nature communications

Article

https://doi.org/10.1038/s41467-025-61597-1

Virus-human chromatin interactions reorganise 3D genome and hijack KDM5B for promoting metastasis in nasopharyngeal carcinoma

Received: 11 September 2024

Accepted: 26 June 2025

Published online: 04 August 2025

Check for updates

Dittman Lai-Shun Chung[1,2,3,10], Zhaozheng Hou[1,10], Ying Wang[4,10], Kazi Anisha Islam[1,10], Songran Liu[1,4,5,10], Jiayan Liu[2,3], Larry Ka-Yue Chow[1], Yuki Yuk-Wun Wong[1], Cyrus Paak-Ting Chak[1], Yingpei Zhang[1], Lanqi Gong[1], Ziyang Qi[1], Kiu-Wai Cheng[1], Zhuoyou Yu[1], Ping Feng[6], Zilu Huang[6], Roger Kai-Cheong Ngan[1], Xinyuan Guan[1,7], Wai-Tong Ng[1,7], Zhonghua Liu[8], Anna Chi-Man Tsang[9], Dora Lai-Wan Kwong[1,7], Anne Wing-Mui Lee[1,7], Victor Ho-Fun Lee[1,7], Honglin Chen[2,3], Yunfei Xia[4,6] & Wei Dai[1,7] ✉

Epstein–Barr virus, the first identified human DNA tumour virus, is detectable in more than 90% of nasopharyngeal carcinoma patients in endemic regions. The 3D chromosome conformation analysis reveals that virus–host chromatin interactions induce the spatial reorganisation of loops and compartments, resulting in "enhancer infestation" and switch of "H3K27 bivalency" at EBV-interacting regions. Through this mechanism, EBV hijacks KDM5B, a gatekeeper of genome stability, and its binding targets, leading to aberrant activation of an EBVIR-enhancer-KDM5B signature. Cancer cells with this signature present increased MYC activation, DNA damage responses, and epigenetic plasticity of epithelial-immune dual features with metastatic potential. Our multicentre multiomics study confirms that this signature is the prerequisite for chromosome instability and can be utilised as a risk factor for distant metastasis. This study highlights a mechanism in which latent viral episomes can alter the host high-order epigenotype, promoting transcriptional rewiring and metastasis in NPC.

Epstein–Barr virus (EBV) is associated with ~200,000 cancer cases each year[1]. EBV stays in latently infected cells in the form of episomes (extrachromosomal plasmids), which replicate during the S-phase of the cell cycle and subsequently partition during cell division in a process that requires tethering to host chromosomes. Several studies have suggested a link between DNA tumour viruses and reorganisation of the genome architecture[2–4]. In the 3D genome, chromosomal territories can be divided into active and inactive compartments[5], and CTCF acts as an insulator to form chromatin loops[6]. Disruption of chromatin loops leads to dysregulated genome folding, resulting in aberrant gene regulation, such as the activation of oncogenes and inactivation of tumour suppressors[7]. Multiple studies have described the dysregulation of enhancer-to-promoter contacts in cancer[8]. "Enhancer infestation" in EBV-associated gastric cancer was recently discovered, suggesting that EBV episomal DNA can release latent enhancers in the human genome by remodelling the chromatin topological structure and activating nearby cancer-relevant genes[9]. However, in nasopharyngeal carcinoma (NPC), the contribution of EBV to the formation of enhancer-to-promoter contacts is poorly studied, and whether EBV episomes engage in a molecular dialogue with human

---

A full list of affiliations appears at the end of the paper. ✉e-mail: weidai2@hku.hk

chromatin and the contribution of EBV episomes to proto-oncogene activation in tumour development and cancer progression remain unknown.

In this study, we utilise an in vitro EBV infection model, paired EBV-positive (EBV+) and EBV-negative (EBV−) cell lines derived from the same NPC patient, EBV+ patient-derived xenografts (PDXs), and clinical NPC samples to decipher the role of EBV in regulating human genome organisation and the clinical impact of EBV-associated changes in the high-order epigenotype and transcription.

Results

Characterisation of the 3D genome in NPC revealed chromosome reorganisation in EBV+ NPC

To capture chromosome conformations, Omni-C analysis was performed with technical replicates (Spearman correlation of duplicates $\rho \geq 0.98$) (Supplementary Fig. 1a). In total, 16 samples were obtained for Omni-C analysis; these samples included normal immortalised nasopharyngeal epithelial (NPE) cells (NP69 and NP460), EBV+ NPC cell lines (C17, C666, NPC43, and HK1EBV), NPC cell lines in which EBV was not detected (EBV− NPC) (NPC43noEBV, NPC53, NPC38, and HK1), EBV+ PDXs (C15, Xeno23, Xeno32, and NPC113), EBV+ gastric carcinoma (EBVaGC) cells (YCCEL1), and one EBV+ clinical NPC sample (Supplementary Data 1). The presence of EBV DNA was confirmed in all EBV+ samples (Supplementary Fig. 1b, c). The A (B) compartments are referred to as active (inactive) compartments at gene-rich (gene-poor) regions, respectively. The Omni-C analysis showed the genomic regions that switched from compartment A (active) to compartment B (inactive) and vice versa in these samples. We used active histone H3K4me3 mark by CUT&RUN-seq to further confirm the identification of active (A) and inactive (B) compartments (Supplementary Fig. 1d)[10]. Consistent compartment switching was identified across the EBV+ NPC cell lines, PDXs, and one clinical tissue sample but not in the EBV− NPC cell lines, suggesting that changes in compartment identity were likely associated with EBV infection (Fig. 1a, Supplementary Data 1). Hierarchical clustering demonstrated dynamic switches from the B compartment to the A compartment and vice versa that were associated with EBV infection (Fig. 1b, c). However, the formation of these compartments in the EBVaGC (YCCEL1) cell line differed from that in the EBV+ NPC samples. Moreover, YCCEL1 cells did not exhibit chromatin accessibility changes specific to EBV derived from NPC cells (Supplementary Fig. 2a, b).

To examine the role of EBV infection for compartment changes, HK1, an EBV-negative cell line, was infected by EBV in vitro, and the results indicated that in vitro EBV infection was sufficient to induce compartment switching and reduce the complexity of chromatin compartments at select chromosomes and regions (Fig. 1d, Supplementary Fig. 2c, Supplementary Data 1). A similar reduction in chromatin compartment complexity was also observed in EBV+ NPC43 cells compared with matched EBV− NPC43 cells derived from the same patient (Fig. 1d, Supplementary Fig. 2c). A human 3D genome on chromosome 4 was constructed from the Omni-C data with a maximum likelihood algorithm in 3DMax and visualised in three dimensions in paired EBV+ and EBV− NPC cell lines, and the selected regions were unified as an active A compartment in the EBV+ cell lines (Supplementary Video 1–4). Compared to that in EBV− NPC cells (NPC43noEBV and HK1), genome folding in EBV+ NPC cells (NPC43 and HK1EBV) was characterised by less defined separation and a reduction of the number of AB compartments in these regions. This dynamic switch between compartments and the enlarged size of the AB compartment indicated the reorganisation of high-order genome folding after EBV infection (Fig. 1d, e, Supplementary Fig. 2d). Given we have demonstrated that EBV episomes can reshape and unify genome organisation at the compartment level and studies showed the depletion of CTCF reduces chromatin loops, leading to vanish the topologically associating domains (TADs)[11,12], we then examined whether EBV infection could induce the large-scale dysregulation of chromatin looping. Aggregate peak analysis (APA) revealed fewer loop–loop interactions in two EBV+ NPC cell lines than in their matched EBV− NPC cell lines (Fig. 1f). Similarly, a reduction in loop–loop interactions was detected in all the EBV+ NPC samples compared with the EBV− NPC and NPE cells (Fig. 1g, Supplementary Fig. 2e, Supplementary Data 1). These results suggest that EBV infection can strongly affect 3D genome organisation by inhibiting chromatin looping.

Virus–host chromatin interactions induce a switch between active and inactive genome compartments

To delve deeper into the mechanisms that drive A/B compartment transitions induced by EBV infection, we extracted Omni-C reads from both the human and EBV genomes and discovered that the EBV episome could interact with select regions on host chromatin, named EBV-interacting regions (EBVIRs) (Fig. 2a). The number of EBVIRs in the human genome was positively correlated with the EBV DNA copy number (Supplementary Fig. 3a, Supplementary Data 2). We further categorised the active and inactive compartments as either enhanced or reduced (A-to-A or B-to-B, respectively) compartments. In total, six types of A/B compartment changes were included in the analysis to examine whether EBV–host interactions contributed to the switching of compartments (Supplementary Fig. 3b). EBV–host interactions were constantly enriched at inactive (B) to active (A) compartments in cell lines and PDXs (Fig. 2b, Supplementary Data 2), suggesting virus–host chromatin interactions is one of the mechanisms underlying B-to-A transition in EBV+ NPC. The active compartment (compartment A) contains more enhancers and active promoters[13]. Our results indicate that EBV–host interactions led to enhancer activation (an increase in H3K27ac) and transcriptional upregulation (Fig. 2c). Surprisingly, we also identified EBVIRs in which H3K27ac was lost but H3K27me3 remained in EBV+ NPC cells (enhancer inactivation). These two histone signatures (enhancer activation and enhancer inactivation) were presented in EBV+ NPC43 cells but not in EBV- NPC43 cells (Fig. 2c). Given that both H3K27me3 and H3K27ac at these EBVIRs were detected in normal cells and EBV− NPC cells, these cells are likely heterogeneous at these EBVIRs, with dynamic changes in both H3K27 markers ("H3K27 bivalency").

Strikingly, we revealed three non-random regions on the EBV episome covering oriP (EBV–oriP–EBVIRs), the EBNA1 intron and oriLyt (EBV–EBNA1–EBVIRs), and the long noncoding RNA (lncRNA) RMPS1 (EBV–RPMS1–EBVIRs), which frequently interacted with host chromatin. This result was consistent across all the EBV+ NPC cell lines and PDXs, as further confirmed by an in vitro experiment with the EBV-infected NPC cell line HK1 (Fig. 2d, Supplementary Data 2). We then characterised the genomic and epigenomic features of EBVIRs. In the human genome, EBV episomes were preferentially tethered at chromosomes 3–6, and 13 in regions of low gene density with an AT-rich DNA sequence (Supplementary Fig. 3c-d, Supplementary Data 2). Notably, EBV–oriP–EBVIRs and EBV–EBNA1–EBVIRs more frequently interacted with chromosome 1q31.1–32.1 (Fig. 2a, Supplementary Fig. 3d, Supplementary Data 2).

Transcriptional rewiring of KDM5B and its binding targets associated with enhancer infestation induced by EBVIRs

We investigated the expression of genes near the EBVIRs associated with enhancer activation using single-cell RNA sequencing (scRNA-seq) data from a previous study[14–17], in which we identified two clusters (EBV-high and EBV-low clusters) of cancer cells with high and low EBV latent gene expression that were distinguishable by *LMP1* and *RPMS1* expression (Fig. 3a, Supplementary Fig. 4a). Genes whose expression was aberrantly upregulated by enhancer activation were identified in the EBV-high cluster. The integrative results further indicated that EBV preferentially tethers to the host chromatin and activates enhancers to promote transcription at KDM5B and its binding targets, named as EBVIR–enhancer–KDM5B signature (Fig. 3b, c, Supplementary

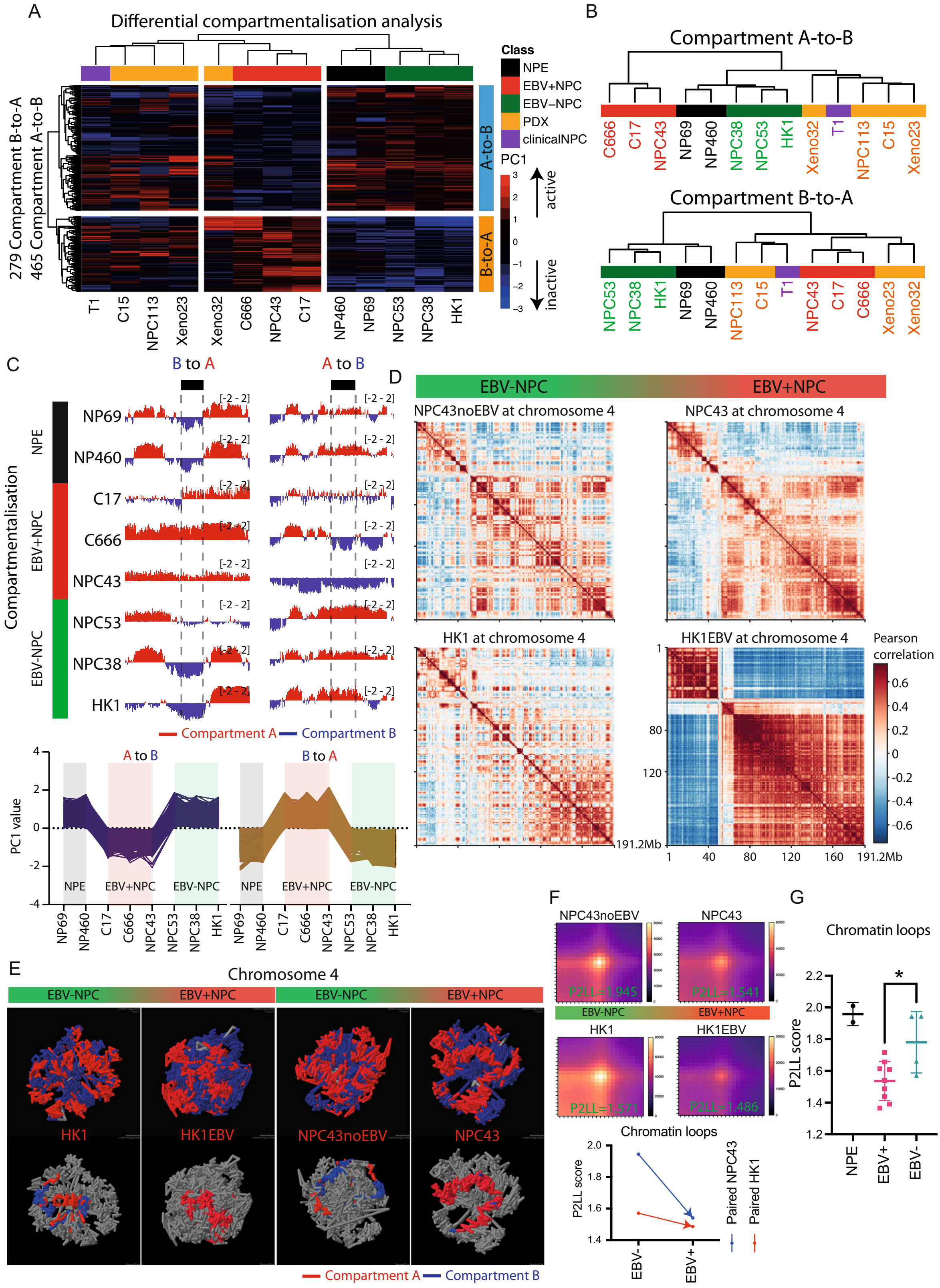


Fig. 4b–e, Supplementary Data 2). This signature was defined as a gene set including KDM5B and KDM5B binding targets that were upregulated in EBV+ NPC with enhancer activation through EBV–host chromatin interactions. KDM5B is located at chromosome 1q32.1, containing a significant number of EBVIRs (Fig. 2a). This region harbours a newly formed topologically associated domain (neoTAD), reflecting altered chromatin architecture. Within this neoTAD, the promoter of KDM5B and its associated cis-regulatory elements exhibit increased chromatin accessibility, facilitating enhanced interactions with a local enhancer. Consistent with this open chromatin state, KDM5B expression was significantly upregulated (Fig. 3d, e, Supplementary Fig. 4f). Furthermore, the expression levels of KDM5B and the EBVIR–enhancer–KDM5B signature were assessed in clinical specimens, revealing a significant elevation of these genes in the EBV-high

Fig. 1 | EBV-associated 3D genome alterations. A EBV-associated compartment features in NPC cells. EBV+ NPC cell lines, EBV+ PDXs and an EBV+ clinical tissue sample were clustered separately from NPE and EBV− NPC cell lines. Differential compartment analysis based on principal component 1 (PC1) was carried out with the EBV+ NPC versus EBV− NPC cell lines. A total of 279 regions switched from compartment A to B, and 465 regions switched from compartment B to A. **B** Unsupervised hierarchical clustering of B-to-A and A-to-B compartment switching in NPC cells. **C** Hierarchical clustering of EBV-associated compartment switching from B to A and A to B in NPE, EBV+ NPC and EBV− NPC cells visualised on the basis of PC1 value (a 50 kbp window size) through the IGV browser. A positive (negative) PC1 value indicates compartment A (compartment B). **D** Redistribution of AB compartments in two paired EBV+ and EBV− NPC cell lines at chromosome 4. A Pearson correlation matrix was normalised by the Knight–Ruiz (KR) method. **E** Visualisation of the 3D genome of chromosome 4 in matched NPC cell line pairs. A selective genomic region showed that compartments A and B were clearly separated from each other without EBV infection, while EBV-infected cells have reorganised compartments with B-to-A transitions. The red colour indicates A (active) compartments, and the blue colour indicates B (inactive) compartments. **F** Reduction in chromatin loops in EBV-infected NPC cells. Aggregate peak analysis (APA) plots based on the loop interactions called from Moustache. A total of 66,100 regions were identified as loops in NPE, EBV+ NPC and EBV− NPC cells and are plotted in 20 × 20 square pixels. APA plots of paired EBV+ and EBV− NPC cells. P2LL denotes the peak-to-lower-left of the interaction matrices. The higher P2LL score indicates stronger chromatin interactions. **G** Reduction in chromatin loops in EBV+ cells. P2LL denotes peak-to-lower-left. Two-tailed Student's *t*-test was performed EBV+ ($n = 7$) versus EBV− ($n = 3$) cells (p-value = 0.018), *p-value ≤ 0.05. The error bar indicates mean ± standard deviation (SD) of the data. NPE ($n = 2$).

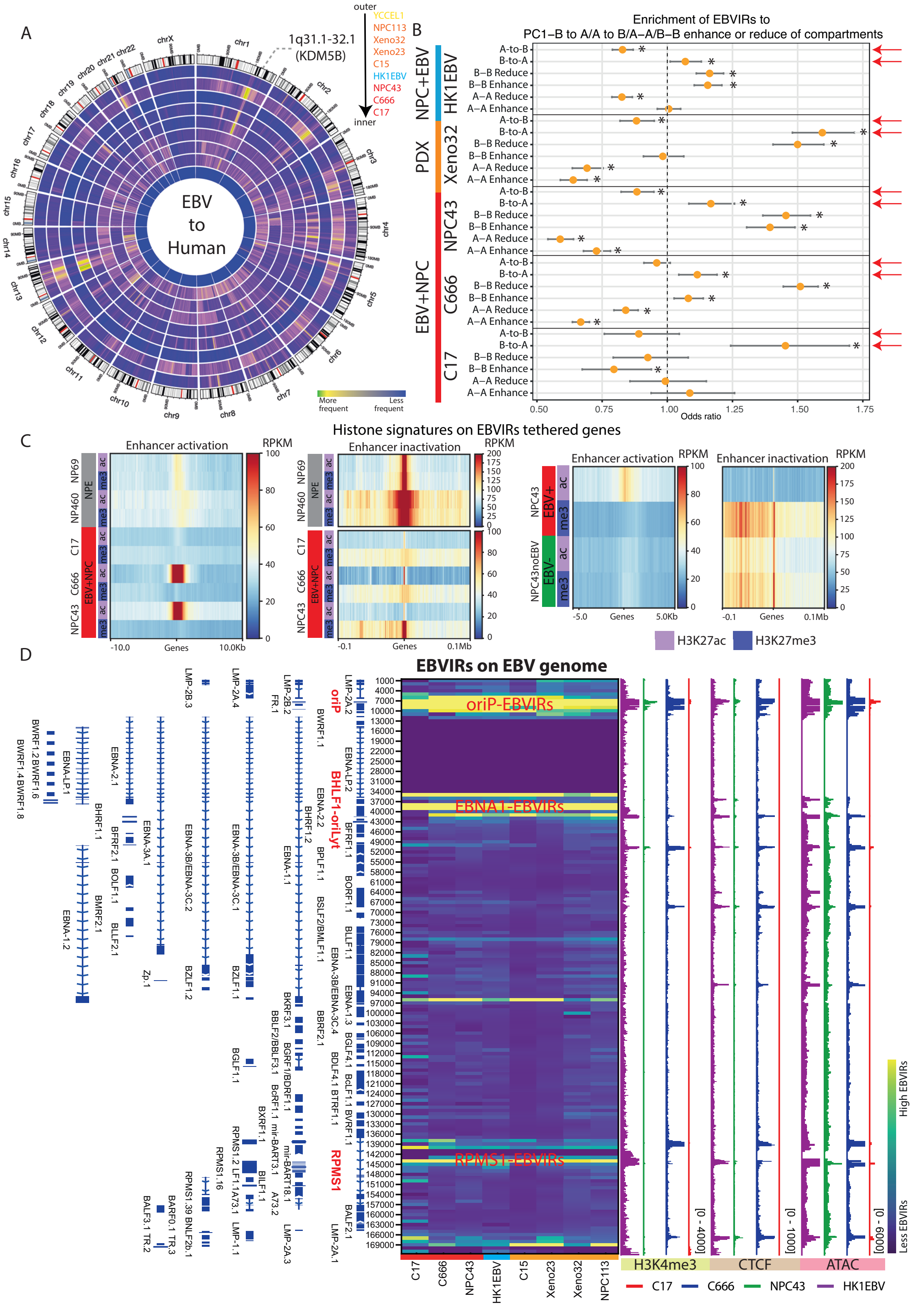
NPC cell cluster compared with the EBV-low NPC cell cluster (Fig. 3f, g, Supplementary Fig. 5a, b, Supplementary Data 3).

To further confirm the association of latent EBV infection with the activation of EBVIR–enhancer–KDM5B signals, in situ hybridisation probes targeting *RPMS1* and *LMP1* were synthesised and applied to clinical samples for single-cell resolution analysis using the NanoString Spatial Molecular Imaging (SMI) platform (Fig. 4a). We identified a total of ten subclusters of epithelial cells (Supplementary Fig. 6a, b), of which three epithelial NPC cell clusters (Epi-1, Epi-2 and Epi-3) had hyperactivated EBVIR–enhancer–KDM5B signals with increased proliferative capacity (MKI67) and activation of MYC and BCL2 expression, concordant with expression of the EBV lncRNA *RPMS1* (Fig. 4b–e, Supplementary Fig. 7a). We found that these three cell clusters, which accounted for 10-20% of the total epithelial cells (Fig. 4f), expressed high levels of *CD74*, a major histocompatibility complex class (MHC) II chaperone, *CIITA*, a transactivator of MHC class II, and immunosuppressive genes, including *LGALS9* (Galectin-9) and *CD47* (a "do-not-eat-me" surface glycoprotein), and this finding was confirmed with scRNA-seq data (Fig. 4c–f, Supplementary Fig. 7a, b). Analysis of both the scRNA-seq and NanoString SMI data revealed that the upregulated expression of *RPMS1* in NPC cells with an active EBVIR–enhancer–KDM5B signature promoted molecular functions involved in MHC class II receptor activity, protein complex binding, and DNA and RNA binding (Supplementary Fig. 7c, Supplementary Data 3). The cancer cells with an elevated EBVIR–enhancer–KDM5B signature also presented increased activity of pathways involving DNA repair, the G2/M checkpoint, the mitotic spindle and protein secretion (Fig. 4d), which is consistent with the functional annotation of KDM5B binding targets (Supplementary Fig. 4c). In addition, vascular cell adhesion molecule 1 (*VCAM1*), which is associated with the aggressiveness and metastasis of cancer cells[18], was significantly upregulated in these cells. The SMI data further suggested that two clusters (Epi-2 and Epi-3) of these cells with strong epigenetic dysregulation involving the Polycomb complex 2 (PRC2) key member EZH2 and the DNA methyltransferase DNMT1 were more likely to have disseminated into the stromal area (Fig. 4a, c). These data highlight the importance of an elevated EBVIR–enhancer–KDM5B signature in NPC cell subsets with epigenetic plasticity and high proliferative capacity that are under strong DNA damage stress with metastatic potential. In addition, we further examined the KDM5B expression level by immunohistochemistry (IHC) staining in the clinical specimens, and the results showed that the patients with high expression of KDM5B and a low number of tertiary lymphoid structures (TLS) were more likely to develop metastasis after conventional treatment (Supplementary Fig. 8a).

A multicentre and multiomics study revealed that the EBVIR–enhancer KDM5B signature is associated with chromosome instability and distant metastasis in NPC

On the basis of these data, we hypothesised that aberrant activation of KDM5B and its binding targets through enhancer activation at EBVIRs is associated with the aggressiveness of cancer cells in NPC, and patients with a strong EBVIR–enhancer–KDM5B signature are likely to have poor clinical outcomes. To test this hypothesis, we carried out a multicentre and multiomics study to examine somatic mutations and chromosomal instability (CIN) in NPC via whole-exome sequencing (WES), the expression of the EBVIR–enhancer–KDM5B signature and cancer-related pathway activity via RNA sequencing in NPC, and the clinical impact of the EBVIR–enhancer–KDM5B signature after conventional chemoradiotherapy. Primary tumours of 77 samples were obtained from treatment-naïve NPC patients recruited from Queen Mary, Queen Elizabeth, Princess Margaret and Pamela Youde Nethersole Eastern Hospitals in Hong Kong (HK) and of 100 samples from the Sun Yat-Sen University Cancer Centre in Guangzhou (GZ). The patients were classified based on 3-year distant metastasis-free survival (DMFS) as MET NPC patients (HK & GZ: 39 and 36 patients) or non-MET NPC patients (HK & GZ: 38 and 64 patients). The clinical characteristics of the HK and GZ cohorts are summarised in Table 1. Gene set enrichment analysis (GSEA) revealed significant enrichment of the EBVIR–enhancer–KDM5B signature in MET patients in both the HK and GZ cohorts (NES = 1.89, FDR *q* value = 0.0111 and NES = 4.45, FDR *q* value = 0.0001) (Fig. 5a).

In addition, we identified differentially expressed genes in cancer cells (PanCK-enriched segments) from the MET patients and non-MET patients via spatial analysis of whole-transcriptome data (Fig. 5b, Supplementary Data 4), and the associations of these genes with metastasis were subsequently evaluated in the HK and GZ cohorts (Supplementary Tables 1 and 2). Notably, the validated metastasis-associated genes were significantly enriched for KDM5B and MYC binding targets (Fig. 5c). The Gene set variation analysis (GSVA) has demonstrated that KDM5B activity is upregulated in MET patients compared to non-MET and normal control samples (Supplementary Fig. 9a, Supplementary Data 4). The EBVIR–enhancer–KDM5B signature was highly correlated with the activity of KDM5B targets derived from this independent analysis of both study cohorts (HK cohort, $r = 0.613$; GZ cohort, $r = 0.898$; Supplementary Fig. 9b, Supplementary Data 4). Chromatin modulators, such as YY1, P300 and CTCF binding targets, were also enriched in metastasis-associated genes in both the HK and GZ cohorts (Fig. 5c, Supplementary Data 4). CTCF binding is correlated with enhancer–promoter interactions for gene regulation, and YY1 is a structural regulator of enhancer–promoter loops[19,20]. Thus, the results imply that enhancer-to-promoter interactions and chromatin looping are more likely to be dysregulated in MET patients. Functional annotation of the metastasis-associated genes revealed the involvement of dysregulated DNA damage and repair through the binding of BRCA2 to RAD51 and the G2/M checkpoint, as well as the involvement of epigenetic dysregulation through PRC2 and DNA methylation, in the development of NPC distant metastasis (Fig. 5d, Supplementary Data 4). Integrative analysis of WES and RNA-seq data revealed that chromosome instability was associated with MET patients (Supplementary Fig. 9a, c, Supplementary Data 4), and the patients with increased chromosome instability always



presented high levels of the EBVIR–enhancer–KDM5B signature (Fig. 5e, Supplementary Fig. 9d).

Multivariable logistic regression analysis further revealed that high KDM5B pathway activity is a risk factor for metastasis, independent of overall tumour stage and plasma EBV DNA copy number before and after treatment (Table 1, Fig. 6a). Furthermore, KDM5B activity can be used in combination with the overall stage and the plasma EBV DNA level at baseline (AUC = 0.77, 95% CI: 0.67–0.87) or after treatment (AUC = 0.82, 95% CI: 0.72–0.92), as a promising predictor of risk (Supplementary Fig. 9e, Supplementary Tables 3–8). The late-stage patients (stage IV) with high KDM5B pathway activities and post-treatment EBV DNA levels have a much higher risk for metastasis (OR = 5.8, 95% CI 2.2–15.4, *p*-value = 0.00004, specificity = 0.9 and sensitivity = 0.6, Supplementary Fig. 9f).

Fig. 2 | Identification of three EBV-interacting regions (EBVIRs) associated with enhancer activation and inactivation. **A** EBV-interacting events across human chromosomes in EBV+ NPC cells, PDX, and EBVaGC (YCCEL1) cells. The frequency of EBV tethering to the human genome was calculated with a 50 kb window size and visualised as a heatmap on a Circos plot. **B** Consistent B-to-A compartment switching was found at EBVIRs. The odds ratio (OR) was used to calculate the association of EBVIRs with six types of compartment switching (A-to-B, B-to-A, B–B/A–A enhance, B–B/A–A reduce) in EBV+ NPC cells, PDXs and EBV-infected NPC cells. The NPE cell control was used as a reference. One-tailed Fisher's Exact Test has been performed to test if EBVIRs are associated with subsets of compartment switches, * p-value ≤ 0.05. The total number of events (C17 = 110658, C666 = 110648, NPC43 = 110660, HK1EBV = 110520 and Xeno32 = 110490) has been examined for the significance. The error bar indicates a 95% confidence interval (CI) of the data. Compartment A-to-B (B-to-A) switching is indicated by a change in the PC1 value from positive (negative) to negative (positive). The enhanced and reduced compartments indicate the magnitude of the change in the PC1 value within the same compartment. The arrows indicate that more EBV was consistently situated at B-to-A compartments than at A-to-B compartments across all samples. **C** EBVIRs are associated with enhancer activation and enhancer inactivation. Euclidean clustering resulted in two clusters of EBVIRs, which are named enhancer activation and enhancer inactivation EBVIRs. The CUT&RUN-seq signals were normalised in Reads Per Kilobase per Million mapped reads (RPKM). **D** Discovery of the three most frequent contact regions in the EBV genome. The EBV genome is annotated in the left panel. A two-tailed Z-score was calculated with data from individual samples, Z-score > 1.645, 90% confidence level (p-value ≤ 0.1); number of regions used for statistical test $n_{C17} = 1232$, $n_{C666} = 18{,}405$, $n_{NPC43} = 7063$, $n_{HK1EBV} = 25{,}796$, $n_{C15} = 22{,}348$, $n_{xeno23} = 1{,}11{,}888$, $n_{xeno32} = 6464$, and $n_{NPC113} = 31{,}400$. Three regions were significantly enriched across C17, C666, NPC43, Xeno32 and HK1EBV cells. The RPKM normalised chromatin accessibility (ATAC), CTCF binding, and H3K4me3 histone modification (CUT&RUN-seq) with a 10 bp window size and summarised signal tracks along the EBV genome in EBV+ NPC cells are shown.

KDM5B is one of the key host factors to promote cancer cell growth and maintain EBV DNA copies

To further understand the functional role of KDM5B, we inhibited expression of KDM5B through knockdown (KD) in the EBV+ NPC cells and observed a reduction of cancer cell growth (Fig. 7a, b). Similarly, treatment of the KDM5B inhibitor (GSK-467) has demonstrated a strong suppression of cancer cell proliferation in EBV+ NPC and EBV+ gastric cancer cells (Fig. 7c, Supplementary Fig. 10a, Supplementary Data 5). As we discovered, the cancer cells with a strong EBVIR–enhancer–KDM5B signature were highly proliferative and associated with metastasis in the clinical samples, we further investigated the cancer cell growth upon a genome-wide loss-of-function (LOF) analysis using the CRISPR screening approach. Targeting the EBVIR–enhancer–KDM5B signature, a noticeable effect on suppression of cancer cell growth was observed in this experiment (NES = −1.36, FDR q-value = 0.039) (Fig. 7d). This result further confirmed that not only KDM5B alone but EBVIR–enhancer–KDM5B targets together also strongly influence tumour cell growth. The bulk RNA-seq data analysis for KDM5B–KD further revealed that KDM5B regulated the genes in this EBVIR–enhancer–KDM5B signature, controlling for viral life cycle, regulation of viral entry into host cells, and oxidative stress-induced cell death (Fig. 8a). To investigate the role of KDM5B for regulating EBV life cycle, we examined the KDM5B binding signals on the EBV genome. As previously reported, EBNA1 is important for binding to the EBV–oriP region for persistent, stable EBV infection[21]. Interestingly, the colocalisation of KDM5B, EBNA1 bindings and H3K27ac signals at the EBV–oriP region indicates that KDM5B has a potential role in regulating EBV latency and is involved in EBV–host chromatin interactions to maintain the EBV genome (Fig. 8b). Consistent with this finding, the KDM5B–KD cells showed an obvious reduction in cellular EBV copies (Fig. 8c, Supplementary Data 5). This result was further confirmed by the clinical specimens, where cellular EBV copies were positively correlated with KDM5B pathway activity in the tumour tissues ($r = 0.633$, Fig. 8d, Supplementary Data 5). Collectively, these results showed that EBV hijacked KDM5B to promote cancer cell growth and maintain its survival in a cellular context.

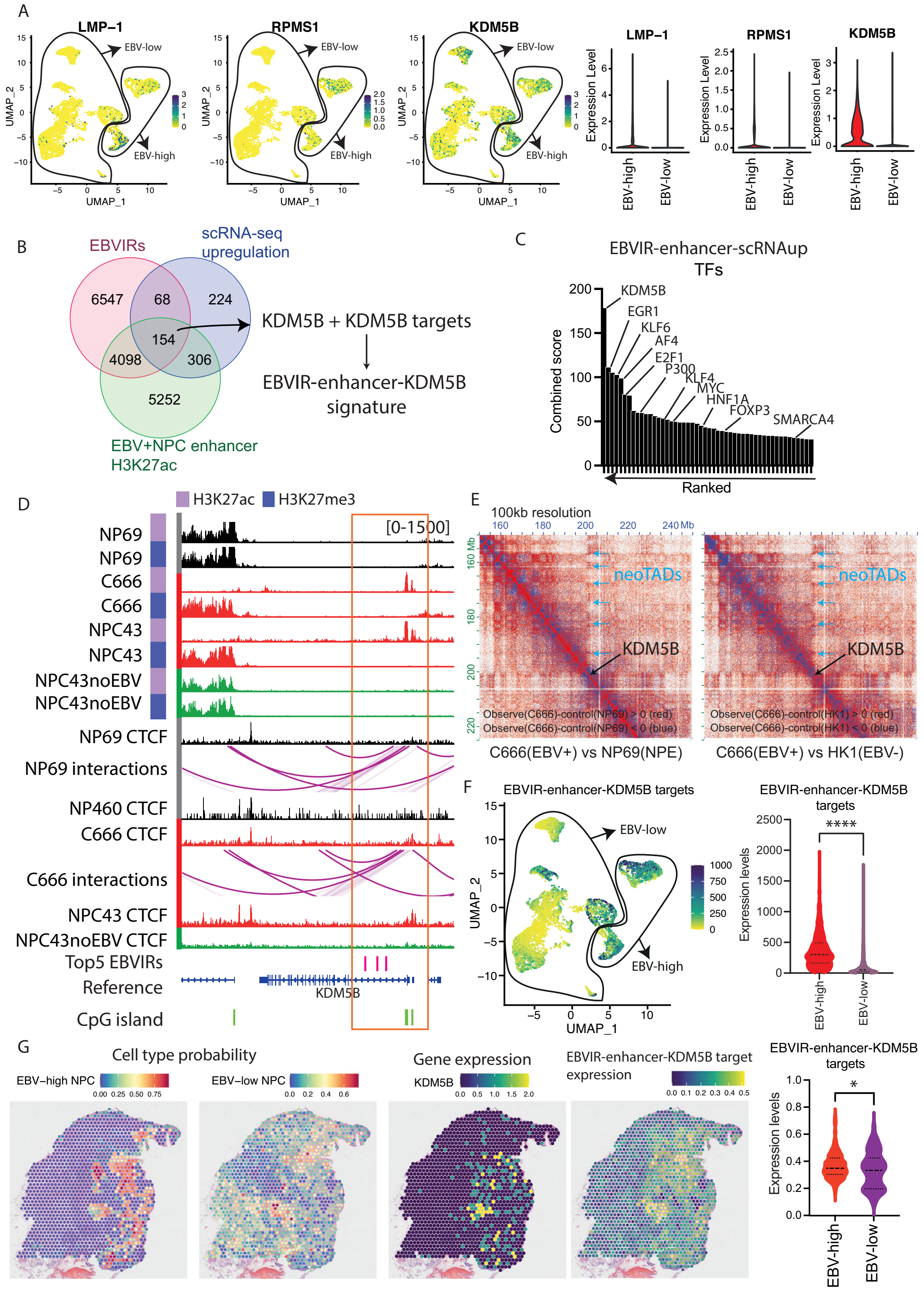
To explore the role of KDM5B on metastasis, the differential expressed genes (DEGs) analysis to compare KDM5B–KD and KDM5B–WT showed that expression of *VCAM1* and *VEGFA*, which have been identified from our SMI data, were suppressed by KDM5B–KD (Fig. 4c, Supplementary Fig. 10b, Supplementary Data 5). VCAM1, encoding Vascular Cell Adhesion Molecule 1, is known to play a significant role in various cancers, promoting tumour cell invasion, metastasis, and cellular immune response[18,22,23]. VEGFA, encoding Vascular Endothelial Growth Factor A, has a crucial role in metastasis by promoting tumour angiogenesis and inhibiting T-cell functions[24,25]. Functional enrichment analysis further revealed that the other genes repressed by KDM5B–KD were highly involved in the metastasis-related pathways, such as hedgehog signalling, epithelial–mesenchymal transition (EMT), and hypoxia (Supplementary Fig. 10c, Supplementary Data 5).

In addition, the KDM5B-regulated targets were enriched for several transcription factors (TFs), such as BRD4 (Supplementary Fig. 10d, Supplementary Data 5), suggesting the coordination of multiple TFs for regulating the transcription of these genes. BRD4, encoding Bromodomain Containing 4, is an enhancer epigenomic reader, regulating histone acetylation to control transcriptions for cancer cell proliferation[26]. Given the important role of enhancer activation in promoting metastasis discovered in this study, we investigated the therapeutic potential of a BRD inhibitor (JQ1), which is known to inhibit the enrichment of BRD4 at enhancer sites[27]. The treatment of the BRD inhibitor showed much higher efficacy than the KDM5B inhibitor and a dramatic inhibitory effect on cancer cell growth across both EBV+ NPC and EBVaGC cells (Fig. 8e, Supplementary Data 5).

Discussion

In this study, we demonstrated reorganisation of the human 3D genome upon EBV infection, with compartmentalisation changes and a reduction in chromatin looping, in NPC. Notably, EBV, in the form of episomes, can interact with the host genome to induce B-to-A compartment switching and enhancer infestation at EBVIRs, leading to transcriptional rewiring in a cis-regulatory manner. Although the exact mechanisms underlying these chromatin interactions remain unclear, it has been reported that the EBV oriP regions tether to host chromosomes through the AT-hook DNA-binding capacity of Epstein–Barr nuclear antigen 1 (EBNA1) to maintain viral genomes in latently infected cells[21,28,29]. Consistently, we found that host EBVIRs have an increased AT content in regions with low gene density, suggesting the role of EBNA1 in EBV–host chromatin interactions. In addition, scRNA-seq and spatial SMI data suggest that the EBV lncRNA *RPMS1* may be another key player that facilitates EBV tethering on host chromosomes to maintain viral survival during latent infection. Besides the EBV–oriP–EBVIR region, we identified two additional EBV regions, EBV–EBNA1–EBVIRs and EBV–RPMS1–EBVIRs, that are likely to interact with host chromosomes. The EBV–EBNA1–EBVIRs cover oriLyt, the lytic origin of DNA replication by BZLF1. This region is surrounded by CTCF-binding sites with increased chromatin accessibility, suggesting that the interaction of EBV with host chromosomes through these EBV–EBNA1–EBVIRs may facilitate the insulation of latent/lytic gene expression.

A phylogenetic study revealed that the EBV genome derived from EBVaGC cells (YCCEL1) occupies a subbranch under type I EBV that is separate from the EBV strain derived from NPC[30,31]. We observed a discrepancy in compartment switching and chromatin accessibility between EBV-associated gastric cancer and NPC. This finding implies that the ability of EBV to induce cancer-specific compartment switching in the human genome depends on its cell and tissue specificity. Moreover, EBV infection in gastric cancer and NPC is associated with



different latency programmes. EBV in NPC is mainly associated with type II latency, expressing *LMP1*, *LMP2A*, *EBNA1*, *EBERs1/2*, and *BART* transcripts, whereas EBV infection in gastric cancer is associated with type I latency[32,33]. Differences in EBV sequences and latency programmes may also contribute to the distinct impacts on host chromosome reorganisation. However, despite the tissue specificity of the epigenomic profile, EBV docking to the human genome may utilise a similar mechanism leading to enhancer infestation to promote tumour growth in both EBVaGC and EBV+ NPC, which is supported by the findings in this study that inhibition of KDM5B and enhancer blockade could suppress cancer cell growth in both EBV+ cancer types.

Our study further revealed that EBV episomes preferentially tether to KDM5B and its binding targets, leading to enhancer and aberrant upregulation of the expression of these genes in NPC.

Fig. 3 | Multiomics integrated data analysis identified KDM5B and its targets related to EBV infection. A The expression levels of *LMP1*, *RPMS1*, EBV latency genes and *KDM5B* in epithelial cells were determined by scRNA-seq. *LMP1* and *RPMS1* were used to stratify the cells into EBV-high and EBV-low clusters. The normalised KDM5B expression is examined and linked to high expression of EBV latent genes. **B** Venn diagram following the multiomics integrated data analysis. Upregulated genes in EBV-high clusters were identified from differential gene expression analysis by comparing the EBV-high ($n = 5335$) and EBV-low ($n = 1876$) clusters in scRNA-seq data. A two-tailed Mann–Whitney *U*-test has been performed, with an adjusted p-value ≤ 0.05. **C** Putative transcription factors involved in regulating upregulated genes in EBVIRs through enhancer activation. TFs were ranked on the basis of their combined score ($\log_{10}$ (p-value) from Fisher's exact test × the the *Z* score for deviation from the expected rank). **D** Integrated IGV browser signal tracks at the KDM5B locus. Histone modifications, including H3K27ac and H3K27me3, CTCF binding signals at the same scale, together with Omni-C chromatin interactions for C666 and NP69 cells, and EBVIRs are included. The orange frame highlights EBVIRs at the KDM5B promoter region, with chromatin contacts forming CTCF chromatin loops in nearby regions. **E** neoTAD formation at the KDM5B gene locus. Omni-C data were used to visualise TAD formation, and data were compared between EBV+ NPC and NPE and EBV- NPC cells using Juicebox. The plot was further normalised by sequencing coverage visualising at 100 kbp resolution. If the observed deduce control data is greater than zero (red), it indicates more chromatin contacts in C666, and vice versa. **F** EBVIR–enhancer–KDM5B target expression determined by scRNA-seq. The sum of normalised expression levels in EBVIR–enhancer–KDM5B signature has been calculated in EBV-high and EBV-low cells ($n_{\text{cells}} = 7211$). The expression levels in the EBV-high ($n = 5335$) and EBV-low ($n = 1876$) clusters were statistically evaluated by a two-tailed unpaired t-test, ****p-value ≤ 0.001. **G** Spatial transcriptomes of KDM5B and EBVIR–enhancer–KDM5B target expression levels in the EBV-high and EBV-low regions in a clinical sample. The average expression levels of genes in EBVIR–enhancer–KDM5B signature were calculated for each circle ($n_{\text{spots}} = 264$), visualised by image and quantified by the violin plots. The expression levels in the EBV-high ($n = 63$) and EBV-low ($n = 201$) clusters were statistically evaluated by two-tailed unpaired t-test (p-value = 0.04), *p-value ≤ 0.05.

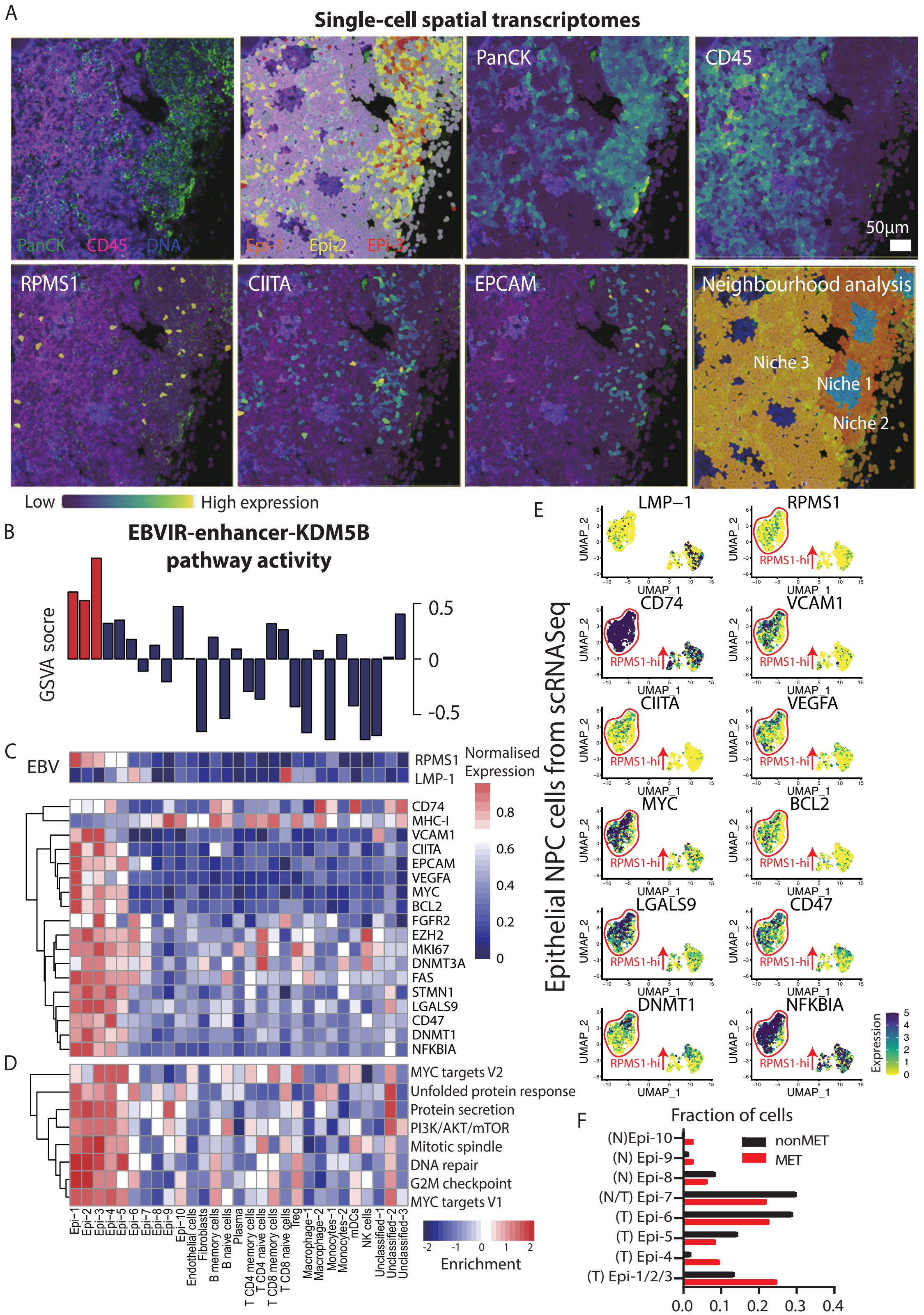
KDM5B, a lysine demethylase, has multiple functions as a transcriptional repressor and activator depending on the cellular context, and a previous study suggested that KDM5B plays a role in regulating epigenetic plasticity to facilitate the activation or repression of transcriptional programmes[34]. KDM5B targets with EBV tethering are mainly involved in the G2/M checkpoint, mitotic spindle formation, MYC targets and protein secretion, as revealed by functional annotation and analysis of SMI data. Notably, cancer cells with the EBVIR–enhancer–KDM5B signature have a highly proliferative capacity in the clinical specimens, and our functional data together indicate inhibition of KDM5B alone and EBVIR-associated KDM5B targets dramatically suppresses tumour growth in vitro, suggesting the cancer cells with EBVIR–enhancer–KDM5B signature have strong growth advantages. Functional study showed that KDM5B regulates the metastasis-promoting genes, such as VCAM1 and VEGFA. These results are consistent with the findings from SMI data in the clinical specimens and bulk RNA sequencing in the multi-centre study, that the cancer cell clusters with this signature have highly proliferative capacity with metastatic potential.

In addition, the cancer cells with EBVIR–enhancer–DM5B signature have elevated expression of class II transcriptional activator (CIITA). CIITA is a transcriptional activator that interacts with the enhanceosome complex of MHC-II genes[35,36]. This interaction is sufficient to induce the expression of MHC class II genes in MHC-II-deficient cells without interferon-γ stimulation[37]. Similarly, data have shown that in vitro EBV infection of pseudostratified epithelium in a 3D primary nasopharyngeal air–liquid interface cell culture model induces the expression of MHC-II genes, including *CD74*, in suprabasal cells[38]. The invariant CD74 chain physically mediates the subcellular trafficking of the MHC-II complex to promote the expression of antigen-bound MHC-II molecules on antigen-presenting cells (APCs), such as B cells, dendritic cells and macrophages. The MHC-II molecules on APCs evoke adaptive T-cell immunity by boosting CD4 helper T cells and tumour-reactive CD8+ T cells[39,40]. However, a previous single-cell RNA sequencing study in NPC revealed the dual epithelial-immune features of NPC cells with high expression of the MHC-II molecule HLA-DR; these cells are highly proliferative and can reduce the cytotoxicity of CD8+ T cells and are thus associated with poor prognosis[41]. In this study, we report that NPC cells with a strong EBVIR–enhancer–DM5B signature have similar dual epithelial-immune features and express multiple immunosuppressive molecules, such as CD47 and Galectin-9 (LGALS9). But transcription of these immunosuppressive genes and MHC-II genes was not altered upon KDM5B–KD. It is likely that the elevation of these genes may be directly associated with enhancer activation, but independent of KDM5B regulation. Taken together, these data indicate that the high-order epigenotype changes associated with EBV–host chromatin interactions promote tumour growth and increase the epigenetic plasticity of epithelial cells with immune cell features for immune evasion in NPC, which likely contributes to the increased risk of metastasis in NPC.

The SMI data suggested that cancer cells with an elevated EBVIR–enhancer–KDM5B signature induce strong DNA damage stress. Our multicentre multiomics study confirmed that NPC patients with a strong EBVIR–enhancer–KDM5B signature have increased chromosome instability and are more likely to develop distant metastasis after conventional treatment. The roles of EBV in genomic and chromosomal instability have been demonstrated in several recent studies. Li et al. suggested that EBNA1 binds fragile DNA on human chromosome 11, triggering the breakage of this DNA in a dose-dependent manner, thereby inducing genomic instability and alterations in the copy numbers of the tumour suppressor ATM and the proto-oncogene MLL after cell division, which might contribute to tumorigenesis[42]. Shumilov et al. reported that EBV particles induce centrosome amplification and chromosomal instability in a B-cell model[43]. Additionally, KDM5B is a key factor in response to DNA damage via the recruitment of Ku70 and BRCA1[44]. Here, we provide more evidence for the potential involvement of EBV in host DNA damage and the DNA repair machinery through cross-species chromatin interactions and the hijacking of KDM5B function.

The clinical implication of this study is the identification of the EBVIR–enhancer–KDM5B signature as an independent risk factor for distant metastasis. Additionally, we established a multivariable model incorporating KDM5B pathway activity and plasma EBV DNA copy number at diagnosis and after treatment, together with the overall stage, for the prediction of distant metastasis. Metastasis occurs in 7–20% of NPC patients, reducing the 5-year survival rate from 80–90% to 20–30%[45–47]. Even in the era of immunotherapy, metastasis remains the leading cause of mortality in NPC patients, with a median progression-free survival (PFS) of less than 1 year[48,49]. The identification and management of metastatic risk would be helpful for improving the prognosis of NPC, but are challenging. Identification of this risk factor could provide molecular evidence for predicting metastasis. The specificity to identify high-risk patients is as high as 0.9 when KDM5B activities combine with overall stage and EBV DNA copy number post-treatment. Our functional data further suggest that inhibition of KDM5B and its targets dramatically suppresses tumour cell growth, indicating the potential of KDM5B inhibitors as the maintenance therapy for this subset of patients. Alternatively, JQ1 treatment targeting enhancer activation could be another promising treatment option to prevent metastasis and improve treatment outcomes in these high-risk patients.

In this work, EBV infection drives host 3D genome reorganisation through virus–host chromatin interactions to increase cellular epigenetic plasticity. Subsequent enhancer infestation aberrantly activates KDM5B and its binding targets. By this mechanism, EBV hijacks host



factor KDM5B, getting involved in the DNA repair pathway and promoting cancer cell proliferation and metastasis. This aberrant KDM5B pathway activity could be used as a risk factor for NPC metastasis in the future. Furthermore, cancer cells exhibiting enhancer-driven metastatic potential display pronounced immunosuppressive traits, suggesting that immune evasion synergises with epigenetic reprogramming to promote EBV-associated NPC progression.

Methods

Ethical approval for the study involving human subjects

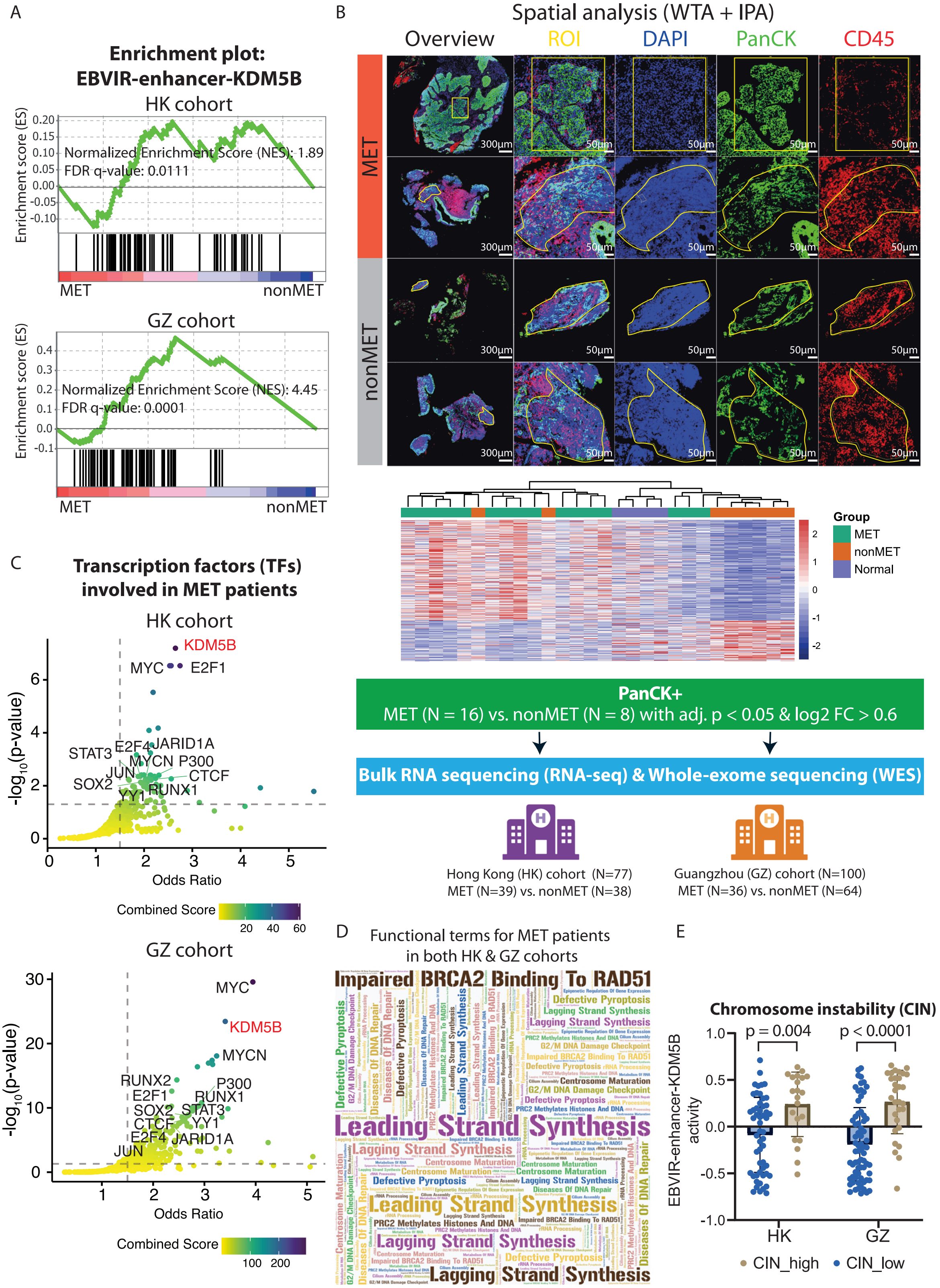
Informed consent for sample collection was written and obtained from patients, and samples were collected according to protocols approved by the Institutional Review Board of the University of Hong Kong (**UW 09-251**) and Hong Kong Clinical Research Ethics Committee (**HKECREC-2021-078**) for the HK cohort and the Institutional Review

Fig. 4 | Single-cell spatial transcriptome analysis revealed subclusters of epithelial NPC cells with aberrant upregulation of the EBVIR–enhancer–KDM5B signature. A Single-cell spatial transcriptome of a metastatic NPC patient sample ($n_{\mathrm{FOV}} = 2883$) illustrated by the AtoMx platform. Two FFPE slides from one MET NPC patient and one non-MET NPC patient have been used to conduct single-cell spatial transcriptomes. PanCK, CD45 and DNA staining were performed. The epithelial cell clusters (Epi-1, Epi-2 and Epi-3) are shown, and select genes, including PankCK, CD45, RPMS1, CIITA and EPCAM, were visualised (highest expression: yellow, moderate expression: green, and lowest expression: blue). Neighbourhood analysis using Jaccard distance stratified cells into three niches in this selected region of interest. **B** Pathway activity of EBVIR–enhancer–KDM5B targets in different cell types. The GSVA score for EBVIR–enhancer–KDM5B pathway activity was calculated for each cluster of cells. Red bars show the most significant pathway activity in epithelial clusters 1–3. **C** Expression levels for selected EBV and human genes in different human cell types. *RPMS1* and *LMP1* were used as indicators for EBV-high cancer cells, as discovered by scRNA-seq data. Expression levels of the genes upregulated in epithelial clusters 1–3 were selected in the plot. Red indicates a higher expression level, and blue indicates a lower expression level. **D** Enrichment scores of NPC-related cancer hallmarks from the Human Molecular Signatures Database (MSigDB) for different cell types. The selective hallmarks with statistical enrichment in epithelial clusters 1–3 were visualised in the plot. Red indicates a positive enrichment score, and blue indicates a negative enrichment score. **E** Ectopic expression of NPC-related genes in RPMS1-expressing EBV+ NPC cells determined via scRNA-seq. EBV-high clusters were identified via scRNA-seq were subclustered ($n_{\mathrm{cells}} = 1696$) via UMAP into RPMS1-high ($n = 1112$) and LMP1-high (or RPMS1-low) ($n = 584$) groups on the basis of expression of the EBV latency genes RPMS1 and LMP1, respectively. Genes commonly found via both scRNA-seq and single-cell spatial transcriptomics were selected. Cells with high expression are coloured blue, and those with low expression are coloured yellow. **F** Epithelial cell proportions in MET and non-MET patients. The proportion of cells in each epithelial cell cluster was examined for MET ($n_{\mathrm{cells}} = 89{,}743$) and non-MET ($n_{\mathrm{cells}} = 14{,}104$) patients. Epi-1/−2/−3 clusters account for about 25% of the total epithelial cells in MET patients. N denotes normal epithelial cells, and T denotes tumour epithelial cells.

Table 1 | Clinical characteristics for the multicentre multiomics study in association with metastasis

Characteristic[a]	HK cohort (*N* = 77)									GZ cohort (*N* = 100)				
	All		MET		nonMET			All		MET		nonMET		
	Number	%	Number	%	Number	%	*P* value	Number	%	Number	%	Number	%	*P* value
Patients enrolled	77	100	39	50.6	38	49.4		100	100	36	36	64	64	
Age at diagnosis (years)							0.511							0.269
Median	51		50		52.5			42		37		44		
Range	31 to 81		31 to 80		32 to 81			8 to 77		8 to 77		24 to 73		
Relapse status	77						0.026*	100		37		63		0.076
Yes	24	31.2	17	70.8	7	29.2		30	30	14	46.7	13	43.3	
No	53	68.8	22	41.5	31	58.5		70	70	22	31.4	51	72.9	
Gender	77						0.358	100		36		64		0.72
Male	62	80.5	33	53.2	29	46.8		75	75	29	38.7	46	61.3	
Female	15	19.5	6	40	9	60.8		25	25	7	28.0	18	72.0	
TNM stage, AJCC (version 7)	77						1	100		36		63	63	0.079
Early stage (I–II)	5	6.5	3	60	2	40		14	14	2	14.3	12	85.7	
Late stage (III–IV)	72	93.5	36	50	36	50		86	86	34	39.5	52	60.5	
EBV DNA copy (baseline)	29						0.163	80		24		56		0.223
Low (≤10,000 IU/mL)	27	93.1	10	37.0	17	63.0		59	73.8	14	23.7	45	76.3	
High (>10,000 IU/mL)	2	6.9	2	100.0	0	0.0		21	26.3	9	42.9	12	57.1	
EBV DNA copy (within 6 months after treatment)	32						0.194	58		23		35		<0.001*
Not detectable (≤40 IU/mL)	26	81.3	9	34.6	17	65.4		46	79.3	13	28.3	33	71.7	
Detectable (>40 IU/mL)	6	18.8	4	66.7	2	33.3		12	20.7	10	83.3	2	16.7	
Chromosome instability (CIN) score	77		39		38		0.143	99		37		62		0.103
Low (≤0.1)	56	72.7	25	44.6	31	55.4		69	68.7	21	65.2	48	34.8	
High (>0.1)	21	27.3	14	66.7	7	33.3		30	30.3	15	43.3	15	56.7	
KDM5B pathway activity	77		39		38		0.004*	100		36		64		0.037*
Low (>−0.3)	31	40.3	9	29	22	71		40	40	9	22.5	31	77.5	
High (<−0.3)	46	59.8	30	65.2	16	35.8		60	60	27	45.0	33	55.0	

[a] Data are presented as number (percentage) for categorical variables and median (range) for continuous variables. Comparisons between MET and non-MET groups within each cohort were performed using the chi-square test or Fisher's exact test (if $n < 5$ in one or more groups) for categorical variables, and the Mann–Whitney *U*-test for continuous variables. A *p*-value ≤ 0.05 was considered statistically significant. Asterisks (*) indicate statistically significant differences.



Board of Sun Yat-Sen University Cancer Centre (**G2021-110-01**) for the GZ cohort.

Cell culture

Two nasopharynx-derived normal immortalised squamous epithelial cell lines, NP69SV40T (NP69, RRID: CVCL_F755) and NP460-tert (NP460, RRID: CVCL_X205), were used in this study. The EBV+ NPC cell lines used included C17 (RRID: CVCL_VT47), C666-1 (RRID: CVCL_M597), and NPC43 (RRID: CVCL_UH64). The EBV− NPC cell lines used included NPC53 (RRID: CVCL_UH65) and NPC38 (RRID: CVCL_UH63). These cell lines were sourced from the NPC tissue bank at the University of Hong Kong. NPC43noEBV, HK1, and HK1EBV cells were kindly provided by Dr. Anna Tsang from The Chinese University of Hong Kong. NPC43noEBV cells originated from the EBV+ NPC43 cell

Fig. 5 | KDM5B is a risk factor for MET, and its targets are involved in chromosomal instability and DNA damage response. A GSEA of EBV tethering-mediated enhancer activation of KDM5B targets in MET patients. Genes were ranked based on differential expression analysis using the bulk RNA-seq data from MET patients and non-MET patients in the HK and GZ cohorts to calculate the enrichment score of EBVIR–enhancer–KDM5B signature. **B** Human whole-transcriptome atlas (WTA) spatial analysis of 26 clinical specimens (MET: $n = 16$, nonMET $n = 8$). The yellow frame indicates the region of interest (ROI). A heatmap shows differentially expressed genes (DEGs) between the MET and non-MET samples. A two-tailed Mann–Whitney *U*-test with log2FC > 0.6 or < −0.6 and adjusted p-values ≤ 0.05 estimated by permutation test identified DEGs. The DEGs in both cohorts were evaluated. **C** Putative transcription factors associated with metastasis-related genes. The putative transcription factors among metastasis-associated genes were evaluated in the HK ($n_{\text{genes}} = 236$) and GZ ($n_{\text{genes}} = 291$) cohorts. A scatter plot shows the odds ratio against $-\log_{10}$ (p-value). The combined score was calculated as $\log_{10}$ (p-value) from one-tailed Fisher's exact test × the Z score for deviation from the expected rank. **D** Functional annotation of genes expressed in the MET patients. A metastasis-associated gene set ($n = 82$) common to both cohorts was identified. The hypergeometric test for Reactome signalling pathways and GO terms for biological processes with Benjamini–Hochberg correction has been used to examine the significance. All significant terms had an adjusted p-value ≤ 0.05, and the font size is based on the odds ratio. **E** Chromosomal instability (CIN) scores for EBV tethering-mediated enhancer activation of KDM5B targets in both the HK and GZ cohorts. A cut-off of 0.1 was assigned and used to identify CIN-high (HK: $n = 21$ and GZ: $n = 30$) and CIN-low patients (HK: $n = 13$ and GZ: $n = 29$). Student's t-test was used to test the significance of the difference between the two groups. A two-tailed unpaired t-test was performed to examine the significance between CIN-high and CIN-low patients in both cohorts. The bar plot with error bars indicates mean ± standard deviation (SD) of the data. CIN-high and CIN-low patients have shown a significant difference in HK and GZ cohorts (p-value = 0.04 and p-value < 0.001, respectively).

line with undetectable EBV DNA. HK1EBV cells were infected with EGFP-neo-resistant EBV-infected Akata cells, and G418 was used for EBV+ cell selection[50]. Namalwa cells (RRID: CVCL_1841) were obtained from the NPC tissue bank at the University of Hong Kong.

NP69 cells were cultured in Keratinocyte SFM (GIBCO Cat# 10744019) supplemented with bovine pituitary extract (BPE), 1 ng/ml epidermal growth factor (EGF), and 1% penicillin. NP460 cells were cultured in a mixture of serum-free defined keratinocyte medium and EpiLife (1:1) with additional supplements (GIBCO Cat# MEPI500CA), 10% foetal bovine serum (FBS) (GIBCO Cat# 26140079), and 1% penicillin (GIBCO Cat# 10378016). C17, NPC43, NPC43noEBV, NPC53, and NPC38 cells were cultured in RPMI-1640 medium supplemented with 10% FBS, 1% penicillin, and the ROCK inhibitor Y-27632 at 2 mM (Enzo Life Sciences Cat# BML-EU299). YCCEL1 and C666 cells were cultured in RPMI-1640 medium (GIBCO, Cat# 31800022) supplemented with 10% FBS and 1% penicillin. HK1 cells were cultured in RPMI-1640 medium (GIBCO, Cat# 31800022) supplemented with 10% FBS and 1% penicillin. HK1EBV cells were cultured in RPMI-1640 (GIBCO, Cat# 31800022) medium supplemented with 10% FBS, 1% penicillin, and 500 µg/ml geneticin, a selective antibiotic (G418 sulfate Cat# 10131027). The cells were maintained in a 37 °C incubator with 5% CO_2. Cell authentication by short tandem repeat (STR) profile analysis (Genetica & the AoE NPC Cell Line Repository) was performed, and the cultured cells were regularly tested and found to be free from mycoplasma contamination.

NPC patient-derived xenografts (PDXs)

The EBV+ PDXs were established previously and provided by Professor George Tsao's laboratory, including C15, Xeno23, Xeno32, Xeno47, Xeno76 and NPC113[31,51]. The EBV+ PDXs have been used in OmniC and ATAC-seq experiments. Experimental procedures and all animal welfare were approved by the Centre for Comparative Medicine Research (CCMR), the University of Hong Kong. Animals and the equipment used on animals are under routine monitoring with up to the performance standards as listed by the Guide for the Care and Use of Laboratory Animals, 8th Edition, National Research Council. Animals were housed in individually ventilated cages under a 12:12 dark–light cycle within environmentally controlled rooms and were fed ad libitum with laboratory diet manufactured by LabDiet, USA. Temperature is regulated at 16–26 °C with a relative humidity range of 30–70%. Inbred NOD.CB17-Prkdcscid/J (NOD/SCID) from The Jackson Laboratory, USA feeding with LabDiet® 5LG4 (Breeding Diet), LabDiet® 5053 (Maintenance Diet) were used under an AAALAC International accredited programme at the Centre for Comparative Medicine Research, HKU under Specific Pathogen Free (SPF) conditions. Either sex can be used for NOD/SCID mice at the age of 8–12 weeks for the subcutaneous implantation by surgical procedures for EBV+ PDXs. A skin incision was introduced dorsally at the flank of the mice to insert the explanted xenograft. For an animal carrying a single tumour, the maximum permitted mean diameter in mice was 12 mm, and it was not exceeded. NOD/SCID mice were maintained under CCMR for 1–4 months to reach enough tumour size to conduct the experiment until euthanasia. NOD/SCID mice were euthanised by ketamine (300–360 mg/kg) combined with xylazine (30–40 mg/kg) by intraperitoneal (IP) injection. The volume administered will not exceed 10 ml/kg. Research only proceeded following review and approval from the HKU Committee on the Use of Live Animals in Teaching and Research (CULATR# 23–213).

Measurement of EBV DNA copy numbers in cell lines and clinical tumour tissues

A known amount of cellular EBV DNA from Namalwa cells (RRID: CVCL_1841) was used to establish a standard calibration curve. This calibration curve was then employed to estimate the EBV copy number in the cell lines, utilising qPCR conducted in triplicate. The BamHI-W target sequence from the EBV genome (forward primer: 5′-CCCAACACTCCACCACACC-3′, reverse primer: 5′-TCTTAGGAGCTGTCCGAGGG-3′) was specifically amplified during qPCR[52].

Omni-C experiment

EBV+ PDXs, including C15, Xeno23, Xeno32 and NPC113 have been used for Omni-C experiment. Instead of using restriction enzymes to cut in the traditional Hi–C, the genome-wide fragmented chromatin was created by an endonuclease enzyme in Omni-C. A Dovetail® Omni-C® Library Preparation Kit (Dovetail Genomics, Cat# 21005) was used for sequencing library preparation. A million cells were freshly harvested and subjected to chromatin cross-linking using disuccinimidyl glutarate (DSG) (Thermo Scientific, Cat# 20593) and 37% formaldehyde. Chromatin digestion was carried out using an endonuclease enzyme, the digestion efficiency was monitored, and quality control was performed with a bioanalyzer. With approximately 80% of the fragments falling within the 100–2500 bp range, the captured and ligated chromatin underwent reverse cross-linking and DNA purification. Libraries were subsequently constructed using the Dovetail™ Library Module for Illumina (Dovetail Genomics, Cat# 25004). The final libraries were sequenced with 150 bp paired-end reads on the NovaSeq 6000 platform at the Centre for PanorOmic Sciences (CPOS) at the University of Hong Kong.

Omni-C analysis pipeline

The BWA (v0.7.17)-MEM alignment method was employed to align reads to both the human genome (hg19) and the EBV genome (NC_007605). Valid ligation events were recorded using the parse function in pairtools (v1.1.3), with ligation junctions documented with specific parameters, including --min-mapq 40, --walks-policy 5unique, and --max-inter-align-gap 30. Then, pairtools was utilised to sort reads, eliminate PCR duplicates, and generate.pairs and.bam files via SAMtools (v1.10). Quality control and reproducibility checks were performed with replicate samples. The quality and reproducibility

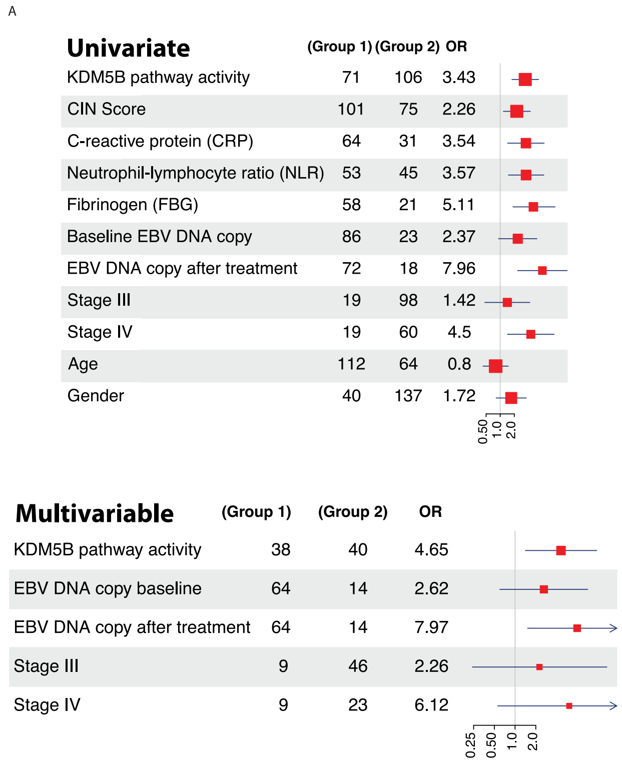


Fig. 6 | Univariate and multivariate logistic regression analyses revealed that KDM5B pathway activity is an independent risk predictor for NPC metastasis. **A** The HK ($n = 77$) and GZ ($n = 100$) cohorts were included in the analysis. Univariate logistic regression analyses were performed to assess the association between individual variables and the binary outcome. KDM5B pathway activity; OR = 3.43, 95% CI (2.54, 4.63) with p-value = 0.0003, CIN Score; OR = 2.26, 95% CI (1.68, 3.05) with p-value = 0.0076, C-reactive protein (CRP); OR = 3.54, 95% CI (2.31, 5.44) with p-value < 0.0001, Neutrophil–lymphocyte ratio (NLR); OR = 3.57, 95% CI (2.40, 5.31) with p-value = 0.3173, Fibrinogen (FBG); OR = 5.11, 95% CI (3.10, 8.42) with p-value < 0.0001, Baseline EBV DNA copy; OR = 2.37, 95% CI (1.50, 3.75) with p-value < 0.0001, EBV DNA copy after treatment; OR = 7.96, 95% CI (4.75, 13.34) with p-value < 0.0001, Stage III; OR = 1.42, 95% CI (0.87, 2.32) with p-value < 0.0001, Stage IV; OR = 4.5, 95% CI (2.69, 7.54) with p-value < 0.0001, Age; OR = 0.8, 95% CI (0.59, 1.09) with p-value < 0.0001, Gender; OR = 1.72, 95% CI (1.21, 2.45) with p-value < 0.0001. Multivariable logistic regression analysis was conducted to evaluate the independent associations between selected clinical and molecular variables and the binary outcome. Group 1 is the reference group for analysis. KDM5B pathway activity; OR = 4.65, 95% CI (2.98, 7.25) with p-value < 0.0001, EBV DNA copy baseline; OR = 2.62, 95% CI (1.47, 4.67) with p-value = 0.001, EBV DNA copy after treatment; OR = 7.97, 95% CI (4.47, 14.21) with p-value < 0.0001, Stage III; OR = 2.26, 95% CI (1.11, 4.62) with p-value < 0.0001, Stage IV; OR = 6.12, 95% CI (2.83, 13.23) with p-value < 0.0001. The plots show odds ratios (ORs) in the centre with 95% confidence interval (CI) of the data. The size of the red square denotes the sample size.

checking for the Omni-C data were then merged into a single file for subsequent analysis. To ensure comparability, the sample reads were normalised through random sampling, resulting in approximately 200 million paired reads.

Differential compartment analysis of Omni-C data

HOMER (v5.1), with the analyzeHiC tool, was used to construct a background model at a 50 kbps resolution for normalisation. Subsequently, runHiCpca.pl was employed for principal component analysis (PCA), calculating the first principal component (PC1) with additional parameters at a 50 kbps resolution and a 50 kbps window size. The active histone modification H3K4me3 from CUT&RUN-seq and gene density have been used to segment the active and inactive compartments, where the active compartment is PC1 > 0 and the inactive compartment is PC1 < 0. We included the two NPE (NP69 and NP460), three EBV+ NPC (C17, C666 and NPC43), and three EBV− NPC (NPC53, NPC38 and HK1). For the differential compartment analysis, it has been performed in comparison between EBV+ NPC and EBV− NPC cells to identify differential compartmentalisation. In this context, B-to-A (A-to-B) compartment switching was denoted by the transformation of a negative (positive) PC1 value in the reference samples into a positive (negative) PC1 value when the sample was compared to the reference. A–A (B–B) enhancement of the compartment was indicated when the compared samples exhibited a greater magnitude change in PC1 values than the reference samples when both were positive (negative) PC1 values. Conversely, A–A (B–B) reduction of the compartment was indicated when the compared samples demonstrated a lower magnitude change in PC1 values than the reference samples when both were positive (negative) PC1 values, with a significance threshold of adjusted p-value ≤ 0.05. We verified the identification of six different classifications of compartments in Supplementary Fig. 3b.

Identification of EBVIRs at the oriP, EBNA1 and RPMS1 tethering regions

A contact frequency matrix was extracted via hic-straw[53]. The matrix underwent normalisation through the Knight–Ruiz (KR) matrix balancing method. Normalised contact counts from both the EBV genome and human DNA interchromosomal interactions at a 1 kbp resolution were extracted. Cross-species interacting reads were then ranked on the basis of their normalised contact frequency. The high confident top 5% of contact frequencies was selected, and a chi-square test was individually performed to assess significant binding (p-value ≤ 0.01) to identify oriP–EBVIRs, EBNA1–EBVIRs and RPMS1–EBVIRs.

Odds ratio, enrichment of EBVIRs in the compartments

The odds ratio (OR) of high confident EBVIRs with enhanced or reduced compartments was calculated as follows:

$$\mathrm{ORs} = \frac{\frac{\text{Number of peaks in EBVIRs with a change of compartment}}{\text{number of peaks that are not in EBVIRs with a change of compartment}}}{\frac{\text{Number of peaks in EBVIRs with no change of compartment}}{\text{Number of peaks that are not in EBVIRs with no change of compartment}}} \quad (1)$$

Identification of chromatin loops and Aggregated peak analysis (APA)

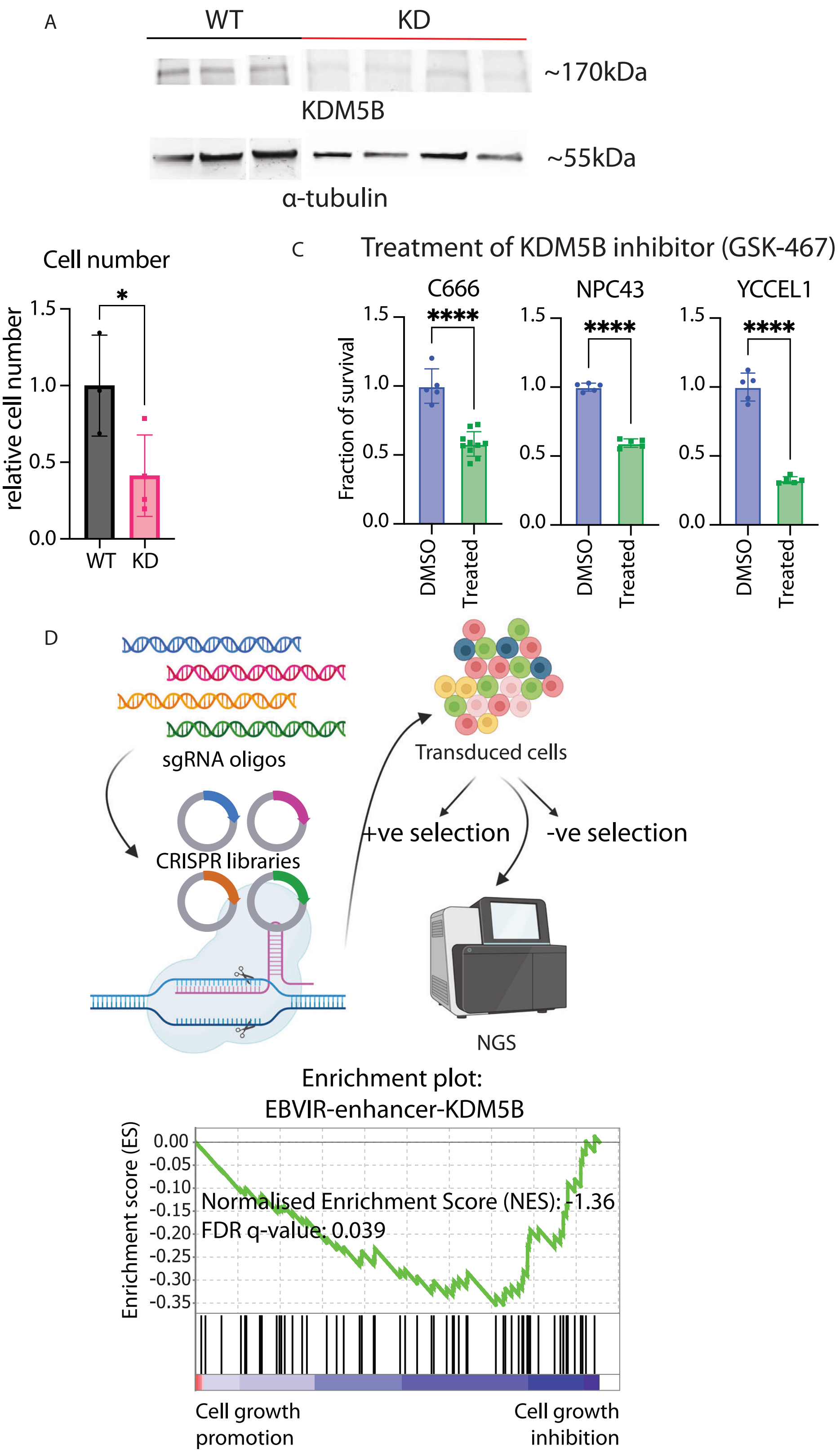
Juicer (v1.6).hic files were used with Moustache (v1.0.1)[54] for loop calling at multiple levels (10, 100, 250 and 500 kbps) across samples with the parameters -pt 0.1 -st 0.3 -norm KR. The loop calling results from the NPE, EBV+ NPC, and EBV− NPC cell lines were summarised by bedtool (v2.31.0) merged. The summarised loop calling regions were used for APA. APA is a standard method to measure enrichment of chromatin loops in a contact matrix. The APA was conducted by Juicer tools with the parameters of -r 10,000 -k KR. The APA peak-to-lower-left (P2LL) score was calculated on the basis of the ratio of the central pixel to the mean of the pixels at the lower left corner. Higher scores indicate higher enrichment of interactions[55,56].

Calculation of the AT content at EBV–oriP–EBVIRs, EBV–EBNA1–EBVIRs and EBV–RPMS1–EBVIRs

The high-confidence EBVIRs were extracted from each sample. EBV annotation was based on the gene transfer format (GTF) in the EBV portal[57]. Individual tethering sites were selected, and the AT content was calculated with HOMER (v4.9) and the -cpg parameter (1-CG%).

Transposase-accessible chromatin with sequencing (ATAC-seq) experimental procedure and data analysis

This study includes NP69, NP460, C17, C666, NPC43, NPC53, NPC38, HK1, C15, Xeno23, Xeno32, Xeno47, Xeno76, NPC113, YCCEL1 cells for ATAC-seq. The procedure has been previously described in detail[17]. Briefly, 50,000 cells were lysed in a cold lysis buffer. The cell lysates were incubated with Tn5 transposase (Illumina, Cat# E7645S) at 37 °C for 30 min. Libraries were sequenced on the NovaSeq 6000 platform. Data analysis was performed following the ENCODE ATAC-Seq pipeline



(v1.0). The ATAC-Seq RPKM signals were normalised with deeptools bamCoverage with -binsize 1[58].

Cleavage under targets & release using nuclease sequencing (CUT&RUN-seq)

This study includes NP69, NP460, C17, C666, NPC43, NPC53, NPC38, HK1 cells for CUT&RUN-seq. The CUTANA™ ChIC/CUT&RUN Kit (EpiCypher, Cat# 14-1048) was used in this experiment. A total of 500,000 cells were harvested and bound to concanavalin A-coated (ConA) magnetic beads. The cell membranes were permeabilised with 5% digitonin, allowing the primary antibody to bind histones with posttranslational modifications (PTMs) or transcription factors (TFs). CUTANA™ pAG-MNase (EpiCypher, Cat# 15-1016) at a volume of 2.5 μL was incubated with the samples for 10 minutes at room temperature to

Fig. 7 | EBV hijacks host factor–KDM5B to provide host cancer cell survival. **A** Generating KDM5B wild-type (WT) ($n = 3$) and knockdown (KD) ($n = 4$) cells in C666. A western blot result for KDM5B at ~170 kDa and α-tubulin control at ~55 kDa demonstrate the success of establishing KDM5B–KD clones. **B** Quantification of cells in KDM5B–WT and –KD cells. An equal number of cells were cultured for 7 days. A cell count chamber was used to calculate the number of cells in WT and KD cells. Biological replicates have been performed for WT clones ($n = 3$) and KD clones ($n = 4$). The number of cells was normalised to WT. An unpaired t-test was used to test the significance, showing KD cells have approximately 60% cell growth reduction with p-value = 0.047. * p-value ≤ 0.05. The bar plot with error bars indicates mean ± standard deviation (SD) of the data. **C** MTT assay results for 7-day treatment of KDM5B inhibitor (GSK-467) in C666, NPC43 and YCCEL1 cells. The experiments were performed with technical replicates ($n = 5$) for both DMSO control and KDM5B treatment groups. The KDM5B inhibitor results were normalised to their matched DMSO control. A two-way ANOVA test was performed with Bonferroni correction, ****adjusted p-value ≤ 0.0001. A concentration of 75 μM of KDM5B inhibitor treatment showed significant suppression of cancer cell growth in EBV+ NPC and EBVaGC cells. The bar plot with error bars indicates mean ± standard deviation (SD) of the data. **D** CRISPR screening revealed that the EBVIR–enhancer–KDM5B signature suppresses tumour cell growth. A genome-wide loss-of-function (LOF) analysis employing the CRISPR screening approach was conducted. CRISPR libraries were transduced into the cells, and cells were subsequently selected by puromycin. The sgRNA sequences were amplified and sequenced by the Illumina platform. The LOF of a gene conferring a growth advantage, characterised by a higher abundance of sgRNAs, was quantified based on the sequencing. Conversely, the LOF of a gene lacking a growth advantage was determined by sequencing with a significant reduction of the sequencing signals (Created in BioRender. Dai, W. (2025) https://BioRender.com/7e447e1). The GSEA result showed significant negative enrichment of EBVIR–enhancer–KDM5B signature ($n = 64$), NES = −1.36, FDR q-value = 0.039, suggesting LOF of EBVIR–enhancer–KDM5B signature strongly inhibited cancer cell growth.

target the primary antibody heavy chain at the region of interest. A 100 mM calcium chloride solution was used to initiate DNA digestion for 2 h at 4 °C. The enriched DNA was purified and prepared for sequencing libraries. One to ten nanograms of enriched DNA was collected to generate the sequencing libraries using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (NEBNext, Cat# E7370L). SNAP-CUTANA™ K-MetStat Panel (EpiCypher, Cat# 19–1002) spike-in controls with mono-, di-, and trimethylation at H3K4, H3K9, H3K27, H3K36 and H4K20 and unmodified controls were spiked into positive control (H3K4me3) and negative control (IgG) antibodies to monitor assay success. A volume of 1 μL of CTCF (Millipore, Cat# 07-729), H3K4me3 (EpiCypher, Cat# 13-0041), H3K27ac (Abclonal, Cat# A7253), H3K27me3 (Invitrogen, Cat# MA5-11198), and IgG (EpiCypher, Cat# 13-0042) antibodies were used for CUT&RUN experiment. The libraries were sequenced with approximately 20 million (150 bp paired-end reads) sequencing reads using the NovaSeq 6000 platform at the Centre for PanorOmic Sciences (CPOS) at the University of Hong Kong.

CUT&RUN analysis pipeline

Sequencing adaptors were trimmed via Trimmomatic (v0.39). Bowtie2 (v2.5.0) was used to align reads to the human genome (hg19) and EBV genome (NC_007605) with the parameters -p 2 --dovetail --phred33. The reads were sorted, and PCR duplicates were removed with Picard in GATK (v4.2.3.0). MACS2 (v2.2.7.1) peak calling with matched control IgG was performed for each sample at -q 0.05. The CUT&RUN RPKM signals were normalised by deeptools bamCoverage with --binsize 1.

Spatial transcriptomic analysis with the 10X Visium spatial platform

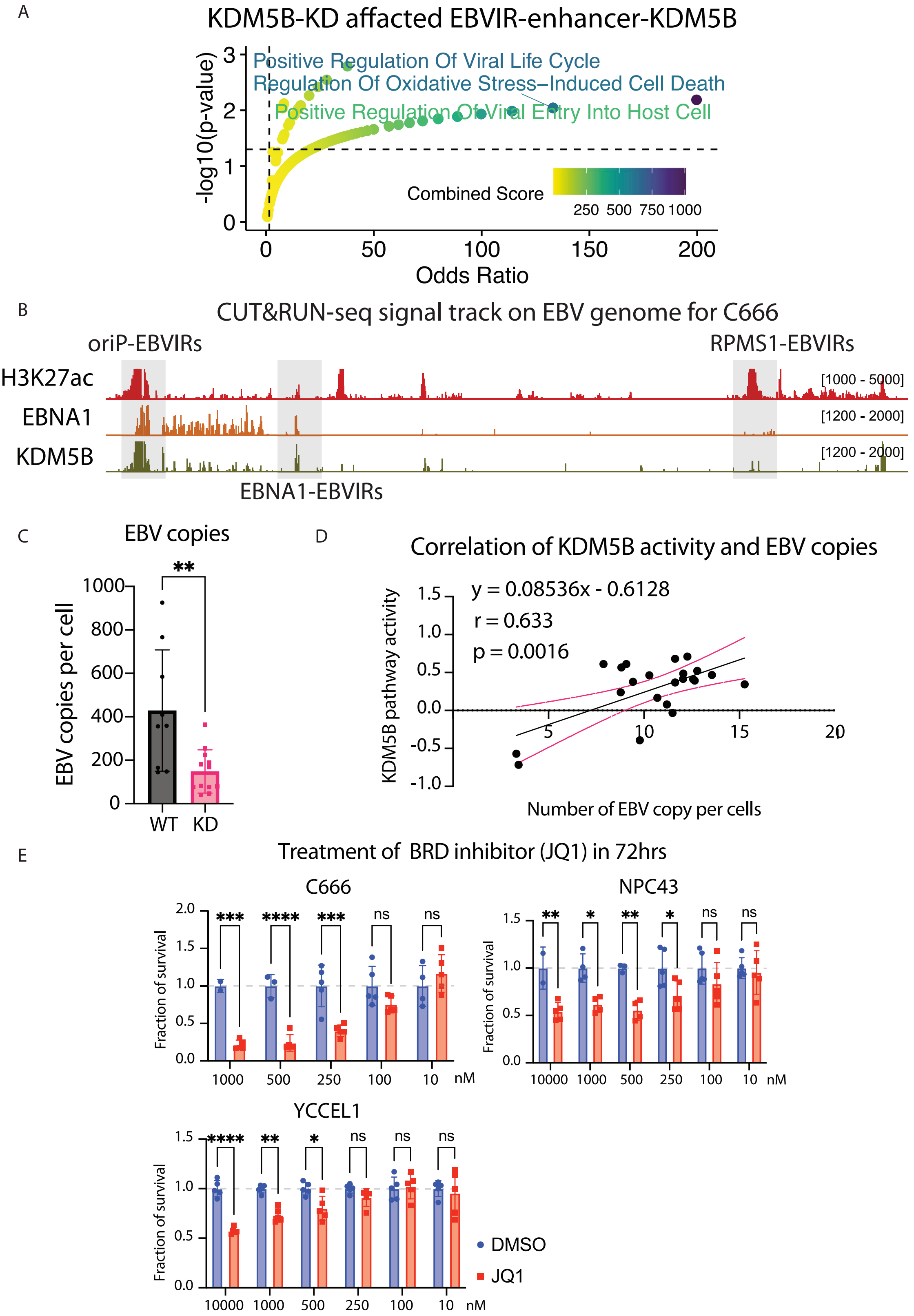
Spatial transcriptome data were generated with the 10X Visium spatial platform. Seurat (v4.1.2) with SCTransform was used to generate normalised data based on regularised negative binomial regression. The gene markers and parameters from the scRNA-seq data in a previous publication[17] were adopted to deconvolute the spatial data using FindTransferAnchors and TransferData with the SCT normalised data. The cell-type probability of the spatial transcriptome data is estimated by the 'anchor'-based integration method[59] based on the integrated scRNA-seq data from three studies, which were downloaded from CNGBdb database - CNP0000428[16], GEO database–GSE150825[15] and GSE162025[14]. Three scRNA-seq datasets were aggregated and well-annotated, including removal of batch effect, doublets removal, copy number variation calling and identification of cell types[17]. The expression of EBV-latent genes *LMP1* and *RPMS1* was used to identify the EBV-high NPC cells and EBV-low NPC cells. This 'anchor'-based integration predicts the probabilities of each cell type for each spot. The expression of KDM5B and average expression of the EBVIR–enhancer–KDM5B signature for each spot were calculated and visualised using the normalised data.

Spatial transcriptome analysis with GeoMx digital spatial profiler (DSP)

Clinically biopsied FFPE NPC samples were obtained from the Sun Yat-sen University Cancer Centre in Guangzhou and used to prepare a tumour microarray (GZ TMA) for spatial transcriptome sequencing (16 MET samples and 8 non-MET samples). The sex, age and clinical stage for all patients were equally distributed between the MET and non-MET groups. Spatial omics profiling of the GZ TMA was performed on the NanoString GeoMx® DSP platform. After slicing and pretreatment, 5-μm formalin-fixed paraffin-embedded (FFPE) tissue sections were fluorescently stained with DNA dye (SYTO 13) and the primary antibodies targeting CD45 (NanoString Technologies, clone ID: D9M8I) and PanCK (NanoString Technologies, clone ID: AE-1/AE-3), as well as the secondary antibody (Thermofisher Goat Anti-rabbit, A-11012). RNA probes coupled with photocleavable oligonucleotide barcodes were used to bind specifically to target mRNAs during the incubation process on the DSP platform on the basis of in situ hybridisation. Then, each segment was exposed to UV radiation, detaching the barcodes from the probes. The cleaved barcodes in units of segments were then collected into multiwell plates for next-generation sequencing on an Illumina sequencer. The counts of the oligonucleotide tags were finally used to calculate the abundance of the corresponding mRNAs. The GZ TMA contained 18,000+ RNAs. After quality control, batch effect removal, and Q3 normalisation using the top 25% of expressers to normalise across region of interest (ROI), differential gene expression analysis using Mann–Whitney U-test was performed between MET and non-MET groups. The genes with permutation q-value less than 0.05 were considered as significant candidates.

Single-cell spatial analysis with the CosMx™ spatial molecular imager (SMI)

Although the DSP provides spatial information for mRNA and protein expression, it does not achieve single-cell accuracy. To obtain more precise mRNA distribution data, SMI transcriptomics containing 1000+ RNA probes together with custom in situ hybridisation probes targeting RPMS1 and LMP-1 was conducted with two FFPE slides from one MET NPC patient and one non-MET NPC patient from Hong Kong Queen Elizabeth Hospital. SMI sequencing was also based on in situ hybridisation of RNA probes photochemically coupled with fluorescent barcodes in fields of view (FOVs). After incubation, the slides with hybridised probes were scanned to generate three-dimensional images with a subcellular resolution. The cleaved oligonucleotide tags were then directly decoded into transcripts in situ with registered x, y, and z spatial positions. Afterwards, the locations of the nucleus and cell boundaries were defined with DAPI staining and antibodies. Finally, each transcript was assigned to a defined individual cell. The data were processed via the AtoMx Spatial Information Platform (version 1.3.2) with the standard data analysis pipeline including



quality control, normalisation, principal component analysis, cell typing, marker gene identification, co-localisation analysis, neighbourhood analysis and ligand-receptor analysis. The normalisation was performed using Seurat "LogNormalize" default setting and adjusting for library size factors to ensure that cell specific total transcript abundance and distribution of counts did influence downstream data analysis. The cell types were identified using semi-supervised method based on the reference gene signature matrix ("ImmuneTumor_safeTME") provided by NanoString. The cells that did not fit in the signature matrix were split into ten clusters. The gene expression profile for each cluster were directly downloaded from AtoMx platform after cell typing. In the neighbourhood network analysis, k-Nearest Neighbour (KNN) was constructed based on the Euclidean distance in PCA space, then Shared Nearest Network (SNN) graph was constructed with

Fig. 8 | EBV hijacks the host factor−KDM5B to maintain its genome copies. **A** KDM5B regulates genes in response to the EBV life cycle and cell death. Bulk RNA-seq has been conducted for KDM5B−WT ($n = 3$) and KDM5B−KD ($n = 4$) with biological replicate clones. Pre-ranked genes based on $\log_2$ fold change of differentially expressed genes were subjected to GSEA. The downregulated genes in the EBVIR−enhancer−KDM5B signature were extracted for GO−BP analysis. A scatter plot shows the odds ratio against $-\log_{10}$ (p-value). The combined score ($\log_{10}$ (p-value) from one-tailed Fisher's exact test × the Z score for deviation from the expected rank) is indicated by colour. The terms with $p\text{-value} \le 0.05$ are highlighted. **B** Profile of EBNA1, KDM5B binding, and H3K27ac on EBV genome visualised in IGV browser. The RPKM normalised EBNA1 and KDM5B binding signals show the involvement of EBNA1 and KDM5B at EBV−oriP−EBVIR with H3K27ac signals. **C** Quantification of EBV copy per cell in KDM5B−WT and −KD cells. qPCR was performed in triplicate for each clone ($n_{\mathrm{WT}} = 9$ and $n_{\mathrm{KD}} = 12$). A two-tailed unpaired t-test was used to test the significance, showing a significant reduction of EBV copies in KD cells with $p\text{-value} = 0.004$. ** $p\text{-value} \le 0.01$. The error bar indicates mean ± standard deviation (SD) of the data. **D** Pearson correlation plots between the KDM5B pathway and cellular EBV copies in the patient biopsies. The Pearson correlation tests showed a strong positive correlation of KDM5B pathway activity to EBV copies ($n_{\mathrm{patients}} = 22$) with $r = 0.633$, $y = 0.08536x - 0.6128$, and $p\text{-value} = 0.0016$. Correlation is significant at the 0.05 level (two-tailed). The error bar indicates mean ± standard deviation (SD) of the data. **E** MTT assay results for 72 h of treatment of BRD inhibitor (JQ1) in C666, NPC43 and YCCEL1 cells. The experiments were performed with technical replicates ($n = 5$) for both drug treatment and DMSO control. A two-way ANNOVA test was performed with Bonferroni correction, *adjusted $p\text{-value} < 0.05$, **adjusted $p\text{-value} < 0.01$, ***adjusted $p\text{-value} < 0.001$ and ****adjusted $p\text{-value} < 0.0001$. The bar plot with error bars indicates mean ± standard deviation (SD) of the data.

edge weights between two cells based on the shared overlap in local neighbourhoods (default Jaccard distance cutoff: 0.06) and distant edges. Subsequently, Leiden clustering was applied to identify distinctive neighbourhood clusters (niches) with default radium of 50 μm.

Study subjects for the multicentre and multiomics experiment

Patients who were diagnosed with NPC at the Sun Yat-Sen University Cancer Centre in Guangzhou between 28 November 2011 and 11 November 2021 or at four public oncology hospitals in Hong Kong (the Queen Mary Hospital, Queen Elizabeth Hospital, Princess Margaret Hospital, and Pamela Youde Nethersole Eastern Hospitals) between 13 September 2010 and 16 November 2018 were recruited as the GZ cohort and the HK cohort, respectively, if they fulfilled the following criteria: a pathological diagnosis of nasopharyngeal undifferentiated nonkeratinising carcinoma; availability of prospectively collected clinical data; and availability of qualified samples from primary nasopharyngeal foci. All the enroled patients had received first-line treatments according to NCCN guidelines and were followed through medical visits and telephone interviews. Cancer progression was confirmed by histopathological examination or at least two other tests, including cytopathologic diagnosis, computed tomography, magnetic resonance imaging, positron emission tomography, single-photon emission computed tomography, ultrasound examination and endoscopy. Patients with any confirmed distant site were diagnosed with distant metastasis. All participants in the GZ cohort were followed until 31 May 2023, and all participants in the HK cohort were followed until 1 December 2021. DMFS was defined as the period from the diagnosis of NPC to the first onset of metastasis or death or the last follow-up date. Patients with NPC who developed distant metastasis within three years after treatment were classified as metastatic NPC (MET-NPC), while those without early metastasis were designated as non-metastatic NPC (nonMET-NPC).

DNA/RNA extraction for WES and RNA sequencing

DNA and RNA were extracted from fresh-frozen clinical NPC biopsies using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen Cat# 80224), whereas DNA was extracted from blood samples using the AllPrep Blood Mini Kit (Qiagen). The integrity of the DNA samples was evaluated using 0.8% agarose gel electrophoresis. The quality and quantity of the DNA/RNA samples were evaluated with a Nanodrop 1000 instrument (Thermo Scientific), a Qubit fluorometer and a bioanalyzer (Agilent). DNA samples without degradation, with an OD260/280 ratio greater than 1.8 and a concentration greater than 20 ng/μl, were subjected to WES with 200× coverage and 24 GB raw data output per sample. The RNA samples with an OD260/280 ratio greater than 2.0 and whose concentration was greater than 50 ng/μl with acceptable integrity (RIN) were subjected to library preparation and RNA sequencing via the ribosomal RNA depletion approach, with 15 GB of raw data output per sample. Sequencing was performed using the paired-end 150 strategy on the Illumina NovaSeq platform by Novogene.

WES data processing

The Genome Analysis Toolkit (GATK) Best Practices pipeline (v4.1.8) was used to process the WES data. In brief, the raw data were aligned with the human genome (hg38) via the Burrows–Wheeler-Alignment (BWA)-MEM method, and recalibrated BAM files with marked duplicates were obtained. The quality of the BAM files, specifically in terms of the percentage of unique reads passing the Illumina filter, the percentage of off-bait bases, the mean bait coverage, the percentage of usable bases on bait and the 80-base penalty, was verified via Picard (v2.17.4) in the GATK pipeline. Germline variant calling was performed on both tumour and blood WES data using GATK HaplotypeCaller in GVCF mode. Somatic mutations in matched tumour-normal mode were identified using the Mutect2 pipeline. Biallelic somatic mutations were selected using SelectVariants for downstream analysis. The functional impact of somatic mutations was annotated via ANNOVAR (v2019Oct24).

Estimation of the weighted genome instability index (wGII) for the CIN score

The genome instability index (GII)[60,61] is the percentage of single nucleotide polymorphisms (SNPs) across the genome present at an aberrant copy number relative to the baseline ploidy of the sample, considering chromosome size variations to ensure that large chromosomes do not have greater effects on scores than small chromosomes do. The percentage of aberrant SNPs was calculated separately for each chromosome, and the mean percentage aberration was then calculated across all 22 chromosomes. This adapted wGII was used as the CIN score in this study.

Bulk RNA sequencing data analysis

The quality of the raw data was evaluated via FastQC (v0.11.8). The raw reads were aligned with the reference human genome (hg38) and EBV genome (NC_007605), and BAM files were generated by STAR (v2.7.5)[62] and salmon (v1.3.0)[63]. The quality of the BAM files, specifically in terms of the percentage of mRNA bases, percentage of ribosomal bases, percentage of usable bases and median 5′ to 3′ bias, was evaluated using Picard. The raw read counts for each gene were estimated by HTSeq (v0.9.1)[64], and the tumour purity of all the samples was estimated via the R package estimate (v1.0.13).

Immunohistochemical (IHC) staining

Formalin-fixed paraffin-embedded (FFPE) blocks from the enroled primary NPC lesions were sectioned at 4 μM for immunohistochemical (IHC) staining of anti-KDM5B (dilution 1:400, Cat# HPA027179, Sigma-Aldrich). Briefly, FFPE slides were deparaffinised and baked at 95 °C for 45 min for epitope retrieval. Then the slides were stained using the two-step method. Finally, the HKU pathologist quantified the expression of KDM5B in the whole slide scanning images, blinded from grouping information.

Differentially expressed gene (DEG) analysis

DEG analysis was carried out with DESeq2 (v1.36.0). Raw read counts were normalised with the median of ratios method in DESeq2. Analysis of differential expression in each cohort was performed with the relapse condition as a design factor to compare expression levels in the MET and non-MET patient groups.

Estimation of HALLMARK cancer pathway activities by GSVA

Pathway activities were calculated with the R package GSVA (1.50.1) on the basis of the raw read count for samples from each cohort via the single-sample GSEA (ssGSEA) normalisation method. Hallmark gene lists from the MSigDB Human collection (2023.1.Hs) were used.

Univariate and multivariable logistic regression analyses

Univariate and multivariate logistic regression analyses were performed via IBM SPSS (v26). In the univariate logistic regression analysis, KDM5B pathway activity, the CIN score, the neutrophil–lymphocyte ratio (NLR), C-reactive protein (CRP), fibrinogen (Fgb), age, sex, overall stage and plasma EBV DNA copy number at baseline and after treatment were used as categorical variables. The cut-offs for KDM5B pathway activity and the CIN score in the HK cohort were determined using non-MET and blood samples as the reference group, respectively, and were directly applied to the GZ cohort. In the multivariable logistic regression analysis, only unrelated variables, including KDM5B pathway activity, overall stage, and plasma EBV DNA copy number at diagnosis and after treatment, were considered.

Gene set enrichment analysis (GSEA)

GSEA (v4.3.2) was used for the analysis. A normalised vsd gene expression counts matrix from the RNA-seq results from both the HK and GZ cohorts was used to test the enrichment of the EBVIR–enhancer–KDM5B signature over background genes. The significance of enrichment in MET-NPC was tested with 10,000 permutations. For the GSEA in the CRISPR screening result of C666, Preranked GSEA was carried out based on the log2-fold change (L2FC) of the genes to test the enrichment of the EBVIR–enhancer–KDM5B signature over background genes. The significance of enrichment in cell growth inhibition was tested with 10,000 permutations.

MTT assay

An equal number of 20k cells was seeded into the 96-well plates in culture for treatment and control samples. Each condition of the experiments was performed in 5 replicates for treatment and control samples. After cells were cultured for 3 days and 7 days, MTT reagent (5 mg/ml) (Invitrogen, Cat# M6494) mixed with distilled water was added into each well for the assay. Followed by incubating at 37 °C, 5% CO_2 for 1 h, the cells were washed with PBS. A 100 μL of DMSO is added to the MTT assay to dissolve the formazan crystals. Detection is done at 570 nm wavelength in the spectrophotometer. The calculation is normalised to concentration-matched DMSO controls.

Doxycycline-inducible shRNA cell line construction

To establish a doxycycline-inducible knockdown system, shRNA sequences specifically targeting KDM5B were cloned into Tet-pLKO-puro vectors. Lentiviral particles were then generated by co-transfecting the individual vectors with psPAX2 and pMD2.G into HEK293T cells. The supernatant was collected 48 h post-transfection. Subsequently, C666-1 cells were infected with these lentiviral particles, which contained 8 mg/ml polybrene, for 48 h to facilitate the creation of stable KDM5B knockdown cell lines. These cells were then subjected to a selection process using puromycin at a final concentration of 2 mg/ml for an additional 5 days. To induce the transcription of shRNAs and the cleavage of endogenous KDM5B, the cells were treated with doxycycline at a final concentration of 1 mg/ml. The medium, containing doxycycline, was refreshed every 24 h to maintain the activity of the inducible promoter. The primers used in this procedure are provided in the Table below[44].

Targets	Sequence (5′−3′)
shRNA-KDM5B-1-Forward primer	CCGGGGAGACTAGTAAGCACTATCTCGAGATAGTGCTTACTAGTCTCC TTTTTG
shRNA-KDM5B-1-Reverse primer	AATTCAAAAAGGAGACTAGTAAGCACTATCTCGAGATAGTGCTTACTAGTCTCCC
shRNA-KDM5B-2- Forward primer	CCGGGGCAGTAAAGGAAATCGAACTCGAGTTCGATTTCCTTTACTGCCTTTTTG
shRNA-KDM5B-2- Reverse primer	AATTCAAAAAGGCAGTAAAGGAAATCGAACTCGAGTTCGATTTCCTTTACTGCC
shRNA-KDM5B-3- Forward primer	CCGGGGGTCTCGGACCGAGCGGACTCGAGCCGCTCGGTCCGAGACCCTTTTTG
shRNA-KDM5B-3- Reverse primer	AATTCAAAAAGGGTCTCGGACCGAGCGGACTCGAGTCCGCTCGGTCCGAGACCCC
shRNA-KDM5B-4- Forward primer	CCGGGCCTCGGTCCCGGAGCCCACTCGAGTGGGCTCCGGGACCGAGGCTTTTTG
shRNA-KDM5B-4- Reverse primer	AATTCAAAAAGCCTCGGTCCCGGAGCCCACTCGAGTGGGCTCCGGGACCGAGGCC

CRISPR library screening

C666 cells were transduced with a lentivirus expressing Cas9 to generate C666–Cas9 cells. The Brunello Human CRISPR Knockout Pooled Library (Addgene #73178) was introduced into the C666–Cas9 cells via lentiviral transduction at a low multiplicity of infection (MOI) of approximately 0.3, and transduced cells were selected using puromycin. After 48 h of culture, genomic DNA was extracted from a sufficient number of cells to ensure coverage of at least 400 cells per sgRNA. Then, sgRNA sequences were amplified for library preparation, and next-generation sequencing (NGS) was performed by the Illumina Novaseq platform to quantify the abundance of each sgRNA. The abundance of sgRNAs in sequencing data reflects the proportion of cells with specific knockouts at the time of DNA collection to infer the cell growth promotion or inhibition. The resulting data were analysed by MAGeCK to identify significant gene knockouts affecting cell growth.

Statistics and reproducibility

All statistical analyses were performed using two-tailed tests with an alpha level of 0.05, unless otherwise specified. Statistical analyses of the data were conducted using Prism 10 version 10.2.3 and R version 4.3.2. Data of majority of the plots were measured three times independently unless it is stated otherwise. No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomised. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

Data availability

The Omni-C data for NPE, EBV + NPC, and EBV-NPC cell lines are deposited into the Gene Expression Omnibus database under GEO accession GSE274581 and are available at the following URL [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE274581]. The ATAC-seq data for NPE, EBV+ NPC, and EBV− NPC cell lines are deposited into the Gene Expression Omnibus database under GEO accession GSE278605. The CUT&RUN-seq data for NPE, EBV+ NPC, EBV− NPC cell lines of H3K4me3, CTCF, and H3K27ac for cell lines are deposited into the Gene Expression Omnibus database under GEO accession GSE278605 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE278607]. The spatial transcriptome data for the 10× visum platform are deposited into the Gene Expression Omnibus database under the GEO accession GSE299776. The spatial transcriptome data for GeoMx-DSP are deposited into the Gene Expression Omnibus database under GEO accession GSE299775. The raw sequence data (WES and bulk RNA-seq) from the clinical samples generated in this study have been deposited in the Genome Sequence Archive (GSA) of the National Genomics Data Centre (NGDC), Beijing Institute of Genomics (China National Centre for Bioinformation), Chinese Academy of Sciences, under accession number HRA008959 [https://ngdc.cncb.ac.cn/gsa-human/browse/dac/HDAC004793]. The raw data are available under controlled access due to data privacy laws related to patient consent for data sharing. The data should be used for research purposes only. According to the guidelines of GSA-human, all non-profit researchers are allowed access to the data, and the Principal Investigator of any research group can apply for the data following the guidelines at the GSA database portal (https://ngdc.cncb.ac.cn/gsa-human/). The response time for access requests is approximately 10 working days. Once access has been granted, the data will be available for download within one month. The user can also contact the corresponding author directly for enquiries. The NPC scRNA-Seq publicly available data used in this study are available in the GEO database under accession code GSE150825[15] and GSE162025[14], and from CNGBdb database with accession number CNP0000428[16]. Source data are provided with this paper. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

References

1. Shannon-Lowe, C. & Rickinson, A. The global landscape of EBV-associated tumors. *Front. Oncol.* **9**, 713 (2019).
2. Wang, C. et al. A DNA tumor virus globally reprograms host 3D genome architecture to achieve immortal growth. *Nat. Commun.* **14**, 1598 (2023).
3. Friedman, M. J. et al. Dynamics of viral and host 3D genome structure upon infection. *J. Microbiol. Biotechnol.* **32**, 1515–1526 (2022).
4. Karimzadeh, M. et al. Human papillomavirus integration transforms chromatin to drive oncogenesis. *Genome Biol.* **24**, 142 (2023).
5. Hildebrand, E. M. & Dekker, J. Mechanisms and functions of chromosome compartmentalization. *Trends Biochem. Sci.* **45**, 385–396 (2020).
6. Oti, M. et al. CTCF-mediated chromatin loops enclose inducible gene regulatory domains. *BMC Genomics* **17**, 252 (2016).
7. Feng, Y. & Pauklin, S. Revisiting 3D chromatin architecture in cancer development and progression. *Nucleic Acids Res.* **48**, 10632–10647 (2020).
8. Zhao, J. & Faryabi, R. B. Spatial promoter-enhancer hubs in cancer: organization, regulation, and function. *Trends Cancer* **9**, 1069–1084 (2023).
9. Okabe, A. et al. Cross-species chromatin interactions drive transcriptional rewiring in Epstein-Barr virus-positive gastric adenocarcinoma. *Nat. Genet.* **52**, 919–930 (2020).
10. Vara, C. et al. Three-dimensional genomic structure and cohesin occupancy correlate with transcriptional activity during spermatogenesis. *Cell Rep.* **28**, 352–367 e9 (2019).
11. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
12. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
13. Harris, H. L. et al. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat. Commun.* **14**, 3303 (2023).
14. Liu, Y. et al. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat. Commun.* **12**, 741 (2021).
15. Gong, L. et al. Comprehensive single-cell sequencing reveals the stromal dynamics and tumor-specific characteristics in the microenvironment of nasopharyngeal carcinoma. *Nat. Commun.* **12**, 1540 (2021).
16. Chen, Y. P. et al. Single-cell transcriptomics reveals regulators underlying immune cell diversity and immune subtypes associated with prognosis in nasopharyngeal carcinoma. *Cell Res.* **30**, 1024–1042 (2020).
17. Ka-Yue Chow, L. et al. Epigenomic landscape study reveals molecular subtypes and EBV-associated regulatory epigenome reprogramming in nasopharyngeal carcinoma. *EBioMedicine* **86**, 104357 (2022).
18. Zhang, D. et al. VCAM1 promotes tumor cell invasion and metastasis by inducing EMT and transendothelial migration in colorectal cancer. *Front. Oncol.* **10**, 1066 (2020).
19. Ren, G. et al. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol. Cell* **67**, 1049–1058 e6 (2017).
20. Hsieh, T. S. et al. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022).
21. Hodin, T. L., Najrana, T. & Yates, J. L. Efficient replication of Epstein-Barr virus-derived plasmids requires tethering by EBNA1 to host chromosomes. *J. Virol.* **87**, 13020–13028 (2013).
22. Chen, Q. & Massague, J. Molecular pathways: VCAM-1 as a potential therapeutic target in metastasis. *Clin. Cancer Res.* **18**, 5520–5525 (2012).
23. VanHeyst, K. A. et al. Ectopic tumor VCAM-1 expression in cancer metastasis and therapy resistance. *Cells* **11**, 3922 (2022).
24. Carmeliet, P. VEGF as a key mediator of angiogenesis in cancer. *Oncology* **69**, 4–10 (2005). Suppl 3.
25. Bourhis, M. et al. Direct and indirect modulation of T cells by VEGF-a counteracted by anti-angiogenic treatment. *Front. Immunol.* **12**, 616837 (2021).
26. Hu, J. et al. Regulation of programmed cell death by Brd4. *Cell Death Dis.* **13**, 1059 (2022).
27. Sengupta, D. et al. Disruption of BRD4 at H3K27Ac-enriched enhancer region correlates with decreased c-Myc expression in Merkel cell carcinoma. *Epigenetics* **10**, 460–466 (2015).
28. Chakravorty, A. & Sugden, B. The AT-hook DNA binding ability of the Epstein Barr virus EBNA1 protein is necessary for the maintenance of viral genomes in latently infected cells. *Virology* **484**, 251–258 (2015).
29. Kim, K. D. et al. Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus. *Nat. Commun.* **11**, 877 (2020).
30. Kim, D. N. et al. Characterization of naturally Epstein-Barr virus-infected gastric carcinoma cell line YCCEL1. *J. Gen. Virol.* **943**, 497–506 (2013).
31. Lin, W. et al. Establishment and characterization of new tumor xenografts and cancer cell lines from EBV-positive nasopharyngeal carcinoma. *Nat. Commun.* **9**, 4663 (2018).

32. Ribeiro, J. et al. Epstein-Barr virus gene expression and latency pattern in gastric carcinomas: a systematic review. *Future Oncol.* **13**, 567–579 (2017).
33. Oh, S. T. et al. A naturally derived gastric cancer cell line shows latency I Epstein-Barr virus infection closely resembling EBV-associated gastric cancer. *Virology* **320**, 330–336 (2004).
34. He, R. et al. H3K4 demethylase KDM5B regulates cancer cell identity and epigenetic plasticity. *Oncogene* **41**, 2958–2972 (2022).
35. Masternak, K. et al. CIITA is a transcriptional coactivator that is recruited to MHC class II promoters by multiple synergistic interactions with an enhanceosome complex. *Genes Dev.* **14**, 1156–1166 (2000).
36. Devaiah, B. N. & Singer, D. S. CIITA and its dual roles in MHC gene transcription. *Front. Immunol.* **4**, 476 (2013).
37. Steimle, V. et al. Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science* **265**, 106–109 (1994).
38. Ziegler, P. et al. A primary nasopharyngeal three-dimensional air-liquid interface cell culture model of the pseudostratified epithelium reveals differential donor- and cell type-specific susceptibility to Epstein-Barr virus infection. *PLoS Pathog.* **17**, e1009041 (2021).
39. Mensali, N. et al. Antigen-delivery through invariant chain (CD74) boosts CD8 and CD4 T cell immunity. *Oncoimmunology* **8**, 1558663 (2019).
40. Kilian, M. et al. MHC class II-restricted antigen presentation is required to prevent dysfunction of cytotoxic T cells by blood-borne myeloids in brain tumors. *Cancer Cell* **41**, 235–251 e9 (2023).
41. Jin, S. et al. Single-cell transcriptomic analysis defines the interplay between tumor cells, viral infection, and the microenvironment in nasopharyngeal carcinoma. *Cell Res.* **30**, 950–965 (2020).
42. Li, J. S. Z. et al. Chromosomal fragile site breakage by EBV-encoded EBNA1 at clustered repeats. *Nature* **616**, 504–509 (2023).
43. Shumilov, A. et al. Epstein-Barr virus particles induce centrosome amplification and chromosomal instability. *Nat. Commun.* **8**, 14257 (2017).
44. Li, X. et al. Histone demethylase KDM5B is a key regulator of genome stability. *Proc. Natl Acad. Sci. USA* **111**, 7096–7101 (2014).
45. Zhang, L. et al. Gemcitabine plus cisplatin versus fluorouracil plus cisplatin in recurrent or metastatic nasopharyngeal carcinoma: a multicentre, randomised, open-label, phase 3 trial. *Lancet* **388**, 1883–1892 (2016).
46. Zhang, Y. et al. Final overall survival analysis of gemcitabine and cisplatin induction chemotherapy in nasopharyngeal carcinoma: a multicenter, randomized phase III trial. *J. Clin. Oncol.* **40**, 2420–2425 (2022).
47. Chen, Y. P. et al. Nasopharyngeal carcinoma. *Lancet* **394**, 64–80 (2019).
48. Mai, H. Q. et al. Toripalimab or placebo plus chemotherapy as first-line treatment in advanced nasopharyngeal carcinoma: a multicenter randomized phase 3 trial. *Nat. Med.* **27**, 1536–1543 (2021).
49. Liu, G. Y. et al. Nab-paclitaxel, cisplatin, and capecitabine versus cisplatin and gemcitabine as first line chemotherapy in patients with recurrent or metastatic nasopharyngeal carcinoma: randomised phase 3 clinical trial. *Br. Med. J.* **385**, e077890 (2024).
50. Lo, A. K. et al. Epstein-Barr virus infection alters cellular signal cascades in human nasopharyngeal epithelial cells. *Neoplasia* **8**, 173–180 (2006).
51. Busson, P. et al. Establishment and characterization of three transplantable EBV-containing nasopharyngeal carcinomas. *Int. J. Cancer* **42**, 599–606 (1988).
52. Lin, J. C. et al. Quantification of plasma Epstein-Barr virus DNA in patients with advanced nasopharyngeal carcinoma. *N. Engl. J. Med.* **350**, 2461–2470 (2004).
53. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101 (2016).
54. Roayaei Ardakany, A. et al. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol.* **21**, 256 (2020).
55. Huang, W. et al. ChIAPoP: a new tool for ChIA-PET data analysis. *Nucleic Acids Res.* **47**, e37 (2019).
56. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
57. Arvey, A. et al. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell Host Microbe* **12**, 233–245 (2012).
58. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
59. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 e21 (2019).
60. Chin, S. F. et al. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* **8**, R215 (2007).
61. Bakhoum, S. F. et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* **553**, 467–472 (2018).
62. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
63. Patro, R. et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
64. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

Acknowledgements

This work was supported by General Research Fund (17102619) and Theme-based Research Scheme (T12-703/22-R; T123-70323-N) from Research Grant Council, Health and Medical Research Grant from Health Bureau (08192296) from Hong Kong (SAR), P. R. China, The University of Hong Kong Start-up Fund, Science and Technology Programme of Guangdong Province in China (2021A0505110010), Shenzhen Science and Technology Programme in China (JCYJ20220818103014030 and ZDSYS20210623091811035), as well as Health@InnoHK, Innovation and Technology Commission, The government of Hong Kong (SAR), P. R. China. We would like to acknowledge the Nasopharyngeal Carcinoma (NPC) Tissue Bank from Li Ka Shing Faculty of Medicine at the University of Hong Kong for the NPC cell lines and clinical specimens, Centre for PanorOmic Sciences (CPOS) of Li Ka Shing Faculty of Medicine at the University of Hong Kong for the genomic service and Ufuk Degirmenci, Sivanisswary Manoharan and Fang Jun Wen for the GeoMx DSP support. We acknowledge Prof. Maria Li Lung for the establishment of the NPC tissue bank for the HK cohort and Prof. George Sai-Wah Tsao for the establishment of the NPC PDXs.

Author contributions

Conceptualised and designed the study: W.D. and Y.X. Experiments and sample collections: D.L.S.C., Y.W., S.L., K.A.I., J.L., L.G., K.K.W.C., Z.Y., P.F., A.H., R.K.C.N., X.G., W.T.N., A.C.M.T., D.L.K.W., A.W.M.L. and V.H.F.L. Bioinformatic data analysis and interpreted the results: D.L.S.C., Z.H. (Zhaozheng Hou), K.A.I., L.K.Y.C., C.P.T.C., Y.Z., Z.L. and Y.X. Draft the paper: D.L.S.C., Z.H. (Zhaozheng Hou), Y.W., K.A.I., S.L., W.D. and Y.X. Review and editing: D.L.S.C., Z.H. (Zhaozheng Hou), Y.W., K.A.I., S.L., J.L., L.K.Y.C., C.P.T.C, Y.Z., L.G., K.K.W.C., Z.Y., P.F., Z.H. (Zilu Huang), R.K.C.N., X.G., W.T.N., Z.L., A.C.M.T., D.L.W.K., A.W.M.L., V.H.F.L., H.C., Y.X. and W.D. Supervision and mentorship: W.D., Y.X. and H.C.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-61597-1.

Correspondence and requests for materials should be addressed to Wei Dai.

Peer review information *Nature Communications* thanks Pinaki Bose, N. Gopalakrishna Iyer, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025

---

[1]Department of Clinical Oncology, Centre of Cancer Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong (SAR), PR China. [2]Centre for Virology, Vaccinology and Therapeutics, Hong Kong Science and Technology Park, Hong Kong, Hong Kong (SAR), PR China. [3]Department of Microbiology and State Key Laboratory for Emerging Infectious Diseases, The University of Hong Kong, Hong Kong, Hong Kong (SAR), PR China. [4]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, The Sun Yat-sen University Cancer Center, Guangzhou, PR China. [5]Department of Pathology, The Sun Yat-sen University Cancer Centre, Guangzhou, PR China. [6]Department of Radiation Oncology, Sun Yat-sen University Cancer Centre, Guangzhou, PR China. [7]The University of Hong Kong-Shenzhen Hospital, Shenzhen, PR China. [8]Department of Biostatistics, Columbia University, New York, NY, USA. [9]Department of Anatomical and Cellular Pathology and State Key Laboratory of Translational Oncology, The Chinese University of Hong Kong, Hong Kong, Hong Kong (SAR), PR China. [10]These authors contributed equally: Dittman Lai-Shun Chung, Zhaozheng Hou, Ying Wang, Kazi Anisha Islam, Songran Liu. ✉e-mail: weidai2@hku.hk