

## Original Article



# Using low-coverage whole genome sequencing technique to analyze the chromosomal copy number alterations in the exfoliative cells of cervical cancer

Tong Ren ,<sup>1,\*</sup> Jing Suo ,<sup>2,\*</sup> Shikai Liu ,<sup>3</sup> Shu Wang ,<sup>1</sup> Shan Shu ,<sup>1</sup>  
Yang Xiang ,<sup>1</sup> Jing-He Lang

<sup>1</sup>Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China

<sup>2</sup>Department of Obstetrics and Gynecology, The Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China

<sup>3</sup>Department of Obstetrics and Gynecology, Cangzhou Central Hospital, Cangzhou, China

## OPEN ACCESS

**Received:** Feb 22, 2018

**Revised:** May 19, 2018

**Accepted:** Jun 11, 2018

### Correspondence to

Shu Wang

Department of Obstetrics and Gynecology,  
Peking Union Medical College Hospital,  
Chinese Academy of Medical Science and  
Peking Union Medical College, 1 Shuaifuyuan,  
Dongcheng District, Beijing 100730, China.  
E-mail: wangshu219@hotmail.com

\*Tong Ren and Jing Suo contributed equally  
to this work.

**Copyright** © 2018. Asian Society of  
Gynecologic Oncology, Korean Society of  
Gynecologic Oncology

This is an Open Access article distributed  
under the terms of the Creative Commons  
Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>)  
which permits unrestricted non-commercial  
use, distribution, and reproduction in any  
medium, provided the original work is properly  
cited.

### ORCID iDs

Tong Ren   
<https://orcid.org/0000-0002-6284-2137>  
Jing Suo   
<https://orcid.org/0000-0002-3237-9604>  
Shikai Liu   
<https://orcid.org/0000-0003-2659-4014>

## ABSTRACT

**Objectives:** We analyzed the chromosomal-arm-level copy number alterations (CNAs) in the cervical exfoliative cell and tissue samples by using the low-coverage whole genomic sequencing technique.


**Methods:** In this study, we retrospectively collected 55 archived exfoliated cervical cell suspension samples and the corresponding formalin-fixed and paraffin-embedded tissue section samples including 27 invasive cervical cancer and 28 control cases. We also collected 19 samples of the cervical exfoliative cells randomly from women to verify the new algorithm model. We analyzed the CNAs in cervical exfoliated cell and tissue samples by using the low-coverage next generation of sequencing.

**Results:** In the model-building study, multiple chromosomal-arm-level CNAs were detected in both cervical exfoliated cell and tissue samples of all cervical cancer cases. By analyzing the consistency of CNAs between exfoliated cells and cervical tissue samples, as well as the heterogeneity in individual patient, we also established a C-score algorithm model according to the chromosomal-arm-level changes of 1q, 2q, 3p, 7q. The C-score model was then validated by the pathological diagnosis of all 74 exfoliated cell samples (including 55 cases in model-building group and 19 cases in verification group). In our result, a cutoff value of C-score >6 showed 100% sensitivity and 100% specificity in the diagnosis of cervical cancer.


**Conclusion:** In this study, we found that CNAs of cervical exfoliated cell samples could robustly distinguish invasive cervical cancer from cancer-free tissues. And we have also developed a C-score algorithm model to process the sequencing data in a more standardized and automated way.

**Keywords:** Uterine Cervical Neoplasms; Mass Screening; High-Throughput Nucleotide Sequencing; DNA Copy Number Variation

Shu Wang 
<https://orcid.org/0000-0001-5447-0946>

Shan Shu 
<https://orcid.org/0000-0003-4748-0227>

Yang Xiang 
<https://orcid.org/0000-0002-7324-2954>

Jing-He Lang 
<https://orcid.org/0000-0002-4452-1727>

### Conflict of Interest

No potential conflict of interest relevant to this article was reported.

### Author Contributions

Conceptualization: W.S.; Data curation: S.S.;  
Formal analysis: R.T.; Funding acquisition:  
W.S.; Investigation: R.T.; Methodology: R.T.,  
S.J., L.S.; Project administration: W.S., X.Y.,  
L.J.; Resources: S.J., L.S., S.S.; Supervision:  
W.S., X.Y., L.J.; Validation: W.S., X.Y., L.J.;  
Visualization: W.S., X.Y., L.J.; Writing - original  
draft: R.T., S.J., L.S.; Writing - review & editing:  
W.S., X.Y., L.J.

## INTRODUCTION

In the past 50 years, Papanicolaou cytology test has been the mainstay of cervical cancer screening, which has made a significant difference in reducing the incidence and mortality rate of cervical cancer. However, cytology has significant limitations in many aspects, most notably is its low sensitivity and poor reproducibility [1]. Compared to cytology, human papillomavirus (HPV) test is better in its early and upstream detection of cervical cancer, based on the understanding of the carcinogenesis of HPV infection. Furthermore, HPV test has a higher sensitivity and reproducibility. It also has many advantages in the automation, centralization, and better quality-checking technology for large specimen throughput. However, the false positive rate of HPV screening results has far outnumbered the true positive ones, leading to many unnecessary referrals to the invasive colposcopy procedures and making patients feel stressful [2]. In the post-vaccination era, it is foreseeable that a fully automated, computerized, and high efficient molecular diagnostic test would be applied in the cervical cancer screening.

Previous study showed that the chromosomal copy number changes are commonly seen in cervical cancer [3,4]. Besides, low-coverage (<20×) next generation sequencing (NGS) are proved to be a flexible and powerful method to assess the copy number alterations (CNAs) in tumor cells [5,6]. Kader et al. [5] indicated that CNA from formalin-fixed and paraffin-embedded (FFPE)-derived DNA using low-coverage whole genome sequencing (LC WGS) (even 0.1–2×coverage) has been reported although point mutations and loss of heterozygosity were not assayed. In this study, we analyzed the CNAs in cervical exfoliated cell and tissue samples by using the low-coverage NGS, and explored the value of this method in future cervical cancer screening.

## MATERIALS AND METHODS

### 1. Objectives and samples

In this study, 32 cases of cervical cancer who were treated in our hospital from January 2017 to June 2017 were retrospectively recruited as the study group, while 30 participants without cervical diseases as the control group. The FFPE tissue sections and the archived exfoliated epithelial cell samples were collected from both groups.

The inclusive criteria of the study group were as follows: the patients should be pathologically diagnosed with cervical cancer by either surgical excision or biopsy specimen at Peking Union Medical College Hospital. The inclusive criteria of the control group were as follows: the patients should have received hysterectomy for the indications of benign uterine or ovarian disease and were pathologically proved of being absence of cervical intraepithelial neoplasm (CIN) or invasive cervical carcinoma. The exclusion criteria were as follows: 1) The patients who had received any physical or chemical treatments before surgery, such as radiotherapy, immunotherapy, and chemotherapy. 2) The patients who were pregnant during treatment. 3) The patients who had autoimmune diseases, considering the possible copy number variation (CNV) related to some immune diseases themselves [7]. 4) The patients who had other concurrent malignant tumors.

In the verification study, before knowing the histological diagnosis, we randomly collected 19 surplus samples of cervical exfoliative cells from women who were receiving cervical

cancer screening at the outpatient department in our hospital in June 2017. These women have all received ThinPrep® Pap testing (Hologic, Inc., Bedford, MA, USA) and the Cobas® 4800 System HPV Genotyping Test (Roche Molecular Diagnostics, Pleasanton, CA, USA). If indicated, colposcopy and specimen pathological examinations were performed. Furthermore, as per standard practice, loop electrosurgical excision procedure (LEEP) or conization surgery were performed to treat CIN3 and a majority of CIN2 cases which were histologically confirmed in colposcopy and biopsy. Written consents were obtained from all participants enrolled in the study. The study was approved by the Institutional Review Board of Peking Union Medical College Hospital (No. S-K341).

In this study, all suspension cells were preserved in PreservCyt solution (Hologic, Inc.). FFPE tissue samples were obtained from 32 cervical cancer cases and 30 cases without CIN or cervical cancer. All of the histological slides were independently reviewed by 2 pathologists. The specimens of 8–10 freshly cut FFPE tissue sections (10 µm thick) were kept in Eppendorf containers.

## 2. Experiments procedures

Total genomic DNA was isolated by using QIAamp DNA mini kit (Qiagen, Hilden, Germany) from cell suspension samples and the DNeasy FFPE Kit (Qiagen) from FFPE tissue samples respectively, as literature described [8]. Notably, the DNA starting material extracted from 5 cervical cancer cases and 2 cancer-free cases (both cell suspension and FFPE) did not meet the quality control criteria. The qualified DNA extractions from the remaining 27 cases in the study group and 28 cases in the control group were preceded in the NGS procedure using low-coverage (10×).

The whole genomic sequencing was performed as previously described [9,10]. The genomic DNA was fragmented into an average size of 300 bp, and then 100 ng of the fragmented genomic DNA was used for the preparation of sequencing libraries (NEBnext Ultra II). The 8 bp barcoded sequencing adaptors were then ligated with DNA fragments and amplified by polymerase chain reaction. Purified sequencing libraries were massively parallel sequenced by Illumina HiSeqX platform.

## 3. Bioinformatics pipeline

At least 3G raw sequencing data per sample were obtained and aligned to human reference genome hg19 by using hardware-accelerated Xiphias™ aligner. Aligned bases with quality higher than 30 (duplicates removed) were counted according to the non-overlapping 200 kb bins. Then self-normalization was performed by dividing the raw base counts by the mean of all bins across the whole genome, aiming at offsetting the effects of different sequencing depths.

## 4. Abnormal chromosome-arms

Chromosome arm-level Z-scores were calculated with the mean of the base counts of all bins on that chromosome arm, and with the mean and variance of the same chromosome arm among the healthy panel samples:  $z = (x - \mu) / \sigma$ ,  $x$  is a raw score,  $\mu$  and  $\sigma$  are the mean and the standard deviation (SD) of the healthy panel samples. An absolute Z-score greater than 4 was defined as abnormal and the gain (Z-score >4), normal ( $-4 \leq$  Z-score  $\leq 4$ ) or loss (Z-score < -4) were also defined in this study.

## 5. The sample of verification model.

In the verification part of study, we have used all 74 exfoliated cell samples including 55 cases in the model-building study group and prospectively collected 19 new cases. Nine out of the 19 patients were found to have positive hrHPV results (among whom 6 cases were HPV16/18 positive). After the pathological examination, 10 cases were confirmed as cancer-free, while 3 cases as CIN1, 1 case as CIN2, 4 cases as CIN3, and 1 patient was diagnosed with SCC of the cervix.

In the model verification phase, we blindly renumbered the previous 55 samples and the new collection 19 samples, recalculated the C-score of each sample as described in the **Results**. By using these 74 cases, we test the cervical cancer detection efficiency of the C-score model by comparing it with the histopathological results.

## RESULTS

### 1. Objectives and samples

In the model-building group of this study, the qualified DNA extraction was acquired from the exfoliated cervical cells and FFPE cervical tissue samples in 27 cases with invasive cervical cancer and 28 individuals with no malignant tumor neither CIN. As shown in **Table 1**, the histology subtypes of cervical cancer include 20 cases of squamous cancer, 4 cases of adenocarcinoma (including 1 endometrioid and 1 clear cell carcinoma), 3 cases of adenosquamous cancer and 1 case of carcinosarcoma. The mean ( $\pm$ SD) age of cervical cancer

**Table 1.** The clinical characters of the cervical cancer group

Case number	Age	FIGO stage	Pathology type	Degree of histological differentiation	HPV test
C1	47	Ib1	SCC	Well	HPV16
C2	61	Ila	ASC	Poor	HPV16,18
C3	29	Iib	SCC	Well	HPV16,18
C4	54	Ib1	SCC	Well	HPV16
C5	51	Iib	ASC	Poor	HPV16,45
C6	54	Ib1	SCC	Well	HPV16
C7	39	Ib2	SCC	Poor	HPV16
C8	35	Ib2	SCC	Well	HPV16
C9	49	Ib1	AC (EC)	Moderate	Negative
C10	34	Ib1	SCC	Poor	HPV16
C11	53	IIib	SCC	Well	HPV16,18
C12	38	Ib1	SCC	Well	HPV16
C13	23	Ib1	SCC	Well	HPV16
C14	47	Iib	SCC	Moderate	HPV16
C15	44	Ia1	SCC	Poor	HPV16,52
C16	61	Ib2	SCC	Poor	HPV16
C17	48	Ib2	AC	Moderate	HPV18
C18	59	Ib1	AC (CCC)	Poor	Negative
C19	31	Ib2	SC	Well	HPV16
C20	48	Ib1	SC	Poor	Negative
C21	67	Ib1	AC	Moderate	HPV16
C22	54	Ib1	SC	Well	HPV16
C23	55	Ib2	SC	Poor	HPV16,58
C24	32	Ib2	CS	Poor	HPV18
C25	53	Ib1	SC	Moderate	HPV16
C26	39	Ib1	SC	Moderate	HPV16
C27	44	Ib2	SC	Moderate	HPV16

ASC, adenosquamous carcinoma; AC, adenocarcinoma; CCC, clear cell carcinoma; CS, carcinosarcoma; EC, endometrioid carcinoma; FIGO, International Federation of Gynecology and Obstetrics; HPV, human papillomavirus; SCC, squamous cell carcinoma.

group was 46.87 ( $\pm 10.37$ ) years (range 23–67 years), and was 48.83 ( $\pm 8.39$ ) years (range 37–71 years) in the control group. No significant difference in age was found in the above 2 groups. According to the FIGO stage of cervical cancer, 1 case was classified as stage Ia1, 13 cases as stage Ib1, 8 cases as stage IIa, 3 cases as stage IIb, and 1 case as stage III. In the verification part of our study (19 cases), 18 patients were initially diagnosed with normal or cervical intraepithelial lesions, and 1 patient was diagnosed with stage Ia2 cervical cancer by the pathological examination after cold knife conization.

Moreover, all patients in our research have received high-risk HPV (hrHPV) detection of the cervical exfoliated cells. In the study group, 23 patients had positive results. All of which showed HPV16/18. However, hrHPV test was negative in 1 case of cervical endometrioid carcinoma, 1 case of cervical clear cell carcinoma and 1 case of cervical squamous cell carcinoma (SCC). In the control group, 2 cases were HPV16 positive and the remaining 26 cases were all hrHPV-negative. All patients with cervical cancer have received standard surgical or radiological treatments after the pathological diagnosis.

## 2. Chromosome arm-level Z-score analysis

We analyzed the DNA samples of both cervical exfoliated cell suspension and FFPE tissue from 27 cervical cancer patients and 28 individuals with no cervical invasive carcinoma or any intraepithelial neoplasm (as seen in **Supplementary Tables 1-4**). Z-score of each chromosome arm was calculated as mentioned above in Materials and Methods, and demonstrated on a heat map. More precisely, we identified the chromosomal gain or loss for the cervical cancer and control group. All cervical cancer cases were found to have multiple chromosomal-arm-level CNAs in both exfoliated cell and FFPE tissue samples (**Fig. 1A**). No CNAs were detected in the control group, except one case in which arm-level abnormal CNA on chr9p were found both in exfoliated cell and FFPE tissue samples (**Fig. 1B**).

## 3. Distribution of chromosome arm of CNA of cervical tumor

The distributions of CNAs in exfoliated cervical cells and tissue samples of the cervical cancer group were shown in **Fig. 2**. By analyzing the data of the FFPE tissue samples of cervical cancer patients, we detected 3p loss in 70.4% cases, 6q loss or 20p gain in 63% cases, 7q loss in 59.2% cases, 1q gain or 3q gain in 55.6% cases, 2q loss or 4p loss in 51.8% cases, 5q loss and 17q gain in 48.1% cases, 1p gain, 11p loss or 18q loss in 44.4% cases. Regarding the results of exfoliated cells samples of the cervical cancer group, we detected 3p loss in 74.1% of cases, 1p gain, 2q loss, 3q gain, 5q loss, 6q loss, 11q gain, or 20p gain in 51.8% cases, 20q gain in 48.1% cases, 4q loss and 7q loss in 44.4% cases.

By analyzing the consistency of CNAs between the cervical tissue and cell suspension samples as well as the heterogeneity in different patients, we also found that the analysis of the combined CNAs of 1q gain, 2q loss, 3p loss and 7q loss could achieve 100% of accuracy rate in the pathological diagnosis of cervical cancer in both cytology and tissue samples. The consistency of CNV of each arms detected by FFPE and cell suspension samples was judged by kappa coefficient, and the results showed that the consistency of the 2 samples was fairly good (**Supplementary Tables 1-4**).

## 4. C-score model based on accumulation of arm-level Z-score of CNA on 4 chromosomes in exfoliated cervical cells

By analyzing the most common copies and the complementarity of chromosomes distributions of CNA in the cytological samples of cervical cancer, we selected the arm-level

changes of 1q, 2q, 3p, 7q to establish the C-score algorithm model (**Supplementary Table 5**). According to the reference literature [11,12], C-score is defined as the  $\log_2$  ratios of the sum of squares of arm-level Z-scores on significant chromosome arms:

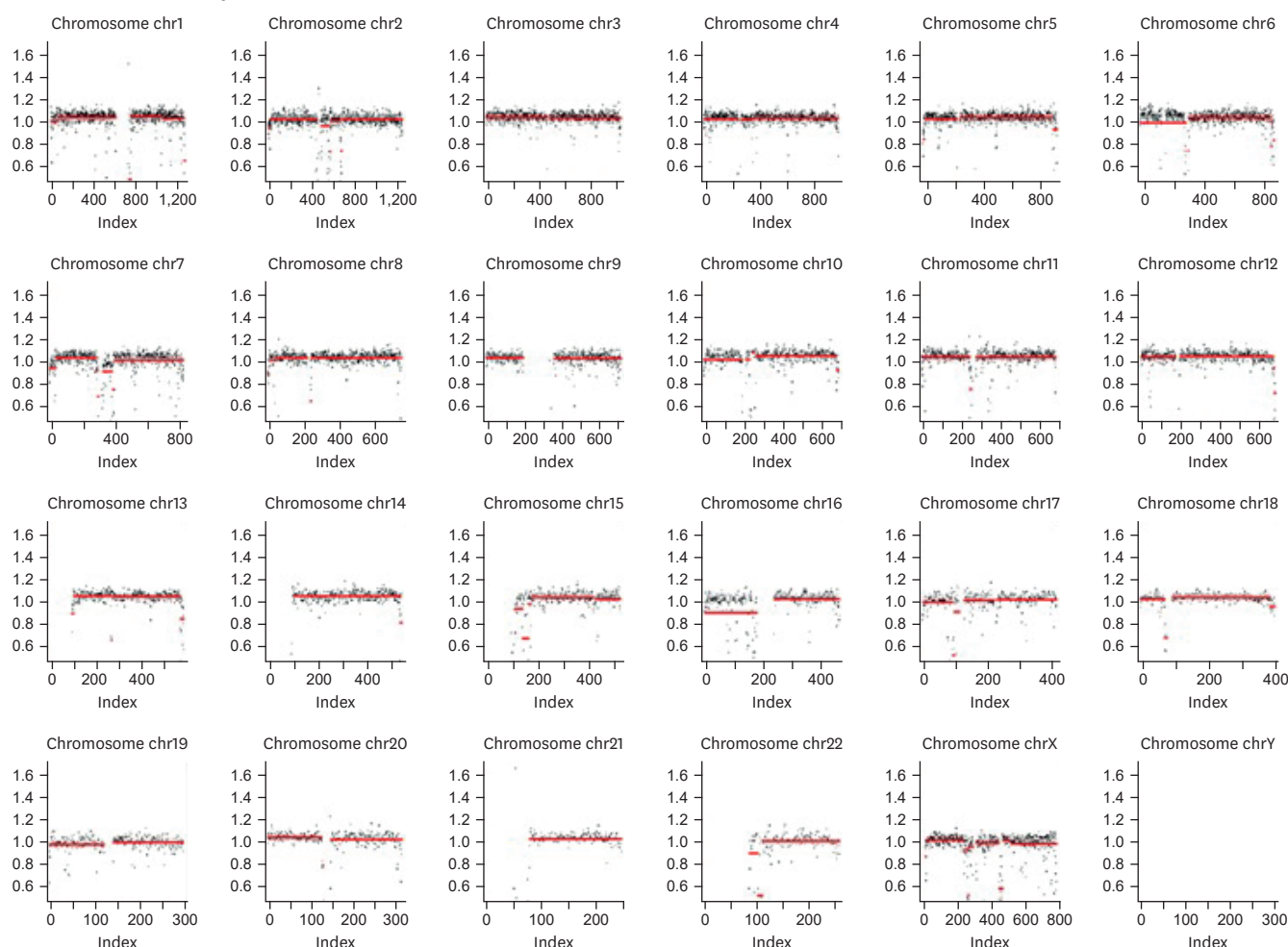
$$\text{Log2}(\text{C-score} = \sum_{\text{arm}=\{1q,2q,3p,7q\}} \text{Z-score}^2)$$

This algorithm could be considered as a parallel version of the raw arm-level Z-score. In this model, an absolute C-score greater than 6 was defined as abnormal. All cervical cancer patients were found to have abnormal C-scores in both exfoliated cell and tissue samples. All cases in the control group were found to have normal C-scores.

### 5. C-score model verification

In the verification part of study, the C-score model was validated by using all 74 exfoliated cell samples as described in **Material and Method**. A cutoff value of C-score >6 were found to

#### A Cancer-free sample

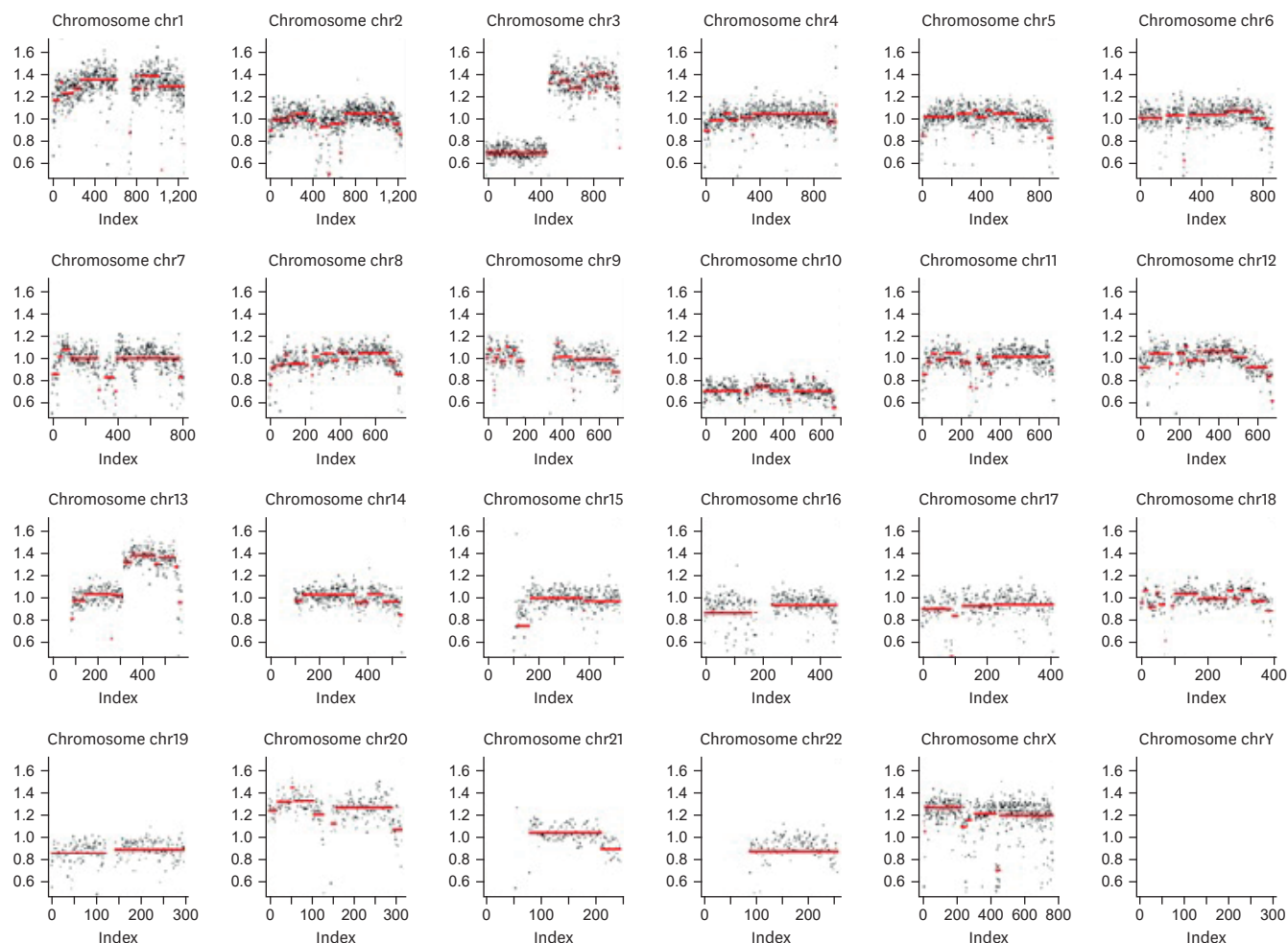


**Fig. 1.** The chromosome copy number variation in cytological samples randomly selected for illustration purposes. (A) No abnormal CNA was detected in the cytological sample of one case in the control group. (B) The multiple arm-level CNAs were found in the cytological sample of one cervical cancer case by using low-coverage NGS. The most obvious variations were mainly distributed at Chr1p(gain), chr1q(gain), chr3p(loss), chr3q(gain), chr7p(loss), chr8(loss), chr10p(loss), chr10q(loss), chr13q(gain), chr20p(gain), chr20q(gain), chrXp(gain), chrXq(gain).

(continued to the next page)



## B Cervical cancer sample

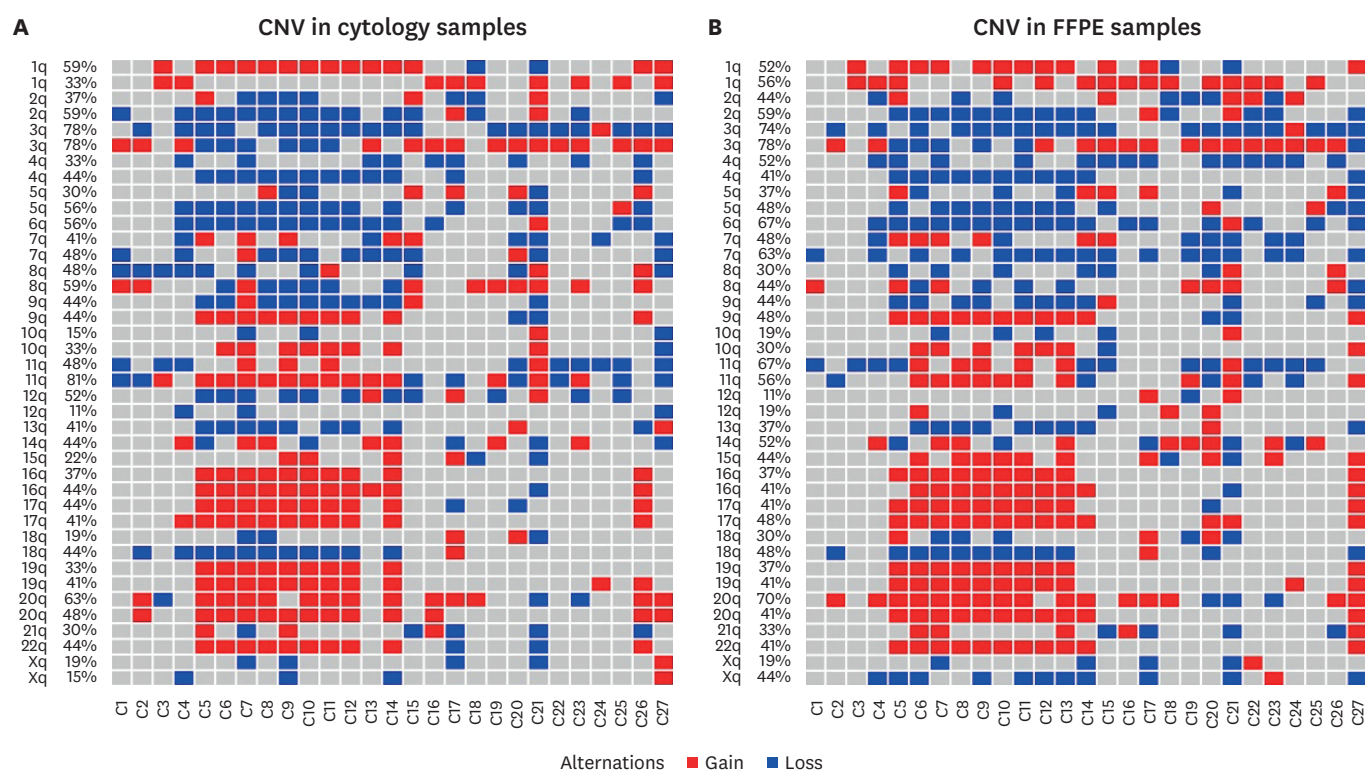


**Fig. 1. (Continued)** The chromosome copy number variation in cytological samples randomly selected for illustration purposes. (A) No abnormal CNA was detected in the cytological sample of one case in the control group. (B) The multiple arm-level CNAs were found in the cytological sample of one cervical cancer case by using low-coverage NGS. The most obvious variations were mainly distributed at Chr1p(gain), chr1q(gain), chr3p(loss), chr3q(gain), chr7p(loss), chr8(loss), chr10p (loss), chr10q(loss), chr13q(gain), chr20p(gain), chr20q(gain), chrXp(gain), chrXq (gain).

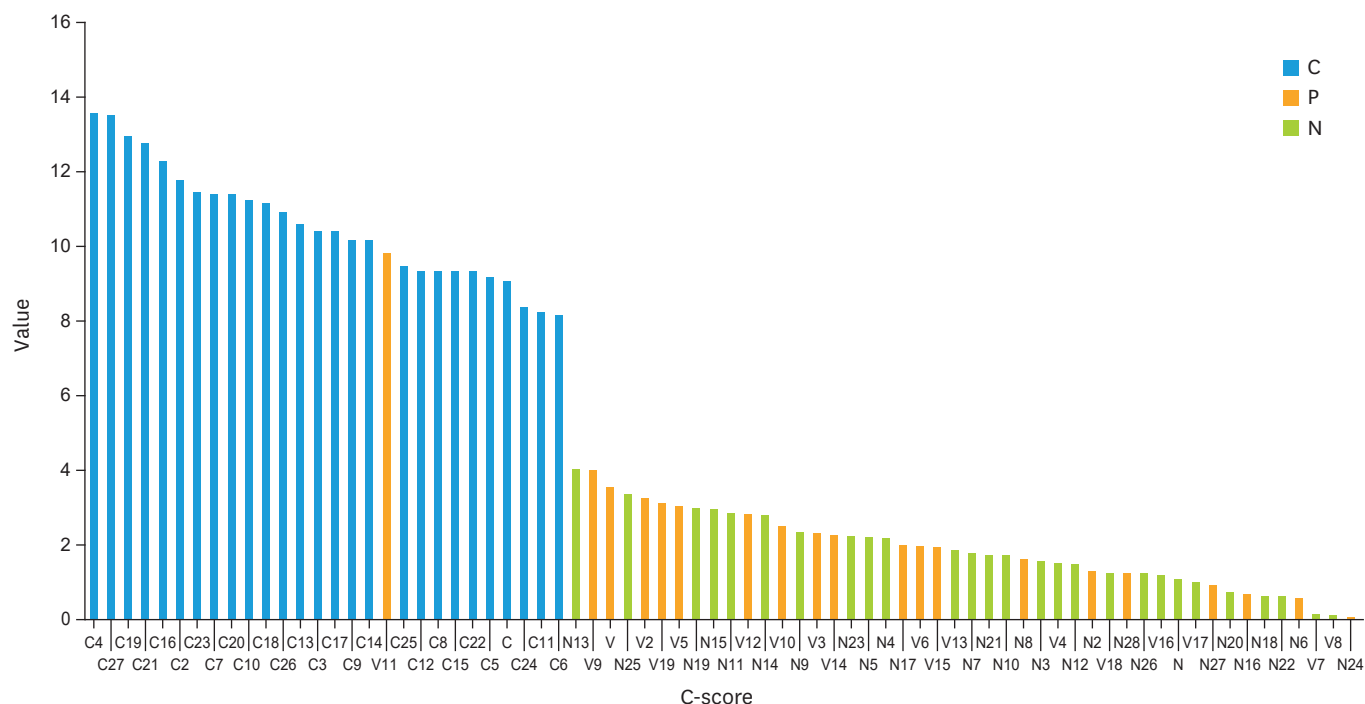
have 100% sensitivity and 100% specificity in the diagnosis of cervical cancer (**Fig. 3**). More detailed information about the distribution of CNAs contributing to C-score in the exfoliated cervical cells of all cases was shown in **Supplementary Table 5**.

## DISCUSSION

In the current model-building study, we analyzed the chromosome copy number changes in the liquid-based cervical cytology samples and FFPE tissue samples among cervical cancer cases and the control group. Low-coverage NGS platforms were used to produce personalized data of CNAs. Besides, we developed a Z-score of chromosome arm-level to establish the primary algorithm according to the raw data of NGS and the data of the panel healthy samples. By performing the above data processing method, the abnormal amplification or loss of signals by systemic errors should be minimized. Our results showed all cervical cancer patients have multiple arm-level CNAs detected in both exfoliated cell and FFPE samples.



**Fig. 2.** Genomic instability in the samples of 27 cervical cancer cases. Red represents gains and blue represents losses. Columns are samples and rows are positions of the probes along the 22 autosomes. (A) Heatmap of CNAs in cytological samples. (B) Heatmap of CNAs in FFPE tissue samples. CAN, copy number alteration; FFPE, formalin-fixed and paraffin-embedded.



**Fig. 3.** C-score values of all 74 exfoliated cell samples according to the algorithms newly developed in this study. Red bar represents the C-score value for cervical cancer cases (C group, n=27), blue represents tumor-free cases (N group, n=28) and yellow represents the cases in verification group (V group, n=19).



Whereas, only one woman in the control group has arm-level abnormal CNA on chr9p both in exfoliated cell and tissue samples. And none of the remaining cases in the control group had abnormal Z-score of CNA. The Z-score of CNA showed a high sensitivity and specificity in distinguishing invasive cervical cancer from cancer-free cases (100% and 96.6%), irrespective of the histological subtypes such as squamous carcinoma or adenocarcinoma. Moreover, it is interesting to note that among the 27 cervical cancer patients, 3 cases with negative hrHPV also showed an abnormal Z-score in both cervical smear and FFPE samples.

We also analyzed the distribution of the CNV on the 23 pairs of chromosomes in order to further investigate the arm-level copy number changes in cervical cancer. In accordance with previous reports, our results indicated that CNA in chr3 was commonly observed in cervical cancer. 3p loss and 3q gain were respectively found in about 70% and 50% of cervical cancer in FFPE tissue or cytology samples. Thomas et al have reviewed 32 study and concluded that the most common alterations in cervical SCC were 3q gains (0.55; 95% confidence interval [CI]=0.43–0.70), and 3p losses (0.36; 95% CI=0.27–0.48) [13]. They also presented that gains at 3q were particularly frequent in HPV16-positive cervical SCC cases (0.84; 95% CI=0.78–0.90), while the frequency was 84.2% in our study. Literature had also reported that the gene TERC harbored on chr3q might be used as a biomarker in cervical cancer screening [14]. Previous TCGA study demonstrated that chr3q (TERC, MECOM, TP63) gain was observed in more than 75% of cervical cancer patients, and 30% patients had a non-reported chr3p (TGFBR2) loss [15]. Besides, chr3p26 loss detected by our research was reported to be associated with HPV infection [16]. Another frequent loss region identified in our research was the chr2q33–q37, which was previously reported as well [17,18]. More specifically, CNAs in chr2q was observed in the early stage of various types of cancer. We also identified the deletion of certain regions on 2q in our cervical cancer samples, which may contain tumor suppression genes related to the progression of invasive cancer development [19,20]. For other chromosome abnormalities we used in our model, the loss of chr3p and the gain of chr1 were reported to be related to malignancy development [21,22]. The loss of chr4p, chr7q, chr5q, and the gain of chr20p have been commonly observed in cervical cancer across multiple studies [15,20–22].

Furthermore, based on the findings of our model-building study and the relative information provided by other literature, we established a new C-score algorithms model and defined the cutoff value as 6 to summarize the extent of CNA in each sample. By pooling the cytology samples of the verification and model-building group of our study, the C-score model was validated by using all 74 exfoliated cell samples. And it showed 96.6% sensitivity and 100% specificity in distinguishing the invasive cervical cancer from CIN or normal cervix samples (**Fig. 3**). Particularly, C-score values were far earlier obtained in patients from the verification study group. Among the 19 patients in the verification group, 9 patients received invasive colposcopy procedure, 1 received LEEP, and 4 received conization surgery. After the pathological confirmation one woman was diagnosed with invasive cervical cancer. It is noticeable that this patient happened to be the only one in the verification group who had a C-score higher than the cutoff value (>6). She was also found to have multiple arm-level CNAs detected in the exfoliated cell sample. However, this model failed to triage CIN3 from other lower grade CINs or normal cervical tissue, probably due to the limited number of CIN3 cases enrolled in this study. A verification study of a larger sample size in our hospital is under way. Providing the decreasing cost and increasing efficiency of the sequencing technology, we expect that the CNV analysis will become a valuable method in future cervical cancer screening. For instance, the recently launched sequencing machines (Illumina MiSeq, Life

Technologies PGM) have reduced its costs to be within the budget of individual laboratory. Meanwhile NGS low-coverage data have shown the potential to become a common application and have partially replaced the hybridization-based technologies such as array comparative genomic hybridization and single nucleotide polymorphism arrays [6,7].

In recent years, many researchers have been exploring a variety of potential molecular biomarkers to further stratify patients with positive HPV test results. Following the Pap smear and HPV test, it is possible that we will enter a new era of molecular cervical cancer screening. In this study, we found that the chromosomal CNAs of cervical cytological samples can robustly distinguish invasive cancers from tumor-free or CINs tissues. And we have also developed a C-score algorithm model to process the sequencing data in a more standardized and automated way. However, the limitation of this study is the somewhat small samples which might lead to some statistical bias of the result. And the variable sequencing platform and the bioinformatics process which might also affect end-results. Besides, the possibility of distinguishing CIN2+ with invasive cancer or lower degree CIN was not explored in this article. A larger scale of verification study to detect the chromosomal CNA in exfoliated cervical cells is in progress by our team, in which we also intended to evaluate the CNV of CIN in different degree. All in all, we proposed that molecular cytology smear by using NGS technology might become one of the non-invasive and efficient methods in cervical cancer screening in the future.

## ACKNOWLEDGEMENT

This study was funded by Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences(CIFMS)(No. 2016-I2M-1-002)

## SUPPLEMENTARY MATERIALS

### Supplementary Table 1

Cross tabulation of 1q between cytology sample and tissue sample

[Click here to view](#)

### Supplementary Table 2

Cross tabulation of 2q between cytology sample and tissue sample

[Click here to view](#)

### Supplementary Table 3

Cross tabulation of 3p between cytology sample and tissue sample

[Click here to view](#)

### Supplementary Table 4

Cross tabulation of 7q between cytology sample and tissue sample

[Click here to view](#)

## Supplementary Table 5

The distribution of CNA contributing to C-score in the exfoliated cervical cells of all cases

[Click here to view](#)

## REFERENCES

1. Blatt AJ, Kennedy R, Luff RD, Austin RM, Rabin DS. Comparison of cervical cancer screening results among 256,648 women in multiple clinical practices. *Cancer Cytopathol* 2015;123:282-8.  
[PUBMED](#) | [CROSSREF](#)
2. Wright TC, Stoler MH, Behrens CM, Sharma A, Zhang G, Wright TL. Primary cervical cancer screening with human papillomavirus: end of study results from the ATHENA study using HPV as the first-line screening test. *Gynecol Oncol* 2015;136:189-97.  
[PUBMED](#) | [CROSSREF](#)
3. Melsheimer P, Vinokurova S, Wentzensen N, Bastert G, von Knebel Doeberitz M. DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. *Clin Cancer Res* 2004;10:3059-63.  
[PUBMED](#) | [CROSSREF](#)
4. Bodelon C, Vinokurova S, Sampson JN, den Boon JA, Walker JL, Horswill MA, et al. Chromosomal copy number alterations and HPV integration in cervical precancer and invasive cancer. *Carcinogenesis* 2016;37:188-96.  
[PUBMED](#) | [CROSSREF](#)
5. Kader T, Goode DL, Wong SQ, Connaughton J, Rowley SM, Devereux L, et al. Copy number analysis by low coverage whole genome sequencing using ultra low-input DNA from formalin-fixed paraffin embedded tumor tissue. *Genome Med* 2016;8:121.  
[PUBMED](#) | [CROSSREF](#)
6. Gusnanto A, Taylor CC, Nafisah I, Wood HM, Rabbitts P, Berri S. Estimating optimal window size for analysis of low-coverage next-generation sequence data. *Bioinformatics* 2014;30:1823-9.  
[PUBMED](#) | [CROSSREF](#)
7. Yim SH, Jung SH, Chung B, Chung YJ. Clinical implications of copy number variations in autoimmune disorders. *Korean J Intern Med* 2015;30:294-304.  
[PUBMED](#) | [CROSSREF](#)
8. Gorringer KL, Hunter SM, Pang JM, Opeskin K, Hill P, Rowley SM, et al. Copy number analysis of ductal carcinoma in situ with and without recurrence. *Mod Pathol* 2015;28:1174-84.  
[PUBMED](#) | [CROSSREF](#)
9. Chiu RW, Chan KC, Gao Y, Lau VY, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci USA* 2008;105:20458-63.  
[PUBMED](#) | [CROSSREF](#)
10. Liang D, Lv W, Wang H, Xu L, Liu J, Li H, et al. Non-invasive prenatal testing of fetal whole chromosome aneuploidy by massively parallel sequencing. *Prenat Diagn* 2013;33:409-15.  
[PUBMED](#) | [CROSSREF](#)
11. Bhat S, Kabekkodu SP, Varghese VK, Chakrabarty S, Mallya SP, Rotti H, et al. Aberrant gene-specific DNA methylation signature analysis in cervical cancer. *Tumour Biol* 2017;39:1010428317694573.  
[PUBMED](#) | [CROSSREF](#)
12. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 2007;23:657-63.  
[PUBMED](#) | [CROSSREF](#)
13. Thomas LK, Bermejo JL, Vinokurova S, Jensen K, Bierkens M, Steenbergen R, et al. Chromosomal gains and losses in human papillomavirus-associated neoplasia of the lower genital tract - a systematic review and meta-analysis. *Eur J Cancer* 2014;50:85-98.  
[PUBMED](#) | [CROSSREF](#)
14. Li T, Tang L, Bian D, Jia Y, Huang X, Zhang X. Detection of hTERT and c-MYC genes in cervical epithelial exfoliated cells for cervical cancer screening. *Int J Mol Med* 2014;33:1289-97.  
[PUBMED](#) | [CROSSREF](#)

15. Cancer Genome Atlas Research Network, Albert Einstein College of Medicine, Analytical Biological Services, Barretos Cancer Hospital, Baylor College of Medicine, Beckman Research Institute of City of Hope. Integrated genomic and molecular characterization of cervical cancer. *Nature* 2017;543:378-84.  
[PUBMED](#) | [CROSSREF](#)
16. Heselmeyer K, Macville M, Schröck E, Blegen H, Hellström AC, Shah K, et al. Advanced-stage cervical carcinomas are defined by a recurrent pattern of chromosomal aberrations revealing high genetic instability and a consistent gain of chromosome arm 3q. *Genes Chromosomes Cancer* 1997;19:233-40.  
[PUBMED](#) | [CROSSREF](#)
17. Rao PH, Arias-Pulido H, Lu XY, Harris CP, Vargas H, Zhang FF, et al. Chromosomal amplifications, 3q gain and deletions of 2q33-q37 are the frequent genetic changes in cervical carcinoma. *BMC Cancer* 2004;4:5.  
[PUBMED](#) | [CROSSREF](#)
18. Narayan G, Pulido HA, Koul S, Lu XY, Harris CP, Yeh YA, et al. Genetic analysis identifies putative tumor suppressor sites at 2q35-q36.1 and 2q36.3-q37.1 involved in cervical cancer progression. *Oncogene* 2003;22:3489-99.  
[PUBMED](#) | [CROSSREF](#)
19. Dasgupta S, Chakraborty SB, Roy A, Roychowdhury S, Panda CK. Differential deletions of chromosome 3p are associated with the development of uterine cervical carcinoma in Indian patients. *Mol Pathol* 2003;56:263-9.  
[PUBMED](#) | [CROSSREF](#)
20. Cortés-Gutiérrez EI, Dávila-Rodríguez MI, Muraira-Rodríguez M, Said-Fernández S, Cerda-Flores RM. Association between the stages of cervical cancer and chromosome 1 aneusomy. *Cancer Genet Cytogenet* 2005;159:44-7.  
[PUBMED](#) | [CROSSREF](#)
21. van den Tillaart SA, Corver WE, Ruano Neto D, ter Haar NT, Goeman JJ, Trimbos JB, et al. Loss of heterozygosity and copy number alterations in flow-sorted bulky cervical cancer. *PLoS One* 2013;8:e67414.  
[PUBMED](#) | [CROSSREF](#)
22. Mitra AB. Genetic deletion and human papillomavirus infection in cervical cancer: loss of heterozygosity sites at 3p and 5p are important genetic events. *Int J Cancer* 1999;82:322-4.  
[PUBMED](#) | [CROSSREF](#)