nature communications

Article

https://doi.org/10.1038/s41467-025-60836-9

Genomic landscape of virus-associated cancers

Received: 8 May 2024

Accepted: 5 June 2025

Published online: 01 July 2025

Check for updates

Yoonhee Nam[1,9], Karen Gomez[1,9], Jean-Baptiste Reynier[1,2], Cole Khamnei[1], Michael Aitken[1,3], Vivian Zheng[1], Tenzin Lhakhang[1], Milena Casula[4], Giuseppe Palmieri[4,5], Antonio Cossu[6], Arnold Levine[7], Enrico Tiacci[8] & Raul Rabadan[1,2]

It has been estimated that 15%-20% of human cancers are attributable to infections, mostly by carcinogenic viruses. The incidence varies worldwide, with a majority affecting developing countries. Here, we conduct a comparative analysis of virus-positive and virus-negative tumors in nine cancers linked to five viruses. We observe a higher frequency of virus-positive tumors in males, with notable geographic differences in incidence. Our genomic analysis of 1971 tumors reveals a lower somatic burden, distinct mutation signatures, and driver gene mutations in virus-positive tumors. Compared to virus-negative cases, virus-positive cases have fewer mutations of *TP53, CDKN2*A, and deletions of 9p21.3/*CDKN2*A-*CDKN1A* while exhibiting more mutations in RNA helicases *DDX3X* and *EIF4A1*. Furthermore, an analysis of clinical trials of PD-(L)1 inhibitors suggests an association of virus-positivity with higher treatment response rate, particularly evident in gastric cancer and head and neck squamous cell carcinoma. Both cancer types also show evidence of increased CD8 + T cell infiltration and T cell receptor clonal selection in virus-positive tumors. These results illustrate the epidemiological, genetic, and therapeutic trends across virus-associated malignancies.

An estimated 15-20% of cancers are attributable to infections[1,2], and 8–10% are caused by viruses[3,4]. To date, seven viruses are known to be associated with the development of cancers in humans (oncoviruses): human gammaherpesvirus 4 (HHV-4, also known as Epstein-Barr virus [EBV]), human herpesvirus 8 (HHV-8), human papillomavirus (HPV), human T-cell lymphotropic virus type 1 (HTLV-1), hepatitis B virus (HBV), hepatitis C virus (HCV), and Merkel cell polyomavirus (MCPyV)[5]. In addition, previous studies have observed potential associations of adeno-associated virus 2 (AAV2) with hepatocellular carcinomas (HCC)[6], cytomegalovirus (CMV) with glioblastoma multiforme (GBM)[7], and a "hit-and-run" mechanism of viral involvement in classical Hodgkin lymphoma (CHL)[8].

While the mechanisms of malignant transformation caused by oncoviruses differ, there are some general patterns that are observed[9]. First, oncoviruses cause a persistent, long-term infection, and tumors develop years after the initial infection. For example, most individuals are infected with EBV by early childhood (in developing countries) or adolescence (in developed countries)[10], but an EBV-associated cancer may not develop until old age. Hepatocellular carcinoma develops 10–30 years after infection with HBV or HCV[11], and cervical cancer

[1]Program for Mathematical Genomics and Department of Systems Biology, Columbia University, New York, NY, USA. [2]Department of Biomedical Informatics, Columbia University, New York, NY, USA. [3]Department of Physics, Columbia University, New York, NY, USA. [4]Unit of Cancer Genetics, Institute of Genetic Biomedical Research (IRGB), National Research Council (CNR), Sassari, Italy. [5]Immuno-Oncology & Targeted Cancer Biotherapies, University of Sassari, Sassari, Italy. [6]Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy. [7]Simons Center for Systems Biology, Institute for Advanced Study, Princeton, NJ, USA. [8]Institute of Hematology and Center for Hemato-Oncology Research, Department of Medicine and Surgery, University and Hospital of Perugia, Perugia, Italy. [9]These authors contributed equally: Yoonhee Nam, Karen Gomez. e-mail: rr2579@cumc.columbia.edu

develops 25–30 years following infection with HPV[12]. Second, oncoviruses encode proteins that directly contribute to malignant transformation. In HPV infected cells, the E6 and E7 oncoproteins inhibit the tumor suppressors p53 and Rb, respectively[13]. The vGPCR protein encoded by HHV8 induces angiogenesis and promotes cell transformation[14]. EBV expresses different genes depending on the viral latency program. Type I latency is characterized by the expression of EBV nuclear antigen 1 (EBNA1), essential for viral DNA replication and potentially inhibiting apoptosis, along with EBV-encoded small RNAs (EBERs) and BamHI-A rightward transcript (BART) miRNAs[15]. In contrast, type II latency involves the expression of latent membrane proteins (LMP1 and LMP2), which activate the NF-κB and PI3K/AKT pathways, in addition to the markers of latency I[15]. Latency I is commonly observed in Burkitt lymphoma (BL), while latency II is seen in classical Hodgkin lymphoma (cHL) and nasopharyngeal carcinoma, with EBV-positive gastric cancer being associated with either latency type I or II[15,16]. Third, viral infection is necessary, but not sufficient, for malignant transformation. Many oncoviruses are highly prevalent in the general population: 90–95% of people worldwide are infected with EBV[10], 80% of individuals will acquire an HPV infection by age 45[17], and MCPyV is detected in 80% of individuals in the general population by age 50[18]. However, only a small fraction of those infected with oncoviruses will develop cancer, suggesting additional genetic and/or environmental factors are required. The factors that contribute to the malignant transformation of virus-infected cells remain incompletely understood, but are known to include a combination of environmental, immune, inherited, and somatic components. While many of these components have been described for individual cancer types, relatively little has been reported about the clinical and genetic factors that are common across virus-associated cancers.

In this work, we utilize previously reported datasets in addition to newly acquired data (Kaposi sarcoma) to include all seven known oncoviruses identified in 923 patients across 14 viral-positive cancer types. We identify patterns of common phenotypic characteristics, somatic drivers, and therapeutic responses among these malignancies. This study provides a comprehensive analysis of human cancers that develop in the context of viral infection and key factors related to their pathogenesis.

Results

Virus-associated cancer show unique epidemiological trends

Virus-associated cancers are known to follow unique epidemiological patterns compared to non-virus-associated cancers. For example, EBV-positive HL is more frequently observed in cases with mixed cellular histology, males, children, older adults, and in developing countries. In these regions, HL incidence shows an earlier peak, primarily in children under 15 years, and is characterized by a high prevalence of EBV positivity. In contrast, young-adult onset nodular sclerosis (NS) HL, typical of the 15 to 39 age group in industrialized nations, is usually EBV-negative[19,20]. BL has been traditionally classified into two clinical variants, i.e., endemic BL (eBL) and sporadic BL (sBL), that present different demographic (eBL tend to be younger), geographic (most eBL are from equatorial Africa), virus status (nearly all eBL are EBV +, while a small fraction of sBL are) and somatic mutation characteristics[21] (this should be taken into account in BL comparisons along the manuscript).

In order to illustrate other common demographic characteristics of virus-associated malignancies, we analyzed data from the Global Cancer Observatory (GLOBOCAN 2020)[22] and published incidence rates in 48 studies of 11 cancer types linked to 5 viruses and 13 non-virus-associated cancers[23–104]. First, we compared the number of viral cancers in males versus females (M/F) reported in select published studies[23–104] (Fig. 1A and Supplementary Data 1). We found that the M/F ratio was greater overall in virus-associated cancers compared to nonviral cancers ($p = 2.2e\text{-}16$, Fisher's exact test). Among studies that reported M/F ratio for virus-positive and virus-negative tumors specifically, virus-positive cases tended to have a greater M/F ratio than virus-negative cases ($p = 5.23e\text{-}8$, Fig. 1B and Supplementary Data 2). This trend was consistent in gastric cancer (GC, $p = 1.04e\text{-}10$) and HL ($p = 0.015$), both of which have been reported previously[105–107]. In contrast, no difference in the M/F ratio and viral status was observed for BL, despite higher M/F incidence ratios of 2:1 to 4:1 being reported in both eBL and sBL[108]. A lower M/F ratio was observed in MCPyV-positive Merkel cell carcinoma (MCC) compared to virus-negative MCC, as noted previously[4]. Digital papillary adenocarcinoma, which has been recently associated to HPV42, is reported to be more frequent in males compared to females at a ratio of 4:1[109].

To examine how the incidence of virus-associated cancers differs by geographic location, we compared the age-standardized incidence rates (ASR) of 4 cancers in 185 countries reported in GLOBOCAN 2020. In HL, EBV-positive cases occur most frequently in North Africa, the Middle East, and South America, with the lowest incidence occurring in East Asia (Fig. 1C, D). In contrast, most cases of EBV-positive nasopharyngeal carcinoma (NPC) occur in China and Southeast Asia (Fig. 1E, F). Similarly, Kaposi sarcoma and cervical cancer (nearly all of which are virus-positive) show disparities in incidence by geographic location (Supplementary Fig. S1). These results illustrate that the locations of global hot spots of virus-positive tumor incidence vary by virus and even among cancers associated with the same virus. These disparities reflect differences in risk factors for virus-positive tumors among human populations, both genetic (e.g., inherited susceptibility polymorphisms[110,111]) and environmental (e.g., oncovirus prevalence[112], and lifestyle factors such as smoking or diet, which affect overall cancer risk[113]).

Virus-positive tumors have fewer somatic mutations than virus-negative tumors

We aggregated somatic mutation data from 1971 tumors in published studies of 9 cancers subjected to whole-exome sequencing/WES (plasmablastic lymphoma [PBL][114] [$n = 15$], cHL[115,116] [$n = 69$], cervical cancer [CC][117] [$n = 178$], hepatocellular carcinomas [HCC][118] [$n = 196$], GC[105] [$n = 440$], and head and neck squamous cell carcinoma [HNSCC][119] [$n = 487$]), targeted DNA sequencing (BL[21] [$n = 29$], PBL[114,120] [$n = 36$], primary central nervous system lymphoma [PCNSL][121] [$n = 58$], MCC[122] [$n = 71$], and cHL[123] [$n = 293$]), and/or whole genome sequencing/WGS (BL[21] [$n = 91$] and cHL[115] [$n = 24$]) (Supplementary Data 3). In general, virus-positive tumors had a lower count of nonsynonymous mutations than virus-negative tumors, as reported in individual cancer types[116,120,124–127] (Figs. 2A, 2B, Supplementary Data 3 and Supplementary Data 5), including in particular PCNSL[125] (Wilcoxon test $p = 1.4e\text{-}7$), cHL (targeted, $p = 2.4e\text{-}6$; WES, $p = 0.032$; WGS, $p = 0.045$), PBL[120] ($p = 0.044$), HNSCC[126] ($p = 6.5e\text{-}6$) and, as a trend, GC[127] ($p = 0.086$), CC ($p = 0.17$), and MCC[124] ($p = 0.072$) (Fig. 2A).

Conversely, in HCC and BL, virus-positive cases had a greater mutation load than virus-negative cases ($p = 0.026$ and $p = 5.4e\text{-}4$, respectively), but the count of nonsynonymous mutations in genes previously described as BL drivers[21] still trended towards lower in virus-positive BL cases ($p = 0.12$, Fig. 2B and Supplementary Fig. S2 and Supplementary Data 5), consistent with that report[21]. Furthermore, when restricting the analysis to driver genes (Supplementary Data 4), the lower mutation count in virus-positive cases became statistically significant in MCC ($p = 0.015$, Supplementary Fig. S2), while virus-positive GC and HCC had more driver gene mutations than their virus-negative counterparts (GC: $p = 0.036$; HCC: $p = 8.8e\text{-}4$). Overall, the total mutation count and/or driver mutation count was lower in virus-positive compared to virus-negative tumors in most cancers studied (Fig. 2B).

In addition, we observed that several cancer subtypes and virus strains differed in mutation load (Supplementary Data 6). The NS subtype, the most common subtype of cHL in our datasets (targeted,

https://doi.org/10.1038/s41467-025-60836-9

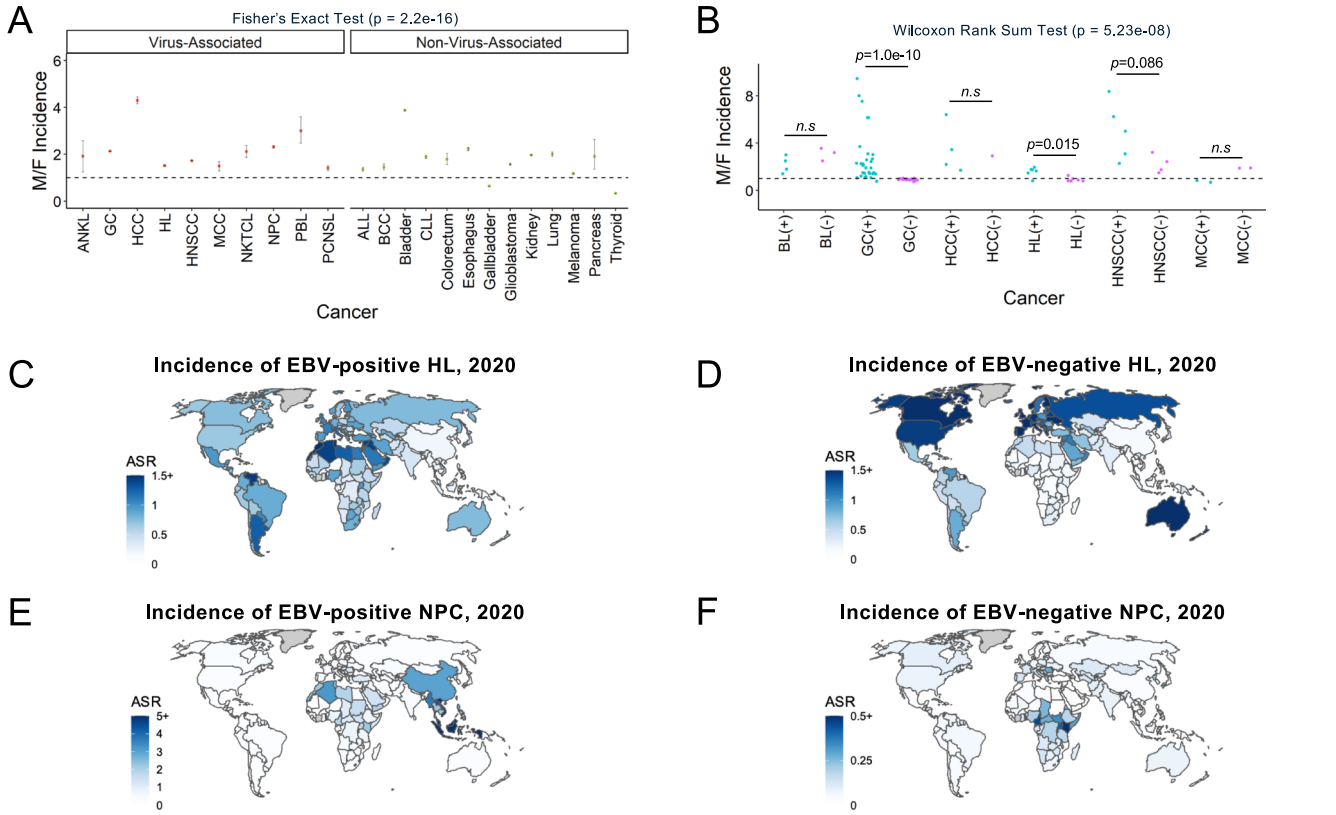

Fig. 1 | Epidemiological trends of virus-associated cancers. A Incidence ratios of virus-associated and non-virus-associated cancers analyzed using two-sided Fisher's Exact Test. Data are presented as point estimates (M/F incidence ratios) with error bars indicating 95% confidence intervals. ANKL (92/48), GC (659302/309048), HCC (16091/3744), HL (60913/40219), HNSCC (266342/154358), MCC (521/352), NKTCL (787/370), NPC (29769/12855), PBL (443/148), PCNSL (1814/1286), ALL (3239/2380), BCC (121701/51529), Bladder (487885/125906), CLL (9908/5249), Colorectum (622/348), Esophagus (13067/5857), Gallbladder (47920/74542), Glioblastoma (31071/19801), Kidney (287986/146433), Lung (2861/1416), Melanoma (21469/18141), Pancreas (105/55), Thyroid (207549/613624). **B** Virus-positive and virus-negative tumors in virus-associated cancers in males compared to females (M/F) reported in selected published studies, analyzed using two-sided Wilcoxon Rank Sum Test. Each point corresponds to an incidence ratio reported in a published study in Supplementary Data 2. ANKL, aggressive NK-cell leukemia; GC, gastric cancer; HCC, hepatocellular carcinoma; HL, Hodgkin lymphoma; HNSCC, head and neck squamous cell carcinoma; MCC, Merkel cell carcinoma; NKTCL, Natural killer/T-cell lymphoma; NPC, nasopharyngeal carcinoma; PBL, plasmablastic lymphoma; PCNSL, primary central nervous system lymphoma; ALL, acute lymphoblastic leukemia; BCC, basal cell carcinoma; CLL, chronic lymphocytic leukemia; BL, Burkitt lymphoma. **C**, **D** Estimated incidence rates of EBV-positive HL (**C**) and EBV-negative HL (**D**) by country. **E**, **F** Estimated incidence rates of EBV-positive NPC (**E**) and EBV-negative NPC (**F**) by country. Map data from Natural Earth (https://www.naturalearthdata.com/, public domain), produced by rnaturalearth R package[194]. ASR, age-standardized rate. Source data are provided as a Source Data file.

84.6%, 204/241; WES, 73.4%, 47/64), exhibited the highest mutation load relative to other subtypes in both targeted and WES cohorts (p = 1.6E-06 and 0.010, respectively). However, after adjusting for subtypes, sex, and age, only EBV status remained significantly associated with a lower mutation load in the targeted and WES cohorts, consistent with a previous report[116] (coefficient = − 0.44, − 2.1; p = 0.009, 0.001). As expected[21], eBL had a higher total mutation load compared to sBL, reflecting a similar trend with EBV status (p = 0.00029). Since EBV-positivity was highly associated with eBL (p = 7.68e-12), when both subtype and EBV status were considered simultaneously, neither variable showed a significant independent association with total mutation load.

In CC, there was no significant difference in mutation load between squamous carcinoma and adenocarcinoma subtypes. However, age at diagnosis positively correlated with mutation load (Spearman's ρ = 0.31, p = 4.1E-05, Supplementary Data 6), consistent with a previous study[128]. Notably, multiple HPV strains in CC differed significantly in mutation load, with HPV31 exhibiting the highest levels (Fold change = 7.07, p = 0.02, Supplementary Fig. S3 and Supplementary Data 6). To account for age, HPV strains, viral status, on mutation load, we used generalized linear model regression model (GLM) and found that HPV31 and virus-positive status remained significantly associated with elevated mutation load (coefficient = 1.81 and 1.09, p = 0.013 and < 0.001, respectively).

In HNSCC, females exhibited higher driver gene mutation counts than males (p = 0.0058), and age correlated positively with both total and driver mutation counts (Spearman's ρ = 0.19 and 0.27, p = 2.4E-05 and 1.2E-09, respectively; Supplementary Data 6), consistent with previous findings[129]. Unlike CC, there were no significant differences in mutation load between HPV strains in HNSCC (Supplementary Fig. S4), as reported in a prior study[130].

In HCC, no difference in mutation load was evident between HBV and HCV related HCC. Neither Activated B-cell (ABC) nor Germinal Center B-cell (GCB) subtypes of PCNSL differed in their mutation loads.

Virus-associated cancers display unique mutation signatures

To detect and quantify the relative contribution of COSMIC mutation signatures[131] within the virus-associated cancers, we applied SigProfilerExtractor[132] to extract de novo mutational signatures and decompose them into COSMIC signatures across 7 available cancers. Virus-positive tumors exhibited different activities of mutation signatures compared to virus-negative tumors of the same cancer type (Fig. 3, S5 and Supplementary Data 7). UV light is known to be the major etiological agent of MCC tumors in the absence of viral

https://doi.org/10.1038/s41467-025-60836-9

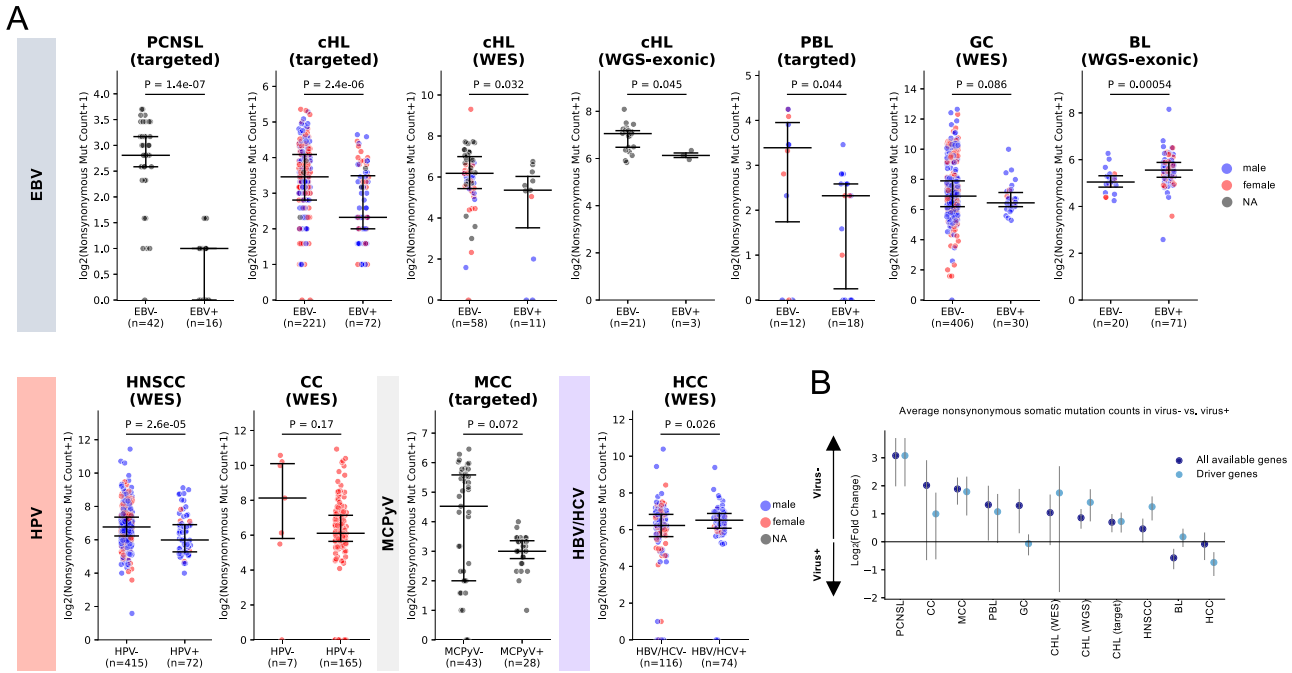

Fig. 2 | Mutation burden of virus-positive and virus-negative tumors in 9 cancers. A Counts of somatic nonsynonymous mutations in virus-positive and virus-negative tumors in the same cancers. Data are presented as median values with interquartile range (25th–75th percentile). Actual median values for virus-negative and positive samples in the shown cancer types are PCNSL: 6 and 1, cHL (targeted): 10 and 4, cHL (WES): 126.5 and 65, cHL (WGS): 132 and 69, BL: 32 and 46, PBL: 9.5 and 4, GC: 117.5 and 86, CC: 279 and 68, HNSCC: 108 and 62.5, MCC: 22 and 7, HCC: 74 and 90.5. *P*-values are calculated by a two-sided Wilcoxon rank-sum test. **B** Log_2(fold change) of average number of somatic nonsynonymous mutations (all genes: dark blue, driver genes: light blue) in virus-negative tumors compared to virus-positive tumors. PCNSL ($n = 58$), CC ($n = 172$), MCC ($n = 71$), GC ($n = 436$), PBL ($n = 51$), cHL (target, $n = 293$; WES, $n = 69$; WGS, $n = 24$), HNSCC ($n = 487$), BL ($n = 91$ for all genes; 68 EBV-positive eBL, 6 EBV-negative eBL, 3 EBV-positive sBL, 14 EBV-negative sBL), BL ($n = 120$ for driver genes) and HCC ($n = 190$). GC, gastric cancer; HCC, hepatocellular carcinoma; cHL, classical hodgkin lymphoma; HNSCC, head and neck squamous cell carcinoma; MCC, Merkel cell carcinoma; PBL, plasmablastic lymphoma; PCNSL, primary central nervous system lymphoma; CC, cervical cancer; BL, Burkitt lymphoma. Data are presented as log_2(fold change) with error bars indicating 95% confidence intervals. Source data are provided as a Source Data file.

infection. Accordingly, the relative contribution of mutations attributed to SBS7a/7b/UV-light was lower in MCPyV-positive compared to MCPyV-negative, as reported previously[122] (Fig. 3A, B and Supplementary Data 7).

In GC, the proportion of mutations associated with SBS15/mismatch-repair (MMR) deficiency was higher in EBV-negative than EBV-positive cases (Fig. 3C, D and Supplementary Data 7). Microsatellite instability (MSI), as assessed by standard methods, is a defining characteristic of a GC subtype that is exclusively EBV-negative and comprised 73/406 (18%) of EBV-negative patients in the TCGA cohort. Accordingly, the relative contribution of mutations attributed to SBS15 and SBS20/MMR-deficiency were higher in the conventionally defined MSI subtype compared to other EBV-negative cases (Supplementary Fig.S6 and Supplementary Data 8). This suggests that the difference in MMR signatures between EBV-positive and EBV-negative GC are largely driven by the MSI subtype.

In HNSCC, HPV-positive HNSCC had the higher relative contribution of SBS2/APOBEC mutations (Fig. 3E, F and Supplementary Data 7), as reported in a previous study[2]. This is consistent with the hypothesis that HPV oncoproteins may increase APOBEC3A and APOBEC3B expression and mutagenic activity[133,134]. In a multivariate analysis of HPV-positive HNSCC, which accounted for sex, HPV strain, age, and total mutation count, HPV33-associated HNSCCs showed significantly lower SBS2/APOBEC mutation counts, while total mutation count was positively associated with SBS2/APOBEC mutation counts, both findings consistent with previous reports[130,135] (coefficient = −1.37 and 0.012, $p = 0.003$ and < 0.001, respectively; Supplementary Data 9, Supplementary Fig. S7).

The mutation counts associated with the SBS5/clock-like signature were higher in HPV-negative cervical cancer (CC) compared to HPV-positive CC ($p = 0.013$, Supplementary Data 7), consistent with our observation that total mutation counts are positively correlated with age and tend to be higher in HPV-negative cases. Notably, the levels of SBS2/13/APOBEC-associated signatures varied between HPV strains, with HPV31 significantly exhibiting the highest mutation counts (Supplementary Fig. S8 and Supplementary Data 8). A similar trend observed in HNSCC was also apparent in CC, where the HPV33 strain exhibited lower APOBEC signature levels compared to HPV16, although this difference was not statistically significant in CC. These strain-specific differences aligned with our findings on total mutation load. Importantly, these differences appear to be primarily driven by variations in total mutation load, as HPV strain-specific effects were no longer significant in a multivariable model that accounted for virus strain, virus status, total nonsynonymous mutation load, and age. In this model, virus-positive status and total nonsynonymous mutation load were positively associated with APOBEC signature mutations (Supplementary Data 10), consistent with previous findings[117].In BL exomes[21] (Supplementary Fig. S5B), a de novo mutational signature, SBS96B, was identified. Although SBS96B did not meet the cosine similarity threshold for decomposition into known COSMIC signatures, potentially related COSMIC signatures included SBS46, SBS5, SBS17, and SBS15. This signature showed a higher proportion in EBV-positive BL samples compared to EBV-negative BL samples (Supplementary Data 7), consistent with the original genome-wide mutation signature analysis on SBS5, SBS17, and SBS15 in the same cohort[21]. However, in partial contrast to the original analysis, we

https://doi.org/10.1038/s41467-025-60836-9

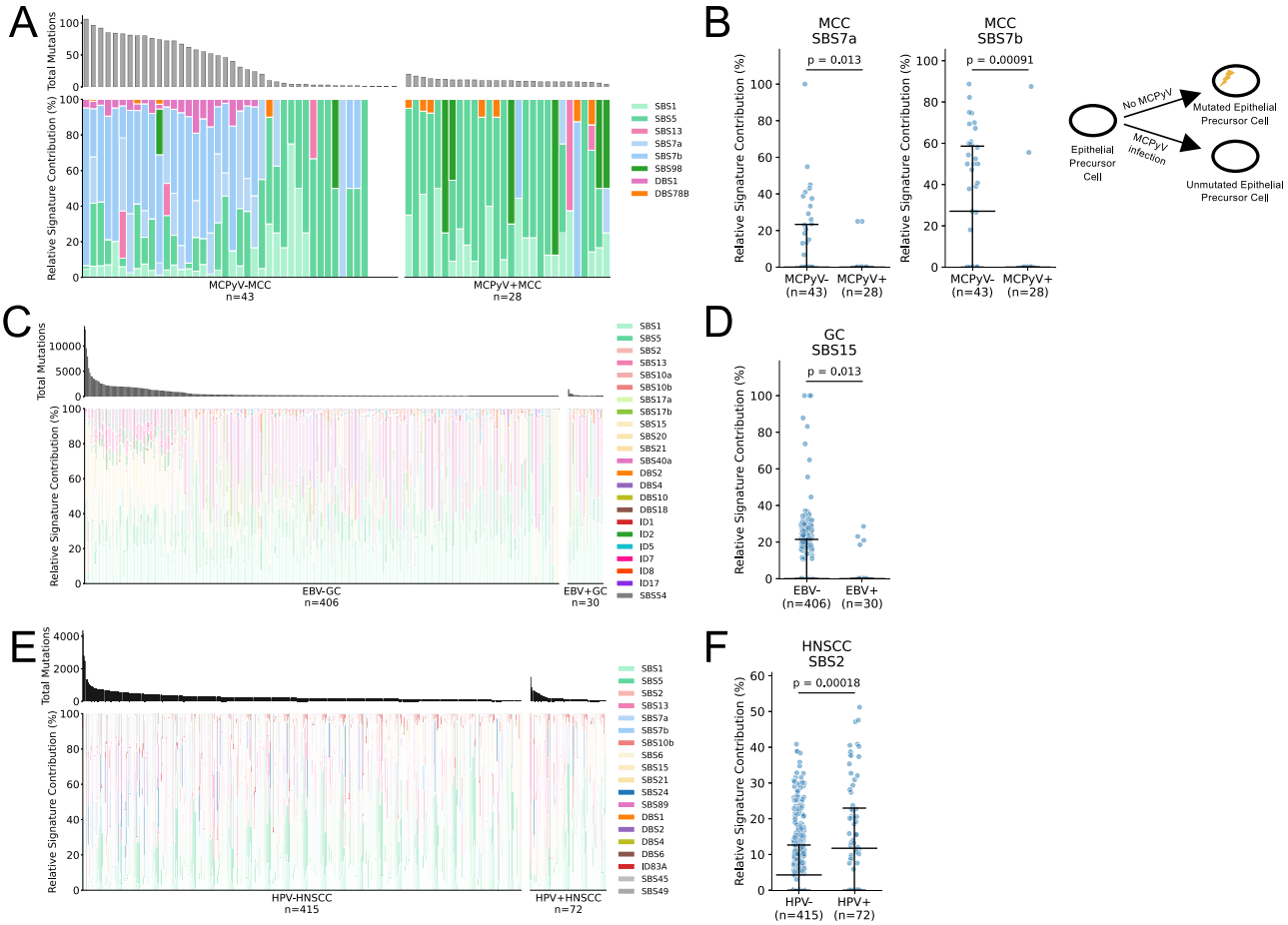

Fig. 3 | Mutation signatures in virus-associated cancers. A, **B** MCC ($n = 71$), (**C**, **D**) GC ($n = 436$), (**E**, **F**) HNSCC ($n = 487$). Total mutations (top bar plot) and proportion of mutations associated with each signature (bottom stacked plot) in virus-positive compared to virus-negative cases. Signatures identified from SigProfilerExtractor are shown. Signature names with SBS96, DBS78, or ID83 followed by a capital letter are de novo mutational signatures identified from the cancer cohort, which could not be decomposed into known COSMIC (v.3.4) signatures. The rest are decomposed COMSIC signatures. Schematic representation of the effect of the absence of processes behind key mutation signatures in virus-associated cases is shown in (**B**) MCC (MCPyV-positive). MCC, Merkel cell carcinoma; GC, gastric cancer; HNSCC, head and neck squamous cell carcinoma. **B**, **D**, **F** Data are presented as median values with interquartile range (25th–75th percentile). *P* values are calculated by a two-sided Wilcoxon rank-sum test. Source data are provided as a Source Data file.

did not detect the SBS9/polη signature, which is associated with non-canonical AID activity and was previously identified in BL genomes[21].

In HCC, there was no difference in absolute count of mutations attributed to mutation signatures (Supplementary Fig. S5C), consistent with the similar mutation burden in virus-positive and -negative HCC overall. However, HCC tumors positive for HBV had a greater proportion of mutations due to SBS24/aflatoxin, an environmental carcinogen known to predispose to HBV -mediated cirrhosis, as observed in previous studies[136–138] ($p = 0.010$) (Supplementary Data 8). There was no significant difference in the absolute counts or proportions of signatures in EBV-positive versus EBV-negative PBL, likely due to the limited number ($n = 1$) of EBV-negative cases (Supplementary Fig. S5D). Overall, these results illustrate that the signatures of somatic mutation processes vary depending on infection status for each cancer, highlighting differing selective pressures on the cancer genomes in the presence or absence of viral oncoproteins.

To identify differences in structural variant (SV) and copy number (CN) signatures between virus-positive and virus-negative tumors, we conducted a signature analysis using SigProfilerExtractor[132], following the same methodology as in SBS analysis. This analysis was performed on WGS samples from HCC, HNSCC, CC, and GC obtained from PCAWG, as well as BL and cHL (Supplementary Figs. S9, 10 and Supplementary Data 11). Among SV signatures, no significant differences were observed between virus-positive and virus-negative tumors, except for an enrichment of the SV2 signature, which consists of non-clustered translocations (COSMIC v3.4), in HPV-positive HNSCC ($q = 0.0025$, Supplementary Fig. S9G). In the CN signature analysis, several de novo signatures that could not be decomposed into known COSMIC CN signatures were identified as differing in multiple cancer types (Supplementary Fig. S10).

Furthermore, we compared differences in chromosomal instability (CIN) signatures between virus-positive and virus-negative tumors using CIN signature data from a recent study on TCGA cohorts, including HCC, HNSCC, CC, and GC (Supplementary Data 3)[139]. Notably, the CX1 signature, which involves whole-arm or whole-chromosome changes and is associated with chromosome mis-segregation due to defective mitosis and/or telomere dysfunction[139], showed a significantly higher relative contribution in virus-positive cases compared to virus-negative cases in HNSCC, GC, and HCC ($p = 7.9e\text{-}6$, 1.3e-4, 0.0032, respectively; Supplementary Fig. S11, 12 and Supplementary Data 12, 13). Although CC exhibited a similar trend, the difference was not statistically significant, likely due to the small number of HPV-negative cases ($n = 4$, Supplementary Fig. S12). This observation is consistent with studies showing that HPV's E6 and E7

https://doi.org/10.1038/s41467-025-60836-9

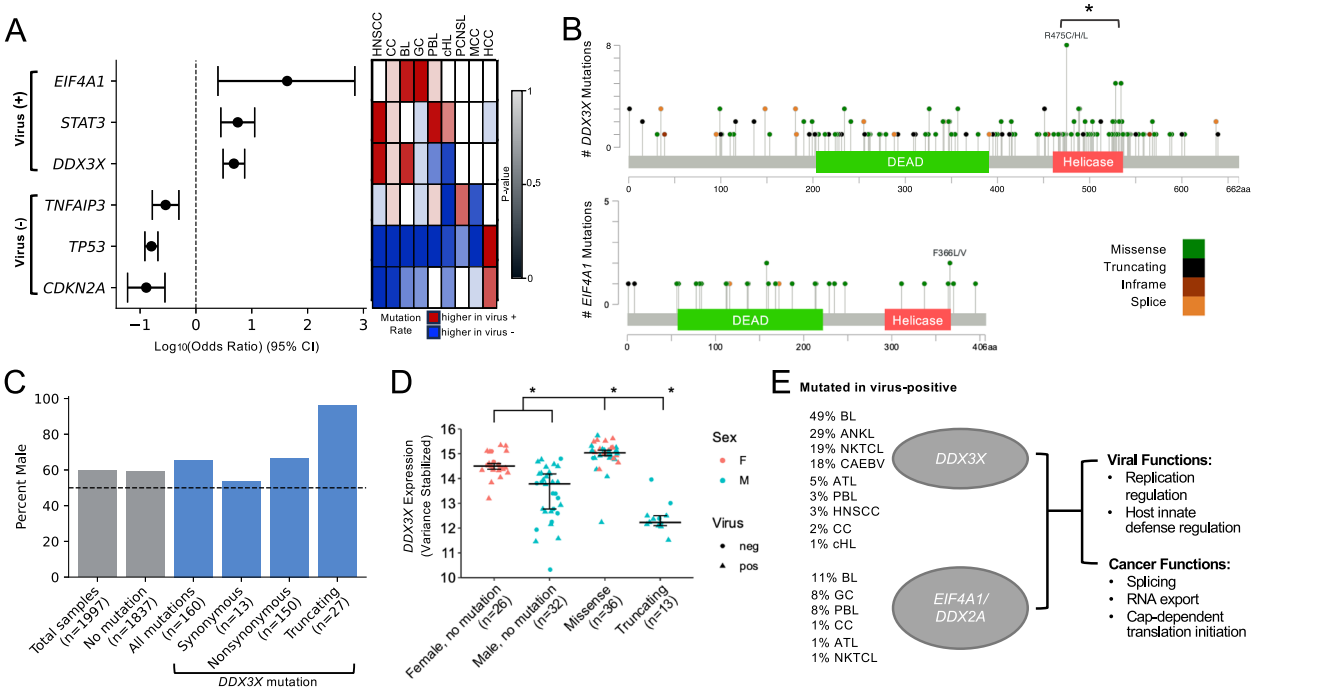

Fig. 4 | Somatic mutations in *EIF4A1* and *DDX3X*, both RNA helicases of the DEAD (Asp-Glu-Ala-Asp) box protein family, are recurrent genetic lesions associated with virus-positive status. A Combined $\log_{10}$(odds ratio) of mutation in genes associated with virus-positive (top) and virus-negative (bottom) status ($q < 0.005$) from pooled data of 1971 tumors across 9 virus-associated cancers. Data are presented as $\log_{10}$(odds ratio) values with error bars indicating 95% confidence intervals. The heatmap on the right displays the cancer cohorts included in the pooled data for the calculation of each gene, with colors representing mutation rate trends in each cohort (red: higher in virus-positive; blue: higher in virus-negative) and shades indicating the two-sided Fisher's exact test *p*-value. HNSCC, head and neck squamous cell carcinoma; CC, cervical cancer; BL, Burkitt lymphoma; GC, gastric cancer; PBL, plasmablastic lymphoma; cHL, classical Hodgkin lymphoma; PCNSL, primary central nervous system lymphoma; MCC, Merkel cell carcinoma; HCC, hepatocellular carcinoma. **B** Mutations in *DDX3X* and *EIF4A1* in 2488 tumors. * $p < 0.05$, two-sided binomial test. **C** Fraction of patients that are male by *DDX3X* mutation status. **D***DDX3X* expression by *DDX3X* mutation status and sex in Burkitt lymphoma ($n = 117$). * $p < 0.05$, two-sided MWU test. Data are presented as median values with interquartile range (25th–75th percentile). **E** Frequencies of mutation of *DDX3X* and *EIF4A1* in virus-positive tumors overall and summary of key biological functions. ANKL, aggressive NK-cell leukemia; NKTCL, Natural killer/T-cell lymphoma; CAEBV, chronic active Epstein-Barr virus disease; ATL, Adult T-cell leukemia/lymphoma. Source data are provided as a Source Data file.

proteins disrupt normal cell cycle regulation and mitotic progression, leading to chromosomal abnormalities[140].

Virus-positive tumors harbor frequent mutations in RNA helicases *DDX3X* and *EIF4A1*

To identify genomic loci preferentially mutated in virus-positive tumors, we compared the rate of nonsynonymous mutations and copy number aberrations in the pooled cohort of 602 virus-positive and 1369 virus-negative tumors from 9 cancers (1,971 total cases; Supplementary Data 3 and Supplementary Fig. S13; see Methods). We found that three genes had significantly elevated odds of mutation in virus-positive tumors: *EIF4A1* (OR = 43.07, 95% CI = 2.48-701.39, $q = 1.65e\text{-}3$, Fisher's exact test, BH corrected), *STAT3* (OR = 5.61, 95% CI = 2.79-11.28, $q = 4.80e\text{-}7$), and *DDX3X* (OR = 4.78, 95% CI = 3.06-7.46, $q = 5.57e\text{-}9$) (Fig. 4A, S14, and Supplementary Data 14).

When looking at individual cancer types, *EIF4A1* mutations were significantly more frequent in EBV-positive GC compared to EBV-negative GC ($p = 5.95e\text{-}3$). This trend was consistently observed across all other cancer types analyzed, including BL, PBL, and CC (Supplementary Fig. S14 and Supplementary Data 14). Similarly, *DDX3X*, an RNA helicase in the same family as *EIF4A1*, was more frequently mutated in EBV-positive HNSCC ($p = 0.044$), as previously reported[141], and a similar trend was observed in BL and CC, though not statistically significant (Supplementary Fig. S14 and Supplementary Data 14). Specifically for BL, we were able to increase the statistical power by extending the analysis to incorporate data on *DDX3X* from six different genomic studies of BL[21,142–146] including eBL ($n = 144$) and sBL ($n = 177$) cases, of whom 142 known to be EBV-positive and 119 EBV-negative (Supplementary Data 15). Notably, we found that *DDX3X* mutation was strongly (though not exclusively) associated with EBV-positive status ($p = 3.78e\text{-}6$), which has been discussed in previous studies but without statistical significance[21,142–146], and endemic subtype ($p = 4.25e\text{-}7$, Supplementary Data 14, Supplementary Data 16). No differences were observed between EBV type 1 and type 2 (Supplementary Data 16). *STAT3* mutations were significantly more frequent in virus-positive PBL ($p = 0.050$), as noted in previous studies[147], and in virus-positive HNSCC ($p = 0.0052$), with cHL and CC showing similar trends compared to their virus-negative cases (Supplementary Data 14).

Furthermore, an analysis of recurrent copy number aberrations in 6 available cancers (HCC, PBL, CC, BL, HNSCC, GC) revealed 14q32.33 loss (OR = 2.79, 95% CI = 2.14-3.65, $q = 2.44e\text{-}13$) and 11q23.3 loss (OR = 2.55, 95% CI = 1.98-3.30, $q = 2.57e\text{-}12$) more frequent in virus-positive tumors (Supplementary Fig. S16). A recent study reported that EBV's protein EBNA1 binds to a specific region on human chromosome 11q23.3, inducing breakage and structural variations[148]. However, the loss of 11q23.3 was enriched in virus-positive HNSCC, CC, and HCC, all of which associated with non-EBV virus, while the EBV-associated BL, GC and PBL showed the opposite trend (Supplementary Fig. S16B).

Meanwhile, virus-negative tumors had higher odds of *TP53* mutations (OR = 8.20, 95% CI = 4.79–8.20, $q = 4.60e\text{-}52$), which were significant and known for most cancer types individually, including HNSCC[2], GC[105], MCC[18], BL[21], PBL[120], among others. However, HCC exhibited the opposite trend, with *TP53* mutations being significantly associated with HBV/HCV-positive HCC, as previously reported[149]. Mutations in *CDKN2A* were also more frequently observed in virus-negative tumors (OR = 7.75, 95% CI = 3.59–16.76, $q = 9.42e\text{-}12$), and

were significant in HNSCC individually, as noted in a prior study[2]. Consistently, loss of 9p21.3/*CDKN2A,CDKN1A* were observed in virus-negative cancers (OR = 4.96, 95% CI = 3.65–6.73, q = 1.77e-25) (Figure S16C). Lastly, *TNFAIP3* mutations were enriched in virus-negative cases (OR = 3.49, 95% CI = 2.02–6.04, q = 3.72e-7), and were significant in cHL individually, as reported previously[150] (Fig. 4A, S15, and Supplementary Data 14).

EIF4A1 and *DDX3X* are RNA helicases of the DEAD (Asp-Glu-Ala-Asp) box protein family, which are known to play a role in splicing, RNA export, and cap-dependent translation initiation[151]. To further explore the role of these genes in virus-associated cancers, we expanded the analysis from the 1971 initial cases to include 201 cases from the extended BL cohort mentioned above (Supplementary Data 15) and also 316 cases from tumors that are virus-associated in almost 100% of instances: Kaposi sarcoma (KS, 10 newly sequenced cases), aggressive NK cell lymphoma (ANKL)[152] (n = 14), adult T cell leukemia/lymphoma (ATL)[153] (n = 81), extranodal NK T-cell lymphomas (NKTCL)[154] (n = 100), and NPC[155] (n = 111) (Supplementary Data 2), for a total of 2,488 cases. Overall, we identified 29 *EIF4A1* nonsynonymous mutations (25 in virus-positive, 4 in virus-negative cases) and 150 *DDX3X* nonsynonymous mutations (93 in virus-positive, 38 in virus-negative, 11 in both, 8 in virus status unknown cases) (Supplementary Data 17). Focusing on the more numerous *DDX3X* mutations, among virus-positive tumor samples they were detected in 49% (67/136) BL, 29% (4/14) ANKL, 19% (19/100) NKTCL, 1.2% (1/86) cHL, 5% (4/81) ATL, 3% (1/30) PBL, 3% (2/60) HNSCC, and 2% (3/144) of CC. *DDX3X* mutations occurred in the helicase domain more frequent than expected by chance (p = 1.97e-16 overall; p = 4.56e-9 for virus-positive only and p = 2.90e-6 for virus-negative only cases, binomial test) (Fig. 4B and Supplementary Data 17). The *DDX3X* gene is located on the X chromosome and was previously reported to escape X inactivation in females[156]. Both truncating events and at least some missense mutations in *DDX3X* have been previously described as causing functional loss of protein activity[146]. While only 60% of patients were male (1193/1997), truncating *DDX3X* mutations occurred almost exclusively in males (26/27, 96%; 14/14 virus-positive, 11/12 virus-negative, 1 male with unknown virus status) (Fig. 4C), consistent with a previous study in BL[157]. As most of *DDX3X*-mutated cases were from the BL cohort, we focused on this disease to evaluate the relationship between mutation status and *DDX3X* expression, using published RNA-sequencing data[21]. We observed a significantly lower *DDX3X* expression in unmutated male versus female cases (median = 13.79 and 14.50, p = 2.4e-6, MWU test), consistent with escape from X inactivation[156]. Cases with *DDX3X* missense mutations had an elevated *DDX3X* expression irrespective of sex compared to unmutated cases (median = 15.04 and 14.24, p = 1.2e-9), potentially suggesting that overexpression of missense mutants may favor their ability to decrease *DDX3X* function, while cases with truncating mutations (9 EBV-positive; 4 EBV-negative) had a lower expression (median = 12.23 and 14.24, p = 1.79e-5) (Fig. 4D), consistent with them being loss-of-function events. Similarly, in the TCGA study of HNSCC, *DDX3X* expression was lower in unmutated male versus female cases (median = 12.58 and 12.99, p < 2.2e-16), and was also lower in cases (3 HPV + ; 2 HPV-) with truncating mutations compared to unmutated cases (median = 10.37 and 12.71, p = 1.5e-4) (Supplementary Fig. S17). Together, these results suggest that mutations in *DDX3X* and *EIF4A1* may play a role in virus-positive tumors in various types of cancer (Fig. 4E).

Virus-associated cancers exhibit more frequent responses to immunotherapy

PD-L1 overexpression has been linked to better overall survival in patients treated with immune checkpoint inhibitors (ICI) in several tumor types, including GC[158], HNSCC[159], and MCC[160]. *PD-L1* expression has been associated with infection by oncoviruses including EBV[161], HPV[162], HBV[163], and MCPyV[160]. To determine whether virus positivity might be a useful marker for response to ICI therapy, we evaluated the correlation of viral status with response to ICI therapy with anti-PD(L)1 in 32 cohorts reported on ClinicalTrials.gov that had available therapy response and virus infection status data, representing four virus-linked cancers (Supplementary Data 18 and Supplementary Data 19).

Virus positivity was significantly associated with ICI therapy response in GC (OR = 2.27, 95% CI = 1.17-4.29, p = 0.011, Fisher's exact test) and HNSCC (OR = 1.85, 95% CI = 1.24-2.74, p = 0.0018), with a similar trend in HCC (OR = 1.30, 95% CI = 0.95-1.78, p = 0.098), but not MCC (p = 0.85) (Fig. 5A). The same cancer types displayed significant association between PD-L1 expression and ICI therapy response, including GC (p = 1.55e-6), HCC (p = 0.0068), and HNSCC (p = 0.034), and a trend was observed also in MCC (p = 0.15). Higher tumor mutational burden (TMB) was associated with ICI therapy response in GC (p = 1.55e-6), exhibiting a similar pattern in HNSCC (p = 0.14), the only two cancer types with such data available (Supplementary Data 19).To determine the relationship between virus positivity and other prognostic markers of ICI therapy response, we compared the expression of PD-L1 [CD274] in TCGA's studies of GC[105], HCC[118], and HNSCC[119]. CD274 expression was higher in virus-positive GC compared to virus-negative GC (median = 6.74 and 5.21, p = 1.6e-07), but this was not the case in HCC or HNSCC (Fig. 5B). Using results of CIBERSORT[164,165] deconvolution of the same TCGA samples, we found evidence of higher CD8 + T cell infiltration in EBV-positive versus EBV-negative GC (median = 0.19 and 0.09, p = 7.4e-09), and in HPV-positive versus HPV-negative HNSCC (median = 0.13 and 0.07, p = 4.3e-10), but not between virus-positive and virus-negative HCC (Fig. 5C). These results were replicated in all other deconvolution approaches tested[166–169] (Supplementary Fig. S18). Similarly, analysis of TRUST4[170,171]-extracted T cell receptors showed higher T cell β receptor clonal selection, measured by counts per clonotypes per thousand reads (CPK), in both virus-positive GC (median = 185.16 and 310.73, p = 7.8e-10) and HNSCC (median = 345.72 and 444.44, p = 0.017) but not in HCC (Fig. 5D). We observed similar trends for the α chain, as well as with the metric of clonality (Supplementary Fig. S19). The effect of virus positivity appears independent from TMB, as virus-positive GC and HNSCC have fewer mutations than virus-negative GC and HNSCC, respectively (Fig. 2A), while in HCC, virus-positive and negative cases have similar mutation load (p = 0.12). These results suggest that virus-positive status may be a positive prognostic marker for patients undergoing ICI therapy for GC and HNSCC, which may be correlated with higher CD8 + T cell infiltration, T cell receptor clonal selection, and T cell exhaustion (for GC only).

Discussion

This study builds upon and extends prior research while providing insights into the epidemiological, somatic, and immune components commonly implicated in the pathogenesis of oncovirus-associated cancers. The observed higher male incidence of virus-associated cancers confirms earlier epidemiological findings in certain malignancies, such as EBV-positive HL and GC[105–107]. Likewise, the finding of a reduced mutation burden in virus-positive tumors relative to virus-negative counterparts is consistent with previous reports in individual tumor types[116,120,124–127], including HNSCC[126] and PCNSL[125]. We also identified frequent RNA helicase mutations (*DDX3X* and *EIF4A1*) in virus-positive tumors spanning multiple cancer types; although *DDX3X* mutations have been reported in HNSCC[141], our results demonstrate their broader relevance across a pan-cancer cohort. Moreover, by aggregating data from 252 BL cases, we found a significant association between *DDX3X* mutations and EBV-positive BL, an association that had been noted previously but not shown to be statistically significant[21,142–146]. Finally, our analysis of immunotherapy response suggests that virus-positive status is associated with enhanced response to checkpoint blockade in gastric cancer and head and neck squamous cell carcinoma.

https://doi.org/10.1038/s41467-025-60836-9

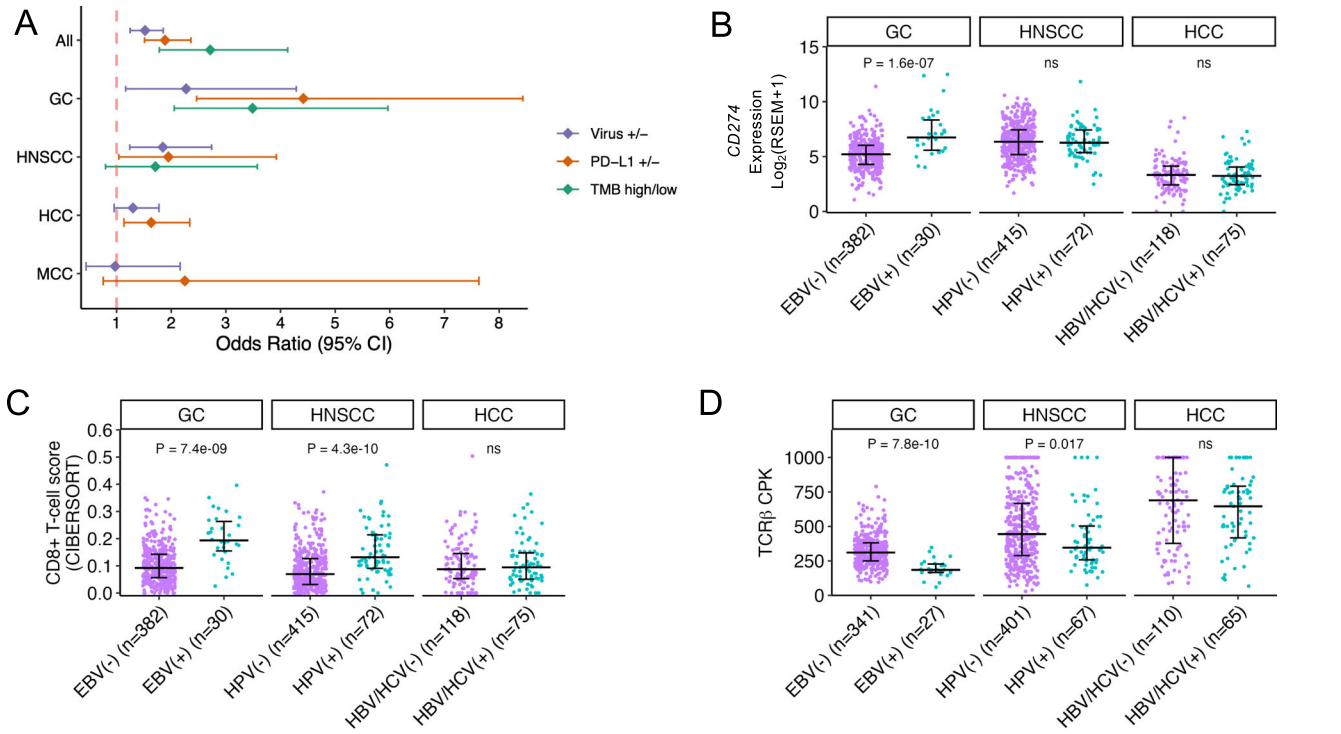

Fig. 5 | Analysis of biomarkers for immunotherapy response in virus-associated cancers. A Odds ratio of positive response to treatment with PD-1/PD-L1 inhibitors with virus-positive status, PD-L1 positive status, and/or high tumor mutation burden (TMB) in 32 studies representing four types of cancer, Fisher's exact test. Data are presented as odds ratio values with error bars indicating 95% confidence intervals. **B** Log_2(RSEM + 1) expression of *PD-L1* (*CD274*), (**C**) CIBERSORT CD8 + T cell infiltration score, and (**D**) TCRβ clonotypes per thousand reads (CPK), versus viral status of tumors in TCGA studies of GC (TCGA-STAD), HCC (TCGA-LIHC), and HNSCC (TCGA-HNSC). GC, gastric cancer; HCC, hepatocellular carcinoma; HNSCC, head and neck squamous cell carcinoma; MCC, Merkel cell carcinoma. *P* values are calculated by a two-sided Wilcoxon rank-sum test. Data are presented as median values with interquartile range (25th–75th percentile). Source data are provided as a Source Data file.

Through analysis of cancer incidence rates reported in a selection of published studies, we noted virus-associated cancers display greater incidence in males compared to females relative to non-virus-associated cancers. This may be caused in part by immunologic predisposition towards viral infection in male compared to females. In general, females have a more robust immune response to infection, which has been attributed to X-chromosome inactivation and regulation of the immune response by genetic, hormonal, and environmental mediators[172,173].

By a large-scale analysis of DNA sequencing data from 1971 tumors collected from different studies, we found that virus-positive tumors generally display a lower mutation load compared to virus-negative tumors. It has been hypothesized that the oncogenic activity of virus-encoded proteins removes selective pressure for somatic mutations. However, unlike the other virus-associated cancers, virus-positive HCC and BLs (comprised mostly but not exclusively of eBL) have a greater mutation load compared to virus-negative cases. In HCC, this may reflect the activity of HCV oncoproteins that inhibit DNA repair and induce double-stranded breaks in the DNA[174], while in EBV-positive BL, it is more difficult to propose hypotheses on the underlying reason(s). Yet, EBV-positive BL had a lower driver mutation load compared to EBV-negative BL, which may be attributed to the activity of viral proteins such as EBNA1 that reduce selective pressure for the acquisition of driver genetic alterations seen commonly in virus-negative BL[21].

Our analysis noted HPV strain-specific differences in mutation patterns across CC and HNSCC, with consistent trends observed in both cancers. APOBEC signatures were strongly correlated with total mutation load in both cancers[117,135], and HPV strains exhibited distinct APOBEC-associated mutation loads. In particular, HPV33 consistently demonstrated lower APOBEC-associated mutation loads compared to HPV16 in both HNSCC[130] and CC. In CC, the strain-specific differences in APOBEC signature levels appeared to be largely driven by total mutation load. In contrast, in HNSCC, HPV33 retained significantly lower APOBEC-associated mutation counts[130] even after adjusting for total mutations, suggesting additional strain-specific mechanisms beyond overall mutational burden. These findings indicate that APOBEC-mediated mutagenesis plays a central role in shaping the mutational landscape of HPV-positive CC and HNSCC, with HPV strain-specific effects potentially reflecting shared viral-host interaction mechanisms across these cancer types. However, the small sample size for certain HPV strains, such as HPV31 ($n = 2$) in CC, limits the generalizability of these findings. Larger studies are needed to confirm these strain-specific effects and to determine their potential implications for tumor progression and clinical outcomes.

Consistent with Zapatka et al. [2], we observed a lack of TP53 mutations in HPV-positive HNSCC. In addition, we observed that somatic mutation of the RNA helicase protein *DDX3X* was more frequent in virus-positive tumors compared to virus-negative tumors. *DDX3X* additionally functions as a component of the innate immune signaling pathway and is known to inhibit replication of viruses such as HBV by activating production of IFN-beta[156,175]. Some RNA viruses, including HCV and HIV, exploit functions of *DDX3X* to aid in viral replication[156,175]. In cancer, *DDX3X* has been described as both a tumor suppressor and an oncogene in different cancer types and even among different tumors of the same cancer type[176]. *DDX3X* is expressed in many tissues of the body and escapes X chromosome inactivation[156,176]. The relatively high frequency of mutations in *DDX3X* in virus-positive tumors and the near-exclusive male bias for truncating mutations suggests that loss of function of *DDX3X* may contribute to the pathogenesis of some virus-associated cancers, particularly BL, which had

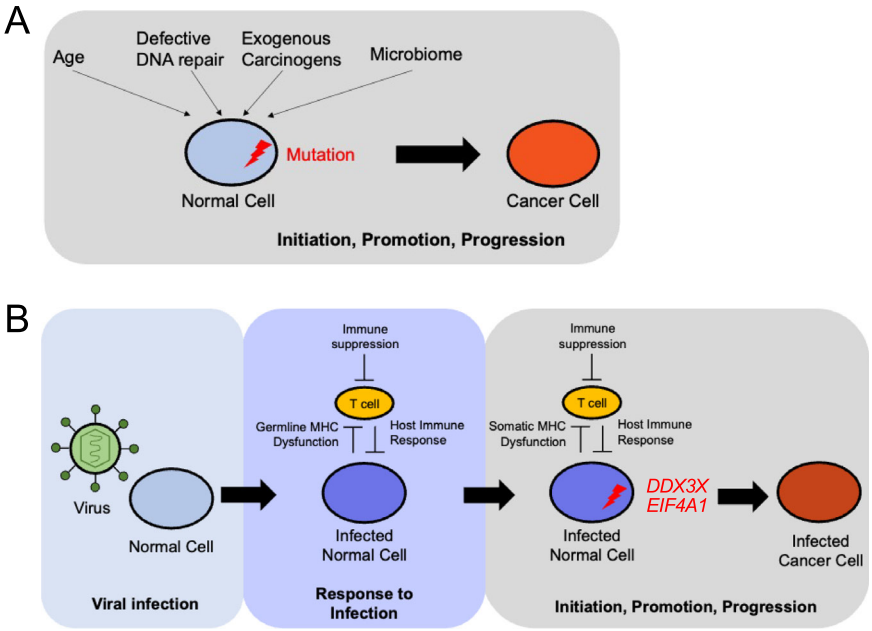

Fig. 6 | Models of oncogenesis for virus-associated and non-virus-associated cancers. A Model for oncogenesis in the absence of viral infection. A normal cell accumulates driver mutations as a result of age, defective DNA repair, exogenous carcinogens, or microbiome interactions, leading under selective pressure to initiation, promotion, and progression that ends in the malignant transformation of the cell. **B** Model for oncogenesis in the presence of viral infection. A normal cell is infected with a virus, and a latent infection is established as a result of inadequate host immune response, potentially associated with germline MHC dysfunction or other inherited risk factors. The infected normal cell acquires somatic mutations in specific genes, such as chromatin modifiers like RNA helicases *DDX3X* and *EIF4A1*, leading to initiation, promotion, and progression that ends in the malignant transformation of the infected cell.

the highest frequency of *DDX3X* mutations in this study and for which similar findings were recently reported in another study[157]. It is worth noting that *DDX3X* mutations, although enriched, did not occur exclusively in EBV-positive BLs. Gong and colleagues reported[146] the role of *DDX3X* and its Y-chromosome paralog *DDX3Y* in facilitating MYC-driven lymphomagenesis. Given the pleiotropic role of *DDX3X* in viral recognition and RNA processing and its higher mutational frequency in virus-associated tumors beyond MYC-driven lymphomas, it will be interesting to molecularly dissect the dual role of these mutations in viral and cell processes leading to tumor development. Lastly, while our findings on *EIF4A1* and *DDX3X* remain robust and underscore the potential importance of these RNA helicases in virus-positive tumors, they are derived from a limited number of cancer types. Future studies incorporating larger and more diverse cohorts will help expand our understanding and validate their roles across a broader range of malignancies.

In this study, we focused on the most significantly differentially mutated genes across all virus-positive versus virus-negative cancers, which may have led to the omission of well-established somatic mutation associations in specific cancer types. For instance, the depletion of *CTNNB1* mutations in HBV-positive HCC cases, a known association[2,149], was not highlighted because the *p*-value for *CTNNB1* mutations in the combined cohort of all eligible cancer types did not meet our significance threshold. In contrast, HCV-positive HCC cases are associated with *CTNNB1* mutations[118], and grouping HBV- and HCV-positive cases together as virus-positive masked the HBV-specific trend. Similarly, *TERT* promoter mutations, well-documented in virus-positive HNSCCs[2,177], were not captured in our study because the analysis focused on exonic somatic mutations, excluding promoter regions.

Analysis of ICI clinical trials revealed that virus-positive status could represent a positive biomarker for ICI therapy response in GC and HNSCC[178]. The improved response to immunotherapy of EBV-positive GC patients compared to EBV-negative GC patients is hypothesized to be due to increased expression of *PD-L1*, potentially through activation of the NF-κB pathway by viral protein LMP2A[179]. However, the association between HPV infection and *PD-L1* expression is less clear: some studies report a link between HPV status and *PD-L1* expression[180,181], while others find no association[182,183], the latter of which is consistent with our results from the TCGA HNSCC dataset. In both GC and HNSCC, we observed a marked increase in CD8 + T cell infiltration as well as an increase in T cell receptor clonality in virus-positive tumors. The high immunogenicity of viral antigens expressed in these tumors might elicit a bigger and more clonal T cell response, which could be reactivated through ICI therapy.

While this study does not focus on epigenetic alterations, previous studies have highlighted significant differences in the epigenetic landscapes of virus-associated cancers. For example, HPV-positive CC exhibit higher promoter methylation and increased gene silencing compared to virus-negative cases[117]. Similarly, HPV-positive HNSCC is widely reported to be hypermethylated compared to HPV-negative cases[184,185], with genes such as *CDKN2A*, *RASSF1*, and *CCNA1*, which are involved in cell cycle regulation and apoptosis, frequently affected[184,186,187]. In GC, EBV-positive cancers are characterized by DNA hypermethylation[105]. These prior findings underscore the important role of epigenetic alterations in the pathogenesis of virus-associated cancers.

Our integrative analysis highlights two distinct routes to oncogenesis, in the absence and presence of viral infection (Fig. 6A, B). Further studies will be needed to understand how this model may be incorporated into the development of targeted therapies.

Methods

Experimental model and subject details

The Kaposi sarcoma patients ($n = 10$) were enrolled for study at the University of Sassari, Italy. Skin lesion tumor samples and adjacent non-neoplastic cells were surgically resected for DNA sequencing. All biological samples from Kaposi patients (tissue and blood specimens),

along with clinical information including sex (based on self-report) and age, were obtained after written consent from the patients and the sample IDs are anonymized. The study was approved by the Committee for the Ethics of the Research and Bioethics of the National Research Council (CNR n.12629). In addition, clinical and genomic data of 2293 cancer patients was obtained from 17 published studies[21,105,114–117,119–123,152–155,188]. The combined cohort consists of 788 females, 1101 males, and 414 individuals of unknown or unreported sex. The ages range from 1 year to 90 years.

Epidemiological analysis

Sex ratio in incidence rates of virus-associated and non-virus associated cancers, as well as virus-positive and virus-negative cases of virus-associated cancers, were obtained from studies listed in Supplementary Data 1 and Supplementary Data 2[23–104]. Global age-standardized incidence rates of cancers by country in 2020 were obtained from GLOBOCAN 2020 Cancer Today online portal (https://gco.iarc.fr/today/home). Attributable fraction of cancer cases for each region were obtained from de Martel et al. [3].

Virus infection status calling

Virus-infection status of patients were reported in the original studies[17,21,105,114–123,152–155,188], or obtained via cBioportal (https://www.cbioportal.org/) for TCGA cervical[117], gastric[105], and head and neck squamous cell carcinoma[119] data sets.

Single nucleotide and indel variant calling pipeline

WES data from 10 Kaposi sarcoma samples were aligned to GRCh37 using the Burrows-Wheeler aligner. Samples were pre-processed by indel realignment, duplicate removal, and base recalibration with GATK following the GATK best practices workflow. SAVI-v2[189] was used to call somatic variants. The variant list was filtered for variants with a minimum total depth 10 and maximum total depth 700 in both tumor and normal, strand bias p-value > 0.001 in tumor and normal and called as significant somatic variants by SAVI (p-value < 0.05, and confidence interval for the significance of the tumor/normal comparison > 0). Variants were excluded if they were found in an in-house supernormal created from 186 normal samples from the TCGA, if they were in the cohort supernormal constructed from variants in the ten normal samples, or if they were common SNPs found at a frequency ≥ 5% in the 1000 Genomes Project.

Mutation load analysis

The mutations in each tumor were obtained from the variant lists reported in the original studies (for previously published cases[21,105,114–117,119–123,152–155,188]), or from the mutation calling pipeline described above (for newly sequenced Kaposi sarcoma cases). Mutations in driver genes were defined as mutations that occurred within genes described as cancer-specific drivers and/or recurrently mutated genes in the original studies[21,105,114,116,117,119–122,152–155,188].

Mutation signature analysis

Mutation signatures were called from somatic variants separately for each cancer type using SigProfilerExtractor (v.1.1.24)[132], an NMF-based mutational signature extraction tool. The analysis was performed with default parameters, with the minimum and maximum number of signatures set to 1 and 10, respectively. The tool performed a de novo extraction of mutational signatures and decomposed them into COSMIC (v3.4) signatures. Suggested solutions from SigProfilerExtractor were used in our analysis, which included decomposed COSMIC signatures as well as de novo signatures in cases where the reconstruction did not achieve a cosine similarity above 0.8. Signatures were obtained from somatic variants called from whole genome sequencing data when available (Burkitt) or whole exome sequencing data (other cancers).

Structural variant signature and copy number signature analysis

Structural variant (SV) and copy number (CN) data were obtained from the International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO) data platform (Legacy ICGC 25 K Data, https://platform.icgc-argo.org/) for the Pan-Cancer Analysis of Whole Genomes (PCAWG) WGS cases, including HCC, HNSCC, CC, and GC. For BL and cHL, SV and CN data were retrieved from their respective publications[21,115]. SV and CN signatures were identified using SigProfilerExtractor (v1.1.24), following the same approach as the mutational signature analysis with default parameters.

Copy number segmentation and variant calling

Copy number segmentation of Kaposi sarcoma samples was conducted using Sequenza[190] for each pair of tumor of normal samples sequenced by whole-exome sequencing for each case. Copy number segmentation of plasmablastic lymphoma samples was performed with Oncoscan[114]. Copy number segmentation data of other cancers was obtained from the original published studies (Burkitt[21], NKTCL[154]) or cBioportal (TCGA samples[105,117–119]).

GISTIC analysis

In order to define significant regions of recurrent CNAs across all available virus-associated cancers, GISTIC 2.0[191] was applied to pooled copy number segmentation data of 1557 tumors, using a significance threshold of $q = 0.1$, a maximum segmentation threshold of 10,000, and all other parameters default via the GenePattern server (https://www.genepattern.org/) (Supplementary Data 20). GISTIC peaks of gain or amplification were counted as present in a patient if the maximum inferred tumor copy number within the wide peak limit region was > 2.3 (gain) or > 3.6 (amplification). GISTIC peaks of deletion were counted as present in a patient if the minimum inferred tumor copy number within the wide peak limit region was < 1.7 (heterozygous loss) or < 0.8 (homozygous loss). Arm and whole-chromosome level CNAs were defined as lesions of the same type (i.e., gain or loss) that covered > 75% of the chromosome arm or chromosome, respectively.

Analysis of recurrently mutated genes in virus-associated cancers

Nonsynonymous mutations in protein-coding genes were analyzed in 1971 patients with available DNA sequencing data. A cancer type or sequencing cohort was included in the combined analysis of a gene only if the gene was covered by the respective sequencing panel and was mutated in at least one sample from that cohort. To minimize noise from hypermutated samples, only cases with fewer than 300 mutations were included across all cancer types.

Significant genes were identified based on an odds ratio (OR) greater than 1 for mutations in virus-positive versus virus-negative cases in the combined dataset, a Benjamini-Hochberg (BH) corrected p-value below 0.005, and recurrence in at least two virus-positive or virus-negative cases across at least three unique cancer types. In addition, the final combined direction of mutation prevalence had to be consistently supported by at least half of the included cancer types.

Significant copy number alterations were identified using the pan-cancer GISTIC analysis (see "GISTIC Analysis" section) and required an odds ratio greater than 1 for mutations in virus-positive versus virus-negative cases, with a BH corrected p-value below 0.0001.

Analysis of immunotherapy trials

The comparative analysis of response to immunotherapy was performed using data from ClinicalTrials.gov and did not include any unpublished clinical trial data. Nine checkpoint inhibitors targeting either PD-1 or PD-L1 were included: Nivolumab, Pembrolizumab, Cemiplimab, Atezolizumab, Avelumab, Durvalumab, Camrelizumab, Sintilimab, Toripalimab. On September 22nd 2022, trials were collected using the following query: "((NOT NOTEXT) [CITATIONS]) AND (< ICI

drug name 1 > OR < ICI drug name 2 > OR ...)", where ICI drug names correspond to the ones listed above as well as any known synonyms (e.g., Nivolumab: Opdivo, ONO-4538, BMS-936558, MDX1106). Only single-arm, interventional studies where the RECIST1.1 objective response rate was available specifically for PD-1/PD-L1 inhibitors (with no other non-ICI combination therapies) were included. Whenever possible, response stratified by virus status, PD-L1 expression and tumor mutational burden was collected. "PD-L1 positive" refers to patients that were classified as PD-L1 positive in the original study (most often, ≥1% of all cells). "Tumor mutational burden (TMB) high" refers to patients classified as TMB high in the original study (with different cut-offs used depending on the study and/or tumor type). Due to the lack of individual patient information, each biomarker for response was evaluated independently (Fisher's exact test).

Immune infiltration and T cell receptor analysis

TCGA RSEM expression data was obtained from cBioportal. TCGA deconvolution results were downloaded from the TIMER2.0 website (http://timer.cistrome.org/)[164]. For the T cell receptor (TCR) clonality analysis, we focused on the complimentary determining region 3 (CDR3) of both the α and β chains, which is the most highly variable sequence of the TCR and the most important for determining antigen affinity[192]. We obtained the previously published TRUST4 TCR calls from the authors, and we applied the two diversity metrics from the original study[171], namely the number of clonotypes per thousand CDR3 reads (CPK), and clonality, defined as (1 − normalized Shannon entropy). All comparisons were performed using the Wilcoxon rank sum test / Mann-Whitney U test.

Statistics

Analyses of the significance of mutation counts and frequencies were performed using the two-sided Wilcoxon rank sum test (Mann-Whitney U test) and two-sided Fisher's exact test, respectively. Odds ratios of mutation by virus status were computed with Haldane-Anscombe correction when applicable. The 95% confidence interval of odds ratios were estimated using the normal approximation (Wald). Multiple hypothesis corrections were applied using the Benjamini-Hochberg Procedure and reported as q-values. Combined *p*-values were computed using Fisher's method (unweighted, based on one-sided *p*-values) and wFisher method (weighted by sample size, based on two-sided *p*-values) as implemented in the metapro R package[193]. No data were excluded from the statistical tests, except in the analysis of recurrently mutated genes in virus-associated cancers, where only cases with fewer than 300 mutations were included across all cancer types to minimize noise from hypermutated samples.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

WES raw data (FASTQ and BAM files) of Kaposi sarcoma generated in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI database under accession code PRJEB76508. The TCGA (HNSCC, CC and GC) cohort's clinical data, mutation and copy number alteration calls are available at cBioportal (HNSCC:https://www.cbioportal.org/study/summary?id=hnsc_tcga, CC:https://www.cbioportal.org/study/summary?id=cesc_tcga_pan_can_atlas_2018, GC:https://www.cbioportal.org/study/summary?id=stad_tcga), raw and normalized chromosomal instability (CIN) signature activities are available within the supplementary information of Drews et al.'s publication[139]. SV and CN data of PCAWG WGS cases are available from the International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO) data platform (Legacy ICGC 25 K Data, https://docs.icgc-argo.org/docs/data-access/icgc-25k-data#open-release-data-object-bucket-details; SV: s3://icgc25k-open/PCAWG/consensus_sv/; CN: s3://icgc25k-open/PCAWG/consensus_cnv/). Clinical and genomic data including SNV, CN, and/or SV calls of other cancer types are available within the supplementary information of their published studies[21,105,114–117,119–123,152–155,188]. EBV status of the Hodgkin lymphoma cohort from Alig et al. [123] has been acquired from the authors directly. Source data are provided in this paper.

Code availability

All analyses were performed using publicly available software: for alignment of Kaposi sarcoma sample WES data to GRCh37, Burrows-Wheeler aligner v.0.7.17 (https://github.com/lh3/bwa); for calling somatic variants, SAVI version 2 (https://github.com/WinterLi1993/SAVI); for calling mutation signatures from somatic variants, SigProfilerExtractor v.1.1.24 (https://github.com/AlexandrovLab/SigProfilerExtractor); for copy number segmentation of Kaposi sarcoma samples, Sequenza (https://sequenzatools.bitbucket.io/); for defining significant regions of recurrent CNAs, GISTIC version 2.0 (https://broadinstitute.github.io/gistic2/); for combining *p*-values using the weighted Fisher method (wFisher), metapro (http://github.com/unistbig/metapro). For general coding, R (version 4.4.1) and Python (version 3.9.21) were used. Detailed information on the software used is also provided in the respective sections of the Methods.

References

1. van Elsland, D. & Neefjes, J. Bacterial infections and cancer. *EMBO Rep.* **19**, e46632 (2018).
2. Zapatka, M. et al. The landscape of viral associations in human cancers. *Nat. Genet.* **52**, 320–330 (2020).
3. de Martel, C., Georges, D., Bray, F., Ferlay, J. & Clifford, G. M. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob. Health* **8**, e180–e190 (2020).
4. Schrama, D. et al. Merkel cell polyomavirus status is not associated with clinical course of Merkel cell carcinoma. *J. Invest. Dermatol.* **131**, 1631–1638 (2011).
5. White, M. K., Pagano, J. S. & Khalili, K. Viruses and human cancers: a long road of discovery of molecular paradigms. *Clin. Microbiol. Rev.* **27**, 463–481 (2014).
6. Nault, J. C. et al. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* **47**, 1187 (2015).
7. Cobbs, C. S. et al. Human cytomegalovirus infection and expression in human malignant glioma. *Cancer Res.* **62**, 3347–3350 (2002).
8. Mundo, L. et al. Frequent traces of EBV infection in Hodgkin and non-Hodgkin lymphomas classified as EBV-negative by routine methods: expanding the landscape of EBV-related lymphomas. *Mod. Pathol.* **33**, 2407–2421 (2020).
9. McLaughlin-Drubin, M. E. & Munger, K. Viruses associated with human cancer. *Biochim. Biophys. Acta* **1782**, 127–150 (2008).
10. Bakkalci, D. et al. Risk factors for Epstein Barr virus-associated cancers: a systematic review, critical appraisal, and mapping of the epidemiological evidence. *J. Glob. Health* **10**, 010405 (2020).
11. Ananthakrishnan, A., Gogineni, V. & Saeian, K. Epidemiology of primary and secondary liver cancers. *Semin Interv. Radio.* **23**, 47–63 (2006).
12. Byun, J. M. et al. Persistent HPV-16 infection leads to recurrence of high-grade cervical intraepithelial neoplasia. *Medicine* **97**, e13606 (2018).
13. Pal, A. & Kundu, R. Human papillomavirus E6 and E7: The cervical cancer hallmarks and targets for therapy. *Front. Microbiol.* **10**, 3116 (2019).
14. Cousins, E. & Nicholas, J. Molecular biology of human herpesvirus 8: novel functions and virus-host interactions implicated in viral pathogenesis and replication. *Recent Results Cancer Res.* **193**, 227–268 (2014).

15. Shechter, O., Sausen, D. G., Gallo, E. S., Dahari, H. & Borenstein, R. Epstein-Barr virus (EBV) epithelial associated malignancies: Exploring pathologies and current treatments. *Int. J. Mol. Sci.* **23**, https://doi.org/10.3390/ijms232214389 (2022).
16. Yang, J., Liu, Z., Zeng, B., Hu, G. & Gan, R. Epstein-Barr virus-associated gastric cancer: A distinct subtype. *Cancer Lett.* **495**, 191–199 (2020).
17. Chesson, H. W., Dunne, E. F., Hariri, S. & Markowitz, L. E. The estimated lifetime probability of acquiring human papillomavirus in the United States. *Sex. Transm. Dis.* **41**, 660–664 (2014).
18. Amber, K., McLeod, M. P. & Nouri, K. The Merkel cell polyomavirus and its involvement in Merkel cell carcinoma. *Dermatol Surg.* **39**, 232–238 (2013).
19. Vockerodt, M., Cader, F. Z., Shannon-Lowe, C. & Murray, P. Epstein-Barr virus and the origin of Hodgkin lymphoma. *Chin. J. Cancer* **33**, 591–597 (2014).
20. Massini, G., Siemer, D. & Hohaus, S. EBV in Hodgkin lymphoma. *Mediterr. J. Hematol. Infect. Dis.* **1**, e2009013 (2009).
21. Grande, B. M. et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood* **133**, 1313–1324 (2019).
22. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN Estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
23. Abdulamir, A., Hafidh, R., Abdulmuhaimen, N., Abubakar, F. & Abbas, K. The distinctive profile of risk factors of nasopharyngeal carcinoma in comparison with other head and neck cancer types. *BMC Public Health* **8**, 1–16 (2008).
24. Adham, M. et al. Nasopharyngeal carcinoma in Indonesia: epidemiology, incidence, signs, and symptoms at presentation. *Chin. J. Cancer* **31**, 185 (2012).
25. Agelli, M. & Clegg, L. X. Epidemiology of primary Merkel cell carcinoma in the United States. *J. Am. Acad. Dermatol.* **49**, 832–841 (2003).
26. Ajila, V., Shetty, H., Babu, S., Shetty, V. & Hegde, S. Human papilloma virus associated squamous cell carcinoma of the head and neck. *J. Sex. Transm. Dis.* **2015**, 791024 (2015).
27. Aka, P. et al. Incidence and trends in Burkitt lymphoma in northern Tanzania from 2000 to 2009. *Pediatr. Blood Cancer* **59**, 1234–1238 (2012).
28. Akhtar, S., Oza, K. K. & Wright, J. Merkel cell carcinoma: report of 10 cases and review of the literature. *J. Am. Acad. Dermatol.* **43**, 755–767 (2000).
29. Alipov, G. et al. Epstein-Barr virus-associated gastric carcinoma in Kazakhstan. *World J. Gastroenterol.* **11**, 27 (2005).
30. Andres, C., Belloni, B., Puchta, U., Sander, C. A. & Flaig, M. J. Prevalence of MCPyV in Merkel cell carcinoma and non-MCC tumors. *J. Cutan. Pathol.* **37**, 28–34 (2010).
31. Bassig, B. A. et al. Subtype-specific incidence rates of lymphoid malignancies in Hong Kong compared to the United States, 2001-2010. *Cancer Epidemiol.* **42**, 15–23 (2016).
32. Bosch, F. X., Ribes, J., Cléries, R. & Díaz, M. Epidemiology of hepatocellular carcinoma. *Clin. Liver Dis.* **9**, 191–211 (2005).
33. Carrascal, E. et al. Epstein-Barr virus-associated gastric carcinoma in Cali, Colombia. *Oncol. Rep.* **10**, 1059–1062 (2003).
34. Castillo, J. J., Bibas, M. & Miranda, R. N. The biology and treatment of plasmablastic lymphoma. *Blood J. Am. Soc. Hematol.* **125**, 2323–2330 (2015).
35. Chang, M.-H. et al. Hepatitis B vaccination and hepatocellular carcinoma rates in boys and girls. *Jama* **284**, 3040–3042 (2000).
36. Chang, M. S., Lee, H. S., Kim, C. W., Kim, Y. I. & Kim, W. H. Clinicopathologic characteristics of Epstein-Barr virus-incorporated gastric cancers in Korea. *Pathol. Res. Pract.* **197**, 395–400 (2001).
37. Chen, W. et al. Esophageal cancer incidence and mortality in China, 2009. *J. Thorac. Dis.* **5**, 19 (2013).
38. Chong, J. M. et al. Expression of CD44 variants in gastric carcinoma with or without Epstein-Barr virus. *Int. J. Cancer* **74**, 450–454 (1997).
39. Claviez, A. et al. Impact of latent Epstein-Barr virus infection on outcome in children and adolescents with Hodgkin's lymphoma. *J. Clin. Oncol.* **23**, 4048–4056 (2005).
40. Conte, S. et al. Population-based study detailing cutaneous melanoma incidence and mortality trends in Canada. *Front. Med.* **9**, 830254 (2022).
41. Corvalan, A. et al. Epstein-Barr virus in gastric carcinoma is associated with location in the cardia and with a diffuse histology: a study in one area of Chile. *Int. J. Cancer* **94**, 527–530 (2001).
42. Czopek, J. P. et al. EBV-positive gastric carcinomas in Poland. *Pol. J. Pathol. Off. J. Pol. Soc. Pathol.* **54**, 123–128 (2003).
43. Deo, S. et al. Colorectal cancers in low-and middle-income countries—demographic pattern and clinical profile of 970 patients treated at a tertiary care cancer center in India. *JCO Glob. Oncol.* **7**, 1110–1115 (2021).
44. Diepstra, A. et al. Latent Epstein-Barr virus infection of tumor cells in classical Hodgkin's lymphoma predicts adverse outcome in older adult patients. *J. Clin. Oncol.* **27**, 3815–3821 (2009).
45. Divaris, K. et al. Oral health and risk for head and neck squamous cell carcinoma: the Carolina Head and Neck Cancer Study. *Cancer Causes Control* **21**, 567–575 (2010).
46. Enblad, G., Sandvej, K., Sundstrom, C., Pallesen, G. & Glimelius, B. Epstein-Barr virus distribution in Hodgkin's disease in an unselected Swedish population. *Acta Oncol.* **38**, 425–429 (1999).
47. Fedder, M. & Gonzalez, M. F. Nasopharyngeal carcinoma. Brief review. *Am. J. Med.* **79**, 365–369 (1985).
48. Galetsky, S. A. et al. Epstein-Barr-virus-associated gastric cancer in Russia. *Int. J. Cancer* **73**, 786–789 (1997).
49. Gulley, M. L., Pulitzer, D. R., Eagan, P. A. & Schneider, B. G. Epstein-Barr virus infection is an early event in gastric carcinogenesis and is independent of bcl-2 expression and p53 accumulation. *Hum. Pathol.* **27**, 20–27 (1996).
50. Hao, Z. et al. The Epstein-Barr virus-associated gastric carcinoma in Southern and Northern China. *Oncol. Rep.* **9**, 1293–1298 (2002).
51. Harn, H.-J. et al. Epstein-Barr virus-associated gastric adenocarcinoma in Taiwan. *Hum. Pathol.* **26**, 267–271 (1995).
52. Herrera-Goepfert, R. et al. Epstein-Barr virus-associated gastric carcinoma: Evidence of age-dependence among a Mexican population. *World J. Gastroenterol.* **11**, 6096 (2005).
53. Hjalgrim, H. et al. *The Epidemiology of EBV and Its Association with Malignant Disease* (Cambridge University Press, 2007).
54. Hsu, J. L. & Glaser, S. L. Epstein–Barr virus-associated malignancies: epidemiologic patterns and etiologic implications. *Crit. Rev. Oncol. Hematol.* **34**, 27–53 (2000).
55. Iscovich, J., Boffetta, P., Franceschi, S., Azizi, E. & Sarid, R. Classic Kaposi sarcoma: epidemiology and risk factors. *Cancer* **88**, 500–517 (2000).
56. Jarrett, R. et al. The Scotland and Newcastle epidemiological study of Hodgkin's disease: impact of histopathological review and EBV status on incidence estimates. *J. Clin. Pathol.* **56**, 811–816 (2003).
57. Johansson, S. L. & Cohen, S. M. Epidemiology and etiology of bladder cancer. *Semin. Surg. Oncol.* **13**, 291–298 (1997).
58. Kang, G. H. et al. Epstein-barr virus-positive gastric carcinoma demonstrates frequent aberrant methylation of multiple genes and constitutes CpG island methylator phenotype-positive gastric carcinoma. *Am. J. Pathol.* **160**, 787–794 (2002).
59. Karim, N. & Pallesen, G. Epstein-Barr virus (EBV) and gastric carcinoma in Malaysian patients. *Malays. J. Pathol.* **25**, 45–47 (2003).
60. Kassem, A. et al. Frequent detection of Merkel cell polyomavirus in human Merkel cell carcinomas and identification of a unique deletion in the VP1 gene. *Cancer Res.* **68**, 5009–5013 (2008).

61. Keegan, T. H. et al. Epstein-Barr virus as a marker of survival after Hodgkin's lymphoma: a population-based study. *J. Clin. Oncol.* **23**, 7604–7613 (2005).
62. Koriyama, C. et al. Epstein-Barr virus-associated gastric carcinoma in Japanese Brazilians and non-Japanese Brazilians in Sao Paulo. *Jpn. J. Cancer Res.* **92**, 911–917 (2001).
63. Kume, T. et al. Low rate of apoptosis and overexpression of bcl-2 in Epstein–Barr virus-associated gastric carcinoma. *Histopathology* **34**, 502–509 (1999).
64. Lopes, L. et al. Epstein-Barr virus infection and gastric carcinoma in São Paulo State, Brazil. *Braz. J. Med. Biol. Res.* **37**, 1707–1712 (2004).
65. Lu, S. N. et al. Secular trends and geographic variations of hepatitis B virus and hepatitis C virus-associated hepatocellular carcinoma in Taiwan. *Int. J. Cancer* **119**, 1946–1952 (2006).
66. Mbulaiteye, S. M. et al. Trimodal age-specific incidence patterns for Burkitt lymphoma in the United States, 1973–2005. *Int. J. Cancer* **126**, 1732–1739 (2010).
67. McGlynn, K. A. & London, W. T. Epidemiology and natural history of hepatocellular carcinoma. *Best. Pract. Res. Clin. Gastroenterol.* **19**, 3–23 (2005).
68. McNeil, D. E., Coté, T. R., Clegg, L. & Mauer, A. SEER update of incidence and trends in pediatric malignancies: acute lymphoblastic leukemia. *Med. Pediatr. Oncol.* **39**, 554–557 (2002).
69. Mimi, C. Y. & Yuan, J.-M. Epidemiology of nasopharyngeal carcinoma. *Semin. Cancer Biol.* **12**, 421–429 (2002).
70. Molica, S. Sex differences in incidence and outcome of chronic lymphocytic leukemia patients. *Leuk. lymphoma* **47**, 1477–1480 (2006).
71. Moritani, S., Kushima, R., Sugihara, H. & Hattori, T. Phenotypic characteristics of Epstein-Barr-virus-associated gastric carcinomas. *J. Cancer Res. Clin. Oncol.* **122**, 750–756 (1996).
72. Murphy, G., Pfeiffer, R., Camargo, M. C. & Rabkin, C. S. Meta-analysis shows that prevalence of Epstein–Barr virus-positive gastric cancer differs based on sex and anatomic location. *Gastroenterology* **137**, 824–833 (2009).
73. Nogueira, C. et al. Prevalence and characteristics of Epstein–Barr virus-associated gastric carcinomas in Portugal. *Infect. Agents Cancer* **12**, 1–8 (2017).
74. Ogwang, M. D., Bhatia, K., Biggar, R. J. & Mbulaiteye, S. M. Incidence and geographic distribution of endemic Burkitt lymphoma in northern Uganda revisited. *Int. J. Cancer* **123**, 2658–2663 (2008).
75. Ojima, H., Fukuda, T., Nakajima, T., Takenoshita, S. & Nagamachi, Y. Discrepancy between clinical and pathological lymph node evaluation in Epstein-Barr virus-associated gastric cancers. *Anticancer Res.* **16**, 3081–3084 (1996).
76. Pallagani, L. et al. Epidemiology and clinicopathological profile of renal cell carcinoma: a review from tertiary care referral centre. *J. Kidney Cancer VHL* **8**, 1 (2021).
77. Qiu, K. et al. Epstein-Barr virus in gastric carcinoma in Suzhou, China and Osaka, Japan: Association with clinico-pathologic factors and HLA-subtype. *Int. J. Cancer* **71**, 155–158 (1997).
78. Ragin, C., Modugno, F. & Gollin, S. The epidemiology and risk factors of head and neck cancer: a focus on human papillomavirus. *J. Dent. Res.* **86**, 104–114 (2007).
79. Rahbari, R., Zhang, L. & Kebebew, E. Thyroid cancer gender disparity. *Future Oncol.* **6**, 1771–1779 (2010).
80. Randi, G., Franceschi, S. & La Vecchia, C. Gallbladder cancer worldwide: geographical distribution and risk factors. *Int. J. Cancer* **118**, 1591–1602 (2006).
81. Rawla, P. & Barsouk, A. Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol.* **14**, 26–38 (2019).
82. Rowlands, D. et al. Epstein-Barr virus and carcinomas: rare association of the virus with gastric adenocarcinomas. *Br. J. Cancer* **68**, 1014–1019 (1993).
83. Sakuma, K. et al. Cancer risk to the gastric corpus in Japanese, its correlation with interleukin-1β gene polymorphism (+ 3953* T) and Epstein-Barr virus infection. *Int. J. Cancer* **115**, 93–97 (2005).
84. Sellam, F. et al. Delayed diagnosis of pancreatic cancer reported as more common in a population of North African young adults. *J. Gastrointest. Oncol.* **6**, 505 (2015).
85. Shibata, D. & Weiss, L. Epstein-Barr virus-associated gastric adenocarcinoma. *Am. J. Pathol.* **140**, 769 (1992).
86. Shin, W. S. et al. Epstein-Barr virus-associated gastric adenocarcinomas among Koreans. *Am. J. Clin. Pathol.* **105**, 174–181 (1996).
87. Souza, E. M. et al. Impact of Epstein–Barr virus in the clinical evolution of patients with classical Hodgkin's lymphoma in Brazil. *Hematol. Oncol.* **28**, 137–141 (2010).
88. Takano, Y. et al. The role of the Epstein-Barr virus in the oncogenesis of EBV (+) gastric carcinomas. *Virchows Arch.* **434**, 17–22 (1999).
89. Tamási, L. et al. Age and gender specific Lung cancer incidence and mortality in hungary: Trends from 2011 through 2016. *Pathol. Oncol. Res.* **88**, 17–22 (2021).
90. Tamimi, A. F. & Juweid, M. *Epidemiology and Outcome of Glioblastoma*. (2017).
91. Tavakoli, A. et al. Association between Epstein-Barr virus infection and gastric cancer: a systematic review and meta-analysis. *BMC Cancer* **20**, 1–14 (2020).
92. Tokunaga, M. et al. Epstein-Barr virus in gastric carcinoma. *Am. J. Pathol.* **143**, 1250 (1993).
93. van Beek, J. et al. EBV-positive gastric adenocarcinomas: a distinct clinicopathologic entity with a low frequency of lymph node involvement. *J. Clin. Oncol.* **22**, 664–670 (2004).
94. Venook, A. P., Papandreou, C., Furuse, J. & Ladrón de Guevara, L. The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective. *Oncologist* **15**, 5–13 (2010).
95. Villano, J., Koshy, M., Shaikh, H., Dolecek, T. & McCarthy, B. Age, gender, and racial differences in incidence and survival in primary CNS lymphoma. *Br. J. Cancer* **105**, 1414–1418 (2011).
96. Wang, X. M. et al. Clinical analysis of 1629 newly diagnosed malignant lymphomas in current residents of Sichuan province, China. *Hematol. Oncol.* **34**, 193–199 (2016).
97. Wang, Y. et al. Quantitative methylation analysis reveals gender and age differences in p16 INK 4a hypermethylation in hepatitis B virus-related hepatocellular carcinoma. *Liver Int.* **32**, 420–428 (2012).
98. Wei, K.-R. et al. Nasopharyngeal carcinoma incidence and mortality in China in 2010. *Chin. J. cancer* **33**, 381 (2014).
99. Wu, M. S. et al. Epstein–Barr virus—associated gastric carcinomas: relation to H. pylori infection and genetic alterations. *Gastroenterology* **118**, 1031–1038 (2000).
100. Wu, S., Han, J., Li, W.-Q., Li, T. & Qureshi, A. A. Basal-cell carcinoma incidence and associated risk factors in US women and men. *Am. J. Epidemiol.* **178**, 890–897 (2013).
101. Yanai, H. et al. Endoscopic and pathologic features of Epstein-Barr virus–associated gastric carcinoma. *Gastrointest. Endosc.* **45**, 236–242 (1997).
102. Yoshiwara, E. et al. Epstein-Barr virus-associated gastric carcinoma in Lima, Peru. *J. Exp. Clin. Cancer Res.* **24**, 49–54 (2005).
103. Zhou, L. et al. Global, regional, and national burden of Hodgkin lymphoma from 1990 to 2017: estimates from the 2017 Global Burden of Disease study. *J. Hematol. Oncol.* **12**, 1–13 (2019).
104. Zhu, Z.-Z. et al. Sex-related differences in DNA copy number alterations in hepatitis B virus-associated hepatocellular carcinoma. *Asian Pac. J. Cancer Prev.* **13**, 225–229 (2012).
105. The Cancer Genome Atlas Research Network Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
106. Murphy, G., Pfeiffer, R., Camargo, M. C. & Rabkin, C. S. Meta-analysis shows that prevalence of Epstein-Barr virus-positive

gastric cancer differs based on sex and anatomic location. *Gastroenterology* **137**, 824–833 (2009).
107. Lee, J. H., Kim, Y., Choi, J. W. & Kim, Y. S. Prevalence and prognostic significance of Epstein-Barr virus infection in classical Hodgkin's lymphoma: a meta-analysis. *Arch. Med. Res.* **45**, 417–431 (2014).
108. Dozzo, M. et al. Burkitt lymphoma in adolescents and young adults: management challenges. *Adolesc. Health, Med. Therapeutics* **8**, 11–29 (2016).
109. Rismiller, K. & Knackstedt, T. J. Aggressive digital papillary adenocarcinoma: Population-based analysis of incidence, demographics, treatment, and outcomes. *Dermatol Surg.* **44**, 911–917 (2018).
110. Li, X., Fasano, R., Wang, E., Yao, K.-T. & Marincola, F. M. HLA associations with nasopharyngeal carcinoma. *Curr. Mol. Med.* **9**, 751–765 (2009).
111. Huang, X. et al. HLA-A* 02: 07 is a protective allele for EBV negative and a susceptibility allele for EBV positive classical Hodgkin lymphoma in China. *PLoS ONE* **7**, e31865 (2012).
112. Schottenfeld, D. & Beebe-Dimmer, J. The cancer burden attributable to biologic agents. *Ann. Epidemiol.* **25**, 183–187 (2015).
113. Jemal, A., Center, M. M., DeSantis, C. & Ward, E. M. Global patterns of cancer incidence and mortality rates and trendsGlobal patterns of cancer. *Cancer Epidemiol. Biomark. Prev.* **19**, 1893–1907 (2010).
114. Liu, Z. et al. Genomic characterization of HIV-associated plasmablastic lymphoma identifies pervasive mutations in the JAK-STAT pathway. *Blood Cancer Discov.* **1**, 112–125 (2020).
115. Maura, F. et al. Molecular evolution of classic hodgkin lymphoma revealed through whole-genome sequencing of hodgkin and reed sternberg cells. *Blood Cancer Discov.* **4**, 208–227 (2023).
116. Tiacci, E. et al. Pervasive mutations of JAK-STAT pathway genes in classical Hodgkin lymphoma. *Blood* **131**, 2454–2465 (2018).
117. The Cancer Genome Atlas Research Network Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378–384 (2017).
118. The Cancer Genome Atlas Research Network Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341 (2017).
119. The Cancer Genome Atlas Network Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
120. Ramis-Zaldivar, J. E. et al. MAPK and JAK-STAT pathways dysregulation in plasmablastic lymphoma. *Haematologica* **106**, 2682–2693 (2021).
121. Gandhi, M. K. et al. EBV-associated primary CNS lymphoma occurring after immunosuppression is a distinct immunobiological entity. *Blood* **137**, 1468–1477 (2021).
122. Starrett, G. J. et al. Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma. *Genome Med.* **12**, 30 (2020).
123. Alig, S. K. et al. Distinct Hodgkin lymphoma subtypes defined by noninvasive genomic profiling. *Nature* **625**, 778–787 (2024).
124. Harms, P. W. et al. The distinctive mutational spectra of polyomavirus-negative merkel cell carcinoma. *Cancer Res.* **75**, 3720–3727 (2015).
125. Kaulen, L. D. et al. Integrated genetic analyses of immunodeficiency-associated Epstein-Barr virus- (EBV) positive primary CNS lymphomas. *Acta Neuropathol.* **146**, 499–514 (2023).
126. Ghasemi, F. et al. Mutational analysis of head and neck squamous cell carcinoma stratified by smoking status. *JCI Insight* **4**, https://doi.org/10.1172/jci.insight.123443 (2019).
127. Panda, A. et al. Immune activation and benefit From avelumab in EBV-positive gastric cancer. *J. Natl. Cancer Inst.* **110**, 316–320 (2018).
128. Zhao, X., Fan, X., Lin, X., Guo, B. & Yu, Y. Deciphering age-specific molecular features in cervical cancer and constructing an angio-immune prognostic model. *Medicine* **103**, e37717 (2024).
129. Stransky, N. et al. The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).
130. Chatfield-Reed, K., Gui, S., O'Neill, W. Q., Teknos, T. N. & Pan, Q. HPV33+ HNSCC is associated with poor prognosis and has unique genomic and immunologic landscapes. *Oral. Oncol.* **100**, 104488 (2020).
131. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
132. Islam, S. M. A. et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genom.* **2**, https://doi.org/10.1016/j.xgen.2022.100179 (2022).
133. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. & Fenton, T. R. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep.* **7**, 1833–1841 (2014).
134. Warren, C. J., Westrich, J. A., Van Doorslaer, K. & Pyeon, D. Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression. *Viruses* **9**, 233 (2017).
135. Gillison, M. L. et al. Human papillomavirus and the landscape of secondary genetic alterations in oral cancers. *Genome Res.* **29**, 1–17 (2019).
136. Chu, Y.-J. et al. Aflatoxin B1 exposure increases the risk of hepatocellular carcinoma associated with hepatitis C virus infection or alcohol consumption. *Eur. J. Cancer* **94**, 37–46 (2018).
137. Chu, Y. J. et al. Aflatoxin B1 exposure increases the risk of cirrhosis and hepatocellular carcinoma in chronic hepatitis B virus carriers. *Int. J. Cancer* **141**, 711–720 (2017).
138. Schulze, K. et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.***47**, 505–511 (2015).
139. Drews, R. M. et al. A pan-cancer compendium of chromosomal instability. *Nature* **606**, 976–983 (2022).
140. Mallick, S., Choi, Y., Taylor, A. M. & Cosper, P. F. Human papillomavirus-induced chromosomal instability and aneuploidy in squamous cell cancers. *Viruses* **16**, https://doi.org/10.3390/v16040501 (2024).
141. Seiwert, T. Y. et al. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin. Cancer Res.* **21**, 632–641 (2015).
142. López, C. et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. *Nat. Commun.* **10**, https://doi.org/10.1038/s41467-019-08578-3 (2019).
143. Abate, F. et al. Distinct viral and mutational spectrum of endemic Burkitt lymphoma. *PLOS Pathog.* **11**, e1005158 (2015).
144. Zhou, P. X. et al. Sporadic and endemic Burkitt lymphoma have frequent FOXO1 mutations but distinct hotspots in the AKT recognition motif. *Blood Adv.* **3**, 2118–2127 (2019).
145. Kaymaz, Y. et al. Comprehensive transcriptome and mutational profiling of endemic Burkitt lymphoma reveals EBV type-specific differences. *Mol. Cancer Res.* **15**, 563–576 (2017).
146. Gong, C. et al. Sequential inverse dysregulation of the RNA helicases DDX3X and DDX3Y facilitates MYC-driven lymphomagenesis. *Mol. Cell* **81**, 4059–4075 (2021).
147. Garcia-Reyero, J. et al. Genetic lesions in MYC and STAT3 drive oncogenic transcription factor overexpression in plasmablastic lymphoma. *Haematologica* **106**, 1120–1128 (2021).
148. Li, J. S. Z. et al. Chromosomal fragile site breakage by EBV-encoded EBNA1 at clustered repeats. *Nature* **616**, 504–509 (2023).

149. Amaddeo, G. et al. Integration of tumour and viral genomic characterisations in HBV-related hepatocellular carcinomas. *Gut* **64**, 820 (2015).
150. Weniger, M. A. & Kuppers, R. Molecular biology of Hodgkin lymphoma. *Leukemia* **35**, 968–981 (2021).
151. Rocak, S. & Linder, P. DEAD-box proteins: the driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell Biol.* **5**, 232–241 (2004).
152. Dufva, O. et al. Aggressive natural killer-cell leukemia mutational landscape and drug profiling highlight JAK-STAT signaling as therapeutic target. *Nat. Commun.* **9**, 1567 (2018).
153. Kataoka, K. et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat. Genet.* **47**, 1304–1315 (2015).
154. Xiong, J. et al. Genomic and transcriptomic characterization of natural killer T cell lymphoma. *Cancer Cell* **37**, 403–419 (2020).
155. Zhang, L. et al. Genomic analysis of nasopharyngeal carcinoma reveals TME-based subtypes. *Mol. Cancer Res.* **15**, 1722–1732 (2017).
156. Mo, J. et al. DDX3X: structure, physiologic functions and cancer. *Mol. Cancer* **20**, 38 (2021).
157. Thomas, N. et al. Genetic subgroups inform on pathobiology in adult and pediatric Burkitt lymphoma. *Blood* **141**, 904–916 (2022).
158. Formica, V. et al. A systematic review and meta-analysis of PD-1/PD-L1 inhibitors in specific patient subgroups with advanced gastro-oesophageal junction and gastric adenocarcinoma. *Crit. Rev. Oncol. Hematol.* **157**, 103173 (2021).
159. De Meulenaere, A. et al. Turning the tide: Clinical utility of PD-L1 expression in squamous cell carcinoma of the head and neck. *Oral. Oncol.* **70**, 34–42 (2017).
160. Lipson, E. J. et al. PD-L1 expression in the Merkel cell carcinoma microenvironment: association with inflammation, Merkel cell polyomavirus and overall survival. *Cancer Immunol. Res* **1**, 54–63 (2013).
161. Derks, S. et al. Abundant PD-L1 expression in Epstein-Barr virus-infected gastric cancers. *Oncotarget* **7**, 32925–32932 (2016).
162. Yang, W. F., Wong, M. C. M., Thomson, P. J., Li, K. Y. & Su, Y. X. The prognostic role of PD-L1 expression for survival in head and neck squamous cell carcinoma: A systematic review and meta-analysis. *Oral. Oncol.* **86**, 81–90 (2018).
163. Li, B. et al. Anti-PD-1/PD-L1 Blockade immunotherapy employed in treating hepatitis B virus infection-related advanced hepatocellular carcinoma: A literature review. *Front. Immunol.* **11**, 1037 (2020).
164. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514 (2020).
165. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
166. Li, B. et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* **17**, 174 (2016).
167. Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **11**, 34 (2019).
168. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
169. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
170. Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).
171. Song, L. et al. Comprehensive characterizations of immune receptor repertoire in tumors and cancer immunotherapy studies. *Cancer Immunol. Res.* **10**, 788–799 (2022).
172. Fish, E. N. The X-files in immunity: sex-based differences predispose immune responses. *Nat. Rev. Immunol.* **8**, 737–744 (2008).
173. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).
174. Tornesello, M. L. et al. Mutations in TP53, CTNNB1 and PIK3CA genes in hepatocellular carcinoma associated with hepatitis B and hepatitis C virus infections. *Genomics* **102**, 74–83 (2013).
175. Riva, V. & Maga, G. From the magic bullet to the magic target: exploiting the diverse roles of DDX3X in viral infections and tumorigenesis. *Future Med. Chem.* **11**, 1357–1381 (2019).
176. He, Y. et al. A double-edged function of DDX3, as an oncogene or tumor suppressor, in cancer progression (Review). *Oncol. Rep.* **39**, 883–892 (2018).
177. Mork, J. et al. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* **344**, 1125–1131 (2001).
178. Özay, Z. İ., Sütcüoğlu, O., Özdemir, N. & Yazıcı, O. Review of immunotherapy efficacy in virus-associated cancers. *EJMO* **6**, https://doi.org/10.14744/ejmo.2022.86807 (2022).
179. Miliotis, C. N. & Slack, F. J. Multi-layered control of PD-L1 expression in Epstein-Barr virus-associated gastric cancer. *J. Cancer Metastasis Treat.* **6** https://doi.org/10.20517/2394-4722.2020.12 (2020).
180. Ukpo, O. C., Thorstad, W. L. & Lewis, J. S. Jr. B7-H1 expression model for immune evasion in human papillomavirus-related oropharyngeal squamous cell carcinoma. *Head. Neck Pathol.* **7**, 113–121 (2013).
181. Lyford-Pike, S. et al. Evidence for a role of the PD-1:PD-L1 pathway in immune resistance of HPV-associated head and neck squamous cell carcinoma. *Cancer Res.* **73**, 1733–1741 (2013).
182. Kim, H. S. et al. Association between PD-L1 and HPV status and the prognostic value of PD-L1 in oropharyngeal squamous cell carcinoma. *Cancer Res. Treat.* **48**, 527–536 (2016).
183. Badoual, C. et al. PD-1-expressing tumor-infiltrating T cells are a favorable prognostic biomarker in HPV-associated head and neck cancer. *Cancer Res.* **73**, 128–138 (2013).
184. Sartor, M. A. et al. Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. *Epigenetics* **6**, 777–787 (2011).
185. van Kempen, P. M. et al. Differences in methylation profiles between HPV-positive and HPV-negative oropharynx squamous cell carcinoma: a systematic review. *Epigenetics* **9**, 194–203 (2014).
186. Ekanayake Weeramange, C. et al. DNA Methylation changes in human papillomavirus-driven head and neck cancers. *Cells* **9**, https://doi.org/10.3390/cells9061359 (2020).
187. Nakagawa, T. et al. DNA Methylation and HPV-associated head and neck cancer. *Microorganisms* **9**, https://doi.org/10.3390/microorganisms9040801 (2021).
188. Wienand, K. et al. Genomic analyses of flow-sorted Hodgkin Reed-Sternberg cells reveal complementary mechanisms of immune evasion. *Blood Adv.* **3**, 4065–4080 (2019).
189. Trifonov, V., Pasqualucci, L., Tiacci, E., Falini, B. & Rabadan, R. SAVI: a statistical algorithm for variant frequency identification. *BMC Syst. Biol.* **7**, S2 (2013).
190. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
191. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
192. La Gruta, N. L., Gras, S., Daley, S. R., Thomas, P. G. & Rossjohn, J. Understanding the drivers of MHC restriction of T cell receptors. *Nat. Rev. Immunol.* **18**, 467–478 (2018).
193. Yoon, S., Baik, B., Park, T. & Nam, D. Powerful p-value combination methods to detect incomplete association. *Sci. Rep.* **11**, 6980 (2021).
194. rnaturalearth: World Map Data from Natural Earth. (2023).

Acknowledgements
We would like to express our sincere thanks to Laura Pasqualucci for her invaluable suggestions, insightful contributions and expertise in the field of lymphoma, which significantly enriched this study. This work was funded by the National Institutes of Health, National Cancer Institute grants R35CA253126 and U01CA243073 (R.R.), Fondazione AIRC (Investigator Grant no. 23732 to E.T.) and SU2C Convergence Program (K.G., Y.N., J.R. and R.R.).

Author contributions
Conceptualization, Y.N., K.G. and R.R.; Methodology, K.G. and R.R.; Investigation, Y.N., K.G., J.R., C.K., M.A., V.Z. and T.L.; Formal Analysis, Y.N., K.G. and J.R.; Writing – Original Draft, K.G.; Writing – Review & Editing, Y.N., K.G., J.R. and R.R.; Resources, M.C., A.L., G.P., A.C., E.T. and R.R.; Supervision, R.R.

Competing interests
R. Rabadan is the founder of Genotwin, a member of the advisory board of Diatech Pharmacogenetics and Flahy. None of these activities are related to the results in the current manuscript. The remaining authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-60836-9.

Correspondence and requests for materials should be addressed to Raul Rabadan.

Peer review information *Nature Communications* thanks Eric Letouzé, Lawrence Young and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.