Epigenomic landscape study reveals molecular subtypes and EBV-associated regulatory epigenome reprogramming in nasopharyngeal carcinoma

Larry Ka-Yue Chow,a Dittman Lai-Shun Chung,a Li Hua Tao,a Kui Fat Chan,a Stewart Yuk Tung,a Roger Kai Cheong Ngan,a,b,c,d Wai Tong Ng,a,d,e,f Anne Wing-Mui Lee,a,c,d,e,f Chun Chung Yau,a,d,e,f Dora Lai-Wan Kwong,a,d Victor Ho-Fun Lee,a,d Ka-On Lam,a Jiayan Liu,b Honglin Chen,b Wei Dai,a,d,e,f and Maria Li Lung*a,**

aDepartment of Clinical Oncology, University of Hong Kong, Hong Kong (SAR), PR China
bDepartment of Clinical Pathology, Tuen Mun Hospital, Hong Kong (SAR), PR China
cDepartment of Clinical Oncology, Tuen Mun Hospital, Hong Kong (SAR), PR China
dDepartment of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong (SAR), PR China
eDepartment of Clinical Oncology, Pamela Youde Nethersole Eastern Hospital, Hong Kong (SAR), PR China
fUniversity of Hong Kong-Shenzhen Hospital, Shenzhen, PR China
gDepartment of Oncology, Princess Margaret Hospital, Hong Kong (SAR) PR China
hDepartment of Microbiology, University of Hong Kong, Hong Kong (SAR), PR China

Summary

Background Epstein-Barr virus (EBV) latent infection is associated with genome-wide epigenomic changes in several malignancies, but its role in epigenetic dysregulation remains unclear in nasopharyngeal carcinoma (NPC).

Methods To investigate EBV-associated epigenetic dysregulation, we performed a multi-omics study by integrating whole-genome bisulfite sequencing (WGBS), assay for transposase-accessible chromatin using sequencing (ATAC-Seq), whole-exome sequencing (WES), and single-cell RNA sequencing (scRNA-Seq) data.

Findings In addition to the known global DNA hypermethylated subtype, we discovered a novel subtype with global hypomethylation in EBV + NPC. The consistent EBV-specific differentially methylated regions (EBV-DMRs) in the human genome were identified from both subtypes and associated with loss of CTCF binding (P < 2.2e-16). Important, CTCF is a master chromatin regulator and CTCF protein was reduced in 45% of NPC cases, especially in those with advanced NPC (Stage IV vs. others: 62% vs. 38%, P = 0.034). This result links EBV with chromatin changes. The ATAC-Seq data suggest regulatory epigenome reprogramming through chromatin accessibility changes in EBV + NPC with altered CTCF binding and the switch of transcription factor binding from differentiation-associated KLF/SP family to the innate and adaptive immunity-related NF-κB and IRF families. Detailed chromatin accessibility analysis identified a potential EBV target gene CD74, which mediated EBV-specific cell-cell communications in the tumor microenvironment (TME) and was strongly correlated with T cell exhaustion (r² = 0.55).

Interpretation Our study reveals the unexpected epigenetic heterogeneity, providing insights into NPC pathogenesis and highlighting the involvement of host factors in virus-associated epigenetic changes. EBV infection is associated with epigenome reprogramming and may promote immune evasion.

Funding This study was funded by the Hong Kong Research Grants Council grant (AoE/M-06/08) to MLL, General Research Fund (17103218 and 17102619) and seed funding for basic research (20161159158) to WD, and General Research Fund (17119618) to HC.

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Epstein-Barr virus; Nasopharyngeal carcinoma; Host-virus interaction; Whole-genome bisulfite sequencing; Chromatin accessibility; Tumor microenvironment
Research in context

Evidence before this study
NPC has a highly distinct geographical distribution, endemic in Southeast Asia but rare worldwide. In the endemic regions, EBV is detected in >90% of NPC cases. Genome-wide association studies and epidemiology models suggest that genetic predisposition, environmental factors, and EBV high-risk subtypes are strong risk factors of NPC. EBV establishes Type II latency in NPC and is associated with global DNA hypermethylation in EBV + gastric cancer. Previous methylation studies in NPC using Illumina HumanMethylation450K BeadChip array profiled a subset (~2%) of CpG-dense sites of the human genome, suggesting a dominant hypermethylation pattern, which is a unique feature in EBV-associated epithelial malignancy and is proposed to be a critical step in tumorigenesis in the current NPC pathogenesis model. However, comprehensive whole-genome methylation profiling is still lacking in NPC. The detailed role of EBV in epigenetic dysregulation in NPC remains unclear: to what extent EBV is associated with epigenetic changes, how these changes arise, how they link to host factors, and their biological impact are largely unknown in NPC.

Added value of this study
To the best of our knowledge, this multi-omics landscape study revealed a new NPC epigenomic subtype with global hypomethylation. We confirmed the existence of the CpG island methylator phenotype in EBV + NPC regardless of global methylation status. Distinct from gastric cancer, in vitro EBV infection of normal immortalized cell lines derived from nonmalignant nasopharyngeal epithelial tissues alone does not induce hypermethylation observed in NPC and EBV + gastric cancer. Our studies now report a reduction of CTCF protein and loss of CTCF binding is linked to EBV-associated hypermethylation in NPC. These results highlight the involvement of cellular context in shaping the host epigenome together with EBV, providing novel insights into the current NPC pathogenesis model. The chromatin accessibility associated with EBV was comprehensively evaluated. Importantly, we provide evidence that EBV latent infection is associated with regulatory epigenome reprogramming, leading to the EBV-specific cell communications between cancer cells and immune cells in the tumor microenvironment (TME), which maybe relevant to immune suppression. Importantly, this study links EBV-associated epigenetic changes, TME, and immune evasion together.

Implications of all the available evidence
Our study highlights the importance of molecular subtyping of EBV + NPC due to genetic and epigenetic heterogeneity. Whether this newly discovered global hypomethylation subtype is relevant to prognosis, diagnosis, and treatment response warrants further investigation. Host factors and EBV are associated with regulatory epigenome reprogramming. Further studies to elucidate detailed mechanisms underlying these epigenetic changes are necessary. In addition, the EBV-specific cell communications in the immune suppressive TME may be therapeutic targets in NPC.

Introduction
The ubiquitous Epstein-Barr virus (EBV) is detected in a variety of lymphoepithelial malignancies including Burkitt’s lymphoma, Hodgkin’s lymphoma, NK cell lymphomas, T cell lymphomas, diffuse large B cell lymphoma, and epithelial tumors such as nasopharyngeal carcinoma (NPC) and gastric cancers, of which about 10% are associated with EBV (EBVaGC). The critical involvement of host genetics in EBV-associated malignancies is reflected by the uneven geographical distribution, different ethnic incidence rates, germline single nucleotide polymorphisms (SNPs) at the major histocompatibility complex (MHC) regions associated with immune recognition of viruses, and recurrent genomic alterations (e.g., PIK3CA and CDKN2A) that are frequently associated with the EBV-associated cancer.1–3

In NPC, 95% of cases from the endemic regions are EBV-positive. EBV establishes type II latency in NPC; the EBV genome replicates slowly and synchronously with the host chromosomes, while expressing a subset of oncogenic viral proteins (e.g., LMP1) that modulate host signaling pathways and epigenetic machinery. NPC has a distinctive hypermethylation pattern presumably associated with EBV, as reported by others and us,3,4 while the somatic mutation rate is relatively low (only 1–2 somatic mutations per Mb).6,7 Thus, global hypermethylation has been proposed to be a critical step during NPC tumorigenesis.1

The previous findings emphasize the importance of epigenetics in NPC pathogenesis. However, with the multifactorial nature of NPC pathogenesis, the EBV-associated epigenetic dysregulation, their functional impacts, and their interactions with genomic alterations remain unclear in NPC. Chromatin accessibility is a critical element of the regulatory epigenome, regulating and being regulated by epigenetic mechanisms, transcription factor (TF) binding, and viruses.8,9 Yet, the impact of EBV infection on chromatin accessibility has not been comprehensively investigated in NPC. Here, we report a multi-omics study of integrating whole-genome bisulfite sequencing (WGBS) and matched whole-exome sequencing (WES) in NPC clinical specimens, together with comprehensive profiling of EBV+ and EBV- NPC cells by the assay for
 transposase-accessible chromatin using sequencing (ATAC-Seq), existing single-cell RNA-Seq (scRNA-Seq) and EBVaGC methylation data, to systematically investigate the changes of DNA methylation, regulatory epigenome, and their associations with EBV infection and genomic alterations in NPC.

**Methods**

**NPC clinical specimens**

The fresh frozen NPC biopsies from 15 sporadic patients and 9 non-tumor adjacent tissues were provided by the Area of Excellence (AoE) NPC Tissue Bank for WGBS analysis. The tumor samples were randomly selected and examined by tissue sectioning and hematoxylin and eosin (H&E) staining to estimate the carcinoma content, as described previously. The WGBS results were validated in an independent patient cohort including 48 NPC patients. A total of 51 NPC patients were included in a tumor tissue microarray (TMA) obtained from the AoE NPC Tissue Bank. The age of the 51 NPC patients ranged from 25 to 101, with a mean of 50.75. The male to female ratio was 3.6. The summary of clinical information of the NPC patients is shown in Supplementary Table S1.

**Cell culture**

The normal immortalized squamous epithelial cell lines derived from nasopharynx included NP361-htert-cyclinD (NP361, RRID: CVCL_B3Q5), NP550-htert-cyclinD (NP550, RRID: CVCL_B3Q9), NP460-tert (NP460, RRID: CVCL_X205), and NP69 (RRID: CVCL_7084). The NPC cell lines C666-1, C17 and NPC43. The target regions included global CGIs from the UCSC database (hg19) and the complete EBV genome (NC_007605). The probes were designed using NimbleDesign (Roche Cat# 6350429001).

**Whole-genome bisulfite sequencing**

Library preparation of the tumor biopsies and non-tumor adjacent tissues was performed using Accel-NGS Methyl-Seq Kit (Swift Biosciences Cat# 30024). Each tumor biopsy had one library and three non-tumor adjacent tissues were pooled together for construction of one library. EpiTect Bisulfite Kit (Qiagen Cat# 59104) was used for bisulfite conversion. The 0.5% Unmethylated Lambda DNA (Promega Cat# D1321) was spiked in for assessing the bisulfite conversion efficiency. For the cell lines, WGBS was performed by Novogene.

**Targeted bisulfite sequencing**

The SeqCap Epi Enrichment System (Roche Cat# SEQ100146) was used for library preparation of target bisulfite sequencing in the normal NP cell lines NP550 and NP361, with and without Akata EBV infection, and the NPC cell lines C666-1, C17 and NPC43. The target regions include global CGIs from the UCSC database (hg19) and the complete EBV genome (NC_007605). The probes were designed using NimbleDesign (Roche Cat# 6350429001).

**ATAC-seq experimental procedure**

A total of 50,000 cells were lysed in 50μL cold lysis buffer (10 mM Tris-Cl pH7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP40) spinning at 500×g centrifugation for 10 min at 4 °C. Tn5 Transposase (Illumina Cat# E7645S) was added, and the transposition reaction took place at 37 °C for 30 min. A portion of samples was used to monitor the GC saturation of the signals and the size bias in PCR. The libraries were quantified by qPCR and evaluated by fragment analyzer before sequencing on the Illumina HiSeq X Ten platform.

**Immunohistochemistry (IHC) staining for CTCF in NPC**

The NPC tissue microarray (TMA) was constructed with the clinical specimens from 51 patients. Two biological replicates were performed for each patient. The median follow-up time was 77 months. An additional five paraffin-embedded tissue specimens containing morphologically normal nasopharyngeal epithelium were included as controls. IHC staining was performed as previously described. Briefly, the TMA slide was incubated with anti-CTCF antibody (Merck Millipore Cat# 07–729, 1:100 dilution) overnight. The standard streptavidin–biotin–peroxidase complex method was
used for detection. The CTCF protein staining was viewed by a trained pathologist (LT). A score between 1 and 2 was given to tumor cells as detectable (CTCF+) and undetectable (CTCF-) expression. The primary CTCF antibody has been validated by the vendor.

**qPCR for determining CTCF copy number**
The primers targeting the intronic region of CTCF at chr16:67,666,800-67,666,859 (FP: GCATGTTAATAACCCGTGCTT; RP: GCTGCGCACAATATGTA) and the control primers (FP: TTTCACAAAGAGGACATGAAAG, RP: TGAAATCTGCTTATGGGAGACA) were used to measure the relative genome copy numbers by Quantitative Polymerase Chain Reaction (qPCR), as described previously. A total of 50 ng genomic DNA was used and the experiments were performed in biological triplicate. The signals were normalized to the average level of normal samples. The copy number loss was used and the experiments were performed in biological triplicate. The signals were normalized to the average level of normal samples. The copy number loss of CTCF was determined by a cut-off ratio <0.8.

**qPCR for CD74 expression**
RNA for each cell lines was extracted by AllPrep DNA/RNA/miRNA Universal Kit (Qiagen Cat# 80284). A 2000 ng of RNA was used for reverse transcription. MMLV reverse transcriptase (Promega Cat# M1701) was used for making cDNA under 37 °C 1 h and 75 °C for 10 min. Two independent experiments with triplicates calculating the average delta–delta CT for CD74 expression through qPCR (CD74, FP: TGGCAACATGACAGAGGACC, RP: CTGCCCTCACTTGGGGACTG) (HPRT1 housekeeping gene, FP: TGACCTTGATTATTTGCCATACC, RP: CGAGCAAGACGTTCAGTCC) were performed.

**Western blotting**
The Western blotting experiments were performed as described. The primary antibody against CD74 Recombinant Rabbit Monoclonal Antibody (JF40-10) (Invitrogen, RRID: AB_2809806, 1:2000 dilution) and the secondary antibody rabbit IgG antibody (HRP) (GeneTex Cat # GTX213110-01, RRID: AB_10618573, 1:2000 dilution) and the secondary antibody rabbit IgG antibody (GeneTex Cat # GTX213110-01, RRID: AB_10618573, 1:2000 dilution) and the secondary antibody rabbit IgG antibody (HRP) (GeneTex Cat # GTX213110-01, RRID: AB_10618573, 1:2000 dilution) and the secondary antibody rabbit IgG antibody (GeneTex Cat # GTX213110-01, RRID: AB_10618573, 1:2000 dilution) were used to detect the expression of CD74. The antibody targeting CD74 has been validated by the vendor.

**Bisulfite pyrosequencing**
PyroMark Q96 CpG LINE-1 assay (Qiagen Cat# 973043) was used to quantify the methylation of LINE1 element. A set of methylation-specific primers (FP: TTGGAGTTGAGAAGTTAGTATGTT, RP: CTTAACTTCTAACCACAAATTACT) was used for validating the hypomethylated region exclusively found in HypoNPC. Bisulfite conversion of DNA was performed using EpiTect Bisulfite Kits (Qiagen Cat# 59104). The bisulfite pyrosequencing was performed in biological duplicate by the Centre for Panoromic Sciences (CPOS) in the University of Hong Kong. The NPC tissues were obtained from AoE NPC Tissue Bank.

**Identification of CTCF binding sites using CUT & RUN sequencing**
ChiC/CUT&RUN kit (Epicypher Cat# 15–1016) was used to profile the CTCF binding sites, with 500,000 cells. NEBNext® UltraTM II DNA Library Prep Kit for Illumina (Illumina Cat# E7645S) was used to make the sequencing libraries. Sequencing was performed using Illumina NovaSeq platform for 150bp paired-end reads. Sequencing data was evaluated by FastQC and the adaptor sequence was trimmed using Trimmomatic (v0.39), followed by Bowtie2 alignment and removal of PCR duplicates by GATK (v4.2.3.0), MACS2 software was used for the narrow peak calling with the matched IgG control with the parameter ‘p 0.05’.

**CUT & RUN-qPCR for detecting the binding signals of IRF1 at CD74 promoter**
ChiC/CUT&RUN kit (Epicypher Cat# 15–1016) and IRF1 Rabbit pAb (Abclonal Cat# A7692) were used to capture the IRF1 binding regions, followed by qPCR using primers targeting CD74 promoter region (FP: CTTAAAGTGTTGCTGGAGAG, RP: CAAAAAGCAGCTTACCAAAAG). The experiment was done in duplicates to quantify the IRF1 binding signals at the CD74 promoter region. The amount of input DNA was normalized for the duplicate qPCR experiment.

**Data analysis for WGBS and targeted bisulfite sequencing by SeqCap Epi enrichment system**
Trimmomatic was utilized to remove the adapter sequences, poor quality bases and reads. BSMAP was used for aligning the trimmed reads to the human genome (hg19) with EBV genome (NC_007605). The duplicated reads were discarded using Picard. BSMAP was used for CpG site methylation ratio quantification. FastQC and Picard were used to assess the quality of FASTQ and BAM files. Metilene was used to identify the differentially methylated regions (DMRs). Specifically, for each sample the CpG sites with less than 10x coverage were excluded from the analysis. Only the DMRs with Bonferroni corrected p-value < 0.05, an absolute difference of average methylation greater or equal to 20% between the groups and at least 10 CpG sites were included. For detection of DMRs in EBV, the same parameters were used except that the number of CpG sites required was lowered to five, since the EBV genome has fewer CpG sites compared to human genome. Unmethylated regions (UMRs) and lowly methylated regions (LMRs) in
the normal and NPC tissues were identified by MethylSeekR using default parameters with PMD filtering. DMVs were defined as UMRs longer than 5 kb. GREAT (v4.0.4) is a regional-based annotation tool. It was used for annotating the hypermethylated regions with the GO terms. The pathway enrichment analysis was performed with EnrichR R package. False discovery rate (FDR) was estimated using Benjamini-Hochberg correction. Only FDRs lower than 0.05 were considered as statistically significant.

**ATAC-seq data analysis for differential chromatin accessibility and footprint analysis**

Trimmomatic was used to remove the adapter sequences and poor-quality bases and reads. The ENCODE ATAC-Seq pipeline (v1.0) was applied for alignment, peak calling, and quality evaluation. The peaks for differential analysis were obtained by merging the narrowPeaks files of the same condition. The bioinformatics tool featureCounts was used for creating the count matrices. DESeq2 was used for identifying the differentially accessible regions. Only the regions with absolute log₂ fold change ≥ 2 and q-value < 0.01 were considered as significant. The function rlog was used for normalization. For the motif analysis, the bam files of the same condition were merged by samtools. The differential TF binding was identified by TOBAIS with the ATAC-Seq signal correction. The TF binding was calculated by TOBIAS BIndetect-motifs and based on the JASPAR database. The software clusterProfiler was used for the Gene Ontology analysis of the genes with differentially accessible regions. The FDR was estimated based on the Benjamini-Hochberg method. The top 200 categories are shown in a network by emapplot. ClueGO was used for visualizing the interaction of GO terms in the analysis of putative IRF1 and RELA targets.

**scRNA-seq data analysis**

The NPC scRNA-Seq datasets were downloaded from the CNGDb database with accession number CNP0000428 and from the GEO database with accession numbers GSE150825 and GSE162025, processed and aggregated by CellRanger. Scrublet was used on each sample independently to remove artificial doublets. Seurat was used for downstream analysis. Harmony was applied to remove the batch effect from different studies. MAST was used for detecting differentially expressed genes. Epithelial cells were identified using EPCAM and KRT19. T cells were identified using CD3D and CD3G. Myeloid cells were defined using LYZ, CD68, and LAMP3. B cells were defined using MS4A1, CD79A, and CD79B. Fibroblasts were identified using COL1A1 and COL3A1. InferCNV was used to profile CNV and identify NPC cells in the epithelial clusters with default parameters. CellPhoneDB (v.2.0) was used for inference of cell–cell interaction using the default parameters.

**Motif enrichment analysis**

HOMER (v4.9) was used for identifying the motifs enriched in different clusters of the hypermethylated regions. The background sequences were adjusted for the CpG content with the -cpg parameter.

**Annotation of genes and genomic regions**

The genes, promoters, exons, introns, and UTRs were defined using the GENCODE (v19) annotations with the pseudogenes removed. The promoters were defined as the ± 2 kb region around the transcription start sites. The gene bodies were defined for the region from the transcription start sites to the transcription end sites. The CGIs were downloaded from the UCSC genome browser. The FANTOM5 enhancers were downloaded from the official website. LINE1 regions were defined using RepeatMasker downloaded from USCS genome browser. ENCODE phase III transcription factor binding sites (TFBS) were downloaded from USCS genome browser. ChIP-Seq signal of CTCF and SMARCA4 were downloaded from GSE30263 and GSE74716 respectively.

**Analysis of Illumina 450K methylation array datasets**

The Gene Expression Omnibus (GEO) datasets GSE52068 and GSE62336 IDAT files were preprocessed using Functional Normalization in MiniFi. Thirty-seven normal nasopharyngeal tissues and 50 NPC tissues were included in the analysis. The probes that have a detection p-value > 0.05 in 5% of the samples, the sex chromosomes or reported to be cross-hybridizing were removed using DMRcate. The batch effect was adjusted with ComBat. The STAD methylation datasets were obtained from TCGA using the R package RTCGA in which the methylation ratios were normalized.

**Analysis of Affymetrix U133 plus 2.0 array datasets**

The GSE12452, GSE34573, and GSE64634 CEL files were processed using the rma function of oligo R Package. Totally, 17 normal nasopharyngeal tissues and 59 NPC tissues were included in the analysis. ComBat was used for the batch effect removal. Only the probes with a q-value lower than 0.01 were considered as significantly different. The probe IDs were converted to the gene symbols using the hgu133.plus.db R package. Limma was utilized to identify the differentially expressed genes.

**Analysis of bulk RNA-seq datasets**

The bulk RNA-Seq datasets GSE102349 and SRP158866 were downloaded from NCBI. STAR aligner (2.5.3a) was used for mapping the reads to the transcriptome for both
human (GENCODE, v27) and EBV (NC_007605). HTSeq (0.9.1) was used for counting the reads. SRP158866 contains the RNA-Seq data for NPC cell lines. The software L1EM was used for estimating the LINE1 expression with the default parameters and realignment. Gene expression correlation was computed using variance-stabilizing transformed normalized counts from DESeq2 R package. We used GSE102349 to examine the correlation between KLF4 expression and EBV gene expression. Only the correlations with absolute Pearson coefficient greater than 0.1 were considered. Data from GSE102349 was used to investigate the correlation of $CD74$ and immune inhibitory genes. The exhaustion scores of the NPC patients were estimated using a generalized linear model including ten marker genes (LAG3, GZMB, HAVCR2, PTMS, CXCL13, VCA1, PRF1, TNFRSF9, TIGIT and PDCD1) identified from the exhausted T cells. This model has been validated in three independent datasets previously.

Statistics
Two-tailed tests were used to determine the statistical significance throughout the study. Chi-square test, $t$-tests, Fisher’s exact tests, determination of Pearson correlation coefficients, were performed using R (v3.6.3) and GraphPad Prism. Standard error of mean was used to estimate the variation of qPCR biological replicates. The odds ratios (OR) between the DMRs and genomic features were calculated as previously described. Mathematically,

$$\text{Odds Ratios} = \frac{\frac{\text{Number of DMR CpG sites overlaps with the genome feature}}{\text{Number of DMR CpG sites do not overlap with the genome feature}}}{\frac{\text{Number of non-DMR CpG sites overlap with the genome feature}}{\text{Number of non-DMR CpG sites do not overlap with the genome feature}}}$$

The ORs between DMRs and TFBS were computed using the same formula, except for using lengths of regions instead of number of CpG sites. 95% confidence intervals (CI) were computed as previously described.

Ethics
Informed consent for clinical sample collection was obtained from the study participants according to the protocol approved by the Institutional Review Board (Reference No.: UW12-239) of the University of Hong Kong.

Role of funders
The funders were not involved in study design, data collection, data analyses, interpretation, or writing of the article.

Results
Identification of two distinct epigenomic subtypes of EBV-positive NPC by WGBS
WGBS was performed on 15 EBV + NPC primary tissues and 9 non-tumor adjacent tissues to comprehensively profile the NPC methylome (Supplementary Table S1). The WGBS dataset had an average coverage of 43-fold, with more than 95% of annotated CpG sites sequenced by at least 10-fold coverage (Supplementary Table S2). The technical duplicates showed concordant results (Pearson’s $r^2 = 0.91$ and Supplementary Fig. S1a), and the data were confirmed by an independent method using the Illumina Infinium Human-Methylation 450K BeadChip (Pearson’s $r^2 = 0.90$–0.91; Supplementary Fig. S1b).

The unsupervised clustering revealed the inter-patient DNA methylation heterogeneity in NPC, identifying two distinct methylation subtypes (Fig. 1a and Supplementary Fig. S2). Specifically, 80% (12/15) of NPC (HyperNPC) showed global hypermethylation and 20% (3/15) of NPC (HypoNPC) demonstrated global hypomethylation (Fig. 1a and b). The long-interspersed element-1 (LINE1) sequences are dispersed across the genome and are regarded as a surrogate marker of global DNA methylation level. Lower LINE1 methylation level extracted by WGBS in HypoNPC thus confirmed the global hypomethylation status in this group (Fig. 1b; Supplementary Fig. S3). We performed bisulfite pyrosequencing targeting the LINE1 elements in an independent cohort. Similarly, 6 out of 33 (18%) NPC tissues were HypoNPC, showing lower methylation levels compared with the normal tissues (Supplementary Table S1; Supplementary Fig. S4).

We further examined the methylome in two normal immortalized nasopharyngeal cell lines (NP69, NP460) as well as three EBV + NPC cell lines (C666-1, C17, and NPC43) and three EBV- NPC cell lines (NPC53, NPC38, and HK1) by WGBS. Noticeably, the EBV + NPC cell line NPC43 had lower global methylation similar to HypoNPC group, while C666-1 and C17 resemble the globally hypermethylated NPC (Fig. 1c; Supplementary Fig. S5). Additionally, the EBV methylome was compared between the two groups, detecting 30 hypomethylated regions in the HypoNPC (Fig. 1d; Supplementary Table S3). NPC43 also demonstrated a similar EBV methylation pattern as the HypoNPC. Altogether, these results raise the possibility that NPC43
Fig. 1: The global methylation and chromatin landscape of NPC. (a) The average methylation of primary NPC tissues and normal tissues of the whole genome. The genome was unbiasedly segmented into consecutive tiles of 1 Mbps. (b) The global methylation levels, methylation ratio at base-pair resolution, and methylation mean in CGIs and LINE1 regions. (Welch’s t-test, ns *P < 0.05, **P < 0.01, ***P < 0.001) (c) The average methylation in the 1 Mbps tiles across the NP and NPC cell lines. The dashed line denotes the median of methylation of the tiles in the NP cell lines. (d) The methylation levels of the differentially methylated regions between the EBV genome of HyperNPC and HypoNPC. (e) Mutational patterns in HyperNPC and HypoNPC. The single nucleotide substitutions are divided into six categories (different colours) with 16 possible flanking changes surrounding the mutated base. The vertical axis displays the percentage of mutations attributed to a specific mutation type. The trinucleotide contexts of the mutated bases are shown on the x-axis. (f) C to T mutations in ACG and TCA trinucleotide contexts are significantly different between the two groups. (Welch’s t-test, *P < 0.05) (g) Contribution of the APOBEC mutational signature between two groups. Only the primary tumours are included in the analysis (Welch’s t-test, **P < 0.01).
is derived from an NPC tumor with global hypomethylation, as observed in the HypoNPC. Previously, the NPC genomic landscape studies identified the mutational signatures relevant to aging, defective DNA mismatch repair (MMR) and APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of cytidine deaminases and homologous recombination deficiency in NPC, as reported by others and us.\textsuperscript{52,53} We compared the mutational load and profiles by WES between HyperNPC and HypoNPC (Fig. 1e; Supplementary Fig. S6) and deconvoluted the mutational signatures against 30 known COSMIC mutational signatures (version 2). Noticeably, the HypoNPC has lower C > T mutations at the ACpG trinucleotides and higher C > T mutations at the TCA trinucleotides (Fig. 1f), with a distinct APOBEC-related signature (COSMIC signature 2) compared to the HyperNPC (Fig. 1g).

**Enrichment of chromatin master regulator CTCF binding sites in EBV-specific methylated regions**

We used Metilene\textsuperscript{21} to identify 9609 differentially methylated regions (NPC-DMRs) between normal tissues and all NPC including both HypoNPC and HyperNPC in the WGBS cohort (Fig. 2a; Supplementary Fig. S7a-b; Supplementary Table S4). These NPC-DMRs, enriched in the regulatory regions with reduced chromatin accessibility were able to distinguish 89% of NPC tissues from the normal samples (Supplementary Figs. S7c-e and S8) using the Illumina 450K BeadChip including 50 normal and 37 NPC tissues obtained from the previous studies from others and us.\textsuperscript{45} The NPC-DMRs were also confirmed by the SeqCap Epi dataset in the EBV + NPC cell lines C666-1, C17, and NPC43 cells (Fig. 2b; Supplementary Fig. S1c; Supplementary Table S2). A total of 1023 hypomethylated regions (Hypo-DMRs) were identified by comparing HypoNPC and the normal tissues (Fig. 2a). To verify the hypomethylated regions identified in the HypoNPC, we examined the methylation level of a locus specifically hypomethylated in HypoNPC by bisulfite pyrosequencing in 15 additional NPC biopsies. Similarly, we identified 26.7% of NPC cases with reduced methylation at this selected locus (Supplementary Fig. S9). HyperNPC was also compared with normal tissues, and the comparison of genomic features between HyperNPC and HypoNPC was shown in Supplementary Fig. S10. HypoNPC contained more hypomethylated regions (HypoNPC: 47% vs HyperNPC <1%), primarily located in intronic, intergenic, and gene body regions.

It was reported that \textit{in vitro} EBV infection of the non-neoplastic gastric epithelial cell line (GES1) and the natural EBV-gastric cancer cell line (MKN7) was associated with the hypermethylation in regions exclusively in the EBV + gastric cancer tissues.\textsuperscript{22,23} It is surprising that no dramatic increase in methylation was observed in the EBV-infected NP550 and NP361 cells (Fig. 2b), indicating that EBV infection alone cannot dramatically shape the methylation changes in these cells and the specific cellular context may be required to modulate methylation changes, which are absent in these non-malignant cells. To identify the potential key cellular factors likely involved in the methylation changes associated with EBV, we further examined the EBV+/EBV- stomach adenocarcinoma (STAD)\textsuperscript{4} and EBV+/EBV- NPC cell lines. We used the NPC hypermethylated regions identified by WGBS to perform unsupervised clustering of STAD cases, identifying four sets of NPC hypermethylated regions (Rows A-D) and three major clusters of STAD cases (Clusters 1–3) (Fig. 2c; Supplementary Table S5). Cluster 1 showed extensive hypermethylation across Rows A-D and higher overall methylation ratios compared with other clusters, and it included 100% of EBV cases and 71% of microsatellite instability (MSI) cases (Supplementary Fig. S11a-b). Also, the EBV + NPC cell lines showed substantially higher methylation in Row A, but not in Rows B-D, compared to the EBV- NPC cell lines (Supplementary Fig. S11c). Both datasets suggest that Row A was highly specific to EBV (EBV-DMRs). Correspondingly, chromatin accessibility in the EBV-DMRs was reduced in the EBV + NPC cell lines compared to the EBV- NPC cell lines (Supplementary Fig. S12).

We examined TF bindings by the enrichment analysis using the ENCODE ChIP-Seq data and motif analysis in Rows A-D, respectively. Row A was significantly enriched for the binding sites of multiple dysregulated chromatin modulators, including the key cohesion complex players, CTCF, RAD21, and SMC3, histone lysine demethylase PHF8, histone deacetylase HDAC2, SWI/SNF related chromatin regulator SMARCA4 and chromatin reader protein BRD4 (Fig. 2d; Supplementary Fig. S13a). This indicated the importance of cellular chromatin modulators in the methylation changes associated with EBV. The CTCF binding signal and CTCF motif occurrence were notably higher in Row A than in other rows (Fig. 2e; Supplementary Fig. S13b-c). The CUT&RUN data showed that CTCF binding signal was exclusively high in Row A in NP69 but abolished in C666-1, which had the hemizygous loss of CTCF (Supplementary Fig. S13d). These consistent results support that loss of CTCF binding frequently occurred at the EBV-DMRs in NPC. CTCF expression was negatively correlated with expression of EBV latent genes RPMS1 and EBER2, suggesting a potential link between CTCF loss and EBV latent gene expression in NPC (Supplementary Fig. S14). In addition, the binding signals of SMARCA4 were also highly associated with the EBV-DMR. SMARCA4 was found to be predominantly overexpressed in the malignant cells by the scRNA-Seq data\textsuperscript{29} (Fig. 2f; Supplementary Fig. S15).
Fig. 2: The characteristics of DMRs in NPC. DMRs were identified by comparing normal tissues with NPC tissues. (a) The methylation level of the clinical specimens in the hypermethylated regions identified in the comparison between all NPC and normal tissues (NPC-DMRs) and the hypomethylated regions identified by comparing HypoNPC to normal tissues (HypoNPC-DMRs). (b) The methylation means of the EBV ± immortalized NP, NPC cell lines, and methylation controls in the NPC-DMRs. (c) Unsupervised clustering of the NPC-DMRs in the TCGA stomach adenocarcinoma (STAD) cohort. Five subtypes of STAD were defined according to TCGA previous analysis, including EBV (Epstein–Barr virus: n = 29), MSI (microsatellite instability: n = 59), CIN (Copy number instability: n = 202), GS (Genomically stable: n = 46), POLE (DNA polymerase epsilon: n = 7). The columns and rows were cut into three and four clusters respectively, with different enrichment of STAD subtypes and methylation levels. Cluster 1 showed extensive hypermethylation across Rows A–D and higher overall methylation ratios compared with other clusters, mainly including EBV and MSI cases. Row A was identified as the EBV-specific differentially methylated regions (EBV-DMRs). (d) The percentage of hypermethylation regions in Rows A–D overlaps with the ENCODE TFBS. The red bars indicate the transcription factors with more overlaps with Row A. The blue bars denote TFs with more overlaps with Rows B–D. (e) CTCF binding signals and (f) SMARCA4 binding signals in Rows A–D. The Row A (EBV-DMRs) was highly enriched at CTCF and SMARCA4 binding sites. The NHEK CTCF and K562 SMARCA4 ChIP-Seq data were downloaded from ENCODE.
Hemizygous loss and reduced protein expression of CTCF frequently detected in NPC

CTCF is a master regulator of chromatin structure, located on chromosome 16q22.1. At the genomic level, CTCF mutations are extremely rare. Out of 216 NPC patients analyzed previously, we only detected one missense mutation (NM_006565, p.E478K), whereas hemizygous loss of CTCF was detected in 41.3% of NPC patients (Supplementary Fig. S16). The WGS study suggests that copy number loss is often found at 16q. Reduction of CTCF protein level was observed in 45% (23/51) of the NPC patients and was more obvious in stage IV than in other stages (Chi-square test, 62% vs 38%, Fig. 3a). We examined the copy number variations (CNVs) of three NPC patient-derived xenografts (PDXs) by WES previously. The WES revealed that hemizygous loss of the region encompassing CTCF was found in two NPC PDXs (Fig. 3a). The CNVs at CTCF were also detected in two EBV + NPC cell lines, C666-1 and C17, three EBV-NPC cell lines, NPC38, NPC53, and HK1, and three out of six (50%) NPC clinical specimens by qPCR (Fig. 3b and c). Hemizygous deletion at CTCF was detected in 7 out of 11 NPC patients in a scRNA-Seq cohort (Welch’s t-test, P < 4.6e-6, Fig. 3d and e) and had a significant impact on its expression, which was also confirmed by the TCGA STAD data (Fig. 3f). We then examined the CTCF protein level in a cohort of 51 NPC specimens by immunohistochemical (IHC) staining using a NPC tissue microarray (TMA) (Supplementary Table S1). Reduction of CTCF protein level was observed in 45% (23/51) of the NPC patients and was more obvious in stage IV than in other stages (Chi-square test, 62% vs 38%, P = 0.034, Fig. 3g–i).

Landscape analysis for chromatin accessibility reveals reprogrammed regulatory epigenome in EBV + NPC

The significant enrichment of the binding sites of chromatin modulators at EBV-DMRs strongly suggested that the chromatin changes were involved in EBV + NPC. To further examine the EBV-associated epigenetic changes at the chromatin level, the genome-wide chromatin accessibility profiles of the EBV + NPC cell lines and normal NP cell lines were evaluated by ATAC-Seq. Differentially accessible regions (DARs) were identified (Fig. 4a). NPC cells showed a more compact chromatin structure than the normal cells, with 4199 open and 10,743 closed DARs identified in the NPC cells, compared to the normal cells (Tumor specific). We also identified 7201 open and 6757 closed DARs in the EBV + NPC cells compared to the EBV-NPC cells (EBV specific). NPC43 exhibited a more distinct chromatin accessibility status, which deviated from the other two EBV + NPC cell lines (C666-1 and C17). At the global level, the chromatin accessibility of NPC43 resembled the EBV- cell lines (Supplementary Fig. S17). Specifically, NPC43 showed increased chromatin accessibility in the EBV-specific closed regions and decreased chromatin accessibility in the EBV-specific open regions. This result is consistent with the findings from the WGS analysis that there are two epigenomic subtypes in the EBV + NPC (Fig. 1c). The CTCF binding sites were completely lost at these closed chromatin regions in both EBV-specific regions and NPC-associated regions, while the gain of CTCF binding signals was subtle in the open regions in C666-1 compared to the normal cell line NP69 (Fig. 4b). These suggested CTCF binding sites were altered in EBV + NPC.

There is increased chromatin accessibility in the enhancers, CGIs, and promoters in the EBV + NPC cell lines, especially in C666-1 and C17 (Fig. 4c), suggesting that regulation of DARs was associated with EBV and these changes have important regulatory functions. To examine the TFs involved in the regions with chromatin changes, we performed the differential TF binding analyses and compared the binding activities of the TFs between the NPC and normal cell lines, as well as between the EBV+ and EBV- NPC cell lines (Fig. 4d). We found that although the binding of the Krüppel-like factor (KLF) and specificity protein (SP) family was generally increased when comparing all the NPC cell lines to the normal cell lines, reduction of binding activities of the KLF family members was specifically observed in the EBV + NPC compared with EBV- NPC. The KLF/SP family, for example, KLF4, plays critical roles in stem cell maintenance, cell proliferation, tissue differentiation, and metabolism, and it is important for maintaining the EBV life cycle in epithelial cells. KLF4 upregulation was shown to activate the EBV lytic phase in epithelial cells. We detected that KLF4 expression was inversely correlated with the expression of EBV latent genes in NPC (Supplementary Fig. S18), so suppression of the KLF/SP family binding by EBV is likely one of the important mechanisms involved in the regulation of the EBV latency in NPC. The enhanced binding of the TFs, such as NFκB2 (p52), RELA (p65), RELB, and REL, from the NF-κB signaling pathways was identified in DARs associated with EBV (Fig. 4d; Supplementary Table S6). This is consistent with the previous findings that constitutive activation of NF-κB pathway is important for tumorigenesis in EBV + NPC. Interestingly, enrichment of the binding sites of the type I interferon regulatory transcription factor (IRF) family was particularly identified when comparing the EBV + NPC to the EBV- NPC cell lines (Fig. 4d). This family is critical in regulating the immune response of cancer cells and immune cell development. The GO analysis further revealed that the EBV-specific regions with differential accessibility were enriched in immune response, cell differentiation and development, metabolism, ERBB pathway, and calcium signaling transduction (Supplementary Fig. S19). These results demonstrate that remodeling of transcriptional activities associated with EBV has an important impact on the biological functions, which are critical in
maintaining virus survival and activating cancer-relevant pathways contributing to NPC tumorigenesis.

In addition, the analysis showed that the closed regions in NPC were often the binding targets of FOS/JUN/ATF (AP-1 components), SMAD2/SMAD3, CEBP family, and ETS-related TFs, while the open chromatin areas were associated with increased binding of the TFs with the forkhead-box (FOX) and SRY-related high mobility group box (SOX) of DNA-binding domains (Fig. 4d). The GO analysis found that the genes associated with chromatin accessibility changes in NPC were primarily involved in cell differentiation and development, immune response, and peptidyl-tyrosine phosphorylation (Supplementary Fig. S19).

**EBV-associated chromatin accessibility at the immune-related genes and shapes cellular interactions in the NPC tumor microenvironment**

The enhanced motif binding in the NF-κB-related TFs was detected in NPC associated with EBV. The NF-κB pathway regulates cellular differentiation, survival, and proliferation and is important for both innate and adaptive immune responses. In our analysis, the increased binding signals were identified in both canonical and non-canonical NF-κB pathway members (Supplementary Fig. S20). Our analysis revealed a number of potential NF-κB targets, of which CD74 is one of the candidates with significant expression changes. Increased chromatin accessibility in the promoter of CD74 was found in the EBV + NPC cell lines C666-1 and C17 but not in NPC43 and EBV- NPC (Fig. 5a). We identified a list of potential CD74 upstream TFs from the differential motif binding analysis (Fig. 5b). The enhanced binding of IRF1 at the promoter of CD74 was confirmed by the CUT-RUN-qPCR (Fig. 5c; Supplementary Fig. S21a). The elevation of CD74 expression was consistently found at the mRNA and protein levels in C666-1 and C17 (Fig. 5d). A strong positive correlation between IRF1 and CD74 at the mRNA level was detected in NPC (Pearson correlation coefficient $r = 0.672$, Supplementary Fig. S21b). These indicate IRF1 plays an important role in regulating CD74 expression in EBV + NPC. BCL3 is a predominant component of NF-κB complex involved in the constitutive activation of NF-κB pathway in NPC, and CD74 encodes for the invariant chain of major histocompatibility complex class II (MHC-II). The expression of BCL3 and HLA-DRA (a member of MHC-II) was examined in NPC cells using the scRNA-Seq data. The result showed that CD74 co-expressed with HLA-DRA but not BCL3 (Supplementary Fig. S21c).

The scRNA-Seq cohorts from three published studies further showed that a higher level of CD74 expression was found in the NPC cells with increased expression of the EBV latent gene, LMP-1. CD74 specifically mediated cellular interactions between the EBV + NPC cells and multiple cell types in the TME, with CD74-MIF/COPA as the predominant interaction (Fig. 6a). There is evidence that CD74-MIF is linked to cancer development and is regarded as a potential drug target. We further integrated the bulk RNA-Seq cohort with 88 NPC patients to investigate the role of CD74 in the TME. The CD74 expression strongly correlated with multiple inhibitory genes typically expressed on T cells, such as HAVCR2, TIGIT, CTLA4, and LAG3 (Fig. 6b). In a previous single-cell study in NPC, a linear model incorporating ten genes identified from the exhausted T cells was established to evaluate the T cell functions. We applied this model in the bulk RNA-Seq data and found that CD74 expression showed a strong correlation with T cell exhaustion (Fig. 6c). The scRNA-Seq data further indicated that the patients with higher CD74 expression had a dramatic increase of CD8+LAG3+ exhausted T cells (Welch’s t-test, $P < 0.01$, Fig. 6d).

**Discussion**

The current model of NPC tumorigenesis suggests the inactivation of tumor suppressors, for example, via downregulation of CDKN2A and RASSF1A, occurs in the precancerous lesion and allows latent EBV infection that triggers genome-wide epigenetic changes including global hypermethylation that presumably contribute to NPC tumorigenesis. Certain EBV subtypes, so-called high-risk strains, carry distinctive single nucleotide polymorphisms (SNPs) in the EBV genes and are strongly associated with a higher risk of NPC. These reports highlight that it is pivotal to understand the role of EBV in NPC pathogenesis, especially how EBV involves epigenetic dysregulation, how the EBV latency is maintained, and their biological impacts during NPC tumorigenesis. In our study, about 80% of the NPC patients were categorized as HyperNPC. DNA methylation is mediated through DNA methyltransferases (DNMTs). A significant upregulation of DNMT1 and DNMT3A was found in NPC (Supplementary Fig. S22), which is potentially relevant to global hypermethylation. It is reported that DNMT1 and DNMT3A were shown to be activated or upregulated by EBV latent gene LMP1. However, the normal immortalized NP cell lines infected with EBV did not show global hypermethylation change regardless of LMP1 expression and upregulation of DNMT3A (Supplementary Fig. S22), suggesting upregulation of DNMTs by EBV in these cells maybe not sufficient to induce global hypermethylation. There are two DNA demethylation pathways: one is through TET-TDG-base excision repair (BER) pathway and the other one is through APOBEC/AID-MBD4 or TDG-dependent BER. In NPC, TET-TDG-BER could be suppressed by multiple
Fig. 3: Dysregulation of CTCF at genomic, mRNA, and protein levels in NPC. (a) The CNV alterations of CTCF for the patient-derived xenografts (PDXs), Xeno23, Xeno47, Xeno32, and NPC cell line NPC43, revealed by WES (SRP158745). Hemizygous deletion (in red) was detected on chromosome 16q22.1 containing CTCF in two of three PDXs. (b) CTCF CNV in cell lines measured by qPCR in duplicate. The bars denote the average normalized signal of the replicates. The error bars denote the standard error of the mean (SEM). NP69 was used as the reference for signal normalization. The samples with an average signal below 0.8 were considered harboring CTCF hemizygous loss. (c) CTCF CNV in NPC patients measured by qPCR in triplicate. T1 to T6 denote six patients, respectively. B1 denotes the matched blood from a NPC patient. The signals were normalized against the blood from healthy individuals (HB1 and HB2). The bars represent the average normalized signal of the replicates. The error bars denote the standard error of the mean (SEM). The samples with an average signal below 0.8 were considered having CTCF hemizygous loss. (d) The genome-wide CNV and expression of CTCF of two NPC patients in the scRNA-Seq cohort. The violin plots visualize CTCF expression in the normal epithelial and tumor cells. Significant downregulation of CTCF was found in the NPC cells with copy number loss in the region (16q22.1) containing CTCF. (Welch’s t-test, ****P < 0.0001). (e) The expression of CTCF in NPC patients with or without CTCF copy number loss detected by scRNA-Seq (Welch’s t-test, ns P > 0.05, ****P < 0.0001). (f) The CNV and expression of CTCF in
mechanisms in NPC, for example, TET1 gene silencing by promoter methylation and downregulation of TDG. It has been shown that reduction of TET activities promoted hypermethylation in cancer and the reduction of TET activities could synergistically induce hypermethylation in vitro with elevated DNMT activities in the mouse gastritis model. We proposed that suppressed DNA demethylation together with elevated DNA methyltransferase may contribute to HyperNPC, when the chromatin is vulnerable for methylation changes due to loss of CTCF bindings. The hypothetic model for the proposed mechanisms is illustrated in Supplementary Fig. S23a.

Surprisingly, our WGBS analysis identified a subgroup of EBV + NPC with global hypomethylation, accounting for 20% of the total patients, with in vitro EBV + NPC cell line models likely representing these two distinct subtypes. Integration of WGBS and WES data suggests different underlying mechanisms for pathogenesis in two EBV + NPC epigenomic subtypes, as summarized in Supplementary Fig. S23a. HypoNPC has an increased APOBEC-associated mutational signature. Upregulation of APOBEC family members in a subset of NPC has been reported by us previously and there is functional evidence that APOBEC3A may efficiently demethylate methylcytosine. The footprint of the APOBEC mutational signature is likely associated with the active APOBEC/AID-mediated demethylation process in HypoNPC. In addition to MMR, repair of 5 mC deamination damage also involves major glycosylases MB4 or TDG-dependent base excision repair (BER). It has been shown that MB4-deficient leukemia demonstrated higher C to T mutations in the context of AcG than CpG, C CpG, and G CpG. Based on this evidence, the reduced mutations in the AcG context in the HypoNPC are likely associated with elevated activities of MB4-dependent BER at these sites in this group. These results suggest the involvement of host DNA repair machinery and APOBEC deaminases in NPC demethylation in HypoNPC. When this global hypomethylation arises during NPC development and its impacts on NPC clinical outcomes will be critical research questions for further investigation in NPC. Moreover, the detailed biological mechanisms underlying two epigenomic molecular subtypes in EBV + NPC remain inconclusive, although there is evidence for the involvement of DNMTs, TETs, and APOBECs for methylation changes in other cancer types or biochemical studies. The mechanistic experiments are warranted to elucidate the functional roles of these pathways in the changes of global methylation pattern in EBV + NPC.

Apart from the global methylation changes, the EBV-associated differentially methylated regions (i.e., EBV-DMRs) were identified from both epigenomic subtypes. We provided evidence to support the involvement of loss of CTCF binding in EBV-associated epigenome reprogramming. CTCF regulates higher-order chromatin organization and accumulating evidence shows that it is the driver of oncogenic epigenetic changes in cancer. Our study detected frequent loss of copy number and protein of CTCF in NPC and reduced CTCF binding at EBV-associated DNA methylation and closed chromatin regions. It is reported that hemizygous loss of CTCF destabilizes DNA methylation in vivo and predisposes to cancer in a wide range of tissues. In mouse, Ctcf hemizygous loss downregulated Ctcf protein level and primarily reduced genome-wide Ctcf binding. Moreover, CTCF copy number loss primarily resulted in DNA hypermethylation in prostate cancer, and loss of CTCF binding induced cancer-specific hypermethylation at the binding sites. Importantly, CTCF deletion syndrome, characterized by microdeletion and haploinsufficiency of CTCF, was associated hypermethylation at the CTCF binding sites. Collectively, we speculate that loss of CTCF at genetic and protein levels is one of the mechanisms underlying the reduced expression of CTCF in NPC, and subsequently tunes down its DNA binding, which may induce instability of methylation at these sites, and thus, these sites are susceptible to methylation changes. EBV latent infection likely further enhance the methylation at these selective sites and subsequently inhibit the methylation sensitive CTCF binding events at certain selective regions. So loss of CTCF binding observed in EBV + NPC maybe a joint effect of loss of CTCF expression or activity at the cellular level together with prevention of CTCF binding at the methylation sensitive sites associated with EBV infection. Undetectable CTCF at the protein level is more frequently seen in the stage IV patients, indicating that with disease progression, NPC patients have more dramatic loss of CTCF that is likely associated with more intensive hypermethylation at the relevant regions.

One of the limitations of our study is lack of detailed mechanistic investigation. The causal link among the loss of CTCF, EBV infection and DNA methylation is not elucidated. It also remains unclear how the reduced CTCF protein expression is associated with advanced...
Fig. 4: The regulatory role and the relevant transcription factors of the differentially accessible regions. (a) The tumor-specific open and closed regions and the EBV-specific open and closed regions identified by differential peak analysis. Tumor-specific denotes the comparison of NPC cell lines and normal cell lines. EBV specific refers to the comparison of EBV + NPC cell lines with EBV- NPC cell lines (b) The CTCF binding signals in NP69 and C666-1 in the differentially accessible regions. (c) The chromatin accessibility at the promoters, CpG islands, and enhancers in the EBV + NPC compared to the EBV- NPC. (Student’s t-test, ****P < 0.0001). (d) The motif binding analysis at the DARs between analysis at the DARs between tumor versus normal cells and EBV + versus EBV- NPC. The top 5% of the TFs associated with open or closed regions were highlighted in the volcano plots with -log10 (p-value) above the 95% quantile. The expression and promoter methylation of TFs were examined by DESeq2 and Metilene.
disease. The negative correlation between CTCF expression and the EBV latent gene expression in NPC clinical specimens suggests the potential link among CTCF, EBV and advanced disease. Comprehensive functional experiments are needed to further disentangle and the causal factors in these associations observed in this study.

Reprogrammed binding of differentiation- and immune-relevant transcription factors and dysregulated epigenetics were observed in EBV + NPC by chromatin accessibility profiling. We observed that the EBV-specific DMRs and differentially closed regions are strongly associated with the development and differentiation pathways, contributing to the histologically undifferentiated phenotype in the EBV + NPC and EBVaGC, which is believed to be required for the establishment of EBV latency.\textsuperscript{55,56,76} The target genes of KLF/SP members are often observed in the EBV-specific closed regions. Additionally, reduced binding of the AP-1 components was found in the NPC-specific regions associated with EBV. The AP-1 members and KLF/SP members were reported to promote reactivation of the EBV lytic cycle.\textsuperscript{55,56,77–79} This is consistent with the binding of these TFs was suppressed to maintain EBV latency in NPC.

Integration of ATAC-Seq, RNA-Seq, and scRNA-Seq datasets identified multiple EBV specific cell–cell communications, particularly the MIF cytokine-CD74 receptor interaction (Supplementary Fig. S23b), which was shown to stimulate epithelial cell growth and release cytokines to regulate stromal cells in the TME.\textsuperscript{58} Though CD74 was widely investigated in immune cells,

---

**Fig. 5: EBV regulates NF-κB signaling in NPC promoting activation of the immune-related gene CD74 associated with T cell exhaustion.**

(a) The ATAC-Seq signal tracks of CD74 visualized using IGV (black track: normal immortalized cell lines, red track: EBV + NPC, green track: EBV-NPC). The DAR is highlighted near the promoter region. Increased chromatin accessibility at CD74 promoter region was observed in C666-1 and C17. (b) The merged signal tracks and footprints of TFs at the CD74 promoter. The ATAC-Seq signals were adjusted by the expected Tn5 binding frequency, and the footprint scores were measured by the difference of the cutting site signal subtracting the signals in the middle of motif binding regions. The numbers in the brackets denote the upper and lower score limits for the signal tracks. Two comparisons between tumor versus normal cell lines and EBV + versus EBV- NPC cell lines were performed. The possible TF binding sites are selectively listed below the signal tracks. CD74 is the target of the canonical NF-κB members RELA and REL, IRF1, and KLF/SP family. (c) Differential binding score and the motif clustering for two sets of comparisons between tumor versus normal cell lines and EBV + versus EBV-NPC cell lines. The TF binding sites with similar sequences were clustered together in the same color. Increased binding of the canonical NF-κB members RELA and REL, IRF1, and reduced motif binding by KLF/SP family were identified in EBV + NPC compared to EBV- NPC. (d) The mRNA and protein expression of CD74 were upregulated in C666-1 and C17 measured by RT-qPCR in triplicate and Western blot, respectively. The relative mRNA expression level was calculated and normalized to the housekeeping gene and the normal immortalized nasopharyngeal cell line (NP69).
The EBV-specific cell-cell communications in NPC TME. (a) The scRNA-Seq profiles were obtained from three published studies. The cellular interactions between cancer cells and TME were evaluated. EBV-high and -low clusters were identified according to the LMP-1 expression level. The cellular interaction between the receptor–ligand pairs was estimated across different cell types. Only the cellular interactions associated with the genes upregulated at the mRNA level with increased chromatin accessibility in the EBV + NPC cell lines are illustrated in the non-malignant, EBV-high, and EBV-low NPC clusters in the graph. The result showed that EBV modulated TME through specific cell-cell communications. (b) The gene expression correlation of CD74, MIF, and immune inhibitory genes in the bulk RNA-Seq cohort of 88 NPC patients. Gene expression of the genes was estimated by variance-stabilizing transformed normalized counts which account for sequencing depth across libraries. (c) The correlation of CD74 expression and T cell exhaustion in 88 NPC patients. The exhaustion score was estimated using the bulk RNA-Seq data by a linear model including ten marker genes identified from the exhausted T cells from a single cell study in NPC (GSE102349). (d) The proportion of CD8+LAG3+ T cells in CD74-high and CD74-low patients from the published scRNA-Seq cohorts (Welch’s t-test, **P < 0.01).

Fig. 6: The EBV-specific cell-cell communications in NPC TME.
its immunological role in viral-associated epithelial cancer cells is largely inconclusive and likely to be cell-type and context-dependent. This study identified IRF1 as an important TF regulating CD74. IRF1 was shown to open the chromatin of interferon-stimulated genes and was reported to induce PD-L1 to promote immune evasion. Therefore, our discovery of enriched IRF1 and NF-kB binding sites in the EBV-open region on CD74 suggests that EBV latent infection are associated with upregulation of CD74 and IRF1 may have a role modulating expression of this gene. A strong correlation between CD74 and T cell exhaustion was observed in our study and our previous NPC single-cell RNA sequencing data suggested an immune suppressive TME in NPC. However, the detailed functional impact of CD74 upregulation in immune evasion in the EBV + NPC cells remain unclear. CD74 encodes for the invariant chain of major histocompatibility complex class II (MHC-II) and is co-expressed with MHC-II HLA-DRA in our study. Co-culturing tumor-infiltrating lymphocytes with C17 cells showed that upregulation of MHC-II was associated with higher tumorigenicity and T cell exhaustion in NPC. This functional study provides some indirect evidence to support the potential role of CD74 associated with immune evasion. Since progressive T cell exhaustion is frequently detected presumably due to chronic EBV infection, our studies also highlight that EBV-specific communications could be the therapeutic targets for reinvigorating exhausted T cells in EBV-associated malignancies.

In summary, our work reveals the unexpected heterogeneity of DNA methylation in EBV + NPC that provides novel insights into NPC pathogenesis and highlights importance of dysregulation of host factor CTCF in EBV-associated epigenetic changes. Our study shows the reprogramming of regulatory epigenome in EBV + NPC, which involves in controlling cellular differentiation and proliferation, and likely having the impact on the immune cell functions in the TME. The cell communications are highly specific between the EBV-positive NPC cells and TME may have the potential to be a therapeutic target or a biomarker in EBV-associated malignancies.

**Contributors**

L.K.Y.C., D.L.S.C., W.D., and M.L.L. wrote the manuscript. L.K.Y.C., D.L.S.C., and W.D. performed the bioinformatics analysis and performed the experiments. D.L.S.C., J.L., and H.C. performed the ATAC-seq experiments. K.F.C constructed the TMA and L.T. performed the ATAC-seq experiments. L.K.Y.C., D.L.S.C., W.D., and M.L.L. directed, designed, and supervised the study. L.K.Y.C., D.L.S.C., W.D., and M.L.L. verified the data.

**Data sharing statement**

The WGBS and SeqCap Epi data generated in this study have been deposited at the Gene Expression Omnibus (GEO) database with accession number GSE173551 and GSE173559, respectively.

**Declaration of interests**

R.K.C.N. is an Advisory Board member or a Data Safety Monitoring Board member of Nuance (ShenZhen, China), received consultation fees from Pfizer, Novartis, Sanofi, AstraZeneca, Eli Lilly, Merck Sharp & Dohme, ZaiLab, Roche, Eisai, and Merck, received payment for lectures, presentations, speakers bureaus, manuscript writing, or educational events from Novartis, AstraZeneca, Sanofi, Pfizer, ZaiLab, Eisai, Eli Lilly, Merck Sharp & Dohme, and received funding for travel or attending meetings from Pfizer, Astellas, Novartis, MSD, Roche, Eisai, Merck, Sanofi, Bristol-Myers Squibb, and ZaiLab. V.H.F.L. is an Advisory Board member or a Data Safety Monitoring Board member of Amgen, AstraZeneca, Merck Sharp & Dohme, Pfizer, and Takeda, received consultation fees from AUXILAB, received payment and honoraria for lectures, presentations, speakers bureaus, manuscript writing or education events from Amgen, AstraZeneca, Boston Scientific, Eli Lilly, Merck Sharp & Dohme, Novartis, Pfizer, and Takeda, received traveling funding from AstraZeneca and Takeda, received receipt of equipment, materials, drugs, medical writing, gifts, or other services from AstraZeneca, and received grants and contracts from AstraZeneca for his institution. The remaining authors have declared that no conflict of interest exists.

**Acknowledgments**

We thank Prof. George Sai Wah Tsao and his team in the School of Biomedical Sciences in the University of Hong Kong for their dedication to establish the normal immortalized nasopharyngeal cell lines (NP361 and NPS56) with and without EBV infection, the normal immortalized NP cell lines NP69 and NP460 and NPC cell lines NPC43, C17, NPC53 and NPC8, and for the advice for culturing the EBV-NPC cell lines and normal NP cell lines. We thank the NPC AoE Tissue Bank clinical and research support teams for providing the clinical specimens and the AoE NPC Cell Line Repository in the University of Hong Kong for providing the NPC cell lines. We acknowledge the Centre for PanOmics Sciences (CPOSS) in the University of Hong Kong for the bisulfite pyrosequencing service.

**Appendix A. Supplementary data**

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2022.104357.

**References**

1. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet. 2019;394(10192):64–80.
2. Shannon-Lowe C, Rickinson AB, Bell AI. Epstein-Barr virus-associated lymphomas. Philos Trans R Soc Lond B Biol Sci. 2017;372(1731).
3. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202–209.
4. Jiang W, Liu N, Chen XZ, et al. Genome-wide identification of a methylation gene panel as a prognostic biomarker in nasopharyngeal cancer. Mol Cancer Ther. 2015;14(12):2864–2873.
5. Dai W, Cheung AK, Ko JM, et al. Comparative methylene analysis in solid tumors reveals aberrant methylation at chromosome 6p in nasopharyngeal carcinoma. Cancer Med. 2015;4(7):1079–1090.
6. Zheng H, Dai W, Cheung AK, et al. Whole-exome sequencing identifies multiple loss-of-function mutations of NF-kappaB pathway regulators in nasopharyngeal carcinoma. Proc Natl Acad Sci USA. 2016;113(40):11283–11288.
7. Li YY, Chung GT, Lui VW, et al. Exome and genome sequencing of nasopharynx cancer identifies NF-kappaB pathway activating mutations. Nat Commun. 2017;8:14121.
8. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet. 2019;20(4):207–220.
9. Yang Y, Ye X, Dai R, et al. Phage separation of Epstein-Barr virus EBNA2 protein reorganizes chromatin topology for epigenetic regulation. Commun Biol. 2021;4(1):967.
60 Hui KF, Chan TF, Yang W, et al. High risk Epstein-Barr virus variants characterized by distinct polymorphisms in the EBER locus are strongly associated with nasopharyngeal carcinoma. Int J Cancer. 2019;144(12):3031–3042.

61 Leong MML, Lung ML. The impact of Epstein-barr virus infection on epigenetic regulation of host cell gene expression in epithelial and lymphocytic malignancies. Front Oncol. 2021;11:629780.

62 Leong MML, Cheung AKL, Dai W, et al. EBV infection is associated with histone bivalent switch modifications in squamous epithelial cells. Proc Natl Acad Sci U S A. 2019;116(28):14144–14153.

63 Li L, Li C, Mao H, et al. Epigenetic inactivation of the CpG demethylase TET1 as a DNA methylation feedback loop in human cancers. Sci Rep. 2016;6:26591.

64 Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell. 2010;18(6):553–567.

65 Thienpont B, Steinbacher J, Zhao H, et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. Nature. 2016;537(7618):63–68.

66 Schutsky EK, Nabel CS, Davis AKF, DeNizio JE, Kohli RM. APO-BEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA. Nucleic Acids Res. 2017;45(13):7655–7665.

67 Bellacosa A, Drohat AC. Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. DNA Repair. 2015;32:33–42.

68 Sands MA, Chew E, Flensburg C, et al. MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML. Blood. 2018;132(14):1526–1534.

69 Kemp CJ, Moore JM, Moser R, et al. CTCF haplinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell Rep. 2014;7(4):1020–1029.

70 Wang H, Maurano MT, Qu H, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 2012;22(9):1680–1688.

71 Fang C, Wang Z, Han C, et al. Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation. Genome Biol. 2020;21(1):247.

72 Atiken SJ, Ibarra-Soria X, Kanteponzidou E, et al. CTCF maintains regulatory homeostasis of cancer pathways. Genome Biol. 2018;19(1):106.

73 Aitken SJ, Ibarra-Soria X, Kentepozidou E, et al. CTCF loss mediates unique DNA hypermethylation landscapes in human cancers. Clin Epigenet. 2020;12(1):80.

74 Damaschke NA, Gawdzik J, Avilla M, et al. CTCF loss mediates unique DNA hypermethylation landscapes in human cancers. Clin Epigenet. 2020;12(1):80.

75 Hori I, Kawamura R, Nakayashiki K, et al. CTCF deletion syndrome: clinical features and epigenetic delineation. J Med Genet. 2017;54(12):836–842.

76 Tsao SW, Tsang CM, Lo KW. Epstein-Barr virus infection and nasopharyngeal carcinoma. Philos Trans R Soc Lond B Biol Sci. 2017;372(1732).

77 Murata T. Regulation of Epstein-Barr virus reactivation from latency. Microbiol Immunol. 2014;58(6):307–317.

78 Lyons DE, Yu KP, Vander Heiden MA, et al. Mutant cellular AP-1 proteins promote expression of a subset of Epstein-barr virus late genes in the absence of lytic viral DNA replication. J Virol. 2018;92(19).

79 Li H, Liu S, Hu J, et al. Epstein-Barr virus lytic reactivation regulation and its pathogenic role in carcinogenesis. Int J Biol Sci. 2016;12(11):1309–1318.

80 Imaoka M, Tanese K, Masugi Y, Hayashi M, Sakamoto M. Macrophase migration inhibitory factor–CD74 interaction regulates the expression of programmed cell death ligand 4 in melanoma cells. Cancer Sci. 2019;110(7):2273–2283.

81 Zheng YX, Yang M, Rong TT, et al. CD74 and macrophase migration inhibitory factor as therapeutic targets in gastric cancer. World J Gastroenterol. 2012;18(18):2253–2261.

82 Schroder B. The multifaceted roles of the invariant chain CD74–More than just a chaperone. Biochim Biophys Acta. 2016;1863(6 Pt A):1269–1281.

83 Karwacz K, Miraldi ER, Pokrovski M, et al. Critical role of IRF1 and BATF in forming chromatin landscape during type 1 regulatory cell differentiation. Nat Immunol. 2017;18(4):412–421.

84 Shao LL, Hou WZ, Scharping NE, et al. IRF1 inhibits antitumor immunity through the upregulation of PD-L1 in the tumor cell. Cancer Immunol Res. 2019;7(8):1258–1266.

85 Garber M, Yosef N, Goren A, et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol Cell. 2012;47(5):810–822.

86 Shao LL, Hou WZ, Scharping NE, et al. IRF1 inhibits antitumor immunity through the upregulation of PD-L1 in the tumor cell. Cancer Immunol Res. 2019;7(8):1258–1266.

87 Lee SJ, Jang BC, Lee SW, et al. Interferon regulatory factor-1 is prerequisite to the constitutive expression and IFN-gamma-induced upregulation of B7-H1 (CD274). FEBS Lett. 2006;580(3):755–762.

88 Jin S, Li R, Chen MY, et al. Single-cell transcriptomic analysis defines the interplay between tumor cells, viral infection, and the microenvironment in nasopharyngeal carcinoma. Cell Res. 2020;30(11):950–965.