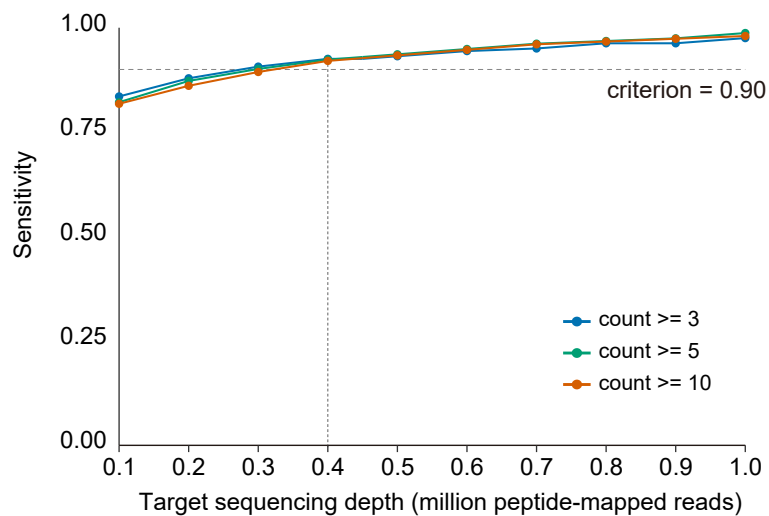


Epitope-resolved Epstein-Barr virus antibody mapping identifies disease-stratifying signatures in nasopharyngeal carcinoma

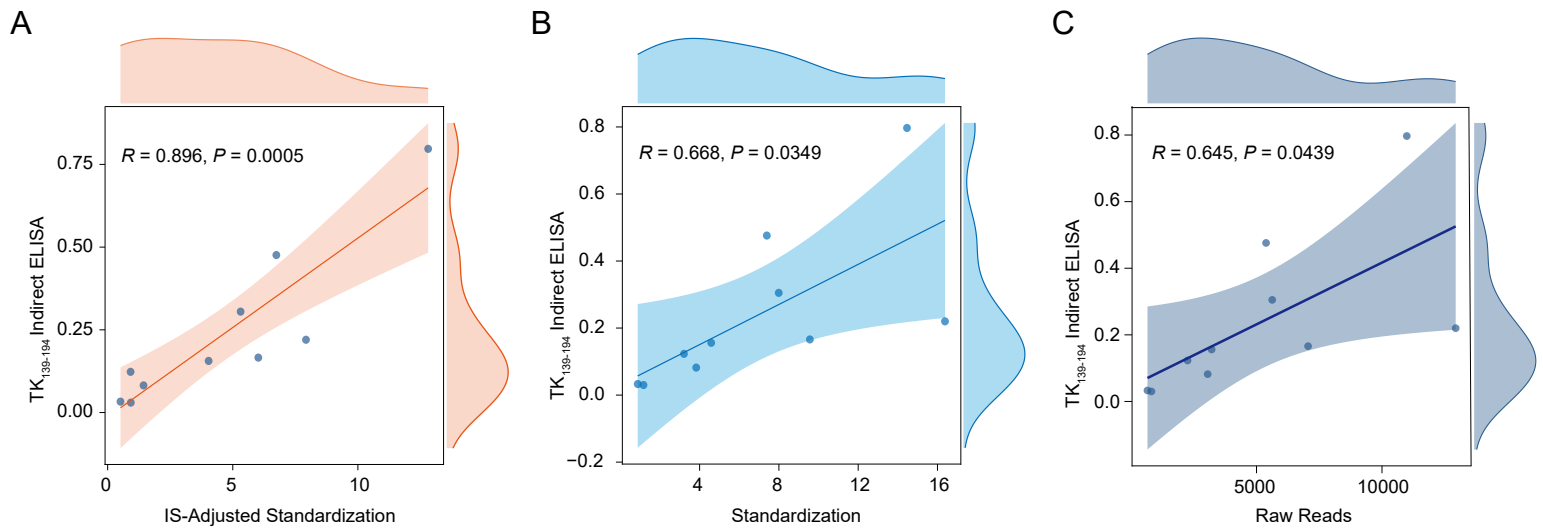
Supplementary Figure 1



Supplementary Figure 1. Evaluation of the minimum sequencing depth by downsampling.

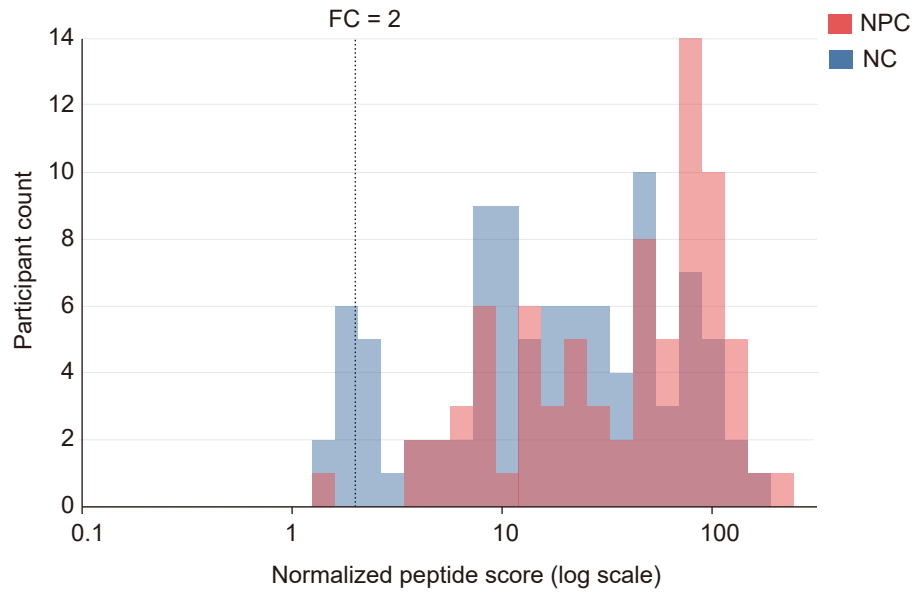
A sample was subjected to 100 nested, without-replacement downsampling replicates at each sequencing depth. Positive peptides were defined as FC > 2 with peptide counts ≥ 3 , ≥ 5 , or ≥ 10 . Sensitivity was calculated relative to the positive-peptide set detected at full depth, and points represent median sensitivity across replicates. The dashed line indicates the predefined sensitivity criterion of 0.90. A conservative minimum depth of 4.0×10^5 peptide-mapped reads was adopted.

Supplementary Figure 2



Supplementary Figure 2. Orthogonal ELISA validation of the relative semi-quantitative performance of qPhIP-Seq for TK₁₃₉₋₁₉₄. Correlation of indirect ELISA reactivity against TK₁₃₉₋₁₉₄ with (A) internal-reference-adjusted qPhIP-Seq values, (B) conventionally standardized qPhIP-Seq values, and (C) raw sequencing reads in NPC serum samples. Pearson correlation coefficients and two-sided P values are shown in each panel. The solid line represents the fitted linear regression and the shaded band represents its 95% confidence interval; marginal density plots show the distributions of the paired measurements. The internal-reference-adjusted values showed the strongest agreement with indirect ELISA (Pearson's $r = 0.896$, $P = 0.0005$, 95%CI = 0.610-0.975; Spearman's $\rho = 0.879$, $P = 0.0008$), compared with conventional standardization (Pearson's $r = 0.668$, $P = 0.0349$, 95%CI = 0.065-0.913; Spearman's $\rho = 0.830$, $P = 0.0029$) and raw reads (Pearson's $r = 0.645$, $P = 0.0439$, 95%CI = 0.027-0.907; Spearman's $\rho = 0.830$, $P = 0.0029$) ($n = 10$).

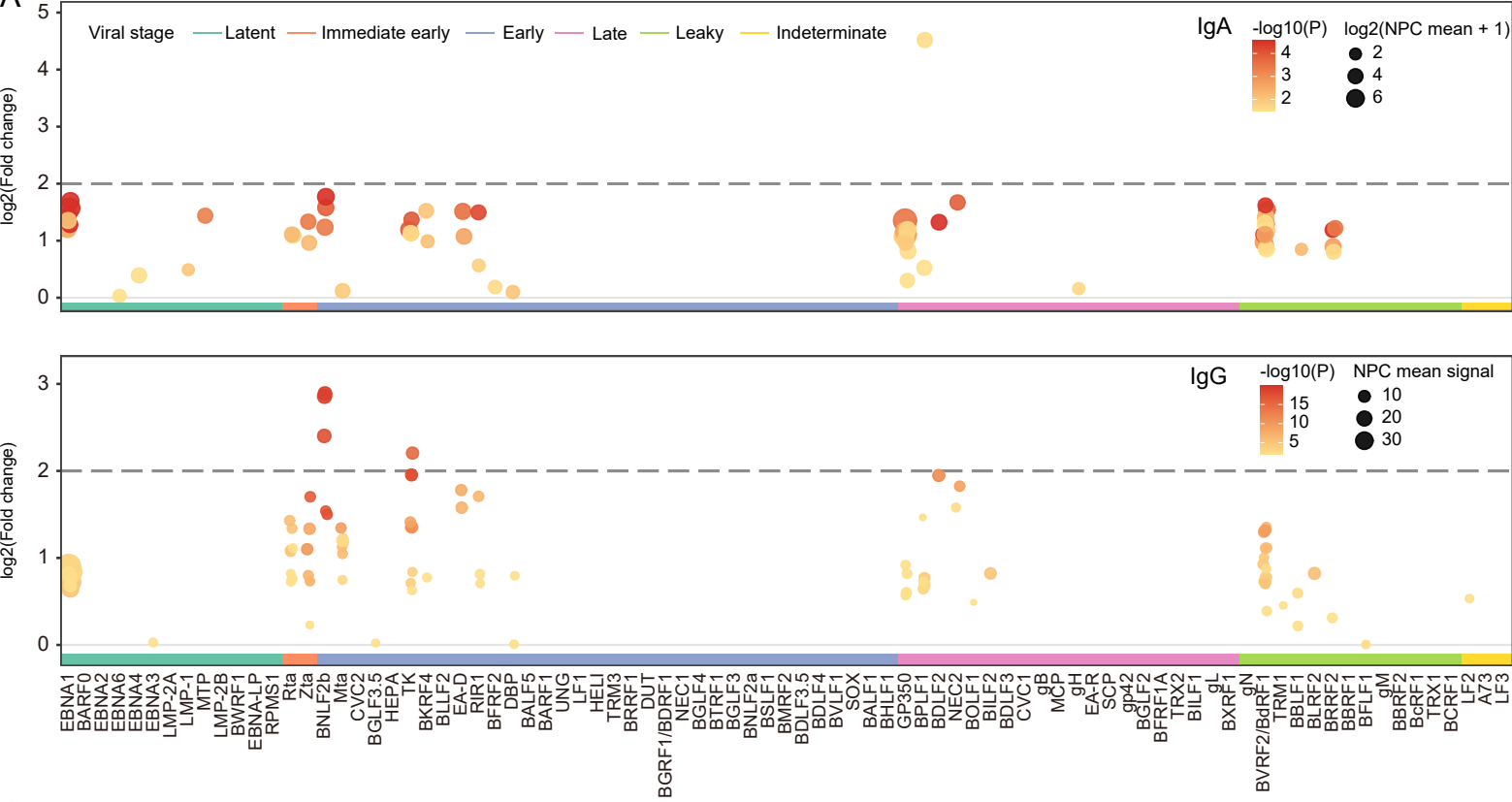
Supplementary Figure 3



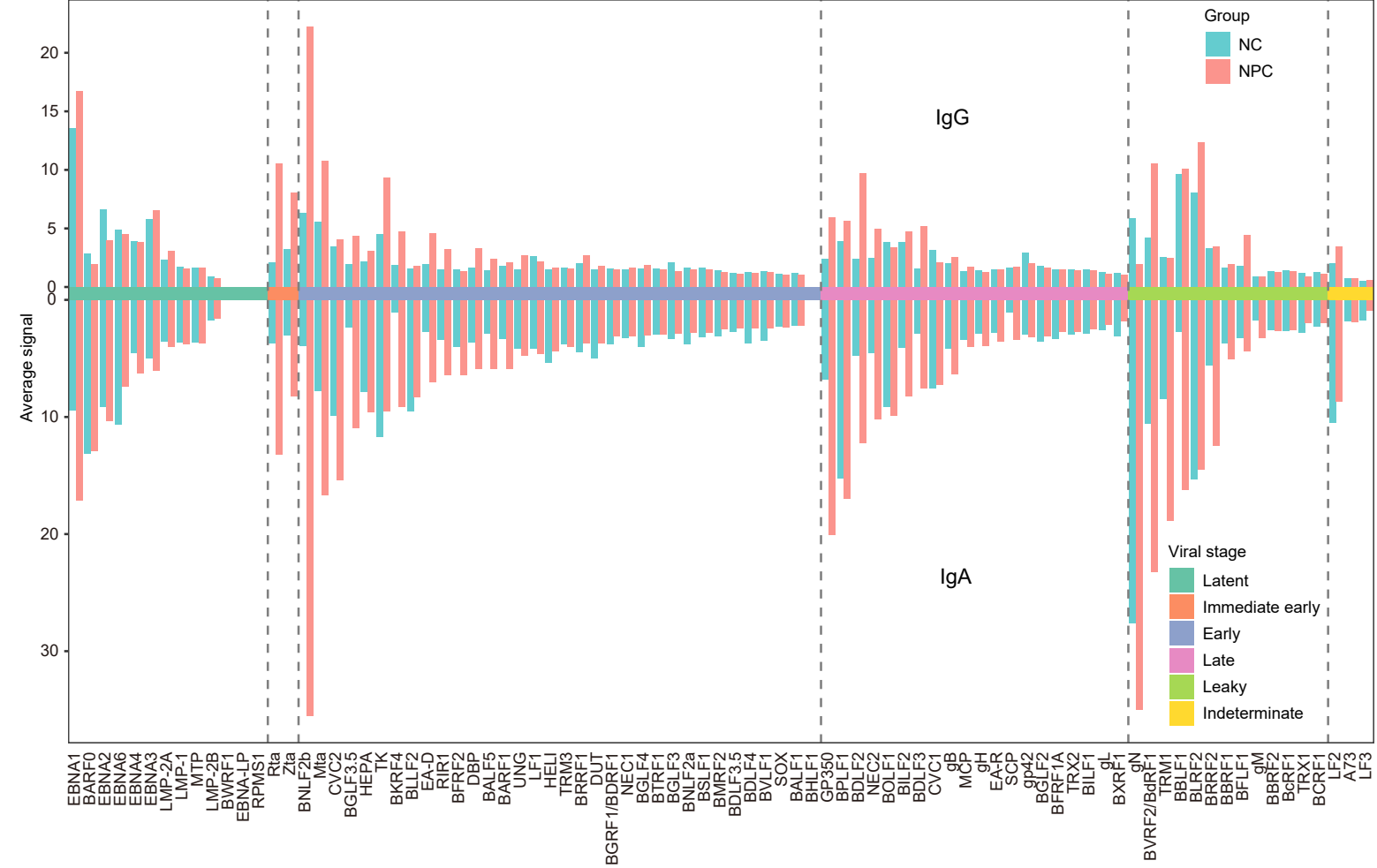
Supplementary Figure 3. EBNA1-derived peptide IgG reactivity indicates widespread prior EBV exposure in both healthy controls and patients with NPC.

Distribution of the maximum normalized IgG reactivity score among all EBNA1-derived peptides for each participant in the NPC (n = 78) and healthy-control (NC n = 93) groups. Scores are displayed on a logarithmic scale; the vertical dotted line denotes the operational cutoff of 2 used to define EBNA1 reactivity, corresponding to a signal-to-background ratio exceeding 2.

A Supplementary Figure 4



B

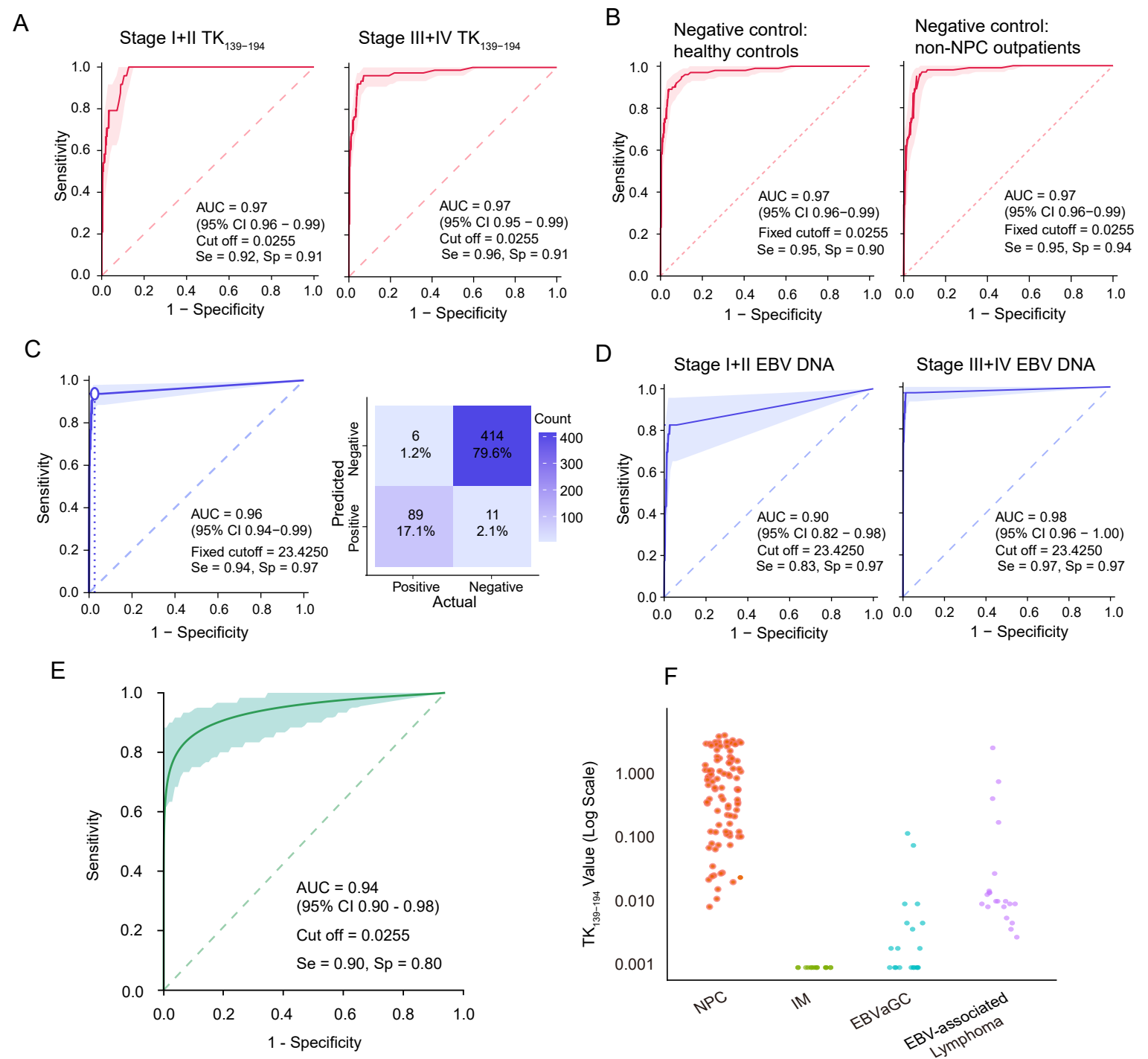


Supplementary Figure 4. Deciphering the dual-isotype (IgG and IgA) humoral atlas across the EBV proteome.

(A) Manhattan plots illustrating differential antibody enrichment across the entire Epstein-Barr virus (EBV) proteome, captured at epitope resolution via EBScope. The top panel shows mucosal IgA signatures, and the bottom panel shows systemic IgG signatures. Peptides are ordered along the x-axis according to their genomic position, with colors representing different viral lifecycle stages: Latent (teal), Immediate-early (orange), Early (blue), Late (pink), Leaky-late (light green), and Indeterminate (yellow). The y-axis represents the $\log_{10}$ fold change (FC) of antibody abundance in NPC patients versus healthy controls. Dotted horizontal lines indicate the significance threshold. Node sizes correspond to the relative abundance of each peptide.

(B) Comparative analysis of average signal intensities for IgG (systemic) and IgA (mucosal) antibodies across individual EBV proteins. The bar plots display the average signal for NPC patients (red) and healthy controls (NC, teal). Proteins are grouped by their respective viral stages as defined in (A). The mirrored visualization (top for IgG, bottom for IgA).

Supplementary Figure 5



Supplementary Figure 5. Stage- and control-stratified diagnostic performance and external validation of TK₁₃₉₋₁₉₄ and EBV-DNA benchmarking.

(A) Receiver operating characteristic (ROC) analyses of TK₁₃₉₋₁₉₄ for early-stage NPC (stages I-II; n = 24) and advanced-stage NPC (stages III-IV; n = 76), each compared with the combined non-NPC control group (n = 690; 495 healthy controls and 195 non-NPC outpatients).

(B) ROC analyses of TK₁₃₉₋₁₉₄ in patients with NPC (n = 100) using healthy controls (n = 495) or non-NPC outpatients (n = 195) as separate negative-control groups.

(C) ROC curve and corresponding confusion matrix for circulating EBV DNA in participants with available EBV-DNA results (95 patients with NPC and 425 controls).

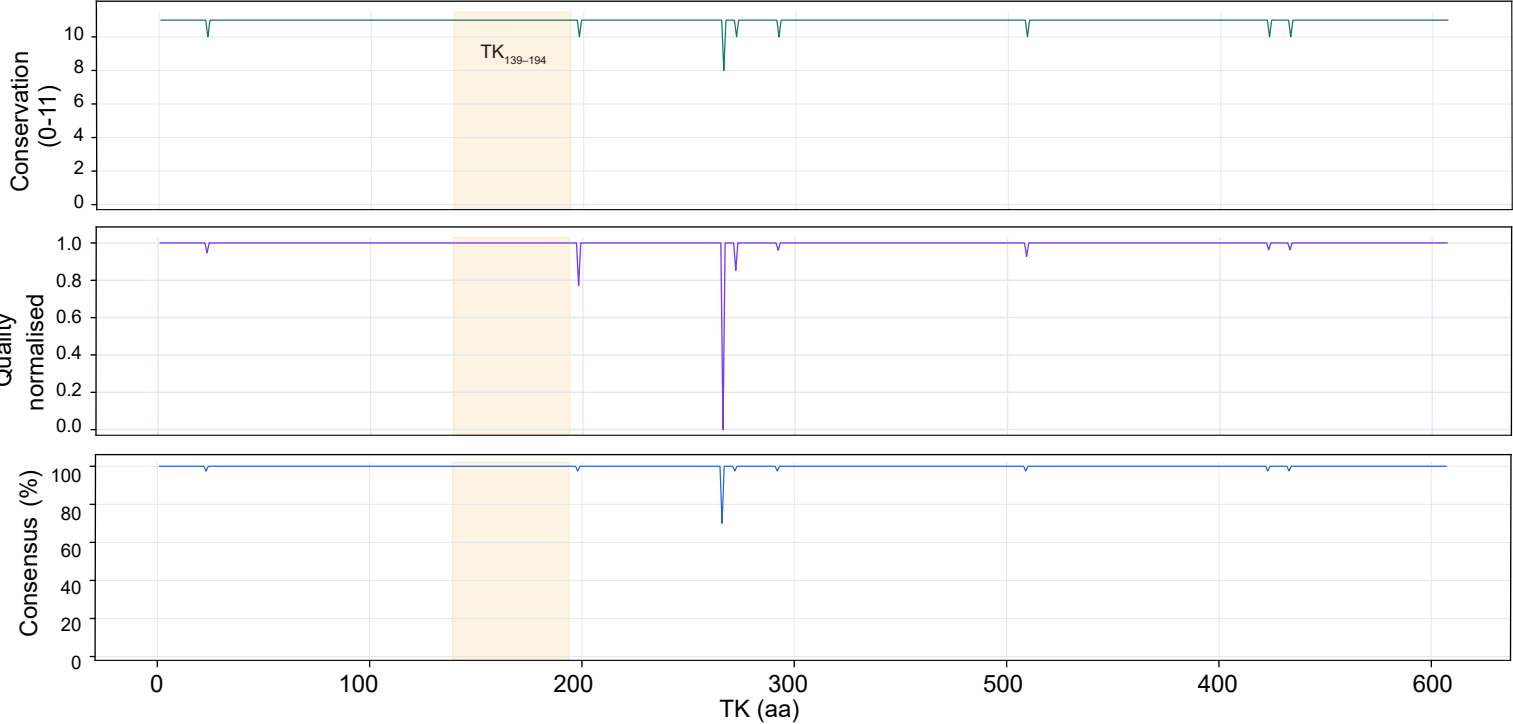
(D) Stage-stratified ROC analyses of circulating EBV DNA for early-stage NPC (stages I-II; n = 23) and advanced-stage NPC (stages III-IV; n = 72).

(E) ROC curve for TK₁₃₉₋₁₉₄ in an independent external cohort from Sichuan comprising 60 patients with NPC and 60 healthy controls.

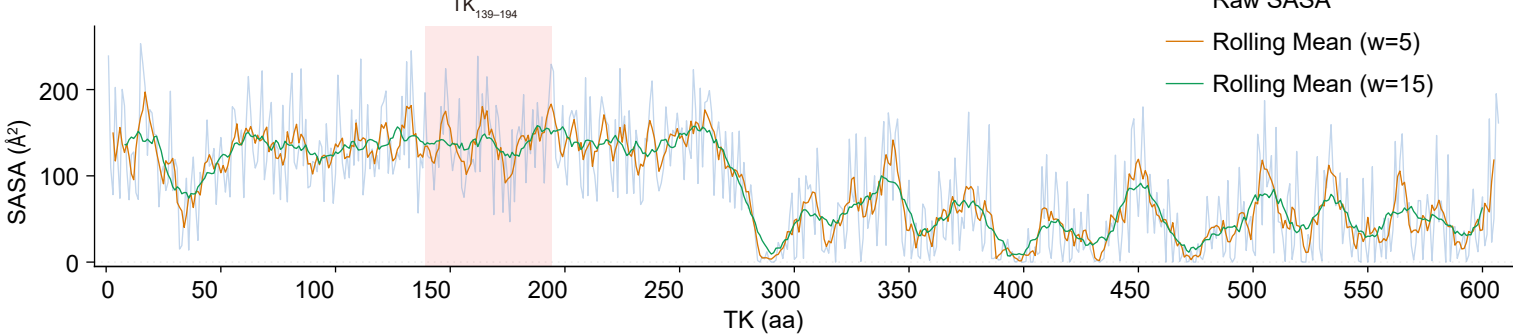
(F) Distribution of TK₁₃₉₋₁₉₄ values in patients with NPC, infectious mononucleosis (IM; n = 15), EBV-associated gastric carcinoma (EBVaGC; n = 20), and EBV-associated lymphoma (n = 21). Each point represents an individual participant, and TK₁₃₉₋₁₉₄ values are displayed on a logarithmic scale. In the confusion matrices, columns represent the actual classification and rows represent the predicted classification; values are shown as counts and percentages of the corresponding cohort. Shaded areas around the ROC curves indicate 95% confidence intervals. AUC, area under the ROC curve; Se, sensitivity; Sp, specificity.

Supplementary Figure 6

A



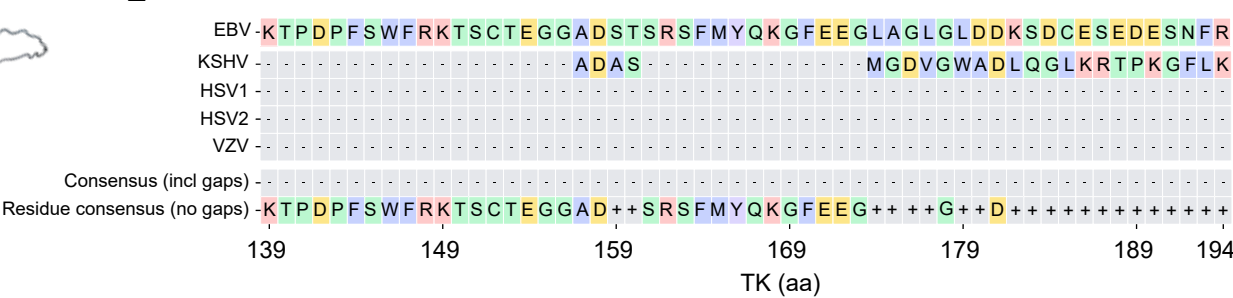
B



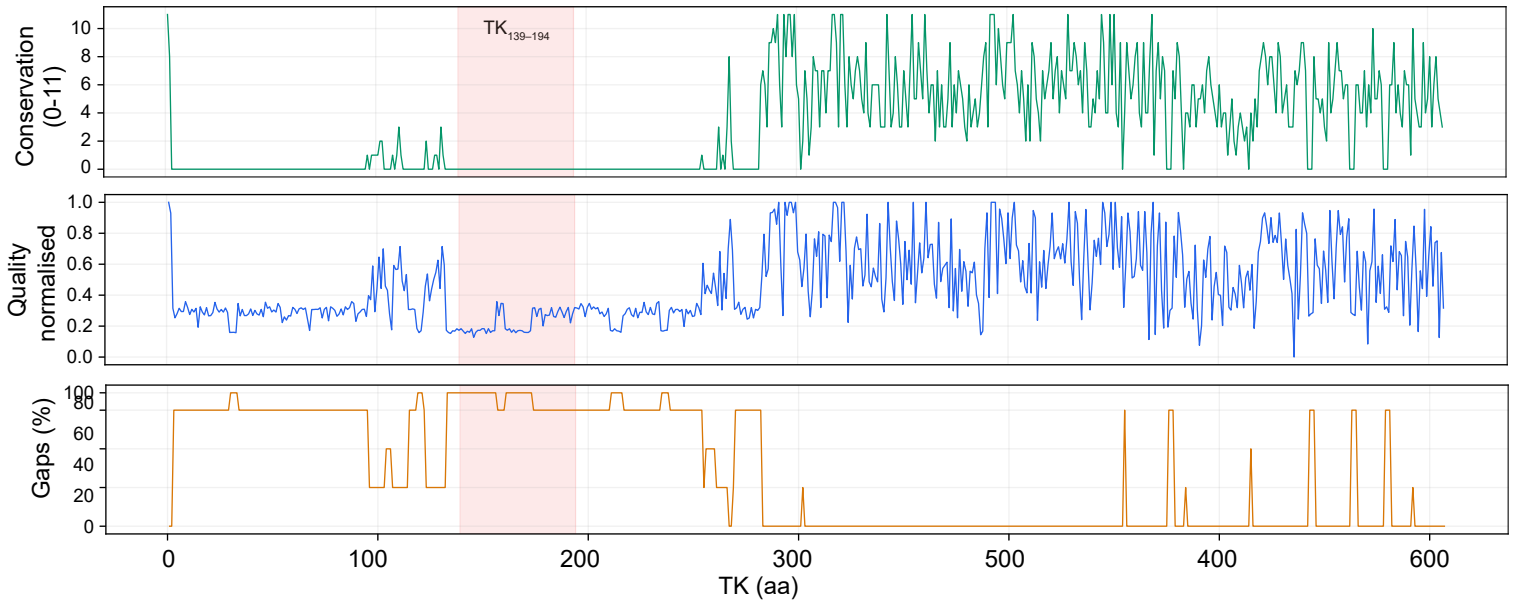
C



E



D



Supplementary Figure 6. Conservation, predicted accessibility, and cross-herpesvirus homology of TK₁₃₉₋₁₉₄.

(A) Residue-level conservation of EBV thymidine kinase (TK) across 40 quality-controlled EBV strain sequences, summarized by Jalview conservation scores (0-11), normalized quality scores, and consensus percentages. The TK₁₃₉₋₁₉₄ region is highlighted.

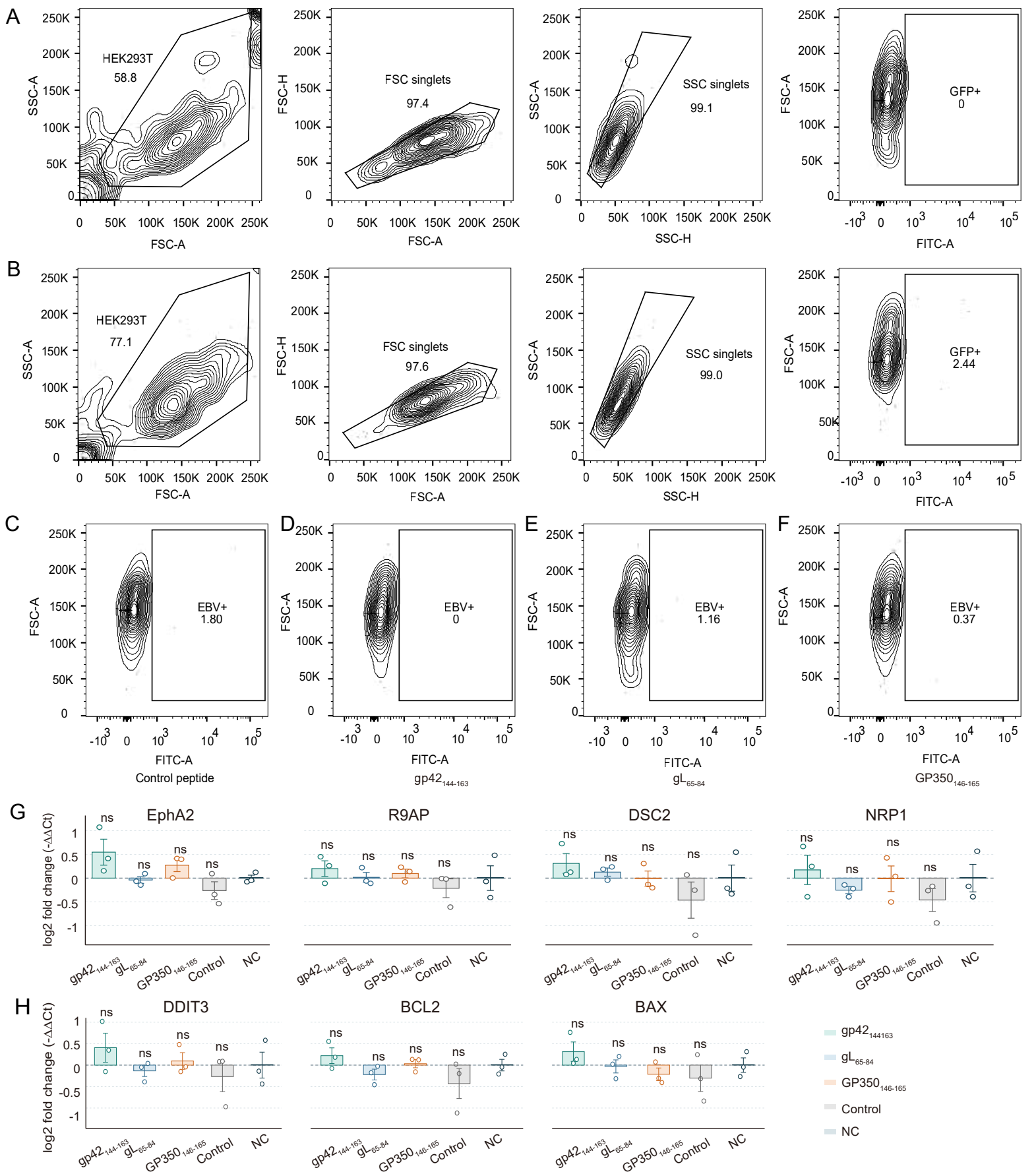
(B) Solvent-accessible surface area (SASA) along the EBV TK sequence, shown as raw residue-level values and rolling means with windows of 5 and 15 residues.

(C) Predicted three-dimensional structure of EBV TK (retrieved from ModelArchive, accession ID: ma-jd-viral-05874) showing the peripheral location of the TK₁₃₉₋₁₉₄ region.

(D) Conservation, normalized quality, and gap frequency across an alignment of canonical human-herpesvirus TK homologs; the region aligned to EBV TK₁₃₉₋₁₉₄ is highlighted.

(E) Residue-level alignment of TK₁₃₉₋₁₉₄ with the corresponding regions of KSHV, HSV-1, HSV-2, and VZV TK homologs. SASA, solvent-accessible surface area; KSHV, Kaposi's sarcoma-associated herpesvirus; HSV, herpes simplex virus; VZV, varicella-zoster virus.

Supplementary Figure 7



Supplementary Figure 7. Effects of peptide blockade on EBV infection and host-cell gene expression.

(A, B) Representative flow-cytometry contour plots illustrating the sequential gating strategy used to identify EBV-positive HEK293T cells in the cells-only control (A) and after EBV exposure (B). HEK293T cells were gated sequentially using FSC-A/SSC-A, FSC-A/FSC-H and SSC-A/SSC-H, followed by identification of GFP-positive events as a readout of EBV infection. (C-F) Representative EBV-positive frequencies after viral exposure in the presence of the control peptide (C), gp42₁₄₄₋₁₆₃ peptide (D), gL₆₅₋₈₄ peptide (E), or GP350₁₄₆₋₁₆₅ peptide (F). Panels A-F show the gating strategy and representative flow-cytometry data underlying the infection-inhibition measurements in Fig. 5G. Representative flow-cytometry plots from three independent experiments with similar results are shown. (G, H) RT-qPCR analysis after peptide treatment for candidate cell-surface genes EphA2, R9AP, DSC2, and NRP1 (G) and cellular stress- and apoptosis-related genes DDIT3, BCL2, and BAX (H). GAPDH was used as the reference gene. Relative expression was calculated by the $2^{-\Delta\Delta C_t}$ method and is presented as log₂ fold change (- $\Delta\Delta C_t$), with the mean of the untreated NC group normalized to zero. Bars show the mean, error bars indicate the SEM. Statistical comparisons were performed on ΔC_t values using two-sided Welch's t-tests against the NC group; all comparisons were not significant ($P \geq 0.05$). NC, untreated control; ns, not significant. n=3 independent biological experiments