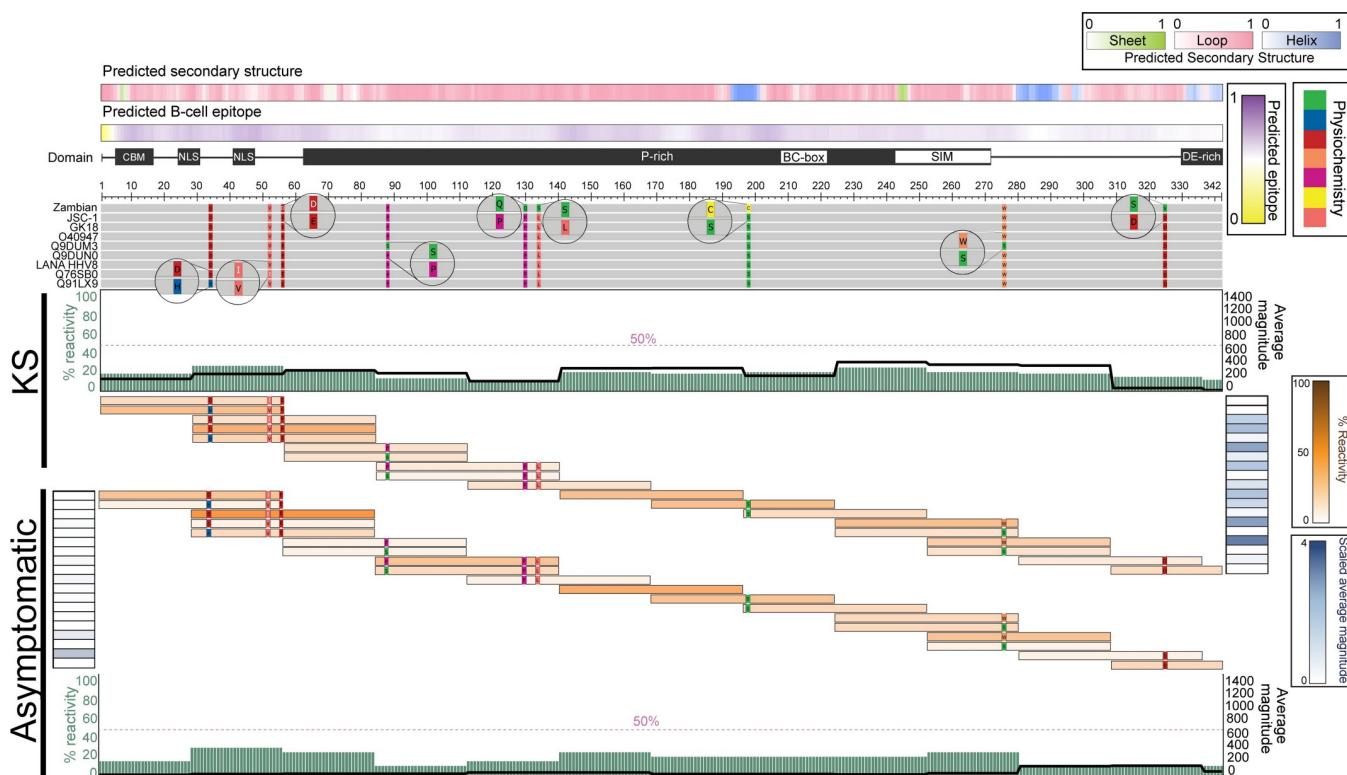


Fig 4. KSHV-LANA N-terminus epitope mapping in KS and asymptomatic individuals. The multiple sequence alignment (MSA) of reference sequences followed by the input library sequences is shown where gray represents conservation among the sequences. Mismatched residues are colored by their physiochemical properties as defined by the Zappo color scheme (green: hydrophilic, salmon: aliphatic/aromatic, orange: aromatic, fuchsia: conformationally special, yellow: cysteine only, red: negatively charged, blue: positively charged). Predicted secondary structures are represented in color gradients representing the probabilities (minimum [0; white] and maximum [1; sheet (green), loop (pink), or helix (blue)]). BepiPred-2.0 scores representing predicted B-cell epitope probabilities, along with functional motifs/domains are annotated directly above the aligned sequences. In addition, each peptide that was targeted by at least one individual was mapped and colored based on the percent cohort reactivity (brown gradient). Adjacent to the peptide responses of each group, scaled average magnitudes are shown in blue gradient. Moreover, percent patient reactivity and average magnitude of responses are calculated at the amino-acid level, represented by the sage bars and black lines in each group, respectively. CBM: Chromosome binding motif, NLS: Nuclear localization sequence, P-rich: Proline-rich region, BC-box: associated to Elongin B/C motif, SIM: Small ubiquitin-related modifier (SUMO) Interacting motif, DE-rich: Aspartic/Glutamic acid-rich region.

<https://doi.org/10.1371/journal.ppat.1011033.g004>

reactive. The same logic was applied to the magnitude, such that, for all patients considered reactive to a residue, the magnitudes of the individual's responses to each of the peptides encompassing that residue were averaged. The per residue percent patient reactivity increased slightly, but remained <30%, confirming that there were no consistently targeted epitopes in the LANA N-terminus (Fig 4). Similarly, the average magnitude was low across the entire N-terminus at both the peptide and amino acid levels (Fig 4). Regions with the highest percent patient reactivity in the N-terminus were the NLS and a portion of the Proline-rich region in both KS and asymptomatic (Fig 4). Broad, low magnitude to the N-terminus of LANA was detected with a slight increase in KS, suggesting the N-terminal region of the LANA protein is not stimulating an efficient Ab response in either group. Additionally, peptides containing the five point mutations in the N-terminus showed similar percent reactivities, indicating that they did not skew the antigenicity profiles.

The first half of the KSHV-LANA C-terminus contains a consistently targeted epitope

The overall Ab-response to the C-terminus of LANA was higher than the response to the N-terminus (Fig 3), and like the N-terminus, the C-terminal peptides were mapped to the reference sequences to ascertain if the C-terminal reactivity was focal or broad. This mapping revealed that the first half of the C-terminus was highly targeted in KS patients (Fig 5). While the same recognition pattern was detected in the asymptomatic controls, it was less consistent and occurred at lower magnitude (Fig 5). More specifically, the nine overlapping peptides spanning LANA₉₇₄₋₉₉₇ [EDEMEVDYPVVSTHEQIASSPPGD] were the most consistently targeted and had high average magnitudes across KS patients. LANA₉₇₄₋₉₉₇ was also highly recognized in asymptomatic controls, but in reactivity was similar to LANA₉₄₁₋₉₇₃ [QELEEVEEQQGVEQQEQETVEEPIILHGSSS]. While neither region was predicted to be highly antigenic by the BepiPred 2.0 tool, our data demonstrate these are the most antigenic LANA epitopes in KSHV infected individuals (see Predicted B cell epitope, Fig 5). B cell epitope prediction at this region was only marginally positive since the BepiPred2.0 tool associates confident random coil structure to higher likelihood of a potential epitope, and the structural predictions included both helices and beta-strands in addition to random coil. The predicted secondary structure of LANA₉₇₄₋₉₉₇ was mostly random coil with a beta-strand near the beginning, while LANA₉₄₁₋₉₇₃ was mostly helical with a beta-strand towards the end. To conclude, the Ab repertoire against the LANA C-terminus is against a pair of specific epitopes rather than the low, broad response to the N-terminus (Fig 5). The focused response against the C-terminus is likely targeting the highly antigenic epitope within LANA₉₄₁₋₉₉₇, a response that may be refined or reinforced in KS.

The effect of HIV co-infection on the anti-LANA response

To investigate the effect of HIV infection on the Ab response, we stratified the KS and asymptomatic groups by HIV status and comparatively evaluated the percent patient reactivity and average magnitude at the residue level (Fig 6); where a high percent reactivity indicates the given region was consistently targeted by patients across the cohort, and a high average magnitude indicates the given region was highly targeted within the patients. Notably, all four groups followed a similar pattern of reactivity across both the N- and C-termini with varying strength and consistency (Fig 6). While the overall N-terminal breadth and magnitude were not significantly different between EnKS and EpKS (S3 Fig), a more focused response was evident in the EnKS patients as indicated by peaks in percent patient reactivity and average magnitude (Fig 6). In contrast, EpKS patients followed a similar pattern but showed less consistency with

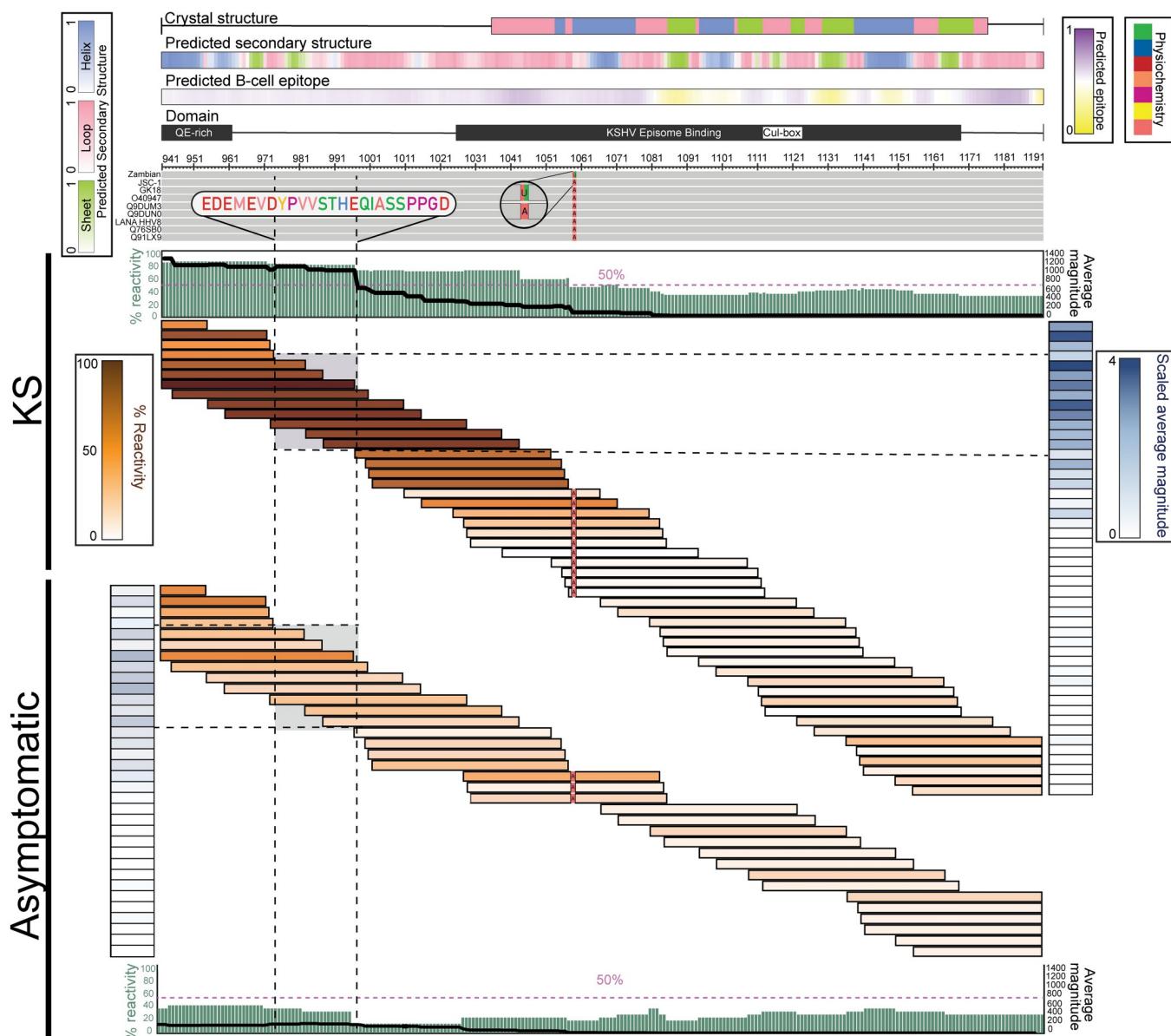


Fig 5. KSHV-LANA C-terminus epitope mapping in KS and asymptomatic individuals. The multiple sequence alignment (MSA) of reference sequences followed by the input library sequences is shown where gray represents conservation among the sequences. Mismatched residues are colored by their physiochemical properties as defined by the Zappo color scheme (green: hydrophilic, salmon: aliphatic/aromatic, orange: aromatic, fuchsia: conformationally special, yellow: cysteine only, red: negatively charged, blue: positively charged). Predicted secondary structures are represented in color gradients representing the probabilities (minimum [0; white] and maximum [1; sheet (green), loop (pink), or helix (blue)]). BepiPred-2.0 scores representing predicted B-cell epitope probabilities, along with functional motifs/domains are annotated directly above the aligned sequences. In addition, each peptide that was targeted by at least one individual was mapped and colored based on the percent cohort reactivity (brown gradient). Adjacent to the peptide responses of each group, scaled average magnitudes are shown in blue gradient. Moreover, percent patient reactivity and average magnitude of responses are calculated at the amino-acid level, represented by the sage bars and black lines in each group, respectively. LANA₉₇₄₋₉₉₇ is highlighted and the amino acid sequence is shown. **QE-rich:** Glutamine/Glutamic acid-rich region, **Cul-box:** associated with Cullin-5.

<https://doi.org/10.1371/journal.ppat.1011033.g005>

lower, more uniform magnitude (Fig 6). This suggests that in the absence of HIV infection and recognized immune suppression, EnKS patients were able to develop a more specific, focused response that would not have been evident without the epitope mapping. Moreover, the HIV⁺ asymptomatic individuals followed the pattern of EnKS patients and had a more focal response in the N-terminus, while the HIV⁻ asymptomatic controls had scattered

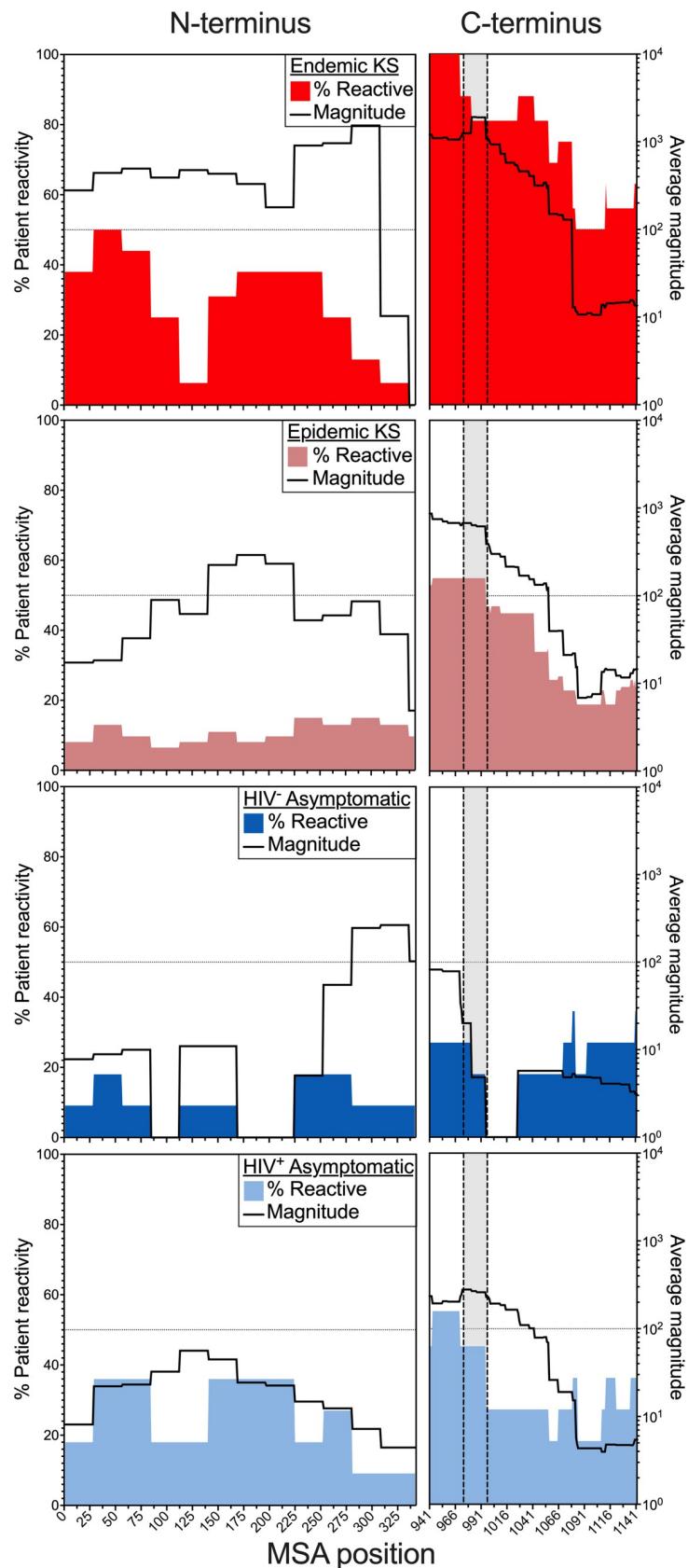


Fig 6. Comparison of percent patient reactivity and the average magnitude of Ab response per residue by KS and HIV status. The colored areas represent the percent of patients that were reactive to at least one peptide containing that residue, while the black lines represent the log-scaled average magnitude [$\log(-\log(p))$] at that residue for EpKS, EnKS, HIV⁺ asymptomatic, and HIV⁻ asymptomatic individuals in both the N- and C-termini. The vertical bars highlighted within the dotted lines span the epitope, LANA₉₇₄₋₉₉₇.

<https://doi.org/10.1371/journal.ppat.1011033.g006>

patient-specific reactivity across the N-terminus (Fig 6). However, there was no significant difference in either breadth or magnitude between the HIV⁻ and HIV⁺ asymptomatic individuals (S3 Fig) suggesting that HIV infection had a subtle effect in the absence of KS.

Importantly, all four groups had more consistent responses to the C-terminus compared to the N-terminus (Fig 6). EnKS patients had the highest reactivity across the C-terminus as denoted by a very consistent response (Fig 6) and a significantly greater breadth (S3 Fig). The magnitude of responses was also significantly elevated in the EnKS patients compared to HIV⁻ asymptomatic controls, suggesting that in the absence of HIV infection, the anti-LANA C-terminal response is elevated in disease (S3 Fig). Furthermore, EnKS patients responded more consistently to the first half of the C-terminus than the second half, as did EpKS patients and HIV⁺ asymptomatic individuals (Fig 6), while the HIV⁻ asymptomatic controls had an inconsistent, low response. The highly antigenic public epitope LANA₉₇₄₋₉₉₇ identified in the KS and asymptomatic groups (Fig 5) had the best combination of high percent patient reactivity and average magnitude across the four groups, but LANA₉₄₁₋₉₇₃ should also be noted as a region of high responsiveness (Fig 6).

In summary, the overall pattern of responsiveness was similar in EpKS, EnKS, HIV⁺ asymptomatic, and HIV⁻ asymptomatic individuals (Fig 6). The first half of the C-terminus (LANA₉₄₁₋₉₉₇) was the most likely target of LANA antibodies in patient plasma regardless of HIV and KS status (Figs 5 and 6). Nonetheless, EnKS patients showed the highest reactivity, and the HIV⁻ asymptomatic controls showed the lowest reactivity across both termini (Fig 6). Thus, HIV infection in the context of KS appeared to hamper the Ab response, while HIV⁺ asymptomatic individuals had stronger, more consistent anti-LANA responses that better mirrored those detected in KS patients compared to HIV⁻ asymptomatic controls. Overall, these high-throughput phage display data demonstrate that HIV co-infection and disease (KS) do have effects on the overall anti-LANA response and highlight the value of epitope mapping to elucidate the breadth and depth of the complete KSHV Ab repertoire in infection and neoplastic disease.

Discussion

Using a novel and high-throughput phage-display technology, VirScan, we generated the first high-resolution epitope map of the KSHV-LANA protein using cohorts representing KS and asymptomatic KSHV infection, each with and without HIV co-infection. VirScan supported fine resolution epitope mapping of the N- and C-terminal domains of LANA, which recapitulates previous work, but yields higher resolution mapping that demonstrates the utility of this phage display technique to identify antigenic epitopes with therapeutic potential in other KSHV proteins. The beginning of the LANA C-terminus immediately following the central repeat region, LANA₉₄₁₋₉₉₇, was consistently recognized in both KS and asymptomatic KSHV-infected individuals, with or without HIV co-infection, suggesting it is the predominant epitope responsible for LANA's antigenicity. Although KS patients had higher reactivity compared to asymptomatic individuals, the two cohorts had a similar pattern of reactivity across the LANA protein. Importantly, while it has been previously shown that KS patients have higher KSHV-specific Ab titers [7–10], we demonstrated that total KSHV Ab titer had a weak-

to-moderate association with the breadth or magnitude of the anti-LANA response, respectively.

Our findings support several prior studies that utilized recombinant proteins to characterize the overall reactivity to LANA and the antigenicity of its three constituent domains. For example, Labo, et al. systematically expressed and purified 72 KSHV ORFs in recombinant systems and found LANA (ORF73) to be the most consistently targeted KSHV protein [2]. Similarly, Zheng, et al. used a protein microarray to compare the Epstein-Barr Virus (EBV) and KSHV humoral responses among healthy controls, EpKS patients, and lymphoma patients [12]. Due to the inability to express the central repeat region in yeast, the N- and C-termini of LANA were expressed separately on the microarray slide, enabling comparisons of reactivity between the two regions. The C-terminal region of LANA was not only highly antigenic, but it was recognized considerably more frequently than the N-terminus, as in our study [12]. Lastly, in an effort to design a sensitive assay to screen KSHV-infected individuals, Olsen, et al. measured the reactivity to 17-mer biotinylated peptides derived from LANA using ELISA [13]. Although their assays did not reach the desired sensitivity, they observed higher reactivity against the LANA central repeat and C-terminal regions compared to the N-terminus.

Although anti-LANA breadth indicated that the central repeat region was highly antigenic, individual peptide sequences were not consistently targeted and the peptides from this region could not be confidently mapped due to the inability to discriminate recognition of one iteration of a repeat from another. For example, the most consistently targeted peptide within the repeat region was only recognized by 56% of KS patients, whereas the most consistently targeted peptide containing LANA₉₇₄₋₉₉₇ was recognized by 74%. Interestingly, LANA₉₇₄₋₉₉₇ is unique to KSHV LANA, sharing no sequence similarity with proteins from other viral or non-viral species. Since epitopes primarily reside on the surface of proteins, they are found often in random coils which promote exposure [14], and numerous studies on the sequence composition of B cell epitopes have demonstrated that epitopes are enriched in regions composed of polar and charged amino acids [15]. Consistent with these concepts, LANA₉₇₄₋₉₉₇ predominantly consists of polar and charged amino acids and is predicted to adopt random coil structure (~63%). Moreover, LANA₉₄₁₋₉₉₇, was consistently recognized in both KS patients and asymptomatic controls, regardless of HIV co-infection. High-resolution epitope mapping revealed a subtle increase in anti-LANA reactivity in HIV co-infected asymptomatic individuals compared to those that were infected with KSHV alone, an aspect of the response profiles that was otherwise indistinguishable. This increase could result from immune activation and/or increased KSHV lytic reactivation in HIV⁺ individuals. An opposite trend was observed in KS, where the HIV⁻ EnKS patients exhibited higher anti-LANA responses. The lower response to LANA₉₄₁₋₉₉₇ in EpKS patients may be explained by the progression of HIV infection to an HIV-associated malignancy prior to initiation of ART or the continued disease progression and immune suppression in untreated subjects within the cohort. Of note, prior immunosuppression is more likely to affect the Ab response, and as this study used a retrospective cohort, we did not have access to the participants' nadir CD4 counts. However, there was a weak-to-moderate correlation of breadth and magnitude with CD4 count (n = 53) measured at the time of sample collection, suggesting that reconstituted CD4 cells support a stronger anti-KSHV Ab response (S4 Fig). In a small subset of samples with HIV viral load available and detectable (n = 16), we observed no substantial association of breadth and magnitude of responses with HIV viral load (S4 Fig).

Phage display has long been used to map epitopes of proteins, but most often, short peptides (<20 amino acids) that lack potential to form secondary structure are expressed. The high-throughput phage display library used in this study has partially addressed this drawback by expressing 56-mer peptides with 28 amino acid overlap to tile the entire LANA protein.

The ideal peptide loaded on a BCR is shorter than the 56-mer peptides used in this study, however, longer peptides enable autonomous folding and adoption of pseudo-conformational epitopes. Thus, the peptides in this study are likely to have adopted more native conformations, but since only continuous epitopes were displayed, epitopes requiring tertiary or quaternary structure would have been missed in our analysis. Additionally, as in other phage display approaches, the linear T7-phage expressed peptides were amplified in bacterial cells, thus it is likely they will lack appropriate post-translational modifications. Therefore, our study is limited to analyzing linear epitopes that lack post-translational modifications. Nonetheless, there are several examples where linear epitopes can be effectively used to induce protective humoral Ab responses [16], and our findings clearly recapitulate those of previous LANA antigenicity studies.

In summary, a high-throughput phage library expressing systematically derived peptides from KSHV LANA was successfully used to generate a fine resolution epitope map and elucidate the epitopes driving LANA's high antigenicity. Further studies are warranted for investigating why LANA₉₄₁₋₉₉₇ is highly antigenic and if it acts as a potential decoy mechanism or has any therapeutic or diagnostic potential. Additionally, we intend to apply this technique to characterize the linear epitopes in the remainder of the KSHV proteome which will have broader implications, such as the identification of epitopes that can potentially be used to induce protective humoral immune responses.

Materials and methods

Ethics statement

The Institutional Review Boards of the University of Nebraska, the Louisiana State University Health Sciences Center-New Orleans, the University of Zambia Biomedical Research Ethics Committee, Tanzania National Institute for Medical Research, and the Ocean Road Cancer Institute approved the study. Written informed consent was obtained from all participants.

Sample collection

KS patients and asymptomatic KSHV-infected individuals were recruited from Zambia and Tanzania, as described previously [17–20]. Briefly, informed consent was obtained, whole blood samples were collected in EDTA tubes (BD), and plasma was isolated by centrifugation (545xg for 15 minutes).

KSHV serology

To determine KSHV serostatus and titer of the anti-KSHV Ab, we performed a monoclonal-enhanced immunofluorescence assay (mIFA) using stimulated BC3 cells, as previously described [7–10]. Briefly, BC3 cells were fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100, and spotted on Teflon-coated 12-well slides. Plasma was heat-inactivated at 56°C for 30 minutes and diluted at 1:40 in phosphate-buffered saline (PBS) for the determination of serostatus. If the plasma sample was positive at 1:40 dilution, further two-fold dilutions were carried out to ascertain the end-point titer. The appropriate dilution of plasma was added to the slide and incubated at 37°C for 30 minutes, followed by the addition of mouse anti-human IgG (CRL 1786, American Type Culture Collection) and Cy2-conjugated donkey anti-mouse IgG. Additionally, Evans blue was used as a reference stain to visualize cells. Plasma was considered positive if two of three independent readers called the sample positive.

HIV viral load

HIV-1 plasma viral load was measured as previously described [21,22]. Briefly, the QIAamp viral RNA mini kit with on-column DNase I treatment (Qiagen, cat #52904) was used to extract viral RNA from plasma. HIV-1 LTR was quantified using qPCR and AcroMetrix HIV-1 High Control samples for standards (Thermo Scientific, cat #964001). HIV-1 copies/mL was calculated based on the standard curve in QuantStudio Design and Analysis Software (ThermoFisher). Statistical analyses were performed in GraphPad Prism v9.3.1.

Phage Immunoprecipitation and Sequencing (PhIP-Seq)

The T7-Vir3 phage library was kindly provided to us by Dr. Stephen J. Elledge at Harvard Medical School. The library has been previously described [5]. Once the library was received, it was amplified and titered as directed by the manufacturer using BLT5403 (Novagen T7Select System) [5,6,23]. Total IgG was quantified using ELISA (Invitrogen, #BMS2091) according to the manufacturer's instructions. PhIP-Seq protocol was performed as previously described [5,6] and consists of three major steps: complex formation, immunoprecipitation, and library DNA preparation. Eight mock precipitations (mock IPs; PBS) were run on each plate, and each sample was run in duplicate to ensure reproducibility. First, the plasma samples were diluted to 0.5 μ g/ μ L total IgG and 96-well deep well plates were blocked with 3% BSA in TBST (Tris-buffered saline, 0.1% Tween-20). The phage library was thawed at 4°C and diluted to 2 \times 10¹⁰ pfu/mL, such that 10⁵ pfu per library member would be present for each 1mL reaction. The library was distributed to each well of the blocked deep well plate, and 2 μ g of IgG (4 μ L of 0.5 μ g/ μ L dilution) from each plasma sample was added to the appropriate well. This mixture of phage library and plasma was incubated rocking end-over-end at 4°C for 20 hours. Next, 40 μ L of a 1:1 mixture of Protein A and Protein G Dynabeads (Invitrogen, 10008D/9D) was added to each well. The mixture of phage library, plasma, and magnetic beads was incubated at 4°C for 4 hours rocking end-over-end. Then, we placed the plate on a magnetic separation rack (NEB, S1511S) and waited for 2 minutes. After the beads were sequestered, we removed the supernatant consisting of unbound phage and added 400 μ L of PhIP-Seq Wash Buffer (150 mM NaCl, 50 mM Tris-HCl, and 0.1% (vol/vol) Triton X-100, pH 7.5). This was repeated for a total of three washes, and after the second wash, the bead mixture was transferred to a fresh blocked 96-well deep well plate. After the third wash, the bead mixture was resuspended in 40 μ L of nuclease-free water and transferred to a 200 μ L PCR plate. The bead mixture was then heated to 95°C for 10 minutes in a thermocycler, and then stored at -20°C until the next step.

The DNA was prepared for sequencing using three rounds of PCR. The first round performed 30 cycles of amplification, and the second round utilized eight cycles of amplification to add the Illumina adaptors and barcode each well of the 96-well plate. After the second round of PCR, quantitative PCR was performed to ensure that each well had a similar amount of DNA. Once it was determined that the DNA amount was consistent across wells, the bar-coded samples were pooled into a single tube and subjected to a third round of PCR consisting of a single round of amplification using replenished primers to ensure the library was of uniform length. The third round PCR product was gel purified using the E.Z.N.A. Gel Extraction Kit (Omega Bio-tek, D2500-02). Finally, the gel-purified DNA was sent to the UNMC Genomics core for quality check and sequencing.

The final pool of gel-purified DNA was quantified using Qubit DS DNA HS Assay reagents and a Qubit Fluorometer (Life Technologies), and the size was confirmed using an Agilent 2100 Bioanalyzer. Each of the gel-purified, barcoded DNA pools was the correct amplicon size (\approx 376bp) and of sufficient DNA concentration. The custom Read1 and Index1 primers were diluted to 0.3 μ M with sequencing buffer (following the Illumina custom primer guide) and

added to the reagent cartridge. The DNA pool was denatured with 0.2N NaOH and adjusted to a final concentration of 1.4pM. The clustering and sequencing were performed on an Illumina NextSeq550 using the 50-cycle, single-end protocol (Mid-output flow-cell). The run was monitored by Illumina Sequence Analysis Viewer and the final FASTQ files were generated after de-multiplexing.

Data processing and analysis

Phage library annotation. The database for the phage library includes annotations for each peptide. These include, but are not limited to, the amino acid sequence, taxonomic identification from strain to kingdom level, protein names, as well as their functional descriptions. Each peptide is assigned a unique ID based on the oligonucleotide sequences, such that entries with identical amino acids can be identified. Duplicate rows, and peptides corresponding to obsolete, redundant, or deleted UniPROT IDs were either removed from the database, or their accession IDs were updated based on the UniProtKB, where possible. Peptides with duplicate amino acid sequences were apparent in the phage library. This was an artifact of including multiple isolates of the same protein where the sequence was identical for a subset of peptides. Thus, another round of curation was applied to the peptides with duplicate amino acid sequences since Abs would not have been able to distinguish between them. The raw read counts for these duplicate peptides were summed, ensuring that the Ab responses quantified represented unique phage-Ab interactions.

Analysis of PhIP-Seq data

A reference fasta file that only contained unique oligonucleotides was built, and a bowtie index was generated in preparation for aligning the sequencing reads to the library reference sequences. The PhIP-stat tool (<https://github.com/lasersonlab/phiP-stat>) was used to preprocess the raw PhIP-seq data. First, the raw counts for each phage were normalized. A generalized linear model using the Gamma-Poisson distribution was applied to obtain the $-\log_{10}(\text{residual } p\text{-value})$, MLXP, representing the frequency with which a peptide is targeted. Thus, the MLXP is the magnitude of a phage-Ab response and will be referred to as such throughout this study. Pearson correlation coefficients (ρ) between each subject's paired plasma replicates were calculated using the raw count data, and samples with $\rho < 0.7$ were excluded from downstream analyses.

A given peptide was considered reactive if and only if its magnitude was > 1.3 (i.e., $p < 0.05$) in both replicates. A presence-absence matrix was generated, where a value of one indicated the plasma had an Ab that reacted to the given peptide, and a value of zero indicated a lack of response to the given peptide. The sum of reactive peptides per sample represents the breadth of response. This dichotomization of the data provides a broad overview of patient reactivity against LANA. Subsequently, for each peptide, the magnitude was averaged across the replicates, representing an individual's magnitude of response against the peptide. Peptides that were non-reactive across all individuals or mock IPs were filtered out. Furthermore, low-quality peptides (i.e., peptides that showed reactivity in $> 25\%$ of the mock IPs) were filtered out to eliminate spurious non-specific binding events. All statistical analyses were performed in GraphPad Prism v.9.3.1, and additional downstream analyses were performed using the R programming language. All relevant data is included in the supplementary information ([S1 Dataset](#)).

Epitope mapping

Full-length protein sequences of KSHV-LANA variants presented in the library were obtained from the Protein Knowledgebase (UniProtKB). The reference sequences used were the

Human herpesvirus 8 strain GK18 (GenBank Accession with version number: AF148805.2; NCBI Reference Sequence: NC_009333.1) [24,25], Human herpesvirus 8 strain JSC-1 clone BAC16 (GenBank Accession with version number: GQ994935.1) [26]. MView was used to generate a 70% sequence identity consensus of the LANA sequences of Zambian patient-derived KSHV [27,28].

The COBALT multiple sequence alignment (MSA) tool was used to generate LANA reference and library MSAs, which highlight the conserved N- and C- termini, and the central repeats. The N- and C-terminal regions were mapped for corresponding reactive peptides using their relative alignment loci. Due to internal repeat region variability across different LANA isolates, the relative start and end locations of the C-terminal region were calculated based on the MSAs. A relative start (or end) site of a reactive peptide was defined by subtracting the start (or end) location of the MSA for the specific LANA sequence from the start location of the peptide itself. After sorting by their relative start sites, the peptides were realigned using COBALT, employing high gap introduction and extension penalties, but without query clusters under advanced parameter options. The outputs were exported in scalable vector graphics (SVG) format for downstream annotation and data visualization.

On both the peptide and amino acid levels, percent patient reactivity and average magnitude of Ab responses were measured. The peptides mapped to the MSAs were displayed as a heatmap to visualize the range of percent patient reactivity against each peptide. For each peptide against which an Ab response was evident, the average magnitude of the Ab response among all responsive patients was calculated. Single amino acid polymorphisms were highlighted in the MSA based on color designations representing the physiochemical properties of the amino acid (Zappo color scheme—green: hydrophilic, salmon: aliphatic/aromatic, orange: aromatic, fuchsia: conformationally special, yellow: cysteine only, red: negatively charged, blue: positively charged—JalView [29]). Where available, the secondary structure defined by previously characterized crystal structure was annotated on the alignment (PDB: 4UZB) [11], along with the predicted secondary structures derived from PredictProtein—Secondary Structure feature (RePROF) [30,31].

Functional motifs and empirically determined domains of LANA were annotated. These include the Histone 2A/B binding domain (the chromosome binding motif), nuclear localization signals (NLS) [32–34], the SUMO-interacting motif (SIM) [33,35], the Elongin B and C box (BC-box) [33,36], and the proline- and the aspartic-glutamic acid-rich regions within the N-terminal region, along with the glutamine-glutamic acid-rich region and the Cullin 5 box (Cul-Box) within the DNA binding domain (DBD) on the C-terminus [33,36]. Finally, results from a B-cell epitope predictor were included using default parameters (BepiPred-2.0) [14], where the epitope prediction scores are represented in probabilities mapped on the alignments as a heatmap.

Supporting information

S1 Dataset. The relevant data used for analysis throughout this study. The sample metadata, LANA-peptide reactivities, and the corresponding magnitude of each reactive peptide are included. Reactive peptides are denoted by “1” and non-reactive peptides by “0”. The sum of reactive peptides per sample represents the breadth of response. The magnitude of response (i.e., $-\log(p)$ obtained by the Gamma-Poisson model) represents the frequency with which a reactive peptide was targeted. The UniProtKB accession, amino acid sequence, and LANA protein region are provided for each peptide.

(XLSX)

S1 Fig. Representation of KSHV-LANA peptides in the input VirScan library. Each plot represents all the peptides from a UniProt entry, and each bar represents the log-scaled counts of a given peptide present in the library from the corresponding UniProt entry. The green bars represent the peptides filtered out due to significant binding in the mock-IPs. The missing peptides (no label on the x-axis) were non-unique amino acid sequences and were thus merged with the peptide that had the matching amino acid sequence.
(TIF)

S2 Fig. The relationship between KSHV Ab titer and KSHV-LANA breadth and magnitude. The KSHV Ab titer was determined using mIFA and categorized: low (1:40, 1:80, 1:160), medium (1:320, 1:640, 1:1280), and high (1:2560, 1:5120, 1:10240). (A) LANA Breadth, the number of reactive KSHV-LANA peptides per patient, was compared between titer levels (*Kruskal-Wallis*) and correlated with titer (*Spearman*). (B) LANA magnitude, the average magnitude (MLXP) of the reactive KSHV-LANA peptides per patient, was compared between titer levels (*Kruskal-Wallis*) and correlated with titer (*Spearman*). The 95% confidence interval of the linear regression is shown. Significance levels are indicated with asterisks, where $p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$, $p < 0.0001^{****}$.
(TIF)

S3 Fig. KSHV-LANA breadth and magnitude compared between EnKS, EpKS, and HIV⁻/HIV⁺ asymptomatic individuals. The number of reactive peptides and their magnitudes per individual was calculated and compared across disease states, in each group, and across each region within LANA, respectively. Comparison of (A) KSHV-LANA breadth and (B) magnitude across KS and asymptomatic individuals (*Mann-Whitney*) and 2D visualization of the PCA plot. Comparison of (C) KSHV-LANA breadth and (D) magnitude across EnKS, EpKS, KSHV⁺/HIV⁻ and KSHV⁺/HIV⁺ individuals within each region (*Kruskal-Wallis*). Significance levels are indicated with asterisks, where $p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$, $p < 0.0001^{****}$.
(TIF)

S4 Fig. Associations of KSHV-LANA breadth and magnitude with HIV plasma viral load and CD4 counts in HIV⁺ KS/Asymptomatic individuals. Both KSHV breadth and magnitude of Ab responses were tested for association with (A) CD4 count (n = 53) and (B) HIV viral load (n = 41).
(TIF)

Acknowledgments

The authors acknowledge the generosity of the participants for making this study possible, Danielle Shea for her crucial support of the study, Dr. Steven J. Elledge for making the VirScan phage library available to us, and Dr. Alok Dhar, Dr. James D. Eudy, and the rest of the University of Nebraska DNA Sequencing Core for the sequencing data collection and initial data processing.

Author Contributions

Conceptualization: Sydney J. Bennett, John T. West, Charles Wood.

Data curation: Sydney J. Bennett, Dicle Yalcin, Sara R. Privatt.

Formal analysis: Sydney J. Bennett, Dicle Yalcin, John T. West.

Funding acquisition: Salum J. Lidenge, John T. West, Charles Wood.

Methodology: Sydney J. Bennett, Dicle Yalcin.

Project administration: Sydney J. Bennett, Owen Ngalamika, Salum J. Lidenge, John T. West, Charles Wood.

Software: Dicle Yalcin.

Supervision: John T. West, Charles Wood.

Visualization: Sydney J. Bennett, Dicle Yalcin.

Writing – original draft: Sydney J. Bennett, Dicle Yalcin.

Writing – review & editing: Sydney J. Bennett, Dicle Yalcin, Sara R. Privatt, Owen Ngalamika, Salum J. Lidenge, John T. West, Charles Wood.

References

1. Cesarman E, Damania B, Krown SE, Martin J, Bower M, Whitby D. Kaposi sarcoma. *Nat Rev Dis Prim* 2019 51. 2019; 5: 1–21. <https://doi.org/10.1038/s41572-019-0060-9> PMID: 30705286
2. Labo N, Miley W, Marshall V, Gillette W, Esposito D, Bess M, et al. Heterogeneity and Breadth of Host Antibody Response to KSHV Infection Demonstrated by Systematic Analysis of the KSHV Proteome. *PLOS Pathog.* 2014; 10: e1004046. <https://doi.org/10.1371/journal.ppat.1004046> PMID: 24675986
3. Yan L, Majerciak V, Zheng ZM, Lan K. Towards Better Understanding of KSHV Life Cycle: from Transcription and Posttranscriptional Regulations to Pathogenesis. *Virol Sin.* 2019; 34: 135. <https://doi.org/10.1007/s12250-019-00114-3> PMID: 31025296
4. Domsic JF, Chen HS, Lu F, Marmorstein R, Lieberman PM. Molecular Basis for Oligomeric-DNA Binding and Episome Maintenance by KSHV LANA. *PLOS Pathog.* 2013; 9: e1003672. <https://doi.org/10.1371/journal.ppat.1003672> PMID: 24146617
5. Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, et al. Comprehensive serological profiling of human populations using a synthetic human virome. *Science* (80-). 2015; 348: aaa0698–aaa0698. <https://doi.org/10.1126/science.aaa0698> PMID: 26045439
6. Mohan D, Wansley DL, Sie BM, Noon MS, Baer AN, Laserson U, et al. PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat Protoc* 2018 139. 2018; 13: 1958–1978. <https://doi.org/10.1038/s41596-018-0025-6> PMID: 30190553
7. Kumar P, Kuwa NY, Minhas V, Marimo C, Shea DM, Kankasa C, et al. Higher Levels of Neutralizing Antibodies against KSHV in KS Patients Compared to Asymptomatic Individuals from Zambia. Schulz TF, editor. *PLoS One*. 2013; 8: e71254. <https://doi.org/10.1371/journal.pone.0071254> PMID: 23967174
8. Minhas V, Crosby LN, Crabtree KL, Phiri S, M'soka TJ, Kankasa C, et al. Development of an Immunofluorescence Assay Using Recombinant Proteins Expressed in Insect Cells To Screen and Confirm Presence of Human Herpesvirus 8-Specific Antibodies. *Clin Vaccine Immunol*. 2008; 15: 1259–1264. <https://doi.org/10.1128/CVI.00487-07> PMID: 18508931
9. Crabtree KL, Wojcicki JM, Minhas V, Smith DR, Kankasa C, Mitchell CD, et al. Risk Factors for Early Childhood Infection of Human Herpesvirus-8 in Zambian Children: The Role of Early Childhood Feeding Practices. *Cancer Epidemiol Biomarkers Prev*. 2014; 23: 300–308. <https://doi.org/10.1158/1055-9965.EPI-13-0730> PMID: 24296855
10. Poppe LK, Kankasa C, Wood C, West JT. Relationships Between Maternal Antibody Responses and Early Childhood Infection With Kaposi Sarcoma-Associated Herpesvirus. *J Infect Dis*. 2020; 222: 1723–1730. <https://doi.org/10.1093/infdis/jiaa288> PMID: 32459337
11. Hellert J, Weidner-Glunde M, Krausze J, Lünsdorf H, Ritter C, Schulz TF, et al. The 3D structure of Kaposi sarcoma herpesvirus LANA C-terminal domain bound to DNA. *Proc Natl Acad Sci U S A*. 2015; 112: 6694–6699. <https://doi.org/10.1073/pnas.1421804112> PMID: 25947153
12. Zheng D, Wan J, Cho YG, Wang L, Chiou CJ, Pai S, et al. Comparison of humoral immune responses to Epstein-Barr virus and Kaposi's sarcoma-associated herpesvirus using a viral proteome microarray. *J Infect Dis*. 2011; 204: 1683–1691. <https://doi.org/10.1093/infdis/jir645> PMID: 21990424
13. Olsen SJ, Sarid R, Chang Y, Moore PS. Evaluation of the Latency-Associated Nuclear Antigen (ORF73) of Kaposi's Sarcoma-Associated Herpesvirus by Peptide Mapping and Bacterially Expressed Recombinant Western Blot Assay. *J Infect Dis*. 2000; 182: 306–310. <https://doi.org/10.1086/315689> PMID: 10882613

14. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* 2017; 45: W24–W29. <https://doi.org/10.1093/nar/gkx346> PMID: 28472356
15. Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol.* 2013; 53: 24. <https://doi.org/10.1016/j.molimm.2012.06.001> PMID: 22784991
16. Malonis RJ, Lai JR, Vergnolle O. Peptide-Based Vaccines: Current Progress and Future Challenges. *Chem Rev.* 2020; 120: 3210–3229. <https://doi.org/10.1021/acs.chemrev.9b00472> PMID: 31804810
17. Poppe LK, Wood C, West JT. The Presence of Antibody-Dependent Cell Cytotoxicity–Mediating Antibodies in Kaposi Sarcoma–Associated Herpesvirus–Seropositive Individuals Does Not Correlate with Disease Pathogenesis or Progression. *J Immunol.* 2020; 205: 2742–2749. <https://doi.org/10.4049/jimmunol.2000489> PMID: 32998986
18. Lidenge SJ, Tso FY, Mortazavi Y, Ngowi JR, Shea DM, Mwaiselage J, et al. Viral and Immunological Analytes are Poor Predictors of the Clinical Treatment Response in Kaposi's Sarcoma Patients. *Cancers* 2020, Vol 12, Page 1594. 2020; 12: 1594. <https://doi.org/10.3390/cancers12061594> PMID: 32560243
19. Lidenge SJ, Tran T, Tso FY, Ngowi JR, Shea DM, Mwaiselage J, et al. Prevalence of Kaposi's sarcoma-associated herpesvirus and transfusion-transmissible infections in Tanzanian blood donors. *Int J Infect Dis.* 2020; 95: 204–209. <https://doi.org/10.1016/j.ijid.2020.04.018> PMID: 32294540
20. Lidenge SJ, Tso FY, Ngalamika O, Ngowi JR, Mortazavi Y, Kwon EH, et al. Similar Immunological Profiles Between African Endemic and Human Immunodeficiency Virus Type 1–Associated Epidemic Kaposi Sarcoma (KS) Patients Reveal the Primary Role of KS-Associated Herpesvirus in KS Pathogenesis. *J Infect Dis.* 2019; 219: 1318–1328. <https://doi.org/10.1093/infdis/jiy654> PMID: 30452681
21. Lidenge SJ, Kossenkov A V., Tso FY, Wickramasinghe J, Privatt SR, Ngalamika O, et al. Comparative transcriptome analysis of endemic and epidemic Kaposi's sarcoma (KS) lesions and the secondary role of HIV-1 in KS pathogenesis. *PLOS Pathog.* 2020; 16: e1008681. <https://doi.org/10.1371/journal.ppat.1008681> PMID: 32706839
22. Tso FY, Kang G, Kwon EH, Julius P, Li Q, West JT, et al. Brain is a potential sanctuary for subtype C HIV-1 irrespective of ART treatment outcome. *PLoS One.* 2018; 13: e0201325. <https://doi.org/10.1371/journal.pone.0201325> PMID: 30040863
23. Larman HB, Zhao Z, Laserson U, Li MZ, Ciccia A, Gakidis MAM, et al. Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol.* 2011. <https://doi.org/10.1038/nbt.1856> PMID: 21602805
24. Glenn M, Rainbow L, Auradé F, Davison A, Schulz TF. Identification of a Spliced Gene from Kaposi's Sarcoma-Associated Herpesvirus Encoding a Protein with Similarities to Latent Membrane Proteins 1 and 2A of Epstein-Barr Virus. *J Virol.* 1999; 73: 6953–6963. <https://doi.org/10.1128/JVI.73.8.6953-6963.1999> PMID: 10400794
25. Rezaee SAR, Cunningham C, Davison AJ, Blackbourn DJ. Kaposi's sarcoma-associated herpesvirus immune modulation: An overview. *J Gen Virol.* 2006; 87: 1781–1804. <https://doi.org/10.1099/vir.0.81919-0> PMID: 16760382
26. Brulois KF, Chang H, Lee AS-Y, Ensser A, Wong L-Y, Toth Z, et al. Construction and Manipulation of a New Kaposi's Sarcoma-Associated Herpesvirus Bacterial Artificial Chromosome Clone. *J Virol.* 2012; 86: 9708–9720. <https://doi.org/10.1128/JVI.01019-12> PMID: 22740391
27. Olp LN, Jeanniard A, Marimo C, West JT, Wood C. Whole-Genome Sequencing of Kaposi's Sarcoma-Associated Herpesvirus from Zambian Kaposi's Sarcoma Biopsy Specimens Reveals Unique Viral Diversity. *J Virol.* 2015; 89: 12299–12308. <https://doi.org/10.1128/JVI.01712-15> PMID: 26423952
28. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 2022; gkac240–gkac240. <https://doi.org/10.1093/nar/gkac240> PMID: 35412617
29. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009; 25: 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033> PMID: 19151095
30. Bernhofer M, Dallago C, Karl T, Satagopam V, Heinzinger M, Littmann M, et al. PredictProtein—Predicting Protein Structure and Function for 29 Years. *Nucleic Acids Res.* 2021; 49: W535–W540. <https://doi.org/10.1093/nar/gkab354> PMID: 33999203
31. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* 2014; 42: W337–W343. <https://doi.org/10.1093/nar/gku366> PMID: 24799431
32. Cherezova L, Burnside KL, Rose TM. Conservation of Complex Nuclear Localization Signals Utilizing Classical and Non-Classical Nuclear Import Pathways in LANA Homologs of KSHV and RFHV. *PLoS One.* 2011; 6: e18920. <https://doi.org/10.1371/journal.pone.0018920> PMID: 21559489

33. Wei F, Gan J, Wang C, Zhu C, Cai Q. Cell Cycle Regulatory Functions of the KSHV Oncoprotein LANA. *Front Microbiol.* 2016; 7: 334. <https://doi.org/10.3389/fmicb.2016.00334> PMID: 27065950
34. Juillard F, Pires de Miranda M, Li S, Franco A, Seixas AF, Liu B, et al. KSHV LANA acetylation-selective acidic domain reader sequence mediates virus persistence. *Proc Natl Acad Sci U S A.* 2020; 117: 22443–22451. <https://doi.org/10.1073/pnas.2004809117> PMID: 32820070
35. Kerscher O. SUMO junction—what's your function? New insights through SUMO-interacting motifs. *EMBO Rep.* 2007; 8: 550. <https://doi.org/10.1038/sj.embor.7400980> PMID: 17545995
36. Cai QL, Knight JS, Verma SC, Zald P, Robertson ES. EC5S Ubiquitin Complex Is Recruited by KSHV Latent Antigen LANA for Degradation of the VHL and p53 Tumor Suppressors. *PLOS Pathog.* 2006; 2: e116. <https://doi.org/10.1371/journal.ppat.0020116> PMID: 17069461