Materials and Methods

Tumor-free Survival
In Eμ-LMP1 and Eμ-LMP2A (Tg6 line) transgenic mice, LMP1 and LMP2A are expressed under immunoglobulin (Ig) heavy-chain promoter and intronic enhancer (Eμ), respectively. In λ-MYC mice, human MYC is overexpressed under Ig λ-chain enhancer. All mice are in the C57BL/6 background. Mice were assigned for the observation of tumor incidence before they develop tumors. When tumors are observed in lymph nodes including cervical, axillary, inguinal and abdominal areas, the time after birth during no sign of tumor was determined as tumor-free survival.

Immunoblots
Pretumor spleens were isolated from 4- to 6-week-old mice. Magnetic activated cell sorting using Mouse Pan-B Cell Isolation Kit (StemCell Technologies) were used to purify splenic B cells. Tumor-bearing lymph nodes were dissociated and treated with 155 mM ammonium chloride to lyse red blood cells. Approximately 90% of tumor cells in tumor-bearing lymph nodes were of B-cell lineage and thus did not required further cell sorting. Isolated pretumor B cells or tumor cells were lysed in lysis buffer with protease and phosphatase inhibitor cocktails (Roche). Lysates were electrophoretically separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (Bio-Rad). Protein was transferred to a nitrocellulose membrane (Bio-Rad) and incubated with primary antibodies. Membranes were probed with IRDye secondary antibodies (Li-Cor Biosciences). The protein bands were visualized with an Odyssey Fc Western Blotting Imager and analyzed with Image Studio version 2.0 (Li-Cor Biosciences). Primary antibodies used in immunoblots included JNK, phospho-JNK, ERK1/2, phospho-ERK1/2, p38, phospho-p38 (Cell Signaling) and Gapdh (Abcam). Precision Plus Protein Standards (Bio-Rad) were used as molecular size markers.

Differential Gene Expression Analysis
Raw RNA sequencing files from 66 primary Burkitt lymphoma tumors (28 sporadic Burkitt lymphoma and 38 endemic Burkitt lymphoma) from three different studies were downloaded using the Sequencing Read Archive (SRA) Toolkit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software)1-3. Reads were trimmed of adapter sequences and low quality reads were filtered using fastp4. Processed reads were aligned to both human and EBV reference genomes using STAR5. Samples were classified as either EBV positive or negative based on the presence or absence of EBV reads. EBV positive samples were further classified as LMP-1 positive or negative and LMP-2A positive or negative if they had reads aligning to those genes. Differential gene expression analysis was performed using DESeq2 with the study origin included as a condition to minimize batch effects6. Differentially expressed genes were visualized using EnhancedVolcano7. Gene set enrichment analysis was performed using ClusterProfiler with input data from DESeq2 differential gene expression analysis8.

Mutation Analysis
Processed reads were aligned to the human genome using STAR in 2-pass mode5. BAM files were processed using the Genome Analysis Toolkit (GATK) according to the best practices workflow for RNA sequencing somatic mutation calling9. Duplicate reads were marked using the
MarkDuplicates function and reads spanning splicing junctions were split using the SplitNcigarReads function. Bases were recalibrated using the BaseRecalibrationSpark and ApplyBQSRSpark functions. Variants were called using Mutect2 with default settings with known germline variants defined by data from gnomAD\textsuperscript{10}. Variants were filtered using the FilterMutectCalls function and annotated for significance using CRAVAT 4\textsuperscript{11}. Mutations in 11 genes known for being commonly mutated in Burkitt lymphoma were counted if they occurred in coding regions and were predicted to be deleterious (VEST score > 0.5) or were known deleterious mutations with corresponding COSMIC IDs\textsuperscript{12}. Mutation heatmaps were constructed using ComplexHeatmap\textsuperscript{13}.

1. Abate F, Ambrosio MR, Mundo L, et al. Distinct Viral and Mutational Spectrum of Endemic Burkitt Lymphoma. \textit{PLoS Pathog.} 2015;11(10):e1005158.
2. Lombardo KA, Coffey DG, Morales AJ, et al. High-throughput sequencing of the B-cell receptor in African Burkitt lymphoma reveals clues to pathogenesis. \textit{Blood Adv.} 2017;1(9):535–544.
3. Schmitz R, Young RM, Ceribelli M, et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. \textit{Nature}. 2012;490(7418):116–120.
4. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. \textit{Bioinformatics}. 2018;34(17):i884–i890.
5. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. \textit{Bioinformatics}. 2013;29(1):15–21.
6. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. \textit{Genome Biology}. 2014;15(12):550.
7. Blighhe K. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. 2022.
8. Yu G, Wang L-G, Han Y, He Q-Y. \textit{clusterProfiler}: an R Package for Comparing Biological Themes Among Gene Clusters. \textit{OMICS: A Journal of Integrative Biology}. 2012;16(5):284–287.
9. Auwera GAV der, O’Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. Sebastopol, CA: O’Reilly Media; 2020.
10. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. \textit{Nature}. 2020;581(7809):434–443.
11. Masica DL, Douville C, Tokheim C, et al. CRAVAT 4: Cancer-Related Analysis of Variants Toolkit. \textit{Cancer Res.} 2017;77(21):e35–e38.
12. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. \textit{Nucleic Acids Research}. 2019;47(D1):D941–D947.
13. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. \textit{Bioinformatics}. 2016;32(18):2847–2849.
Figure Legends

Supplemental Figure 1
Statistics of tumor-free survival in murine BL model.
(A) Mice with the genotypes of LMP1/λ-MYC, LMP2A/λ-MYC, LMP1/LMP1/λ-MYC and λ-MYC were analyzed for tumor-free survival and observed for tumor incidences. (B) Primary sites where the lymphomas were first observed were indicated as percentages in each mouse genotypes. (C) Mice were classified with primary tumor locations and compared with tumor-free survival by log-rank (Mantel-Cox) test. *P < .05, **P < .01, ***P < .001, and ****P < .0001.

Supplemental Figure 2
Immunoblots of p53 and p19ARF expression in tumor cells.
Lysates of tumor cells prepared from indicated mice were immunoblotted with p53 (Cell Signaling) and p19ARF (GeneTex) antibodies. Arrow heads indicate appropriate p53 or p19 proteins. A lysate from λ-MYC mice (#5037) was used in all blots for the control of p53++ and p19+ expression.

Supplemental Figure 3
Immunoblots of MAPK-pathway proteins in pretumor B cells and tumor cells.
Lysates of isolated pretumor B cells and tumor cells prepared from indicated mice were immunoblotted with primary antibodies against JNK, phospho-JNK, ERK1/2, phospho-ERK1/2, p38, phospho-p38 (Cell Signaling) and Gapdh (Abcam). Closed arrow heads indicate appropriate MAPK-pathway proteins while open arrow heads indicate non-specific bands.

Supplemental Figure 4
Differential mutations in BL-associated genes
Primary BL tumors (28 sBL and 38 eBL) from three different studies1–3 were reanalyzed for somatic gene mutations using RNA-Seq analysis. They were classified as (A) EBV-positive or negative, (B) LMP1-positive or negative, and (C) LMP2A-positive or negative BLs based on their expression of log2 transcripts per million (TPM) as indicated in bar plots (red to blue gradient) on the top of panels (yellow or green). Mutation incidence in genes known for being commonly mutated in BL is indicated in heatmaps (red: single or multiple mutations, blue: no mutation). Prevalence of mutated tumors are indicated in right-hand stacked bar plots as additive percentages of EBV-positive and negative, LMP1-high and low, or LMP2A-high and low groups, respectively. P-values of prevalent differences between classified groups were calculated with Fisher’s exact test (** P<0.01 and * P<0.05).

Supplemental Figure 5
Differential gene expression in human BL
Significantly different gene expression of (A) EBV-positive vs negative, (B) LMP1-positive vs negative, and (C) LMP2A-positive vs negative primary BL tumors were shown in volcano plot using DESeq2. Global gene downregulation was observed in LMP1-High BLs when compared to LMP1-Low BLs. Two-fold change of magnitude in gene expression (as shown in log2 fold change in X axis) and P<0.05 of statistical significance (as shown in log10 adjusted P-value in Y axis) were used as cutoffs to display significant (red) and non-significant (green) differences in gene expression.
Supplemental Figure 6
Gene Set Enrichment Analysis (GSEA) in human BL

Significantly activated or suppressed groups of genes were calculated by comparing (A) LMP1-positive vs negative and (B) LMP2A-positive vs negative primary BL tumors using ClusterProfiler. Top groups with highest adjusted P-values are displayed in dot plots. Multiple replication-related gene groups were activated in LMP1-High BL, while multiple gene groups in varied pathways were suppressed in LMP1-High BL. Count indicates the number of activated/suppressed genes in each group. Gene ratio (horizontal axis) indicates the ratio of activated/suppressed genes to total genes in each group.
Supplemental Figure 3
Ikeda et al.

A. Pretumor

- p-JNK
- JNK
- p-ERK
- ERK
- p-p38
- p38
- Gapdh

B. Tumor

- p-JNK
- JNK
- p-ERK
- ERK
- p-p38
- p38
- Gapdh
Supplemental Figure 4
Ikeda et al

A

EBV Positive | EBV Negative

MYC | LMP1 TPM
TP53 | Mutated Tumors (%*%)
ID3 |
DDX3X |
RHOA |
ARID1A |
MSH6 |
CCND3 |
SMARCA4 |
PIK3R1 |
TCF3 |

B

LMP1 expression

MYC | LMP1 High
TP53 |
ID3 |
DDX3X |
RHOA |
ARID1A |
MSH6 |
CCND3 |
SMARCA4 |
PIK3R1 |
TCF3 |

C

LMP2A expression

MYC | LMP2A High
TP53 |
ID3 |
DDX3X |
RHOA |
ARID1A |
MSH6 |
CCND3 |
SMARCA4 |
PIK3R1 |
TCF3 |

Mutated tumors
Non-mutated tumors
