Genome-wide profiling of the human papillomavirus DNA integration in cervical intraepithelial neoplasia and normal cervical epithelium by HPV capture technology

Ying Liu1,*, Chaoting Zhang1,*, Weijiao Gao2, Limin Wang3, Yaqi Pan1, Yunong Gao2, Zheming Lu4,* & Yang Ke1

HPV integration plays an important role in cervical carcinogenesis. HPV genotypes and the exact integration sites were investigated using HPV capture technology combined with next generation sequencing in 166 women. Three, one and six integration sites were verified in 7 HPV-positive ‘normal cervical epithelium’, 6 HPV-positive CIN2 and 15 HPV-positive CIN 3 samples, respectively. Of the 10 integrations, one and nine were involved with HPV33 and HPV16, respectively. Our study accurately evaluated HPV integration level in CINs and normal cervical tissues using high-throughput viral integration detection method providing basic evidence for HPV integration-driven cervical carcinogenesis.

Human papillomavirus (HPV) infection has been recognized as an important cause of cervical precancerous lesions or cancer, and yet is necessary but not sufficient for cervical carcinogenesis process1,2. Therefore, in addition to HPV infection, HPV integration could contribute to the cervical carcinogenesis process.

HPV integration could upregulate the expression of viral oncogenes E6 and E7 and eventually promotes host genomic instability, which could be a crucial event of cervical carcinogenesis process3,4. Additionally, the level of HPV integration was positively associated with cervical intraepithelial neoplasia (CIN) grades and was proposed as a marker for cervical disease progress5,6. Therefore, comprehensively and accurately identifying the sites and level of HPV integrations from normal cervical epithelium to CIN and cervical cancer is necessary to assess HPV-induced carcinogenesis process. However, cervical HPV integrations have only been comprehensively investigated in invasive cancer and limited data were available in CINs and normal cervical epithelium7,8. Moreover, most previous assays could not identify HPV integrations sensitively, which could lead to the underestimation of HPV integration level7,9,10. Recently, the study by Hu et al. reported 3,667 HPV integration events in 26 CINs, 104 cervical cancer samples, and five cell lines based on high-throughput viral integration detection

1*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Z.L. (email: luzheming@bjmu.edu.cn) or Y.K. (email: keyang@bjmu.edu.cn)
However, Nigel Dyer et al. claimed that 87% of the integration breakpoints reported by Hu et al. were likely to be experimental and computational artifacts according to their own data analysis pipeline\(^1\). Meanwhile, we used the same HPV capture technology and the next generation sequencing as in Hu’s study to detect cervical HPV integrations in 39 HPV-positive primary cervical tumor samples and 2 cell lines and yet only identified 117 unique validated HPV integration breakpoints\(^2\). Moreover, we found that the Sanger sequencing validation rate based on one, two, three, or more than three discordant paired-end reads was 3.7\%, 47.8\%, 44.4\%, and 83.3\%, respectively, indicating that HIVID could be a sensitive method to detect integrated HPV and yet has a high false positive rate with fewer supporting reads. HPV integration rates in the study by Hu et al. could be overestimated in cervical cancer and even in CINs. Therefore, given that comprehensive and accurate data regarding HPV integrations in CINs and normal cervical tissues were limited, we detected HPV integrations in a series of CINs and normal cervical epithelium samples using base-resolution HPV capture technology and the next generation sequencing as previously reported\(^2\).

In this study we enrolled 166 participants with CIN or ‘normal cervical epithelium’ in order to investigate the level of cervical HPV integration in CINs and normal cervical epithelium.

Result

**Distribution of HPV types.** As shown in Fig. 1 and Table 1, of the 64 ‘normal cervical epithelium’, 62 CIN 1, 19 CIN 2 and 21 CIN 3 samples, we detected HPV in 7 (10.9\%, 95% CI: 4.5–21.2\%), 8 (12.9\%, 95% CI: 5.7–23.9\%), 6 (31.6\%, 95% CI: 12.6–56.6\%) and 15 (71.4\%, 95% CI: 47.8–88.7\%) samples, respectively. Of the 166 samples, we detected HPV16 (n = 23), HPV18 (n = 2), HPV33 (n = 2) and HPV58 (n = 5); four samples harbored two types of HPV with HPV45 and HPV58 in CIN 1–42, HPV 33 and HPV18 in CIN 1–42, HPV 82 and HPV52 in CIN 2–12, and HPV58 and HPV52 in CIN 3–5 (Fig. 2). HPV infection rate in younger women (age < 50) and in older women (age > 50) were 25.2\% and 12.8\% (P = 0.096 by Fisher’s exact test).

**Determination of potential HPV integration sites.** As described in our previous study\(^2\), if a specific site had one or more discordant reads mapped on one end to the HPV reference genome and the other to human chromosome, it would be considered as a potential HPV integration locus. A total of 37, 21, and 45 potential HPV integrations were identified in 7 HPV-positive ‘normal cervical epithelium’, 8 HPV-positive CIN 1, 6 HPV-positive CIN 2 and 15 HPV-positive CIN 3 samples, respectively (Table 1).

**Validation of HPV integration sites.** In order to confirm the potential HPV integration sites and to further identify the sequence between cellular and viral genome, all potential HPV integration positions were verified by targeted PCR amplification and Sanger sequencing. The validation rate of Sanger sequencing on the basis of one, two, three, or more than three discordant paired-end reads, was 0% (0/131), 33.3\% (95\% CI: 0.8–90.6\%), 50\% (95\% CI: 1.3–98.7\%), and 72.7\% (95\% CI: 39.0–94.0\%), respectively (Table 1). After validation, only three, zero, one, and six integration sites were verified in 7 HPV-positive ‘normal cervical epithelium’, 8 HPV-positive CIN 1, 6 HPV-positive CIN 2 and 15 HPV-positive CIN 3 samples, respectively. Of the 10 integrations, one was involved with HPV33 and nine with HPV16, respectively (Table 1).

**Mapping and characterization of cellular-viral junction sequences.** 10 unique HPV integration breakpoints were distributed in six samples, including ‘normal cervical epithelium’ (n = 1), CIN2 (n = 1), and CIN3 (n = 4) (Fig. 3 and Table 1). Of the 10 integrations, four samples (CIN2-5, CIN3-15, CIN3-20, and CIN3-21) had one integration site and two samples (Control-31 and CIN3-21) had three integration sites, respectively (Fig. 3 and Table 1). The viral genome regions of the 10 validated integration positions in six samples were E1 (n = 5),
E2 (n = 2), E2/E4 (n = 1), L1 (n = 1), and L2 (n = 1). Due to the limited number of integration events, we did not find hot spots in the human genome (Table 1).
All integration positions were examined for the presence of fragile sites in the human genome. Of the 10 integration positions, one was located in a fragile site and three were close to a fragile site (Supplementary Table S1). Meanwhile, the human genomic sequences within 50 kb of an integration locus were investigated. Seven integration sites were located in cellular genes with six in introns and one in an exon (Supplementary Table S1).

**Discussion**

Our study only found three, zero, one and six verified integration sites in 7 HPV-positive ‘normal cervical epithelium’, 8 HPV-positive CIN1, 6 HPV-positive CIN2 and 15 HPV-positive CIN 3 samples, respectively. HPV integration sites were mainly located in the E1 and E2 regions of the viral genome and in cellular genes of the human genome.
We found that HPV integration rates in HPV-positive CIN 1, CIN 2, and CIN 3 samples were 0% (0/8), 16.6% (95% CI: 0.4–64.1%), and 26.8% (95% CI: 7.8–55.1%), respectively, which were similar to those reported previously.12,13,14 However, the study by Hu et al.15 reported that HPV integration rates in CIN 1, CIN 2, and CIN 3 were 50% (95% CI: 18.7–81.3%), 44.4% (95% CI: 13.7–78.8%), and 71.4% (95% CI: 29.0–96.3%), respectively, which were much higher than the validated integration rates in our study and yet similar to what we considered as the “potential” HPV integration rates (37.5% [95% CI: 8.5–75.5%], 66.7% [95% CI: 22.3–95.7%], and 53.3% [95% CI: 26.6–78.7%]) based on HPV capture and the next generation sequencing technology without Sanger sequencing validation. Notably, the validation rate of Sanger sequencing on the basis of one, two, three, or more than three different paired-end reads was 0% (0/131), 33.3% (95% CI: 0.8–90.6%), 50% (95% CI: 1.3–98.7%), and 72.7% (95% CI: 39.0–94.0%) in our study, respectively. Moreover, HPV integration rates in cervical cancer was significantly higher in Hu's study15 than in our previous study12. The overall HPV integration rates in the study by Hu et al. could be overestimated in cervical cancer and CINs, since only selected HPV integrations in Hu's study were validated by Sanger sequencing.

It is worth noting that HPV integrations could occur in cervical tissue with normal epithelium and the integration rate in CIN 3 was significantly higher than those in CIN 1 or CIN 2. This indicated that HPV integrations could play an important role in the early stage of cervical carcinogenesis, although our results were lack of the statistical analysis and the sample size may not be enough for the analysis. In addition, we found similar characteristics of the HPV integration sites in cervical cancer and non-cancer specimens.12 For example, HPV integration sites were mainly located in the E1 and E2 regions of the viral genome and in cellular genes of the human genome.

Two different types of mechanisms are presumed to explain cervical carcinogenesis process induced by HPV integration, i.e. altering viral gene expression or disrupting cellular transcripts. In order to determine the effect of these two mechanisms, it is necessary to comprehensively profile the host and viral genome. However, detection methods of HPV integration in most previous studies were low-throughput and lower sensitivity. In order to better understand the cervical carcinogenesis induced by HPV integration, this approach is able to discern fusion breakpoints accurately at single-base resolution for further elucidating the effect of HPV integration on viral and its flanking cellular transcripts. In addition, since HPV integration could lead to the viral persistence and moreover HPV persistent infection plays role in cervical carcinogenesis, this approach provides unbiased, genome-wide integration information to monitor the persistent or permanent infection.

However, there are some limitations in our study. Firstly, since HPV integration rates in ‘normal cervical epithelium’ and CIN were low and moreover only 36 HPV-positive women were involved in HPV integration analysis, comprehensively evaluating the sites and the level of HPV integrations was limited to some extent. Secondly, the cross-sectional study did not investigate the temporal relationship between HPV integration and CINs. Thirdly, in our study, CIN enriched tissue was not sampled by laser microdissection, which did not rule out contamination from normal adjacent epithelium or the underlying stroma. This would overestimate HPV integration rate in CINs to some extent although this effect was small due to significantly lower HPV integration rate in ‘normal cervical epithelium’ than in CINs. Fourthly, Since HPV DNA was detected using a highly sensitive PCR primer set (SPF1/GP6+) amplifying a 184-bp fragment of the L1 open-reading frame before performing HPV capture and sequencing, this might produce HPV false negatives from L1 breakpoints among the 166 samples. However, since proportion of breakpoints occurring in this targeted region of L1 was low16 and in most situations, HPV viral genome may be existed in both episomal and integrated forms, HPV false negative probability due to L1 breakpoints was small.

In summary, the accurate identification of HPV integrations in CINs and normal cervical tissues could provide basic evidence for HPV integration-driven cervical carcinogenesis and be served as individualized markers in cervical cancer screening in the future.

Materials and Methods
Study population and specimen collection. A total of 166 cervical biopsy specimens were collected and diagnosed with normal cervical epithelium or acute/chronic cervicitis without atypical hyperplasia (n = 64), CIN 1 (n = 62), CIN 2 (n = 19) and CIN 3 (n = 21) from Beijing Cancer Hospital, Beijing, China, between 2014 and 2015. All biopsy specimens were reviewed by two experienced pathologists who confirmed the diagnosis of CIN. Cervical biopsies were histologically diagnosed using criteria defined by the World Health Organization.16 All cases have no histological evidence of epithelial malignancy of the cervix in this study. Normal cervical epithelium or acute/chronic cervicitis without atypical hyperplasia is defined as “normal cervical epithelium” in this study. Punch biopsy samples were divided into two parts; one was kept for histopathological analysis, and the other one was used for HPV typing and integration analysis. Individual informed consents had been collected from all participants. This study received ethical approval from the Institutional Review Board of the Peking University School of Oncology, China. All experiments were performed in accordance with relevant guidelines and regulations.

The specimens were stored at −80°C and genomic DNA was extracted from the frozen tissues using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The β-globin gene was evaluated in all specimens by PCR.

HPV typing and integration detection. HPV DNA in valid (β-globin positive) specimens was detected using a highly sensitive PCR primer set (SPF1/GP6+) amplifying a 184-bp fragment of the L1 open-reading frame. Specimens showing the PCR amplification product were used to identify HPV genotypes and integrations. HPV probes were designed according to the full-length genome of 17 HPV types (6, 11, 16, 18, 31, 33, 35, 39, 45, 52, 56, 58, 59, 66, 68, 69, and 82) by MyGenostics (MyGenostics, Baltimore, MD, USA). Details of HPV typing and the detection of HPV integrations, as well as Sanger sequencing validation of potential HPV integration sites were described previously.12 In brief, the whole-genomic libraries were hybridized with HPV probes (MyGenostics GenCap Technology), adsorbed onto the beads via biotin and streptavidin magnetic beads, and...
the uncaptured DNA fragments were removed by washing. Then the eluted fragments containing the targeted gene were enriched by PCR to generate libraries for sequencing. Libraries were quantified and sequenced for paired-end 125bp using the Illumina HiSeq 2500 sequencer (Illumina Inc., San Diego, CA, USA). Illumina clean reads were mapped to human genome (GRCh37/hg19) and HPV genome of 17 types using the BWA program. The paired-end read, uniquely mapped with one end to a human chromosome and the other to the HPV reference genome, is identified as a discordant read pair. If a specific position has one or more discordant read pairs, it would be considered as a potential HPV integration site. PCR and Sanger sequencing were used to verify all the potential HPV integration breakpoints. All sequences of the fusion genes were characterized by the NCBI human mega Blast database alignment tool and the UCSC Blat database.

Statistical analyses. Fisher’s exact test was used to determine the relationship between age and HPV infection. Statistical analyses were performed using STATA version 12.0 software (STATA Corporation, College Station, TX, USA). P values less than 0.05 (two-sided) were considered to be significant.

References
1. zur Hausen, H. Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer 2, 342–350, doi: 10.1038/nrc798 (2002).
2. Crobie, E. J., Einstein, M. H., Francveschi, S. & Kitchener, H. C. Human papillomavirus and cervical cancer. Lancet 382, 889–899, doi: 10.1016/S0140-6736(13)60022-7 (2013).
3. Akagi, K. et al. genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome Res 24, 185–199, doi: 10.1101/gr.164806.113 (2014).
4. Pett, M. & Coleman, N. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? J Pathol 212, 356–367, doi: 10.1002/path.2192 (2007).
5. Hudeelist, G. et al. physical state and expression of HPV DNA in benign and dysplastic cervical tissue: different levels of viral integration are correlated with lesion grade. Gynecologic oncology 92, 873–880 (2004).
6. Arias-Pulido, H., Peyton, C. L., Jostte, N. E., Vargas, H. & Wheelere, C. M. Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. Journal of clinical microbiology 44, 1755–1762 (2006).
7. Woodman, C. R., Collins, S. I. & Young, L. S. The natural history of cervical HPV infection: unresolved issues. Nature Reviews Cancer 7, 11–22 (2007).
8. Hu, Z. et al. genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. Nat Genet 47, 158–163, doi: 10.1038/ng.3178 (2013).
9. Thorland, E. C. et al. Human papillomaviruses type 16 integrations in cervical tumors frequently occur in common fragile sites. Cancer research 60, 5916–5921 (2000).
10. Raybould, R., Fandere, A., Wilkinson, G. W. & Hibbitts, S. HPV integration detection in CaSki and SiHa using detection of integrated papillomavirus sequences and restriction-site PCR. Journal of virological methods 206, 51–54 (2014).
11. Dyer, N., Young, L. & Ott, S. Artifacts in the data of Hu et al. Nat Genet 48, 2–4, doi: 10.1038/ng.3392 (2016).
12. Liu, Y., Lu, Z., Xu, R. & Ke, Y. Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology. Oncotarget 7, 5852–5864, doi: 10.18632/oncotarget.6809 (2016).
13. Li, H. et al. Preferential sites for the integration and disruption of human papillomavirus 16 in cervical lesions. J Clin Virol 56, 342–347, doi: 10.1016/j.jcv.2012.12.014 (2013).
14. Klaes, R. et al. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res 59, 6132–6136 (1999).
15. Kuri, R. J., Carcangiu M. L. & Herrington C. S. WHO Classification of Tumours of Female Reproductive Organs. Lyon: International Agency for Research on Cancer, 172–176 (2014).
16. Huang, S.-L. et al. Comparison between the Hybrid Capture II Test and an SPF1/GP6+ PCR-based assay for detection of human papillomavirus DNA in cervical swab samples. Journal of clinical microbiology 44, 1733–1739 (2006).

Acknowledgements
This work was supported by grants from Natural Science Foundation of China [grant number 81321003], Beijing Natural Science Foundation [grant number 7152032], and Beijing Municipal Science and Technology [grant numbers Z151100001615022, Z141100002114046].

Author Contributions
Z.L. and Y.K. were involved in the design and supervision of the survey. Y.L., C.Z., W.G., L.W., Y.P. and Y.G. contributed to the sample collection and analyzed the data. Y.L., C.Z. and Z.L. performed laboratory testing of specimens and drafted the manuscript text. All authors approved the final draft.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liu, Y. et al. Genome-wide profiling of the human papillomavirus DNA integration in cervical intraepithelial neoplasia and normal cervical epithelium by HPV capture technology. Sci. Rep. 6, 35427; doi: 10.1038/srep35427 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016