Genome-Wide Analysis of DNA Methylation Dynamics During Cervical Carcinogenesis

Jianfang Li
Guangzhou Medical University
Wen Teng
Guangzhou Medical University
Yuan Wang
Guangzhou Medical University
Yuan Tan
Guangzhou Medical University
Dongmei Zhou
Guangzhou Medical University
Ting Cao
Guangzhou Medical University
Yuhua Ou
Guangdong Women and Children Hospital
Caizhou Zhong
Guangzhou Medical University
Yingqing Deng
Guangzhou Medical University
Wei Li
Guangzhou Medical University
Jieying Xu
Guangzhou Medical University
Ting Wang
Guangzhou Medical University
Yongkun Ji
Guangzhou Medical University
Huilong Long
Guangzhou Medical University
Linlang Guo
Southern Medical University
Qianqian Liu
Guangzhou Medical University
Lipai Chen

Guangzhou Medical University

Ruiyu Xie

University of Macau

Xiping Luo

Guangdong Women and Children Hospital

Xiujie Sheng

Guangzhou Medical University

Xinke Zhou

Guangzhou Medical University

Deqiang Sun (deqiangs@zju.edu.cn)

zhejiang university https://orcid.org/0000-0002-1136-8551

Primary research

Keywords: Genome, DNA, methylation, carcinogenesis

Posted Date: September 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-895713/v1

Genome-wide analysis of DNA methylation dynamics during cervical carcinogenesis

Jianfang Li[1,2#], Wen Teng[1,2,3#,] Yuan Wang[1,2#], Yuan Tan[1,2#], Dongmei Zhou[4], Ting Cao[4], Yuhua Ou[5], Caizhou Zhong[1,2], Yingqing Deng[1,2], Wei Li[1,2], Jieying Xu[1,2], Ting Wang[1,2], Yongkun Ji[1,2], Huilong Long[1,2], Linlang Guo[6], Qianqian Liu[7], Lipai Chen[7], Ruiyu Xie[8*], Xiping Luo[5*], Xiujie Sheng[4*], Xinke Zhou[1,2*], Deqiang Sun[1,2*]

# These authors contributed equally to this study.

* Correspondence: deqiangs@zju.edu.cn (leading contact); zxkstar@126.com; 2008691150@gzhmu.edu.cn; 2488692399@qq.com; ruiyuxie@um.edu.mo

Affiliations:

1. Innovation centre for Advanced Interdisciplinary Medicine, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou 510700, China

2. Guangzhou Key Laboratory of Enhanced Recovery after Abdominal Surgery, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou 510700, China

3. Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China.

4. Department of Obstetrics and Gynecology, Center for Reproductive Medicine/Department of Fetal Medicine and Prenatal Diagnosis/BioResource Research Center, Key Laboratory for Major Obstetric Disease of Guangdong Province, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, 510150, China

5. Department of Obstetrics and Gynecology, Guangdong Women and Children Hospital, Guangzhou, 511400, Guangdong, China

6. Department of Pathology, Zhujiang Hospital, Southern Medical University, Guangzhou, China

7. Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou, China

8. Centre of Reproduction, Development, and Aging, Faculty of Health Sciences, University of Macau, Taipa, Macau SAR.

Abstract

Background

Cervical cancer is the second most common gynaecological tumor of women worldwide, however, the molecular mechanism for the cervical carcinogenesis remains elusive. Current study provides a series of genome-wide DNA methylation blueprints of normal cervix and cervical cancers using Whole Genome Bisulfite Sequencing.

Results

DNA methylation dynamic alternations during cervical carcinogenesis were focused on the signal pathway of TGF-beta and epidermal growth factor, which could be used to monitor the treatment response and Tumorigenesis. Transcription factor of E2F6, MBD2 and STAT3 were interfered by aberrant methylation in cancer development. Furthermore, those identified novel methylation markers for the risk of progression along the spectrum of lesion grades could provide new insights into the prevention and treatment.

Conclusion

DNA methylation signature in cervical cancers can serve as valuable epigenetic markers to guide the clinical treatment. The epigenetic features detected in this study can be exploited for previously unidentified biomarker and prognostic marker development.

Background

Cervical cancer is the second most common gynaecological tumor of women worldwide, accounting for more than 500,000 new cases and 300,000 deaths each year [1]. Current therapies include surgery and radiotherapy, but the 5-year survival rate is only about 15% among advanced patients, or worse, there has been little progress in the development of new therapies in recent years [2]. Infection with carcinogenic human papillomavirus (HPVs) is occurred in nearly all the cases of cervical cancer and are considered as the high-risk factor in the etiology of cervical cancer [3]. However, HPV alone is insufficient to cause cervical carcinoma, because most HPV infections could be cleared within months. Only a few of HPVs persist in the host and inactivate p53 and retinoblastoma proteins through HPV E6 and E7 oncoproteins, increasing the genomic instability

and even integrating HPV into the host genome in some cases [4]. The molecular mechanism for the HPV initiated cervical carcinogenesis remains elusive.

Over the past decades, conventional Papanicolaou (Pap) smears testing as a cytological screening method has been widely used to identify cervical cancer and its precursors. However, the specificity and sensitivity of Pap testing for high-grade squamous intraepithelial lesion (HGSIL) was 96.8% and only 55.4% [5]. The knowledge of the relationship HPVs and cervical neoplasia has contributed to the high-risk HPVs as surrogate markers in clinical screening. Although the sensitivity of HPV testing was much higher for HGSIL patients (94.6%) than Pap testing, the high false-positive ratio might result in unnecessary medical and psychological burdens, especially in early stages of disease [5,6]. To make the early detection of cervical cancer much more feasible and effective, a screening test with a better diagnostic performance needs to be developed. DNA methylation biomarker detection seems to be a promising approach with more sensitivity, higher specificity and more reproducible, which need further development [7,8].

DNA methylation is one of the well-recognized epigenetic modifications in human cells. Heritable changes regulated by abnormal DNA methylation play a critical role in development, differentiation, aging and human diseases such as schizophrenia and most cancers [9]. Both hypo- and hypermethylation of DNA methylation are increasingly observed in human malignancies in recent years. Hypomethylation of proto-oncogenes, such as *RAS* and *WNT*, leads to the gene reactivation and chromosomal instabilities and finally results in tumorigenesis [10,11]. On the contrary, the promoter hypermethylation could inactivate the tumor suppressor genes (TSGs) and other cancer associated genes, which are involved in many biological functions, such as cell-cycle control, apoptosis and DNA repair [12,13]. Aberrant promoter methylation changes of oncogenes and TSGs also detected in cervical cancer, including *p16*, *CADM1*, *MAL*, *DAPK1*, *MGMT*, *MYOD1*, *TIMP3* and so on [14][15]. However, most previous studies in cervical cancer restricted to specific loci in a limited number of genes that are known in other cancers.

Over the past decade, many microarray and high-throughput sequencing methods have emerged and promote the analyses of DNA methylation from few loci to a genome-wide scale. The bisulfite pyrosequencing has developed as a quantitative method for evaluating DNA methylation level at single CpG sites resolution. As the first version BeadChip for human methylation analysis, Infinium 27 K array (HM27) was used to discover the abnormal DNA methylation

changes between pooled SCC patients and normal cervical scraping cells [16]. A series of 255 squamous cell carcinomas (SCC), 49 adenocarcinomas (CCA) and 3 Healthy controls has been analyzed with HM450 array (TCGA; https://tcga-data.nci.nih.gov/tcga). A recent work using Illumina Infinium Methylation EPIC array (HM850) identified new methylation markers along the continuum of cervical intraepithelial neoplasia (CIN) to cervical cancer, including 54 normal, 50 LGSIL, 40 CIN2 and 42 HGSIL samples [17]. Although those microarrays characterize the methylation status of 27,000 ~ 850,000 CpG sites across the multiple gene promoters and enhancer repertoires, they only covered less than 3% of the total of 28 million CpG sites known in human genome, the DNA methylation profile of whole genome remain unclear and need further demonstrations.

In the current study, we obtained DNA methylomes of different stages of human cervical cancer tissues and normal cervix control employing Whole Genome Bisulfite Sequencing (WGBS), which was considered as the golden standard of true methylation states. Herein, we first provided the genome-wide DNA methylation landscapes of normal cervix, as well as cervical cancers. We elucidated the DNA methylation dynamic alternations during cervical carcinogenesis and discussed their underlying mechanisms in cervical cancer. We identified novel methylation markers for the risk of progression along the spectrum of lesion grades. Taken together, our study could further our understanding of cervical carcinogenesis and might provide new insights into the prevention and treatment.

RESULTS

DNA methylation landscapes of normal cervix and cervical cancers

To profile the DNA methylation patterns in cervical cancers, we collected precancerous lesion samples from patients with LGSIL (n = 6) and patients with LGSIL (n = 5), and cervical cancer tissues from patients with squamous cell carcinoma (n = 5) and patients with adenocarcinoma (n = 2) (Fig. 1a). Meanwhile, we collected normal cervical epitheliums from healthy contributors (n = 5) used as control samples. The ages of patients used in this study were ranging from 31 ~ 50 with an average of 38.7 years old (Table S1). All the total of 23 samples were successfully sequenced by deeply WGBS and an average of 431.4 million 150 pair-end reads were obtained with high bisulfite conversion efficiencies (BCE $\geq$ 0.987), suggesting the successful run of library

construction and sequence. Over 85.6% of sequencing reads were uniquely aligned to genome reference with an average high sequencing depth of 23× on CpG dinucleotides, and covering 27.0 - 28.1 million CpG sites sequenced by at least three times, with classical bimodal distribution of methylation ratio in each sample (Table S1 and Fig. S1a, b). To further evaluate the data quality, we performed Pearson correlation analysis among individuals from five groups. The results showed that there are relatively high Pearson correlation coefficients (PCCs) between samples come from the group of healthy control (PCC ≥ 0.87), followed by precancerous lesions (PCC ≥ 0.82). While the PCCs between individuals from the group of cervical cancers were much lower (average PCC < 0.79), might suggest an apparent increasing global DNA methylation disorder in cervical tumor tissues (Fig. S1c). Among groups, patients with precancerous lesions and tumors were obviously separated from normal cervix controls with decreased PCC values during cervical cancer development (Fig. S1c). The average DNA methylation ratio revealed that normal cervix samples had the similar and higher genome-wide methylation ratio (0.78 ± 0.005). However, the methylation ratio of cervical precancerous lesion and cancer individuals were relatively low and varied within groups, ranging from 0.74 ± 0.01 to 0.76 ± 0.02, which also suggested the methylation disorder in cervical patients (Fig. 1b and Fig. S1d). The decreased genome-wide methylation contributes to genomic instability and result in the occurrence of tumor, which has been reported in multiple cancers [18]. At the annotated gene coding regions, methylation ratio was almost depleted at transcription start site (TSS) but enriched through gene body at each group (Fig 1C and Fig. S1e). To further claim the DNA methylation pattern, we calculated the distribution of methylation ratio within various genome elements (Fig. 1d and Fig. S2). Except for the whole genome level, we also observed that the methylation ratio was descended in precancerous lesion and tumor samples at gene body, gene intron, intergenic genes and repeat elements, such as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and long terminal repeats (LTRs). Oppositely, the methylation level at CGI and gene promoter region was ascended in precancerous lesions and tumors compared to normal cervix.

To validate the methylation ratio of WGBS data, we compared our result with previous public Human Methylation Array data from hundreds of individuals [19]. We found that our identified CpG sites (depth ≥ 3) covered over 99.7% of the CpGs designed in methylation arrays in each group, respectively (Fig. S1f). Within those CpG sites detected in both WGBS and array, high Pearson correlation (average values > 0.8) between matched samples suggest that our WGBS result

is reliable and the genome globally CpG sites are much powerful for uncovering the methylation regulation mechanism in cervical cancer than methylation array data (Fig. S1g).

DNA methylation signature in precancerous lesion and cervical cancer

We next identified the differentially DNA methylation cytosines (DMCs) in precancer lesions (LGSIL and HGSIL) and cervical cancers (SCC and CCA) compared to normal cervix tissue to unveil tumor driving methylation alternation during cervical carcinogenesis. For the detail, 141,903 Hyper- and 384,265 Hypo-DMCs were detected in LGSIL patients, 156,862 Hyper- and 179,793 Hypo-DMCs in HGSIL, 395,275 Hyper- and 311,657 Hypo-DMCs in SCC and 253,737 Hyper- and 373,935 Hypo-DMCs in CCA with different methylation change (Fig. 2a). Notably, the methylation level of much more CpGs were changed in tumors than that in precancer lesions (Fig. 2a). Based on the methylation ratio of the total of identified DMCs across all 23 samples, principal component analysis significantly separated samples from normal cervix, precancerous lesions, and cervical tumor, except one CCA patient was closed to precancerous lesions (Fig. 2b). Similar result was also detected in the common DMCs detected in both our WGBS and previous HM450/HM850 array data (Fig. S3a). Furthermore, the methylation changes of both SCC and CCA tumors in WGBS data were highly consistent with previous array data with PCC equal to 0.91 and 0.89, respectively (Fig. S3b). Those results indicated our identified DMCs in cervical patients is reliable. Next, functional genome analysis divulged that the percentage of Hyper-DMCs across CpGs located in gene exon, 3'UTR and low complexity repeat region, as well as highly methylated CpG Islands in normal cervix tissue (CGI $\geq$ 0.9), compared to global genome reference. In contrast, Hypo-DMCs tend to frequently occur in repeat elements (Fig. S3c). These findings were consistent with previously observation, which said that a significant fraction of the hypermethylated CpG loci located in CGI in contrast to the hypomethylated ones located in repetitive regions in tumor cells [20]. Notably, we then integrated compared the DMCs between each two patient groups and found that very few DMC was switched from hypermethylation to hypomethylation, and vice versa, indicating that the methylation changes in LGSIL, HGSIL and cervical cancers are consistent (Fig. S3d). Separately, over 90.59% Hyper-DMCs and 66.95% Hypo-DMCs were detected in any cervical tumors (SCC or CCA), suggesting that the most methylation variation occurred in tumor cells, through there are 207,992 (22.95%) Hypo-DMCs only changes in LGSIL patients (Fig. 2c). This DMC information might could serve as potential biomarkers to report cervical carcinogenesis during clinical early screening. For example, we

observed hypermethylation at the *HIST1H4* locus in squamous cell carcinoma (Fig. 2d), similar abnormal hypermethylation has been reported in 17 types of cancers, including in non-small-cell lung cancer to classify high- and low-risk stage I and patients with shorter relapse-free survival [21].

DNA methylation dynamics in cervical cancer carcinogenesis

To identify the essential DNA methylation patterns and genes that initiate and maintain the cervical carcinogenesis events, we analyzed the differentially methylated regions (DMRs) based on previously called DMCs using MOABS. We totally observed 11,376 Hyper- and 24,370 Hypo-DMRs in LGSIL; 12,185 Hyper- and 9,336 Hypo-DMRs in HGSIL, 33,920 Hyper- and 15,885 Hypo-DMRs in SCC and 19,471 Hyper- and 17,187 hypo-DMR in CCA (Fig. 3a). In agreement with DMCs, Hyper-DMRs tend to locate in CGI while Hypo-DMRs tend to locate in repeat element regions (Fig. 3b). We next classified these Hyper/Hypo-DMRs into seven clusters based on the DNA methylation dynamics in cervical cancer development (Fig. 3c). Most of the DMRs undergo consistent one direction change (consistently Hyper/Hypo-DMRs (LGSIL-to-CC), gained Hyper/Hypo-DMRs (HGSIL-to-CC), and loss Hyper/Hypo-DMRs (LGSIL specific and LGSIL-to-HGSIL)) (Fig. 3d and Fig. S4a). Motif analysis revealed that many reported tumorigenesis-related transcription factors (TF) motifs were enriched within DMRs (Fig. S4b). For example, bHLH, Forkhead, Homeobox and MADS were mostly associated with LGSIL-to-CC cluster, HGSIL-to-CC and CC-specific clusters [22–25]. Functional enrichment with the GREAT indicated these consistently Hyper-DMRs (LGSIL-to-CC) were significantly enriched for the transforming growth factor beta (TGFB) signaling pathway, while consistently Hypo-DMRs (LGSIL-to-CC) were enriched for the Notch signaling pathway and epidermal growth (Fig. 3e) [26]. We chose the IRF3 gene as an example and found that the promoter region of IRF3 gene is significantly hypomethylated in precancerous lesions and cervical tumor samples than normal tissue. (Fig. 3g). IRF3 is a well-known transcription factor that plays an essential role in the process of the innate immune system's response to viral infection, such as the high-risk human papillomavirus E6 (hrHPV E6) [27–29]. To check whether these DMRs subsequently have an impact on gene regulation during cervical tumorigenesis, we next employed gene expression data of adenocarcinoma from GSE145372. We found that gene's expression is reversely correlated with the DNA methylation changes within gene transcription start site (TSS) up/downstream 1 kb (Fig. S4C). Considering bisulfite sequencing cannot discriminate between 5-methylcytosine (5mC) and

5-hydroxymethylcytosine (5hmC), we further compared our 5mC data with public 5hmC data downloaded from E-MTAB-7810. We found 57.9% of CpGs within Hyper-DMRs of squamous cell carcinoma in our data showed reduced 5hmC in early-stage (I to IIa) squamous cell carcinoma, and 65.2% of CpGs within Hyper-DMRs of squamous cell carcinoma in our data showed reduced 5hmC in advanced stage (IIb to IV) squamous cell carcinoma. Meanwhile, there were 42.1% of CpGs within Hypo-DMRs of squamous cell carcinoma in our data had increased 5hmC in early-stage (I to IIa) squamous cell carcinoma, and 36.4% of CpGs within Hyper-DMRs of squamous cell carcinoma in our data had increased 5hmC in advanced stage (IIb to IV) squamous cell carcinoma (Fig. S4D).

Methylation dynamics disruptes Transcription factor binding and target gene expression

In cancer, many signaling pathways were dysregulated, such as 'Wnt signaling pathway', 'Notch signaling pathway', which leads to uncontroled growth of cells [30,31]. The common processes for signaling transduction in these pathways need receptors, proteins that convey the signals and transcription factors (TF) that regulate specific gene expression in the downstream [32]. Therefore, the dysfunction of TF is one of the leading factors in carcinogenesis. To study the effect of methylation on TF in cervical cancer, all cervical cancer related TFs were collected from cistrome database. We found the methylation ratio of E2F6 promoter decreased as the disease progressed from normal cervix to cancer, as well as the expression level of E2F6 was up-regulated (Fig4. A). E2F6 as a member of E2F family paticipates in cell cycle process and promotes cell proliferation by targeting E2F target genes [33–35]. Our results suggest E2F6 as a potential risk factor in cervical cancer in which E2F6 high expression is caused by hypo-methylation in promoter region.

Besides to activate or repress the TF per se, methylation on TF target regions has great effect on binding affinity [36–38]. To find the TFs which binding were significantly disrupted by methylation changes, the proportion of DMC on TF peaks were measured for each TF. Dramatically hyper-methylation was found on MBD2 peaks in cancer patients (Fig4. B and FigS5. A). And the mean methylation levels on MBD2 peaks enhance gradually as disease progressed (Fig. 4 C). In additon, 919 MBD2 target genes were differentially expressed in cancer (449 down-regulated and 470 up-regulated) (Fig. 4 D), and functional enrichment analysis shows that dysfunction of MBD2 was related with many cancer pathways (Fig. 4 E). For example, in Wnt signaling pathway, the abberant methylation on MBD2 targets improve the transmission of

extracellular signal by inhibiting *Rspo* and *FRP* and further promote the cell cycle by activating *Ponthin52* and *TCF/LEF* (Fig. 4 F). Dysregulation of other cancer pathways, such as 'Hippo signaling pathway' and 'cAMP signaling pathway', can also be attributed to methylation alteration on MBD2 targets (Fig. S5b). STAT3 is another TF that shows significantly hypo-methylation on binding peaks in cancer (Fig. S5c). The differential methylation on STAT3 peaks account for 1160 DEGs in cancer (373 down-regulate and 787 up-regulate). Functional enrichment results showed hypo-methylation on STAT3 binding regions activate *hDLG*, *Bak*, *p43* and *hTid-1*, which contribute to cervical cancer development by promoting cell transformation, proliferation and immunosuppression (Fig. S4d, e). STAT3 also invovled in 'JAK-STAT signaling pathway' regulating (Fig. S5f, g), and it has been reported that the binding of STAT3 is related with IL-10 expression [39]. which is an immune suppressor and can mediate the macrophage polarization into M2 phenotypic in cervical cancer [39]. Taken together, abberant methylation play a key role in cancer development by interfering with the function of TFs.

HPV infection causes methylation change in cervical cancer.

To uncover the influence of HPVs infection on the cervical cancer, we counted the sequenced reads come from each HPV virus and compared the result to the clinical HPV testing for each patient, respectively. The detected HPV type was corresponding to the HPV testing in most samples, though some of them were simply tested as 12 high risk HPVs positive (including HPV16 and HPV18), indicating the reliable quality of our WGBS library and the clinical HPV testing (Fig. 5a). We then identified the HPV signature DMCs from at least two patients with the infection of HPV16, HPV18 and HPV31, respectively (HPV52 signature DMC number is too little to show in this study). 1,078 - 4,380 Hyper- and 3,348 - 10,679 Hypo-DMCs were found and those DMCs might be caused by specific HPV infection (Fig. 5b). Motif enrichment analysis revealed that Homeobox motifs were most prominently enriched around Hyper-DMCs, especially in HPV18 signature Hyper-DMCs (Fig. 5c, Top). COUP-TFII motif was associated with the HPV31 signature Hyper-DMCs. Whereas, bZIP and ETS motifs were enriched around HPV signature Hypo-DMCs (Fig. 5c, Bottom). To further validate this observation, we compared our data with published COUP-TFII and CEBPB (bZIP family) ChIP-seq data collected in HeLa cells. Similar, COUP-TFII and CEBPB were mostly enriched around HPV31 Hyper-DMCs and HPV18 Hypo-DMCs, respectively (Fig. 5d, e). Those results indicated those motif family may play a crucial role in the formation and progression of HPVs induced cervical cancer.

Discussion

Previous studies have discussed the importance of DNA methylation in human cancers. To date, no studies have been undertaken to investigate the methylation status across the genome wide in samples of cervical cancer patients. In the present study, we firstly reported the DNA methylation blueprint for a series of cervical tissues, including normal cervix, low-grade squamous intraepithelial lesions (LGSIL), five high-grade squamous intraepithelial lesions (HGSIL), squamous cell carcinoma (SCC) and two adenocarcinoma (CCA). Those results provide a standard methylation status for cervical patients.

Comparing the DNA methylation status between cervical patients and normal cervix, aberrant methylation patterns were found during the multistage pathogenesis of cervical cancer with an increasing trend to methylation with increasing pathological changes, which similar to previous study [40]. Furthermore, genes related to TGF-beta signaling pathways and epidermal growth factors were respectively hypermethylated and hypomethylated in LGSIL, HGSIL and cervical cancers. Those results indicating the disruption of TGF-beta signaling pathway and other pathways might contribute to the malignant progression of human cervical tumors. On the other hand, we also identified DNA methylation dynamics could disrupt the binding of transcription factors, such as E2F6, MBD2 and STAT3, then dysregulate the expression of target genes. Those changes might lead to the uncontrolled growth of cervical tissue cells, which has been reported in previous studies [31,41].

As the knowledge of most cervical cancer were caused by persist infection of different HPV types, we identified the HPV signature DMCs using our dataset, and identified 1,078 - 4,380 Hyper- and 3,348 - 10,679 Hypo-DMCs in samples infected by HPV16, HPV18 and HPV31, respectively. Furthermore, Homeobox and bZIP motif families may play a curial role during the cervical cancer development caused by HPVs. Those markers could be used for the clinical detection and therapy in the further.

Conclusions

Overall, DNA methylation signature in cervical cancers can serve as valuable epigenetic markers to guide the clinical treatment. The epigenetic features detected in this study can be exploited for previously unidentified biomarker and prognostic marker development.

Methods

Patients and Samples Preparation

In this study, five squamous cell carcinoma (SCC) and two adenocarcinoma (CCA) tissues were collected from cervical cancer patients, six low-grade squamous intraepithelial lesions (LGSIL) and five high-grade squamous intraepithelial lesions (HGSIL) were collected from precancerous lesion patients, and five normal cervical epitheliums from healthy contributors were used as control. The total of 23 samples were originated from both The Third Affiliated Hospital of Guangzhou Medical University and Guangdong Women and Children Hospital, which have been subjected to pathological diagnosis. HPV typing of patient specimens was performed using INNO-LiPA Genotyping Extra kit (Innogenetics, Barcelona, Spain). The demographic and clinical information of patient were listed in Table S1. All samples were frozen in liquid nitrogen quickly and preserved in the -80 ^{O}C freezer.

WGBS library preparation and data analysis

Genomic DNA was isolated using the DNeasy Blood & Tissue kit (Qiagen). Library preparation and high-throughput sequencing were conducted by NovaGene (Beijing, China). In brief, purified genomic DNA (with 1% unmethylated lambda DNA spike-in, Promega) was sheared to a fragment size of 100–700 bp (primary size 250 bp) and subjected to bisulfite conversion using the EZ DNA Methylation-gold kit (Zymo Research). Bisulfite-converted DNA was amplified with 13 PCR cycles and purified by AMPure XP beads (Beckman Coulter). Each library was deep sequenced on Illumina HiSeq6000 (150-cycle, paired-end).

For WGBS data analysis, we first generated a reference genome including NCBI Human Reference Genome Build GRCh38 (hg38) and all the types of HPV. Raw FASTQ files of 150 bp paired-end reads were filtered and mapped to the reference genome using bsmap v.2.89 with paired mode, and only uniquely mapped reads were retained [42]. BSeQC was applied to perform quality control [43]. The CpG coverage and DNA methylation ratio were calculated using MCALL

module in MOABS software (--trimWGBSEndRepairPE2Seq 10) [44]. Bisulfite conversion efficiency for each sample was estimated based on unmethylated lambda DNA spike-in. Only CpG sites with coverage of more than three were used for downstream analysis. The significantly differential methylated CpG sites and regions (DMCs and DMRs) were identified using the MCOMP module considering variance among clinical samples (--withVariance 1) in MOABS software. The output bedGraph files from MCALL included single-base resolution DNA methylation ratios, which were transformed into bigwig file format and were uploaded to the UCSC genome browser for visualization. Motif enrichment analysis of DMRs was performed using HOMER software, functional annotation was performed using GREAT with default settings, and many plots related to WGBS was performed using MMINT [45].

To compare the WGBS and methylation assay, the microarray datasets of Illumina HumanMethyaltion450 (TCGA-CSEC) and Illumina Infinium Methylation EPIC array (GSE143752) were downloaded from TCGA and NCBI GEO, respectively. The raw transcriptome sequence data were downloaded from GSE145372 and PRJNA262780. Raw reads were aligned to the hg38 using HISAT2 with default parameters, and only uniquely mapped reads were used for downstream analysis. A count matrix for each gene was generated using htseq-count (HTSeq package). DESeq2 was used to identify significant DEGs between groups (fold change (FC) ≥ 2; False Discovery Rate (FDR) < 0.05).

Transcription factor analysis

ChIP-seq peak files for cervical cancer related transcription factors (TFs) were collected from cistrome database (http://cistrome.org), which involved in 158 TFs from 3 different cell types (HeLa, HeLa-S3 and KB cell) and cervical cancer tissues. The promoter methylation level of these TFs were defined as the mean methylation for all CpG sites situated within 1 Kb upstream and 500 bp downstream of TSS region. To identify the TFs which binding affinity were disrupted by abberant methylation level in cancer, the percentage of differentail methylated CpG sites on peaks were calculated for each TF in each group. Genes with binding peaks in promoter region are defined as TF target genes. The expression of TF and their targets were studied in public TCGA CESC RNA-seq datasets, PRJNA262780 and GSE145372. Differential gene expression analysis was performed with DESeq2 and the functional analysis was carried out with R package 'clusterProfiler'. Deeptools was used to generate the profile of methylation level on TF peak

regions, with 'computeMatrix scale-regions' function [46]. The TF peaks were scaled to 300 bp and the value was calculated on 10 bp windows.

HPV analysis

In this study, LiBis was employed to improve the HPV mapping efficiency and recognize the HPV DNA insertion [47]. After the alignment of BSMAP, LiBis dynamically clipped unmapped reads by 40 bp sliding window (--window 40) at 5 bp step size (--step 5) and remapped obtained clipped reads to reference genome including hg38 and all HPV types. Only uniquely and continuously mapped reads were kept for the following recombination process to obtain longer extended fragments, which were finally considered as the highly confident rescued reads. The reads aligned to each HPV type were counted for each sample. MOABS was used to identify DMCs for each patient with specific HPV infection compared to healthy controls without HPV infection. Strictly, DMCs detected in all patients with specific HPV infection but not detected in other patients, were identified as HPV signature DMCs and used for further analysis. The bigwig files of COUP-TFII and CEBPB were downloaded from GSM1126878 (Hela) and GSM935553 (Hela-S3), and the motif profiles across HPV specific DMCs were calculated using Deeptools [46].

Declarations

Ethics approval and consent to participate

Samples used in this study were approved by the institutional review board of The Third Affiliated Hospital of Guangzhou Medical University (No: 2019–037). Informed consents were obtained from all participants in this study.

Consent for publication

Not applicable.

Data availability

WGBS data from this study were submitted to the China National GeneBank DataBase (CNGBdb; https://db.cngb.org) under accession number CNP0001763. All relevant data supporting the critical findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request.

Competing interests

The other authors declare no other competing interests.

Funding

This work was supported by grants from the National Natural Science Foundation of China (NSFC, Project 81773012) and Guangdong Provincial Natural Science Foundation (2020A1515010082).

Authors' contributions

D.S. and X.Z directed and oversaw the project. Y.T., D.Z., Y.O. and T.C. collected cervical tissue samples. J.L., W.T. and Y.W. performed comprehensive bioinformatics analyses. J.L., W.T., Y.W., D.S. wrote the manuscript with all the other authors participating in discussion, data interpretation and manuscript editing.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021;71:7–33.

2. Feng C, Dong J, Chang W, Cui M, Xu T. The Progress of Methylation Regulation in Gene Expression of Cervical Cancer. Int J Genomics. 2018;2018.

3. Basu P, Joshi S, Sankaranarayanan R. Human Papillomavirus (HPV) Testing for Secondary Prevention of Cervical Cancer. Curr Obstet Gynecol Rep. 2015;4:201–12.

4. Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer. 2010;10:550–60.

5. Mayrand MH, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al. Human Papillomavirus DNA versus Papanicolaou Screening Tests for Cervical Cancer. N Engl J Med. 2007;35706:1579–88.

6. Kulasingam SL, Hughes JP, Kiviat NB, Mao C, Weiss NS, Kuypers JM, et al. Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: comparison of sensitivity, specificity, and frequency of referral. JAMA. 2002;288:1749–57.

7. Chen X, Gole J, Gore A, He Q, Lu M, Min J, et al. Non-invasive early detection of cancer four

years before conventional diagnosis using a blood test. Nat Commun. Springer US; 2020;11:1–10.

8. Wentzensen N, Sherman ME, Schiffman M, Wang SS. Utility of methylation markers in cervical cancer early detection: Appraisal of the state-of-the-science. Gynecol Oncol. 2009;112:293–9.

9. Baylin SB, Jones PA. A decade of exploring the cancer epigenome-biological and translational implications. Nat Rev Cancer. 2011;11:726–34.

10. Patra SK. Ras regulation of DNA-methylation and cancer. Exp Cell Res. 2008;314:1193–201.

11. van der Meide WF, Snellenberg S, Meijer CJLM, Baalbergen A, Helmerhorst TJM, van der Sluis WB, et al. Promoter methylation analysis of WNT/β-catenin signaling pathway regulators to detect adenocarcinoma or its precursor lesion of the cervix. Gynecol Oncol. 2011;123:116–22.

12. Cheung HH, Lee TL, Rennert OM, Chan WY. DNA methylation of cancer genome. Birth Defects Res Part C - Embryo Today Rev. 2009;87:335–50.

13. Esteller M. Dormant hypermethylated tumour suppressor genes: Questions and answers. J Pathol. 2005;205:172–80.

14. Mersakova S, Holubekova V, Grendar M, Visnovsky J, Nachajova M, Kalman M, et al. Methylation of CADM1 and MAL together with hpv status in cytological cervical specimens serves an important role in the progression of cervical intraepithelial neoplasia. Oncol Lett. 2018;16:7166–74.

15. Widschwendter A, Muller HM, Fiegl H, Ivarsson L, Wiedemair A, Muller-Holzner E, et al. DNA methylation in serum and tumors of cervical cancer patients. Clin cancer Res. 2004;10:565–71.

16. Chen YC, Huang RL, Huang YK, Liao YP, Su PH, Wang HC, et al. Methylomics analysis identifies epigenetically silenced genes and implies an activation of b-catenin signaling in cervical cancer. Int J Cancer. 2014;135:117–27.

17. El-Zein M, Cheishvili D, Gotlieb W, Gilbert L, Hemmings R, Behr MA, et al. Genome-wide DNA methylation profiling identifies two novel genes in cervical neoplasia. Int J Cancer. 2020;147:1264–74.

18. Agrawal A, Murphy RF, Agrawal DK. DNA methylation in breast and colorectal cancers. Mod Pathol. 2007;20:711–21.

19. The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017;543:378–84.

20. Lai HC, Lin YW, Huang THM, Yan P, Huang RL, Wang HC, et al. Identification of novel DNA methylation markers in cervical cancer. Int J Cancer. 2008;123:161–7.

21. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. J Clin Oncol. 2013;31:4140–7.

22. Perk J, Iavarone A, Benezra R. Id family of helix-loop-helix proteins in cancer. Nat Rev Cancer. 2005;5:603–14.

23. Bach D, Long NP, Luu T, Anh NH. The Dominant Role of Forkhead Box Proteins in Cancer. Int J Mol Sci. 2018;19:3279.

24. Ounzain S, Micheletti R, Beckmann T, Schroen B, Alexanian M, Pezzuto I, et al. Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs. Eur Heart J. 2015;36:353–68.

25. Mughal W, Nguyen L, Pustylnik S, Da Silva Rosa SC, Piotrowski S, Chapman D, et al. A conserved MADS-box phosphorylation motif regulates differentiation and mitochondrial function in skeletal, cardiac, and smooth muscle cells. Cell Death Dis. 2015;6:e1944.

26. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28:495–501.

27. Tamura T, Yanai H, Savitsky D, Taniguchi T. The IRF family transcription factors in immunity and oncogenesis. Annu Rev Immunol. 2008;26:535–84.

28. Saito H, Morita Y, Fujimoto M, Narazaki M, Naka T, Kishimoto T. IFN Regulatory Factor-1-Mediated Transcriptional Activation of Mouse STAT-Induced STAT Inhibitor-1 Gene Promoter by IFN-γ. J Immunol. 2000;164:5833–43.

29. Shah M, Anwar MA, Park S, Jafri SS, Choi S. In silico mechanistic analysis of IRF3 inactivation and high-risk HPV E6 species-dependent drug response. Sci Rep. Nature Publishing Group; 2015;5:1–14.

30. Polakis P. Wnt signaling in cancer. Cold Spring Harb Perspect Biol. 2012;4:a008052.

31. Maillard I, Pear WS. Notch and cancer: Best to avoid the ups and downs. Cancer Cell.

2003;3:203–5.

32. Nebert DW. Transcription factors and cancer: an overview. Toxicology. 2002;181–182:131–41.

33. Lafta IJ. E2f6 is essential for cell viability in breast cancer cells during replication stress. Turkish J Biol. 2019;43:293–304.

34. Giangrande PH, Zhu W, Schlisio S, Sun X, Mori S, Gaubatz S, et al. A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription. Genes Dev. 2004;18:2941–51.

35. Bertoli C, Klier S, McGowan C, Wittenberg C, de Bruin RAM. Chk1 Inhibits E2F6 Repressor Function in Response to Replication Stress to Maintain Cell-Cycle Transcription. Curr Biol. 2013;23:1629–37.

36. Hughes TR, Lambert SA. Transcription factors read epigenetics. Science . 2017;356:489–90.

37. Schübeler D. Function and information content of DNA methylation. Nature. 2015;517:321–6.

38. Machado ACD, Zhou T, Rao S, Goel P, Rastogi C, Lazarovici A, et al. Evolving insights on how cytosine methylation affects protein-DNA binding. Brief Funct Genomics. 2015;14:61–73.

39. Gabryšová L, Howes A, Saraiva M, O'Garra A. The Regulation of IL-10 Expression BT - Interleukin-10 in Health and Disease. In: Fillatreau S, O'Garra A, editors. Interleukin-10 Heal Dis. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 157–90.

40. Cheng Q, Lau WM, Tay SK, Chew SH, Ho TH, Hui KM. Identification and characterization of genes involved in the carcinogenesis of human squamous cell cervical carcinoma. Int J Cancer. 2002;98:419–26.

41. Liu Y, Guo JZ, Liu Y, Wang K, Ding W, Wang H, et al. Nuclear lactate dehydrogenase A senses ROS to produce α-hydroxybutyrate for HPV-induced cervical tumor growth. Nat Commun. Springer US; 2018;9:1–16.

42. Xi Y, Li W. BSMAP: Whole genome bisulfite sequence MAPping program. BMC Bioinformatics. 2009;10:1–9.

43. Lin X, Sun D, Rodriguez B, Zhao Q, Sun H, Zhang Y, et al. BSeQC: Quality control of bisulfite sequencing experiments. Bioinformatics. 2013;29:3227–9.

496 44. Sun D, Xi Y, Rodriguez B, Park HJ, Tong P, Meong M, et al. MOABS: model based analysis
497 of bisulfite sequencing data. Genome Biol. 2014;15:1–12.

498 45. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of
499 Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for
500 Macrophage and B Cell Identities. Mol Cell. Elsevier Inc.; 2010;38:576–89.

501 46. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. DeepTools: A flexible platform for
502 exploring deep-sequencing data. Nucleic Acids Res. 2014;42:187–91.

503 47. Yin Y, Li J, Li J, Lee M, Zhao S, Guo L, et al. LiBis: an ultrasensitive alignment augmentation
504 for low-input bisulfite sequencing. Brief Bioinform. 2021;22:bbaa332.

Figure And Table Legend

Fig. 1. DNA methylation profile of normal cervical tissues, pre-cancerous cervical lesions, and cervical cancer. (a) Schematic of the experimental design. Normal cervical tissue (*n* = 5), LGSIL (LGSIL, n = 6), HGSIL (HGSIL, n = 5), cervical cancers tissue including SCC (n = 5) and CCA (n = 2) were sampled in this study. (b) Global average DNA methylation levels during cervical carcinogenesis. Cytosine modification in normal cervical tissues is highly concordant than that in pathological tissues. (c) DNA methylation patterns across genes of five groups. Genes are methylated with a sharp drop of methylation at the promoter. (d) Normalized DNA methylation dynamics in genomic features across individuals.

Fig. 2. Cytosine modification changes during cervical carcinogenesis compared to normal cervical. (a) Numbers of DMRs showing the increase and decrease methylation level in LGSIL, HGSIL, SCC and CCA compared to normal cervical tissue. (b) Primary Component of Analysis of methylation ratio using all differentially methylated CpG sites among all cervical samples. (c) UpsetR plot showing the consistent and specific Hyper-DMC (upper) and Hypo-DMC (bottom) from each patient group compared to the normal cohort. (d) UCSC genome browser view of 5mC enrichment at HIST1H4F gene in normal cervical, LGSIL, HGSIL, SCC and CCA samples.

Fig. 3. Distinct DNA methylation changes characterize cervical carcinogenesis process. (a) Numbers of DMR showing the increase and decrease methylation level in LGSIL, HGSIL, SCC and CCA compared to normal cervical tissue. (b) The genomic distribution of DMRs between health person and LGSIL, HGSIL, SCC and CCA patients. Y-axis represents the percentage of DMR numbers overlapping corresponding genomic features across the total DMR in the whole genome. (c) Schematic depicting the cytosine modification dynamic changes during the

carcinogenesis process. LGSIL-to-CC representing consistent Hyper/Hypo-DMR in LGSIL, HGSIL, SCC and CCA. HGSIL-to-CC representing consistent Hyper/Hypo-DMR in HGSIL, SCC and CCA. CC specific representing specific Hyper/Hypo-DMR in SCC and CCA. LGSIL/HGSIL specific representing specific Hyper/Hypo-DMR in LGSIL or HGSIL respectively. LGSIL-to-HGSIL representing consistent Hyper/Hypo-DMR in LGSIL and HGSIL. Non-HGSIL representing consistent Hyper/Hypo-DMR in LGSIL, SCC and CCA. (d) Heatmap showing the cytosine modification dynamic changes in each cluster across all samples. (e) GREAT enrichment analysis of Hyper-DMRs of LGSIL-to-CC. (f) GREAT enrichment analysis of Hypo-DMRs in LGSIL-to-CC.

Fig. 4. Methylation dynamics disruptes Transcription factor binding and target gene expression. (a) The methylation level of TF E2F6 in each condition was shown in the up-panel. The promoter region was highlighed and the quantitative value was shown in the right barplot in down-panel. The E2F6 gene expression in normal, low grade, high grade and cancer were displayed in the boxplot. (b) The scatter plot shows the methylation level of DMCs on MBD2 peaks in normal (x-axis) and (pre-) cancer samples. The color represents the density of DMCs and the percentage of Hyper/Hypo-DMCs were labeled. (c) The methylation distribution around MBD2 peaks (+/-1kbp) were shown in different condition. The ‘start’ and ‘end’ represent the scaled TF peak boundarys. (d) Differential expressed MBD2 target genes were shown in volcano plot. The x-axis indicates log2 fold-change and y-axis is -log10 p-value from DESeq2 method. Red and blue dots were used to denote up-regulated and down-regulated targets. (e) The top 20 enriched gene ontology terms were shown in barplot. Color represents the significance of enrichment analysis. (f) MBD2 target genes were labeled in ‘Wnt signaling pathway’. The blue

549 and red boxes indicate the target genes that were down-regulated and up-regulated in cervical
550 cancer.

551 **Fig. 5. HPV infection causes methylation change in cervical cancer.** (a) Clinical HPV types
552 and the alignment reads ratio onto different HPV types in each sample. (b) Heatmap showing the
553 hierarchical clustering of samples infected by HPVs across signature DMCs caused by HPV16,
554 HPV18 and HPV31, respectively. Each row represents one DMC, and each column represents one
555 sample. The color scale from blue to red represents normalized DNA methylation level from low
556 to high. (c) Transcription factor (TF) motif enrichment analysis of signature Hyper- and Hypo-
557 DMCs across HPV16, HPV18 and HPV31, respectively. (d) The Profile of COUP-TFII TF across
558 upstream and downstream 1 kb of signature Hyper-DMCs across HPV16, HPV18 and HPV31,
559 respectively. (e) The Profile of CEBPB TF across upstream and downstream 1 kb of signature
560 Hypo-DMCs across HPV16, HPV18 and HPV31, respectively.

561 **Fig. S1.** (a) Consistent mean methylation ratio (solid line) and the gradual drop of CpG site
562 numbers (dotted line) under series sequencing depths were found as expected in each sample. (b)
563 All samples exhibit the classic frequency of CpG methylation ratio of human tissue. (c) Heatmap
564 representing the correlation of samples under common CpG methylation ratio. (d) Violin plot
565 showing the distribution of global DNA methylation levels of samples from five cohorts. (e) DNA
566 methylation patterns across genes of samples in each group. (f) Venn plot showing the comparison
567 of detected CpG sites between previously published Illumina HM450/850 array data and our
568 sequencing data. Boxplot showing the correlation of DNA methylation levels between samples in
569 this study and samples with public 450k array data from normal tissue, SCC and CCA, respectively.

Fig. S2. Violin plot showing the methylation ratio (quantified as mean mCG/CG) in various gene structures. The thickness of the bar representing the density of methylation ratio.

Fig. S3. (a) PCA analysis using methylation signals on detected DMCs, including our WGBS data and published Illumina Human-Methylation450 BeadChip data. (b) The scatter plot showing the consistent methylation changes between our sequencing data and published Illumina Human-Methylation450 array data. (c) Percentage of hyperDMC (red bar) and hypoDMC (blue bar) across detected CpGs in genomic features. Horizontal dotted red/blue line showing the percentage of total Hyper/Hypo-DMC across detected CpGs in the whole genome, respectively. (d) UpsetR plot showing the highly consistent methylation changes in LGSIL, HGSIL, and cervical tumor compared to the normal sample.

Fig. S4. (a) Clustering of Hyper- and Hypo-DMRs. (b) Motif enrichment of Hyper- and Hypo-DMRs across clustering, respectively. (c) Dot plots depicting the log2 fold change (CCA over Ctr) of gene expression (y-axis) at DMR regions that were hypermethylated (red) or hypomethylated (blue) in CCA samples compared with normal cervix. Values indicate the fraction of upregulated or downregulated genes in Hyper- and Hypo-DMRs, respectively. (d) Dot plots showing dynamic alternations in DNA hydromethylation within Hyper- (Up) and Hypo-DMRs (Bottom) between normal cervix and SCC patients. X-axis showed 5hmC levels at the normal cervix and the Y-axis showed the 5hmC levels at the I-to-IIa stage SCC (Left) and IIb-to-IV stage SCC (Right). The red dots represent reduced DNA hydromethylation level in Hyper-DMRs in SCC patients. The blue dots represent induced DNA hydromethylation level in Hypo-DMRs in SCC patients. (e) UCSC genome browser view of 5mC enrichment at *IRF3* gene in normal cervical, LGSIL, HGSIL, SCC and CCA samples.

Fig. S5. (a) The TFs were sorted by the percentage of DMCs in TF peaks, and top 100 TFs were shown for each compared group. The height of yellow bar represents the percentage of DMC in peaks and the height of red and blue bars indicate the percentage of hypo and hyper DMCs. The label for each bar under the figure is composed of cistrome file ID and TF name. (b) MBD2 target genes were labeled in 'cAMP signaling pathway'. The blue and red boxes indicate the target genes that were down-regulated and up-regulated in cervical cancer. (c) The scatter plot shows the methylation level of DMCs on STAT3 peaks in normal (x-axis) and (pre-) cancer samples. The color represents the density of DMCs and the percentage of Hyper/Hypo-DMCs were labeled. (d) Differential expressed STAT3 target genes were shown in volcano plot. The x-axis indicates log2 fold-change and y-axis is -Log10 p-value from DESeq2 method. Red and blue dots were used to denote up-regulated and down-regulated targets. (e) STAT3 target genes in were labeled 'HPV pathway'. The blue and red boxes indicate the target genes that were down-regulated and up-regulated in cervical cancer. (f) The methylation distribution around STAT3 peaks (+/-1kbp) were shown in different condition. The 'start' and 'end' represents the scaled TF peak boundarys. (g) STAT3 target genes in were labeled 'JAK-STAT signaling pathway'. The blue and red boxes indicate the target genes that were down-regulated and up-regulated in cervical cancer.

Tables

Table S1. Sample clinical information and WGBS data statistics.

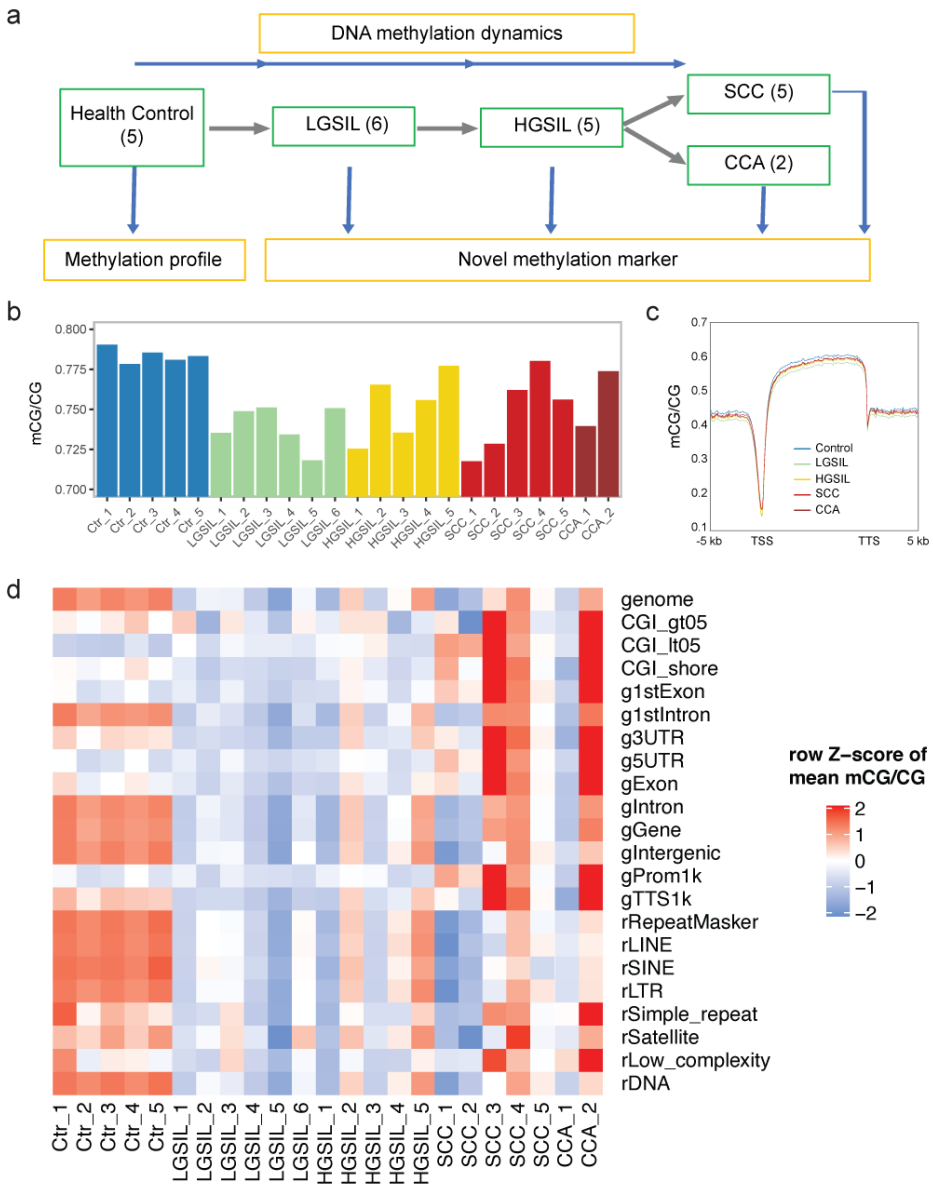

Fig. 1. DNA methylation profile of normal cervical tissues, pre-cancerous cervical lesions, and cervical cancer. (a) Schematic of the experimental design. Normal cervical tissue (*n* = 5), LGSIL (LGSIL, n = 6), HGSIL (HGSIL, n = 5), cervical cancers tissue including SCC (n = 5) and CCA (n = 2) were sampled in this study. (b) Global average DNA methylation levels during cervical carcinogenesis. Cytosine modification in normal cervical tissues is highly concordant than that in pathological tissues. (c) DNA methylation patterns across genes of five groups. Genes are methylated with a sharp drop of methylation at the promoter. (d) Normalized DNA methylation dynamics in genomic features across individuals.

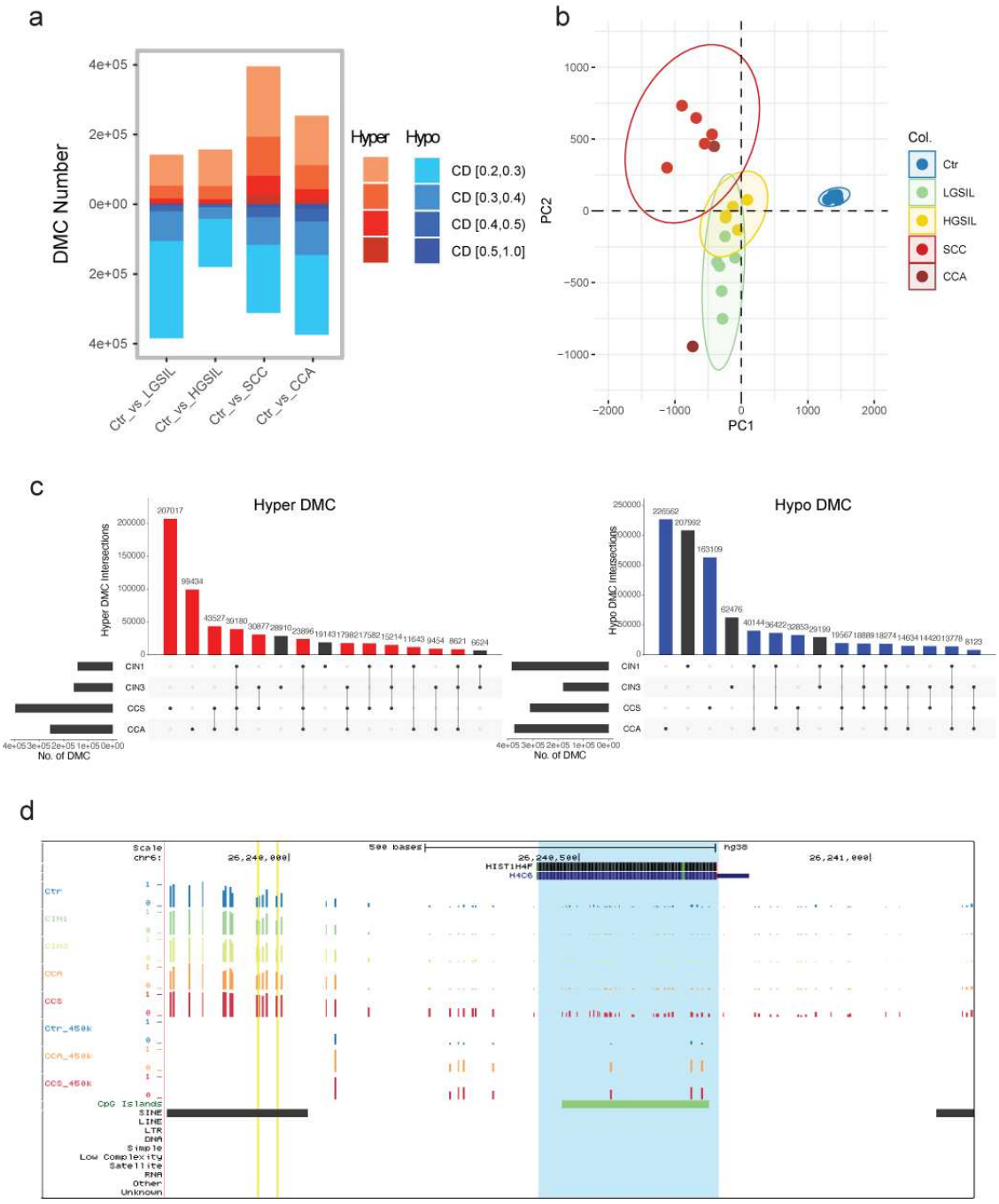

Fig. 2. Cytosine modification changes during cervical carcinogenesis compared to normal cervical. (a) Numbers of DMRs showing the increase and decrease methylation level in LGSIL, HGSIL, SCC and CCA compared to normal cervical tissue. (b) Primary Component of Analysis of methylation ratio using all differentially methylated CpG sites among all cervical samples. (c) UpsetR plot showing the consistent and specific Hyper-DMC (upper) and Hypo-DMC (bottom) from each patient group compared to the normal cohort. (d) UCSC genome browser view of 5mC enrichment at *HIST1H4F* gene in normal cervical, LGSIL, HGSIL, SCC and CCA samples.

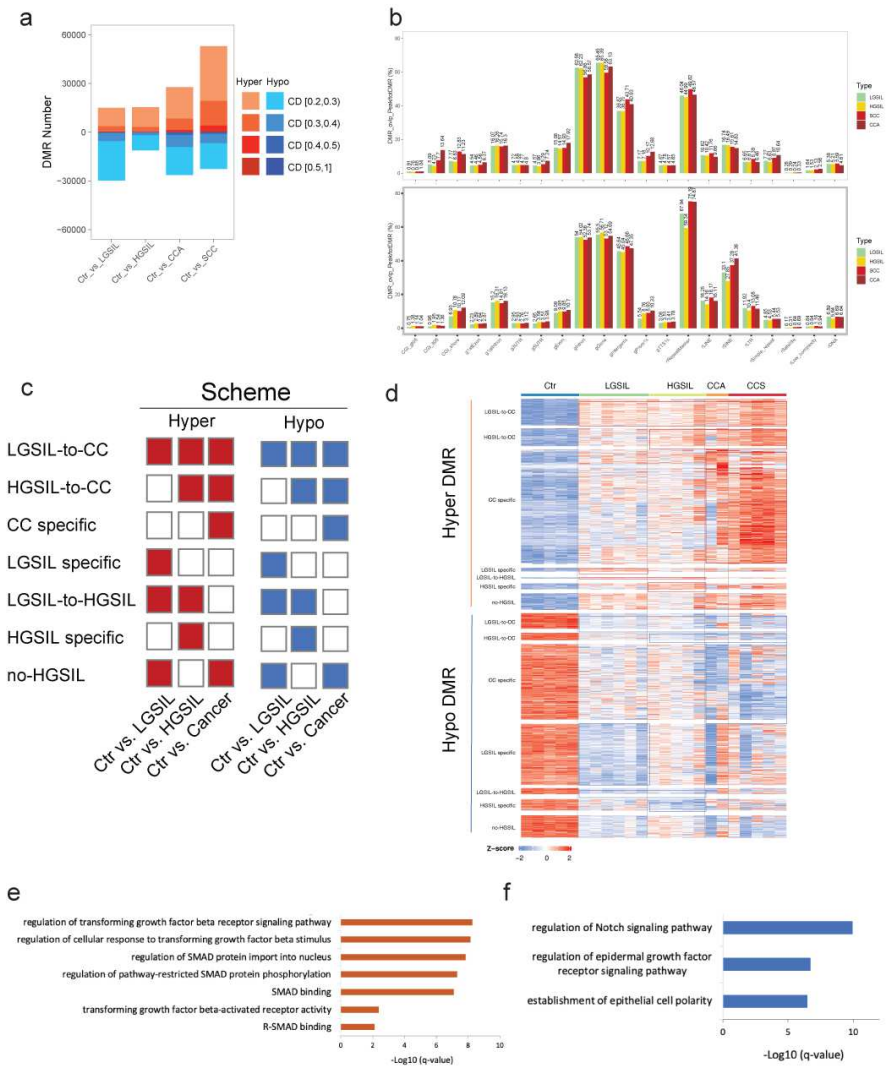

Fig. 3

Fig. 3. Distinct DNA methylation changes characterize cervical carcinogenesis process. (a) Numbers of DMR showing the increase and decrease methylation level in LGSIL, HGSIL, SCC and CCA compared to normal cervical tissue. (b) The genomic distribution of DMRs between health person and LGSIL, HGSIL, SCC and CCA patients. Y-axis represents the percentage of DMR numbers overlapping corresponding genomic features across the total DMR in the whole genome. (c) Schematic depicting the cytosine modification dynamic changes during the carcinogenesis process. LGSIL-to-CC representing consistent Hyper/Hypo-DMR in LGSIL, HGSIL, SCC and CCA. HGSIL-to-CC representing consistent Hyper/Hypo-DMR in HGSIL, SCC and CCA. CC specific representing specific Hyper/Hypo-DMR in SCC and CCA. LGSIL/HGSIL specific representing specific Hyper/Hypo-DMR in LGSIL or HGSIL respectively. LGSIL-to-HGSIL representing consistent Hyper/Hypo-DMR in LGSIL and HGSIL. Non-HGSIL representing consistent Hyper/Hypo-DMR in LGSIL, SCC and CCA. (d) Heatmap showing the cytosine modification dynamic changes in each cluster across all samples. (e) GREAT enrichment analysis of Hyper-DMRs of LGSIL-to-CC. (f) GREAT enrichment analysis of Hypo-DMRs in LGSIL-to-CC.

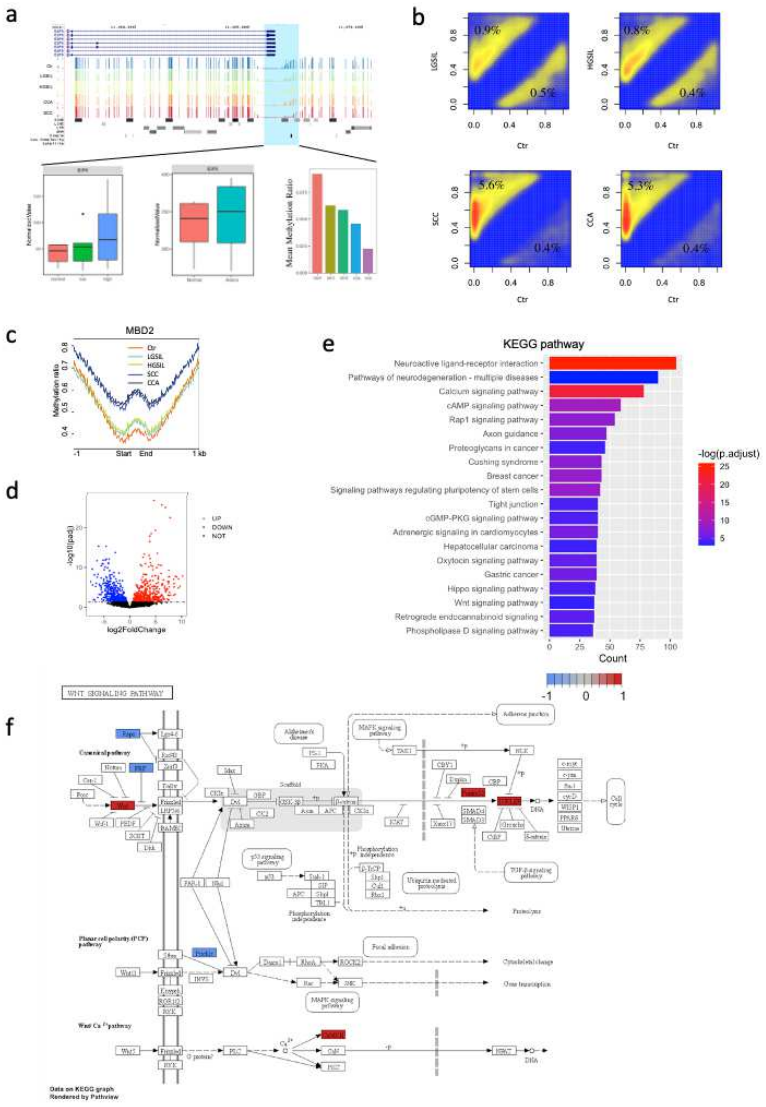

Fig. 4. Methylation dynamics disruptes Transcription factor binding and target gene expression. (a) The methylation level of TF E2F6 in each condition was shown in the up-panel. The promoter region was highlighed and the quantitative value was shown in the right barplot in down-panel. The E2F6 gene expression in normal, low grade, high grade and cancer were displayed in the boxplot. (b) The scatter plot shows the methylation level of DMCs on MBD2 peaks in normal (x-axis) and (pre-) cancer samples. The color represents the density of DMCs and the percentage of Hyper/Hypo-DMCs were labeled. (c) The methylation distribution around MBD2 peaks (+/-1kbp) were shown in different condition. The 'start' and 'end' represent the scaled TF peak boundarys. (d) Differential expressed MBD2 target genes were shown in volcano plot. The x-axis indicates log2 fold-change and y-axis is -log10 p-value from DESeq2 method. Red and blue dots were used to denote up-regulated and down-regulated targets. (e) The top 20 enriched gene ontology terms were shown in barplot. Color represents the significance of enrichment analysis. (f) MBD2 target genes were labeled in 'Wnt signaling pathway'. The blue and red boxes indicate the target genes that were down-regulated and up-regulated in cervical cancer.

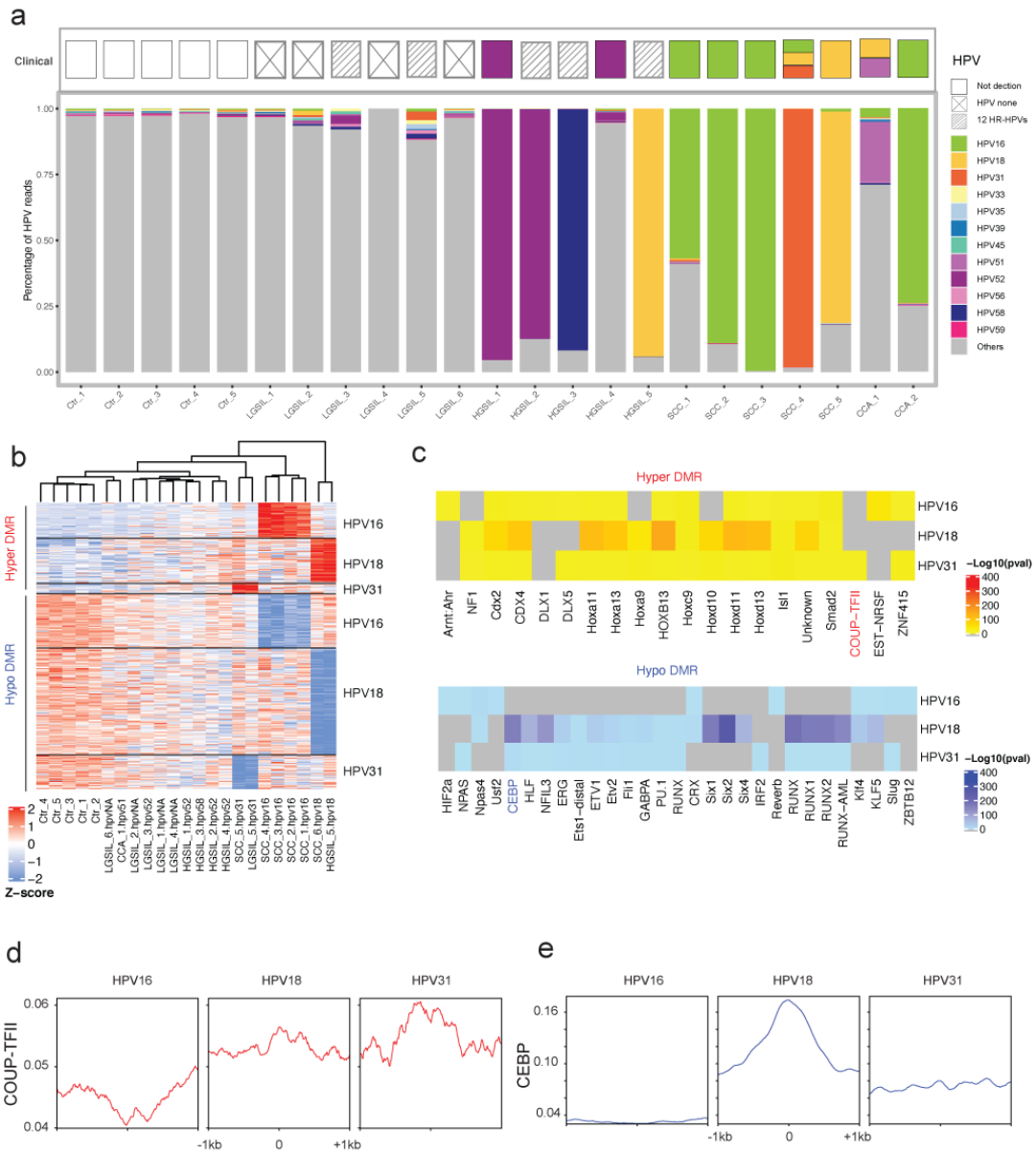

Fig. 5. HPV infection causes methylation change in cervical cancer. (a) Clinical HPV types and the alignment reads ratio onto different HPV types in each sample. (b) Heatmap showing the hierarchical clustering of samples infected by HPVs across signature DMCs caused by HPV16, HPV18 and HPV31, respectively. Each row represents one DMC, and each column represents one sample. The color scale from blue to red represents normalized DNA methylation level from low to high. (c) Transcription factor (TF) motif enrichment analysis of signature Hyper- and Hypo-DMCs across HPV16, HPV18 and HPV31, respectively. (d) The Profile of COUP-TFII TF across upstream and downstream 1 kb of signature Hyper-DMCs across HPV16, HPV18 and HPV31, respectively. (e) The Profile of CEBPB TF across upstream and downstream 1 kb of signature Hypo-DMCs across HPV16, HPV18 and HPV31, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfiles.pdf