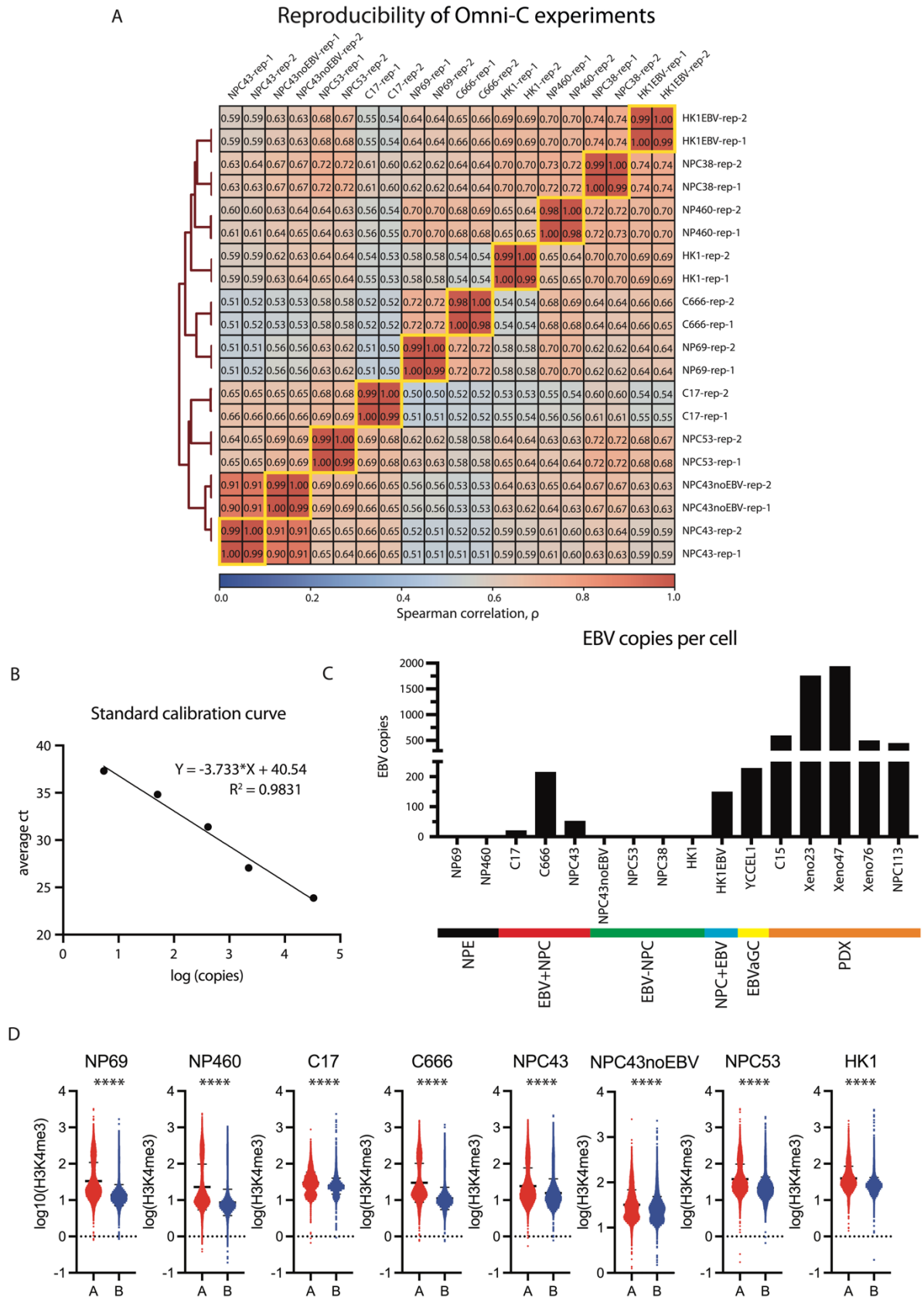
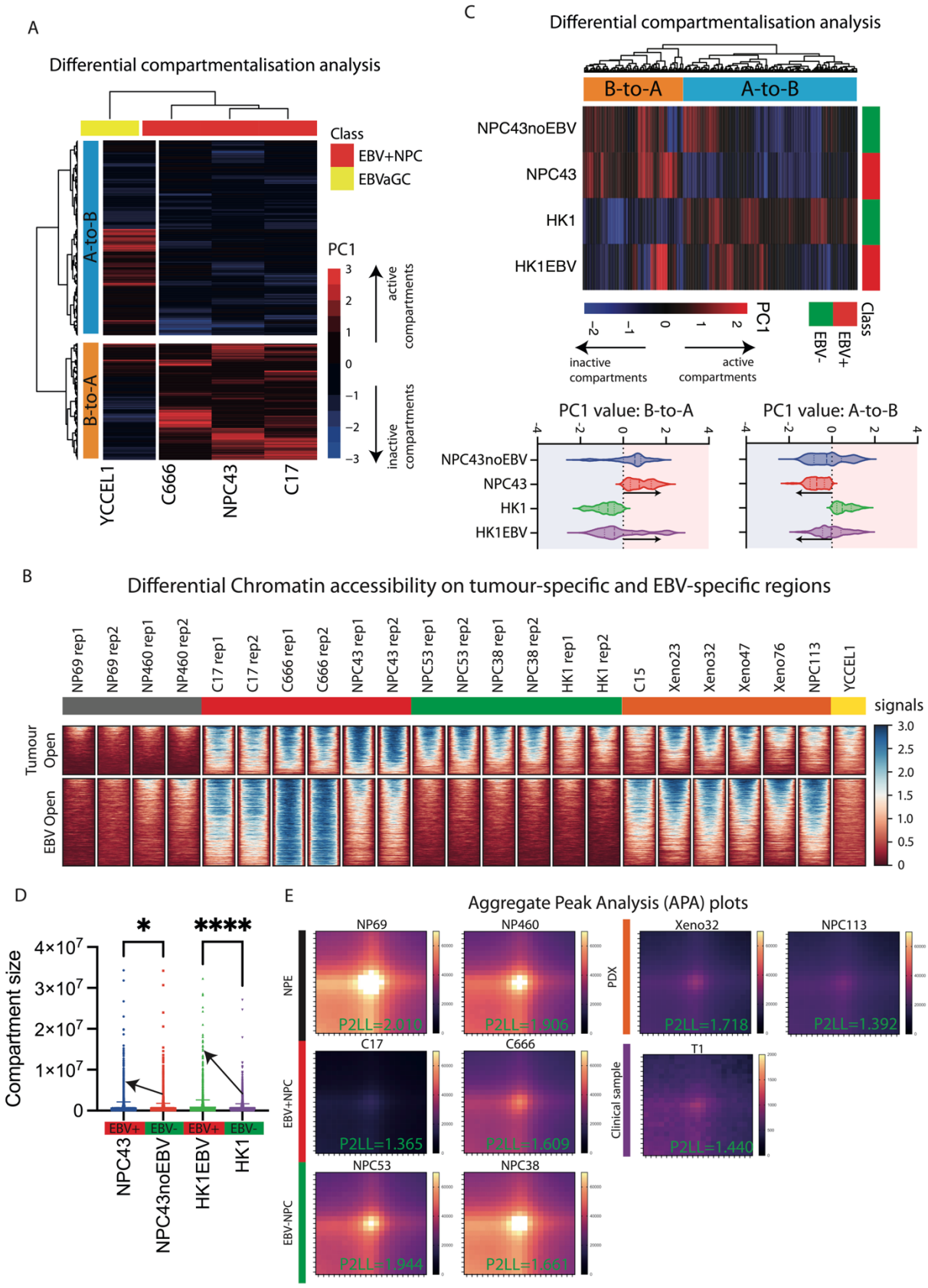

A
Reproducibility of Omni-C experiments
NPC43-rep-1
NPC43-rep-2
NPC43noEBV-rep-1
NPC43noEBV-rep-2
NPC53-rep-1
NPC53-rep-2
C17-rep-1
C17-rep-2
NP69-rep-1
NP69-rep-2
C666-rep-1
C666-rep-2
HK1-rep-1
HK1-rep-2
NP460-rep-1
NP460-rep-2
NPC38-rep-1
NPC38-rep-2
HK1EBV-rep-1
HK1EBV-rep-2
HK1EBV-rep-2
HK1EBV-rep-1
NPC38-rep-2
NPC38-rep-1
NP460-rep-2
NP460-rep-1
HK1-rep-2
HK1-rep-1
C666-rep-2
C666-rep-1
NP69-rep-2
NP69-rep-1
C17-rep-2
C17-rep-1
NPC53-rep-2
NPC53-rep-1
NPC43noEBV-rep-2
NPC43noEBV-rep-1
NPC43-rep-2
NPC43-rep-1
0.0
0.2
0.4
0.6
0.8
1.0
Spearman correlation, ρ
B
Standard calibration curve
Y = -3.733*X + 40.54
R² = 0.9831
average ct
log (copies)
C
EBV copies per cell
EBV copies
NP69
NP460
C17
C666
NPC43
NPC43noEBV
NPC53
NPC38
HK1
HK1EBV
YCCEL1
C15
Xeno23
Xeno47
Xeno76
NPC113
NPE
EBV+NPC
EBV-NPC
NPC+EBV
EBVaGC
PDX
D
NP69
NP460
C17
C666
NPC43
NPC43noEBV
NPC53
HK1
****
log10(H3K4me3)
log(H3K4me3)
A
B

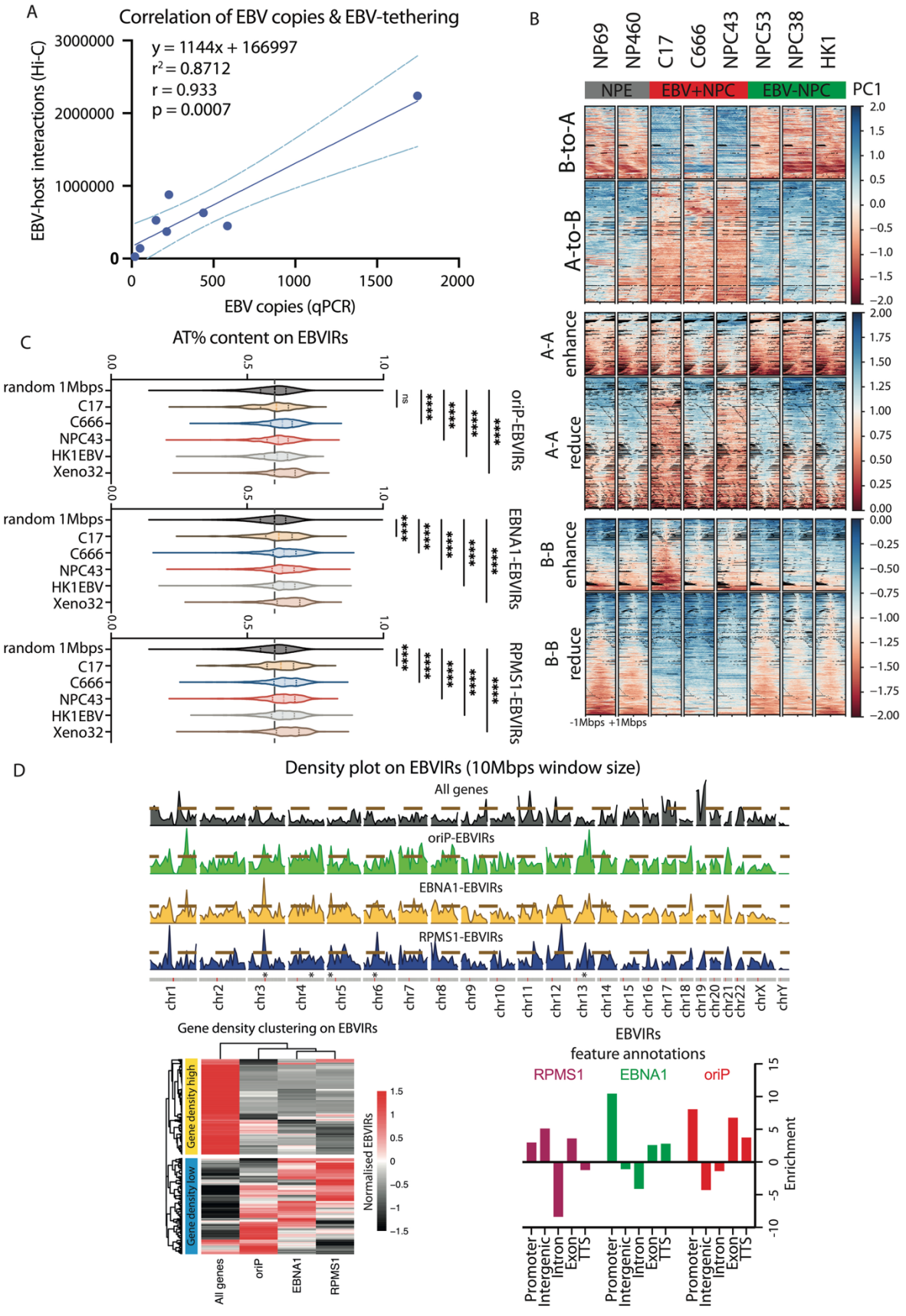
Supplementary Fig. 1 Quality control on Omni-C data and detection of EBV copies in all NPC cell lines and PDXs.

A) Reproducibility of Omni-C experiment. Replicates were performed for all cell lines (n_{total} = 20). Spearman correlation coefficients were calculated based on bam files with a bin size of 10 Mbps. **B) Establishment of a standard calibration curve for the measurement of EBV DNA copies.** To establish a standard calibration curve for measuring EBV DNA copies, Namalwa cells containing EBV DNA copies were serially diluted, and qPCR was performed. The calibration curve was generated by plotting log10-transformed values of EBV DNA copies against the average ct values from qPCR. The experiment contains technical triplicates (n =3) for each data point. Average mean is plotted. The linear equation was established as $y = -3.733x + 40.54$, with a coefficient of determination (r^2) of 0.9831. **C) Estimation of EBV DNA copies per cell.** The ct value obtained was used to calculate EBV copies per cell based on the standard calibration curve established with an equal amount of input. The experiment contains technical triplicates (n =3) for each cell line. Average mean is used for the plot. EBV copies were confirmed in various cell lines, including C17, C666, NPC43, HK1EBV, YCCEL1, C17, Xeno23, Xeno32, Xeno47, Xeno76, and NPC113. This is to confirm the detectable EBV copies present in EBV+NPC and EBV+PDXs. **D) Active histone modification of H3K4me3 at A/B compartments.** CUT&RUN-Seq was performed to investigate the distribution of the active histone mark H3K4me3 in different genomic compartments. The signals were normalised as RPKM (Reads Per Kilobase per Million mapped reads). Compartments were divided into 50kb window sizes and categorised as A (active – red colour) and B (inactive – blue colour) based on the gene density within each compartment. The log_{10}-transformed H3K4me3 signals were summarised by each 50kb window size. The violin plot showed that A compartments exhibit higher H3K4me3 signals compared to their individual B compartments, indicating that A compartments are more transcriptionally active than the B compartments. A two-tailed unpaired t-test has been performed to examine if there is a significant difference between compartment A and B; $n_{NP69compartmentA}$ = 29575 and $n_{NP69compartmentB}$ = 25693, $n_{NP460compartmentA}$ = 27008 and $n_{NP460compartmentB}$ = 28265, $n_{C17compartmentA}$ = 26989 and $n_{C17compartmentB}$ = 28286, $n_{C666compartmentA}$ = 25697 and $n_{C666compartmentB}$ = 29587, $n_{NPC43compartmentA}$ = 26986 and $n_{NPC43compartmentB}$ = 28288, $n_{NPC43noEBVcompartmentA}$ = 28531 and $n_{NPC43noEBVcompartmentB}$ = 26746, $n_{NPC53compartmentA}$ = 26814 and $n_{NPC53compartmentB}$ = 28466, $n_{HK1compartmentA}$ = 26006 and $n_{HK1compartmentB}$ = 29272. Mean and standard deviation were displayed in the plot, ****adjusted *p-value* ≤ 0.0001.


A
Differential compartmentalisation analysis
Class
EBV+NPC
EBVaGC
PC1
active compartments
inactive compartments
A-to-B
B-to-A
YCCEL1
C666
NPC43
C17
C
Differential compartmentalisation analysis
B-to-A
A-to-B
NPC43noEBV
NPC43
HK1
HK1EBV
PC1
inactive compartments
active compartments
Class
EBV+
EBV-
PC1 value: B-to-A
PC1 value: A-to-B
B
Differential Chromatin accessibility on tumour-specific and EBV-specific regions
NP69 rep1
NP69 rep2
NP460 rep1
NP460 rep2
C17 rep1
C17 rep2
C666 rep1
C666 rep2
NPC43 rep1
NPC43 rep2
NPC53 rep1
NPC53 rep2
NPC38 rep1
NPC38 rep2
HK1 rep1
HK1 rep2
C15
Xeno23
Xeno32
Xeno47
Xeno76
NPC113
YCCEL1
signals
Tumour Open
EBV Open
D
Compartment size
EBV+
EBV-
NPC43
NPC43noEBV
HK1EBV
HK1
E
Aggregate Peak Analysis (APA) plots
NP69
NP460
Xeno32
NPC113
C17
C666
T1
NPC53
NPC38
NPE
PDX
EBV+NPC
Clinical sample
EBV-NPC
P2LL=2.010
P2LL=1.906
P2LL=1.718
P2LL=1.392
P2LL=1.365
P2LL=1.609
P2LL=1.440
P2LL=1.944
P2LL=1.661

Supplementary Fig. 2 The 3D genome organisation and 2D chromatin accessibility in EBV+ NPC and EBVaGC (YCCEL1), and loss of chromatin loops leading to increased compartment size in EBV+ NPC samples.

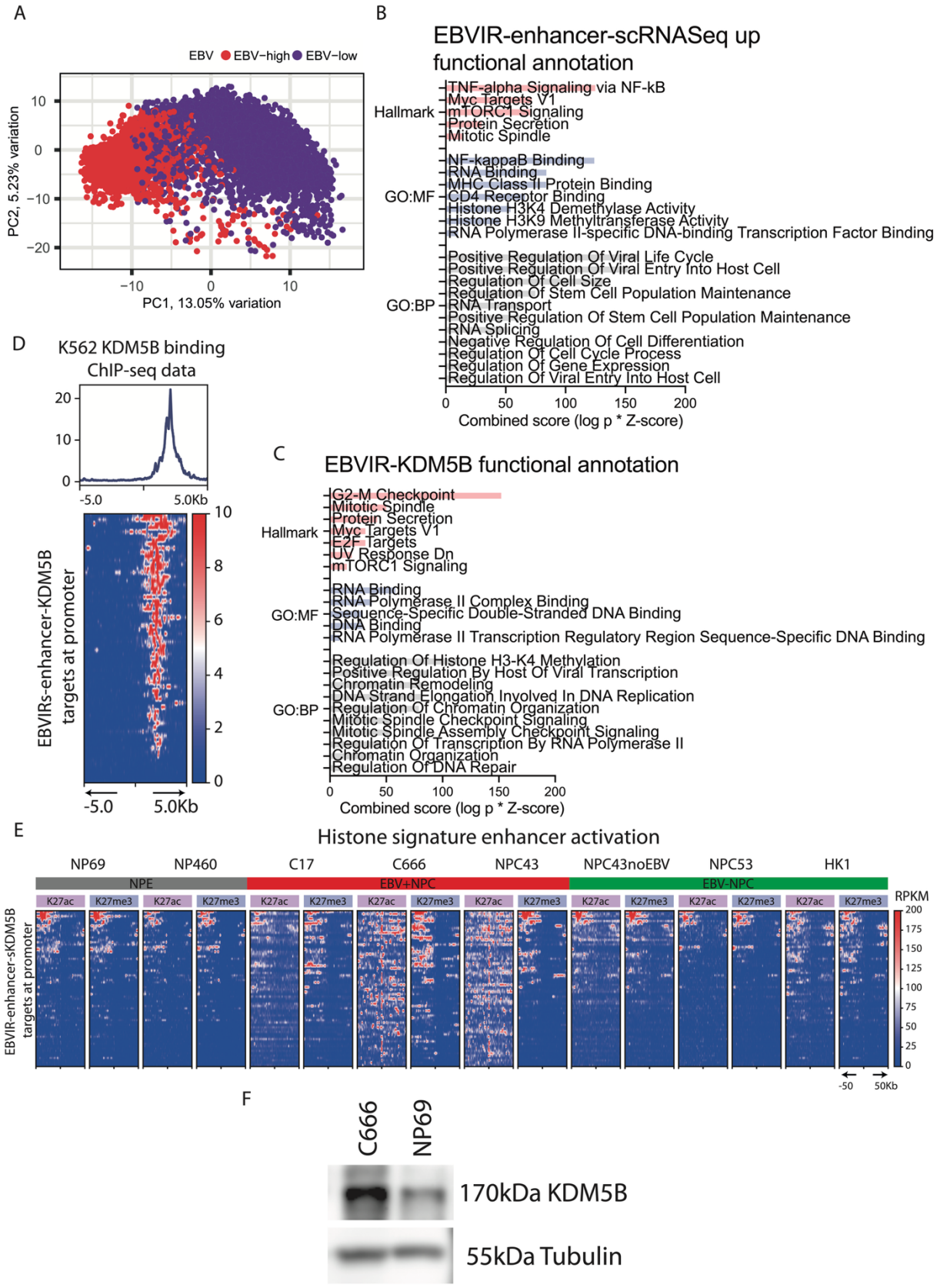
A) Heatmap of A/B compartments in EBV+NPC (C17, C666, NPC43) and EBVaGC (YCCEL1). Compartments are classified as switches of A-to-B (active to inactive) and B-to-A (inactive to active) compartments. A-to-B compartments refer to the compartment changes from A in EBV- NPC to B in EBV+ NPC. B-to-A compartments refer to the compartment changes from B in EBV- NPC to A in EBV+ NPC. YCCEL1 exhibits a distinctive organisation of AB compartments compared to EBV+ NPC samples. **B) Chromatin accessibility of EBVaGC at NPC tumour-specific and EBV-specific open regions.** The differential chromatin accessibility regions were identified from a previous study [1]. YCCEL1 showed open accessibility on tumour-specific open chromatin, similar to other EBV+ samples. However, in EBVaGC EBV did not affect EBV-specific open chromatin accessibility identified from EBV+ NPC samples, suggesting EBV influences the chromatin accessibility in a tissue-specific manner. **C) Heatmap and violin plot of differential A/B compartments in two paired EBV- / EBV+ NPC cells.** The *in vitro* EBV-infected HK1EBV has already shown the selective compartment changes from B to A. The NPC43noEBV has lost the EBV copies from NPC43, showing the loss of EBV episomes could partially reverse the change of A/B compartments. **D) Overall size of compartments in paired EBV+ and EBV- NPC cell lines derived from HK1 and NPC43.** The compartments are segmented into each 50 kbps and classified as active (A) or inactive (B). The same classification of compartments was aggregated, and the compartment size was therefore measured. A two-tailed unpaired t-test has been performed to examine if there is a significant difference between EBV- and EBV+ of matched cell lines ($n_{NPC43noEBV}$ = 4512 and n_{NPC43} = 3788, n_{HK1} = 3747 and n_{HK1EBV} = 2096). *adjusted *p-value* ≤ 0.05, ****adjusted *p-value* ≤ 0.0001. **E) Reduction of chromatin looping in EBV+ NPC, PDX and clinical NPC.** Chromatin loop-loop interactions were generated using Omni-C data by Mustache. The chromatin loops identified from NPE, EBV+NPC, and EBV-NPC samples were combined for the analysis. Aggregate Peak Analysis (APA) was performed on individual samples, calculating the peak-to-lower-left (P2LL) score of the interaction matrices.


A
Correlation of EBV copies & EBV-tethering
y = 1144x + 166997
r² = 0.8712
r = 0.933
p = 0.0007
EBV-host interactions (Hi-C)
EBV copies (qPCR)
B
NP69
NP460
C17
C666
NPC43
NPC53
NPC38
HK1
NPE
EBV+NPC
EBV-NPC
PC1
B-to-A
A-to-B
A-A enhance
A-A reduce
B-B enhance
B-B reduce
-1Mbps +1Mbps
C
AT% content on EBVIRs
random 1Mbps
C17
C666
NPC43
HK1EBV
Xeno32
oriP-EBVIRs
EBNA1-EBVIRs
RPMS1-EBVIRs
D
Density plot on EBVIRs (10Mbps window size)
All genes
oriP-EBVIRs
EBNA1-EBVIRs
RPMS1-EBVIRs
Gene density clustering on EBVIRs
Gene density high
Gene density low
Normalised EBVIRs
EBVIRs
feature annotations
RPMS1
EBNA1
oriP
Enrichment
Promoter
Intergenic
Intron
Exon
TTS

Supplementary Fig. 3 Genomic features of EBVIRs

A) Correlation of the number of EBV-host chromatin interacting sites and EBV DNA copies. The number of reads representing EBV interacting with the human genome was extracted from the Omni-C data. Each dot in the plot represents one sample. A robust correlation is observed between the number of EBV copies and the frequency of EBV-host chromatin interactions, $n_{cell\text{-}lines} = 8$. The dashed line denotes a 95% confidence interval, and the relationship is represented by the linear regression line ($y = 1144x + 166997$, r = 0.933, *p-value* ≤ 0.0007). **B) Compartment alterations in EBV+ NPC cell lines compared to EBV- NPC cell lines for six classifications of the compartments (B-to-A, A-to-B, A-A enhanced, A-A reduced, B-B enhanced, B-B reduced compartments).** The figure illustrates compartment alterations in EBV+ NPC cell lines (EBV+NPC) compared to EBV- NPC cell lines (EBV-NPC) for six classifications of compartments, including B-to-A, A-to-B, A-A enhanced, A-A reduced, B-B enhanced, and B-B reduced compartments. The B-to-A (A-to-B) compartments represent inactive (active) compartments in EBV-NPC that become active (inactive) compartments in EBV+NPC. The A-A (B-B) enhancement of compartments is identified by greater positive (negative) PC1 values in EBV+NPC compared to EBV-NPC in active (inactive) compartments. The A-A (B-B) reduction of compartments is identified by less positive (negative) PC1 values in EBV+NPC compared to EBV-NPC in active compartments. The heatmap displays average signals in the centre with both upstream and downstream of 1 Mbps extension. The heatmap plot is based on the PC1 value of the Omni-C data. **C) EBVIRs are enriched at AT-rich sequences.** The percentage of AT content was calculated for individual samples (n = 5) of the EBVIRs at oriP, EBNA1, and RPMS1. One million random sequences of 1000bp fragment size sampled from the human genome (hg19) were used as background AT content level. One-way ANOVA with Bonferroni correction was performed to test the significance, *adjusted *p-value* ≤ 0.01. **D) Density plots and feature annotations on oriP-, EBNA1-, and RPMS1-EBVIRs** on the human genome in 10Mbps window

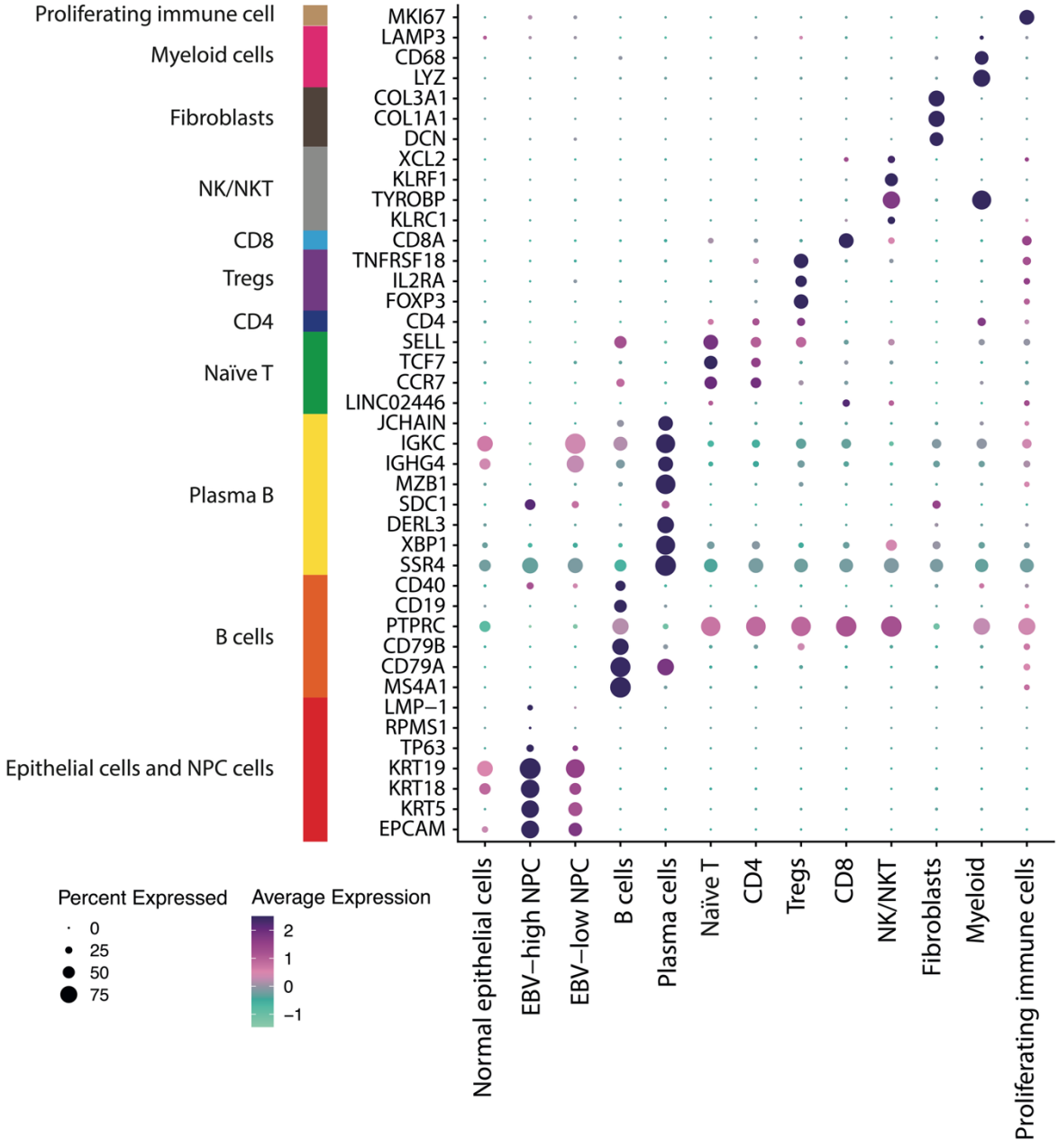
size across chromosomes. Enrichment has been calculated for three EBVIRs according to promoter, intergenic, intron, exon, and TSS regions.


A
EBV EBV−high EBV−low
PC2, 5.23% variation
PC1, 13.05% variation
B
EBVIR-enhancer-scRNASeq up
functional annotation
Hallmark
GO:MF
GO:BP
TNF-alpha Signaling via NF-kB
Myc Targets V1
mTORC1 Signaling
Protein Secretion
Mitotic Spindle
NF-kappaB Binding
RNA Binding
MHC Class II Protein Binding
CD4 Receptor Binding
Histone H3K4 Demethylase Activity
Histone H3K9 Methyltransferase Activity
RNA Polymerase II-specific DNA-binding Transcription Factor Binding
Positive Regulation Of Viral Life Cycle
Positive Regulation Of Viral Entry Into Host Cell
Regulation Of Cell Size
Regulation Of Stem Cell Population Maintenance
RNA Transport
Positive Regulation Of Stem Cell Population Maintenance
RNA Splicing
Negative Regulation Of Cell Differentiation
Regulation Of Cell Cycle Process
Regulation Of Gene Expression
Regulation Of Viral Entry Into Host Cell
Combined score (log p * Z-score)
C
EBVIR-KDM5B functional annotation
Hallmark
GO:MF
GO:BP
G2-M Checkpoint
Mitotic Spindle
Protein Secretion
Myc Targets V1
E2F Targets
UV Response Dn
mTORC1 Signaling
RNA Binding
RNA Polymerase II Complex Binding
Sequence-Specific Double-Stranded DNA Binding
DNA Binding
RNA Polymerase II Transcription Regulatory Region Sequence-Specific DNA Binding
Regulation Of Histone H3-K4 Methylation
Positive Regulation By Host Of Viral Transcription
Chromatin Remodeling
DNA Strand Elongation Involved In DNA Replication
Regulation Of Chromatin Organization
Mitotic Spindle Checkpoint Signaling
Mitotic Spindle Assembly Checkpoint Signaling
Regulation Of Transcription By RNA Polymerase II
Chromatin Organization
Regulation Of DNA Repair
Combined score (log p * Z-score)
D
K562 KDM5B binding
ChIP-seq data
EBVIRs-enhancer-KDM5B
targets at promoter
-5.0
5.0Kb
E
Histone signature enhancer activation
NP69
NP460
C17
C666
NPC43
NPC43noEBV
NPC53
HK1
NPE
EBV+NPC
EBV-NPC
K27ac
K27me3
RPKM
EBVIR-enhancer-sKDM5B
targets at promoter
-50
50Kb
F
C666
NP69
170kDa KDM5B
55kDa Tubulin

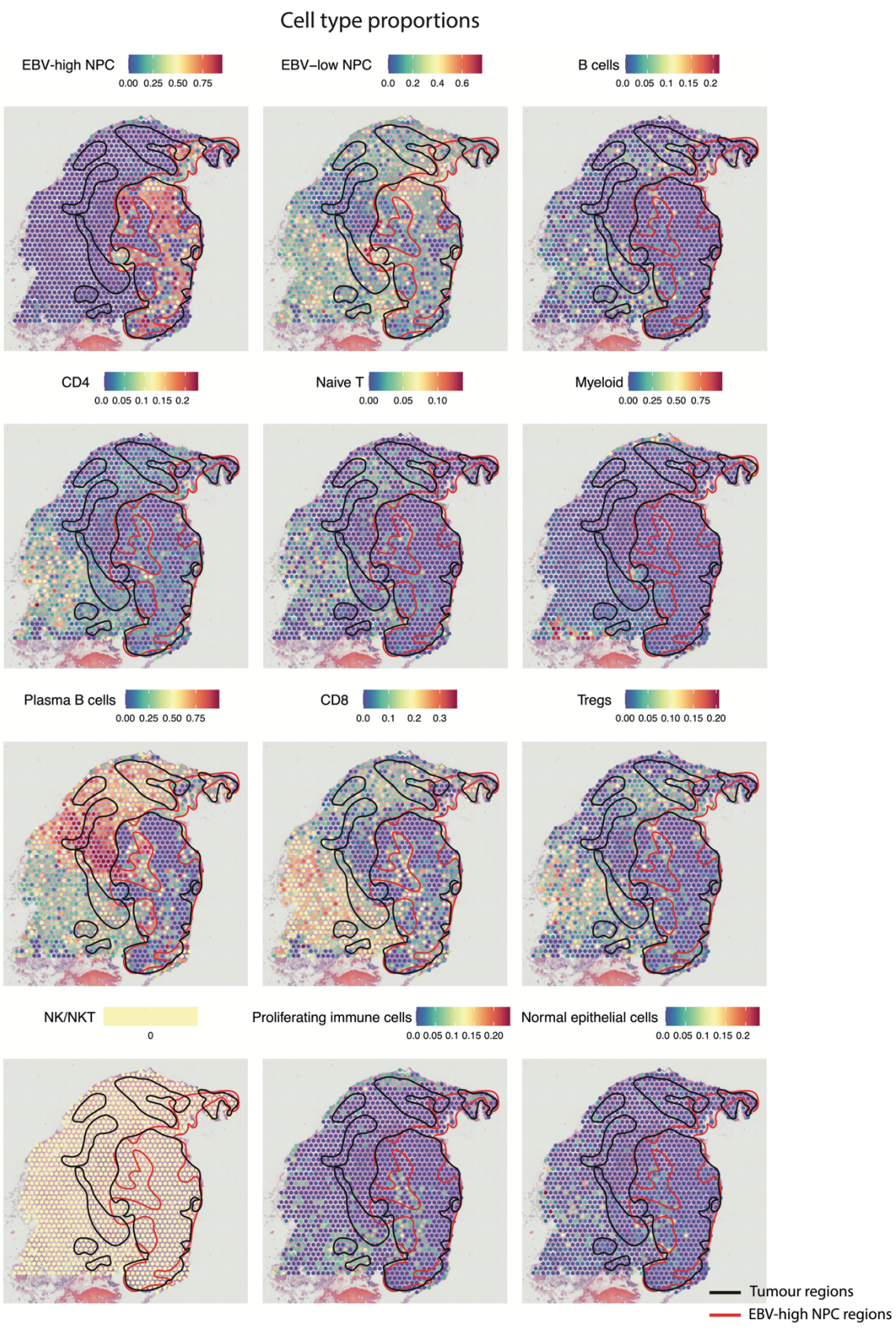
Supplementary Fig. 4 EBVIR-enhancer regulation was associated with KDM5B and KDM5B target activations.

A) Principal component analysis (PCA) of NPC malignant cells shows two distinctive EBV-high and EBV-high NPC cells from scRNA-seq. It contains $n_{features}$ = 7211 and n_{genes} = 1054 for the PCA plot. This dataset was used to identify the genes regulated by enhancer activation induced by EBV tethering in a cis-regulatory manner. **B) Gene Ontology analysis for EBVIR-enhancer-related genes.** Upregulated genes (n = 154) involved in EBVIR-enhancer were performed the functional annotations using the molecular signature database Hallmark 2020, GO Biological Process 2023, and GO Molecular Function 2023 in Enrichr. The Hallmark terms involved in TNF-alpha signalling via NF-κB, Myc targets, protein secretion, and mitotic spindle were identified. The molecular functional terms involved in NF-κB binding, RNA binding, MHC class II protein binding, histone H3K4 demethylase, and histone H3K9 methyltransferase activity were identified. Biological processes involved in the positive regulation of viral entry to the host, cell differentiation, cell cycle process, and gene expression were also identified. Significant terms were selected with *p-value* ≤ 0.1. **C) Gene Ontology analysis for EBVIRs-KDM5B targets (n = 808 genes).** The Hallmark terms involved in the G2/M checkpoint, mitotic spindle, protein secretion, and MYC targets, were identified. The molecular functional terms involved in RNA Binding, RNA Polymerase II complex binding, double-stranded DNA binding, and RNA Polymerase II transcription DNA binding were identified. The biological process involved H3K4 methylation, positive regulation of viral transcription, chromatin remodelling and reorganisation, DNA repair and double-strand break repair for EBV-regulated KDM5B targets were identified. Significant terms were selected with adjusted *p-value* ≤ 0.05. **D) Involvement of KDM5B in EBVIR-enhancer-KDM5B targets.** The ChIP-Seq data (ENCSR000AQA-ENCFF455EXO) was used to confirm the enrichment of KDM5B binding in EBVIR-enhancer-KDM5B targets. **E) Enhancer activation for EBVIR-enhancer-KDM5B signature.** The H3K27ac and H3K27me3 were profiled at the promoter of the genes in the EBVIR-enhancer-KDM5B signature, showing the elevation of H3K27ac in EBV+ NPC cells. The H3K27me3 and H3K27ac bounded the nearby 50 kbps regions from the promoter of these genes in EBV- NPC and NPE samples. **F) Increase KDM5B protein expression level in C666 compared to NP69.** The western blot results showed enhanced KDM5B protein expression in C666 than in the NP69 control sample.

A

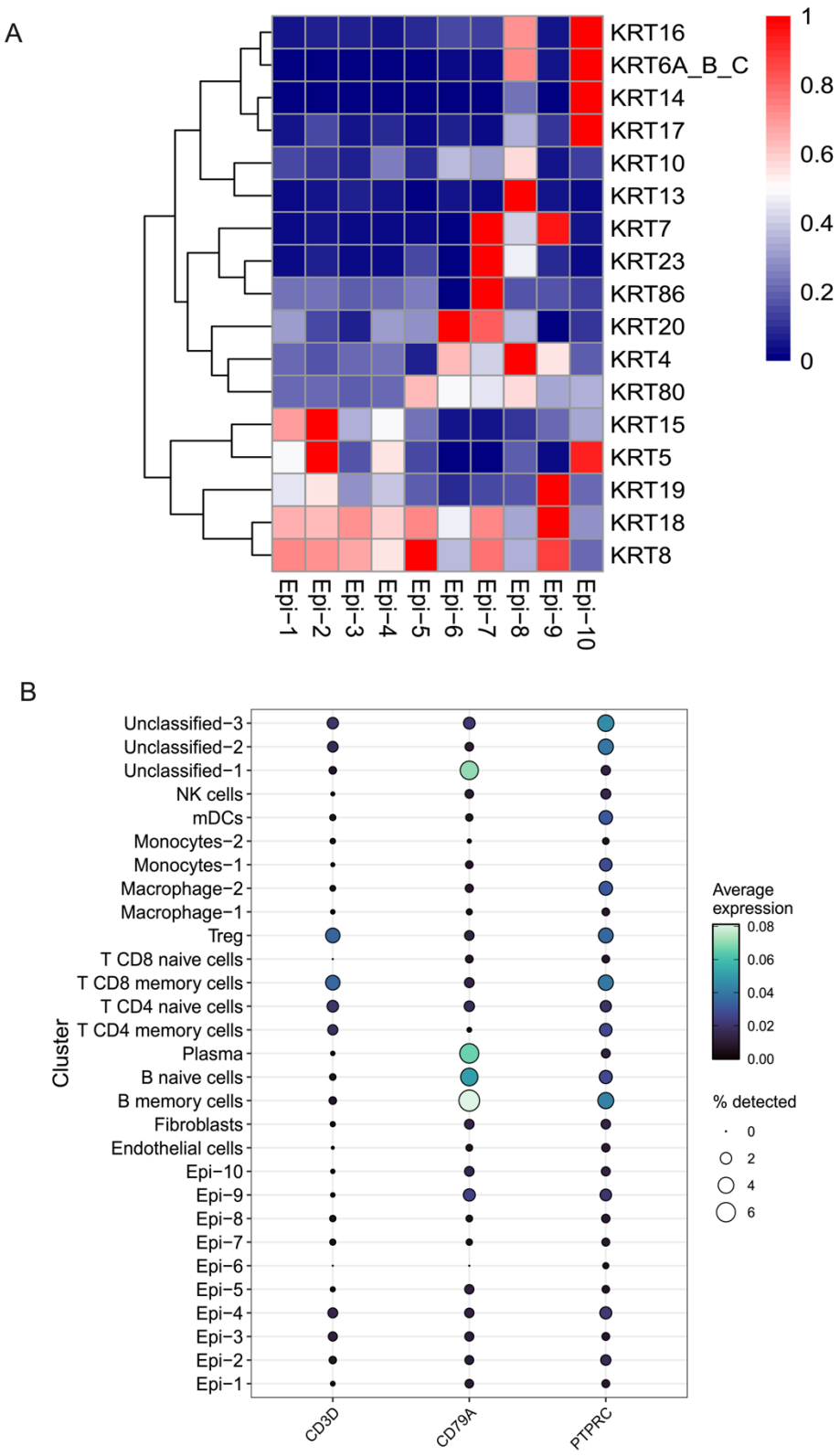

Proliferating immune cell
Myeloid cells
Fibroblasts
NK/NKT
CD8
Tregs
CD4
Naïve T
Plasma B
B cells
Epithelial cells and NPC cells
MKI67
LAMP3
CD68
LYZ
COL3A1
COL1A1
DCN
XCL2
KLRF1
TYROBP
KLRC1
CD8A
TNFRSF18
IL2RA
FOXP3
CD4
SELL
TCF7
CCR7
LINC02446
JCHAIN
IGKC
IGHG4
MZB1
SDC1
DERL3
XBP1
SSR4
CD40
CD19
PTPRC
CD79B
CD79A
MS4A1
LMP−1
RPMS1
TP63
KRT19
KRT18
KRT5
EPCAM
Normal epithelial cells
EBV−high NPC
EBV−low NPC
B cells
Plasma cells
Naïve T
CD4
Tregs
CD8
NK/NKT
Fibroblasts
Myeloid
Proliferating immune cells
Percent Expressed
0
25
50
75
Average Expression
2
1
0
−1

B


Cell type proportions
EBV-high NPC
0.00 0.25 0.50 0.75
EBV–low NPC
0.0 0.2 0.4 0.6
B cells
0.0 0.05 0.1 0.15 0.2
CD4
0.0 0.05 0.1 0.15 0.2
Naive T
0.00 0.05 0.10
Myeloid
0.00 0.25 0.50 0.75
Plasma B cells
0.00 0.25 0.50 0.75
CD8
0.0 0.1 0.2 0.3
Tregs
0.00 0.05 0.10 0.15 0.20
NK/NKT
0
Proliferating immune cells
0.0 0.05 0.1 0.15 0.20
Normal epithelial cells
0.0 0.05 0.1 0.15 0.2
Tumour regions
EBV-high NPC regions

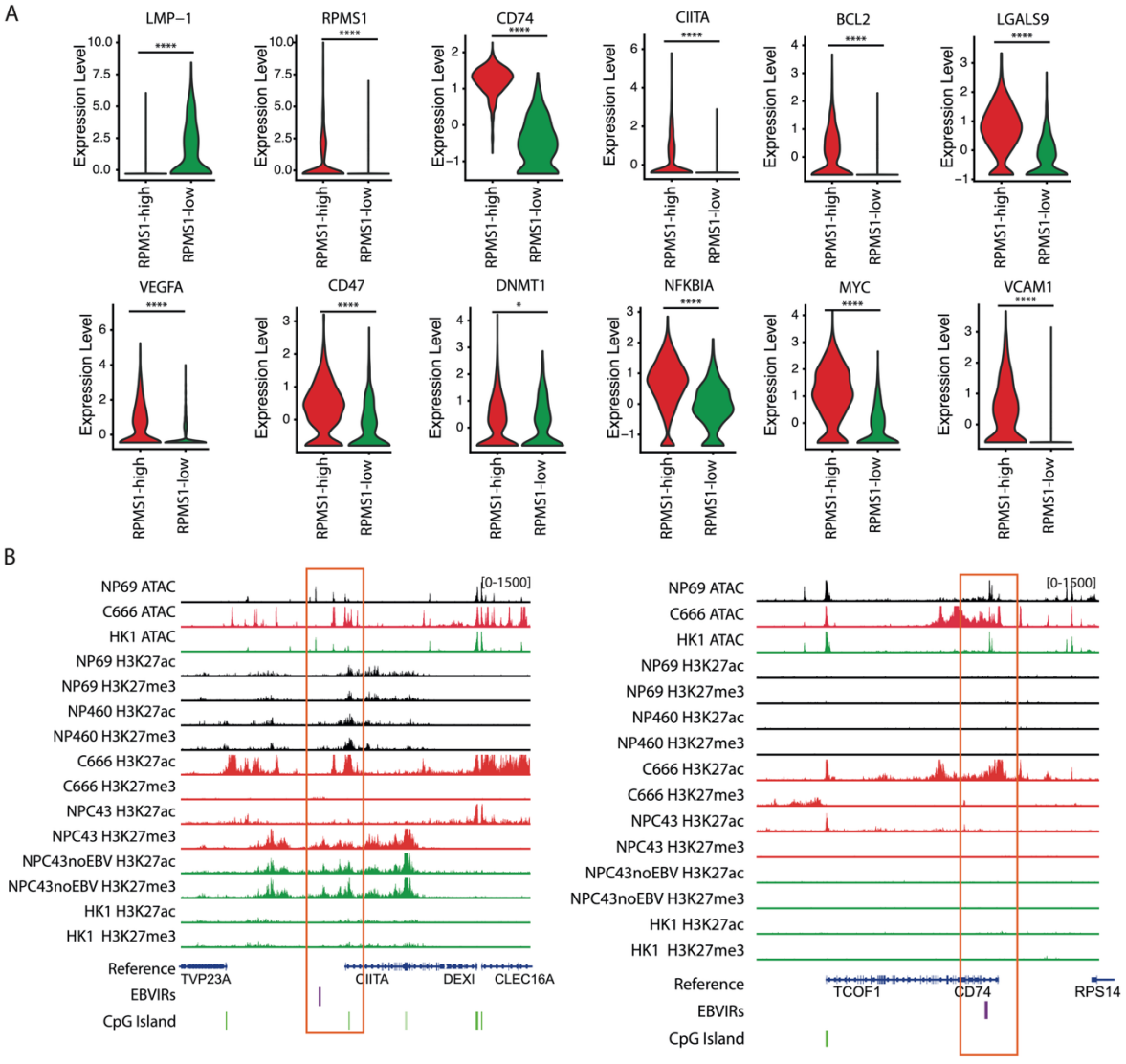
Supplementary Fig. 5 scRNA-seq cell-type classifications and estimated cell-type probability in 10X Visium Spatial Transcriptome data of an NPC patient.

A) Classifications of scRNA-seq clusters. Different cell types were derived from scRNA-seq data based on gene markers. B cells - *MS4A1, CD79A, CD79B, CD45, CD19, CD40; CD4* cells - *CD4*; CD8 cells - *CD8A*; EBV-high NPC - *RPMS1, LMP-1*; Epithelial cells and NPC cells - *EPCAM, KRT5, KRT18, KRT18, KRT19, TP63*; Fibroblasts - *DCN, COL1A1, COL3A1*; Myeloid cells - *LYZ, CD68, LAMP3, CD45*; Naïve T - *LINC02446, CCR7, TCF7, SELL;* NK/NKT - *KLRC1, TYROBP, KLRF1, XCL2*; Plasma B - *SSR4, XBP1, DERL3, SDC1, MZB1, IGHG4, IGKC, JCHAIN*; Proliferating immune cell - *MKI67*; Tregs - *FOXP3, IL2RA, TNFRSF18*. **B) Spatial deconvolution from scRNA-seq data estimating cell type proportion in a clinical NPC sample.** The probability of different cell types was estimated using the scRNA-seq anchor-based method for the deconvolution. The cancer cell regions were confirmed by H&E staining by a pathologist and framed in black. The tumour regions with high EBV latent gene expression (EBV-high NPC region) were framed in red.


A
KRT16
KRT6A_B_C
KRT14
KRT17
KRT10
KRT13
KRT7
KRT23
KRT86
KRT20
KRT4
KRT80
KRT15
KRT5
KRT19
KRT18
KRT8
Epi−1
Epi−2
Epi−3
Epi−4
Epi−5
Epi−6
Epi−7
Epi−8
Epi−9
Epi−10
1
0.8
0.6
0.4
0.2
0
B
Cluster
Unclassified−3
Unclassified−2
Unclassified−1
NK cells
mDCs
Monocytes−2
Monocytes−1
Macrophage−2
Macrophage−1
Treg
T CD8 naive cells
T CD8 memory cells
T CD4 naive cells
T CD4 memory cells
Plasma
B naive cells
B memory cells
Fibroblasts
Endothelial cells
Epi−10
Epi−9
Epi−8
Epi−7
Epi−6
Epi−5
Epi−4
Epi−3
Epi−2
Epi−1
CD3D
CD79A
PTPRC
Average expression
0.08
0.06
0.04
0.02
0.00
% detected
0
2
4
6

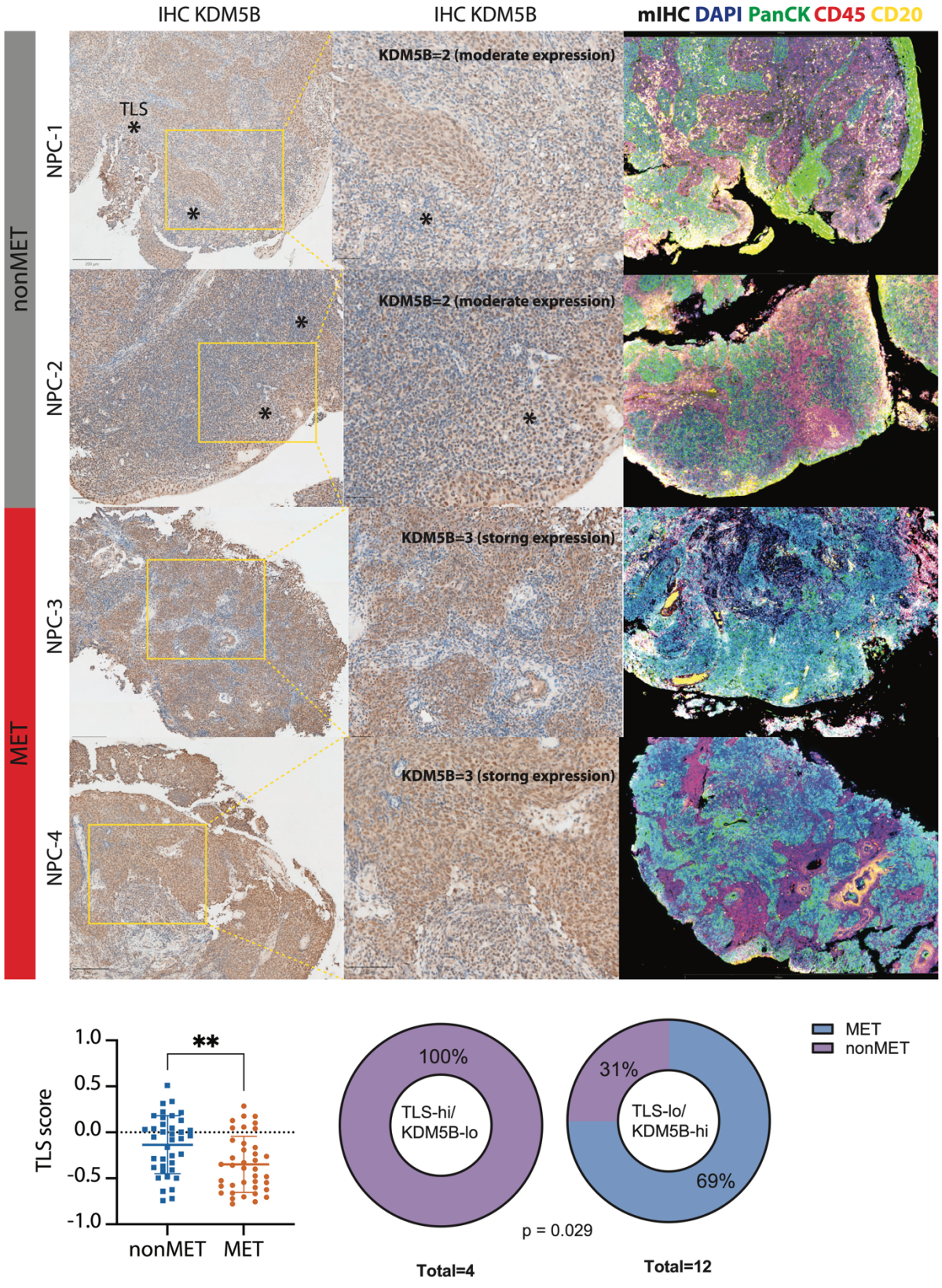
Supplementary Fig. 6 Expression of epithelial and immune cell markers among cell clusters in SMI.

A) KRT gene expression in epithelial cell clusters. The expression of KRT genes was evaluated among 10 epithelial cell clusters (Epi-1-10) for the SMI NanoString platform. Red indicates highly expressed, and blue indicates lowly expressed. **B) expression of immune cell markers among cell clusters.** The expression levels of *CD30D*, *CD79A* and *PTPRC* have been profiled among different cell clusters. Dot size indicates the percentage of detectable cells.


A
LMP−1
RPMS1
CD74
CIITA
BCL2
LGALS9
VEGFA
CD47
DNMT1
NFKBIA
MYC
VCAM1
Expression Level
RPMS1-high
RPMS1-low
B
NP69 ATAC
C666 ATAC
HK1 ATAC
NP69 H3K27ac
NP69 H3K27me3
NP460 H3K27ac
NP460 H3K27me3
C666 H3K27ac
C666 H3K27me3
NPC43 H3K27ac
NPC43 H3K27me3
NPC43noEBV H3K27ac
NPC43noEBV H3K27me3
HK1 H3K27ac
HK1 H3K27me3
[0-1500]
Reference
EBVIRs
CpG Island
TVP23A
CIITA
DEXI
CLEC16A
TCOF1
CD74
RPS14
C
GO
Epidermis Development
Mitochondrial ATP Synthesis Coupled Electron Transport
Skin Development
Aerobic Electron Transport Chain
Epidermal Cell Differentiation
Keratinocyte Differentiation
RPMS1-low-BP
DNA Binding
Sequence-Specific Double-Stranded DNA Binding
RNA Binding
Kinase Binding
Protein Kinase Binding
DNA-binding Transcription Activator Activity, RNA Polymerase II-specific
Ubiquitin-Like Protein Ligase Binding
Transcription Regulatory Region Nucleic Acid Binding
Ubiquitin Protein Ligase Binding
Guanyl Ribonucleotide Binding
RNA Polymerase II-specific DNA-binding Transcription Factor Binding
GTP Binding
Double-Stranded RNA Binding
MHC Class II Protein Complex Binding
Death Receptor Activity
MHC Class II Receptor Activity
Large Ribosomal Subunit rRNA Binding
RPMS1-high-MF
0
20
40
60
80
Odd ratio

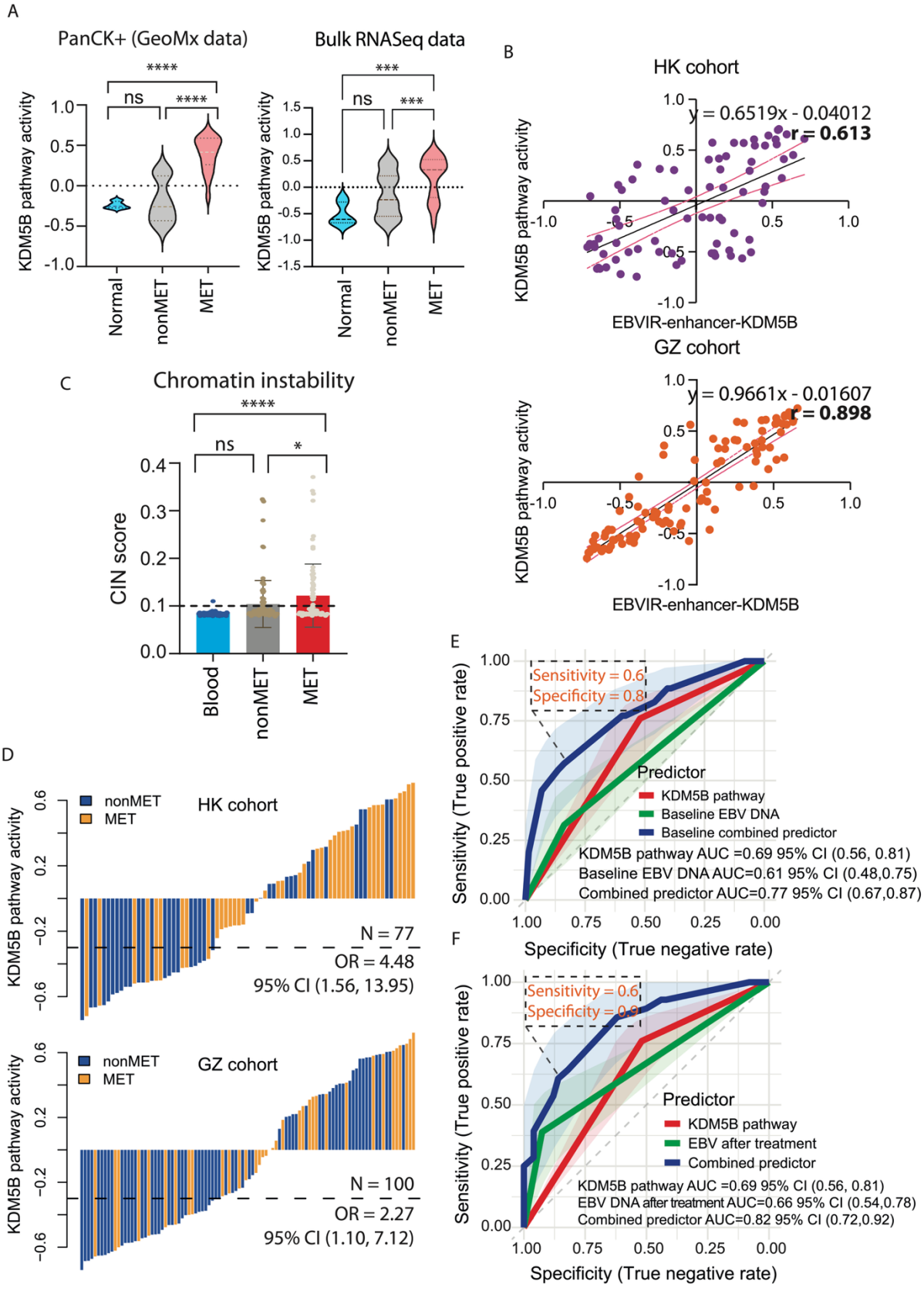
Supplementary Fig. 7 Selective genes upregulated in the cancer cells with elevated EBVIR-enhancer-KDM5B and *RPMS1*.

A) Gene expression profile for selected genes from scRNA-seq. *RPMS1* highly expressed genes identified from the SMI NanoString platform were further confirmed to be upregulated in the RPMS1-high cluster from scRNA-seq data, visualising in violin plots. A two-tailed unpaired t-test was performed to examine the significant difference between RPMS1-high (n = 1112) and -low (n = 584) clusters of individual genes, *adjusted *p-value* ≤ 0.05, ****adjusted *p-value* ≤ 0.0001. **B) Visualisation of EBVIRs at *CD74* and *CIITA* loci from IGV browser.** EBVIRs nearby regions have been associated with increased enhancer histone modification H3K27ac. The EBVIRs have also been associated with increased chromatin accessibility. **C) Functional annotations for RPMS1-high and RPMS1-low clusters of cells.** Genes commonly upregulated (n = 24) from both the SMI NanoString platform and scRNA-seq data were conducted for gene ontology analysis. Functional terms with *p-value* ≤ 0.05 were plotted.


A
IHC KDM5B
IHC KDM5B
mIHC DAPI PanCK CD45 CD20
nonMET
MET
NPC-1
NPC-2
NPC-3
NPC-4
TLS
KDM5B=2 (moderate expression)
KDM5B=2 (moderate expression)
KDM5B=3 (storng expression)
KDM5B=3 (storng expression)
TLS score
1.0
0.5
0.0
-0.5
-1.0
**
nonMET
MET
100%
TLS-hi/
KDM5B-lo
31%
TLS-lo/
KDM5B-hi
69%
p = 0.029
Total=4
Total=12
MET
nonMET

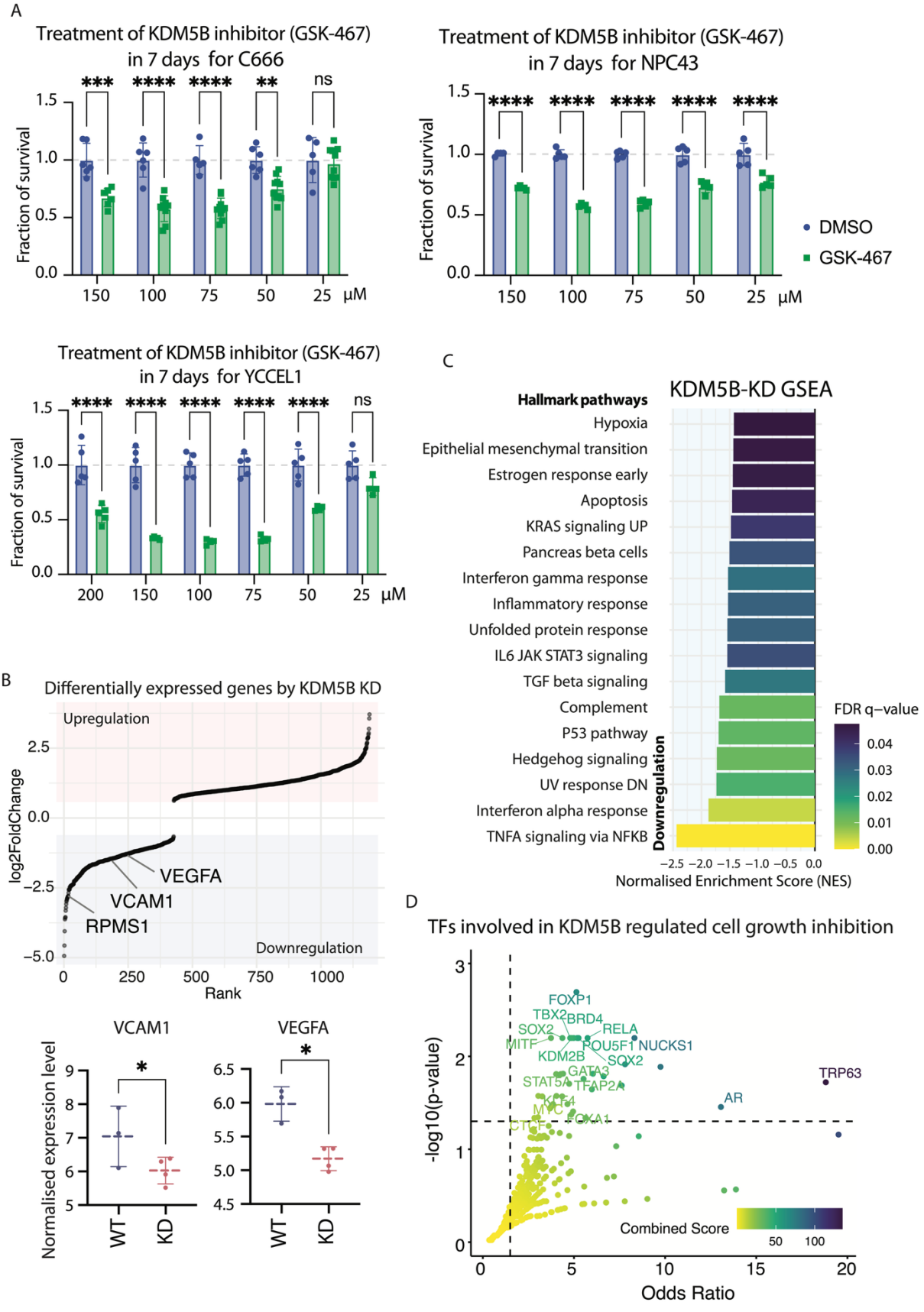
Supplementary Fig. 8 High level of KDM5B expression with low level of tertiary lymphoid structures (TLS) related to NPC metastasis.

A) The high expression levels of KDM5B and low levels of TLS were related to NPC metastasis. The IHC staining of KDM5B and multiplex IHC (mIHC) staining results of PanCK, CD20, CD45 in MET (n = 39) and nonMET (n = 38) patients. The intensity of KDM5B immunohistochemical staining was scored as 0: negative, 1: weak, 2: moderate, and 3: strong expression. A score of 3 was categorised as KDM5B-hi and a score of less than 3 as KDM5B-lo. As shown in the figure, NPC-1 and NPC-2 were scored 2, while NPC-3 and NPC-4 were scored 3. Tertiary lymphoid structures (TLS), marked * asterisk in the figure, is defined as an aggregate consisting of more than 100 CD20+CD45+PanCK- cells, surrounded by CD20-CD45+PanCK- cells, and located inside or adjacent to tumour nests based on multiplex immunofluorescent staining (DAPI: blue, PanCK: green, CD45: red, CD20: yellow). NPC cases (n =77) from HK cohort have been evaluated. TLS of every case were calculated from the whole-slide image and normalised by the tissue area. The expression of KDM5B and were evaluated by the HKU pathologist. TLS score was estimated by GSVA method using the gene signature derived from a previous study [2]. Cases with normalised TLS more than 3 were classified as TLS-hi and less than 3 as TLS-lo. A two-tailed unpaired t-test has been performed for TLS score in MET and nonMET patients, **adjusted *p-value* ≤ 0.01. The box plot showed MET patients has significantly less TLS compared to nonMET patients. The box plot shows mean ± standard deviation (SD) of the data. Fisher's exact test has been performed for KDM5B-hi & -low and TLS-high & -low groups. There is a statistically significant difference between high expression levels of KDM5B (KDM5B-hi) with low TLS (TLS-lo) and low expression levels of KDM5B (KDM5B-lo) with high level of TLS (TLS-hi), *p-value* ≤ 0.029. About 69% of MET patient samples contain high expression levels of KDM5B and low levels of TLS. However, it cannot be found in low expression levels of KDM5B with high TLS samples.


A
PanCK+ (GeoMx data)
Bulk RNASeq data
KDM5B pathway activity
ns
****
***
Normal
nonMET
MET
B
HK cohort
y = 0.6519x - 0.04012
r = 0.613
GZ cohort
y = 0.9661x - 0.01607
r = 0.898
EBVIR-enhancer-KDM5B
C
Chromatin instability
CIN score
*
Blood
D
nonMET
MET
N = 77
OR = 4.48
95% CI (1.56, 13.95)
N = 100
OR = 2.27
95% CI (1.10, 7.12)
E
Sensitivity = 0.6
Specificity = 0.8
Predictor
KDM5B pathway
Baseline EBV DNA
Baseline combined predictor
KDM5B pathway AUC =0.69 95% CI (0.56, 0.81)
Baseline EBV DNA AUC=0.61 95% CI (0.48,0.75)
Combined predictor AUC=0.77 95% CI (0.67,0.87)
Sensitivity (True positive rate)
Specificity (True negative rate)
F
EBV after treatment
Combined predictor
KDM5B pathway AUC =0.69 95% CI (0.56, 0.81)
EBV DNA after treatment AUC=0.66 95% CI (0.54,0.78)
Combined predictor AUC=0.82 95% CI (0.72,0.92)

Supplementary Fig. 9 KDM5B pathway activity.

A) KDM5B pathway activity from PanCK+ segments and bulk RNA-seq data. GSVA has been performed to evaluate the KDM5B pathway activity using bulk RNASeq data and GeoMx WTA data in NPC. One-way ANOVA with Bonferroni's correction was performed to test the significance, ****p-value* ≤ 0.001 and *****p-value* ≤ 0.0001; $n_{PanCK+normal} = 4$, $n_{PanCK+nonMET} = 10$, $n_{PanCK+MET} = 19$ and $n_{RNAseq\ normal} = 6$, $n_{RNAseq\ nonMET} = 38$, $n_{RNAseq\ MET} = 39$. **B) Pearson correlation plot for KDM5B pathway and EBVIR-enhancer-KDM5B activity.** A positive correlation has been established between KDM5B pathway activity and EBVIR-enhancer-KDM5B activity (HK; n = 77, r = 0.613, *p-value* ≤ 0.0001 and GZ; n = 100, r = 0.898, *p-value* ≤ 0.0001). **C) Chromatin instability score (CIN) in blood (n = 100), MET (n = 75) and nonMET (n = 46) from WES data in GZ and HK cohorts.** One-way ANOVA with Bonferroni's correction was performed to test the significance, **p-value* ≤ 0.05 and *****p-value* ≤ 0.0001. The error bar indicates mean ± standard deviation (SD) of the data. **D) KDM5B pathway activity for each patient in both HK (n = 77) and GZ (n = 100) cohorts.** The bar plots are ranked from low to high KDM5B pathway activity. **E) The receiver operating characteristic (ROC) curves for the KDM5B pathway activity, plasma EBV DNA copy at diagnosis and combination of the KDM5B pathway activity, plasma EBV DNA at diagnosis and Stage (n = 100).** The combination of the KDM5B pathway activity and plasma EBV DNA at diagnosis and Stage has an AUC of 0.77 (CI: 0.67 – 0.87) higher than the KDM5B pathway activity (AUC = 0.69, CI: 0.56 – 0.81) and plasma EBV DNA at diagnosis (AUC = 0.61, CI: 0.48 – 0.75) alone. **F) The ROC curves for the KDM5B pathway activity, plasma EBV DNA copy after treatment and combination of the KDM5B pathway activity, plasma EBV DNA at diagnosis, plasma EBV DNA after treatment and Stage (n = 77).** The combination of the KDM5B pathway, and plasma EBV DNA at diagnosis, plasma EBV DNA after treatment and Stage has an AUC of 0.82 (CI: 0.72 – 0.92) higher than the KDM5B pathway (AUC = 0.69, CI: 0.56 – 0.81), and plasma EBV DNA after treatment (AUC = 0.66, CI: 0.54 – 0.78) alone.


A
Treatment of KDM5B inhibitor (GSK-467) in 7 days for C666
Fraction of survival
150 100 75 50 25 µM
Treatment of KDM5B inhibitor (GSK-467) in 7 days for NPC43
DMSO
GSK-467
Treatment of KDM5B inhibitor (GSK-467) in 7 days for YCCEL1
200 150 100 75 50 25 µM
B
Differentially expressed genes by KDM5B KD
Upregulation
Downregulation
log2FoldChange
Rank
VEGFA
VCAM1
RPMS1
VCAM1
VEGFA
Normalised expression level
WT
KD
C
KDM5B-KD GSEA
Hallmark pathways
Hypoxia
Epithelial mesenchymal transition
Estrogen response early
Apoptosis
KRAS signaling UP
Pancreas beta cells
Interferon gamma response
Inflammatory response
Unfolded protein response
IL6 JAK STAT3 signaling
TGF beta signaling
Complement
P53 pathway
Hedgehog signaling
UV response DN
Interferon alpha response
TNFA signaling via NFKB
Downregulation
FDR q−value
Normalised Enrichment Score (NES)
D
TFs involved in KDM5B regulated cell growth inhibition
-log10(p-value)
Odds Ratio
FOXP1
TBX2
BRD4
SOX2
RELA
MITF
POU5F1
NUCKS1
KDM2B
SOX2
GATA3
STAT5A
TFAP2A
MYC
FOXA1
AR
TRP63
Combined Score

Supplementary Fig. 10 KDM5B inhibition and KDM5B knockdown regulated TFs and hallmark pathways.

A) MTT assay results for 7-day treatment of KDM5B inhibitor (GSK-467) in C666, NPC43, and YCCEL1 cells. Multiple concentrations of KDM5B inhibitors have been tested. The concentration-matched DMSO controls were included. The KDM5B inhibitor results were normalised to DMSO control. The experimental was performed for DMSO control (n = 5) and KDM5B treated (n = 5) groups. A two-way ANNOVA test was performed with Bonferroni correction, **adjusted *p-value* < 0.01, ***adjusted *p-value* < 0.001 and ****adjusted *p-value* < 0.0001. The error bar indicates mean ± standard deviation (SD) of the data. **B) The KDM5B-KD experiment identified differentially expressed genes related to metastasis.** The ranked gene list was calculated the log2-fold change (L2FC) from WT controls (n = 3) and KD (n = 4) samples. The positive (negative) value indicates upregulation (downregulation) in KDM5B-KD C666 cells. *VCAM1* and *VEGFA* showed significant downregulation after KDM5B knockdown with log2-fold change = -1.44 and -1.29, respectively, with adjusted *p-value* ≤ 0.05. The *RPMS1* expression has downregulation with L2FC = -2.68, adjusted *p-value* = 0.063. The box plot with the error bar represents the normalised expression level in the KDM5B-WT (n = 3) and -KD (n = 4) cells from RNA-seq data. **C) Functional terms enrichment for KDM5B-KD regulated genes from the molecular signature database (MSigDB).** GSEA was conducted based on the ranked L2FC of KDM5B-KD genes (n = 64) using MSigDB Hallmark 2020. Terms with significant enrichment for downregulated genes, FDR *q-value* ≤ 0.05 were plotted. **D) Putative transcription factors involvement in KDM5B regulating genes to suppress cell growth.** A CRISPR screening result (n = 2079 genes) identified from C666 with L2FC ≤ -2 and adjusted *p-value* ≤ 0.01 indicates the loss of function of genes shown significant inhibition on cancer cell growth compared to non-target gRNA control. Bulk RNA-seq identified differentially downregulated genes by KDM5B-KD in C666 with L2FC ≤ -0.05 and adjusted *p-value* ≤ 0.1. The overlapping (n = 36) genes from the CRISPR screening and knockdown experiments were subjected to TF analysis using ChEA 2022. The combined score was calculated ($\log_{10}$ (*p-value*) from Fisher's exact test × the Z score for deviation from the expected rank.

Supplementary Table 1: Differentially expressed genes for MET patients in both cohorts.

Index	Target name	GeoMx NPC cohort[1]			HK cohort		GZ cohort		Group
		Log2FC	P value	adjusted P value	log2FC	P value	log2FC	P value	
1	KIF18A	1.096	0.014	0.01	0.455	0.034	0.413	0.043	Metastasis-associated genes
2	LDLRAD3	1.457	0.001	<0.01	0.579	0.006	0.428	0.006	Metastasis-associated genes
3	TET3	0.853	0.030	0.01	0.379	<0.001	0.166	0.027	Metastasis-associated genes
4	PRELID3A	0.811	0.034	0.02	0.652	0.013	0.546	0.010	Metastasis-associated genes
5	EZH2	0.980	0.115	0.04	0.453	0.005	0.404	0.005	Metastasis-associated genes
6	LDHA	1.212	0.034	0.04	0.750	0.001	0.376	0.024	Metastasis-associated genes
7	ABTB2	1.160	0.009	0.01	0.613	0.039	0.529	0.018	Metastasis-associated genes
8	ANLN	1.508	0.011	0.01	0.517	0.026	0.735	0.003	Metastasis-associated genes
9	KIAA1217	1.074	0.009	<0.01	0.308	0.040	0.566	0.001	Metastasis-associated genes
10	WDR5	0.966	0.039	0.04	0.253	0.024	0.267	0.020	Metastasis-associated genes
11	TUBGCP4	1.098	0.039	0.04	0.367	<0.001	0.191	0.039	Metastasis-associated genes
12	TMEM120B	1.041	0.021	0.02	0.332	0.029	0.229	0.032	Metastasis-associated genes
13	PRPF38A	1.269	0.008	<0.01	0.151	0.047	0.122	0.008	Metastasis-associated genes
14	THOC5	1.088	0.024	0.03	0.256	0.002	0.112	0.046	Metastasis-associated genes
15	PRKRA	1.057	0.009	<0.01	0.172	0.038	0.195	0.004	Metastasis-associated genes
16	QSER1	0.610	0.062	0.04	0.305	0.013	0.359	0.003	Metastasis-associated genes
17	WDR43	1.302	0.018	0.01	0.251	0.046	0.486	<0.001	Metastasis-associated genes
18	NEDD1	2.161	0.005	0.01	0.609	0.015	0.524	0.020	Metastasis-associated genes
19	NCAPG2	1.011	0.012	0.02	0.372	0.024	0.348	0.014	Metastasis-associated genes
20	ENY2	1.171	0.049	0.03	0.389	0.001	0.229	0.023	Metastasis-associated genes
21	RNASEH1	1.050	0.006	0.01	0.218	0.014	0.162	0.024	Metastasis-associated genes
22	RANBP17	1.390	0.001	0.01	0.608	0.020	0.660	0.008	Metastasis-associated genes
23	DTNB	1.317	0.024	0.01	0.380	0.007	0.509	<0.001	Metastasis-associated genes
24	CYP27C1	1.101	0.003	0.01	0.925	0.012	0.981	0.003	Metastasis-associated genes
25	CSNK1E	1.189	0.014	0.01	0.382	0.001	0.180	0.047	Metastasis-associated genes
26	MMAB	0.852	0.085	0.02	0.330	0.012	0.204	0.045	Metastasis-associated genes
27	DENR	1.043	0.016	<0.01	0.309	0.010	0.188	0.040	Metastasis-associated genes
28	AIF1L	0.811	0.030	0.02	0.710	0.015	0.765	0.002	Metastasis-associated genes
29	CDK5RAP2	1.082	0.018	<0.01	0.274	0.013	0.204	0.008	Metastasis-associated genes
30	E2F8	0.715	0.055	0.04	0.501	0.003	0.420	0.020	Metastasis-associated genes
31	WDR12	1.270	0.011	<0.01	0.265	0.018	0.299	0.006	Metastasis-associated genes
32	TIMM8A	1.045	0.007	<0.01	0.278	0.040	0.404	<0.001	Metastasis-associated genes
33	HUS1	1.013	0.002	0.01	0.170	0.035	0.170	0.004	Metastasis-associated genes
34	DDX55	0.998	0.062	0.02	0.350	0.004	0.278	<0.001	Metastasis-associated genes
35	RRP15	0.835	0.030	0.01	0.278	0.027	0.334	0.002	Metastasis-associated genes
36	LRP6	1.377	0.016	0.02	0.424	0.042	0.399	0.024	Metastasis-associated genes
37	NUMBL	0.771	0.044	0.02	0.336	0.041	0.241	0.017	Metastasis-associated genes
38	STK38L	1.489	0.004	<0.01	0.398	0.017	0.275	0.042	Metastasis-associated genes
39	HMGXB4	1.253	0.024	0.02	0.225	0.045	0.241	0.003	Metastasis-associated genes
40	FOXRED1	0.796	0.012	0.02	0.276	0.009	0.182	0.035	Metastasis-associated genes
41	ANAPC7	1.171	0.005	<0.01	0.293	0.002	0.194	0.020	Metastasis-associated genes
42	UTP6	1.132	0.034	0.01	0.197	0.045	0.226	0.004	Metastasis-associated genes
43	ORC1	0.989	0.027	0.01	0.452	0.034	0.514	0.013	Metastasis-associated genes
44	DNMT3A	1.428	0.009	<0.01	0.365	0.005	0.240	0.028	Metastasis-associated genes
45	BIRC5	1.291	0.016	0.02	0.742	0.006	0.583	0.018	Metastasis-associated genes
46	RALA	0.943	0.014	0.01	0.317	0.027	0.336	0.002	Metastasis-associated genes
47	GPR75	1.038	0.005	<0.01	0.434	0.003	0.456	<0.001	Metastasis-associated genes
48	CMAS	1.078	0.062	0.01	0.381	0.015	0.296	0.016	Metastasis-associated genes
49	KNSTRN	0.898	0.021	<0.01	0.515	0.004	0.469	0.008	Metastasis-associated genes
50	ABCB9	0.990	0.027	0.01	0.459	0.007	0.261	0.042	Metastasis-associated genes
51	EXO1	0.843	0.039	0.02	0.530	0.044	0.512	0.046	Metastasis-associated genes
52	EIF4G2	0.968	0.024	0.02	0.273	0.020	0.218	0.020	Metastasis-associated genes
53	BRSK1	0.776	0.027	0.02	0.816	0.006	0.429	0.042	Metastasis-associated genes
54	ENAH	1.511	0.007	<0.01	0.681	<0.001	0.416	0.030	Metastasis-associated genes
55	ILF3	0.986	0.168	0.04	0.321	0.001	0.190	0.023	Metastasis-associated genes
56	PCNT	0.683	0.034	0.02	0.384	0.007	0.171	0.044	Metastasis-associated genes
57	PUS7	1.184	0.024	0.04	0.478	0.010	0.532	0.003	Metastasis-associated genes
58	C1orf21	0.673	0.094	0.02	0.477	0.005	0.374	0.004	Metastasis-associated genes
59	MYO19	1.200	0.039	0.01	0.517	0.002	0.319	0.033	Metastasis-associated genes
60	PHF14	1.496	0.062	0.02	0.328	0.020	0.247	0.015	Metastasis-associated genes
61	POLA1	1.126	0.014	0.03	0.295	0.017	0.221	0.022	Metastasis-associated genes
62	GRB10	1.046	0.062	0.04	0.525	0.013	0.646	0.003	Metastasis-associated genes
63	TRMT11	1.115	0.012	0.03	0.285	0.025	0.354	0.001	Metastasis-associated genes
64	GPATCH4	1.266	0.005	<0.01	0.361	0.007	0.239	0.029	Metastasis-associated genes
65	RRP1B	1.307	0.044	0.01	0.205	0.028	0.211	0.021	Metastasis-associated genes
66	ZNF519	0.727	0.055	0.04	0.448	0.006	0.282	0.028	Metastasis-associated genes
67	FABP5	1.428	0.007	<0.01	1.432	<0.001	0.607	0.009	Metastasis-associated genes
68	EMP2	1.388	0.014	0.03	0.582	0.003	0.441	0.035	Metastasis-associated genes
69	STAMBP	0.806	0.034	0.01	0.315	<0.001	0.141	0.008	Metastasis-associated genes
70	NOCT	1.117	0.009	<0.01	0.713	<0.001	0.462	0.011	Metastasis-associated genes
71	RFC5	1.016	0.039	0.02	0.356	0.033	0.437	0.002	Metastasis-associated genes
72	BEST3	0.769	0.007	<0.01	1.254	0.032	1.797	<0.001	Metastasis-associated genes
73	DONSON	1.072	0.016	<0.01	0.329	0.007	0.347	0.005	Metastasis-associated genes
74	MRPS25	0.942	0.030	0.03	0.239	0.025	0.231	0.011	Metastasis-associated genes
75	GJB2	1.421	0.034	0.03	1.643	<0.001	0.868	0.013	Metastasis-associated genes
76	FAM180A	0.725	0.030	0.03	0.939	0.024	0.667	0.037	Metastasis-associated genes
77	BRIP1	1.127	0.006	0.01	0.447	0.044	0.449	0.013	Metastasis-associated genes
78	GMPS	1.279	0.024	0.02	0.229	0.042	0.213	0.037	Metastasis-associated genes
79	POP1	1.071	0.004	0.01	0.650	<0.001	0.392	0.009	Metastasis-associated genes
80	DTD1	0.781	0.049	0.03	0.396	0.005	0.241	0.025	Metastasis-associated genes
81	DNAJC2	1.067	0.014	<0.01	0.310	0.010	0.218	0.008	Metastasis-associated genes
82	YES1	1.178	0.034	0.03	0.286	0.043	0.499	0.001	Metastasis-associated genes

[1] A two-tailed Mann–Whitney U test was performed to compare the distributions between the two groups. The p-value was adjusted using a permutation method with 100 iterations.

Supplementary Table 2: Differentially expressed genes for nonMET patients in both cohorts.

Index	Target name	GeoMx NPC cohort[2]			HK cohort		GZ cohort		Group
		Log2FC	P value	adjusted P value	log2FC	P value	log2FC	P value	
1	SIGLEC10	-0.721	0.013	0.01	-0.562	0.011	-0.316	0.038	Non-metastasis-associated genes
2	RNASET2	-0.879	0.025	<0.01	-0.350	0.034	-0.250	0.034	Non-metastasis-associated genes
3	KMO	-0.830	0.034	0.04	-0.705	0.013	-0.359	0.039	Non-metastasis-associated genes
4	MACROD2	-0.824	0.022	0.01	-0.710	0.041	-0.657	0.013	Non-metastasis-associated genes
5	LAX1	-1.025	0.016	0.01	-0.590	0.016	-0.402	0.016	Non-metastasis-associated genes
6	PPP1R16B	-0.882	0.021	0.01	-0.602	0.005	-0.467	0.006	Non-metastasis-associated genes
7	CYLD	-1.196	0.004	<0.01	-0.500	0.002	-0.252	0.032	Non-metastasis-associated genes
8	CD53	-0.822	0.028	0.02	-0.735	0.001	-0.365	0.008	Non-metastasis-associated genes
9	TSPAN4	-0.853	0.011	<0.01	-0.349	0.047	-0.221	0.048	Non-metastasis-associated genes
10	RHOH	-0.865	0.004	0.02	-0.593	0.015	-0.401	0.017	Non-metastasis-associated genes
11	CYBA	-1.040	0.007	<0.01	-0.412	0.018	-0.322	0.018	Non-metastasis-associated genes
12	GPR82	-1.485	0.005	<0.01	-1.142	0.001	-0.613	0.004	Non-metastasis-associated genes
13	LHFPL4	-1.112	0.016	<0.01	-1.574	0.013	-1.983	<0.001	Non-metastasis-associated genes
14	FUCA1	-1.153	0.001	<0.01	-0.748	0.002	-0.462	0.009	Non-metastasis-associated genes
15	TMEM243	-0.675	0.017	0.01	-0.473	0.002	-0.271	0.030	Non-metastasis-associated genes
16	GRIA4	-0.728	0.049	0.01	-1.069	0.049	-0.864	0.031	Non-metastasis-associated genes
17	CCR4	-0.667	0.021	0.01	-0.757	<0.001	-0.479	0.009	Non-metastasis-associated genes
18	CD27	-0.901	0.049	<0.01	-0.545	0.036	-0.538	0.009	Non-metastasis-associated genes
19	SIT1	-1.143	0.027	<0.01	-0.725	0.003	-0.465	0.015	Non-metastasis-associated genes
20	IL21	-0.923	0.004	<0.01	-1.037	0.003	-1.039	<0.001	Non-metastasis-associated genes
21	RAB33A	-0.800	0.010	0.01	-0.746	0.001	-0.458	0.002	Non-metastasis-associated genes
22	NCR3	-0.854	0.006	<0.01	-0.881	0.028	-0.708	0.004	Non-metastasis-associated genes
23	IKZF3	-0.867	0.014	0.01	-0.385	0.037	-0.377	0.007	Non-metastasis-associated genes
24	BTK	-0.951	0.011	<0.01	-0.571	0.019	-0.371	0.022	Non-metastasis-associated genes
25	SASH3	-0.912	0.020	0.01	-0.502	0.019	-0.429	0.008	Non-metastasis-associated genes
26	P2RY14	-0.849	0.021	0.02	-0.924	0.001	-0.540	0.014	Non-metastasis-associated genes
27	ITGAL	-1.108	0.030	<0.01	-0.488	0.018	-0.339	0.029	Non-metastasis-associated genes
28	DNAJB9	-0.768	0.021	0.01	-0.674	0.001	-0.254	0.044	Non-metastasis-associated genes
29	CD1C	-1.042	0.010	0.01	-1.098	0.006	-0.632	0.039	Non-metastasis-associated genes
30	IL10RA	-1.119	0.008	<0.01	-0.499	0.003	-0.313	0.014	Non-metastasis-associated genes
31	TBC1D9	-0.881	0.025	<0.01	-0.379	0.001	-0.299	0.010	Non-metastasis-associated genes
32	P2RY10	-1.097	0.007	<0.01	-0.857	<0.001	-0.401	0.035	Non-metastasis-associated genes
33	SCIMP	-1.194	0.003	<0.01	-0.768	0.011	-0.414	0.046	Non-metastasis-associated genes
34	AGMAT	-0.753	0.005	0.01	-0.585	0.013	-0.427	0.023	Non-metastasis-associated genes
35	RAB30	-0.746	0.030	0.04	-0.429	0.017	-0.283	0.045	Non-metastasis-associated genes
36	SELL	-1.246	0.009	<0.01	-0.718	0.022	-0.478	0.035	Non-metastasis-associated genes

[2] A two-tailed Mann–Whitney U test was performed to compare the distributions between the two groups. The p-value was adjusted using a permutation method with 100 iterations.

Supplementary Table 3: Two-tailed univariate logistic regression analysis.

Univariate logistic regression analysis (GZ+HK cohort)[3]						
Parameters	**Group**	**#Patient**	**HR**	**95% CI**		**P value**
KDM5B pathway activity	Low (≤-0.3)	71	1.00			
	High (>-0.3)	106	3.43	1.78	6.61	<0.001*
CIN Score	Low (≤-0.1)	101	1.00			
	High (>0.1)	75	2.26	1.17	4.39	0.016*
C-reactive protein (CRP) (mg/L)	Low (≤3 mg/L)	64	1.00			
	High (>3 mg/L)	31	3.54	1.44	8.68	0.006*
Neutrophil-lymphocyte ratio (NLR)	Low (≤2.5)	53	1.00			
	High (>2.5)	45	3.57	1.50	8.52	0.004*
Fibrinogen (FBG) (g/L)	Low (≤3.34 g/L)	58	1.00			
	High (>3.34 g/L)	21	5.11	1.76	14.84	0.003*
EBV DNA copy (baseline) (IU/mL)	Low (≤10000 IU/mL)	86	1.00			
	High (>10000 IU/mL)	23	2.37	0.92	6.09	0.074
EBV DNA copy (after treatment) (IU/mL)	Low (=0 IU/mL)	72	1.00			
	High (>0 IU/mL)	18	7.96	2.35	26.92	0.001*
Stage	Stage I-II	19	1.00			
	Stage III	98	1.42	0.47	4.29	0.532
	Stage IV	60	4.50	1.43	14.17	0.01*
Age (years)	≤50	112	1.00			
	>50	64	0.80	0.43	1.49	0.472
Gender	Female	40	1.00			
	Male	137	1.72	0.82	3.61	0.153

[3] Univariate logistic regression analyses were performed using a two-tailed approach to assess the association between each independent variable and the binary outcome. Odds ratios (ORs), 95% confidence intervals (CIs), and p-values were reported. P values < 0.05 were considered statistically significant and are marked with an asterisk (*).

Supplementary Table 4: Two-tailed multivariable logistic regression analysis

Multivariable logistic regression analysis at diagnosis (GZ+HK cohort)[4]						
Parameters	**Group**	**#Patient**	**HR**	**95% CI**		**P value**
KDM5B pathway activity	Low (≤-0.3)	44	1.00			
	High (>-0.3)	46	5.51	1.81	16.8	0.003*
EBV DNA copy (baseline) (IU/mL)	Low (≤40 IU/mL)	72	1.00			
	High (>40 IU/mL)	18	6.22	1.48	26.15	0.013*
Stage	Stage I-II	10	1.00			0.002*
	Stage III	50	2.58	0.27	24.43	0.409
	Stage IV	30	14.63	1.4	152.71	0.025*
Multivariable logistic regression analysis (GZ+HK cohort)[4]						
Parameters	**Group**	**#Patient**	**HR**	**95% CI**		**P value**
KDM5B pathway activity	Low (≤-0.3)	38	1.00			
	High (>-0.3)	40	4.65	1.41	15.35	0.012*
EBV DNA copy (baseline) (IU/mL)	Low (≤10000 IU/mL)	64				0.199
	High (>10000 IU/mL)	14	2.62	0.6	11.36	
EBV DNA copy (after treatment) (IU/mL)	Low (=0 IU/mL)	64	1.00			
	High (>0 IU/mL)	14	7.97	1.52	39.05	0.011*
Stage	Stage I-II	9	1.00			0.198
	Stage III	46	2.26	0.24	21.53	0.478
	Stage IV	23	6.12	0.56	67.23	0.139

[4] A multivariable logistic regression analysis was conducted to assess the association between multiple predictor variables and the binary outcome. Hazard ratios (HRs), 95% confidence intervals (CIs), and p-values were reported for each variable. P values < 0.05 were considered statistically significant and are marked with an asterisk (*).

Supplementary Table 5: Summary of blood test results and clinical parameters in association with distant metastasis in GZ NPC patients.

Characteristic[5]	GZ cohort (N=100)					
	All	MET		nonMET		
	Number	Number	Average±SD	Number	Average±SD	P value
Heamoglobin concentration (HGB) (g/L)	99	36	138.61±16.59	63	145.18±21.96	0.122
Neutrophil-lymphocyte ratio (NLR)	99	36	3.27±1.68	63	2.57±1.5	0.035*
Platelet count (PLT)	99	36	246.9±78.95	63	250.16±83.98	0.850
Albumin concentration (ALB) (g/L)	97	36	43.3±4.92	61	44.53±3.59	0.157
Lactate Dehydrogenase (LDH) concentration (U/L)	96	35	196.59±61.39	61	185.96±66.44	0.440
C-reactive protein concentration (LRP) (mg/L)	95	36	6.45±9.45	59	2.8±3.62	0.032*
Prothrombin time (PT) (sec)	79	27	11.56±0.87	52	11.51±2.9	0.931
Prothrombin time activity (PTA) (%)	75	27	98.85±19.7	48	103.66±21.63	0.344
International normalized ratio (INR)	79	27	1.01±0.08	52	1±0.25	0.854
Activated partial thrmoboplastin time (APTT) (sec)	79	27	26.67±3.71	52	26.23±4.13	0.648
Fibrinogen (Fbg) (g/L)	79	27	3.4±1	52	2.93±0.72	0.038*
Thrombin time (TT) (sec)	79	27	18.1±1.26	52	18±1.44	0.757
D-dimer (D-D) (μg/mL)	46	21	0.7±0.82	25	0.45±0.49	0.204
Fibrinogen degradation products (FDP) (μg/mL)	47	21	2.34±2.16	26	2.1±3.82	0.793
Antithrombin III (ATIII) (%)	41	18	95.04±26.01	23	100.29±9.59	0.376

[5] Data are presented as mean ± standard deviation. Comparisons between the MET and nonMET groups were performed using unpaired two-tailed t-tests. P values < 0.05 were considered statistically significant and are marked with an asterisk (*).

Supplementary Table 6: Summary of blood test results and clinical parameters in association with KDM5B pathway activity in GZ NPC patients.

Clinical characteristics	Total # patient (N)	KDM5B pathway activity[6]				P value
		Low		High		
		#Patient	Average±SD (or %)	#Patient	Average±SD (or %)	
Heamoglobin concentration (HGB) (g/L)	99	40	143.38±15.07	59	142.39±23.35	0.815
Neutrophil-lymphocyte ratio (NLR)	99	40	2.55±1.54	59	3.01±1.63	0.164
Platelet count (PLT)	99	40	257.71±96.25	59	243.05±70.62	0.384
Albumin concentration (ALB) (g/L)	97	39	44.46±4.86	58	43.81±3.63	0.453
Lactate Dehydrogenase (LDH) concentration (U/L)	96	39	201.53±80.37	57	181.83±50.2	0.143
C-reactive protein concentration (CRP) (mg/L)	95	37	2.39±3.03	58	5.32±8.01	0.014*
Prothrombin time (PT) (sec)	79	33	11.26±0.91	46	11.72±3.05	0.401
Prothrombin time activity (PTA) (%)	75	32	102.74±18.84	43	101.33±22.6	0.775
International normalized ratio (INR)	79	33	0.98±0.08	46	1.02±0.26	0.452
Activated partial thrmoboplastin time (APTT) (sec)	79	33	26.12±4.18	46	26.57±3.85	0.626
Fibrinogen (Fbg) (g/L)	79	33	3.09±0.89	46	3.09±0.83	0.999
Thrombin time (TT) (sec)	79	33	18.34±1.41	46	17.82±1.33	0.101
D-dimer (D-D) (μg/mL)	46	18	0.61±0.51	28	0.53±0.75	0.713
Fibrinogen degradation products (FDP) (μg/mL)	47	18	1.59±1.23	29	2.59±3.88	0.296
Antithrombin III (ATIII) (%)	41	17	101.47±10.78	24	95.51±22.45	0.318
EBV DNA copy (baseline)	80	33	41.3	47	58.8	
Low (≤10000 IU/mL)	59	27	45.8	32	54.2	0.264
High (>10000 IU/mL)	21	6	28.6	15	71.4	
EBV DNA copy (within 6 month after treatment)	58	26		32		
Not detectable (≤40 IU/mL)	46	22		24		0.567
Detectable (>40 IU/mL)	12	4		8		
Age (years)			42±12		42±14	0.973
Gender	100	40		60		
Female	25	10	40.0	15	60.0	1
Male	75	30	40.0	45	60.0	
Stage	100	40		60		
I-II	14	7	50.0	7	50.0	0.711
III	42	16	38.1	26	61.9	
IV	44	17	38.6	27	61.4	

[6] Data are presented as mean ± standard deviation or percentage, as appropriate. Comparisons between low and high KDM5B pathway activity groups were performed using unpaired two-tailed Student's t-tests for continuous variables and Chi-square or Fisher's exact tests (if n < 5 in one or more groups) for categorical variables. P values < 0.05 were considered statistically significant and are marked with an asterisk (*).

Supplementary Table 7: Summary of blood test results and clinical parameters in association with overall stage in GZ NPC patients.

Clinical characteristics[7]	P value	Total # patient	Stage I-II		Stage III		Stage IV		P value (Stage I vs. Stage III)	P value (Stage I vs. Stage IV)	P value (Stage III vs. Stave IV)
			# Patient	Mean±SD (or %)	# Patient	Mean±SD (or %)	# Patient	Mean±SD (or %)			
Heamoglobin concentration (HGB) (g/L)	0.056	99	14	154.22±30.29	41	142.62±18.88	44	139.32±16.68	0.098	0.022*	0.394
Neutrophil-lymphocyte ratio (NLR)	0.046*	99	14	1.92±0.6	41	2.8±1.58	44	3.13±1.74	0.004*	<0.001*	0.365
Platelet count (PLT)	0.150	99	14	210.24±47.07	41	251.87±96.18	44	258.6±73.36	0.127	0.024*	0.717
Albumin concentration (ALB) (g/L)	0.154	97	13	46.12±2.27	41	43.9±3.57	43	43.62±4.93	0.040*	0.014*	0.770
Lactate Dehydrogenase (LDH) concentration (U/L)	0.205	96	13	168.51±34.92	41	184.27±66.44	42	201.87±68.27	0.418	0.025*	0.238
C-reactive protein concentration (CRP) (mg/L)	0.026*	95	13	2.04±4.06	40	2.74±3.18	42	6.22±8.94	0.526	0.110	0.021*
Prothrombin time (PT) (sec)	0.701	79	14	11.12±0.65	34	11.76±3.56	31	11.47±0.85	0.514	0.185	0.662
Prothrombin time activity (PTA) (%)	0.734	75	13	101.78±14.04	31	104.07±26.41	31	99.85±17.24	0.769	0.723	0.458
International normalized ratio (INR)	0.737	79	14	0.97±0.06	34	1.02±0.3	31	1±0.07	0.535	0.141	0.764
Activated partial thrmoboplastin time (APTT) (sec)	0.823	79	14	26.19±4	34	26.14±4.65	31	26.73±3.19	0.975	0.626	0.549
Fibrinogen (Fbg) (g/L)	0.004*	79	14	2.65±0.36	34	2.94±0.61	31	3.46±1.08	0.108	0.001*	0.023*
Thrombin time (TT) (sec)	0.936	79	14	17.91±1.4	34	18.06±1.48	31	18.06±1.29	0.747	0.732	0.992
D-dimer (D-D) (μg/mL)	0.073	46	7	0.23±0.15	15	0.39±0.33	24	0.77±0.83	0.242	0.101	0.099
Fibrinogen degradation products (FDP) (μg/mL)	0.183	47	7	3.74±7.17	16	1.19±0.79	24	2.43±2.16	0.384	0.650	0.034
Antithrombin III (ATIII) (%)	0.963	41	6	96.57±10.27	15	98.93±9.57	20	97.7±25.1	0.621	0.916	0.858
Chromosome instability (CIN) score	0.705	99	14	0.1±0.02	42	0.11±0.06	43	0.11±0.07	0.398	0.429	0.928
KDM5B pathway activity	0.973	100	14	-0.07±0.54	42	-0.05±0.46	44	-0.07±0.45	0.875	0.989	0.827
Age	0.096	99	14	45±15	41	44±12	44	39±13	0.969	0.164	0.164
Gender	0.778								0.862	1.000	0.709
Female		25	3	12.0	12	48	10	40.0			
Male		75	11	14.7	30	40	34	45.3			

[7] Data are presented as mean ± standard deviation (SD) or percentage (%), as appropriate. Statistical comparisons among cancer stages (Stage I–II, Stage III, and Stage IV) were performed using one-way ANOVA or Kruskal-Wallis test for continuous variables, and chi-square or Fisher's exact test for categorical variables. Pairwise comparisons between stages were conducted with post hoc tests, and p-values < 0.05 were considered statistically significant. Asterisks (*) indicate statistically significant differences.

Supplementary Table 8: Summary of blood test results and clinical parameters in association with plasma EBV DNA copies at diagnosis and after treatment in GZ NPC patients.

Clinical characteristics[8]	Total # patient (N)	EBV DNA at baseline					Total # patient (N)	EBV DNA after treatment				
		Low		High		P value		Low		High		P value
		#Patient	Average±SD (or %)	#Patient	Average±SD (or %)			#Patient	Average±SD (or %)	#Patient	Average±SD (or %)	
Heamoglobin concentration (HGB) (g/L)	80	59	143.51±23.52	21	142.42±13.51	0.841	58	46	141±14.54	12	137.18±19.18	0.453
Neutrophil-lymphocyte ratio (NLR)	80	59	2.77±1.66	21	2.8±1.37	0.936	58	46	3.06±1.9	12	2.85±1.47	0.732
Platelet count (PLT)	80	59	252.2±84.61	21	231.7±60.87	0.312	58	46	249.53±94.08	12	266.14±66.91	0.569
Albumin concentration (ALB) (g/L)	78	57	44.9±3.76	21	43.05±2.52	0.040*	56	44	44.71±3.82	12	42.78±5.65	0.171
Lactate Dehydrogenase (LDH) concentration (U/L)	77	56	184.8±65.82	21	209.89±65.25	0.140	55	44	197.94±77.79	11	167.92±40.23	0.223
C-reactive protein concentration (CRP) (mg/L)	76	55	2.68±3.15	21	5.46±6.16	0.011*	55	43	3.37±4.36	12	6.72±10.45	0.102
Prothrombin time (PT) (sec)	64	47	11.64±3.03	17	11.21±0.73	0.566	44	36	11.67±3.42	8	11.74±0.76	0.958
Prothrombin time activity (PTA) (%)	60	44	102.37±22.23	16	100.12±15.5	0.711	42	35	103.43±21.95	7	102.63±18.75	0.929
International normalized ratio (INR)	64	47	1.01±0.26	17	0.98±0.07	0.598	44	36	1.02±0.29	8	1.01±0.06	0.967
Activated partial thrmoboplastin time (APTT) (sec)	64	47	26.75±4.4	17	25.17±3.25	0.181	44	36	26.72±4.34	8	26.78±3.15	0.973
Fibrinogen (Fbg) (g/L)	64	47	2.88±0.61	17	3.38±1.07	0.023*	44	36	3.02±0.77	8	3.07±1.08	0.885
Thrombin time (TT) (sec)	64	47	18.05±1.45	17	18.28±1.04	0.559	44	36	18.28±1.5	8	17.95±1.51	0.580
D-dimer (D-D) (μg/mL)	34	25	0.6±0.83	9	0.64±0.37	0.896	29	24	0.61±0.85	5	0.75±0.39	0.721
Fibrinogen degradation products (FDP) (μg/mL)	35	26	2.52±4.09	9	2.26±0.93	0.853	30	25	2.47±4.08	5	3.46±2.19	0.606
Antithrombin III (ATIII) (%)	31	23	97.59±23.51	8	99.74±9.86	0.806	26	22	101.66±11.13	4	105.78±9.98	0.498
Chromosome instability (CIN) score	79	58	0.11±0.06	21	0.13±0.07	0.320	57	45	0.11±0.06	12	0.12±0.07	0.780
KDM5B pathway activity	80	59	-0.07±0.49	21	-0.01±0.46	0.606	58	46	-0.09±0.49	12	-0.12±0.37	0.851
Age (years)	80	59	39±11	21	50±16	0.001*	58	46	41±12	12	32±12	0.022*
Gender												
Female	20	16	80	4	20	0.566	16	12	75	4	25	0.72
Male	60	43	71.7	17	28.3		42	34	81	8	19	
Stage												
I-II	17	9	52.9	8	47.1	0.456	8	8	100	0	0	0.063
III	49	29	59.2	20	40.8		23	20	87.0	3	13	
IV	40	22	55	18	45		27	18	66.7	9	33.3	

[8] Data are presented as mean ± standard deviation or percentage, as appropriate. Comparisons between low and high EBV DNA levels at baseline and after treatment were performed using unpaired two-tailed t-tests for continuous variables and Chi-square or Fisher's exact tests for categorical variables. P values < 0.05 were considered statistically significant and are marked with an asterisk (*).

Supplementary References

1. Ka-Yue Chow, L., et al., *Epigenomic landscape study reveals molecular subtypes and EBV-associated regulatory epigenome reprogramming in nasopharyngeal carcinoma.* EBioMedicine, 2022. **86**: p. 104357.

2. Liu, Y., et al., *Single-cell and spatial transcriptome analyses reveal tertiary lymphoid structures linked to tumour progression and immunotherapy response in nasopharyngeal carcinoma.* Nat Commun, 2024. **15**(1): p. 7713.