

## Genomic variations in EBNA3C of EBV associate with posttransplant lymphoproliferative disorder

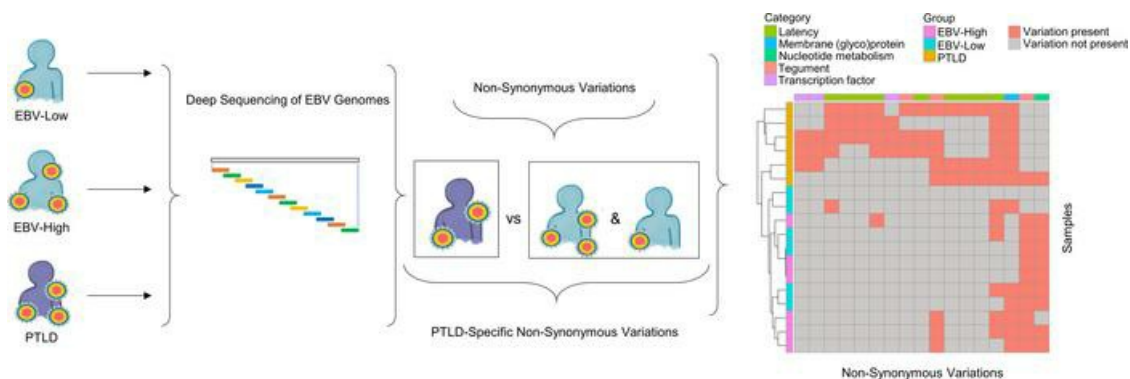
Eden M. Maloney, ... , Carlos O. Esquivel, Olivia M. Martinez

*JCI Insight.* 2020;5(6):e131644. <https://doi.org/10.1172/jci.insight.131644>.

Research Article

Transplantation

### Graphical abstract



Find the latest version:

<https://jci.me/131644/pdf>



# Genomic variations in EBNA3C of EBV associate with posttransplant lymphoproliferative disorder

Eden M. Maloney,<sup>1</sup> Vincent A. Busque,<sup>2</sup> Sin Ting Hui,<sup>2</sup> Jiaying Toh,<sup>1</sup> Marcelo Fernandez-Vina,<sup>3</sup> Sheri M. Krams,<sup>1,2</sup> Carlos O. Esquivel,<sup>2</sup> and Olivia M. Martinez<sup>1,2</sup>

<sup>1</sup>Stanford Immunology, <sup>2</sup>Division of Abdominal Transplantation, Department of Surgery, and <sup>3</sup>Department of Pathology, Stanford University School of Medicine, Stanford, California, USA.

Epstein-Barr Virus (EBV) is a ubiquitous virus linked to a variety of lymphoid and epithelial malignancies. In solid organ and hematopoietic stem cell transplant recipients, EBV is causally associated with posttransplant lymphoproliferative disorder (PTLD), a group of heterogeneous lymphoid diseases. EBV<sup>+</sup> B cell lymphomas that develop in the context of PTLD are generally attributed to the immunosuppression required to promote graft survival, but little is known regarding the role of EBV genome diversity in the development of malignancy. We deep-sequenced the EBV genome from the peripheral blood of 18 solid organ transplant recipients, including 6 PTLD patients. Sequences from 6 EBV<sup>+</sup> spontaneous lymphoblastoid B cell lines (SLCL) were similarly analyzed. The EBV genome from PTLD patients had a significantly greater number of variations than EBV from transplant recipients without PTLD. Importantly, there were 15 nonsynonymous variations, including 8 in the latent cycle gene *EBNA3C* that were associated with the development of PTLD. One of the nonsynonymous variations in *EBNA3C* is located within a previously defined T cell epitope. These findings suggest that variations in the EBV genome can contribute to the pathogenesis of PTLD.

## Introduction

Epstein-Barr Virus (EBV) is a gammaherpesvirus that has infected more than 90% of the world's population. Infection generally occurs via transmission through the saliva of a carrier and is typically asymptomatic but can cause infectious mononucleosis (IM) in adolescents and young adults. EBV is also linked to approximately 120,000 malignancies globally each year (1), and the factors that account for the distinct outcomes following infection remain unclear. EBV-associated malignancies are most commonly of B cell or epithelial cell origin, consistent with the propensity of the virus to infect these cell types, and include Burkitt lymphoma (BL), Hodgkin's lymphoma, non-Hodgkin's lymphomas, nasopharyngeal carcinoma (NPC), gastric adenocarcinoma (GC), and B cell lymphomas in immunocompromised or immunosuppressed individuals. In hematopoietic stem cell and solid organ transplant recipients, EBV<sup>+</sup> B cell lymphomas fall within a spectrum of abnormal lymphoproliferations termed posttransplant lymphoproliferative disorder (PTLD).

The 172-kb double-stranded DNA genome of EBV includes genes that are coordinately expressed during the lytic phase of infection that yields new viral particles, as well as during the latent phase of infection that promotes survival of the host cell and persistence of the viral genome. Distinct groups of latency genes that contribute to transformation and cellular proliferation are expressed in EBV-associated malignancies. EBV<sup>+</sup> B cell lymphomas in PTLD are characterized by expression of the EBV-encoded small RNAs (EBER) *EBNA1*, *EBNA2*, *EBNA3A*, *EBNA3B*, *EBNA3C*, *EBNA-LP*, *LMP1*, and *LMP2*. The latency genes are also important targets of the adaptive immune response to EBV (2). In particular, viral-specific T cells become activated and proliferate following T cell receptor (TCR) recognition of EBV-derived peptides presented by the MHC proteins expressed on antigen presenting cells. Once activated, CD4<sup>+</sup> T cells produce cytokines to support antibody production and cytotoxic T cell (CTL) proliferation and differentiation to mediate elimination of virally infected cells. In the case of EBV<sup>+</sup> PTLD, the immunosuppression required to prevent graft rejection is thought to impair the CD4<sup>+</sup> and CD8<sup>+</sup> T cells

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Copyright:** © 2020, American Society for Clinical Investigation.

**Submitted:** July 14, 2019

**Accepted:** February 26, 2020

**Published:** March 26, 2020.

**Reference information:** *JCI Insight*. 2020;5(6):e131644.  
<https://doi.org/10.1172/jci.insight.131644>.

responsible for controlling the expansion of EBV-infected B cells. Whether EBV genome diversity impacts host immunity or contributes to the development of EBV<sup>+</sup> PTLD is unknown.

The first EBV genome sequenced was derived from the B95-8 strain that originated from a patient with IM. Since then, sequencing studies delineated EBV variation on the basis of distinct strains, geography, and cancer phenotypes (3–9). The sequencing of *EBNA3s* and *EBNA2* from a variety of EBV<sup>+</sup> cell lines led to the development of the type 1 and type 2, also known as type A and type B, classification scheme of EBV strains (10). Another scheme defined the EBV sequence variants Alaskan, China 1, China 2, China 3, Med, and North Carolina (NC) based on differences within *LMPI* (11). With the advent of deep-sequencing, EBV genomes have now been sequenced from in vitro transformed EBV<sup>+</sup> lymphoblastoid B cell lines (LCL), EBV-associated tumors, and saliva of healthy individuals allowing broader analysis of the genome (12). Palser et al. (9) conducted a whole genome sequencing study of EBV isolated from tumors and cell lines, as well as one healthy donor, that focused primarily on geographic differences. EBV genome variations have been reported in viruses isolated from patients with IM, EBV malignancies such as NPC, gastric cancer, lung carcinoma, NK/T cell lymphoma, and PTLD, as well as patient-derived cell lines (4–9, 13–19). However, it has not been possible to establish specific variations characteristic of EBV-associated malignancies due to the paucity of EBV genomes from control subjects without EBV disease. Moreover, several of these studies examined EBV from cancer samples obtained from multiple geographic locations; thus, it is still unclear whether the variations are linked to disease or location (3). For instance, a variation in *EBNA1* originally thought to be linked to the disease BL is instead likely linked to a geographic location — in this case, Africa (20). In this study, we performed the first deep-sequencing of the EBV genome from the blood of transplant recipients with EBV<sup>+</sup> PTLD and PTLD-free transplant recipients to identify genomic variations associated with the cancer phenotype.

## Results

*EBV quantitation and whole genome sequencing.* EBV viral loads were determined from the blood of pediatric transplant recipients ( $n = 18$ ) (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.131644DS1>) by quantitative PCR (qPCR) targeting the *EBNA1* gene (Supplemental Figure 1). Transplant recipients with PTLD ( $n = 6$ ) had a viral load of  $3165 \pm 1956$  (mean  $\pm$  SEM) copies per  $\mu\text{g}$  of DNA. Transplant recipients without PTLD were divided into 2 groups based on EBV viral load, EBV<sup>hi</sup> ( $>1000$  copies per  $\mu\text{g}$  DNA) and EBV<sup>lo</sup> ( $<1000$  copies per  $\mu\text{g}$  DNA). EBV<sup>hi</sup> samples had a mean viral load of  $4946 \pm 3683$  copies per  $\mu\text{g}$  of DNA, and EBV<sup>lo</sup> samples had a mean viral load of  $164 \pm 212$  copies per  $\mu\text{g}$  of DNA. The EBV copy numbers were significantly higher in the PTLD group ( $P < 0.004$ ) and the EBV<sup>hi</sup> group ( $P < 0.01$ ) compared with the EBV<sup>lo</sup> group. Due to the relatively low abundance of EBV DNA compared with human DNA in the peripheral blood, the EBV genome was amplified by PCR primers that targeted nonrepetitive regions to create overlapping amplicons for whole genome sequencing (Supplemental Table 2) (13). A total of  $97 \times 10^6$  sequence reads of 150-bp end reads were generated from the combined pool of amplicons, with an average of  $3 \times 10^6$  sequence reads per sample. A total of 39%–85% of the reads mapped to the EBV genome (Table 1). All genomes were assembled into large contiguous DNA sequences using de novo assembly and did not include the repeat regions. The mean percent GC content for all groups was 59.5%, and the average read depth coverage for all samples was 2645-fold. The mean percent coverage of the EBV genome was 83.75% (range 73%–94%). Higher EBV viral load correlated to greater coverage of the genome. Thus, 18 potentially novel whole EBV genomes isolated from PTLD patients, EBV<sup>hi</sup>, and EBV<sup>lo</sup> transplant recipients were sequenced from peripheral blood.

*Regions of increased variation are present in EBV genomes isolated from PTLD patients compared with transplant recipients without PTLD.* EBV genomes sequenced from healthy individuals that were previously published (9, 18) were used as a control group in order to compare the number of variations in the EBV genome of transplant recipients to nontransplanted, healthy individuals. When compared with the reference EBV genome AJ507799, composed of a combination of sequences from B95-8 and Raji, the EBV genomes from all samples had a mean of 693 total variations. EBV genomes sequenced from the PTLD group were the most variable, with a mean of 1282 single nucleotide variations (SNV), while the EBV genomes from the control group were the least variable, with a mean of 807 SNV (Table 1). EBV genomes from spontaneous B lymphoblastoid cell lines (SLCL) derived from a second, distinct cohort of PTLD subjects had the highest number of multiple nucleotide variations (MNV) and Indel (insertions and deletions) — 128 and 366, respectively — while the control group had the lowest number of MNV and Indel — 16 and 30, respectively.

**Table 1. Summary of MiSeq sequencing of EBV genomes**

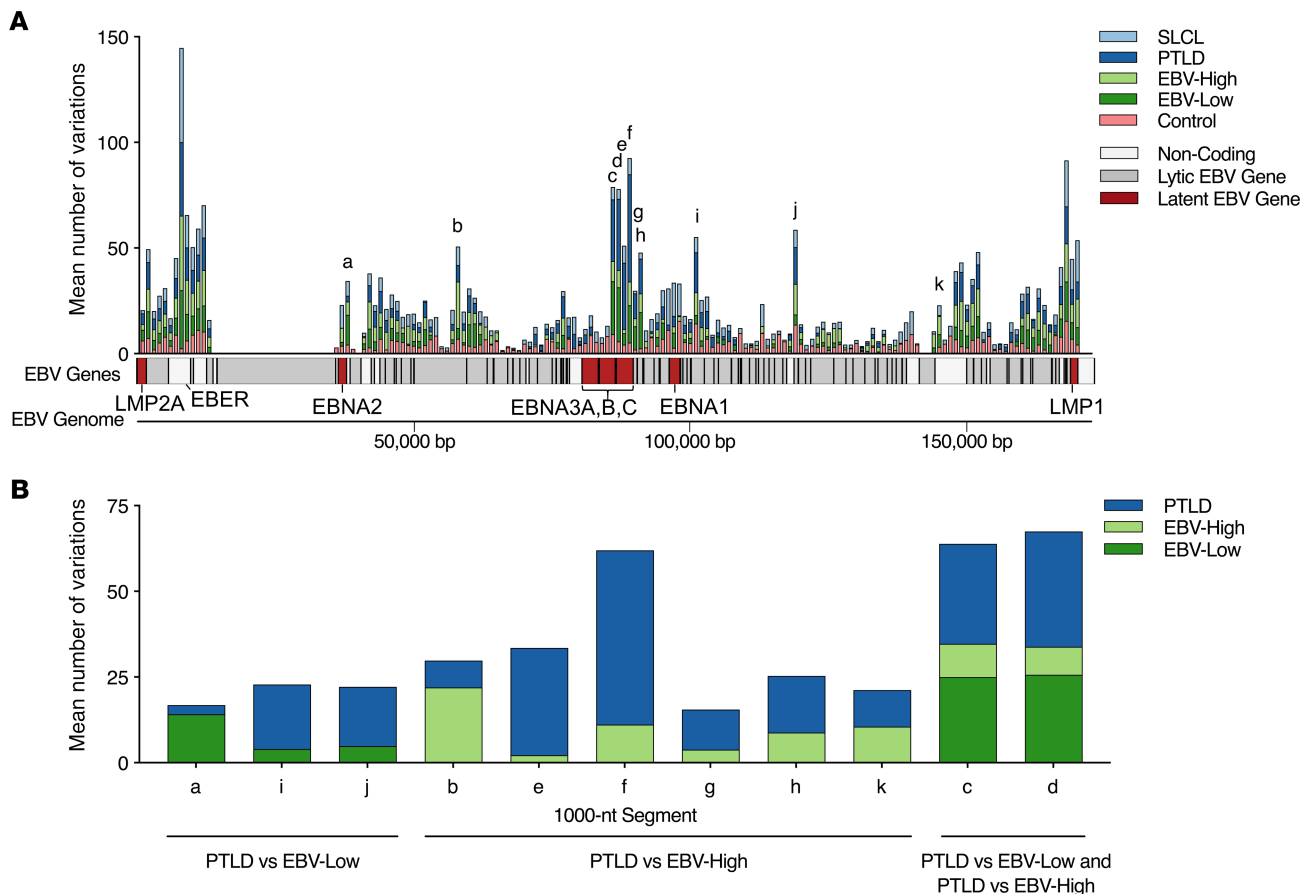
Sample group	Samples (n)	Mean % GC content	Mean % reads that mapped	Mean depth	Mean % coverage of genome	Mean SNV <sup>A</sup> / Coverage <sup>B</sup>	Mean MNV <sup>C</sup> / Coverage	Mean indel <sup>D</sup> / Coverage
SLCL <sup>E</sup>	6	59.5	85%	9817	94%	1172	128	366
PTLD	6	59.5	43%	189	79%	1282	71	128
EBV <sup>hiF</sup>	6	59.5	72%	348	89%	1011	23	89
EBV <sup>loG</sup>	6	59.5	39%	227	73%	1100	62	101
Control <sup>H</sup>	4	NA (9,18)	70%–80% (9), NA (18)	>1000 (9), 112 (18)	82%	807	16	30

<sup>A</sup>SNV, single nucleotide variations. <sup>B</sup>Normalized to number of nucleotides sequenced. <sup>C</sup>MNV, multiple nucleotide variations. <sup>D</sup>Indel, insertions and deletions. <sup>E</sup>Spontaneous lymphoblastoid cell lines derived from patients with EBV<sup>+</sup> PTLD. <sup>F</sup>EBV<sup>hi</sup> is defined as an EBV viral load of > 1000 copies per µg of DNA. <sup>G</sup>EBV<sup>lo</sup> is defined as an EBV viral load of ≤ 1000 copies per µg of DNA. <sup>H</sup>Sequences from healthy individuals previously published (9, 18).

To study the variability across the EBV genomes and the location of hotspots, the number of variations within a 1000 nucleotide sliding window for each sample group was mapped to the EBV genome from 0 bp to 172,000 bp (Figure 1A). Variations included insertions, deletions, SNV and multiple sequential single nucleotide variations (MSV). The EBV genome contains regions of high diversity, and there were differences in the number of variations between the sample groups. We compared the number of variations within the 1000 nucleotide segments in the PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups using a 2-way ANOVA with the Tukey post hoc test and identified 11 regions, designated a–k, with significant differences between groups. Genes that contain 1000 nucleotide segments with a significantly greater number of variations in the PTLD group than the EBV<sup>hi</sup> and/or EBV<sup>lo</sup> group include *EBNA3B* (Figure 1B, region c), *EBNA3C* (Figure 1B, regions b–d, f, and g), *BZLF2* (Figure 1B, region g), *BZLF1* (Figure 1B, regions g and h), *BBLF4* (Figure 1B, region i), *BDLF3* (Figure 1B, region j), *BDLF2* (Figure 1B, region j), and noncoding region 144,000–145,000 bp (Figure 1B, region k). The EBV genomes in the EBV<sup>hi</sup> group had significantly more variations in regions containing the *BPLF1*, *BDLF3*, and *BDLF2* genes (Supplemental Figure 2, regions b and j) than the EBV<sup>lo</sup> group genomes. The EBV<sup>lo</sup> group had significantly more variations in *EBNA2* than the PTLD group (Figure 1, A and B) and the EBV<sup>hi</sup> group (Supplemental Figure 2, region a), and it had more variations in the *EBNA3C* gene than the EBV<sup>hi</sup> group (Supplemental Figure 2, regions d–f). Overall, the significant difference in the number of variations within the PTLD group compared with the EBV<sup>lo</sup> and EBV<sup>hi</sup> groups indicates that there may be selective pressure in specific regions of the EBV genome associated with PTLD.

*Latency and tegument genes contain the highest number of synonymous and nonsynonymous variations.* The number of variations for each sample group was determined in the following gene categories: latency, tegument, membrane glycoprotein, unknown, replication, packaging, capsid, transcription factor, and nucleotide metabolism. Across all sample groups, the latency and tegument genes had the highest number of variations, while transcription factors and nucleotide metabolism genes had the lowest number of variations (Figure 2A). The mean number of nonsynonymous variations within the groups was similar, and there were no significant differences across groups (Figure 2B, left). Similarly, the number of synonymous variations was not significantly different across the groups (Figure 2B, right). Overall for all the groups combined, the EBV coding regions had  $172 \pm 6$  (mean  $\pm$  SEM) nonsynonymous and  $273 \pm 15$  synonymous variations (Figure 2B). The number of nonsynonymous variations in each gene category was not correlated to the number of genes in each category (Figure 2C). However, when the number of nonsynonymous variations was normalized to the total length of the genes in each gene category, the latency and tegument genes had a higher ratio of nonsynonymous variations to gene length compared with all other gene categories (Figure 2D).

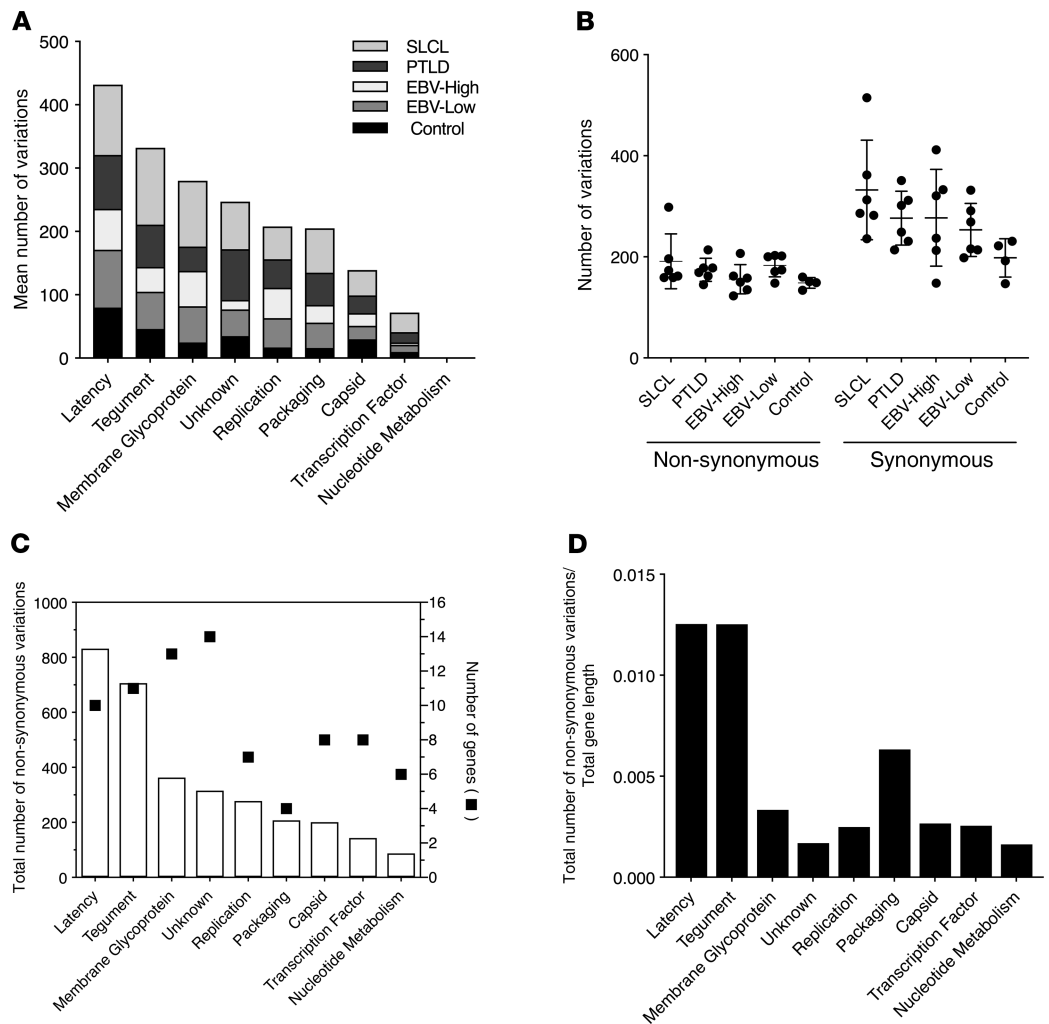
*EBV genomes isolated from PTLD patients have significantly more nonsynonymous variations in EBNA3C compared with EBV genomes isolated from PTLD-free transplant recipients.* Next, we analyzed the variability of the latency genes expressed in EBV<sup>+</sup> B cell lymphomas in PTLD. *EBNA1*, *EBNA3C*, and *LMP1* had the greatest mean number of nonsynonymous variations (Figure 3A). EBV genomes in the SLCL group had significantly more nonsynonymous variations within *EBNA1* than the EBV<sup>hi</sup> and EBV<sup>lo</sup> groups ( $P = 0.04$  and  $P = 0.04$ , respectively), but there



**Figure 1. PTLD-derived EBV genomes have significantly more variations in hotspots than EBV<sup>lo</sup> and EBV<sup>hi</sup> genomes. (A)** The number of variations within a 1000-nt sliding window was determined for each location of the EBV genome from 0–172,000 bp. Variations included insertions, deletions, SNV, and MSV. The bar graph corresponds to the number of variations in each 1000-nt window for SLCL (light blue,  $n = 6$ ), PTLD (dark blue,  $n = 6$ ), EBV<sup>hi</sup> (light green,  $n = 6$ ), EBV<sup>lo</sup> (dark green,  $n = 6$ ), and control (pink,  $n = 4$ ) groups. Gene locations are shown below: latent genes (red), lytic genes (gray), and noncoding regions (white). Segments (1000 nt) that contain significant differences in number of variations between groups are lettered a–k. **(B)** Segments (1,000 nt) of the EBV genome that contain significant differences in number of variations between PTLD and EBV<sup>hi</sup> and/or EBV<sup>lo</sup> groups. Statistical significance between number of variations from PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups was determined using 2-way ANOVA with Tukey’s multiple comparisons test. Lines on the bottom indicate significant comparisons: PTLD versus EBV<sup>lo</sup>, PTLD versus EBV<sup>hi</sup>, PTLD versus EBV<sup>lo</sup>, and PTLD versus EBV<sup>hi</sup>. Region a, includes part of *EBNA2* gene, PTLD versus EBV<sup>lo</sup>,  $P = 0.0009$ ; region i, within *BBLF4* gene, PTLD versus EBV<sup>lo</sup>,  $P = 0.001$ ; Region j, includes part of *BDLF3* and *BDLF2* genes, PTLD versus EBV<sup>lo</sup>,  $P = 0.007$ ; Region b, within *BPLF1* gene PTLD versus EBV<sup>hi</sup>,  $P = 0.02$ ; Region e, within *EBNA3C* gene, PTLD versus EBV<sup>hi</sup>,  $P < 0.0001$ ; Region f, within *EBNA3C* gene, PTLD versus EBV<sup>hi</sup>,  $P < 0.0001$ ; Region g, includes part of *EBNA3C*, *BZLF2*, and *BZLF1* genes, PTLD versus EBV<sup>hi</sup>,  $P = 0.04$ ; Region h, includes part of *BZLF1* gene, PTLD versus EBV<sup>hi</sup>,  $P = 0.04$ ; Region k, within a noncoding region, PTLD versus EBV<sup>hi</sup>,  $P = 0.02$ ; Region c, includes part of *EBNA3B* and *EBNA3C* gene, PTLD versus EBV<sup>lo</sup>,  $P < 0.0001$  and PTLD versus EBV<sup>hi</sup>,  $P < 0.0001$ ; Region d, within *EBNA3C* gene, PTLD versus EBV<sup>lo</sup>,  $P < 0.0001$ , PTLD versus EBV<sup>hi</sup>,  $P < 0.0001$ .

was no difference compared with the PTLD group (Figure 3B). However, PTLD-derived EBV genomes had significantly more nonsynonymous variations in *EBNA3C* compared with genomes from the EBV<sup>hi</sup> ( $P = 0.05$ ), EBV<sup>lo</sup> ( $P = 0.03$ ), and SLCL groups ( $P = 0.02$ ) (Figure 3F). There were no differences in the number of non-synonymous variations in *EBNA2* (Figure 3C), *EBNA3A* (Figure 3D), *EBNA3B* (Figure 3E), *LMP1* (Figure 3G), or *LMP2* (Figure 3H) across the groups. These findings indicate that genomic variations resulting in amino acid changes in the *EBNA3C* protein are particularly prevalent in EBV isolated from PTLD patients.

*Phylogenetic analysis of whole EBV genomes and the latent genes EBNA1, EBNA3B, and EBNA3C.* Phylogenetic trees were constructed to visualize the impact of variations on the relationship between whole EBV genomes and the latency genes within the groups. Nucleotide sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE) alignment, and trees were constructed using neighbor-joining model and Jukes-Cantor genetic distance modeling. Phylogenetic analysis of the entire EBV genome did not reveal any defined cluster by sample group (Figure 4A). However, phylogenetic analysis of *EBNA1* genes (Figure 4B) showed a cluster of EBV<sup>hi</sup> and EBV<sup>lo</sup> samples and a grouping of control samples, whereas *EBNA3B* (Figure 4C) and *EBNA3C* (Figure 4D) gene analysis produced groupings of

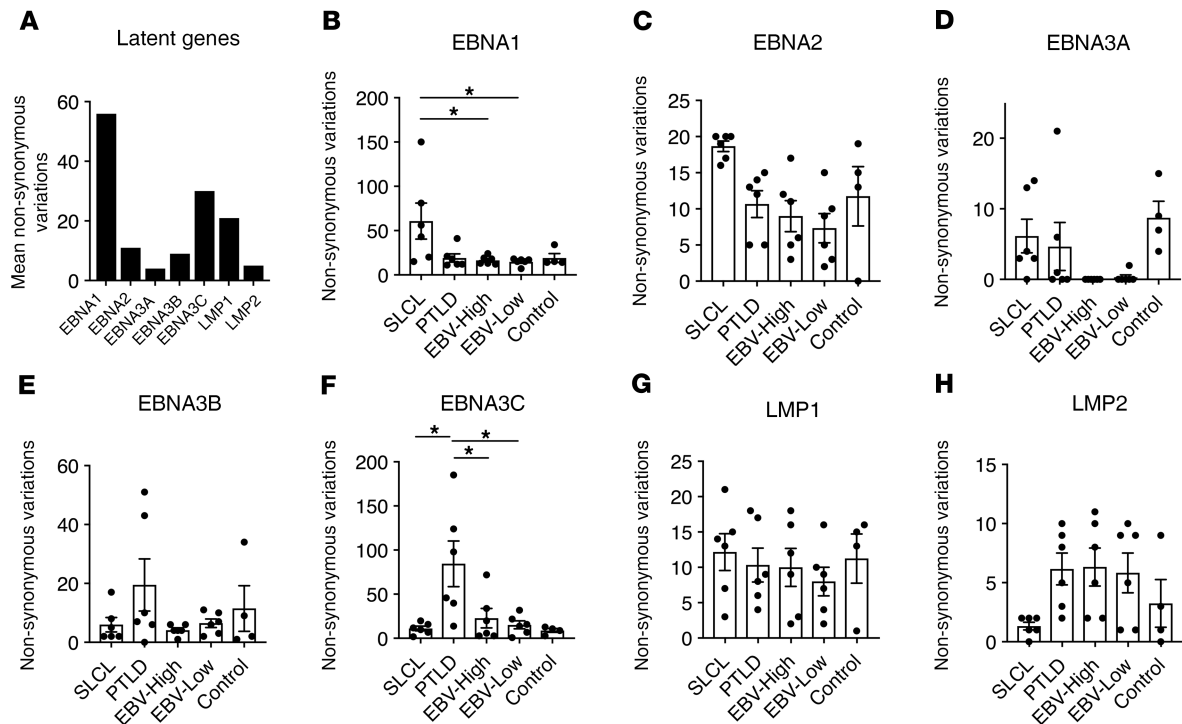


**Figure 2. Latency and tegument genes contain the highest number of synonymous and nonsynonymous variations.** (A) Mean number of variations within EBV gene categories for each group: SLCL ( $n = 6$ ), PTLD ( $n = 6$ ), EBV<sup>hi</sup> ( $n = 6$ ), EBV<sup>lo</sup> ( $n = 6$ ), control ( $n = 4$ ). (B) Number of nonsynonymous and synonymous variations in each sample group. Mean  $\pm$  SEM are shown for each group. (C) Total number of nonsynonymous variations (bars) and number of genes in each category (squares). (D) Total number of nonsynonymous variations divided by the total length of genes in each gene category. There was no statistical significance between the sample groups, as determined by 2-way ANOVA with Tukey's multiple comparisons tests.

PTLD and SLCL samples and a grouping of EBV<sup>hi</sup> and EBV<sup>lo</sup> samples. The other latency genes, *EBNA2*, *EBNA3A*, *LMP1*, and *LMP2*, did not generate group-specific clades (Supplemental Figure 3). Thus, there are variations within the *EBNA3B* and *EBNA3C* genes that are found more frequently in EBV from PTLD and SLCL samples compared with EBV from EBV<sup>hi</sup> and EBV<sup>lo</sup> samples.

We next applied 2 well-established EBV classification schemes to evaluate the relationship of the EBV genomes isolated from PTLD subjects and transplant recipients. Using the scheme of Edwards et al. (11), which focuses on *LMP1* sequences, we determined that 3 of the 6 PTLD samples are China 1, while 1 each are NC, B95-8, and Med. A phylogenetic analysis of *LMP1* sequences indicates that the China 1 and NC strains form a distinct cluster from the B95-8 and Med strains (Supplemental Figure 3C) such that the majority of the PTLD samples (4 of 6) group together in the clade that includes the China 1 and NC strains. Using the type 1 and type 2 classification scheme (10) that is based on *EBNA2* and *EBNA3* sequence diversity, we determined that 6 of 6 PTLD EBV genomes and 6 of 6 EBV<sup>hi</sup> EBV genomes were of the type 1 strain, while 5 of 6 EBV genomes from the EBV<sup>lo</sup> group were type 1 and the sixth was type 2.

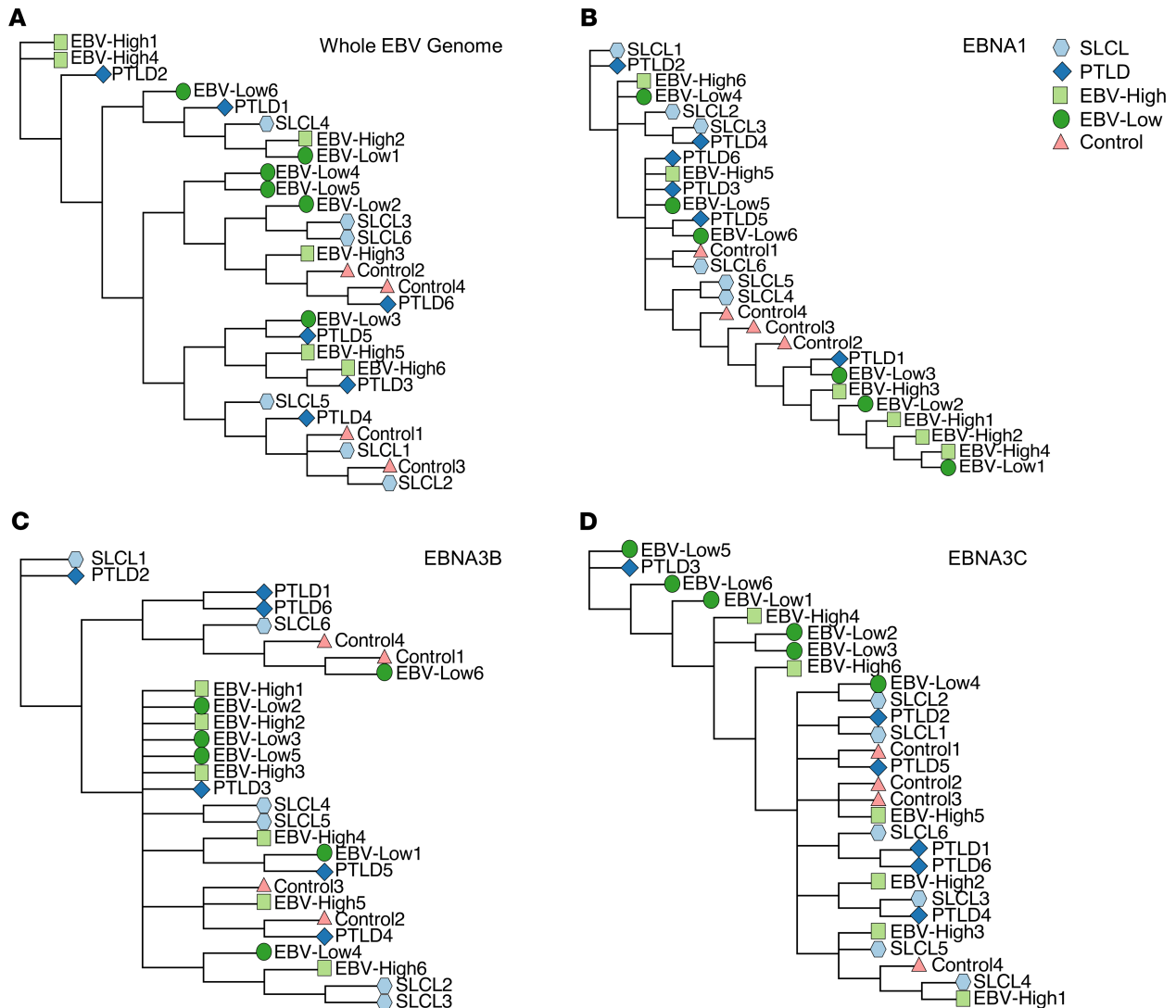
*Identification of nonsynonymous variations associated with PTLD.* To determine the variations in the EBV genome associated with PTLD, we performed Fisher's exact test. We identified 15 nonsynonymous variations



**Figure 3. EBV genomes from PTLTD subjects have more nonsynonymous variations in *EBNA3C* than EBV genomes from PTLTD-free subjects.** SLCL ( $n = 6$ ), PTLTD ( $n = 6$ ), EBV<sup>hi</sup> ( $n = 6$ ), EBV<sup>lo</sup> ( $n = 6$ ), control ( $n = 4$ ) groups are shown. (A) Mean number of nonsynonymous variations within EBV latent genes. (B) Nonsynonymous variations within *EBNA1* for each sample group. SLCL versus EBV<sup>hi</sup>,  $P = 0.04$ ; SLCL versus EBV<sup>lo</sup>,  $P = 0.04$ . (C) Nonsynonymous variations within *EBNA2*. (D) Nonsynonymous variations within *EBNA3A*. (E) Nonsynonymous variations within *EBNA3B*. (F) Nonsynonymous variations within *EBNA3C*. PTLTD versus SLCL,  $P = 0.02$ ; PTLTD versus EBV<sup>hi</sup>,  $P = 0.05$ ; PTLTD versus EBV<sup>lo</sup>,  $P = 0.03$ . (G) Nonsynonymous variations within *LMP1*. (H) Nonsynonymous variations within *LMP2*. (B–H) Mean  $\pm$  SEM are shown for each group. Statistical significance was determined by 2-way ANOVA with Tukey's multiple comparisons test. \* $P < 0.05$ .

that were significantly associated with the PTLTD group compared with the EBV<sup>lo</sup> and EBV<sup>hi</sup> groups. A heatmap with unsupervised analysis was used to visualize the distribution of the PTLTD-associated variations in all the samples (Figure 5). The clustering resulted in distinct grouping of the PTLTD samples from the EBV<sup>hi</sup> and EBV<sup>lo</sup> samples (Figure 5A and Table 2). There were no samples in the EBV<sup>hi</sup> or EBV<sup>lo</sup> groups that contained an equal number or more PTLTD-associated variations than any PTLTD sample, indicating the strength of the association of these variations. Moreover, the samples within the PTLTD group had significantly more PTLTD-associated variations (range 6–12) than the EBV<sup>hi</sup> (range 0–3) ( $P = 0.0007$ ) and EBV<sup>lo</sup> (range 0–3) ( $P = 0.0006$ ) groups (Figure 5B). The PTLTD-associated variations were identified in latency (*EBNA3C* and *LMP2A*), membrane glycoprotein (*BDLF3*), tegument (*BDLF2* and *BPLF1*), and transcription factor (*BARF1* and *BMLF1*) genes (Figure 5A), with the majority in the latency genes (*EBNA3C*,  $n = 8$ ; *LMP2*,  $n = 1$ ). In contrast, 2 genes, *BOLFI* (tegument) and *BLLF3* (nucleotide metabolism), contained non-PTLTD-associated (EBV<sup>hi</sup> and EBV<sup>lo</sup>) variations. Nine of the PTLTD-associated variations, including 6 within *EBNA3C*, were also present in the SLCL group (Figure 5C). Thus, we identified a series of nucleotide variations resulting in amino acid changes in EBV-encoded proteins, some of which are restricted to transplant recipients that develop PTLTD.

The PTLTD-associated variation (*Arg887Thr*) is located within a defined *EBNA3C* peptide epitope and decreases the predicted binding affinity of the peptide to HLA-B7. To determine whether these protein sequence changes could affect immune recognition of EBV, we compared the PTLTD-associated variations to previously established CD8<sup>+</sup> EBV-specific T cell epitopes (21). Two variations were identified within known T cell epitopes, an arginine to threonine change at position 887 of *EBNA3C* (*EBNA3CArg887Thr*) located within the HLA-B7 epitope QPRAPIRPI, and a serine to threonine change at position 444 of *LMP2A* (*LMP2A Ser444Thr*) located within the HLA-A25 epitope VMSNTLLSAW. Using NetMHC 4.0 Server, we calculated the predicted binding affinity concentration to HLA-B7 of the reference peptide QPRAPIRPI of *EBNA3C* and the PTLTD-associated variant *EBNA3CArg887Thr* (Figure 6A). The *EBNA3CArg887Thr* change resulted in an increase in binding affinity concentration for QPRAPITPI

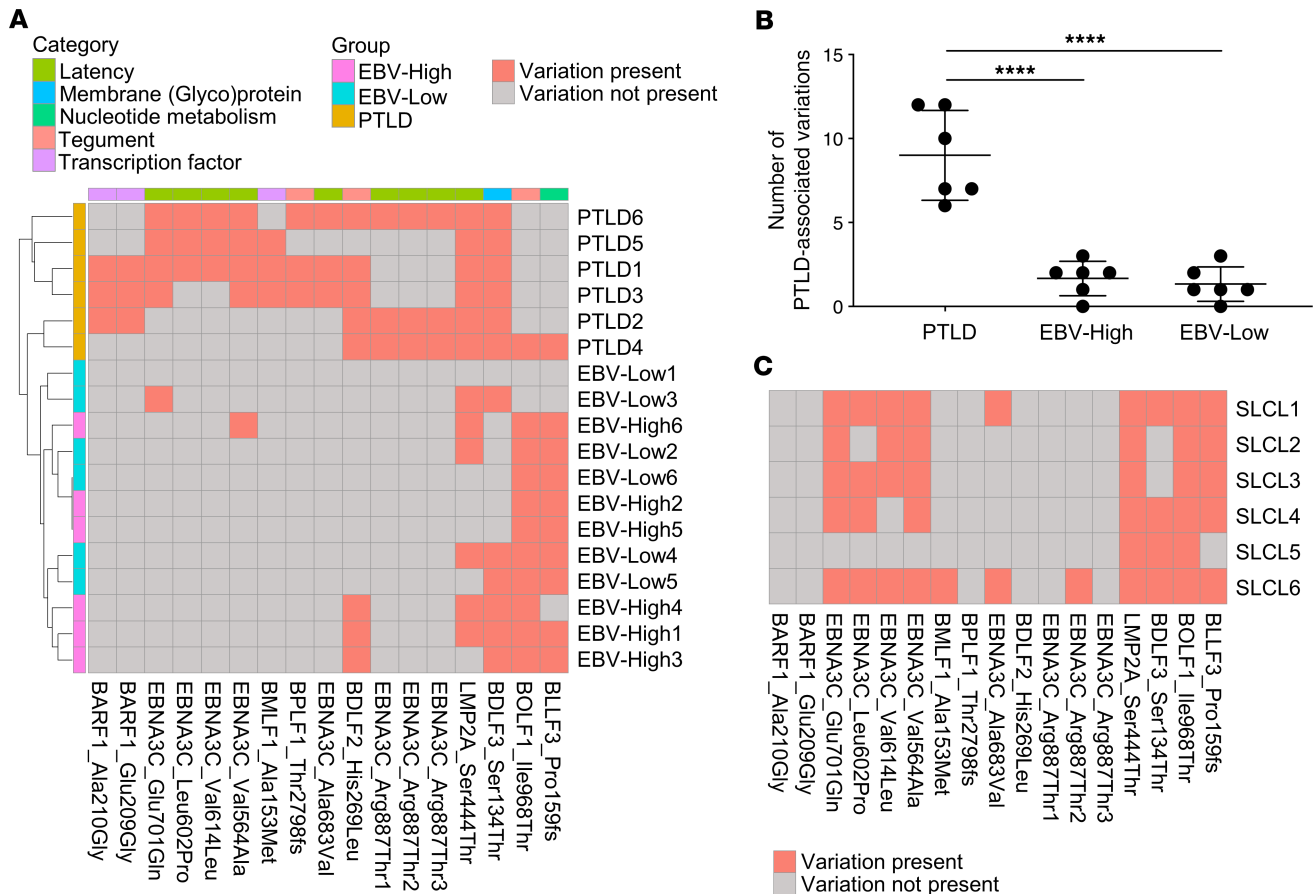


**Figure 4. Phylogenetic analysis of whole EBV genome and latent genes *EBNA1*, *EBNA3B*, and *EBNA3C*.** (A) Phylogenetic analysis of EBV genomes. (B) *EBNA1* genes shows a cluster of EBV<sup>hi</sup> and EBV<sup>lo</sup> samples and a grouping of control samples. (C) *EBNA3B* genes shows groupings of EBV<sup>hi</sup> and EBV<sup>lo</sup> samples and groupings of PTLD and SLCL samples. (D) *EBNA3C* genes shows groupings of EBV<sup>hi</sup> and EBV<sup>lo</sup> samples and groupings of PTLD and SLCL samples. (A–D) Light blue hexagons, SLCL samples; dark blue diamonds, PTLD samples; light green squares, EBV<sup>hi</sup> samples; dark green circles, EBV<sup>lo</sup> samples; and pink triangles, control samples. Nucleotide sequences were aligned using MUSCLE alignment. Trees were constructed using neighbor-joining model and Jukes-Cantor genetic distance modeling.

(13.04 nM) compared with the reference epitope QPRAPIRPI (8.84 nM). Since the binding affinity concentration is inversely proportional to binding affinity, the PTLD-associated variation causes a predicted decrease in binding affinity of the peptide to HLA-B7. Using PyMOL, we modeled the reference QPRAPIRPI and PTLD-associated variation QPRAPITPI in the HLA-B7 binding groove (Figure 6, B and C). For the PTLD-associated epitope, there was a loss of 1 hydrogen bond, which may contribute to the decreased binding affinity caused by the PTLD-associated variation.

*PTLD-associated variations alter binding of predicted epitopes to HLA-A, -B, and -DR alleles.* NetMHC and NetMHCII were used to determine whether the remaining PTLD-associated variations were located within predicted T cell epitopes and how the PTLD-associated variation could affect the binding of the epitope (Figure 7, Table 3, Table 4, Table 5, and Table 6). An amino acid sequence was obtained from the reference genome AJ507799 that contained the location of each PTLD-associated variation. NetMHC4.0 and NetMHCII were used to predict 9mer and 15mer peptides, respectively, that contained the variable amino acid with a sliding window and the corresponding HLA-A, -B, and -DR alleles. NetMHC predicted 85 epitopes for HLA-A alleles and 140 epitopes for HLA-B alleles





**Figure 5. Significant nonsynonymous variations associated with PTLD.** (A) Significant nonsynonymous variations associated with PTLD group ( $n = 6$ ) versus EBV<sup>hi</sup> ( $n = 6$ ) and EBV<sup>lo</sup> ( $n = 6$ ) groups. Cladogram represents unsupervised clustering of PTLD (gold), EBV<sup>hi</sup> (pink), and EBV<sup>lo</sup> (aqua) samples. The variations are color-coded by gene category: latency (green), membrane glycoprotein (blue), nucleotide metabolism (teal), tegument (pink), and transcription factor (purple). (B) The number of PTLD-associated variations in each sample of the PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups. Each dot represents 1 sample. Mean  $\pm$  SEM are shown for each group. Statistical significance between number of PTLD-associated variations in the PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups was determined using unpaired 2-tailed  $t$  test with  $P$  value. PTLD versus EBV<sup>hi</sup>,  $P < 0.0001$ ; PTLD versus EBV<sup>lo</sup>,  $P < 0.0001$ . (C) Presence of PTLD-associated variations in SLCL samples. \*\*\*\* $P < 0.0001$ .

for the reference sequence. NetMHCII predicted 91 epitopes derived from the reference sequence for HLA-DR alleles. Reference peptide epitopes were predicted to be strong binders (SB), weak binders (WB), or nonbinders (NB). To determine the impact on the binding affinity concentration of the epitopes, the SB and WB peptides were analyzed with the PTLD-associated variation (Figure 7). When the PTLD-associated variations were present, 4 distinct predicted changes in binding status of the epitopes were observed (SB to WB, WB to NB, WB to SB, and SB to NB). Twenty-five peptides showed a decrease in binding (SB to WB) (Figure 7, A, D, and H), while 6 SB peptides became NB peptides (Figure 7G) and 76 WB peptides became NB peptides (Figure 7, B, E, and I). Additionally, potential tumor neoantigens were identified, as 17 WB peptides became predicted SB peptides when the PTLD-associated variations were present (Figure 7, C, F, and J). HLA typing was performed on DNA from subject samples (Supplemental Table 3), and several PTLD-associated variants were present in individuals carrying the particular HLA allele that the PTLD variant epitope is predicted to bind (Table 3, Table 4, Table 5, and Table 6).

### Discussion

Full genome sequencing of EBV offers insights into the underlying pathogenesis of EBV-associated malignancies and the impact of viral diversity on the host immune response. In this report, we provide the first genome wide analysis to our knowledge of EBV isolated from the blood of transplant recipients with and without PTLD. We identified 15 PTLD-associated nonsynonymous variations resulting in amino

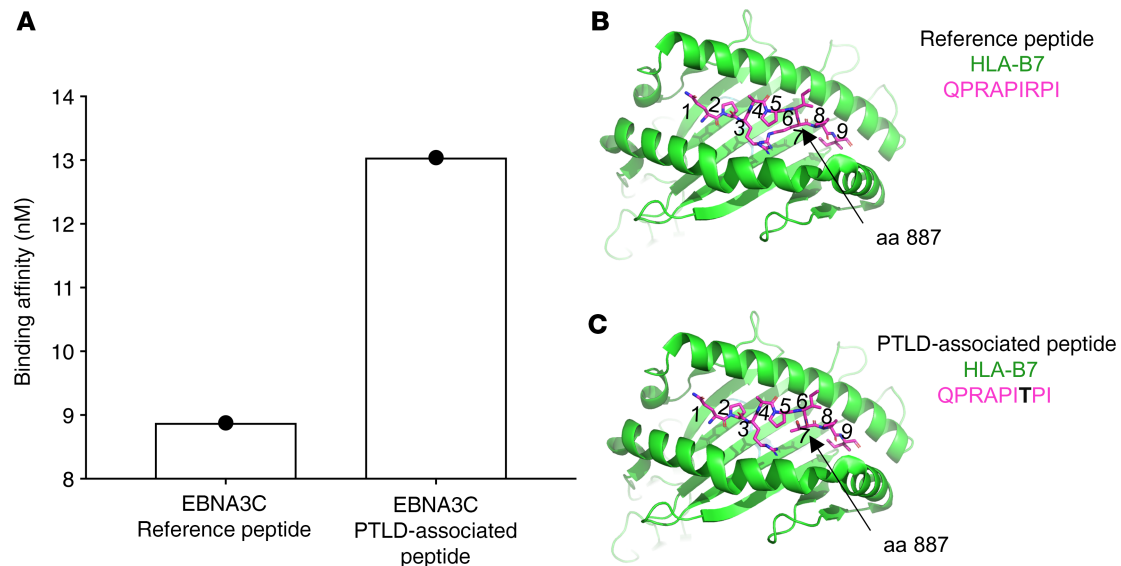
**Table 2. Significant nonsynonymous variations associated with PTLD group versus EBV<sup>hi</sup> and EBV<sup>lo</sup> group.**

Gene	Gene category	Amino acid change	Association <sup>A,B</sup>	PTLD samples <sup>C</sup>	EBV <sup>lo</sup> samples	EBV <sup>hi</sup> samples
BARF1	Transcription factor	Ala210Gly	PTLD	3	0	0
BARF1	Transcription factor	Glu209Gly	PTLD	3	0	0
EBNA3C	Latency	Glu701Gln	PTLD	4	1	1
EBNA3C	Latency	Leu602Pro	PTLD	3	0	0
EBNA3C	Latency	Val614Leu	PTLD	3	0	0
EBNA3C	Latency	Val564Ala	PTLD	3	0	0
BMLF1	Tegument	Ala153Met	PTLD	3	0	0
BPLF1	Tegument	Thr2798fs <sup>D</sup>	PTLD	3	0	0
EBNA3C	Latency	Ala683Val	PTLD	3	0	0
BDLF2	Tegument	His269Leu	PTLD	5	3	0
EBNA3C	Latency	Arg887Thr1 <sup>E</sup>	PTLD	3	0	0
EBNA3C	Latency	Arg887Thr2 <sup>E</sup>	PTLD	3	0	0
EBNA3C	Latency	Arg887Thr3 <sup>E</sup>	PTLD	3	0	0
LMP2A	Latency	Ser444Thr	PTLD	6	3	3
BDLF3	Membrane Glycoprotein	Ser134Thr	PTLD	6	3	4
BOLF1	Tegument	Ile968Thr	EBV <sup>hi</sup> and EBV <sup>lo</sup>	1	6	4
BLLF3	Nucleotide Metabolism	Pro159fs <sup>D</sup>	EBV <sup>hi</sup> and EBV <sup>lo</sup>	1	5	4

<sup>A</sup>Groups that variation is associated with. <sup>B</sup>Fisher's exact test was performed comparing variations in the PTLD group to the variations in the EBV<sup>hi</sup> and EBV<sup>lo</sup> group. <sup>C</sup>Number of samples containing the variation. <sup>D</sup>fs, frame shift. <sup>E</sup>Different nucleotide changes that resulted in the same amino acid change.

acid changes, including 8 in *EBNA3C*. Overall, the *EBNA3C* sequences isolated from the PTLD samples had significantly more nonsynonymous variations than *EBNA3C* sequenced from PTLD-free transplant recipients. *EBNA3C* is a latent cycle EBV gene that is essential for B cell transformation and acts as a transcriptional regulator through interaction with a variety of cellular DNA binding proteins to mediate activation, or repression, of cellular and viral genes (22–29). *EBNA3C* is also an important target of the T cell response during latent infection, with as many as 4% of CD8<sup>+</sup> T cells specific for an immunodominant peptide of *EBNA3C* (30). Previous full genome sequencing studies of EBV in PTLD were restricted to cell lines derived from PTLD patients (9) or FFPE PTLD specimens (16) and did not include non-PTLD transplant controls. Consistent with our findings, Dharnidharka et al. (16) also found that *EBNA3C*, as well as *EBNA1* and *EBNA-LP*, isolated from FFPE specimens had increased numbers of variations compared with other EBV genes; however, the specific variations were not reported. Interestingly, the Ala-683Val and Glu701Gln variations in *EBNA3C* that we identified from the blood of PTLD subjects were also identified in some PTLD-derived B cell lines (9). It will be important in future studies to compare variations in the viral genome isolated from the blood directly with those in the tumor. We also identified PTLD-associated variations in the latency gene *LMP2A* and a series of lytic genes including the tegument genes *BPLF1*, *BDLF2*, and *BOLF1*; the transcription factors *BARF1* and *BMLF1*; the membrane glycoprotein gene *BDLF3*; and the nucleotide metabolism gene *BLLF3*. Of these genes, *BDLF2* and *BDLF3* had significantly more variations in the PTLD group than the EBV<sup>hi</sup> or EBV<sup>lo</sup> groups. To our knowledge, these are the first reported EBV lytic gene sequences from PTLD patients and transplant recipients.

The prevailing paradigm has been that PTLD arises primarily as a consequence of immunosuppression-induced deficiencies in EBV-specific T cells. However, all transplant recipients are required to take immunosuppression, but only a subset develop EBV<sup>+</sup> PTLD, indicating that other factors are also at play. Our results suggest that variations in immunogenic epitopes of viral proteins also contribute to altered host immunity to EBV and promote escape from the immune response. Thus, these findings constitute the basis for a potentially novel mechanism of pathogenesis contributing to the immune impairment that underlies EBV<sup>+</sup> PTLD. Here, we focused on the impact of PTLD-associated variations on the binding of EBV-derived peptides to MHC proteins for presentation to EBV-specific T cells. Two of the PTLD-associated variations we identified, *EBNA3C* Arg887Thr (HLA-B7) and *LMP2A* Ser444Thr (HLA-A25), are found within previously defined T cell epitopes. We were able to model the

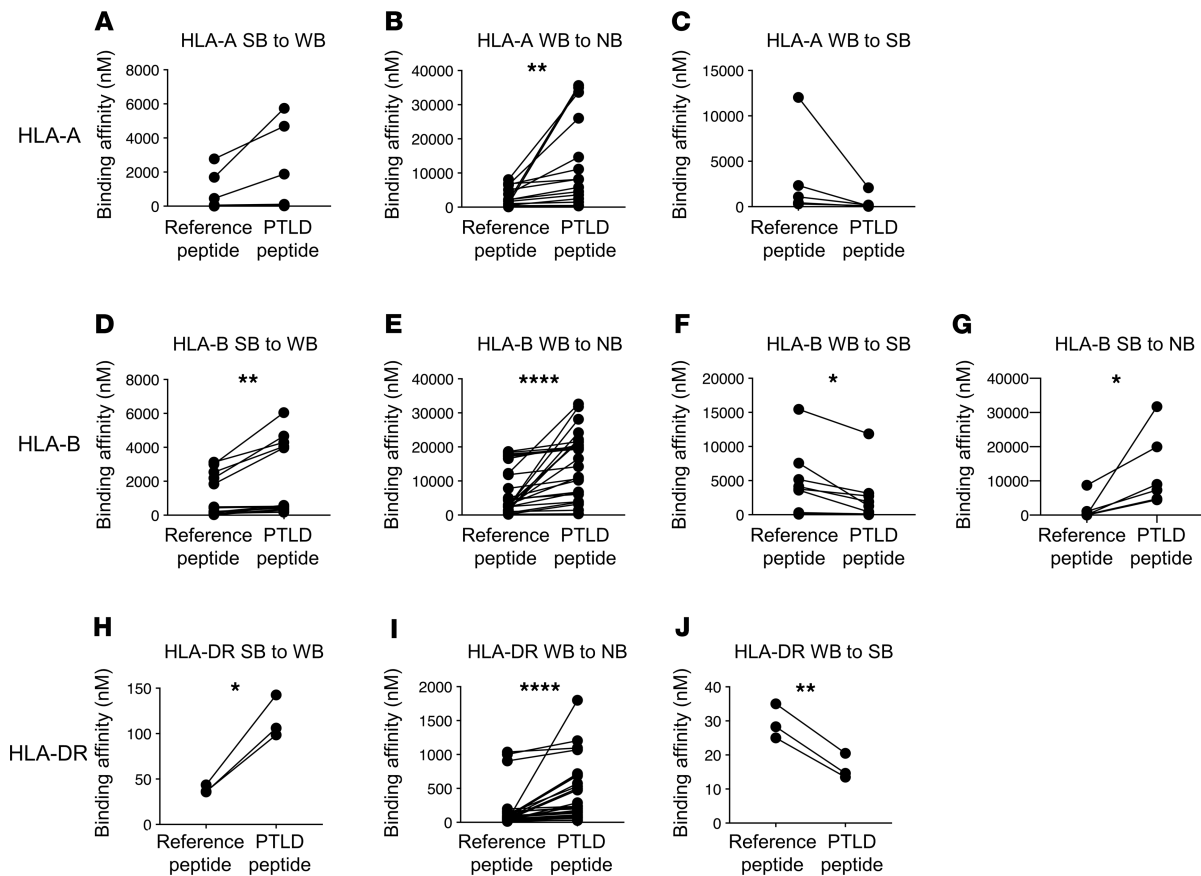


**Figure 6. PTLD-associated variation (Arg887Thr) in HLA-B7-restricted EBNA3C epitope increases the binding affinity concentration of the peptide to HLA-B7. (A)** Binding affinity of HLA-B7 EBNA3C epitope with and without PTLD-associated variation. **(B)** Pymol model of reference EBNA3C epitope, QPRAPIRPI (pink), in HLA-B7 (green). **(C)** Pymol model of EBNA3C epitope with PTLD-associated variation (bolded), QPRAPITPI (pink), in HLA-B7 (green).

binding of the EBNA3C PTLD-associated variation QPRAPIRPI to HLA-B7. Computational analysis indicated that the variation caused a decrease in HLA binding compared with the reference peptide QPRAPITPI. Other PTLD-associated variations that have not yet been established as T cell epitopes were analyzed to determine how these variations might affect binding to HLA class I and class II molecules for antigen presentation. The majority of the EBV-derived peptides from the PTLD group caused a predicted decrease, or loss, in HLA binding, suggesting that the PTLD-associated variations could impact antigen presentation and subsequent T cell response, thereby increasing the risk for PTLD. Moreover, we were able to identify several PTLD-associated variations in PTLD subjects that were predicted to bind to specific HLA alleles that matched the HLA of the subject carrying the variant epitope. These data suggest that the PTLD-associated variants may impact the immune response to EBV and potentially contribute to development of disease. Along these lines, Tao et al. (31) found that CTL generation was significantly diminished when peptides derived from EBNA3C variants were used to stimulate T cells compared with control peptides, while Gottschalk et al. (32) found that adoptive transfer of ex vivo EBV-specific CTL was ineffective because the major activity of the CTL line was directed against an HLA-B restricted epitope of EBNA3B that had been deleted in the EBV found in the tumor. In addition, White et al. (33) reported that LCL produced with EBV lacking *EBNA3B* were more tumorigenic in humanized mice than LCL made with WT EBV, and other examples of *EBNA3B* alterations were identified in clinical PTLD samples. Finally, some EBV peptides from the PTLD group had predicted increases in HLA binding when compared with the reference peptides. These variant peptides may constitute tumor neoantigens that could be exploited to augment cellular immunotherapy approaches.

Using previously established classification schemes for EBV strains, we found that all EBV genomes isolated from the blood of PTLD patients were type 1, consistent with prior reports that type 1 is prevalent in EBV isolated from FFPE specimens of PTLD patients (16, 34). In contrast, 4 of the 6 strain types defined by *LMP1* diversity were represented in PTLD-derived EBV genomes (11).

This study constitutes an important advance in our understanding of EBV genome diversity in health and disease and provides the basis for the application of immunogenomics to PTLD. By comparing the EBV genome from transplant recipients with and without PTLD, we discovered amino acid changes within several genes, including *EBNA3C*, associated with PTLD. Moreover, we determined that these variations alter the predicted binding of EBV-derived peptides to HLA-A, -B, and -DR proteins, elucidating a potential mechanism of reduced immune presentation of EBV epitopes that may promote immune escape in PTLD. These PTLD-associated variations provide insight into how cancer-specific



**Figure 7. PTLD-associated variations alter binding of predicted epitopes to HLA-A, -B, and -DR alleles.** For each PTLD-associated variation, a 29-amino acid sequence from the EBV<sup>hi</sup> and EBV<sup>lo</sup> reference sequence, with the location of the variable amino acid at location 15, was used to predict peptides that bind to HLA-A, -B, and -DR alleles on NetMHC 4.0 Server and NetMHCII. Peptides were predicted to be strong binders (SB), weak binders (WB), or nonbinders (NB). (A–F and H–J) Shown are the binding affinity concentrations for peptides for which the PTLD-associated variation caused a change in binding from a strong binder to a weak binder (A, D, H), a weak binder to a strong binder (B, E, I), or a weak binder to nonbinder (C, F, J). (A–C) HLA-A predicted epitopes for which the PTLD-associated variation changed the binding from strong to weak (A), reference versus PTLD peptides,  $P = 0.009$ ; and weak to strong (C). (D–G) HLA-B predicted epitopes for which the PTLD-associated variation changed the binding from strong to weak (D), reference versus PTLD peptides,  $P = 0.006$ ; weak to no binding (E), reference versus PTLD peptide,  $P < 0.0001$ ; weak to strong (F), reference versus PTLD peptide,  $P = 0.02$ ; strong to no binding (G), reference versus PTLD peptide,  $P = 0.05$ . (H–J) HLA-DR predicted epitopes for which the PTLD-associated variation changed the binding from strong to weak (H), reference versus PTLD peptide,  $P = 0.02$ ; weak to no binding (I), reference versus PTLD peptide,  $P = 0.0005$ ; weak to strong (J), reference versus PTLD peptide,  $P = 0.005$ . (A–J) Statistical significance between number of reference peptide and PTLD peptide binding affinities was determined using paired 2-tailed  $t$  test with  $P$  value. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\*\* $P < 0.0001$ .

EBV variations alter the immune response to EBV, identify potential biomarkers that may help identify risk for PTLD development, and reveal possible tumor neoantigens for cellular immunotherapy.

## Methods

**Patient samples, cell lines, and DNA isolation.** DNA was isolated from 0.5 mL of blood obtained from pediatric liver transplant recipients that had detectable EBV by semi-qPCR (33) but did not develop PTLD ( $n = 12$ ). The 12 patients were separated into 2 groups, EBV<sup>hi</sup> and EBV<sup>lo</sup>, based on their EBV viral loads in whole blood determined by qPCR as described below. EBV<sup>hi</sup> samples were defined as  $> 1000$  copies/ $\mu\text{g}$  of DNA, and EBV<sup>lo</sup> samples were defined as  $\leq 1000$  copies/ $\mu\text{g}$  of DNA. DNA was also isolated from the blood of 6 pediatric liver and/or kidney transplant recipients within 1 month of biopsy-proven PTLD, as previously described (35). Lastly, DNA was isolated from 6 EBV<sup>+</sup> SLCL derived from a separate set of 6 PTLD patients (36). The QIAamp DNA Mini Kit (Qiagen) was used to isolate DNA according to manufacturer's recommendations.

**Quantification of EBV.** EBV viral load quantification was determined using the artus EBV PCR kit, which targets the *EBNA1* gene (Qiagen). PCR was performed according to the manufacturer's instructions, except the total reaction mixture was scaled down to 30  $\mu\text{L}$ . For quantification, a standard curve was

**Table 3. Summary of variations identified in PTLD samples that alter binding of predicted epitopes to HLA-DR alleles**

Gene	Variation	HLA allele	Reference peptide	Reference binding affinity (nM)	Reference peptide binding strength	PTLD-associated peptide	PTLD binding affinity (nM)	PTLD peptide binding strength	Change in binding
LMP2	Ser444Thr	DRB1_1602	AGAVWLTVMSTNTLLS	43.7	SB	AGAVWLTVMSTNTLLS	142.5	WB	SB to WB
LMP2	Ser444Thr	DRB1_1602	AVWLTVMSTNTLLSAW	36.4	SB	AVWLTVMSTNTLLSAW	98.5	WB	SB to WB
LMP2	Ser444Thr	DRB1_1602	GAVWLTVMSTNTLLSA	35.9	SB	GAVWLTVMSTNTLLSA	106.1	WB	SB to WB
LMP2	Ser444Thr	DRB1_0101	LTVMSNTLLSAWILT	20.3	WB	LTVMTNTLLSAWILT	33	NB	WB to NB
<b>LMP2<sup>A</sup></b>	<b>Ser444Thr</b>	<b>DRB1_0101</b>	<b>VAGAVWLTVMSTNTLL</b>	<b>27.5</b>	<b>WB</b>	<b>VAGAVWLTVMSTNTLL</b>	<b>62.1</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0101</b>	<b>AGAVWLTVMSTNTLLS</b>	<b>15.7</b>	<b>WB</b>	<b>AGAVWLTVMSTNTLLS</b>	<b>34.6</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0401</b>	<b>AVWLTVMSTNTLLSAW</b>	<b>158.6</b>	<b>WB</b>	<b>AVWLTVMSTNTLLSAW</b>	<b>205.4</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0401</b>	<b>GAVWLTVMSTNTLLSA</b>	<b>166.3</b>	<b>WB</b>	<b>GAVWLTVMSTNTLLSA</b>	<b>198.2</b>	<b>NB</b>	<b>WB to NB</b>
LMP2	Ser444Thr	DRB1_0402	VWLTVMSTNTLLSAWI	1004.7	WB	VWLTVMSTNTLLSAWI	1204.5	NB	WB to NB
LMP2	Ser444Thr	DRB1_0402	AGAVWLTVMSTNTLLS	904.9	WB	AGAVWLTVMSTNTLLS	1071.2	NB	WB to NB
LMP2	Ser444Thr	DRB1_0403	TMVAGAVWLTVMSTNT	1035.4	WB	TMVAGAVWLTVMSTNT	1097.7	NB	WB to NB
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0701</b>	<b>AVWLTVMSTNTLLSAW</b>	<b>51.7</b>	<b>WB</b>	<b>AVWLTVMSTNTLLSAW</b>	<b>112</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0701</b>	<b>GAVWLTVMSTNTLLSA</b>	<b>43.7</b>	<b>WB</b>	<b>GAVWLTVMSTNTLLSA</b>	<b>91.5</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0701</b>	<b>MVAGAVWLTVMSTNTL</b>	<b>25.3</b>	<b>WB</b>	<b>MVAGAVWLTVMSTNTL</b>	<b>73.4</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0701</b>	<b>VWLTVMSTNTLLSAWI</b>	<b>53.7</b>	<b>WB</b>	<b>VWLTVMSTNTLLSAWI</b>	<b>96.2</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0701</b>	<b>WLTVMSTNTLLSAWIL</b>	<b>50.3</b>	<b>WB</b>	<b>WLTVMSTNTLLSAWIL</b>	<b>73.9</b>	<b>NB</b>	<b>WB to NB</b>
<b>LMP2</b>	<b>Ser444Thr</b>	<b>DRB1_0701</b>	<b>AGAVWLTVMSTNTLLS</b>	<b>34.2</b>	<b>WB</b>	<b>AGAVWLTVMSTNTLLS</b>	<b>69.2</b>	<b>NB</b>	<b>WB to NB</b>
EBNA3C	Leu602Pro	DRB1_1001	AAGPRILAPLSAGPP	40.2	WB	AAGPRILAPPSAGPP	242.6	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1001	AGPRILAPLSAGPPA	35.6	WB	AGPRILAPPSAGPPA	243.9	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1001	GPRILAPLSAGPPAA	36.1	WB	GPRILAPPSAGPPAA	288.3	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1001	PRILAPLSAGPPAAG	37.7	WB	PRILAPPSAGPPAAG	478.4	NB	WB to NB
LMP2	Ser444Thr	DRB1_1501	AVWLTVMSTNTLLSAW	77.8	WB	AVWLTVMSTNTLLSAW	157.5	NB	WB to NB
LMP2	Ser444Thr	DRB1_1501	GAVWLTVMSTNTLLSA	72.1	WB	GAVWLTVMSTNTLLSA	145	NB	WB to NB
LMP2	Ser444Thr	DRB1_1501	VWLTVMSTNTLLSAWI	75.9	WB	VWLTVMSTNTLLSAWI	140	NB	WB to NB
LMP2	Ser444Thr	DRB1_1501	WLTVMSTNTLLSAWIL	99.2	WB	WLTVMSTNTLLSAWIL	148	NB	WB to NB
LMP2	Ser444Thr	DRB1_1501	AGAVWLTVMSTNTLLS	88.9	WB	AGAVWLTVMSTNTLLS	168.7	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1602	AAGPRILAPLSAGPP	80.9	WB	AAGPRILAPPSAGPP	689.2	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1602	AGPRILAPLSAGPPA	65.4	WB	AGPRILAPPSAGPPA	573.9	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1602	GPRILAPLSAGPPAA	55.3	WB	GPRILAPPSAGPPAA	501	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1602	PAAGPRILAPLSAGP	91.9	WB	PAAGPRILAPPSAGP	719.2	NB	WB to NB
EBNA3C	Leu602Pro	DRB1_1602	PRILAPLSAGPPAAG	71.7	WB	PRILAPPSAGPPAAG	701.2	NB	WB to NB
LMP2	Ser444Thr	DRB1_1602	MVAGAVWLTVMSTNTL	128.1	WB	MVAGAVWLTVMSTNTL	529.6	NB	WB to NB
LMP2	Ser444Thr	DRB1_1602	VAGAVWLTVMSTNTLL	54.6	WB	VAGAVWLTVMSTNTLL	180.3	NB	WB to NB
LMP2	Ser444Thr	DRB1_1602	WLTVMSTNTLLSAWIL	150	WB	WLTVMSTNTLLSAWIL	226.9	NB	WB to NB
LMP2	Ser444Thr	DRB3_0202	VWLTVMSTNTLLSAWI	197.9	WB	VWLTVMSTNTLLSAWI	237.6	NB	WB to NB
LMP2	Ser444Thr	DRB3_0301	LTVMSNTLLSAWILT	25.1	WB	LTVMTNTLLSAWILT	1801.1	NB	WB to NB
LMP2	Ser444Thr	DRB3_0301	VAGAVWLTVMSTNTLL	26.2	WB	VAGAVWLTVMSTNTLL	29.3	NB	WB to NB
BDLF2	His269Leu	DRB3_0202	INNVTYYPNNTDSHG	25	WB	INNVTYYPNNTDSL	13.5	SB	WB to SB
BDLF2	His269Leu	DRB3_0202	VINNVTYYPNNTDSH	35	WB	VINNVTYYPNNTDSL	20.5	SB	WB to SB
BDLF2	His269Leu	DRB3_0202	VTYYPNNTDSHG	28.3	WB	VTYYPNNTDSL	14.6	SB	WB to SB

<sup>A</sup>Bolded rows indicate cases where the specific EBV variation and the presenting HLA allele were identified in the same PTLD sample or SLCL. SB, strong binder; WB, weak binder; NB, nonbinder.

created with a 10-fold dilution (24,000–2.4 copies of EBV per  $\mu\text{L}$ ) of the first WHO International Standard EBV for Nucleic Acid Amplification from National Institute for Biological Standards and Control (Hertfordshire, United Kingdom). Additionally, a 10-fold dilution series (500,000–5 copies per  $\mu\text{L}$ ) of EBV Positive Control DNA (Qiagen) was used. EBV viral load was calculated by using the EBV copy number, as determined by qPCR per  $\mu\text{g}$  of DNA of each sample.

*PCR amplification of whole EBV genome.* The EBV genome, excluding repeat regions, was amplified from SLCL DNA using 59 PCR primer pairs described by Kwok et al. (13). Due to a lower EBV copy number in patient blood samples compared with SLCL, 79 additional primer pairs (Table 1, Supplemental Table 2) were designed to amplify the same region of the EBV genome as the Kwok et al. (13) primers and produce

**Table 4. Summary of variations identified in PTLD samples that alter binding of predicted epitopes to HLA-A alleles**

Gene	Variation	HLA allele	Reference peptide	Reference binding affinity (nM)	Reference peptide binding strength	PTLD-associated peptide	PTLD binding affinity (nM)	PTLD peptide binding strength	Change in binding
LMP2	Ser444Thr	HLA-A0203	VMSNTLLSA	8.51	SB	VMTNTLLSA	19.61	WB	SB to WB
LMP2	Ser444Thr	HLA-A0211	VMSNTLLSA	14.78	SB	VMTNTLLSA	42.59	WB	SB to WB
LMP2	Ser444Thr	HLA-A0212	VMSNTLLSA	65.06	SB	VMTNTLLSA	119.42	WB	SB to WB
BDFL2	His269Leu	HLA-A2601	DSHGHHGGF	454.96	SB	DSLGGHGGF	1884.21	WB	SB to WB
EBNA3C	Val614Leu	HLA-A2603	HIVTPPSAR	1697.39	SB	HILTPPSAR	5747.97	WB	SB to WB
EBNA3C	Val614Leu	HLA-A2603	VTPPSARPR	2773.15	SB	LTPPSARPR	4690.01	WB	SB to WB
EBNA3C	Val614Leu	HLA-A6801	HIVTPPSAR	28.97	SB	HILTPPSAR	70.42	WB	SB to WB
LMP2	Ser444Thr	HLA-A0202	VMSNTLLSA	76.62	WB	VMTNTLLSA	345.98	NB	WB to NB
<b>EBNA3C<sup>A</sup></b>	<b>Leu602Pro</b>	<b>HLA-A0205</b>	<b>PLSAGPPAA</b>	<b>6567.58</b>	<b>WB</b>	<b>PPSAGPPAA</b>	<b>26059.16</b>	<b>NB</b>	<b>WB to NB</b>
<b>EBNA3C</b>	<b>Val614Leu</b>	<b>HLA-A0205</b>	<b>PPAAGPHIV</b>	<b>8104.84</b>	<b>WB</b>	<b>PPAAGPHIL</b>	<b>33654.2</b>	<b>NB</b>	<b>WB to NB</b>
EBNA3C	Leu602Pro	HLA-A0216	PLSAGPPAA	931.72	WB	PPSAGPPAA	35047.01	NB	WB to NB
EBNA3C	Leu602Pro	HLA-A0219	PLSAGPPAA	1758.61	WB	PPSAGPPAA	35657.09	NB	WB to NB
EBNA3C	Ala683Va	HLA-A0250	SWAPSVCAL	533.69	WB	SWVPSVCAL	2485.1	NB	WB to NB
EBNA3C	Ala683Va	HLA-A2402	SWAPSVCAL	1676.36	WB	SWVPSVCAL	3654.43	NB	WB to NB
BDLF3	Ser134Thr	HLA-A3002	RSSSTTSAT	1028.18	WB	RSSTTSAT	1485.61	NB	WB to NB
<b>BDLF3</b>	<b>Ser134Thr</b>	<b>HLA-A3201</b>	<b>STTSATTRI</b>	<b>2151.68</b>	<b>WB</b>	<b>TTTSATTRI</b>	<b>5850.37</b>	<b>NB</b>	<b>WB to NB</b>
EBNA3C	Ala683Va	HLA-A3207	TPRPSWAPS	2240.7	WB	TPRPSWVPS	4535.75	NB	WB to NB
BDLF3	Ser134Thr	HLA-A3215	STTSATTRI	5111.18	WB	TTTSATTRI	8251.27	NB	WB to NB
BDLF2	His269Leu	HLA-A6601	HGGHGGFEA	6596.43	WB	LGHHGGFEA	11122.04	NB	WB to NB
LMP2	Ser444Thr	HLA-A6601	LTVMSNTLL	6883.71	WB	LTVMTNTLL	8157.44	NB	WB to NB
LMP2	Ser444Thr	HLA-A6802	TVMSNTLLS	420.99	WB	TVMTNTLLS	466.29	NB	WB to NB
EBNA3C	Leu602Pro	HLA-A6901	APLSAGPPA	3850.28	WB	APPSAGPPA	14663.21	NB	WB to NB
LMP2	Ser444Thr	HLA-A2501	MSNTLLSAW	12041.47	WB	MTNTLLSAW	2094.21	SB	WB to SB
LMP2	Ser444Thr	HLA-A2602	MSNTLLSAW	2337.58	WB	MTNTLLSAW	94.53	SB	WB to SB
EBNA3C	Ala683Va	HLA-A3207	SWAPSVCAL	1095.82	WB	SWVPSVCAL	186.06	SB	WB to SB
EBNA3C	Val614Leu	HLA-A3301	HIVTPPSAR	305.52	WB	HILTPPSAR	101.78	SB	WB to SB
LMP2	Ser444Thr	HLA-A6823	MSNTLLSAW	449.82	WB	MTNTLLSAW	40.82	SB	WB to SB

<sup>A</sup>Bolded rows indicate cases where the specific EBV variation and the presenting HLA allele were identified in the same PTLD sample or SLCL. SB, strong binder; WB, weak binder; NB, nonbinder.

overlapping amplicons, ranging from 1019–2498 bp in length. To amplify the EBV genome, 100 ng of DNA was added to a 100  $\mu$ L reaction mixture with 10  $\mu$ L HotstarTaq Plus 10 $\times$  PCR buffer, 20  $\mu$ L HotstarTaq Plus Q solution, 0.5  $\mu$ L HotstarTaq Plus enzyme (Qiagen), 2  $\mu$ L of 10 mM deoxynucleotide triphosphate (dNTPs) (Thermo Fisher Scientific), and 2.4  $\mu$ L of 25  $\mu$ M forward and reverse primers (Protein and Nucleic Acid Facility). PCR was conducted in an iCycler Thermal Cycler (Bio-Rad) with the following conditions: enzyme activation at 95°C for 5 minutes, 50 cycles of denaturation at 94°C for 45 seconds, annealing at 56°C for 45 seconds, and extension at 72°C for 1 minute per 1 kb of product.

*Next-generation sequencing.* PCR products were visualized by loading a mixture of 45  $\mu$ L of PCR product and 9  $\mu$ L of 6 $\times$  Blue Gel Loading Dye (New England BioLabs) on a 2% TopVision agarose gel (Thermo Fisher Scientific) with 20  $\mu$ L of GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific). PCR product (50  $\mu$ L) was then purified using QIAquick PCR Purification Kit (Qiagen) per manufacturer's instructions. The DNA concentration was determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). The PCR products for each sample were normalized to equal molecular quantity, pooled, and sequenced at the Stanford Functional Genomics Facility. A quality check was performed on the samples using the Agilent High Sensitivity DNA Kit and Agilent 2100 bioanalyzer, and only samples that were successfully fragmented to 150 bp were used. Barcodes and standard Illumina indexed adapters were added using the KAPA DNA library preparation kit for Illumina. The library was then sequenced on an Illumina MiSeq Next Generation Sequencer at the Stanford Functional Genomics Facility. HLA typing was performed by extended/full gene next-generation sequencing as described previously (37).

**Table 5. Summary of PTLD-associated variations that alter binding of predicted epitopes to HLA-B alleles**

Gene	Variation	HLA allele	reference peptide	Reference binding affinity (nM)	Reference peptide binding strength	PTLD-associated peptide	PTLD binding affinity (nM)	PTLD peptide binding strength	Change in binding
EBNA3C	Ala683Va	HLA-B0702	RPSWAPSV	107.47	SB	RPSWVPSVC	185.3	WB	SB to WB
EBNA3C	Ala683Va	HLA-B0702	TPRPSWAPS	81.25	SB	TPRPSWVPS	203.43	WB	SB to WB
EBNA3C	Leu602Pro	HLA-B0702	APLSAGPPA	34.69	SB	APPSAGPPA	309.7	WB	SB to WB
EBNA3C	Ala683Va	HLA-B1402	SWAPSVCAL	1845.44	SB	SWVPSVCAL	3979.38	WB	SB to WB
LMP2	Ser444Thr	HLA-B1509	WLTVMNTL	3007.57	SB	WLTVMNTL	6050.51	WB	SB to WB
LMP2	Ser444Thr	HLA-B3901	WLTVMNTL	488.2	SB	SWVPSVCAL	535.6	WB	SB to WB
BARF1	Glu209Gly	HLA-B4001	KEEAHGYYV	52.04	SB	KEGAHGYYV	358.42	WB	SB to WB
EBNA3C	Ala683Va	HLA-B4201	TPRPSWAPS	178.63	SB	TPRPSWVPS	583.71	WB	SB to WB
EBNA3C	Val614Leu	HLA-B4201	GPHIVTPPS	433.54	SB	GPHILTPPS	514.62	WB	SB to WB
BARF1	Ala210Gly	HLA-B4501	KEEAHGYYV	216.81	SB	KEEGHGYYV	417.24	WB	SB to WB
EBNA3C	Leu602Pro	HLA-B5401	APLSAGPPA	47.52	SB	APPSAGPPA	527.21	WB	SB to WB
EBNA3C	Val564Ala	HLA-B5401	GPPAVGPPA	171.31	SB	GPPAAGPPA	430.16	WB	SB to WB
EBNA3C	Ala683Va	HLA-B8301	APSVCALSV	2535.87	SB	VPSVCALSV	4051.71	WB	SB to WB
EBNA3C	Arg887Thr	HLA-B8301	SPQPRAPIR	2194.94	SB	SPQPRAPIT	4673.9	WB	SB to WB
EBNA3C	Val564Ala	HLA-B8301	GPPAVGPPA	3123.88	SB	GPPAAGPPA	4294.01	WB	SB to WB
EBNA3C	Arg887Thr	HLA-B0702	RPIPTRFPP	250.57	WB	TIPIPTRFPP	3242.03	NB	WB to NB
EBNA3C	Arg887Thr	HLA-B0802	APIRPIPTR	18638.46	WB	APITPIPTR	21591.98	NB	WB to NB
LMP2	Ser444Thr	HLA-B0802	MSNTLLSAW	17911.88	WB	MTNTLLSAW	19364.52	NB	WB to NB
EBNA3C	Leu602Pro	HLA-B1402	AGPRILAPL	2572.07	WB	AGPRILAPP	16617.94	NB	WB to NB
EBNA3C	Ala683Va	HLA-B1502	SWAPSVCAL	4345.56	WB	SWVPSVCAL	6794.18	NB	WB to NB
EBNA3C	Arg887Thr	HLA-B1502	QPRAPIRPI	4793.75	WB	QPRAPITPI	6098.16	NB	WB to NB
LMP2	Ser444Thr	HLA-B1502	WLTVMNTL	4427.29	WB	SWVPSVCAL	6794.18	NB	WB to NB
BDF3	Ser134Thr	HLA-B1503	RSSSTTSAT	202.69	WB	RSSTTSAT	390.71	NB	WB to NB
BARF1	Ala210Gly	HLA-B1801	EEAHGYYVS	2408.39	WB	EEGHGYYVS	6691.09	NB	WB to NB
BARF1	Glu209Gly	HLA-B1801	EEAHGYYVS	2408.39	WB	EGAHGYYVS	31834.36	NB	WB to NB
BARF1	Ala210Gly	HLA-B2720	KEEAHGYYV	3012.87	WB	KEEGHGYYV	11124.93	NB	WB to NB
EBNA3C	Leu602Pro	HLA-B2720	AGPRILAPL	3027.81	WB	AGPRILAPP	21432.07	NB	WB to NB
<b>EBNA3C<sup>A</sup></b>	<b>Glu701Gln</b>	<b>HLA-B3501</b>	<b>QPIESSHLS</b>	<b>1033.48</b>	<b>WB</b>	<b>QPIQSSHLS</b>	<b>1406.64</b>	<b>NB</b>	<b>WB to NB</b>
<b>EBNA3C</b>	<b>Leu602Pro</b>	<b>HLA-B3501</b>	<b>APLSAGPPA</b>	<b>631.5</b>	<b>WB</b>	<b>APPSAGPPA</b>	<b>3751.18</b>	<b>NB</b>	<b>WB to NB</b>
EBNA3C	Leu602Pro	HLA-B4201	AGPRILAPL	1442.47	WB	AGPRILAPP	22228.45	NB	WB to NB
EBNA3C	Glu701Gln	HLA-B4402	IESSLSSM	1372.39	WB	IQSSHLSSM	24196.04	NB	WB to NB
BARF1	Ala210Gly	HLA-B4403	KEEAHGYYV	2519.89	WB	KEEGHGYYV	3890.19	NB	WB to NB
EBNA3C	Glu701Gln	HLA-B4403	IESSLSSM	2552.66	WB	IQSSHLSSM	28117.42	NB	WB to NB
<b>LMP2</b>	<b>Ser444Thr</b>	<b>HLA-B4601</b>	<b>VMSNTLLSA</b>	<b>11855.18</b>	<b>WB</b>	<b>VMTNTLLSA</b>	<b>14243.81</b>	<b>NB</b>	<b>WB to NB</b>
EBNA3C	Leu602Pro	HLA-B4801	AGPRILAPL	12252.4	WB	AGPRILAPP	32636.88	NB	WB to NB
EBNA3C	Val614Leu	HLA-B5101	PPAAGPHIV	5011.03	WB	PPAAGPHIL	10268.81	NB	WB to NB
EBNA3C	Val614Leu	HLA-B5401	PPAAGPHIV	2761.98	WB	PPAAGPHIL	21712.15	NB	WB to NB
BARF1	Ala210Gly	HLA-B7301	EEAHGYYVS	18371.96	WB	EEGHGYYVS	19969.01	NB	WB to NB
EBNA3C	Ala683Va	HLA-B7301	PRPSWAPSV	17439.34	WB	PRPSWVPSV	19893.31	NB	WB to NB
EBNA3C	Leu602Pro	HLA-B7301	APLSAGPPA	16612.19	WB	APPSAGPPA	20546.74	NB	WB to NB
BDF2	His269Leu	HLA-B7301	HGGHGGFEA	17194.64	WB	SGGHGGFEA	19532.01	NB	WB to NB
EBNA3C	Ala683Va	HLA-B8301	SWAPSVCAL	7824.53	WB	SWVPSVCAL	10681.07	NB	WB to NB
LMP2	Ser444Thr	HLA-B0802	WLTVMNTL	15470.62	WB	WLTVMNTL	11859.29	SB	WB to SB
EBNA3C	Glu701Gln	HLA-B1503	IESSLSSM	75.69	WB	IQSSHLSSM	5.03	SB	WB to SB
EBNA3C	Val614Leu	HLA-B1509	PHIVTPPSA	3779.87	WB	PHILTPPSA	2755.41	SB	WB to SB
<b>BDF2</b>	<b>His269Leu</b>	<b>HLA-B3501</b>	<b>YPNNTDSHG</b>	<b>316.85</b>	<b>WB</b>	<b>YPNNTDSL</b>	<b>125.03</b>	<b>SB</b>	<b>WB to SB</b>
EBNA3C	Val614Leu	HLA-B3901	PPAAGPHIV	3610.92	WB	PPAAGPHIL	424.14	SB	WB to SB
BARF1	Glu209Gly	HLA-B4002	KEEAHGYYV	217.4	WB	KEGAHGYYV	128.19	SB	WB to SB
EBNA3C	Ala683Va	HLA-B5101	APSVCALSV	7576.19	WB	VPSVCALSV	1289.17	SB	WB to SB
EBNA3C	Arg887Thr	HLA-B5101	QPRAPIRPI	4133.55	WB	QPRAPITPI	1922.48	SB	WB to SB
LMP2	Ser444Thr	HLA-B8301	MSNTLLSAW	5179.93	WB	MTNTLLSAW	3145.86	SB	WB to SB

<sup>A</sup>Bolded rows indicate cases where the specific EBV variation and the presenting HLA allele were identified in the same PTLD sample or SLCL. SB, strong binder; WB, weak binder; NB, nonbinder.

**Table 6. Summary of variations in PTLD samples that alter epitope binding to HLA-B alleles from strong binders to nonbinders**

Gene	Variation	HLA Allele	Reference Peptide	Reference Binding Affinity (nM)	Reference Peptide Binding Strength	PTLD-Associated Peptide	PTLD Binding Affinity (nM)	PTLD Peptide Binding Strength	Change in Binding
EBNA3C	Glu701Gln	HLA-B1801	IESSHLSSM	41.93	SB	IQSSHLSSM	4459.88	NB	SB to NB
<b>EBNA3C<sup>A</sup></b>	<b>Glu701Gln</b>	<b>HLA-B4001</b>	<b>IESSHLSSM</b>	<b>138.3</b>	<b>SB</b>	<b>IQSSHLSSM</b>	<b>4777.59</b>	<b>NB</b>	<b>SB to NB</b>
EBNA3C	Glu701Gln	HLA-B4002	IESSHLSSM	116.36	SB	IQSSHLSSM	8998.87	NB	SB to NB
EBNA3C	Arg887Thr	HLA-B4013	RPIPTRFPP	1109.12	SB	TPIPTRFPP	7350.82	NB	SB to NB
BARF1	Glu209Gly	HLA-B4501	EEAHGVVVS	111.88	SB	EGAHGVVVS	31759.69	NB	SB to NB
EBNA3C	Arg887Thr	HLA-B7301	PRAPIRPIP	8712.03	SB	PRAPITPIP	19953.88	NB	SB to NB

<sup>A</sup>Bolded rows indicate EBV variations identified in PTLD samples for which the HLA allele and variant are found in the same PTLD sample or SLCL. SB, strong binder; WB, weak binder; NB, nonbinder.

*Data submission.* Raw sequencing reads were uploaded to Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>). The Submission ID is SUB5881103 and the ProjectID is PRJNA551423.

*EBV genome sequence analysis.* Barcodes were trimmed from the reads, and low-quality reads (quality score below 20) were discarded from the data using Fasta packages. The reads were mapped to the reference AJ507799 EBV genome and assembled using Genome Analysis Toolkit (GATK) (Broad Institute) and Burrow Wheeler Aligner (38); insertions and deletions (Indels) were determined using GATK. SNV and base call filtering was conducted using SAMtools (39). SNVs with low scores were removed from the data based on low GenCall and Cluster Separation scores. The number of variations in 1000 nucleotide segments was calculated for each sample group. Two-way ANOVA was used to determine which 1000 nucleotide segment had significant differences in number of variations among the PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups. The number of variations in each gene category was calculated. The numbers of non-synonymous and synonymous variations in each sample group were determined using CLC Genomics Workbench (Qiagen), and differences were analyzed by 2-tailed Student's *t* test. The number of nonsynonymous variations in each gene category was divided by the sum of gene lengths (reference AJ507799 was used to calculate gene lengths) in each category. Two-way ANOVA with the Tukey post hoc test was used to determine the significance in the number of nonsynonymous variations between the groups. Phylogenetic trees were constructed from specific genes, segments of the genome, and the whole EBV genome for all samples by aligning with MUSCLE and constructing trees using neighbor-joining model and Jukes-Cantor genetic distance modeling with CLC Genomics Workbench. Fisher's exact test with CLC Genomics Workbench was used to determine which variations were associated with the PTLD group compared with the EBV<sup>hi</sup> and EBV<sup>lo</sup> groups. A *P* value threshold of 0.05 was set to identify variations that were significantly more common in the PTLD samples than in the EBV<sup>hi</sup> and EBV<sup>lo</sup> samples. The presence or absence of these variations in the PTLD, EBV<sup>hi</sup>, EBV<sup>lo</sup>, SLCL, and control groups was mapped using the pheatmap package in RStudio (RStudio).

*Analysis of EBV peptides and epitopes.* For each PTLD-associated variation, a 29-amino acid sequence from the EBV<sup>hi</sup> and EBV<sup>lo</sup> reference sequence, with the location of the variable amino acid at location 15, was used to predict peptides that bind to HLA-A, -B, and -DR alleles on NetMHC 4.0 Server and NetMHC-II (DTU Bioinformatics). Percentile rank is used for assignment to binder status and is determined from neural network analysis of binding affinity and eluted ligand data. For the PTLD-associated variations in the *BARF1* gene, the amino acid sequence used in the prediction analysis was truncated because the variable amino acids were less than 14 amino acids from the end of the protein. For each predicted epitope that contained the variable amino acid, the PTLD-associated variation replaced the reference amino acid to determine the impact on the binding affinity concentration of the epitope. NetMHC 4.0 Server was used to determine whether PTLD-associated variations altered the binding of previously defined and predicted T cell epitopes. PyMOL was used to model the binding of the EBNA3C epitope, QPRAPIRPI, to HLA-B7 and to predict the number of hydrogen bonds between the peptide and the HLA binding pocket.

*Statistics.* Statistical significance of EBV viral load between the PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups was determined by 2-way ANOVA, Tukey's multiple comparisons test, and a significant *P* value of less than or equal to 0.05. Statistical significance between the number of variations in 1000-nt segments across the EBV



genome from PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups was determined using 2-way ANOVA with Tukey's test and a significant *P* value of less than or equal to 0.05. To determine if there was a significant difference in the number of synonymous and nonsynonymous variations between sample groups, a 2-way ANOVA with Tukey's test and a significant *P* value of less than or equal to 0.05 was used. Phylogenetic trees were constructed using neighbor-joining model and Jukes-Cantor genetic distance modeling. Fisher's exact test with CLC Genomics Workbench was used to determine which variations were associated with the PTLD group compared with the EBV<sup>hi</sup> and EBV<sup>lo</sup> groups and a significant *P* value of less than or equal to 0.05. Statistical significance between the number of PTLD-associated variations in the PTLD, EBV<sup>hi</sup>, and EBV<sup>lo</sup> groups was determined using unpaired 2-tailed *t* test with a significant *P* value of less than or equal to 0.05. Statistical significance between number of reference peptide and PTLD peptide binding affinities was determined using paired 2-tailed *t* test with a significant *P* value of less than or equal to 0.05.

*Study approval.* This study was approved by the IRB at Stanford University. All participants provided written informed consent prior to inclusion in the study.

### Author contributions

EMM contributed to design of research studies, conducted experiments, acquired data, analyzed data, and wrote the manuscript; STH and VAB conducted experiments; JT performed data analysis and edited the manuscript; MFV performed the HLA typing; COE provided patient samples and analysis of clinical information; SMK contributed to design of the study, analyzed data, and edited the manuscript; and OMM conceived of and designed the research studies, analyzed data, supervised the study, and wrote the manuscript. All authors reviewed and approved the manuscript.

### Acknowledgments

The authors acknowledge support from Roche Organ Transplant Research Foundation Grant 739309649, the Tissue and Transplant Engineering Center of Excellence, and the Lucile Packard Foundation. E.M. was supported by the NIH T32 AI007290. J.T. is supported by a Singapore Agency for Science, Technology and Research (A\*STAR) National Science Scholarship.

Address correspondence to: Olivia M. Martinez, Stanford University School of Medicine, 1201 Welch Road, MSLS P312, Stanford, California 94305-5492, USA. Phone: 650.498.6247; Email: omm@stanford.edu.

1. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Health*. 2016;4(9):e609–e616.
2. Rickinson AB, Long HM, Palendira U, Münz C, Hislop AD. Cellular immune controls over Epstein-Barr virus infection: new lessons from the clinic and the laboratory. *Trends Immunol*. 2014;35(4):159–169.
3. Correia S, et al. Sequence Variation of Epstein-Barr Virus: Viral Types, Geography, Codon Usage, and Diseases. *J Virol*. 2018;92(22):e01132-18.
4. Kwok H, et al. Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples. *J Virol*. 2014;88(18):10662–10672.
5. Tu C, et al. Genome-Wide Analysis of 18 Epstein-Barr Viruses Isolated from Primary Nasopharyngeal Carcinoma Biopsy Specimens. *J Virol*. 2017;91(17):e00301-17.
6. Borozan I, Zapatka M, Frappier L, Ferretti V. Analysis of Epstein-Barr Virus Genomes and Expression Profiles in Gastric Adenocarcinoma. *J Virol*. 2018;92(2):e01239-17.
7. Wang S, Xiong H, Yan S, Wu N, Lu Z. Identification and Characterization of Epstein-Barr Virus Genomes in Lung Carcinoma Biopsy Samples by Next-Generation Sequencing Technology. *Sci Rep*. 2016;6:26156.
8. Peng RJ, et al. Genomic and transcriptomic landscapes of Epstein-Barr virus in extranodal natural killer T-cell lymphoma. *Leukemia*. 2019;33(6):1451–1462.
9. Palser AL, et al. Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. *J Virol*. 2015;89(10):5222–5237.
10. Sample J, et al. Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. *J Virol*. 1990;64(9):4084–4092.
11. Edwards RH, Seillier-Moiseiwitsch F, Raab-Traub N. Signature amino acid changes in latent membrane protein 1 distinguish Epstein-Barr virus strains. *Virology*. 1999;261(1):79–95.
12. Farrell PJ. Epstein-Barr Virus and Cancer. *Annu Rev Pathol*. 2019;14:29–53.
13. Kwok H, et al. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. *PLoS ONE*. 2012;7(5):e36939.
14. Liu Y, Yang W, Pan Y, Ji J, Lu Z, Ke Y. Genome-wide analysis of Epstein-Barr virus (EBV) isolated from EBV-associated gastric carcinoma (EBVaGC). *Oncotarget*. 2016;7(4):4903–4914.
15. Yao Y, et al. Genome-wide analysis of Epstein-Barr virus identifies variants and genes associated with gastric carcinoma and

- population structure. *Tumour Biol.* 2017;39(10):1010428317714195.
16. Dharnidharka VR, et al. Metagenomic analysis of DNA viruses from posttransplant lymphoproliferative disorders. *Cancer Med.* 2019;8(3):1013–1023.
17. Lin Z, et al. Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. *J Virol.* 2013;87(2):1172–1182.
18. Santpere G, et al. Genome-wide analysis of wild-type Epstein-Barr virus genomes derived from healthy individuals of the 1,000 Genomes Project. *Genome Biol Evol.* 2014;6(4):846–860.
19. Weiss ER, et al. Early Epstein-Barr Virus Genomic Diversity and Convergence toward the B95.8 Genome in Primary Infection. *J Virol.* 2018;92(2):e01466-17.
20. Correia S, et al. Natural Variation of Epstein-Barr Virus Genes, Proteins, and Primary MicroRNA. *J Virol.* 2017;91(15):e00375-17.
21. Hislop AD, Taylor GS, Sauce D, Rickinson AB. Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. *Annu Rev Immunol.* 2007;25:587–617.
22. Grundhoff AT, et al. Characterization of DP103, a novel DEAD box protein that binds to the Epstein-Barr virus nuclear proteins EBNA2 and EBNA3C. *J Biol Chem.* 1999;274(27):19136–19144.
23. Lin J, Johannsen E, Robertson E, Kieff E. Epstein-Barr virus nuclear antigen 3C putative repression domain mediates coactivation of the LMP1 promoter with EBNA-2. *J Virol.* 2002;76(1):232–242.
24. Zhao B, Sample CE. Epstein-barr virus nuclear antigen 3C activates the latent membrane protein 1 promoter in the presence of Epstein-Barr virus nuclear antigen 2 through sequences encompassing an spi-1/Spi-B binding site. *J Virol.* 2000;74(11):5151–5160.
25. Cotter MA, Robertson ES. Modulation of histone acetyltransferase activity through interaction of Epstein-Barr nuclear antigen 3C with prothymosin alpha. *Mol Cell Biol.* 2000;20(15):5722–5735.
26. Knight JS, Lan K, Subramanian C, Robertson ES. Epstein-Barr virus nuclear antigen 3C recruits histone deacetylase activity and associates with the corepressors mSin3A and NCoR in human B-cell lines. *J Virol.* 2003;77(7):4261–4272.
27. Radkov SA, et al. Epstein-Barr virus nuclear antigen 3C interacts with histone deacetylase to repress transcription. *J Virol.* 1999;73(7):5688–5697.
28. Subramanian C, Hasan S, Rowe M, Hottiger M, Orre R, Robertson ES. Epstein-Barr virus nuclear antigen 3C and prothymosin alpha interact with the p300 transcriptional coactivator at the CH1 and CH3/HAT domains and cooperate in regulation of transcription and histone acetylation. *J Virol.* 2002;76(10):4699–4708.
29. Subramanian C, Robertson ES. The metastatic suppressor Nm23-H1 interacts with EBNA3C at sequences located between the glutamine- and proline-rich domains and can cooperate in activation of transcription. *J Virol.* 2002;76(17):8702–8709.
30. Koehne G, et al. Quantitation, selection, and functional characterization of Epstein-Barr virus-specific and alloreactive T cells detected by intracellular interferon-gamma production and growth of cytotoxic precursors. *Blood.* 2002;99(5):1730–1740.
31. Tao Q, Yang J, Huang H, Swinnen LJ, Ambinder RF. Conservation of Epstein-Barr virus cytotoxic T-cell epitopes in posttransplant lymphomas: implications for immune therapy. *Am J Pathol.* 2002;160(5):1839–1845.
32. Gottschalk S, et al. An Epstein-Barr virus deletion mutant associated with fatal lymphoproliferative disease unresponsive to therapy with virus-specific CTLs. *Blood.* 2001;97(4):835–843.
33. White RE, et al. EBNA3B-deficient EBV promotes B cell lymphomagenesis in humanized mice and is found in human tumors. *J Clin Invest.* 2012;122(4):1487–1502.
34. Ibrahim HA, Menasce LP, Pomplun S, Burke M, Bower M, Naresh KN. Epstein-Barr virus (EBV) genotypes among human immunodeficiency virus (HIV)-related B-cell lymphomas and B-cell post-transplant lymphoproliferative disorders (B-PTLD)-late-onset lymphomas, especially in the HIV setting, are associated with type-B-EBV. *Eur J Haematol.* 2010;85(3):227–230.
35. Rowe DT, et al. Use of quantitative competitive PCR to measure Epstein-Barr virus genome load in the peripheral blood of pediatric transplant patients with lymphoproliferative disorders. *J Clin Microbiol.* 1997;35(6):1612–1615.
36. Beatty PR, Krams SM, Martinez OM. Involvement of IL-10 in the autonomous growth of EBV-transformed B cell lines. *J Immunol.* 1997;158(9):4045–4051.
37. Creary LE, et al. Next-generation sequencing reveals new information about HLA allele and haplotype diversity in a large European American population. *Hum Immunol.* 2019;80(10):807–822.
38. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26(5):589–595.
39. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079.