

Integrated Pan-Cancer Map of EBV-Associated Neoplasms Reveals Functional Host-Virus Interactions



Srishti Chakravorty¹, Bingyu Yan¹, Chong Wang², Luopin Wang³, Joseph Taylor Quaid¹, Chin Fang Lin⁴, Scott D. Briggs¹, Joydeb Majumder⁵, D. Alejandro Canaria⁶, Daniel Chauss⁷, Gaurav Chopra⁵, Matthew R. Olson⁶, Bo Zhao², Behdad Afzali⁷, and Majid Kazemian^{1,3}

Abstract

Epstein-Barr virus (EBV) is a complex oncogenic symbiont. The molecular mechanisms governing EBV carcinogenesis remain elusive and the functional interactions between virus and host cells are incompletely defined. Here we present a comprehensive map of the host cell-pathogen interactome in EBV-associated cancers. We systematically analyzed RNA sequencing from >1,000 patients with 15 different cancer types, comparing virus and host factors of EBV⁺ to EBV⁻ tissues. EBV preferentially integrated at highly accessible regions of the cancer genome, with significant enrichment in super-enhancer architecture. Twelve EBV transcripts, including LMP1 and LMP2, correlated inversely with EBV reactivation signature. Overexpression of these genes significantly suppressed viral reactivation, consistent with a "virostatic" function. In cancer samples, hundreds of novel frequent missense and nonsense variations in virostatic genes were identified, and variant genes failed to regulate

their viral and cellular targets in cancer. For example, one-third of patients with EBV⁺ NK/T-cell lymphoma carried two novel nonsense variants (Q322X, G342X) of *LMP1* and both variant proteins failed to restrict viral reactivation, confirming loss of virostatic function. Host cell transcriptional changes in response to EBV infection classified tumors into two molecular subtypes based on patterns of IFN signature genes and immune checkpoint markers, such as PD-L1 and IDO1. Overall, these findings uncover novel points of interaction between a common oncovirus and the human genome and identify novel regulatory nodes and druggable targets for individualized EBV and cancer-specific therapies.

Significance: This study provides a comprehensive map of the host cell-pathogen interactome in EBV⁺ malignancies.

See related commentary by Mbulaiteye and Prokunina-Olsson, p. 5917

Introduction

EBV is ubiquitous, infecting approximately 90% of the world's adult population (1). Most acute infections are asymptomatic, following which, EBV establishes latency. In the latency phase, EBV infection is associated with substantial morbidity and mortality as it causes a wide range of lymphocytic and epithelial malignancies, such as Burkitt and diffuse large B cell lymphomas (DLBCL), nasopharyngeal carcinoma (NPC), and stomach ade-

nocarcinoma, in both immune-competent and immune-compromised hosts (2). Nearly 150,000 cancer-related deaths can be attributed to EBV annually (3), partially due to the lack of effective, EBV-specific treatment options tailored to individualized cancers. Over the five decades since the discovery of this multidisease-associated oncovirus, various aspects of its life cycle and host immune response in individual tumor types or cell lines have been studied (4–11). However, the generalized cancer and tumor type-specific characteristics of EBV and the host cells' response, that is, the host-virus interaction in a large cohort of patients with cancer, have not been systematically delineated.

Several factors affect the life cycle of EBV, including viral integration into the host genome, expression, and mutation(s) of viral genes and the host response to the presence of virus. Although some of these factors in individual tumor types or cell lines have been studied, how they interact to promote tumorigenesis remains poorly understood. For example, despite typically being maintained in an episomal state within infected cells, several studies have demonstrated EBV integration into the genome of individual cancer types and cell lines (4, 12, 13), but the extent and the location of these integrations across different cancer types remain poorly understood. Moreover, EBV reactivation is one of the major risk factors for developing cancers and several viral and nonviral factors are known to reactivate EBV (8, 14, 15). However, how genetic factors such as variants in viral coding genes influence this has been less explored.

¹Department of Biochemistry, Purdue University, West Lafayette, Indiana.

²Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts.

³Department of Computer Science, Purdue University, West Lafayette, Indiana.

⁴Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, Indiana.

⁵Department of Chemistry, Purdue University, West Lafayette, Indiana.

⁶Department of Biological Science, Purdue University, West Lafayette, Indiana.

⁷Immunoregulation Section, Kidney Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, Maryland.

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corresponding Author: Majid Kazemian, Purdue University, West Lafayette, IN 47906. Phone: 765-494-9350; E-mail: kazemian@purdue.edu

Cancer Res 2019;79:6010–23

doi: 10.1158/0008-5472.CAN-19-0615

©2019 American Association for Cancer Research.

Moreover, to date, there has been no systematic pan-cancer genomic analysis of this widespread, multidisease-associated oncovirus to delineate shared and tumor-specific characteristics of EBV and the host response.

High-throughput DNA- and RNA-sequencing technologies capture genetic and transcriptional profiles in host cells, which can be leveraged to interrogate not only the host genome but also any pathogen genomes infecting the host cells (4, 7, 16, 17). Because EBV establishes a long-term persistence within host cells, we reasoned that large numbers of RNA-sequencing data from multiple EBV-associated tumors would provide a platform to interrogate EBV characteristics and interactions within the tumor environment. Such integrated pan-cancer genomic analysis could establish a host-virus "interactome map" essential for the design of novel generic and/or cancer-specific treatments for EBV-associated cancers. Here, we constructed a comprehensive "interactome map" by analyzing publicly available RNA-sequencing samples from more than 1,000 samples across multiple cancer types. This comprised detailed viral and host expression and mutation profiles for each cancer and cancer type associated with EBV as a resource for discovery-led science. We present our initial findings exploring new facets of host-virus interaction in cancer. Notably, we identified a set of EBV integration loci across multiple cancer types, with association of those integrations to the most active regulators of the human genome, that is, super-enhancers (SE). In addition, we found large numbers of novel missense and nonsense variants in a subset of viral genes that are negatively linked to EBV reactivation. Finally, we uncovered that EBV⁺ cancer types could be dichotomously classified according to activation and repression of IFN signature genes in response to EBV infection.

Materials and Methods

Data sources

RNA-seq samples were collected from different sources publicly available databases as of November 2018 (Supplementary Table S1). The samples include angioimmunoblastic T-cell lymphoma (18), nasal NK/T-cell lymphoma (19), endemic Burkitt lymphoma (5), sporadic Burkitt lymphoma (20), cutaneous T-cell lymphoma (21), mantle cell lymphoma (22), NPC (23), and healthy donor lymph nodes (24, 25). Normal nasopharyngeal tissue samples were from BioProject (accession number: PRJNA283839). The samples from diffuse large B-cell lymphoma, acute myeloid leukemia, and stomach adenocarcinoma were collected from The Cancer Genome Atlas (TCGA; dbGaP study accession number: phs000178). Additional samples for diffuse large B-cell lymphoma were from dbGaP (study accession number: phs000532). The samples for peripheral T-cell lymphoma and anaplastic large cell lymphoma were obtained from dbGaP (study accession number: phs000689). The samples for follicular lymphoma were from dbGaP (study accession number: phs000533) and BioProject (accession number: PRJNA278311). Samples for chronic lymphocytic leukemia were from dbGaP (study accession number: phs000767). EBV-transformed lymphocytes and healthy donor blood samples were from Genotype-Tissue Expression Project (GTEx; dbGaP study accession number: phs000424).

Cell lines and cell culture

The EBV⁻ Gastric Carcinoma (GC) cell line SNU-1 (CRL-5971, ATCC), EBV⁺ GC cell line SNU-719 (CSC-C9698L, Creative

Bioarray), and B-lymphoblastic cell line (LCL-358, catalog no. 1038-3754NV17, Astarte Biologics and GTEX-UPJH-0001-SM-3YRE9, GTEx) were cultured in RPMI1640 (VWRL0105-0500) media supplemented with 10% heat-inactivated FBS (Hyclone) and 1% penicillin/streptomycin (Life Technologies). Cells were freshly purchased, were tested negative for *Mycoplasma* contamination according to PCR Mycoplasma Detection Kit (ABM Inc.), and were used at low passage (<10) number, but were not independently authenticated.

RNA isolation and quantitative real-time PCR

Total RNA was extracted using Direct-zol RNA extraction kit with DNase I treatment (Zymo Research) and reverse transcribed into cDNA using OneScript Plus cDNA Synthesis SuperMix (ABM Inc.). The qRT-PCR was performed on Bio-Rad CFX connect. All experiments were performed in independent triplicates in total reaction volumes of 15 μ L using BrightGreen 2X qPCR MasterMix-No Dye (ABM Inc.). The relative quantities for each sample were first determined using a standard calibration curve generated by four 10-fold serial dilutions, then normalized to that of indicated housekeeping gene (e.g., *UBC*) in the same sample, and finally calibrated to the control samples. All gene-specific primers used in this study are listed in Supplementary Table S2.

Transient transfection

All plasmids were purified using ZymoPURE Plasmid MidiPrep Kit (catalog no. D4201). Lymphoblastoid cell lines (LCL) were resuspended in pre-warmed complete media at confluency of 1 million/mL a day prior to the transfection. At the day of transfection, cells were washed twice with 1 \times PBS to get rid of residual antibiotics. For each experiment, 5–7 \times 10⁶ cells were transfected with concentration of 1 μ g plasmid per million cells using Kit V, Amaxa Nucleofector IIb, program T-20, according to the manufacturer's instructions. Following electroporation, cells were recovered in pre-warmed antibiotic-free RPMI media supplemented with 20% heat-inactivated FBS. After 5–6 hours, equal volume of prewarmed complete media supplemented with 20% heat-inactivated FBS was added to the cells and incubated in 5% CO₂ at 37°C. Twenty-four hours posttransfection, cells were stimulated with doxorubicin.

EBV reactivation and cell treatment

To induce EBV lytic cycle, LCLs were treated with 200 nmol/L doxorubicin (Millipore Sigma, catalog no. 44583) or control carrier DMSO (VWR, catalog no. 97063-136) for 48 hours, after which, cells were collected, fixed, permeabilized, and stained for intracellular antibodies. EBV reactivation in gastric carcinoma cells (SNU-1, SNU-719) was triggered with 20 ng/mL of 12-O-Tetradecanoylphorbol-13-acetate (TPA; Millipore Sigma, catalog no. 1585) and 3 mmol/L sodium butyrate (NaB; Millipore Sigma, catalog no. B5887) for 12 hours. IFN γ treatment was performed using 100 IU/mL Recombinant Human IFN γ (carrier free, BioLegend, catalog no. 570206) for 12 hours prior to staining cells for flow cytometry.

Primary human B-cell culture and infection

Discarded, de-identified human peripheral blood mononuclear cells (PBMC) left in apheresis tube after platelet donation were obtained from Dana-Farber Cancer Institute blood donor center (Boston, MA). PBMCs were first purified by centrifugation over a density gradient. Primary B lymphocytes were negatively selected

using EasySep Human B Cell Enrichment Kits (StemCell Technologies, #19054 and 15064) following the manufacturer's instruction. Cells were suspended in RPMI1640 medium (Life Technologies) supplemented with 10% FCS. Equal amount of EBV^{WT} (B958) and EBV^{ΔEBNA2/ΔEBNALP} (P3HR1; entire EBNA2 and the last two exons of EBNALP deleted) were used to infect B lymphocytes. Total RNAs were prepared three days post infection. qRT-PCR was used to quantitate the expression of lytic genes (see Supplementary Table S2 for primers). EBNA1 was used as internal control to normalize the RNA loading and B958 levels were set to 1.

CRISPR-mediated EBNA1 knockout

sgRNAs targeting EBNA1 were cloned into lentiGuide-Puro vector and packaged into lentiviruses by cotransfecting with VSVG and psPAX2. Harvested lentiviruses were used to infect GM12878 stably expressing CAS9 for 2 days. Cells infected with lentiviruses were selected with puromycin. The efficiency of EBNA1 knockout was validated by Western blotting at 3 days after puromycin selection. After knockout EBNA1 for 7 days, cells were harvested and used for qRT-PCR validations.

Plasmids

EBV genes *EBNA3A*, *EBNA3C*, *LMP1*, and *LMP2A* were PCR amplified from commercially available plasmids from Addgene (catalog no. 37956, 37958, 37962, 37963, respectively). *LMP2B* was separately amplified from LCL. All genes were cloned into pLVX-EF1a-IRES-zGreen (catalog no. 631982, Clontech Laboratories, Inc.) using HiFi DNA Assembly cloning (catalog no. E2621, New England Biolabs; NEB) and verified using Sanger sequencing. The *LMP1* point mutations were generated using specific primers (Supplementary Table S2) by the Q5 Site-Directed Mutagenesis Kit (catalog no. E0554S, NEB) according to the manufacturer's instructions.

Antibodies and flow cytometry

All data acquisition was carried out on Attune NxT Flow Cytometer (Thermo Fisher Scientific) in a final staining volume of 200 μ L. The flow data were analyzed using FlowJo (v10.4.2). Appropriate internal controls were used to assign gates. Fluorescence from multiple antibodies were compensated using AbC Total Compensation beads (Thermo Fisher Scientific, catalog no. A10497). Two percent human serum (H4522, Millipore Sigma) was used for Fc blockade in all experiments before surface and intercellular staining. In all experiments, cells were assessed for viability by staining with Fixable viability dye eFluor780 (Life Technologies, catalog no. 65-0865-14; 1:2,000 dilution) as per the manufacturer's instructions. PD-L1 surface marker staining (CD274; clone 29E.2A3; BioLegend catalog no. 329714; 1:60 dilution) was performed using standard protocol in 1 \times PBS. Cells were then fixed with 4% methanol-free formaldehyde (Thermo Fisher Scientific, catalog no. 28908) for 20 minutes to retain intracellular levels of GFP proteins prior to permeabilization using FoxP3/Transcription Factor Staining Buffer Set (eBioscience, catalog no. 5523). Antibodies for staining for intracellular proteins were IDO1 (clone eyedio; eBioscience, catalog no. 12-9477-41; 1:60 dilution), BZLF1 (Santa Cruz Biotechnology, catalog no. sc-53904; 1:60 dilution), LMP1 (clone LMPO24; Novus Biologicals, catalog no. NBP2-50383; 1:60 dilution), and anti-HA.11 (clone 16B121; BioLegend, catalog no. 901509; 1:500 dilution).

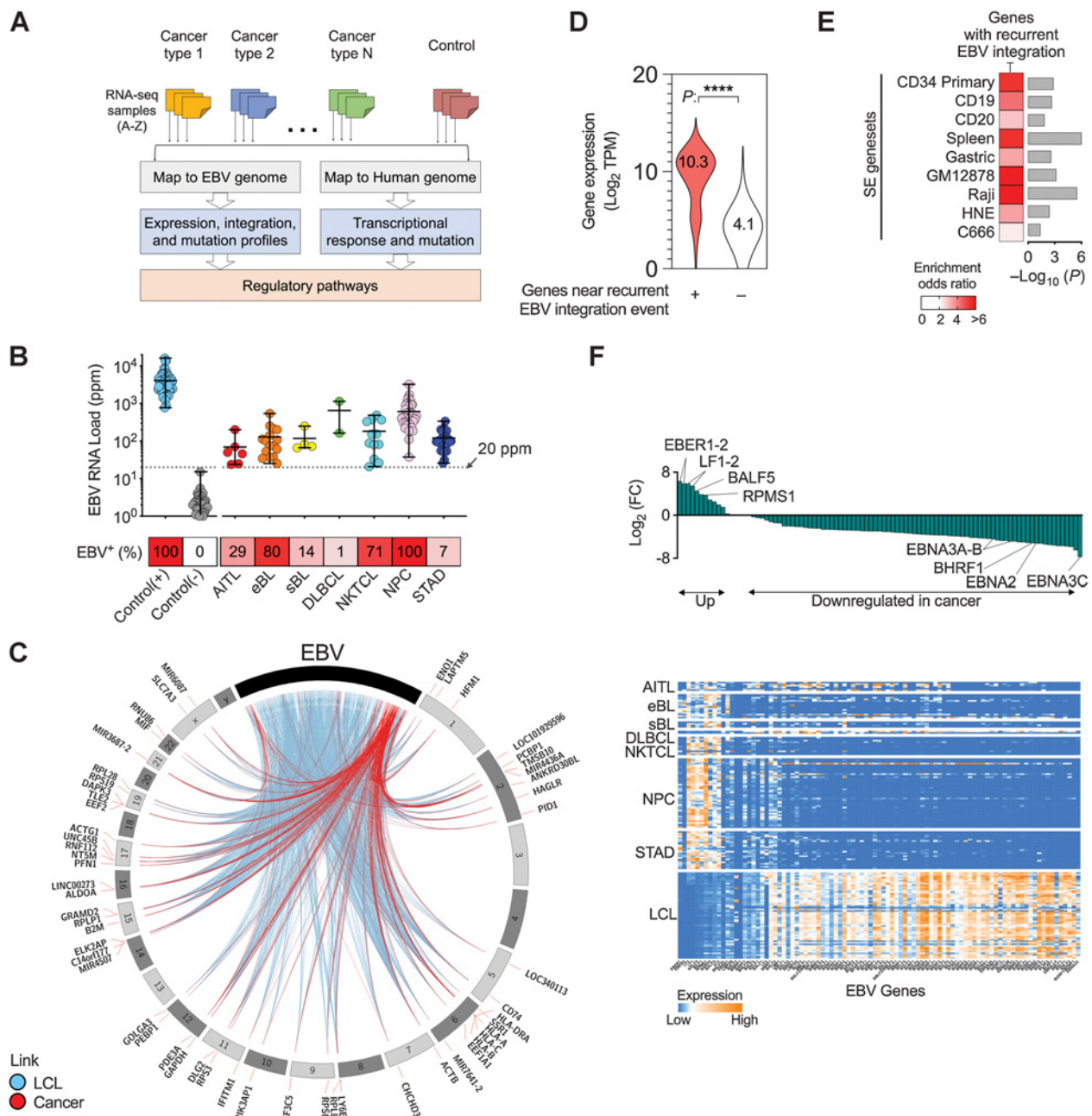
Western blotting

Nearly 4 million cells per sample were collected and lysed using 100 μ L of Pierce RIPA buffer (Thermo Fisher Scientific, catalog no. 89900) containing 1 \times Halt Protease Inhibitor Cocktail, EDTA-free (Thermo Fisher Scientific, catalog no. 78425) as per manufacturer's instructions. The cell lysates were additionally sonicated 30 seconds at 80% amplitude. Protein concentration in the lysate was determined using the Pierce BCA Assay Kit (Thermo Fisher Scientific, catalog no. 23227) and approximately 10 μ g of protein was loaded per well of a 10% SDS-PAGE gel (1 mm). Proteins were then transferred to a polyvinylidene difluoride membrane using the Semi Dry Transfer Apparatus (Bio-Rad). To prevent nonspecific antibody binding, membranes were incubated with 5% milk in 1 \times TBS blocking solution for 1 hour at room temperature. Membranes were washed three times for 10 minutes with 1 \times TBST and were then incubated overnight with primary antibodies. BZLF1 (1:500, sc-53904, Santa Cruz Biotechnology), Vinculin (1:1,000, sc-73614, Santa Cruz Biotechnology), or GAPDH antibodies were commercially obtained (catalog no.: ab8245). OT1X mAb against EBNA1 was a gift from Dr. Jaap Middeldorp (Amsterdam University Medical School, Amsterdam, the Netherlands). The next day, membranes were washed and incubated with HRP-linked anti-mouse secondary antibody (1:3000, 7076S, Cell Signaling Technology) for 1 hour at room temperature, followed by washing and developing the blots using ECL (Thermo Fisher Scientific, catalog no. 32209).

Results

EBV integrates at highly accessible genomic loci of the host enriched in SE architecture

We collected 291 control and 1,051 RNA-seq samples across 15 different cancer types from publicly available databases as of November 2018 (Supplementary Table S1A; refs. 5, 18–25). We constructed a comprehensive host–virus interactome map for each cancer type by analyzing sequencing reads aligned to the EBV and to the human genomes (Fig. 1A and also Materials and Methods). We first defined viral reactivation signature (viral RNA load) by counting the fraction of sequencing reads mapping to the EBV genome (parts per million reads of library size, ppm; ref. 16) in each cancer sample (Supplementary Table S1B). An EBV detection threshold of 20 ppm was selected to denote EBV-infected samples because at this threshold 100% of the 57 positive controls (EBV-transformed lymphoblastoid cell lines; LCL) and 234 negative control samples (healthy blood cells and lymph nodes) were detected as EBV⁺ and EBV⁻ respectively, corresponding to 100% specificity and sensitivity (Supplementary Fig. S1A). In other words, all EBV⁺ samples had ppm > 20 and all negative control samples (healthy donor blood samples from GTEx and healthy donor lymph nodes from; refs. 24, 25) had ppm < 20 (mean ppm = 0.5). The frequency of EBV infection depended on cancer type, being highest in NPC (100%), endemic Burkitt lymphoma (80%), and NKT cell lymphoma (NKTCL; 71%) and absent in eight other tested cancers [peripheral T-cell lymphoma (PTCL), cutaneous T cell lymphoma (CTCL), anaplastic large cell lymphoma (ALCL), follicular lymphoma, mantle cell lymphoma (MCL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and multiple myeloma; Fig. 1B; Supplementary Table S1A]. The percentage of EBV⁺ tumors was consistent with the known frequency of EBV in these

**Figure 1.**

EBV is integrated at limited loci across multiple tissues and a subset of its genes are preferentially expressed in cancer. **A**, Schematic showing pipeline for generation and analysis of EBV-host interaction map. **B**, EBV RNA load (dotplot, top) and the percentage of samples (heatmap, bottom) with ppm ≥ 20 across multiple cancers (shown are mean and min-to-max range). **C**, Circos plot showing integration of EBV genes (black bar) into human DNA (human chromosomes shown in gray). Connecting lines are color-coded to differentiate LCLs (blue) and cancer (red) samples; nearest human genes to the site of integration are annotated. **D**, Expression distribution of all expressed human transcripts adjacent to EBV integration sites across all EBV⁺ samples (median is shown; two-tailed Wilcoxon test). **E**, Enrichment odds ratios and *P* values (Fisher exact test) for SE regulation of genes that are preferred sites for EBV integration. Shown are data from multiple healthy tissue/cell types assessed or EBV⁺ cancer cells (data from refs. 31, 34, 65). **F**, Expression of EBV genes (columns) across multiple cancers and LCL samples (rows; bottom) with bar-chart (top) showing the log-fold change between the mean gene expression in cancer versus LCLs. eBL, endemic Burkitt lymphoma; sBL, sporadic Burkitt lymphoma; STAD, stomach adenocarcinoma. ****, *P* < 0.0001.

tumors (26). In EBV⁺ cancers, the highest and lowest EBV RNA burden was in NPC (mean = 613 ppm) and angioimmunoblastic T-cell lymphoma (AITL; mean = 69 ppm), respectively (Fig. 1B).

EBV typically persists as an episome in the host cells but can also integrate into the cancer genome (4, 12, 27). Chimeric RNA-seq reads that partially map to the EBV genome and partially to the human genome represent viral integration sites (28). We used

these chimeric reads to determine the frequency and location of EBV integration into the human genome across different cancers (see Supplementary Methods). Sixty out of 112 EBV⁺ cancers (corresponding to 56 unique genomic loci) showed evidence of viral integration (median of 3 integration events per sample; Fig. 1C; Supplementary Table S1C and S1D), with significant over-representation near ribosomal RNA genes and in proximity to promoters (3.8-fold enrichment). We annotated these 56 specific integration loci with their nearest genes and analyzed their functional properties. These include integration at *B2M*, *CD74*, and *HLA-C* loci, which are part of the MHC class I and II complexes (Supplementary Table S1C). Because the integration events were inferred from RNA-sequencing data in LCLs or EBV⁺ cancer samples of any origin, we decided to use an orthogonal method to verify these events. To this end, we turned to an existing whole-genome long-read DNA-sequencing (as opposed to RNA-sequencing) data from an independent LCL (GM12878; (29)). We identified chimeric viral-human reads that were mapped to 20 of these integration loci, independently verifying these events (Supplementary Table S1C, see Supplementary Methods). The top three pathways enriched in these genes were viral processes ($P \leq 1e-10$), response to type I IFNs ($P \leq 1e-9$), and regulation of apoptotic pathways in response to DNA damage ($P \leq 1e-5$; Supplementary Fig. S1B). Enrichment for viral process genes was attributable to integration near genes encoding human ribosomal proteins controlling RNA translation (*RPL8*, *RPL28*, *RPS3*, *RPS6*, *RPS19*, and *RPLP1*), which is intriguing because EBV recruits and subverts this machinery to translate its own mRNA as a survival strategy (30). Host genes near preferred EBV integration sites (integration sites supported by more than one sample) were expressed on average 20-fold higher than other expressed genes ($P < 1e-16$; Fig. 1D). Integration events not only did not interrupt the expression of these already highly expressed genes, but seemed to slightly augment their expression (Supplementary Fig. S1C), suggesting that highly accessible chromatin may harbor more recurrent EBV integration events. In cancer cells, clusters of regulatory elements called SEs are generated near highly expressed genes important to tumorigenesis (31). We reasoned, therefore, that SE regions could be attractive to EBV integration, as seen in other viruses, for example, human hepatitis and herpes viruses (32, 33). EBV integration sites significantly overlapped with SEs annotated in multiple immune cells and tissues (31), as well as EBV⁺ cancer cells (Fig. 1E, see Supplementary Methods; ref. 34). For example, SE architecture existed in the vicinity of approximately 12% of all genes in at least one screened tissue type; this compares with 45% (25 of 56) of EBV-integrated genes using the same screening (Fisher exact $P = 1e-9$; OR = 4.6). Collectively, these data indicate that EBV preferentially integrates near highly expressed genes with SE architecture.

Virostatic EBV genes are frequently mutated in cancer

EBV establishes latent and lytic infection during its lifecycle, which is regulated by expression of specific viral genes. To determine how EBV behaves in cancerous tissues, we next compared expression of EBV genes in cancerous cells to EBV-transformed LCLs. Most EBV genes expressed in LCLs derived from healthy donors were either transcriptionally silent or extremely low-expressed in EBV⁺ cancer samples of any origin (Fig. 1F). These included EBV nuclear antigen genes *EBNA2*, *3A*, *3B*, and *3C*, all of which are essential for EBV transformation (35) but potentially

redundant for other EBV life-cycle stages. While expression of some EBV genes appeared to be cancer-specific, such as EBV-encoded small RNAs *EBER1* and *EBER2* in stomach adenocarcinoma and endemic Burkitt lymphoma, a handful of genes including *LF2*, *BALF5*, *BILF1*, *BARF0*, *RPMS1*, and *A73* were ubiquitously expressed in all cancer types (but minimally or infrequently expressed in LCLs) and thus have potential as prognostic and/or diagnostic markers for these diseases (Fig. 1F; Supplementary Table S1E).

EBV controls the lytic phase of its life cycle by orchestrating the expression of various viral genes that have the capacity to prevent EBV reactivation from latency (15, 36, 37). To determine which viral genes are the most influential to this regulation and to what extent, we initially correlated expression of EBV genes to viral RNA load in 57 LCLs. Twelve genes were strongly inversely correlated with viral RNA load in LCLs (Bonferroni-adjusted $P < 0.0005$ of Spearman correlation; Fig. 2A and B), suggesting potential function as inhibitors of virus replication. *EBNA1* had the highest negative correlation with viral RNA load, consistent with its reported function in inhibiting spontaneous viral reactivation (15). Other "virostatic" genes in order of their inverse correlation to the viral RNA load included *EBNA3C*, *EBNA2*, *EBNA-LP*, latent membrane proteins *LMP2B*, *LMP2A*, *BHRF1*, *BARF0*, *RPMS1*, *LMP1*, *BALF5*, and *EBNA3A*, some of which were also previously shown to inhibit virus lytic cycle induction (38, 39). To confirm the ability of these genes to suppress lytic viral replication, we performed three experiments. First, we cloned five of them (*LMP1*, *LMP2A*, *LMP2B*, *EBNA3A*, *EBNA3C*) that represented a mixture of high and low negative correlations with viral RNA load. We overexpressed these genes in LCL-358 and induced EBV lytic replication with doxorubicin (Supplementary Fig. S2A and S2B). All five genes significantly suppressed *BZLF1* (Fig. 2C) and *BRLF1* expression (Fig. 2D), both of which are immediate-early viral lytic genes. The same results were obtained in an independent LCL, GTEX-UIPH (Supplementary Fig. S2C). Second, we infected primary human B cells with P3HR1 EBV, an EBV strain with whole gene deletion of *EBNA2* and deletion of the last two exons of *EBNA-LP*, and showed that P3HR1 expressed higher levels of viral lytic genes, compared with cells infected with wild-type EBV (Supplementary Fig. S2D), confirming the virostatic function of these EBNA genes. Finally, we knocked out *EBNA1* expression in another independent LCL, GM12878, using two separate single guide RNAs, and showed that cells with *EBNA1* knockout have higher levels of viral lytic genes (Supplementary Fig. S2E). Collectively, these data indicate that a subset of EBV-encoded genes have virostatic function.

We reasoned that loss of virostatic genes, for example, via inactivating variations, may be conducive to heightened virus expression in tumors. Five of these genes (*EBNA-2*, *-3A*, *-3C*, *-LP*, *BHRF1*) were either not expressed or very low in most cancers (Figs. 1F and 3A). To determine whether the remaining seven harbor inactivating variations in cancer (i.e., gene mutations that occur in cancer or EBV strains that carry mutations), we cataloged all recurrent variants in EBV genes from 112 EBV⁺ cancer and 57 LCLs (all EBV-subtype I; Supplementary Table S3; ref. 40). Variants were selected to be either homozygous (meaning that all the expressed EBV genomes in the cell contain the variant) or heterozygous (meaning that at least half of all the EBV genomes in the cell contain the variant). We detected 562 recurrent nonsynonymous variations across all EBV⁺ cancers, half of which were shared across two or more cancer types (Fig. 3). As expected, the

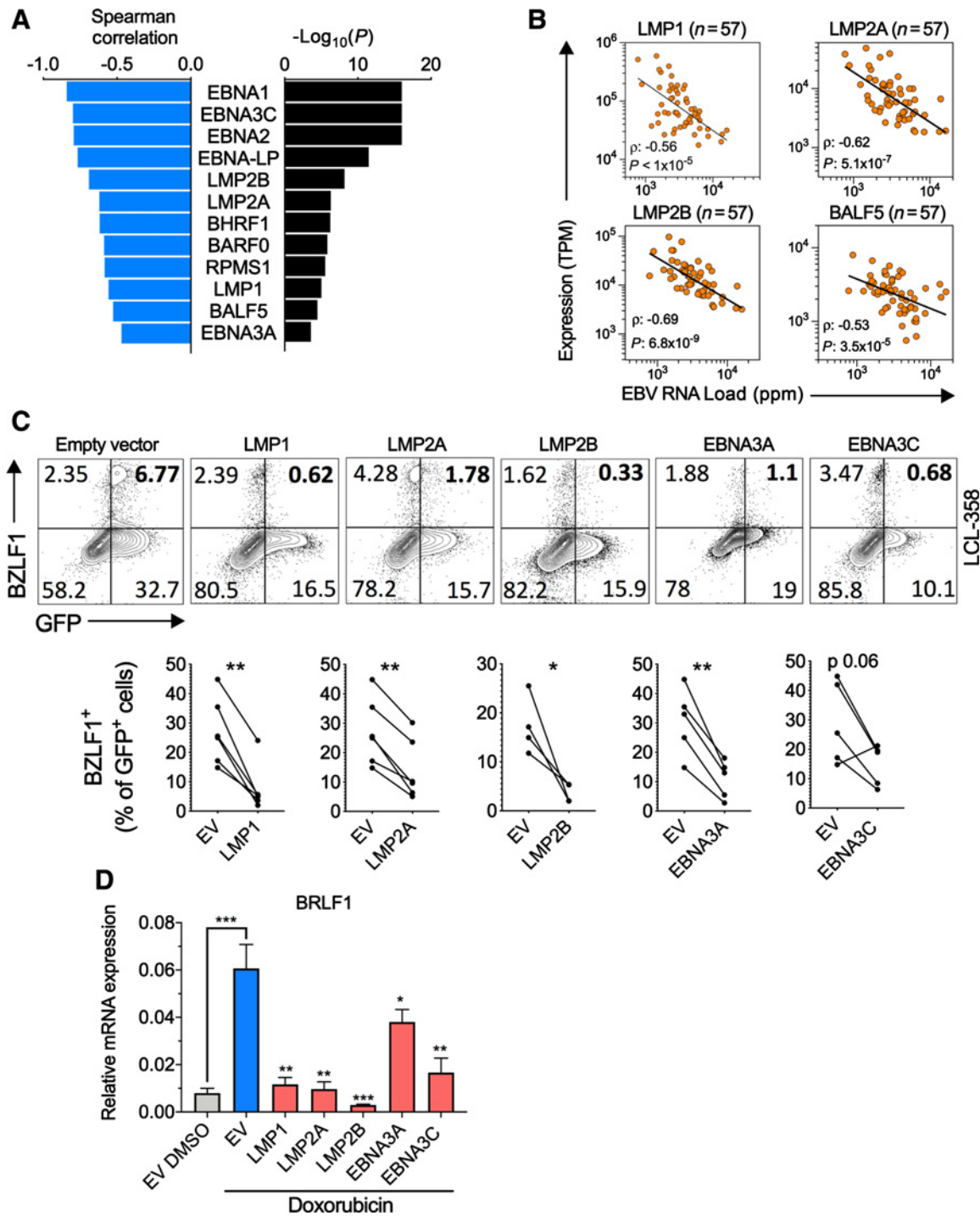


Figure 2. A subset of EBV genes are negatively correlated with EBV RNA load in 57 LCLs. **A** and **B**, Inverse correlation between expression of 12 EBV gene transcripts and EBV RNA load in 57 LCLs. Shown are Spearman rank correlation coefficients (**A**, left) and P values (**A**, right) and representative scatter plots for *LMP1*, *LMP2A*, *LMP2B*, and *BALF5* (**B**). **C**, Immediate early viral lytic gene expression (BZLF1) in LCL-358 transfected with empty vector or vectors encoding five representative virostatic genes. Shown are representative flow cytometry plots (top) and cumulative data showing % BZLF1⁺ of transfected (GFP⁺) LCLs with the indicated overexpression plasmids (bottom). Data are from $n \geq 4$ independent experiments; gating strategy is shown in Supplementary Fig. S2A. **D**, Expression of *BZLF1* mRNA, another immediate early viral lytic gene, in LCLs transfected with empty vector (EV) or vectors encoding indicated genes. Bars show mean \pm SD from $n = 3$ independent experiments; comparisons were against EV treated with doxorubicin. P values are from two-tailed ratio paired t test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

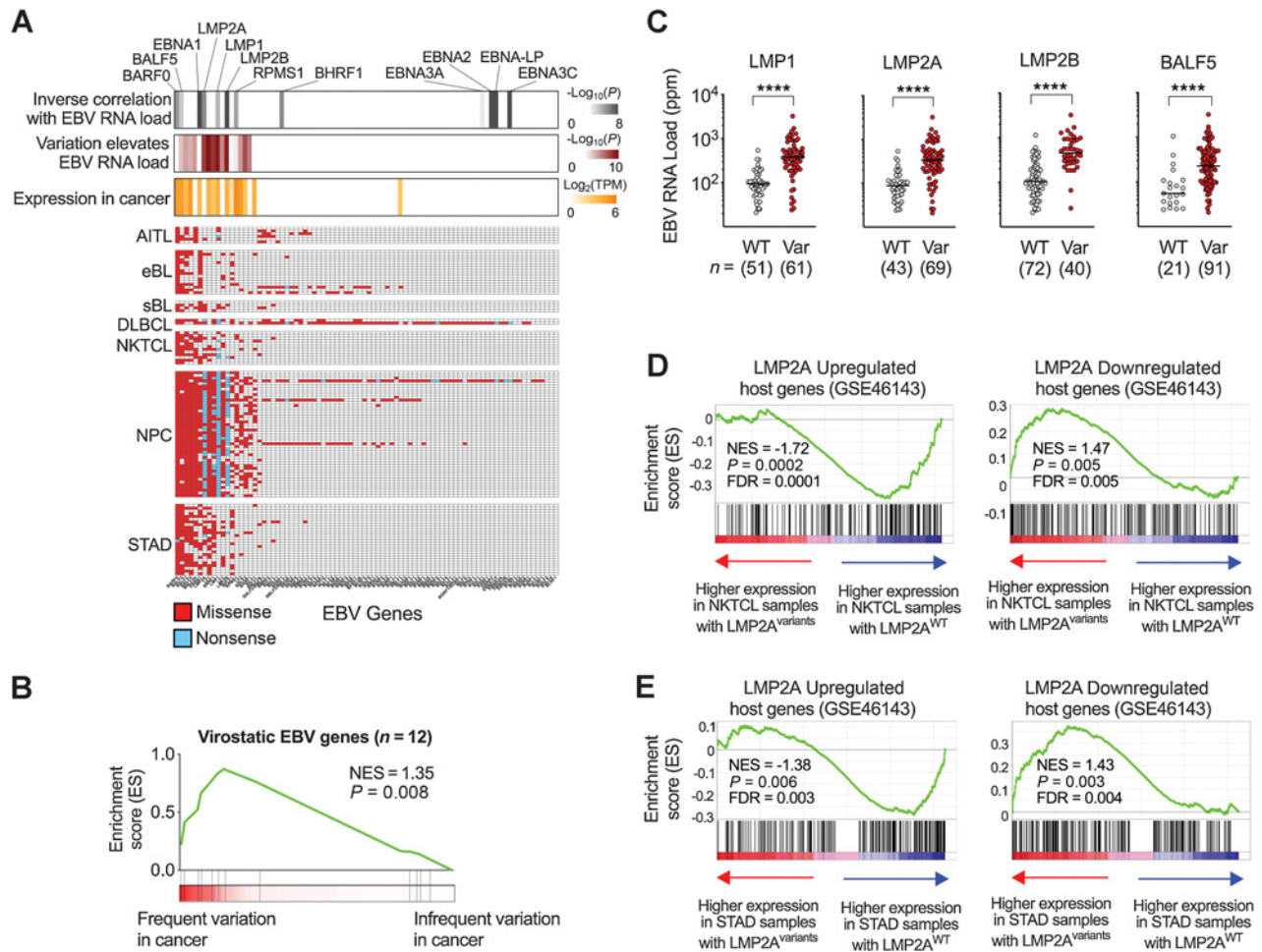


Figure 3. Virostatic EBV genes have frequent variations in cancer and affect expression of host genes. **A**, Bar heatmaps showing EBV genes correlated with increased viral RNA load (first bar; gray coded), samples in which missense or nonsense variation in EBV genes is associated with elevated viral RNA load (second bar; brown coded) and expression of EBV genes in cancer (third bar; orange coded). Table matrix below shows missense (red) or nonsense (blue) variations in all EBV genes (column) across all cancer samples (row) ordered by frequency. **B**, GSEA plot for EBV genes associated with 12 identified protein-coding virostatic genes, comparing EBV genes with frequent or infrequent variations in cancer. Normalized enrichment score (NES) is shown. **C**, Representative example of EBV RNA load in EBV⁺ cancer samples of any origin with (Var) or without (WT) missense or nonsense variants in *LMP1*, *LMP2A*, *LMP2B*, and *BALF5*. Shown are individual samples (dots) with median values. **D** and **E**, GSEA plots for host genes up- (left) and downregulated (right) by *LMP2A* (GSE46143), comparing *LMP2A*^{WT} versus *LMP2A*^{variant} NKTCL (**D**) or stomach adenocarcinoma (STAD; **E**) samples. ****, $P < 0.0001$, two-tailed Wilcoxon test.

majority of LCLs had no, or scant EBV variations (Supplementary Table S3). Across different cancers, EBV had a mean of 5.8 mutations/kb (SD = 2.6), with NPC the highest (8.0 mutations/kb; Supplementary Fig. S3A). Most variations were missense (35.9%), noncoding (15.7%), or synonymous (47.5%) and only rarely nonsense (0.9%; Supplementary Fig. S3B). The mean transition/transversion ratio was 1.36 (Supplementary Fig. S3C). Mutational signatures were not the commonly observed APOBEC-driven pattern (TCW>TT/GW, where W is enriched in T or A; refs. 41, 42) but, rather, substitutions of C>T at NpCpG (which may suggest a possible effect on methylation sites across the EBV genome) and T>C at NpTpG trinucleotides across multiple cancer types (Supplementary Fig. S3D and S3E).

Virostatic genes harbored on average four times more variations than other EBV genes in cancer (Fig. 3A and B; Supplemen-

tary Fig. S3F). Six of the virostatic EBV genes (Fig. 3A; top bar; gray coded), *BARF0*, *BALF5*, *EBNA1*, *LMP1*, *LMP2A*, and *RPMS1*, were both expressed (Fig. 3A; third bar; orange coded) and had frequent variations (Fig. 3A; variation matrix) in cancer. In the majority of these genes, samples with variations had elevated EBV burden (Fig. 3A; second bar; brown coded), suggesting that these are pathogenic viral proteins whose variation in cancer results in elevated viral RNA loads. For example, *LMP1*, *LMP2A*, *LMP2B*, and *BALF5* whose expressions were strongly inversely correlated with EBV RNA load in LCLs (Fig. 2B), had a high rate of nonsense or missense variations and cancer samples with variations in these genes had significantly higher EBV RNA load (Fig. 3C).

To determine the functional consequence of nonsense and missense variations in virostatic EBV genes to the host *in vivo*, we selected *LMP2A* as an exemplar and focused on NKTCL and

stomach adenocarcinoma, which were the cancer types in our dataset for which we had both wild-type and variant LMP2A tissue samples. From the dataset of Vockerodt and colleagues, GSE46143 (43), we sourced the top 200 host genes upregulated and the top 200 host genes downregulated by LMP2A. We determined whether expression of host genes regulated by this protein were different between NKTCL and stomach adenocarcinoma cancers that had wild-type or variant LMP2A, using gene set enrichment analysis (GSEA; ref. 44). We found that NKTCL and stomach adenocarcinoma samples with wild-type LMP2A had higher expression of LMP2A-upregulated genes and lower expression of LMP2A-downregulated genes compared with samples with variant LMP2A, consistent with attenuation of LMP2A function in tissues containing the variant proteins (Fig. 3D and -E). Taken together, these data indicated frequent missense and nonsense variations in virostatic EBV genes that potentially interrupt viral gene function and lead to higher viral RNA load in tumor samples.

Frequent nonsense variations in LMP1 fail to restrict viral reactivation

We showed that LMP1 represses EBV reactivation (Fig. 2B and C), has frequent variation in cancer (Fig. 3A), and tumor samples with variations in LMP1 have higher viral RNA load than those with wild-type LMP1 (LMP1^{WT}). Importantly, nearly all EBV⁺ NPC and DLBCL samples, and one-third of NKTCL samples, had one or two frequent nonsense variations in the LMP1 protein, Q322X and G342X. To elucidate the significance of these LMP1 variations, we reactivated EBV in LCLs engineered by electroporation to overexpress wild-type (LMP1^{WT}) or variant (LMP1^{Q322X} or LMP1^{Q342X}) LMP1 and measured EBV reactivation by expression of BZLF1 protein (Supplementary Fig. S4). Vectors used to overexpress LMP1 were IRES-containing and bicistronic, allowing simultaneous expression of LMP1 proteins and GFP. GFP was evident in LCLs expressing either LMP1^{WT} or its variants, showing successful expression of these genes in LCLs (Fig. 4A). LMP1 protein, however, was only detectable in LCLs electroporated with LMP1^{WT}, whereas cells transfected with variants were only positive for GFP, confirming that neither variant was translated to protein because both were mutants with premature stop codons (Fig. 4A). Neither variant restricted viral reactivation, in contrast to wild-type LMP1, indicating that the consequence of both variants is loss of function (Fig. 4B). These observations were further confirmed by Western blotting (Fig. 4C). In NKTCL cancer transcriptomes, where we could directly compare tissues with wild-type versus any missense or nonsense variants of LMP1, we detected higher EBV RNA load in samples with LMP1 variation (Fig. 4D). These data confirm that these LMP1 proteins are loss-of-function variants *in vivo*.

Higher viral RNA load is associated with mutation in cancer driver genes

To determine how tumor cells respond to EBV, we next switched our focus to study genetic and transcriptional factors of the host cell. We first performed mutational pattern analysis of the host genome in EBV⁺ and EBV⁻ cancer samples. We factorized the mutational patterns into two *de novo* signatures, which we called "signature A" and "signature B" (Supplementary Fig. S5A). Signature A resembled a previously identified observation in Hodgkin lymphoma, known as signature #25 from

Mutational Signatures v2 (https://cancer.sanger.ac.uk/cosmic/signatures_v2). Signature B aligned with signatures #12 and #16 that were described in liver cancers. The etiology of these signatures remains unknown. We did not observe any significant difference in the usage of signature A or B in EBV⁺ versus EBV⁻ cancer types (Supplementary Table S4A; Supplementary Fig. S5B). Of note, the majority of stomach cancer samples show signature A regardless of EBV status, but other cancer types show the signature B mutational pattern.

We then focused on mutations in cancer driver genes. Although variation in cancer driver genes have previously been studied in the context of EBV-associated cancers (5, 20), their prevalence in EBV⁺ versus EBV⁻ cancer is less explored. Moreover, their association with EBV RNA load is unknown. A recent large-scale genomics study in pediatric endemic and sporadic Burkitt lymphoma has shown aberrant somatic hypermutation in EBV⁺ tumors (45) supporting a link between EBV and tumor mutational burden. To estimate host cancer driver gene mutations and their association with viral RNA load, we implemented a previously validated pipeline (46). We compared EBV⁺ with EBV⁻ cancers, focusing on missense/nonsense coding SNPs in approximately 200 cancer driver genes curated by COSMIC (see Supplementary Methods; ref. 47). We found that sixty-three cancer driver genes were mutated in at least 2 (out of 1,051) samples (Supplementary Table S4B and S4C). Among those, eight genes, *DDX3X*, *MYC*, *BCOR*, *ARID1A*, *TRAF3*, *EP300*, *PTEN*, and *CASP8*, were more frequently mutated in 112 EBV⁺ than 939 EBV⁻ cancers (Fisher exact test, $P < 0.05$), with the remainder cancer driver genes (e.g., *TP53* and *ID3*) not significantly different (Fig. 5A). *DDX3X* and *MYC* mutations were specific to EBV⁺ endemic Burkitt lymphoma cancers, consistent with previous reports on mutations of these genes in Burkitt lymphoma (20). Mutations in *DDX3X*, a protein that is highly intolerant of loss of function (ExAC pLOF score = 1; ref. 48) and missense mutations (ExAc Z-score = 5.13; ref. 48), predominantly affected the helicase domains at amino acids with almost universal phylogenetic conservation (Supplementary Fig. S5C, signifying that these are important residues. Indeed, seven of the nine mutations we found were predicted to be deleterious using the Polyphen-2 algorithm (Supplementary Fig. S5D) (49). Of note, endemic Burkitt lymphoma samples with higher EBV RNA load had four times higher mutation rate in either the *DDX3X* or *MYC* genes (Fisher exact test $P < 0.05$; Fig. 5B), suggesting an association between the frequency of mutations occurring in specific cancer driver genes and higher EBV burden.

EBV⁺ tumors can be dichotomously classified according to the host transcriptional response

To understand the molecular mechanisms of host responses to EBV infection, we first compared gene expression profiles of host cells between EBV⁺ and EBV⁻ cancers. NPC, endemic Burkitt lymphoma, and stomach adenocarcinoma had the highest and AITL and DLBCL had the lowest number of differentially expressed genes (DEG; fold change > 1.5; $\text{rpmk} > 4$; $\text{FDR} < 0.05$; Supplementary Fig. S6A; Supplementary Table S5A–S5G). Most DEGs were cancer-type specific; however, 556 upregulated and 591 downregulated genes were differentially expressed in at least two cancers, indicating that some of the response to EBV in cancer is shared (Supplementary Fig. S6A). To functionally categorize the response, we performed integrative pathway analysis using Metascape (50), and focused on the top 20 statistically

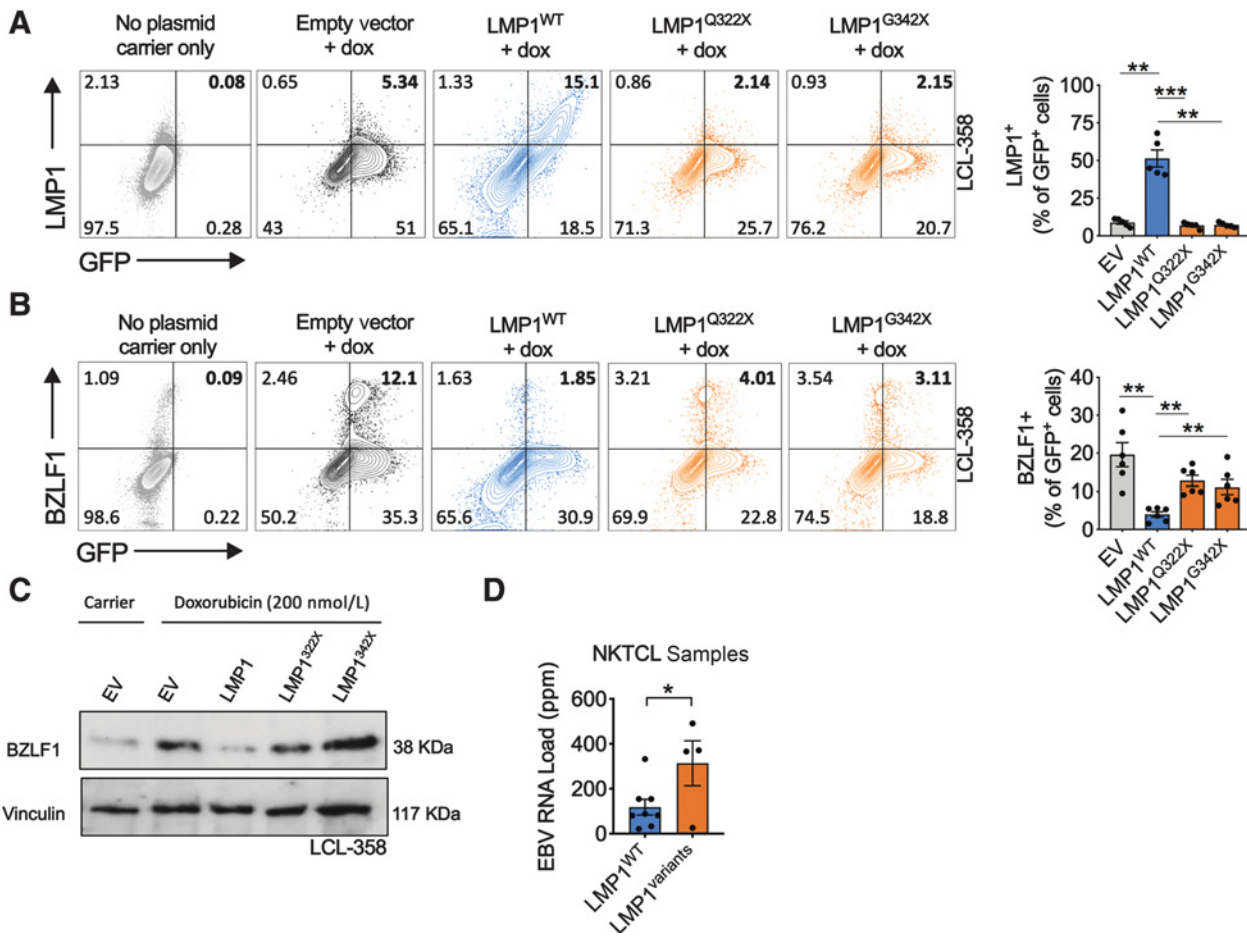


Figure 4. LMP1 variants affect virostatic function. **A**, Representative flow cytometry plots (left) and cumulative data (right) from $n = 5$ independent experiments showing LMP1 and GFP expression in LCLs cultured with doxorubicin (dox) or carrier alone (DMSO) after transfection with the indicated plasmids. **B**, Representative flow cytometry plots (left) and cumulative data (right) from $n = 5$ independent experiments showing % BZLF1⁺ of all live transfected (GFP⁺) LCL-358 with the indicated overexpression plasmids. **C**, Representative immunoblot from $n = 2$ independent experiments showing BZLF1 expression following transfection of LCL-358 cells with the wild-type or mutant LMP1. **D**, EBV RNA load in NKTCL samples with or without missense/nonsense variations in the *LMP1* gene. EV, empty vector. Bars, mean + SEM. *, $P < 0.05$ by paired ratio *t* test (**A** and **B**) and unpaired two-tailed *t* test (**D**). **, $P < 0.01$; ***, $P < 0.001$.

significant pathways (Supplementary Fig. S6B). As expected, notable shared pathways in 4 or more cancer types included regulation of cell death, the immune system and responses to pathogens (Fig. 6A; Supplementary Fig. S6A). Specific gene networks in individual cancer types are shown in Supplementary Fig. S6C–S6I. For example, regulation of mitotic cell-cycle pathway, as recently reported (5), was highly enriched in EBV-responsive DEGs of endemic Burkitt lymphoma ($P = 1e-27$; Supplementary Fig. S6E), consistent with elevated expression of the cell-cycle regulator CDKN2A. Killer Cell Lectin Like Receptor B1 (KLRB1), a protein with both inflammatory and regulatory functions, was markedly downregulated in EBV⁺ NKTCL cancers, an observation that was also made in $\gamma\delta$ T-cell lymphoproliferative disorders when compared with normal $\gamma\delta$ T-cells (Supplementary Fig. S6H; ref. 51). Likewise, in NPC, cell-to-cell adhesion was significantly altered by EBV ($P = 1e-44$; Supplementary Fig. S6D), including elevation of epithelial cell adhesion molecule (EPCAM), providing a rationale for anti-EpCAM antibody as treatment.

To identify how EBV response genes are regulated, we examined active regulatory regions in their vicinity. EBV response genes were significantly enriched in SE architecture (Supplementary Fig. S6J), for example, approximately 25% of DEGs in NPC and approximately 25% of DEGs in NKTCL have associated SE architecture in at least one of the screened immune cells as compared with expected 12% (Supplementary Fig. S6J; see Supplementary Methods; Fisher exact $P < 1e-45$; OR > 2), identifying SEs as a shared feature of gene regulation in EBV-driven cancers (52). Thus, there is a connection between EBV and host SEs at two levels. First, EBV preferentially integrates at highly accessible regions of the genome that are enriched in SEs (Fig. 1E). This may suggest a mechanism by which EBV subverts the highly active transcription machinery of the host to replicate its own genes. Second, genes that are differently expressed between EBV⁺ and EBV⁻ tumors are significantly enriched in SE architecture. These observations suggest that SEs play an important role in regulating EBV-driven host programs within tumor cells, an important observation as SE inhibitors are currently undergoing clinical trials (53).

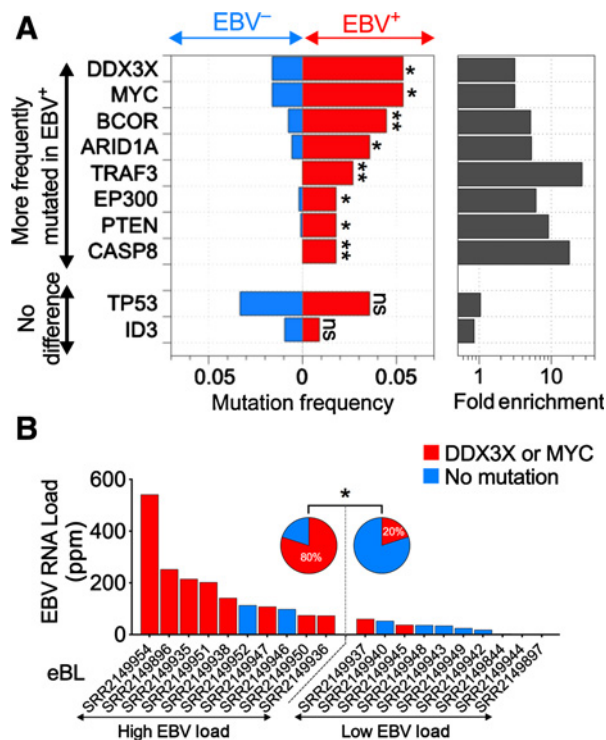


Figure 5. EBV is associated with pathogenic mutations in cancer driver genes. **A**, Mutation frequency (left) and fold enrichment (right) of cancer driver genes in EBV⁺ (red) versus EBV⁻ (blue) cancers. **B**, EBV RNA load across all endemic Burkitt lymphoma cancer samples. Samples with *DDX3X* or *MYC* mutations are highlighted in red. Pie chart shows frequency of *DDX3X* or *MYC* mutations in samples with high and low EBV RNA loads. *, $P < 0.05$; **, $P < 0.01$; Fisher exact test.

To identify upstream regulators [transcription factors (TF) and cytokines] of the EBV response genes, we next performed Ingenuity Pathway Analysis. We found 10 regulators predicted to be significantly activated or inhibited in at least two cancer types (Fig. 6B; $|Z\text{-score}| > 4$; $P < 1e-5$). These were mostly type I and type II IFN pathway proteins (e.g., IFN γ , IFN α , IFN β), TFs that drive or transduce IFN signatures (e.g., STAT1 and IRF7) or that overlap with type I IFN responses (e.g., NF κ B; Fig. 6B). These upstream regulators classified EBV⁺ cancer types into two groups: IFN⁺, consisting of stomach adenocarcinoma, NPC, and DLBCL and characterized by an activated IFN signature and IFN⁻, containing endemic Burkitt lymphoma, AITL, NKTCL, and sporadic Burkitt lymphoma and denoted by inhibited IFN response when comparing EBV⁺ to EBV⁻ tumors of the same origin. Literature-validated IFN γ targets overlapped significantly with EBV response genes in these two clusters and could explain up to 20% of response genes (Supplementary Fig. S6K). To further quantify this observation, we calculated type I and type II IFN activity scores for each individual sample (Supplementary Table S51). EBV⁺ cancers in the IFN-activated cluster overexpressed several immune checkpoint proteins, including the universal checkpoints Programmed Death-Ligand 1 (PD-L1) and indoleamine 2,3-dioxygenase (IDO)-1 (Fig. 6C; ref. 54), which are both IFN γ -induced and mediate immunosuppression and tumor evasion. For example, EBV⁺

stomach adenocarcinoma samples displayed strong activation of the IFN γ pathway ($Z\text{-score} > 8$; Fig. 6B), with an average of 3.9- to 5.6-fold higher mRNA levels of PD-L1 and IDO1 (Fig. 6D). Both these markers had strong predictive power to distinguish EBV⁺ from EBV⁻ stomach adenocarcinoma ($AUC > 0.8$; $P < 0.0001$; Fig. 6E).

To better understand and verify the effect of IFN response pathways and EBV reactivation on PD-L1 and IDO1 expression in EBV⁺ cancers, we treated EBV⁺ (SNU719) and EBV⁻ (SNU1) stomach adenocarcinoma cancer cell lines with exogenous IFN γ . This cytokine induced IDO1 only in the EBV⁺ cells and further induced PD-L1 in the EBV⁺, compared with the EBV⁻ cells (Fig. 6F; see Supplementary Fig. S6L for gating strategy) with no significant effect on EBV replication (measured by expression of the EBV lytic gene *BMRFL1*; Supplementary Fig. S6M), confirming IFN pathways as preferential drivers of these molecules in EBV⁺ cells. PD-L1 was induced by IFN γ in both EBV⁺ and EBV⁻ cells, as reported previously (55). To directly assess the effect of EBV reactivation on PD-L1, we induced EBV lytic cycle by treating cells with 12-O-tetradecanoylphorbol-13-acetate and sodium butyrate (TPA/NaB) and measured PD-L1 expression. EBV reactivation induced PD-L1 only in EBV⁺ cells (Fig. 6G). We next examined two independent transcriptome datasets of DLBCL and NPC cancers (GSE12452 and GSE27255; refs. 56, 57). In both cases, PD-L1 was preferentially overexpressed in EBV⁺, compared with EBV⁻, tumors (Supplementary Fig. S6N and S6O), supporting the role of EBV in inducing immune checkpoint protein expression in EBV⁺ cancers. In summary, these data indicate that the transcriptomes of EBV⁺ tumors identify a shared transcriptional program regulated by IFNs leading to a significant change in expression of immune checkpoint genes.

Discussion

We carried out a large-scale pan-cancer RNA-seq analysis to interrogate various aspects of host-virus interaction in EBV-associated cancers. Collectively, our work provides a comprehensive interactome map of EBV-associated cancers that enables a broad interrogation of host-pathogen interactions, including expression and mutation of the pathogen genome as well as integration into the host genome, host gene mutations, and transcriptional responses. There are other host-pathogen mechanisms that operate in different cancer types that have not been discussed here. However, detailed viral and host expression and mutation profiles for each cancer and cancer type have been provided in the Supplementary Materials as a resource for future discovery and validation. Our approach can be equally applied to the evaluation of other symbiotic, disease-associated pathogens, including cancer-causing viruses (5, 7, 17, 33).

Using our interactome map, we made several novel observations. By evaluating viral gene expression within the host tumor environment, we uncovered pan-cancer and cancer-type-specific expression of several virus genes. The pan-cancer expression of a few EBV genes makes them prime candidates for designing early-onset diagnostics. Our samples were obtained mainly from tumor sites (e.g., lymph node), thus the potential for EBV genes as biomarkers should first be evaluated by whether they can also be detected in blood or urine of patients using sensitive techniques. As EBV itself can be detected in whole blood and urine (58),

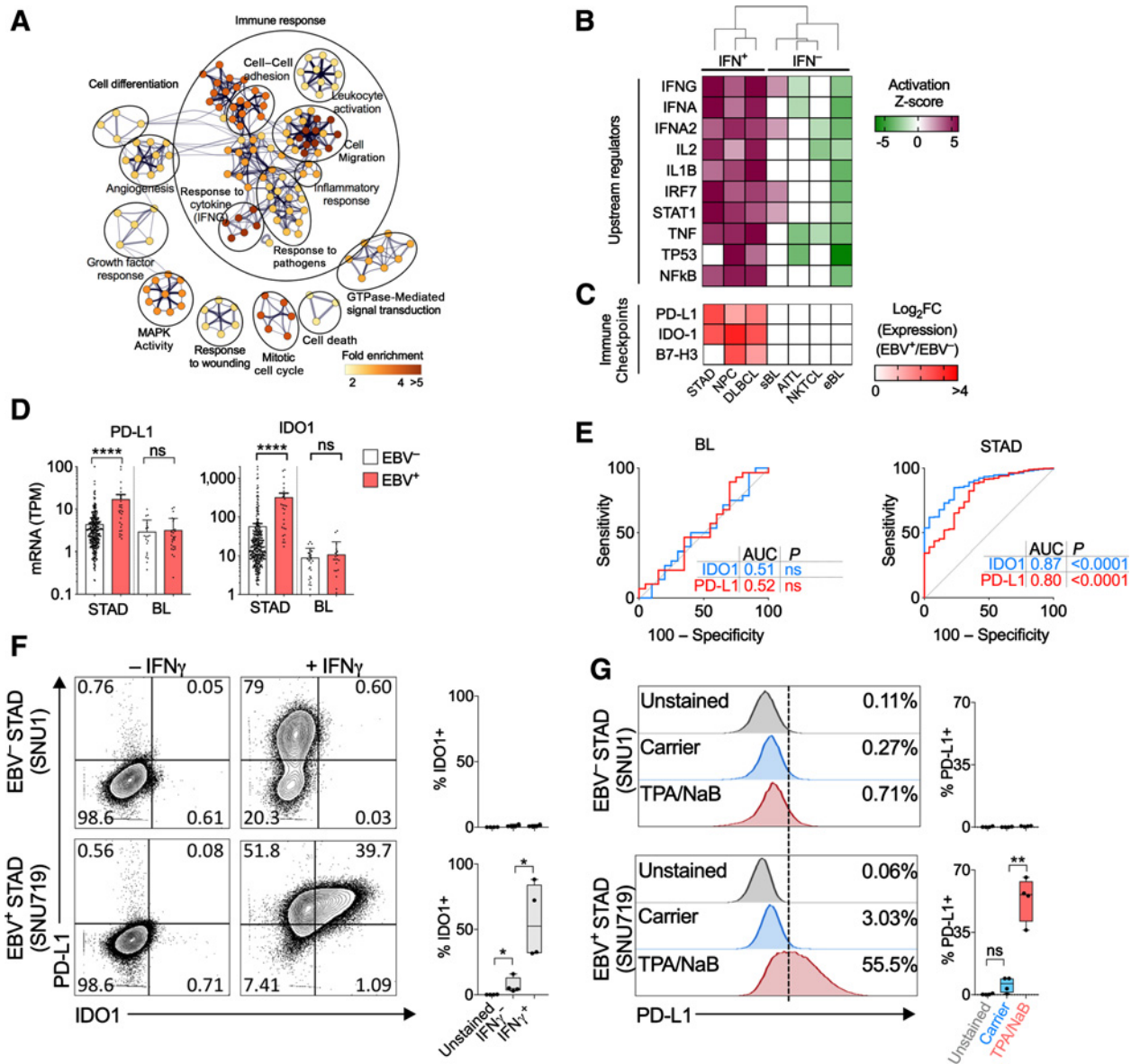


Figure 6. EBV drives shared and cancer-specific gene networks in host cells to dichotomously classify tumors. **A**, Biological pathways significantly enriched in shared EBV response genes (see also Supplementary Fig. S6A). Nodes indicate individual pathways, grouped by ellipses into functional category. Nodes are colored by fold enrichment and band thickness indicates number of shared response genes between nodes. **B** and **C**, Ingenuity Pathway Analysis showing top 10 upstream regulators (TFs and cytokines) of EBV-driven response genes (**B**) and heatmap showing expression of immune checkpoint genes in EBV⁺ compared with EBV⁻ tissues (**C**) in each cancer type. Columns are hierarchically clustered. Indicated are cancer types subclassified into IFN⁺ and IFN⁻ groups. **D**, *PD-L1* (left) and *IDO1* (right) mRNA in EBV⁺ and EBV⁻ stomach adenocarcinomas (STAD) and Burkitt lymphoma (BL, including both sporadic and endemic) samples. **E**, Receiver operator curves show performance of *PD-L1* and *IDO1* mRNA expression to predict EBV⁺ versus EBV⁻ status of samples in Burkitt lymphoma (left) and stomach adenocarcinomas (right). AUC and *P* values are indicated; **F**, Representative flow cytometry plots (left) and cumulative data (right) from *n* = 4 independent experiments showing *PD-L1* and *IDO1* expression by EBV⁺ (SNU719) and EBV⁻ (SNU1) stomach adenocarcinomas cancer cell lines with or without IFN γ treatment. **G**, Representative flow cytometry plots (left) and cumulative data (right) from *n* = 4 independent experiments showing *PD-L1* expression in EBV⁺ and EBV⁻ stomach adenocarcinomas cancer cell lines with or without TPA/NaB-induced EBV reactivation (see also Supplementary Fig. S6M). Bars, mean \pm SEM; *P* values in **D**, **F**, and **G** are two-tailed Wilcoxon test. *, *P* < 0.05; **, *P* < 0.01; ****, *P* < 0.0001; ns, nonsignificant.

we predict that these markers should also be detectable in these samples by PCR (59).

We identified a reliable set of 56 loci that have EBV integration signatures in more than one EBV⁺ cell type. The rate of integration

in each sample could be rare because of two reasons: one, rather than most tumor cells, only a small fraction of them could have the integration, and two, only one or small number of several EBV copies present in each cell may be integrated. Nonetheless, the

high-throughput sequencing method utilized here is extremely sensitive and could detect such rare events and as such the integration signature was evident in almost half of the tumor samples.

This meta-analysis identified frequent missense and non-sense variations in virostatic genes in cancer. These variations were associated with elevated viral RNA load in cancer samples harboring the variation as well as the host transcriptional response to changes in these genes. As high EBV RNA load increases the risk of cancer, inactivating mutation(s) in these genes could heighten the risk of cancer development. There are three issues when considering variations in viral genes. The first issue is that there are generally several copies of EBV genomes per cell and thus variations may be present in a subset of EBV genomes. To address this issue, our variants are selected to be either homozygous (meaning that all the expressed EBV genomes in the cell contain the variant), and in several cases heterozygous (meaning that at least half of all the EBV genomes in the cell contain the variant). The second issue is that since EBV associated cancers occur in different parts of the world, there are known geographic inter-strain variations (60) that could complicate the interpretation of any identified variation. To be able to parse which variations are less likely to be due to strain differences, we have looked at the distribution of all our *de novo* cataloged variations and determined the proportion that are present in a single individual cancer type and the proportion present in more than one cancer type. Those variations present only in a single cancer type could represent strain differences while those shared among different, geographically distinct, cancer types are more likely to be related to biology. Nearly half of our identified variants are shared between multiple cancer types, suggesting potential relevance to cancer rather than inter-strain differences. This is further supported by elevated viral RNA load in tumor samples with variations and host transcriptional responses to these mutations. The third issue is that our analysis relied on RNA-sequencing, primarily due to the wider availability of such data. However, mutations in nonexpressed/low-expressed EBV genes cannot be detected using this approach—the same limitation also exists for the EBV integration analysis. Future whole-genome or targeted sequencing approaches are required to overcome this limitation. Nevertheless, the identification of viral genes that have virostatic function and that are frequently mutated in cancer also raises the possibility that these predicates can be leveraged as diagnostic/prognostic disease markers and/or form the basis of therapeutic drug development, for example, using mimetic agents. We have provided some evidence on the function of virostatic genes and their mutations, but subsequent experimental verifications of these observations directly in primary cancer cells remain to be done.

A shared transcriptional program regulated by IFNs was also evident, leading to a significant change in expression of immune checkpoint genes. This enabled us to propose a dichotomous molecular classifier for EBV⁺ cancers—IFN⁺ or IFN⁻ group. On the basis of this classification, the treatments and mechanisms identified in one cancer type could potentially be applied to the linked group. The upregulation of immune checkpoints in individual cancer types of IFN⁺ group have previously been shown in separate studies (61, 62). There has also been a report on a selective loss of EBNA1-specific T-cell responses in children with

endemic Burkitt lymphoma (63), which could be explained by repression of IFN signatures identified in this IFN⁻ cancer type. Using our unified classification, one could anticipate distinct clinical treatment options. For example, inhibitors of IFN signaling (antibody therapy, or JAK inhibitors) or checkpoint blockade (e.g., PD-L1 inhibitors; ref. 64) would be more efficacious in EBV⁺ cancers in IFN⁺ group, whereas immunostimulants and adjuvants such as IFNs might improve survival in IFN⁻ group. Thus, this distinction could help individualize treatment options and minimize consequent toxicities of treatments such as checkpoint inhibitors, which have a strong association with development of *de novo* autoimmunity.

In conclusion, we have presented here findings from a pan-cancer interactome atlas of host cells with EBV, highlighting the following: a catalog of reliable EBV integration sites genome-wide, an association between regulatory regions with high activity and EBV integration, a list of virostatic EBV genes, and their susceptibility to nonsense and missense variation in cancer. Furthermore, we identified shared and unique host cell transcriptional responses to virus, allowing dichotomous classification of EBV⁺ cancer types. These responses were regulated by IFN pathway signaling, resulting in differential immune checkpoint expression.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: S. Chakravorty, G. Chopra, M.R. Olson, M. Kazemian
Development of methodology: S. Chakravorty, L. Wang, J.T. Quaid, C.F. Lin, S.D. Briggs, D.A. Canaria, B. Zhao, M. Kazemian
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Chakravorty, C. Wang, J.T. Quaid, D.A. Canaria, M.R. Olson, B. Zhao, M. Kazemian
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S. Chakravorty, B. Yan, L. Wang, J.T. Quaid, C.F. Lin, J. Majumder, D. Chauss, B. Afzali, M. Kazemian
Writing, review, and/or revision of the manuscript: S. Chakravorty, B. Yan, J. Majumder, D. Chauss, G. Chopra, M.R. Olson, B. Afzali, M. Kazemian
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. Chakravorty, B. Afzali, M. Kazemian
Study supervision: S.D. Briggs, G. Chopra, M. Kazemian

Acknowledgments

We thank Drs. Fred P. Davis (NIH, Bethesda, MD), Claudia Kemper (NIH, Bethesda, MD), Ourania Andrisani (Purdue University, West Lafayette, IN), and Jeffrey Cohen (NIH, Bethesda, MD) for invaluable comments on the manuscript. This work was supported by Purdue Cancer Center (startup fund to M. Kazemian), National Heart, Lung, and Blood Institute (NIH grant 5K22HL125593 to M. Kazemian), Showalter Trust (research award to M. Kazemian and G. Chopra), and the Wellcome Trust (grant 097261/Z/11/Z to B. Afzali). This work was supported (in part) by the Intramural Research Program of the NIH, The National Institutes of Diabetes and Digestive and Kidney Diseases (NIDDK). The authors thank patients who donated tissue samples and investigators that carried out the RNA-sequencing.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 20, 2019; revised June 24, 2019; accepted August 27, 2019; published first September 3, 2019.

References

- IARC Working Group. IARC monographs on the evaluation of carcinogenic risks to humans. Volume 100B: Biological agents. Lyon, France: IARC: Distributed for the International Agency for Research on Cancer by the Secretariat of the World Health Organization; 2012. Available from: <https://publications.iarc.fr/119>.
- Kutok JL, Wang F. Spectrum of Epstein-Barr virus-associated diseases. *Annu Rev Pathol* 2006;1:375–404.
- Khan G, Hashim MJ. Global burden of deaths from Epstein-Barr virus attributable malignancies 1990–2010. *Infect Agent Cancer* 2014;9:38.
- Peng RJ, Han BW, Cai QQ, Zuo XY, Xia T, Chen JR, et al. Genomic and transcriptomic landscapes of Epstein-Barr virus in extranodal natural killer T-cell lymphoma. *Leukemia* 2018;33:1451–62.
- Abate F, Ambrosio MR, Mundo L, Laginestra MA, Fuligni F, Rossi M, et al. Distinct viral and mutational spectrum of endemic Burkitt lymphoma. *PLoS Pathog* 2015;11:e1005158.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513:202–9.
- Arvey A, Tempera I, Tsai K, Chen HS, Tikhmyanova N, Klichinsky M, et al. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell Host Microbe* 2012;12:233–45.
- Okuno Y, Murata T, Sato Y, Muramatsu H, Ito Y, Watanabe T, et al. Defective Epstein-Barr virus in chronic active infection and haematological malignancy. *Nat Microbiol* 2019;4:404–13.
- Price AM, Dai J, Bazot Q, Patel L, Nikitin PA, Djavadian R, et al. Epstein-Barr virus ensures B cell survival by uniquely modulating apoptosis at early and late times after infection. *Elife* 2017;6. doi: 10.7554/eLife.22509.
- Verma D, Thompson J, Swaminathan S. Spironolactone blocks Epstein-Barr virus production by inhibiting EBV SM protein function. *Proc Natl Acad Sci U S A* 2016;113:3609–14.
- Arvey A, Tempera I, Lieberman PM. Interpreting the Epstein-Barr Virus (EBV) epigenome using high-throughput data. *Viruses* 2013;5:1042–54.
- Ohshima K, Suzumiya J, Kanda M, Kato A, Kikuchi M. Integrated and episomal forms of Epstein-Barr virus (EBV) in EBV associated disease. *Cancer Lett* 1998;122:43–50.
- Hurley EA, Agger S, McNeil JA, Lawrence JB, Calendar A, Lenoir G, et al. When Epstein-Barr virus persistently infects B-cell lines, it frequently integrates. *J Virol* 1991;65:1245–54.
- Hsu CH, Hergenbahn M, Chuang SE, Yeh PY, Wu TC, Gao M, et al. Induction of Epstein-Barr virus (EBV) reactivation in Raji cells by doxorubicin and cisplatin. *Anticancer Res* 2002;22:4065–71.
- Sivachandran N, Wang X, Frappier L. Functions of the Epstein-Barr virus EBNA1 protein in viral reactivation and lytic infection. *J Virol* 2012;86:6146–58.
- Tang KW, Alaei-Mahabadi B, Samuelsson T, Lindh M, Larsson E. The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat Commun* 2013;4:2513.
- Khoury JD, Tannir NM, Williams MD, Chen Y, Yao H, Zhang J, et al. Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. *J Virol* 2013;87:8916–26.
- Yoo HY, Sung MK, Lee SH, Kim S, Lee H, Park S, et al. A recurrent inactivating mutation in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nat Genet* 2014;46:371–5.
- Kucuk C, Jiang B, Hu X, Zhang W, Chan JK, Xiao W, et al. Activating mutations of STAT5B and STAT3 in lymphomas derived from gamma-delta-T or NK cells. *Nat Commun* 2015;6:6025.
- Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M, et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* 2012;490:116–20.
- Choi J, Goh G, Walradt T, Hong BS, Bunick CG, Chen K, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet* 2015;47:1011–9.
- Saba NS, Liu D, Herman SE, Underbayev C, Tian X, Behrend D, et al. Pathogenic role of B-cell receptor signaling and canonical NF-kappaB activation in mantle cell lymphoma. *Blood* 2016;128:82–92.
- Lin DC, Meng X, Hazawa M, Nagata Y, Varela AM, Xu L, et al. The genomic landscape of nasopharyngeal carcinoma. *Nat Genet* 2014;46:866–71.
- Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 2014;13:397–406.
- Kobayashi K, Lin LW, Yeadon JE, Klickstein LB, Smith JA. Cloning and sequence analysis of a rat liver cDNA encoding acyl-peptide hydrolase. *J Biol Chem* 1989;264:8892–9.
- Thompson MP, Kurzrock R. Epstein-Barr virus and cancer. *Clin Cancer Res* 2004;10:803–21.
- Xiao K, Yu Z, Li X, Tang K, Tu C, et al. Genome-wide Analysis of Epstein-Barr Virus (EBV) Integration and Strain in C666-1 and Raji Cells. *J Cancer* 2016;7:214–24.
- Li JW, Wan R, Yu CS, Co NN, Wong N, Chan TF. ViralFusionSeq: accurately discover viral integration events and reconstruct fusion transcripts at single-base resolution. *Bioinformatics* 2013;29:649–51.
- Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 2018;36:338–45.
- Walsh D, Mohr I. Viral subversion of the host protein synthesis machinery. *Nat Rev Microbiol* 2011;9:860–75.
- Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell* 2013;155:934–47.
- Dooley KE, Warburton A, McBride AA. Tandemly integrated HPV16 can form a Brd4-dependent super-enhancer-like element that drives transcription of viral oncogenes. *MBio* 2016;7. doi: 10.1128/mBio.01446-16.
- Zhao LH, Liu X, Yan HX, Li WY, Zeng X, Yang Y, et al. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun* 2016;7:12992.
- Yuan J, Jiang YY, Mayakonda A, Huang M, Ding LW, Lin H, et al. Super-enhancers promote transcriptional dysregulation in nasopharyngeal carcinoma. *Cancer Res* 2017;77:6614–26.
- Sample C, Parker B. Biochemical characterization of Epstein-Barr virus nuclear antigen 3A and 3C proteins. *Virology* 1994;205:534–9.
- Miller CL, Lee JH, Kieff E, Longnecker R. An integral membrane protein (LMP2) blocks reactivation of Epstein-Barr virus from latency following surface immunoglobulin crosslinking. *Proc Natl Acad Sci U S A* 1994;91:772–6.
- Adler B, Schaadt E, Kempkes B, Zimmer-Strobl U, Baier B, Bornkamm GW. Control of Epstein-Barr virus reactivation by activated CD40 and viral latent membrane protein 1. *Proc Natl Acad Sci U S A* 2002;99:437–42.
- Prince S, Keating S, Fielding C, Brennan P, Floettmann E, Rowe M. Latent membrane protein 1 inhibits Epstein-Barr virus lytic cycle induction and progress via different mechanisms. *J Virol* 2003;77:5000–7.
- Rechsteiner MP, Berger C, Weber M, Sigrist JA, Nadal D, Bernasconi M. Silencing of latent membrane protein 2B reduces susceptibility to activation of lytic Epstein-Barr virus in Burkitt's lymphoma Akata cells. *J Gen Virol* 2007;88:1454–9.
- Sample J, Young L, Martin B, Chatman T, Kieff E, Rickinson A, et al. Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. *J Virol* 1990;64:4084–92.
- Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 2013;45:970–6.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–21.
- Vockerodt M, Wei W, Nagy E, Prouzova Z, Schrader A, Kube D, et al. Suppression of the LMP2A target gene, EGR-1, protects Hodgkin's lymphoma cells from entry to the EBV lytic cycle. *J Pathol* 2013;230:399–409.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- Grande BM, Gerhard DS, Jiang A, Griner NB, Abramson JS, Alexander TB, et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood* 2019;133:1313–24.

46. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet* 2013;93:641–51.
47. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 2017;45:D777–D83.
48. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
49. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
50. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang C, Stein DA, et al. Meta- and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. *Cell Host Microbe* 2015;18:723–35.
51. Zhang Y, Ohyashiki JH, Shimizu N, Ohyashiki K. Aberrant expression of NK cell receptors in Epstein-Barr virus-positive gammadelta T-cell lymphoproliferative disorders. *Hematology* 2010;15:43–7.
52. Zhou H, Schmidt SC, Jiang S, Willox B, Bernhardt K, Liang J, et al. Epstein-Barr virus oncoprotein super-enhancers control B cell growth. *Cell Host Microbe* 2015;17:205–16.
53. ClinicalTrials.gov. GZ17-6.02, phase I study evaluating GZ17-6.02 in patients with advanced solid tumors or lymphoma. Available from: <https://ClinicalTrials.gov/show/NCT03775525>.
54. Soliman H, Mediavilla-Varela M, Antonia S. Indoleamine 2,3-dioxygenase: is it an immune suppressor? *Cancer J* 2010;16:354–9.
55. Moon JW, Kong SK, Kim BS, Kim HJ, Lim H, Noh K, et al. IFN-gamma induces PD-L1 overexpression by JAK2/STAT1/IRF-1 signaling in EBV-positive gastric carcinoma. *Sci Rep* 2017;7:17810.
56. Petrich AM, Leshchenko V, Kuo PY, Xia B, Thirukonda VK, Ulahannan N, et al. Akt inhibitors MK-2206 and nelfinavir overcome mTOR inhibitor resistance in diffuse large B-cell lymphoma. *Clin Cancer Res* 2012;18:2534–44.
57. Sengupta S, den Boon JA, Chen IH, Newton MA, Dahl DB, Chen M, et al. Genome-wide expression profiling reveals EBV-associated inhibition of MHC class I expression in nasopharyngeal carcinoma. *Cancer Res* 2006;66:7999–8006.
58. Hohaus S, Santangelo R, Giachelia M, Vannata B, Massini G, Cuccaro A, et al. The viral load of Epstein-Barr virus (EBV) DNA in peripheral blood predicts for biological and clinical characteristics in Hodgkin lymphoma. *Clin Cancer Res* 2011;17:2885–92.
59. Lo YM, Chan WY, Ng EK, Chan LY, Lai PB, Tam JS, et al. Circulating Epstein-Barr virus DNA in the serum of patients with gastric carcinoma. *Clin Cancer Res* 2001;7:1856–9.
60. Correia S, Bridges R, Wegner F, Venturini C, Palser A, Middeldorp JM, et al. Sequence variation of Epstein-Barr virus: viral types, geography, codon usage, and diseases. *J Virol* 2018;92. doi: 10.1128/JVI.01132-18.
61. Fang W, Zhang J, Hong S, Zhan J, Chen N, Qin T, et al. EBV-driven LMP1 and IFN-gamma up-regulate PD-L1 in nasopharyngeal carcinoma: Implications for oncotargeted therapy. *Oncotarget* 2014;5:12189–202.
62. Chen BJ, Chapuy B, Ouyang J, Sun HH, Roemer MG, Xu ML, et al. PD-L1 expression is characteristic of a subset of aggressive B-cell lymphomas and virus-associated malignancies. *Clin Cancer Res* 2013;19:3462–73.
63. Moormann AM, Heller KN, Chelimo K, Embury P, Ploutz-Snyder R, Otieno JA, et al. Children with endemic Burkitt lymphoma are deficient in EBNA1-specific IFN-gamma T cell responses. *Int J Cancer* 2009;124:1721–6.
64. Kelly RJ. Immunotherapy for esophageal and gastric cancer. *Am Soc Clin Oncol Educ Book* 2017;37:292–300.
65. Donato E, Croci O, Sabo A, Muller H, Morelli MJ, Pelizzola M, et al. Compensatory RNA polymerase 2 loading determines the efficacy and transcriptional selectivity of JQ1 in Myc-driven tumors. *Leukemia* 2017;31:479–90.