

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Sequencing data were generated on an Illumina HiSeq instrument using 150-bp paired-end sequencing. Flow cytometry data were acquired on a BD LSRFortessa X-20 flow cytometer using BD FACSDiva software (v8.0.1). RT-qPCR data were acquired on an Agilent AriaMx Real-Time PCR system using Agilent Aria Real-Time PCR software (v1.71)
Data analysis	Statistical analyses were performed in R v4.4.3, and PhIP-Seq sequencing data were processed using custom Python scripts (Python v3.12). Flow-cytometry data were analysed using FlowJo v10.8.1. Sequence alignments were generated using MAFFT v7.520 with the G-INS-i strategy, and conservation metrics were calculated according to the Jalview v2.11.5.1 algorithms. Structural visualizations were generated using PyMOL v3.0.3, and schematic illustrations were created using BioRender.com. Custom analysis scripts and workflows are available at https://doi.org/10.5281/zenodo.21944590 .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession code PRJNA1423001 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1423001>. The Virscan and EBVscan peptide library designs, de-identified participant metadata, and processed peptide-level reactivity matrices are available in Zenodo under DOI 10.5281/zenodo.21944590 <https://doi.org/10.5281/zenodo.21944590>. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Sex was obtained from clinical records. Female and male participant counts are reported in the cohort characteristics tables. Sex was included as a covariate in the adjusted chemotherapy-response models.

Reporting on race, ethnicity, or other socially relevant groupings

Participants were recruited at clinical centres in Fujian and Sichuan, China. Race and ethnicity were not collected and were not inferred from nationality or recruitment location. No race- or ethnicity-stratified analyses were performed.

Population characteristics

The study included 1,139 serum samples: 239 NPC cases, 649 healthy controls, 195 outpatient non-NPC controls, and 56 EBV-associated non-NPC disease cases (15 infectious mononucleosis, 20 EBV-associated gastric cancer, and 21 EBV-associated lymphoma). The chemotherapy-response analysis included 46 patients with NPC (30 responders and 16 non-responders), after exclusion of five patients from the 51-patient eligible subset following quality control.

Recruitment

Serum samples were collected between 2021 and 2025. Participants with NPC were recruited at Fujian Cancer Hospital. Healthy controls and outpatient non-NPC controls were recruited at the First Affiliated Hospital of Fujian Medical University and Fujian Medical University Union Hospital. The external validation cohort was recruited at Sichuan Cancer Hospital, and participants with EBV-associated non-NPC diseases were recruited at Fujian Medical University Union Hospital. Participants were selected according to the cohort-specific eligibility criteria described in the Methods. Recruitment was consecutive. The hospital-based recruitment and differences in age distributions between cohorts, particularly the paediatric infectious-mononucleosis controls, may limit generalizability and influence between-group comparisons. Participants received no compensation.

Ethics oversight

The study complied with the Declaration of Helsinki and was approved by the Ethics Committee of Fujian Medical University (FMU-2019-16, FMU-2022-120 and FMU-2024-31) and the Ethics Committee for Medical Research and New Medical Technology of Sichuan Cancer Hospital (KY-2026-131-02). Written informed consent was obtained from all participants and/or their legal guardians before enrollment.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical method was used to predetermine sample size. The study included 1,139 serum samples available from 2021 to 2025 across four non-overlapping discovery, refinement, independent validation, and disease-specificity cohorts.

Data exclusions

For the VirScan analysis, 13 healthy-control samples were excluded because they yielded fewer than 4×10^5 peptide-mapped reads, leaving 79 participants with NPC and 81 healthy controls. For EBVScan IgG analyses, one NPC sample and one healthy-control sample were excluded because their estimated relative serum input was <0.10 , leaving 78 participants with NPC and 93 healthy controls. For EBVScan IgA analyses, six NPC samples and five healthy-control samples were excluded because calibration factors were non-finite, ≤ 0.03 , or markedly inconsistent across internal references, leaving 73 participants with NPC and 89 healthy controls. For chemotherapy-response analysis, 28 of the original 79 patients were excluded because they did not receive chemotherapy ($n = 22$), declined treatment ($n = 1$), or received non-gemcitabine regimens ($n = 5$). Five of the remaining 51 patients were excluded because of failed signal detection, leaving 46 patients (30 responders and

	16 non-responders). For survival analysis, six patients were excluded because their IgA data failed the calibration quality-control criteria described above, leaving 73 patients.
Replication	The PhIP-Seq workflow included duplicate samples per plate and inter-plate replicates. Technical reproducibility across batches was assessed using 2,637 paired peptide measurements (Pearson's $r = 0.9976$; Spearman's $\rho = 0.9846$). EBV infection experiments were performed in three independent biological replicates. Diagnostic findings were evaluated in non-overlapping Fujian and external Sichuan validation cohorts.
Randomization	Participants were grouped by clinical diagnosis and were not randomized to disease groups. Sample placement on assay plates, assay order and allocation of cell cultures to treatment conditions were randomized.
Blinding	Investigators were not blinded to group allocation during experiments, outcome assessment or data analysis. Clinical diagnosis was required for group assignment. Automated pipelines, fixed analysis thresholds, predefined plate quality-control criteria, and fixed validation-cohort cutoffs were used to limit analytical bias.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	
Blinding	

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	
Location	

Access & import/export

Disturbance

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a Involved in the study
- ☐ ☒ Antibodies
- ☐ ☒ Eukaryotic cell lines
- ☒ ☐ Palaeontology and archaeology
- ☒ ☐ Animals and other organisms
- ☐ ☒ Clinical data
- ☒ ☐ Dual use research of concern
- ☒ ☐ Plants

Methods

- n/a Involved in the study
- ☒ ☐ ChIP-seq
- ☐ ☒ Flow cytometry
- ☒ ☐ MRI-based neuroimaging

Antibodies

Antibodies used

Isotype-specific capture antibodies were used for PhIP-Seq. ELISA detection used HRP-conjugated goat anti-human IgG (WANTAI BioPharm; 1:5,000), HRP-conjugated goat anti-human IgA (KPL; 1:20,000), and streptavidin–HRP (Thermo Fisher Scientific; 1:5,000), as appropriate for the assay. Details of the capture antibodies and detection reagents are provided in the Supplementary Reagents and Sequences table (Table S14).

Validation

HRP-conjugated anti-human IgG and IgA were used for indirect ELISA at dilutions of 1:5000 and 1:20000, respectively. Each ELISA plate included no-serum blanks, healthy negative-control sera and pooled NPC positive-control sera to monitor background and assay performance. These controls provide assay-level quality control rather than independent validation of antibody specificity. Reagent details are provided in Table S14.

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

HEK293T cells were a kind gift from Prof. Shengxiang Ge at the School of Public Health, Xiamen University.

Authentication

HEK293T cells were authenticated by STR profiling.

Mycoplasma contamination

HEK293T cells were not tested for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

No cell lines used in this study are listed in the ICLAC database of commonly misidentified cell lines.

Palaeontology and Archaeology

Specimen provenance

Specimen deposition

Dating methods

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

Wild animals

Reporting on sex

Field-collected samples

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	This retrospective observational diagnostic-biomarker study was not a prospective interventional clinical trial; trial registration was therefore not applicable.
Study protocol	Four non-overlapping cohorts were used: Cohort I, 79 NPC cases and 94 healthy controls for VirScan discovery and EBVScan refinement; Cohort II, 100 NPC cases, 495 healthy controls, and 195 outpatient non-NPC controls for Fujian validation; Cohort III, 60 NPC cases and 60 healthy controls for external validation; and Cohort IV, 15 infectious-mononucleosis, 20 EBV-associated gastric-cancer, and 21 EBV-associated lymphoma cases for disease-specificity evaluation. Methods are described in the manuscript.
Data collection	Serum was collected from 2021 to 2025, processed immediately, aliquoted to avoid repeated freeze-thaw cycles, and stored at -80 degrees C. Clinical diagnosis, TNM stage, treatment, and response information were obtained from participating hospitals. Tumor response after the third neoadjuvant-chemotherapy cycle was assessed according to RECIST version 1.1
Outcomes	Study outcomes included diagnostic performance, disease specificity, inhibition of EBV infection, chemotherapy response and progression-free survival. The LTP ₈₃₆₋₈₉₁ prognostic analysis was exploratory because the marker was selected using the study data.

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☐	☐ Demonstrate how to render a vaccine ineffective
☐	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐ Increase transmissibility of a pathogen
☐	☐ Alter the host range of a pathogen
☐	☐ Enable evasion of diagnostic/detection modalities
☐	☐ Enable the weaponization of a biological agent or toxin
☐	☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	
Novel plant genotypes	
Authentication	

ChIP-seq

Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	
Files in database submission	
Genome browser session (e.g. UCSC)	

Methodology

Replicates	
Sequencing depth	
Antibodies	
Peak calling parameters	
Data quality	
Software	

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	HEK293T cells were seeded at 1×10^4 cells per well in 96-well plates and cultured overnight to 40-50% confluence. Cells were preincubated with candidate peptide for 1.5 h, exposed to EBV for 3 h, given fresh medium, and harvested 24 h after infection.
Instrument	A BD flow cytometer was used. GFP fluorescence was detected in the FITC-A channel. GFP fluorescence was detected in the FITC-A channel.
Software	Data acquisition was performed on a BD LSRFortessa X-20 flow cytometer using BD FACSDiva software v8.0.1. Flow-cytometry data were analysed using FlowJo v10.8.1.
Cell population abundance	GFP-positive infection rates were quantified across three independent biological replicates.
Gating strategy	Cells were selected using FSC-A versus SSC-A, followed by sequential singlet gates using FSC-A versus FSC-H and SSC-A versus SSC-H. GFP-positive events were identified in the FITC-A channel using uninfected controls to establish the positivity threshold. No viability dye was used. The gating strategy is shown in Supplementary Fig. 7a,b.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	
Design specifications	
Behavioral performance measures	

Acquisition

Imaging type(s)	
Field strength	
Sequence & imaging parameters	
Area of acquisition	
Diffusion MRI	☐ Used ☐ Not used

Preprocessing

Preprocessing software	
Normalization	
Normalization template	
Noise and artifact removal	
Volume censoring	

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

(See [Eklund et al. 2016](#))

Correction

Models & analysis

n/a

Involved in the study

☐

☐ Functional and/or effective connectivity

☐

☐ Graph analysis

☐

☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis