Five candidate biomarkers associated with the diagnosis and prognosis of cervical cancer

Running title: Five biomarkers of cervical cancer

Hong-Yan Han1*,#, Jiang-Tao Mou1#, Wen-Ping Jiang1, Xiu-Ming Zhai1, Kun Deng1*

1 Department of Laboratory Medicine, The Third Hospital Affiliated to Chongqing Medical University.

*Correspondence to: Hong-Yan Han
Department of Laboratory Medicine, The Third Hospital Affiliated to Chongqing Medical University, 1st Shuanghu Branch Road, Yubei District, Chongqing 401120, China.
Email: hongyanhan2020@163.com
Tel: +86-18223578600

Kun Deng
Department of Laboratory Medicine, The Third Hospital Affiliated to Chongqing Medical University, 1st Shuanghu Branch Road, Yubei District, Chongqing 401120, China.
Email: dengkun@hospital.cqmu.edu.cn
Tel: +86-15923800009

# These two authors contributed equally to this work and should be considered co-first authors.
Abstract

**Purpose:** Cervical cancer (CC) is one of the most general gynecological malignancies and is associated with high morbidity and mortality. We aimed to select candidate genes related to the diagnosis and prognosis of CC.

**Methods:** The mRNA expression profile datasets were downloaded. We also downloaded RNA-sequencing gene expression data and related clinical materials from TCGA, which included 307 cervical cancer samples and 3 normal samples. Differentially expressed genes (DEGs) were obtained by R software. GO function analysis and KEGG pathway enrichment analysis of DEGs were performed in the DAVID dataset. Using machine learning, the optimal diagnostic mRNA biomarkers for CC were identified. We used qRT-PCR and HPA database to exhibit the differences in gene and protein levels of candidate genes.

**Results:** A total of 313 DEGs were screened from the microarray expression profile datasets. DNMT1, CHAF1B, CHAF1A, MCM2, CDKN2A were identified as optimal diagnostic mRNA biomarkers for CC. Additionally, the GEPIA database showed that the DNMT1, CHAF1B, CHAF1A, MCM2 and CDKN2A were associated with the poor survival of CC patients. HPA databases and qRT-PCR confirmed that these genes were highly expressed in CC tissues.

**Conclusion:** The present study identified five DEmRNAs, including DNMT1, CHAF1B, CHAF1A, MCM2 and KNTC1, as potential diagnostic and prognostic biomarkers of CC.

**Keywords:** cervical cancer, machine learning, differentially expressed genes, biomarkers
Introduction

Cervical cancer (CC) is as one of the most common gynecological malignant tumors worldwide and has become a prominent public health issue [1,2]. According to reports, the incidence of cervical cancer ranks second among female malignant tumors in the world, and the mortality rate ranks first among female malignant tumors of the reproductive system. It is a serious threat to women’s health [3]. Cervical cancer is difficult to diagnose at the early stage, leading to a delay for effective treatment[4]. At present, surgery, chemotherapy and radiotherapy are the most commonly used treatment methods for cervical cancer; however, due to the resistance of cervical cancer cells to therapeutic drugs, chemotherapy drugs are relatively ineffective in treating cervical cancer[5,6]. Therefore, it is of great significance to develop new diagnostic or treatment methods for cervical cancer.

Gene Expression Integration (GEO) was launched in response to the growing demand for public repositories of high-throughput gene expression data [7]. The Cancer Genome Atlas (TCGA) project analyzes DNA copy number, mRNA expression, promoter methylation, microRNA expression of 307 cervical SCCs, and 3 normal cervical tissue samples[8,9]. Therefore, combining the GEO and TCGA datasets may provide an important perspective for the study of new biomarkers. In fact, there are many reports that revealed a series of highly specific and sensitive markers by screening for tumor biomarkers based on GEO and TCGA data recently[10]. Compared with conventional screening methods, analyzing high-throughput data based on bioinformatics methods allows researchers to obtain stable and reliable biomarkers in a larger number of clinical samples.

In this study, we downloaded five original mRNA microarray datasets from the GEO database, GSE7410, GSE7803, GSE9750, GSE55940 and GSE63514, a total of 176 samples containing 68 normal cervical samples and 108 cervical cancer specimens. We also downloaded RNA-sequencing gene expression data and related clinical materials from TCGA, which included 307 cervical cancer samples and 3 normal samples. The differentially expressed genes (DEGs) in normal tissues and tumor
samples were obtained by R software. By evaluating the diagnostic and prognostic value of these DEGs, 5 DEGs associated with the diagnostic and prognosis of CC were screened out.

**Materials and methods**

**Microarray data**

Using the keywords “cervical cancer” to search on the GEO database, the raw gene expression profiles of GSE7410, GSE7803, GSE9750, GSE55940 and GSE63514 were downloaded. RNA-sequencing gene expression data and related clinical materials from TCGA were retrieved on March 24, 2020, and comprised data from 307 cervical cancer samples and 3 normal samples. The inclusion criteria were pathologically confirmed cervical cancer, complete RNA expression data from the patients [11].

**Identification of DEGs and functional enrichment analysis**

Data analysis was conducted using package limma in R language (version 3.4.0)[12]. FDR<0.01 and |Combined.ES|>1.5 were set up to screen DEGs. Heat maps were generated by heatmap package in R. Through the DAVID database [13], functional enrichment analysis was conducted to identify gene ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG). GO terms and KEGG pathways with P < 0.05 were selected.

**Identification of optimal diagnostic mRNA biomarkers for CC**

Modeling and machine learning algorithm were used to identify optimal diagnostic mRNA biomarkers for CC. The procedure were performed as previously described[14]. ROC curve was used to explore the differentially diagnostic capability of the five candidate genes.

**Identification of survival-related DEGs and establishment of the prognostic gene signature**

Using the survival and survminer package in R, Kaplan Meier plots and log-rank tests were performed to elucidate the relationship between 5-year overall survival (OS) rates and DEGs expression levels. DEGs with p<0.01 were considered statistically significant and included in subsequent analyses. For these prognostic genes, LASSO
Cox regression analysis was used to determine the gene signature prognostic model using the following formula to calculate the risk score for each sample.

\[ \text{riskScore} = \sum_{i=1}^{n} \exp(x_i) \beta_i \]

Patients with CC were divided into high-risk and low-risk groups based on the median risk score, and Kaplan-Meier (KM) survival analysis was performed between the low-risk and high-risk groups. ROC curve analysis to evaluate the predictive power of the forecasting model.

**Quantitative RT-PCR confirmation and validation in datasets**
According to the results of GEO integrated analysis, we selected six DEGs (DNMT1, CHAF1B, CHAF1A, MCM2, CDKN2A, KNTC1) in CC versus normal control as candidate genes. A total of 13 blood samples were collected from seven normal subjects, six patients diagnosed with CC. Informed written consent was obtained from all participants, and research protocols were approved by the Ethics Committee of our hospital.

To evaluate mRNA expression of candidate genes, we used GEO database and TCGA database to differentiate expression of candidate genes in CC tissues and normal tissues.

**Evaluation of Immunohistochemical Staining**

To validate the protein expression level of candidate genes in CC tissues, we used Human Protein Atlas (HPA, https://www.proteinatlas.org/) database to obtain immunohistochemical staining.

**Relationship between candidate genes and clinical features in HCC patients**
To further explore the relationship between candidate genes and tumor clinical features, we analyzed the TCGA clinical data using LinkedOmics (http://www.linkedomics.org/) database.

**Results**

**Identification of DEGs in cervical cancer**

After retrieving, we obtained five microarray datasets of mRNA according to the inclusion criteria from the GEO database. The characteristics of the individual database for the integrated analysis are displayed in Table 1. A total of 313
differentially expressed genes (DEGs) were obtained, including 200 up-regulated genes and 113 down-regulated genes. The cluster heatmap of top 100 DEGs in all samples are shown in Figure 1.

**GO and KEGG pathway analysis of DEGs**

GO analysis showed that DEGs were most enriched in cell cycle, chromosome, specific DNA binding, and single-stranded DNA binding. GO functional enrichment of DEGs with a p-value <0.05 was considered statistically significant and the results are presented in Figure 2A. KEGG pathway analysis showed that integrated DEGs were mainly enriched in four pathways consisted by cell cycle, DNA replication, p53 signaling pathway, and Mismatch repair (Fig. 2B).

**Identification of the optimal diagnostic mRNA biomarkers for CC**

Based on the reduced dimension of the data, comparing CC and normal tissues identified 20 DEGs using LASSO algorithm analysis (Table 2). The random forest analysis was used to rank the 20 DEGs, according to the decrease in mean accuracy (Fig. 3A). A 10-fold cross-validation result demonstrated that the average accuracy rate of 10 DEGs, including DNMT1, CHAF1B, CHAF1A, MCM2, CDKN2A, KNTC1, CRISP2, KRT32, SLC5A1 and CRNN exhibited the highest score (Fig. 3B). Therefore, these ten DEGs were selected as the potential optimal diagnostic mRNA biomarkers for CC and were used to establish the random forests, decision tree and SVM models.

The AUC of the random forests model was 0.996 and the specificity and sensitivity of this model were 0.941 and 0.991, respectively (Fig. 4A). The AUC of the decision tree model was 0.931 and the specificity and sensitivity of this model were 0.897 and 0.972, respectively (Fig. 4B). The AUC of the SVM model was 0.995, and the specificity and sensitivity of this model were 0.956 and 0.981 (Fig. 4C). The AUC curves of these ten genes were shown in Fig. 5, indicating strong diagnostic ability for CC.

**Identification of five DEGs associated with OS and establishment of the four-gene prognostic signature**

In Figure 6, to identify the mRNAs which would be potentially associated with
overall survival of cervical cancer patients, we evaluated the association between mRNAs expression and patients' survival using Kaplan-Meier curve and Log-rank test. The results showed that these five mRNAs (DNMT1, CHAF1B, CHAF1A, KNTC1, MCM2) were negatively correlated with overall survival (OS). Five genes were subsequently used to construct a prognostic gene-signature. Finally, 4-gene signature prognostic model was established. Time-dependent ROC and Kaplan-Meier curve were used to assess the prognostic capacity of the four-gene signature. The AUCs (Area under the ROC curve) for 1-year, 3-year, and 5-year OS were 0.617, 0.665, 0.690. Patients in the high-risk group shown significantly poorer OS than patients in the low-risk group (all \( p < 0.001 \)) (Fig. 7A–D)

**Correlation between candidate genes and clinical features in CC patients**

Downloading the TCGA clinical data in LinkedOmics online tool, we analyzed the relationship between selected genes and clinical features in CC patients. The CHAF1B and KNTC1 in CC patients were significantly correlated with tumor purity. The CHAF1B was also significantly correlated with radiation therapy and MCM2 was significantly correlated with histological type (Table 3).

**Quantitative RT-PCR confirmation and validation in datasets**

To indicate the results of integrated analysis, we selected five mRNAs (DNMT1, CHAF1B, CHAF1A, MCM2, KNTC1 and CDKN2A) in CC versus normal control. Based on the results of quantitative RT-PCR, the expression of DNMT1, CHAF1B, CHAF1A, MCM2, KNTC1 and CDKN2A were all up-regulated, which were consistent with the results of our integrated analysis (Fig. 8).

Using the GEO and TCGA data, we analyzed the expression of the five selected up-regulated genes in CC tissues and normal tissues. The results showed that the DNMT1, CHAF1B, CHAF1A, MCM2 and KNTC1 were highly expressed in CC tissues, and the differences were statistically significant (Fig. 9). The immunohistochemical staining is displayed in Figure 10.

**Discussion**

Cervical cancer is one of the most general malignant tumors in gynecology, and it is also one of the main causes of female cancer deaths [15]. Statistically, 80 percent of
patients develop aggressive cancer once diagnosed and the age of diagnosed patients is tardily decreasing[16]. Furthermore, on account of metastasis and recurrence, the incidence and mortality of CC remain high [17]. At the same time, reliable and specific biomarkers for the diagnosis and prognosis of cervical cancer are scarce and lack of exploration. Thus, it is urgent to find diagnostic biomarkers and new therapeutic targets to predict the survival of CC. Gene expression microarrays have been broadly used in the study of genes related to tumors, offering broad prospects for molecular therapy of drugs[18]. A battery of biomarkers have been suggested as potential targets for the diagnosis and prognosis of CC[19,20].

DNA methyltransferase 1 (DNMT1) is a member of the DNA methyltransferase family, responsible for the DNA methylation of cytosine-phosphoguanine (CpG) island upstream of tumor suppressor genes [21,22]. DNMT1 is the most important one. Abnormal expression of DNMT1 can lead to abnormal methylation of some tumor suppressor genes CpG islands, which in turn leads to inactivation of tumor suppressor genes and cell carcinogenesis[23]. Highly expressed DNMT1 is not only detected in a variety of tumor cells, but also appears before DNA methylation[24]. Previous studies have shown that DNMT1 is an essential substance for the maintenance of cancer stem cells (CSCs) in various cancers such as prostate cancer, pancreatic cancer and breast cancer[25-27]. For example, DNMT1 induces histone demethylation of H3K9me3 and H3K27me3 on the promoters of Zeb2 and KLF4 in prostate cancer cells[26]. More importantly, a recent study showed that DNMT1 is up-regulated in breast tumors. DNMT1 inhibition or DNMT1 induced Islet-1 (ISL1) hypermethylation/downregulation limits the number of CSCs in breast cancer cells[25]. In our study, DNMT1 was up-regulated in CC versus normal control, and may be as an oncogene in development of cervical cancer. Furthermore, DNMT1 was significantly associated with overall survival time and one of the diagnostic markers. Thus, we speculated that the expression level of DNMT1 has both diagnostic and prognostic values.

MCM2 is a component of the DNA replication licensing complex (MCM2-7) that has been found to mainly localize to the nucleus in eukaryotic cells. MCM2 has been
recognised as a useful marker in screening for cervical carcinoma oral squamous cell carcinoma and medulloblastoma. Overexpression of MCM2 occurs frequently in CC, especially in cases with persistent high-risk HPV infection[28]. Many studies mainly focus on the analysis of biomarkers related to precancerous lesions, but only a few studies have confirmed the prognostic influence of MCM2 expression in the progression of aggressive cervical cancer[29,30]. Jun Wang et al. found that MCM2 was a prognostic biomarker in cervical cancer[31]. We also observed high expression levels of MCM2 in CC tissues. Aihemaiti et al. reported that cytoplasmic rather than nuclear accumulation of MCM2 is related to improved survival for patients with ovarian clear cell carcinoma, which maybe associated with MCM2-mediated DNA damage-induced apoptosis[32]. It was therefore hypothesized in the present study that MCM2 may serve an important role in CC. In addition, MCM2 was also correlated with diagnosis and prognosis for CC.

Chromatin Assembly Factor 1, subunit A(CHAF1A) which is also called P150 is the subunit of Chromatin Assembly Factor-1(CAF-1) [33-37]. CHAF1A enhances Gfi1-mediated transcriptional repression and occupies Gfi1 target gene promoters in transfected cells[38]. Recently, CHAF1A has been associated with the development and progression of solid tumors, including breast cancer, prostate squamous cell carcinoma, hepatocellular carcinoma, glioma and neuroblastoma[39-43]. A number of studies have shown that CHAF1A is also highly expressed in breast cancer, colon cancer, cervical cancer and other tumors, and can be used as a potential marker for judging the prognosis of tumor patients and a target for tumor treatment[44]. However, the expression and biological function of CHAF1A in CC remains largely unidentified.

Chromatin assembly factor 1 subunit B (CHAF1B), the p60 subunit of CAF-1, plays a vital role in DNA replication and chromatin assembly in proliferating tissues [45]. Tumor cells are usually characterized by a high proliferation rate. Therefore, it is speculated that CHAF1B plays an important role in the pathogenesis of malignant tumors. Researches have shown that up-regulated CHAF1B is significantly correlated with poor outcomes and that CHAF1B has potential in predicting the prognosis in
several cancers, including cervical cancer[46-48]. These reported results were consistent with our study.

Kinetochore-related protein 1 (KNTC1) gene is distributed in the cytoplasm, nucleus, chromosomal centromere, centromere, cytoskeleton, and in the spindle fiber. KNTC1 has been studied in a variety of human malignancies and is related to the pathological grade of tumor tissues[49]. As previously described, the expression of KNTC1 in three esophageal squamous cell carcinoma (ESCC) cell lines and established that all the tested cell lines showed that it is positively expressed[50]. Previous studies have shown that the KNTC1 transcriptional activity changes with the size of tumor in patients with oral squamous cell carcinoma[51]. Also, compared with the healthy control group, the KNTC1 expression in neuroblastoma samples increased statistically. Several bioinformatics studies have also established that the KNTC1 gene could be one of the vital genes associated with cancer development, including Hepatocellular carcinoma (HCC), Pancreatic cancer (PC), and nasopharyngeal carcinoma (NPC)[52]. Huan Chen et al. reported that KNTC1 gene is closely related to the poor prognosis of CC, which was consistent with our findings[49].

In conclusion, we obtained several DEGs in CC and found that overexpression of DNMT1, CHAF1B, CHAF1A, MCM2, KNTC1 in tumor tissues predicted poor survival in CC. These DEGs also have diagnostic value for CC at early stage. We hypothesized that DNMT1, CHAF1B, CHAF1A, MCM2 and KNTC1 may be potential therapeutic targets for CC. We analyzed these genes at the transcriptional and protein levels, validated by qRT-PCR and immunohistochemical staining.

**Declarations**

**Funding**

This study have no funding support.

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the
corresponding author on reasonable request.

**Code availability**

Not applicable.

**Authors' contributions**

HH and KD contributed to the conception of the study. JM performed the data analyses. WJ and XZ contributed significantly in writing the manuscript. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

The present study has been approved by the Third Hospital Affiliated to Chongqing Medical University.

**Consent for publication**

The consent for publication was obtained from all patients.
**References**

1. Marth C, Landoni F, Mahner S, McCormack M, Gonzalez-Martin A, Colombo N (2017) Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology 28 (suppl_4):iv72-iv83. doi:10.1093/annonc/mdx220

2. Luan X, Wang Y (2018) LncRNA XLOC_006390 facilitates cervical cancer tumorigenesis and metastasis as a ceRNA against miR-331-3p and miR-338-3p. 29 (6):e95. doi:10.3802/jgo.2018.29.e95

3. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, Dandona L, Fleming T, Forouzanfar MH, Hancock J, Hay RJ, Hunter-Merrill R, Huynh C, Hosgood HD, Jonas JB, Khubchandani J, Kumar GA, Kutz M, Lan Q, Larson HJ, Liang X, Lim SS, Lopez AD, MacIntyre MF, Marczak L, Marquez N, Mokdad AH, Pinho C, Pourmalek F, Salomon JA, Sanabria JR, Sandar L, Sartorius B, Schwartz SM, Shackelford KA, Shibuya K, Stanaway J, Steiner C, Sun J, Takahashi K, Vollset SE, Vos T, Wagner JA, Wang H, Westerman R, Zeeb H, Zeeckler L, Abd-Allah F, Ahmed MB, Alabed S, Alam NK, Aladdahri SF, Alem G, Alemayohu MA, Ali R, Al-Raddadi R, Amare A, Amoako Y, Artaman A, Asayesh H, Atnafu N, Awasthi A, Saleem HB, Barac A, Bedi N, Bensenor I, Berhane A, Bernabé E, Betsu B, Binagwaho A, Boneya D, Campos-Nonato I, Castañeda-Orjuela C, Catalá-López F, Chiap P, Chibueze C, Chitheer A, Chio JY, Cowie B, Damtew S, Das Neves J, Dey S, Dharmaratne S, Dhillon P, Ding E, Driscoll T, Ekwueme D, Endries AY, Farvid M, Farzadfar F, Fernandes J, Fischer F, TT GH, Gebru A, Gopalan S, Hailu A, Horino M, Horita N, Hussein A, Huybrechts I, Inoue M, Islami F, Jakovljevic M, James S, Javanbakht M, Jee SH, Kasaean J, Khader YS, Khang YH, Kim D, Leigh J, Linn S, Lunevicius R, El Razek HMA, Malekzadeh R, Malta DC, Marcenes W, Markos D, Melaku YA, Meles KG, Mendoza W, Mengiste DT, Meretoja T, Miller TR, Mohammad KA, Mohammad A, Mohammed S, Moradi-Lakeh M, Nagel G, Nand D, Le Nguyen Q, Nolte S, Ogbo FA, Oladimeji KE, Oren E, Pa M, Park EK, Pereira DM, Plass D, Qorbani M, Rahnheim M, Rana SM, Søreide K, Satpathy M, Sawhney M, Sepanlou SG, Shaikh MA, Shi J, Shiue I, Shore HR, Shrimie MG, Soreji S, Stathopoulou V, Stroupoulis K, Sufiyand MB, Sykes BL, Tabarés-Seisdedos R, Tadesse F, Tedla BA, Tessema GA, Thakur JS, Tran BX, Ukwaja KN, Uzochukwu BSC, Vlassov VV, Weiderpass E, Wubs Terefe M, Yebo HG, Yimam HH, Yonemoto N, Younis MZ, Zaidi Z, Zenebe ZM, Murray CJL, Naghavi M (2017) Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. JAMA oncology 3 (4):524-548. doi:10.1001/jamaoncol.2016.5688

4. Noordhuis MG, Fehrmann RS, Wisman GB, Nijhuis ER, van Zanden JJ, Moerland PD, Ver Loeren van Themaat E, Volders HH, Kok M, ten Hoor KA, Hollema H, de Vries EG, de Bock GH, van der Zee AG, Schuuring E (2011) Involvement of the TGF-beta and beta-catenin pathways in pelvic lymph node metastasis in early-stage cervical cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 17 (6):1317-1330. doi:10.1158/1078-0432.ccr-10-2320

5. Rischin D, Narayan K, Oza AM, Milesklin L, Bernshaw D, Choi J, Hicks R, McClure B, Fyles A (2010) Phase 1 study of tirapazamine in combination with radiation and weekly cisplatin in patients with locally advanced cervical cancer. International journal of gynecological cancer : official journal of the International Gynecological Cancer Society 20 (5):827-833. doi:10.1111/IGC.0b013e3181dc8d7e

6. Liu M, Jia J, Wang X, Liu Y, Wang C, Fan R (2018) Long non-coding RNA HOTAIR promotes cervical
cancer progression through regulating BCL2 via targeting miR-143-3p. Cancer biology & therapy 19 (5):391-399. doi:10.1080/15384047.2018.1423921

7. Salvatici M, Achilarre MT, Sandri MT, Boveri S, Vanna Z, Landoni F (2016) Squamous cell carcinoma antigen (SCC-Ag) during follow-up of cervical cancer patients: Role in the early diagnosis of recurrence. Gynecologic oncology 142 (1):115-119. doi:10.1016/j.ygyno.2016.04.029

8. Integrated genomic analyses of ovarian carcinoma (2011). Nature 474 (7353):609-615. doi:10.1038/nature10166

9. Zhu Y, Mohamed ASR, Lai SY, Yang S, Kanwar A, Wei L, Kamal M, Sengupta S, Elhalawani H, Skinner H, Mackin DS, Shiao J, Messer J, Wong A, Ding Y, Zhang L, Court L, Ji Y, Fuller CD (2019) Imaging-Genomic Study of Head and Neck Squamous Cell Carcinoma: Associations Between Radiomic Phenotypes and Genomic Mechanisms via Integration of The Cancer Genome Atlas and The Cancer Imaging Archive. JCO clinical cancer informatics 3:1-9. doi:10.1200/cci.18.00073

10. Beuselinck B, Job S, Becht E, Karadimou A, Verkarre V, Couchy G, Giraldo N, Rioux-Leclercq N, Molinié V, Sibony M, Elaidi R, Teghem C, Patard JJ, Ménard A, Fridman WH, Sautès-Fridman C, de Reyniès A, Oudard S, Zucman-Rossi J (2015) Molecular subtypes of clear cell renal cell carcinoma are associated with sunitinib response in the metastatic setting. Clinical cancer research: an official journal of the American Association for Cancer Research 21 (6):1329-1339. doi:10.1158/1078-0432.ccr-14-1128

11. Arbyn M, Weiderpass E, Brun L, de Sanjosé S, Saraiya M, Ferlay J, Bray F (2020) Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. The Lancet Global health 8 (2):e191-e203. doi:10.1016/s2214-109x(19)30482-6

12. Ritchie ME, Shipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43 (7):e47. doi:10.1093/nar/gkv007

13. Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics (Oxford, England) 28 (13):1805-1806. doi:10.1093/bioinformatics/bts251

14. Yang S, Wang Y, Ren J, Zhou X, Cai K, Guo L, Wu S (2020) Identification of diagnostic and prognostic lncRNA biomarkers in oral squamous carcinoma by integrated analysis and machine learning. Cancer biomarkers : section A of Disease markers 29 (2):265-275. doi:10.3233/cbm-191215

15. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A (2015) Global cancer statistics, 2012. CA: a cancer journal for clinicians 65 (2):87-108. doi:10.3322/caac.21262

16. Dai F, Chen G, Wang Y, Zhang L, Long Y, Yuan M, Yang D, Liu S, Cheng Y, Zhang L (2019) Identification of candidate biomarkers correlated with the diagnosis and prognosis of cervical cancer via integrated bioinformatics analysis. OncoTargets and therapy 12:4517-4532. doi:10.2147/ott.s199615

17. Wang IT, Chou SC, Lin YC (2014) Zoledronic acid induces apoptosis and autophagy in cervical cancer cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 35 (12):11913-11920. doi:10.1007/s13277-014-2460-5

18. Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, Hastings E, Huber W, Jupp S, Keays M, Kryvych N, McMurry J, Marioni JC, Malone J, Megy K, Rustici G, Tang AY, Taubert J, Williams E, Mannion O, Parkinson HE, Brazma A (2014) Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. Nucleic acids research 42 (Database issue):D926-932. doi:10.1093/nar/gkt1270
19. Nahand JS, Taghizadeh-Boroujeni S, Karimzadeh M, Borran S, Pourhanifeh MH, Moghoofei M, Bokharaei-Salim F, Karampoor S, Jafari A, Asemi Z, Tbibzadeh A, Namdar A, Mirzaei H (2019) microRNAs: New prognostic, diagnostic, and therapeutic biomarkers in cervical cancer. 234 (10):17064-17099. doi:10.1002/jcp.28457

20. Otter SJ, Chatterjee J, Stewart AJ, Michael A (2019) The Role of Biomarkers for the Prediction of Response to Checkpoint Immunotherapy and the Rationale for the Use of Checkpoint Immunotherapy in Cervical Cancer. Clinical oncology (Royal College of Radiologists (Great Britain)) 31 (12):834-843. doi:10.1016/j.clon.2019.07.003

21. Yong WS, Hsu FM, Chen PY (2016) Profiling genome-wide DNA methylation. Epigenetics & chromatin 9:26. doi:10.1186/s13072-016-0075-3

22. Samowitz WS, Ogino S (2008) DNA methylation in breast and colorectal cancers. Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc 21 (8):1054; author reply 1054-1055. doi:10.1038/modpathol.2008.18

23. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics 13 (7):484-492. doi:10.1038/nrg3230

24. Wang JT, Ding L, Jiang SW, Hao J, Zhao WM, Zhou Q, Yang ZK, Zhang L (2014) Folate deficiency and aberrant expression of DNA methyltransferase 1 were associated with cervical cancerization. Current pharmaceutical design 20 (11):1639-1646. doi:10.2174/13816128113199990543

25. Pathania R, Ramachandran S, Elangovan S, Padia R, Yang P, Cinghu S, Veeranan-Karmegam R, Arjunan P, Gnana-Prakasam JP, Sadanand F, Pei L, Chang CS, Choi JH, Shi H, Manicassamy S, Prasad PD, Sharma S, Ganapathy V, Jothi R, Thangaraju M (2015) DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. Nature communications 6:6910. doi:10.1038/ncomms7910

26. Lee E, Wang J, Yumoto K, Jung Y, Cackowski FC, Decker AM, Li Y, Franceschi RT, Pienta KJ, Taichman RS (2016) DNMT1 Regulates Epithelial-Mesenchymal Transition and Cancer Stem Cells, Which Promotes Prostate Cancer Metastasis. Neoplasia (New York, NY) 18 (9):553-566. doi:10.1016/j.neo.2016.07.007

27. Zagorac S, Alcala S, Fernandez Bayon G, Bou Kheir T, Schoenhals M, Gonzalez-Neira A, Fernandez Fraga M, Aicher A, Heeschen C, Sainz B, Jr. (2016) DNMT1 Inhibition Reprograms Pancreatic Cancer Stem Cells via Upregulation of the miR-17-92 Cluster. Cancer research 76 (15):4546-4558. doi:10.1158/0008-5472.can-15-3268

28. Zheng J (2015) Diagnostic value of MCM2 immunocytochemical staining in cervical lesions and its relationship with HPV infection. International journal of clinical and experimental pathology 8 (1):875-880

29. Amaro Filho SM, Nuovo GJ, Cunha CB, Ramos Pereira Lde O, Oliveira-Silva M, Russomano F, Pires A, Nicol AF (2014) Correlation of MCM2 detection with stage and virology of cervical cancer. The International journal of biological markers 29 (4):e363-371. doi:10.5301/jbm.5000081

30. Giaginis C, Vgenopoulou S, Vielh P, Theocharis S (2010) MCM proteins as diagnostic and prognostic tumor markers in the clinical setting. Histology and histopathology 25 (3):351-370. doi:10.14670/hh-25.351

31. Wang J, Zheng H, Han Y, Wang G, Li Y (2020) A Novel Four-Gene Prognostic Signature as a Risk Biomarker in Cervical Cancer. International journal of genomics 2020:4535820. doi:10.1155/2020/4535820

32. Aihemaiti G, Kurata M, Nogawa D, Yamamoto A, Mineo T, Onishi I, Kinowaki Y, Jin XH, Tatsuzawa A, Miyasaka N, Kitagawa M, Yamamoto K (2018) Subcellular localization of MCM2 correlates with the
prognosis of ovarian clear cell carcinoma. Oncotarget 9 (46):28213-28225. doi:10.18632/oncotarget.25613

33. Verreault A, Kaufman PD, Kobayashi R, Stillman B (1996) Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell 87 (1):95-104. doi:10.1016/s0092-8674(00)81326-4

34. Takami Y, Ono T, Fukagawa T, Shibahara K, Nakayama T (2007) Essential role of chromatin assembly factor-1-mediated rapid nucleosome assembly for DNA replication and cell division in vertebrate cells. Molecular biology of the cell 18 (1):129-141. doi:10.1091/mbc.e06-05-0426

35. Kadyrova LY, Blanko ER, Kadyrov FA (2011) CAF-1-dependent control of degradation of the discontinuous strands during mismatch repair. Proceedings of the National Academy of Sciences of the United States of America 108 (7):2753-2758. doi:10.1073/pnas.1015914108

36. Doyen CM, Moshkin YM, Chalkley GE, Bezstarosti K, Demmers JA, Rathke C, Renkawitz-Pohl R, Verrijzer CP (2013) Subunits of the histone chaperone CAF1 also mediate assembly of protamine-based chromatin. Cell reports 4 (1):59-65. doi:10.1016/j.celrep.2013.06.002

37. Smith CL, Matheson TD, Trombly DJ, Sun X, Campeau E, Han X, Yates JR, 3rd, Kaufman PD (2014) A separable domain of the p150 subunit of human chromatin assembly factor-1 promotes protein and chromosome associations with nucleoli. Molecular biology of the cell 25 (18):2866-2881. doi:10.1091/mbc.E14-05-1029

38. Wu Z, Cui F, Yu F, Peng X, Jiang T, Chen D, Lu S, Tang H, Peng Z (2014) Up-regulation of CHAF1A, a poor prognostic factor, facilitates cell proliferation of colon cancer. Biochemical and biophysical research communications 449 (2):208-215. doi:10.1016/j.bbrc.2014.05.006

39. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL (2004) Gene expression profiling predicts clinical outcome of prostate cancer. The Journal of clinical investigation 113 (6):913-923. doi:10.1172/jci20032

40. Xu M, Jia Y, Liu Z, Ding L, Tian R, Gu H, Wang Y, Zhang H, Tu K, Liu Q (2016) Chromatin assembly factor 1, subunit A (P150) facilitates cell proliferation in human hepatocellular carcinoma. OncoTargets and therapy 9:4023-4035. doi:10.2147/ott.s107050

41. Bethke L, Webb E, Murray A, Schoemaker M, Johansen C, Christensen HC, Muir K, McKinney P, Hepworth S, Dimitropoulou P, Lophatananon A, Feychtung M, Lönnt S, Ahlbom A, Malmer B, Henriksson R, Auvinen A, Kiuru A, Salminen T, Swerdlow A, Houlston R (2008) Comprehensive analysis of the role of DNA repair gene polymorphisms on risk of glioma. Human molecular genetics 17 (6):800-805. doi:10.1093/hmg/ddm351

42. Peng H, Du B, Jiang H, Gao J (2016) Over-expression of CHAF1A promotes cell proliferation and apoptosis resistance in glioblastoma cells via AKT/FOXO3a/Bim pathway. Biochemical and biophysical research communications 469 (4):1111-1116. doi:10.1016/j.bbrc.2015.12.111

43. Barbieri E, De Preter K, Capasso M, Chen Z, Hsu DM, Tonini GP, Lefever S, Hicks J, Versteeg R, Pession A, Speleman F, Kim ES, Shohet JM (2014) Histone chaperone CHAF1A inhibits differentiation and promotes aggressive neuroblastoma. Cancer research 74 (3):765-774. doi:10.1158/0008-5472.can-13-1315

44. Reese BE, Bachman KE, Baylin SB, Rountree MR (2003) The methyl-CpG binding protein MBD1 interacts with the p150 subunit of chromatin assembly factor 1. Molecular and cellular biology 23 (9):3226-3236. doi:10.1128/mcb.23.9.3226-3236.2003

45. Mattiroli F, Gu Y, Balsbaugh JL, Ahn NG, Luger K (2017) The Cac2 subunit is essential for productive histone binding and nucleosome assembly in CAF-1. Scientific reports 7:46274. doi:10.1038/srep46274

46. Yang S, Long Q, Chen M, Liu X, Zhou H (2020) CAF-1/p150 promotes cell proliferation, migration,
invasion and predicts a poor prognosis in patients with cervical cancer. Oncology letters 20 (3):2338-2346. doi:10.3892/ol.2020.11775

47. Mascolo M, Ilardi G, Merolla F, Russo D, Vecchione ML, de Rosa G, Staibano S (2012) Tissue microarray-based evaluation of Chromatin Assembly Factor-1 (CAF-1)/p60 as tumour prognostic marker. International journal of molecular sciences 13 (9):11044-11062. doi:10.3390/ijms130911044

48. Volk A, Crispino JD (2015) The role of the chromatin assembly complex (CAF-1) and its p60 subunit (CHAF1b) in homeostasis and disease. Biochimica et biophysica acta 1849 (8):979-986. doi:10.1016/j.bbagrm.2015.05.009

49. Chen H, Wang X, Jia H, Tao Y, Zhou H, Wang M, Wang X, Fang X (2020) Bioinformatics Analysis of Key Genes and Pathways of Cervical Cancer. 13:13275-13283. doi:10.2147/ott.s281533

50. Diniz MG, Silva Jde F, de Souza FT, Pereira NB, Gomes CC, Gomez RS (2015) Association between cell cycle gene transcription and tumor size in oral squamous cell carcinoma. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 36 (12):9717-9722. doi:10.1007/s13277-015-3735-1

51. Wolf M, Korja M, Karhu R, Edgren H, Kilpinen S, Ojala K, Mousses S, Kallioniemi A, Haapasalo H (2010) Array-based gene expression, CGH and tissue data defines a 12q24 gain in neuroblastic tumors with prognostic implication. BMC cancer 10:181. doi:10.1186/1471-2407-10-181

52. Chan GK, Jablonski SA, Starr DA, Goldberg ML, Yen TJ (2000) Human Zw10 and ROD are mitotic checkpoint proteins that bind to kinetochores. Nature cell biology 2 (12):944-947. doi:10.1038/35046598
Figure legends

Figure 1 Hierarchical clustering analysis of top 100 differentially expressed genes in cervical cancer based on GEO (fold change >2.0, P < 0.05 in t-test).

Figure 2 Enrichment analysis of GO and KEGG pathway of DEGs in CC. (A) GO enrichment analysis (B) KEGG enrichment analysis.

Figure 3 Identification of mRNA biomarkers for CC. (A) Importance value of each DEGs ranked according to the mean decrease in accuracy by using the random forest analysis. (B) Variance rate of classification performance when increasing numbers of the predictive DEGs.

Figure 4 ROC analysis of five cervical cancer-specific mRNA biomarkers. The ROC results of these five diagnostic mRNA biomarkers (DNMT1, CHAF1B, CHAF1A, MCM2 and CDKN2A) based on (A) support random forest, (B) decision tree model and (C) support vector machine model.

Figure 5 ROC curves of DNMT1, CHAF1B, CHAF1A, MCM2, KNTC1, CRISP2, CRNN, KRT32, SLC5A1 and CDKN2A. (A) CHAF1A, (B) CHAF1B, (C) DNMT1, (D) CRISP2, (E) KRT32, (F)MCM2, (G)KNTC1, (H)CDKN2A, (I)CRNN, (J) SLC5A1.

Figure 6 Overall survival validation of cervical cancer patients grouped by median cutoffs of DNMT1, CHAF1B, CHAF1A, MCM2 and KNTC1. (A) CHAF1A, (B) CHAF1B, (C) DNMT1, (D) KNTC1, (E) MCM2.

Figure 7 Time-dependent ROC analysis, risk score analysis, and Kaplan–Meier analysis for the four-gene signature in CC. (A) LASSO Cox analysis, (B) risk score, (C) Kaplan–Meier curve of the four-gene signature,(D) Time-dependent ROC analysis.

Figure 8 qRT-PCR results of five DEGs (DNMT1, CHAF1B, CHAF1A, MCM2, KNTC1 and CDKN2A) in cervical cancer. *p<0.05, ** p<0.01, *** p<0.001

Figure 9 Validation of the expression of candidate genes in cervical cancer tissues and normal tissues in GEO and TCGA. (A) CHAF1A, (B) CHAF1B, (C) DNMT1, (D) KNTC1, (E) MCM2.

Figure 10 Immunohistochemical staining of candidate genes in cervical cancer.
tissues and normal tissues in the HPA database.
### Table 1 Details for GEO cervical cancer data

| GEO ID  | Samples(Normal:CC) | Type  | Platform  | Year | Author       | Type  |
|---------|--------------------|-------|-----------|------|--------------|-------|
| GSE7410 | 5:21               | mRNA  | GPL1708   | 2008 | Biewenga P   | tissue|
| GSE7803 | 10:21              | mRNA  | GPL96     | 2007 | Zhai Y       | tissue|
| GSE9750 | 24:33              | mRNA  | GPL96     | 2008 | Murty VV     | tissue|
| GSE55940| 5:5                | mRNA  | GPL16238  | 2014 | Ye C         | tissue|
| GSE63514| 24:28              | mRNA  | GPL570    | 2015 | den Boon J   | tissue|

### Table 2 Twenty mRNAs screened by LASSO

| ID     | Symbol | Combined.ES | P.Value | FDR | UpDown |
|--------|--------|-------------|---------|-----|--------|
| 1029   | CDKN2A | 3.491349    | 0       | 0   | Up     |
| 1786   | DNMT1  | 2.594474    | 0       | 0   | Up     |
| 2263   | FGFR2  | -1.54865    | 0       | 0   | Down   |
| 3882   | KRT32  | -1.73084    | 0       | 0   | Down   |
| 4171   | MCM2   | 3.044853    | 0       | 0   | Up     |
| 6523   | SLC5A1 | -2.23921    | 0       | 0   | Down   |
| 6691   | SPINK2 | -1.80647    | 0       | 0   | Down   |
| 7180   | CRISP2 | -1.97518    | 0       | 0   | Down   |
| 8208   | CHAF1B | 2.189046    | 0       | 0   | Up     |
| 8424   | BBOX1  | -2.53337    | 0       | 0   | Down   |
| 9735   | KNTC1  | 3.197516    | 0       | 0   | Up     |
| 9796   | PHYHIP | -2.30627    | 0       | 0   | Down   |
| 10036  | CHAF1A | 2.314881    | 0       | 0   | Up     |
| 10321  | CRISP3 | -3.04948    | 0       | 0   | Down   |
| 10947  | AP3M2  | 1.72574     | 0       | 0   | Up     |
| 23225  | NUP210 | 2.369988    | 0       | 0   | Up     |
| 49860  | CRNN   | -3.22716    | 0       | 0   | Down   |
| 64786  | TBC1D15| 1.697909    | 0       | 0   | Up     |
| 65982  | ZSCAN18| -2.0272     | 0       | 0   | Down   |
| 79875  | THSD4  | -2.47196    | 0       | 0   | Down   |
Table 3 Relationship between selected candidate genes and clinical features in CC

| Item                                | N   | DNMT1 Statistic | DNMT1 Pvalue | CHAF1B Statistic | CHAF1B Pvalue | CHAF1A Statistic | CHAF1A Pvalue | MCM2 Statistic | MCM2 Pvalue | KNTC1 Statistic | KNTC1 Pvalue |
|-------------------------------------|-----|-----------------|--------------|------------------|--------------|------------------|--------------|----------------|-------------|----------------|--------------|
| years_to_birth (Spearman Correlation) | 307 | -0.02           | 0.72         | 0.04             | 0.56         | 0.00             | 0.98         | 0.05           | 0.45       | 0.01           | 0.90         |
| Tumor_purity (Spearman Correlation) | 307 | 0.05            | 0.37         | 0.22             | 0.00         | 0.04             | 0.55         | 0.10           | 0.10       | 0.19           | 0.00         |
| ethnicity (Wilcoxon Test)            |     |                 |              |                  |              |                  |              |                |             |                |              |
| hispanic or latino                  | 24  | -0.01           | 0.12         | -0.03            | 0.16         | -0.01            | 0.56         | -0.01          | 0.37       | -0.02          | 0.42         |
| not hispanic or latino              | 171 |                |              |                  |              |                  |              |                |             |                |              |
| race (Kruskal-Wallis Test)           |     |                 |              |                  |              |                  |              |                |             |                |              |
| White                               | 211 | 2.05            | 0.73         | 7.74             | 0.10         | 1.97             | 0.74         | 7.92           | 0.09       | 4.21           | 0.38         |
| Black or African                    |     |                 |              |                  |              |                  |              |                |             |                |              |
| American                             | 30  |                |              |                  |              |                  |              |                |             |                |              |
| radiation_therapy (Wilcoxon Test)    |     |                 |              |                  |              |                  |              |                |             |                |              |
| Yes                                 | 129 | -0.01           | 0.37         | -0.02            | 0.04         | -0.03            | 0.10         | -0.02          | 0.19       | -0.01          | 0.18         |
| NO                                  | 55  |                |              |                  |              |                  |              |                |             |                |              |
| T1                                  | 141 |                |              |                  |              |                  |              |                |             |                |              |
| pathology_T_stage (Kruskal-Wallis Test) |   | 4.04            | 0.40         | 2.43             | 0.66         | 2.17             | 0.70         | 0.90           | 0.93       | 1.68           | 0.79         |
| T2                                  | 72  |                |              |                  |              |                  |              |                |             |                |              |
| T3                                  | 21  |                |              |                  |              |                  |              |                |             |                |              |
| T4                                  | 10  |                |              |                  |              |                  |              |                |             |                |              |
| pathology_N_stage (Wilcoxon Test)    |     |                 |              |                  |              |                  |              |                |             |                |              |
| NO                                  | 135 | -0.01           | 0.41         | -0.02            | 0.37         | 0.02             | 0.72         | 0.00           | 0.28       | -0.01          | 0.82         |
| N1                                  | 60  |                |              |                  |              |                  |              |                |             |                |              |
| pathology_M_stage (Wilcoxon Test)    |     |                 |              |                  |              |                  |              |                |             |                |              |
| M0                                  | 116 | 0.00            | 0.30         | -0.02            | 0.41         | -0.05            | 0.07         | -0.03          | 0.74       | -0.01          | 0.99         |
| M1                                  | 10  |                |              |                  |              |                  |              |                |             |                |              |
| histological_type (Kruskal-Wallis Test) |   | 1.68            | 0.89         | 5.32             | 0.38         | 11.32            | 0.05         | 26.49          | 0.00       | 3.52           | 0.62         |
| cervical squamous cell carcinoma     | 254 |                |              |                  |              |                  |              |                |             |                |              |
| endocervical type                    |     |                |              |                  |              |                  |              |                |             |                |              |
The bar chart illustrates the log2 fold change for several genes: DNMT1, CHAF1B, CHAF1A, MCM2, CDKN2A, and KNTC1. Each bar represents the fold change with error bars indicating the variability.

- **DNMT1**: Fold change 0.824
- **CHAF1B**: Fold change 0.781
- **CHAF1A**: Fold change 0.134
- **MCM2**: Fold change 0.795
- **CDKN2A**: Fold change 0.829
- **KNTC1**: Fold change 0.186

Statistical significance levels are indicated as ** for p < 0.01 and *** for p < 0.001.
