IJC
International Journal of Cancer

Dysregulation of host cellular genes targeted by human papillomavirus (HPV) integration contributes to HPV-related cervical carcinogenesis

Ruiyang Zhang[1], Congle Shen[1], Lijun Zhao[2], Jianliu Wang[2], Malcolm McCrae[3], Xiangmei Chen[1] and Fengmin Lu[1]

[1] Department of Microbiology and Infectious Disease Center, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, People's Republic of China
[2] Department of Obstetrics and Gynecology, Peking University People's Hospital, Beijing, People's Republic of China
[3] The Pirbright Institute, Pirbright, United Kingdom

Integration of human papillomavirus (HPV) viral DNA into the human genome has been postulated as an important etiological event during cervical carcinogenesis. Several recent reports suggested a possible role for such integration-targeted cellular genes (ITGs) in cervical carcinogenesis. Therefore, a comprehensive analysis of HPV integration events was undertaken using data collected from 14 publications, with 499 integration loci on human chromosomes included. It revealed that HPV DNA preferred to integrate into intragenic regions and gene-dense regions of human chromosomes. Intriguingly, the host cellular genes nearby the integration sites were found to be more transcriptionally active compared with control. Furthermore, analysis of the integration sites in the human genome revealed that there were several integration hotspots although all chromosomes were represented. The ITGs identified were found to be enriched in tumor-related terms and pathways using gene ontology and KEGG analysis. In line with this, three of six ITGs tested were found aberrantly expressed in cervical cancer tissues. Among them, it was demonstrated for the first time that *MPPED2* could induce HeLa cell and SiHa cell G1/S transition block and cell proliferation retardation. Moreover, "knocking out" the integrated HPV fragment in HeLa cell line decreased expression of *MYC* located ~500 kb downstream of the integration site, which provided the first experimental evidence supporting the hypothesis that integrated HPV fragment influence *MYC* expression *via* long distance chromatin interaction. Overall, the results of this comprehensive analysis implicated that dysregulation of ITGs caused by viral integration as possibly having an etiological involvement in cervical carcinogenesis.

Brief description: Genes targeted by human papillomavirus (HPV) integration (ITGs) are concentrated in transcriptionally active, gene-dense regions and enriched in cancer-related functional terms and pathways. Expression of *MYC* was decreased by "knocking out" the integrated HPV fragment in HeLa cell lines. *MPPED2*, one of the ITGs, has, for the first time, been identified as being involved in cervical carcinogenesis. This study highlights aberrant functioning of ITGs as a novel phenomenon of potential importance in HPV-related cancers.

Cervical cancer is responsible for 10–15% of female cancer deaths worldwide.[1] Persistent human papillomavirus (HPV) infection, especially with the high-risk types of HPV (HR-HPV) 16 and/or 18, which account for about 70% of cervical cancers, is the predominant cause of cervical cancers.[2–6] HPV usually infect basal cells, and the viral genome is maintained

Key words: HPV, integration, cervical cancer, functional annotation analysis, MYC, MPPED2

Abbreviations: cITGs: ITGs closest to integration sites; dITGs: ITGs directly disrupted by HPV integration; EST: expressed sequence tag; GO: gene ontology; HBV: hepatitis B virus; HPV: human papillomavirus; ITG: integration-targeted gene; KEGG: Kyoto Encyclopedia of Genes and Genomes; RTG: recurrently targeted host genes; TPM: transcripts per million; HR-HPV: high-risk types of HPV; STR: short tandem repeats

Additional Supporting Information may be found in the online version of this article.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Grant sponsor: 973 Program; **Grant number:** 2012CB518900; **Grant sponsor:** National Natural Science Foundation of China; **Grant number:** 81372679; **Grant sponsor:** The 111 Project

DOI: 10.1002/ijc.29872

History: Received 10 Feb 2015; Accepted 9 Sep 2015; Online 29 Sep 2015

Correspondence to: Fengmin Lu or Xiangmei Chen, Department of Microbiology and Infectious Disease Center, School of Basic Medical Sciences, Peking University Health Science Center, 38 Xueyuan Road, Haidian District, Beijing 100191, People's Republic of China, Tel.: [86-10-82805136], Fax: +[86-10-82805136]. E-mail: lu.fengmin@hsc.pku.edu.cn or xm_chen6176@bjmu.edu.cn

What's new?

Human papillomavirus (HPV) integration into the host genome is associated with increased severity of cervical precancer, suggesting that it is an influential event in cervical carcinogenesis. However, whether the genomic sites affected by HPV integration are relevant to cervical cancer remains unclear. In this study, analysis of 499 loci showed preferred HPV integration into intragenic and gene-rich chromosomal sites. Frequently, these sites were located near transcriptionally active regions. In addition, multiple HPV integration "hotspots" were identified, the majority of which contained cancer-related genes, such as *MYC* and the potentially novel tumor suppressor *MPPED2*.

as an episome in precancerous lesions. In most cases, HR-HPV infection is self-limiting, and the virus is eliminated within several months. However, in approximately 10% of cases, the infection results in a transformation process being initiated. HR-HPV DNA is found integrated into the host cell genome in cervical cancers, and, as this process progresses through precancerous lesions to cervical cancer, the frequency of viral genome integration increases.[6] The E6 and E7 proteins encoded by HPV are among the most well-known viral oncoproteins and seem to accelerate the progression of cervical cancer by abrogating key inhibitors of cell proliferation.[7,8] The expression of E6 and E7 is strictly repressed by the virally encoded E2 protein.[9] Numerous studies have shown that integration of HPV normally involves breaking the viral genome in the E1 and E2 regions, and this results in the inactivation of E1 and E2 due to disruption of their open reading frames.[10–12] The integration-mediated disruption of the E2 gene and the consequent up-regulation of E6 and E7 is generally regarded to be a critical step in HPV-related cervical cancers.

The role(s) of tumor-related DNA viruses in cancer pathogenesis is mediated through various mechanisms. In spite of the strong oncogenic potency of some proteins encoded by such viruses, integration-induced aberrant expression or functional change of host genes has also been shown to play a role in virus-related human malignancies. In line with this, our recent study has demonstrated that the function of integration-targeted genes (ITGs), which were disrupted directly by HPV integration or the ones closest to the HPV integration site as described before, has a role in determining the oncogenicity of hepatitis B virus (HBV) integration in hepatocellular cancer.[13] However, in HPV-related cancers, although HPV DNA integration had been recognized as a major promoting step in malignant transformation, the possible role of ITGs has not received much attention. Several studies have reported the detection of HPV integration into chromosomal sites close to where tumor-relevant genes mapped[14,15]; however, an earlier study that reviewed more than 190 integration events did not support the hypothesis that alteration of critical cellular genes can play a major role in HPV-related cervical carcinogenesis.[16]

In this study, to systematically understand the pathogenic role(s) that ITGs might have on cervical carcinogenesis, all the available integration data for HR-HPV types 16 and 18 were collected and reviewed with respect to the characteristics of the loci in the human genome targeted by the integration events. An attempt was also made to interpret the characteristics of the integration sites on human chromosomes and to analyze the distinct functions of ITGs using gene function annotation analysis.

Material and Methods

Data collection

To collect all available data, an extensive PubMed search was performed. The key words used were "HPV" and "integration sites." All published articles detailing either HPV breakpoints or chromosomal loci or genes disrupted by HPV integration were included. The full database was then filtered to include only those clinical samples from HPV16 or HPV18 associated cervical cancers, with those originating from other HPV types being excluded. Those integration events obtained from the same sample by the same laboratory were treated in this study as a single event.

Random sites collection

The total number of human nucleotide was 3,036,303,846 according to UCSC database, and each nucleotide was represented by a different serial number from 1 to 3,036,303,846. To randomly pick up 200 sites for random control, this serial numbers were imported into SPSS17.0 software (SPSS, Chicago, IL), and came out approximately 6.59×10^{-8} (3,036,303,846 encompassed by 200 random sites) various combinations. Then, a random number was selected from the compute filter = (uniform (total number of nucleotide) $\leq 6.59 \times 10^{-6}$), and its corresponding relative nucleotide sites, defined as random nucleotide sites in this manuscript, were extracted for subsequent analysis.

Expressed sequence tag profiles data source

Expressed sequence tag (EST) profiles of cervix tissues were downloaded from the UniGene database (http://www.ncbi.nlm.nih.gov/unigene/). The EST profiles were inferred from EST counts and the cDNA library sources that showed approximate gene expression patterns. The expression level of each gene was scored as transcripts per million (TPM), that is, the number of transcripts of the gene in every one million clones.

Gene functional annotation analysis

DAVID (http://david.abcc.ncifcrf.gov/) was used to perform the gene functional annotation analysis, and the categories of gene ontology (GO) and KEGG Pathways were chosen as background databases. All genes of *Homo sapiens* obtained from the UCSC database (http://genome.ucsc.edu/) were used as the background gene list.

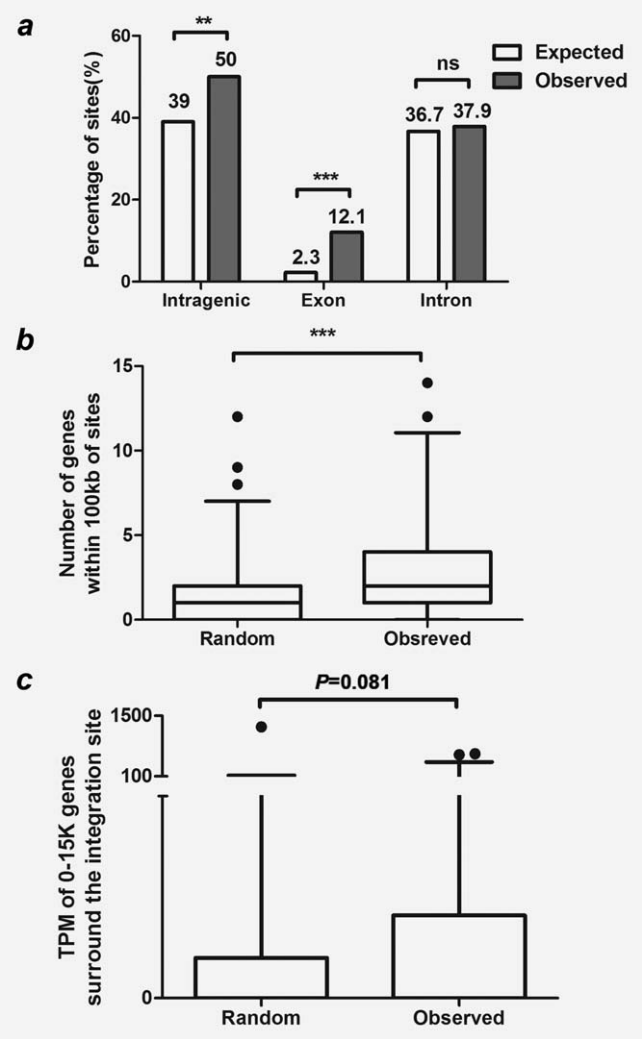

Figure 1. Distribution of HPV integration sites in the human genome. (*a*) The percentage of genes located in intragenic, exon and intron regions. (*b*) Number of genes located within 100 kb up- and downstream of integration sites. Data are presented in boxes and whiskers' style, which represents the medians and ranges of the data. (*c*) The transcriptional activities of host genes surrounding HPV integration sites. "TPM" indicates the expression level of those genes in normal cervix tissues. $^{**}p < 0.01$; $^{***}p < 0.001$.

In vitro gene functional experiments

Thirty-nine cervical cancer tissues and 19 benign lesion tissues were obtained from patients who underwent surgery in Peking University People's Hospital and whose diagnosis had been histologically confirmed. HeLa and SiHa cell lines used in this study were tested in the last 2 months by short tandem repeats (STR) matching analysis of their STR profile using the online STR analysis tools provided by the DSMZ database (http://www.dsmz.de) for human cell lines.

The sgRNA/Cas9 dual expression vector pSpCas9(BB)-2A-GFP was obtained from Addgene (Cambridge, MA). Two specific sgRNA targeted to the HPV genome was designed using the optimized CRISPR design software (http://crispr.mit.edu/). The resulting oligonucleotide sequences are listed as follows: sgRNA5170: Top 5′-CACCGAACTGCAAATGGCCCTACA-3′; Bottom 5′-AAACTGTAGGGCCATTTGCAGTTC-3′. sgRNA36: Top 5′-CACCGCAGGTGGTGGCAATAAGC-3′; Bottom 5′-AAACGCTTATTGCCACCACCTGC-3′. The primers used for detection of knocked out HPV fragment in HeLa cells were as follows: Primer-Fwd: 5′-GTTATTACACAG CTATCAGAGCAA-3′; Primer-Rev: 5′-GGTCTTCTCTGCAAT CCATCTGGAGC-3′. The primers used for detection of wild-type HPV integration in HeLa cells were primer-Fwd and sgRNA5170-Top listed earlier.

HeLa and SiHa cervical cancer cell lines were used for restoring the expression of *MPPED2* using the pLEX-MCS lentivirus expression system (cat no. OHS4735). The plasmid pLEX-MCS-*MPPED2* expressing C-terminally Myc-tagged *MPPED2* was constructed by cloning the full-length coding sequence of *MPPED2* into the eukaryotic expression vector pLEX-MCS. Cell cycle was measured by flow cytometry, with the percentage of cells in G0/G1, S and G2/M phases calculated using the ModFit programs as previously described.[17] The EdU incorporation assay was performed using the Cell-Light EdU DNA cell proliferation kit (RiboBio, Guangzhou, China). Cells (4,000 per well) were grown in 96-well plates, and at 12 hrs after seeding, they were incubated with 100 μl EdU (50 μM). The remaining steps of the assay were carried out according to the manufacturer's instructions, with a fluorescence microscope being used to observe positive cells, which had incorporated EdU. The 3-(4,5-dim-ethylthiazol-2-yl)−2,5-diphenyltetrazolium bromide (MTT) assay method was used for cell proliferation measurements. Cells were seeded at 2,000 cells per well in 96-well plates, and proliferation was measured every 48 hrs by reading the absorbance at 570 nm using a microplate reader as described previously.[17]

Statistical analysis

The χ^2 test and Fisher's exact test were used for statistical analyses to compare variables between two groups based on SAS 9.1 for windows (SAS, Cary, NC). All estimates were accompanied by a 95% confidence interval, and $p < 0.05$ was considered to be statistically significant.

Results

Overview of the data collected

HPV16 and HPV18 are the two most common HR-HPV types found in cervical cancers.[2] After searching the literature, a total of 499 individual integration events with chromosomal locus information were collected from 14 studies, and these were taken into analysis in this study (Supporting Information Tables S1 and S2). They included 379 HPV16 integration events and 120 HPV18 integration events. To investigate the characteristics of HPV integration into the human genome, the integration sites in human chromosomes were analyzed. The methods used collecting the information of HPV integration sites were listed in Supporting Information Table S2. Among the 499 integration events analyzed, 325 events occurred either in the region of a gene or nearby to a gene region. The genes directly disrupted by HPV integration

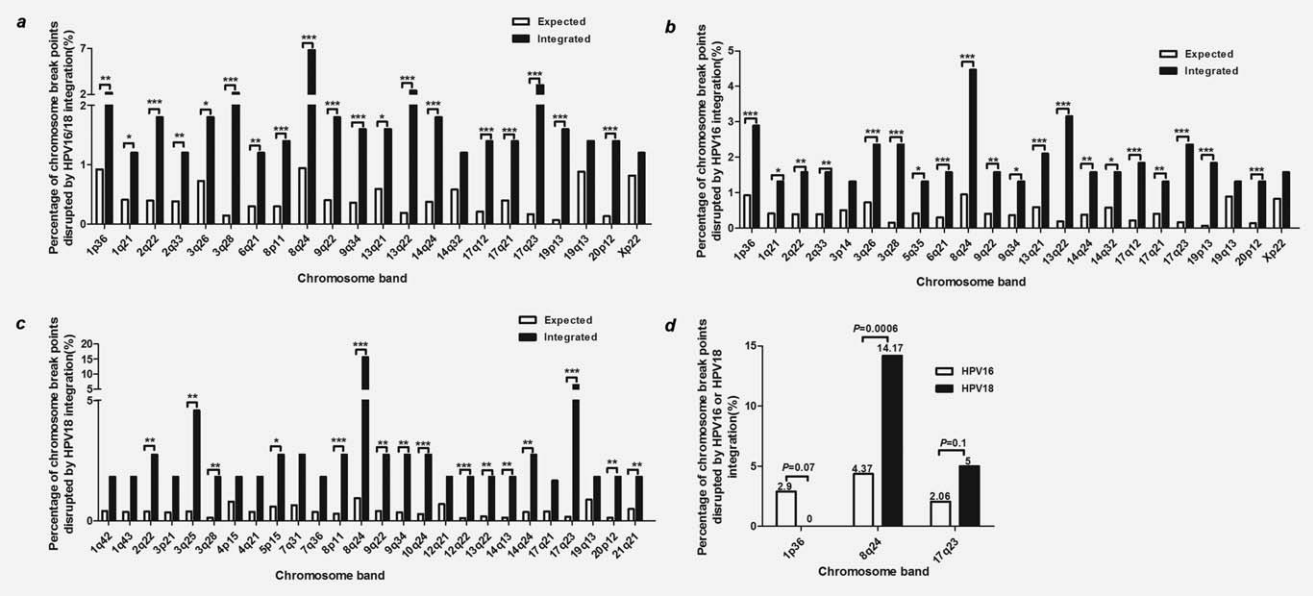

Figure 2. The hotspots of HPV integration into human chromosomes. Comparison of observed chromosome break points in the human genome with expected. Chromosome break points integrated by (*a*) HPV16/18, (*b*) HPV16 and (*c*) HPV18, compared with expected, respectively. (*d*) Different distribution of chromosome break points in different HPV types. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

or those whose transcription start sites were closest to the HPV integration sites were categorized as ITGs as previously reported.[13] A total of 257 ITGs were identified in this study and were used for subsequent analysis. Among these, 38 were named as recurrently targeted host genes (RTGs) because each was targeted by viral integration more than once.

HPV preferentially integrated into intragenic regions and gene-dense regions of the human genome

Given that both HBV and human immunodeficiency virus have been shown to preferentially integrate into the transcription active regions, including intragenic regions and gene-dense regions of human chromosomes,[13,18] it was reasonable to assume that HPV integration might also exhibit such a preference. To clarify if this assumption was correct, 116 integration sites obtained by next generation sequencing derived from two studies (Refs. 5 and 13) with exact positional relationship to a gene were analyzed (Supporting Information Table S1). To avoid any bias, those integration sites identified by other methods were excluded. First, whether the distribution pattern of integration sites was enriched in intragenic regions in human genome was analyzed. As shown in Figure 1*a*, some 50% of the integration sites were located in intragenic regions (including exon and intron regions), which was significantly higher than expected, based on the proportion of the human genome encompassed by such region[13] (50% observed in the integration group *vs.* 39% expected if integration was random; $p < 0.01$; Fig. 1*a*). More detailed analysis of the 116 integration sites revealed that there was a statistically significant preferential localization in exons (12.1% detected in the integration group *vs.* 2.3% expected if integration was random; $p < 0.001$; Fig. 1*a*). By contrast, there was no preferential localization of integration events in introns (37.9% detected in the integration group *vs.* 36.7% expected for random integration; Fig. 1*a*). It was believed that transcriptional active region of chromatins were more accessible for HPV integration. Next, to analyze whether HPV integrates preferentially into the chromatin configuration where genes enriched and were transcriptionally active, the number of genes in a 100 kb region upstream and/or downstream of the 116 precisely localized integration sites were statistically analyzed against 200 sites randomly picked across the whole human genome. A total of 303 genes surrounding the 116 integration sites were catalogued; this is significantly more than the 309 genes identified as surrounding the 200 control sites ($p < 0.001$; Fig. 1*b*). To validate the hypothesis further, the transcription status of host genes surrounding the integration sites were analyzed. For this analysis, TPM values obtained from normal cervix tissue were used to evaluate the transcriptional activity of host cellular genes. Using TPM values from EST profiles, the expression status of the genes surrounding the HPV integration sites were compared with those of the genes surrounding the randomly selected control sites. The expression level of the 79 genes that mapped within 15 kb upstream or downstream of the 116 HPV integration sites were marginally higher than that of the 94 genes located at a similar distance from the 200 randomly selected control sites ($p = 0.081$; Fig. 1*c*).

HPV DNA preferentially integrated into some hotspots in human chromosomes

The results aforementioned demonstrated that HPV preferentially integrates into transcriptionally active regions. It has been suggested that there were hotspots for HPV integration

Table 1. ITGs in hotspots and relationship with tumors

	Integrated times			ITGs in the hotspot and relationship with tumors	
Hotspot	HPV16/18	HPV16	HPV18	Y	N
1p36	11	11	0	***CASZ1, C1orf201, WASF2, ALPQTL2***	*C1orf196*
1q21	6	6	0	***RPS27, USP21, GOLPH3L***	*LOC645166*
2q22	8	5	3	***LRP1B, PABPC1P2***	*LOC151128*
2q33	6	6	0	***ORC2, PARD3B, RTN4IP1***	
3q26	9	9	0	***MECOM, MDS1, ATP11B, SOX2***	*TNIK*
3q28	11	9	2	***TP63, LEPREL1, FAM79B, CLDN16, CLDN1***	
6q21	6	6	0		*FIG4*
8p11	7	4	3		*ZNF703*
8q24	34	17	17	***POU5F1B, MYC, AGO2, GPAA1, RASSF6, DEPTOR***	*LOC727677*
9q22	8	5	3	***FANCC***	*C9orf3*
9q34	8	5	3	***Notch1, EGFL7***	*SLC25A25*
13q21	8	8	0	***DACH1***	
13q22	12	10	2	***KLF5, KLF12***	
14q24	9	6	3	***RNGTT, RAD51B, GLI2***	
14q32	6	6	0	***BCL11B, TNFAIP2, SLC24A4***	
17q12	7	7	0	***MED24, PLXDC1, IKZF3, GRB7, HNF1B***	*C-CR7*
17q21	7	5	2	***ERBB2, BRCA1, STARD3, EIF1, RARA***	*KRT40*
17q23	14	8	6	***TMEM49, CGI-147***	*TUBD1*
19p13	8	7	1	***PRKACA, NANOS3, ARID3A, HMHA1***	*MOB3A, ZNF506, GATAD2A*
19q13	6	4	2	***CEACAM5***	*LOC126235, Q8N6Q3*
20p12	7	6	1	***MACROD2, FLRT3***	
Xp22	6	6	0	***STS, NHS, PDK3***	*FAM48B2, VCX2*

The bold genes have been reported to be tumor-related based on NCBI database, whereas the underlined genes were recurrently integration-targeted genes.
Abbreviations: Y: Yes; means that these genes had been reported to be a tumor-related gene; N: no; means that there was no report about the relationship between this gene and tumors.

in specific human chromosomes.[11,16] To provide further evidences and also to identify whether there were any previously unrevealed hotspots, the distribution characteristics of the integration sites across the human genome were investigated based on this relatively large study cohort. In this study, all integration sites were normalized to standard "chromosome bands." For example, they were normalized to "8q24" irrespective of whether the integration location was "8q24.1" or "8q24.2." To reduce the bias caused by length difference between different chromosome bands, the "chromosome band" was represented by ratio based on the proportion of the human genome encompassed by each chromosome band. The integration sites included in the dataset were assigned to all 23 human chromosomes (Supporting Information Fig. S1), with some regions being targeted more frequently by HPV16/18 integration than expected (Fig. 2*a*). Among these hotspots, there were 34 integration events localized to 8q24, making it the most frequently affected chromosome band, followed by 17q23 (14 events), 13q22 (12 events), 1p36 and 3q28 (11 events each). When the hotspot distribution of HPV16 and HPV18 integration sites were analyzed separately, the overall hotspot distribution remained (Figs. 2*b* and 2*c*). However, there were some differences between the two viruses; thus, 1p36 was only disrupted by HPV16 (integrated by HPV16 *vs.* integrated by HPV18; $p = 0.07$), whereas the HPV18 genome integrated more frequently into 8q24 (integrated by HPV16 *vs.* integrated by HPV18; $p = 0.0006$) and 17q23 (integrated by HPV16 *vs.* integrated by HPV18; $p = 0.1$; Fig. 2*d*).

Most of the ITGs in chromosome hotspots and the RTGs were functionally cancer related

A total of 80 ITGs were mapped to the hotspots as described earlier. It would be interesting to investigate whether these ITGs were functionally cancer-related genes. Here, the term "cancer-related gene" was defined in the National Center for Biotechnology Information (NCBI) database as oncogene or tumor suppressor gene, or that has been reported to be involved in carcinogenesis. As expected, 61 of the 80 ITGs

Table 2. Recurrent targeted genes and relationship with tumors

Gene symbol	Integrated times	Cancer related	Gene function	Reference (PMID)
MYC	13	Y	Negative regulation of apoptotic process; positive regulation of cell proliferation	NCBI gene
TMEM49 (*VMP1*)	5	Y	*VMP1* is involved in multiple cancers as a tumor suppressor gene through regulating autophagy, inhibiting metastasis and proliferation of hepatocellular carcinoma	24549370; 24365149; 22971212
FANCC	4	Y	DNA repair	23028338; 17490643
KLF5	4	Y	Acts as an oncogene in kinds of tumors through positive regulation proliferation	24626089; 23913682
C9orf3	3	N		
BCL11B	3	Y	*BCL11B* tumor suppressor inhibits *HDM2* expression in a p53-dependent manner; mutation and decreased expression of this gene is related with multiple cancers	22450536; 22245141; 21878675
COX4I2	3	Y		22320183
GLS	3	Y	Promotes glioma cell lines proliferation	24276018
HTERT	3	Y	Telomerase reverse trancriotitase	NCBI gene
LOC727677	3	N		
LRP1B	3	Y	Inhibits cell migration	12004004; 23521319
NR4A2	3	Y	*NR4A2* emerges as an important nuclear factor linking gastrointestinal inflammation and cancer	23821160; 23322982
TP63	3	Y	Keratinocyte proliferation; negative regulation of apoptotic process	NCBI gene
RAP2B	3	Y	Belongs to a family of *RAS*-related genes; over expressed in tumors	17316888; 16204058
RPS27	3	Y	Regulates apoptosis	17057733; 2117008.
PABPC1P2	3	N		
ERBB2	2	Y	Amplification and/or overexpression of this gene has been reported in numerous cancers, including breast and ovarian tumors	NCBI gene
CEACAM5	2	Y	Negative regulation of apoptotic process	NCBI gene
FHIT	2	Y	Aberrant transcripts from this gene were found in about half of all esophageal, stomach and colon carcinomas through apoptotic signaling pathway by p53 class mediator	24556720; 24370550
AFP	2	Y	Marker of liver tumor	24425104
CHS1	2	N		
DRAP1	2	N		
IL8	2	Y	Negative regulation of cell proliferation; cell cycle arrest	NCBI gene
KLF12	2	Y	Enhances tumor cell invasive potential	19588488
KLHL3	2	N		
LEPREL1	2	Y	It can be hypermethylated and acts as a tumor suppressor in breast cancer	19436308
LIPC	2	Y		23343765

Table 2. Recurrent targeted genes and relationship with tumors (Continued)

Gene symbol	Integrated times	Cancer related	Gene function	Reference (PMID)
MACROD2	2	Y	Deletion of this gene is found in cancer genomes	23805207
MECOM	2	Y	Positive regulation of proliferation	22372463
SLC25A36	2	N		
ORC2	2	Y	Overexpressed in kinds of tumors; G1/S transition of mitotic cell cycle	16163736; NCBI gene
FER1L3	2	N		
POU5F1P1	2	Y	Overexpressed in kinds of tumors and was correlated with metastasis	20017164; 21748294
RAD51B	2	Y	Overexpression of this gene was found to cause cell cycle G1 delay and cell apoptosis	NCBI gene
DRAP1	2	N		
RPS27	2	Y	Enhances the DNA repair capacity; promote apoptosis through p53 signaling pathway	23826192; 17057733
SH3PXD2A	2	Y	Promotes prostate cancer metastasis and invasion in multiple cancer types	24174371; 23873940
AGO2	2	Y	Promotes angiogenesis, metastasis	24886719; 24427355

Abbreviations: Y: Yes; means that these genes had been reported to be a tumor-related gene. N: no; means that there was no report about the relationship between this gene and tumor.

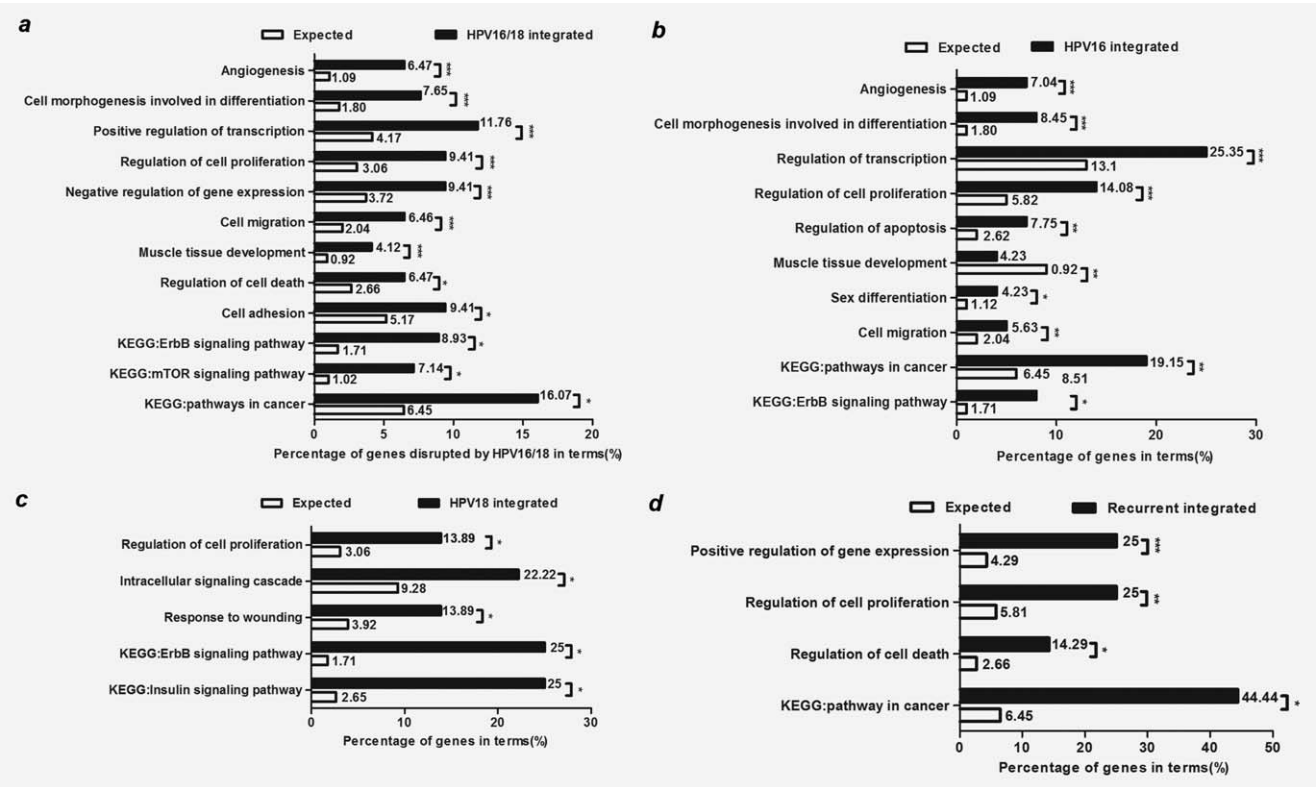

Figure 3. Functional annotation analysis of ITGs. Comparison of gene oncology and KEGG pathway analysis between observed genes integrated by (*a*) HPV16/18, (*b*) HPV16, (*c*) HPV18 and (*d*) recurrent targeted by HPV16/18 integration and expected, respectively. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

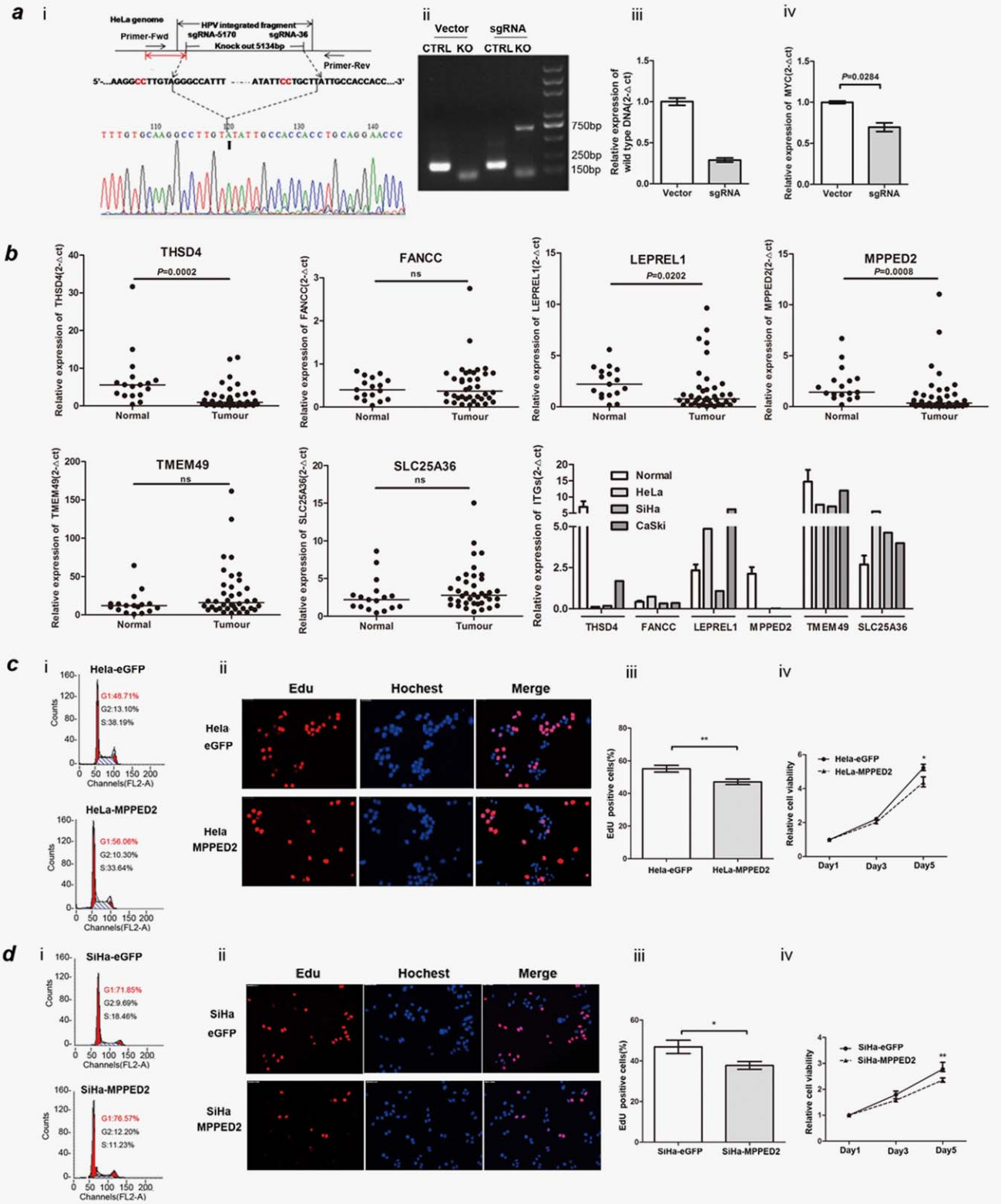
a
i
HeLa genome
HPV integrated fragment
Primer-Fwd
sgRNA-5170
sgRNA-36
Knock out 5134bp
Primer-Rev
5'-...AAGGCCTTGTAGGGCCATTT ATATTCCTGCTTATTGCCACCACC...-3'
ii
Vector
sgRNA
CTRL KO CTRL KO
750bp
250bp
150bp
iii
Relative expression of wild type DNA(2-△ct)
Vector
sgRNA
iv
Relative expression of MYC(2-△ct)
P=0.0284
b
THSD4
P=0.0002
FANCC
ns
LEPREL1
P=0.0202
MPPED2
P=0.0008
TMEM49
SLC25A36
Normal
Tumour
Relative expression of ITGs(2-△ct)
HeLa
SiHa
CaSki
c
i
Hela-eGFP
HeLa-MPPED2
Counts
Channels(FL2-A)
ii
Edu
Hochest
Merge
Hela eGFP
Hela MPPED2
iii
EdU positive cells(%)
iv
Relative cell viability
Day1
Day3
Day5
d
SiHa-eGFP
SiHa-MPPED2
SiHa eGFP
SiHa MPPED2

Figure 4.

analyzed were tumor-related genes (Table 1), which implicated that HPV DNA fragment integrated preferentially into genomic hotspots where tumor-related genes located. Presumably, the cells in which such integration events occurred acquired survival and proliferation advantages during cervical carcinogenesis.

In addition to these 80 ITGs, 38 RTGs recurrently targeted by HPV integration were also reviewed. *MYC* came out to be the gene most frequently affected by viral integration (13 events), followed by trans-membrane protein 49 (*TMEM49*; five events) and *FANCC* and *KLF5* (four events). A further 12 genes were affected three times, leaving 22 of the RTGs affected twice in the dataset analyzed. Consistent with the observation that genes in chromosome hotspots were tumor related, 29 of the 38 RTGs have also been reported to be tumor related (Table 2).

Genes affected by HPV integration were distinctly clustered in tumor-related terms

To gain further insights into the possible roles that the ITGs may play in HPV-related cervical cancer, they were all subjected to functional annotation analysis using DAVID software. In total, 226 of the 257 ITGs could be identified in DAVID analysis. GO analysis revealed that the genes identified were significantly enriched in the following nine terms: "angiogenesis," "positive regulation of transcription," "cell morphogenesis involved in differentiation," "negative regulation of gene expression," "cell migration," "muscle tissue development," "regulation of cell proliferation," "cell adhesion" and "regulation of apoptosis and cell death" ($p < 0.05$; Fig. 3*a*). Further, KEGG pathway annotation analysis showed that three pathways were significantly clustered: the "ErbB signaling pathway," the "mTOR signaling pathway" and the "Pathways in cancer" ($p < 0.05$; Fig. 3*a*), all of which are tumor-related pathways. The term "Bladder cancer" was also marginally enriched ($p = 0.075$; data not shown). To analyze whether there was any difference between HPV16 and HPV18, GO and KEGG analysis was performed on these two groups separately. The ITGs involving integration of HPV18 were enriched in "insulin signaling pathway" terms, whereas HPV16/18 or HPV16 integrations did not show this association (Figs. 3*b* and 3*c*). The RTGs were also subjected to GO and KEGG analysis, significant enrichment in "regulation gene expression," "regulation of cell proliferation" and "regulation of apoptosis and cell death," as well as the term of "pathways in cancer" were identified ($p < 0.05$; Fig. 3*d*).

According to their precise integration site on the chromosome, the ITGs were further grouped into those in which the insertion of viral DNA resulted in *d*isruption of the cellular gene (dITGs) with the viral DNA being inserted directly into the intragenic region of the gene, and those in which integration occurred *c*lose to but not actually within the cellular gene (cITGs). GO and KEGG analysis was performed on these groups separately. This revealed that dITGs were enriched in the following terms: "low-density lipoprotein binding," "transcription factor activity," "intracellular signaling cascade" and "keratinocyte proliferation" ($p < 0.05$; Supporting Information Table S3), whereas cITGs were enriched in: "transcription factor activity," "blood vessel morphogenesis" and "Wnt signaling pathway" ($p < 0.05$). In addition, the cITGs also showed enrichment in the terms "negative regulation of apoptosis" ($p = 0.058$), "epidermis morphogenesis" ($p = 0.061$) and "positive regulation proliferation" ($p = 0.084$) to some extent (Supporting Information Table S3). As expected, PubMed search indicated that tumor suppressors, such as *TP63* and *CDH13*, which negatively regulate keratinocyte proliferation, were linked to the gene structure destroyed dITGs group. Whereas, the cITGs contained known oncogenes such as *BCL11B, NR4A2, MYC* and *ARHGDIA* that could negatively regulate apoptosis.

"Knocking out" of the integrated HPV fragment decreased the expression of *MYC* in HeLa cells

It has been suggested that in the HeLa cell line, the over expression of *MYC* was modulated by the integrated HPV genome ~500 kb upstream of it, *via* a mechanism involving

Figure 4. The cancer relativity of HPV ITGs. (*a*-i) Construction strategy of dual sgRNAs used for "knocking out" the integrated HPV fragment in HeLa cells with CRISPR/Cas9. A 5,134 bp integrated HPV fragment was expected to be "knocked out" by the dual sgRNAs. If this integrated HPV fragment was successfully removed by sgRNA cleavage, a ~750 bp PCR product would be detected using Primer-Fwd and Primer-Rev. The region covering the viral–host junction sequence shown between the red arrows amplified was used for detecting cleavage efficiency. For amplification of this region, primers (Primer-Fwd and sgRNA5170-Top) were set in the host and virus genome, respectively, and the PCR amplicon expected was ~200 bp. (*a*-ii) "Knock-out" of the integrated HPV fragment was confirmed by PCR assay in HeLa cells transfected with sgRNA or vector plasmids. The expected ~750 bp PCR amplicon, labeled sgRNA-KO, was detected using Primer-Fwd and Primer-Rev, indicating the successful knock out of the integrated HPV fragment. Direct sequencing of the junction region of the ~750 bp amplicon is shown in (*a*-i). The ~200 bp PCR amplicon covering the viral–host junction sequence was amplified using Primer-Fwd and sgRNA5170-Top and used as control. (*a*-iii) Cleavage efficiency was evaluated by real-time PCR using the same primer pair covering the viral–host junction. (*a*-iv) Expression of *MYC* was detected by real-time RT-PCR in HPV knocked out HeLa cells and nontransfected controls. (*b*) Expression of *THSD4*, *FANCC*, *LEPREL1*, *MPPED2*, *TMEM49* and *SLC256A36* in normal cervix tissues, cervical cancer tissues and HeLa, SiHa and CaSki cell lines was detected by real-time PCR. (*c*-i, *d*-i) The proportion of cells at different stages of the cell cycle was detected by flow cytometry in HeLa and SiHa cells with ectopic expression of *MPPED2*. (*c*-ii, *d*-ii) The effect of *MPPED2* on cell proliferation was evaluated using EdU incorporation assays in HeLa and SiHa cell lines. (*c*-iii, *d*-iii) Quantitative analysis of EdU incorporation assays. (*c*-iv, *d*-iv) The ability of cell growth in HeLa and SiHa cell lines with ectopic expression of *MPPED2* was measured by MTT assays. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

long distance chromatin interaction.[19] To confirm whether removal of the integrated HPV fragment had any effect on *MYC*, the CRISPR/Cas9 system was used to remove the integrated HPV DNA fragment. Then, the subsequent change in *MYC* expression was detected by real-time reverse transcription polymerase chain reaction (RT-PCR). Figure 4*a*-i shows the construction strategy used, with two sgRNAs targeted to the integrated HPV fragment. To confirm that the dual sgRNA could mediate cleavage of the integrated HPV fragment in HeLa cells, PCR was performed using different primer sets (Fig. 4*a*-i), and the products were detected by agarose gel electrophoresis. First, the specific cleavage of the integrated HPV DNA was analyzed. Forty-eight hours after transfection of the HeLa cells with the sgRNA/Cas9 dual expression vectors, cellular DNA was extracted and PCR was performed using Primer-Fwd and Primer-Rev (Fig. 4*a*-i). As expected, a ~750 bp length PCR amplicon was detected (Fig. 4*a*-ii, shown as "KO"), indicating that the dual sgRNAs had promoted the removal of the 5,134 bp integrated HPV fragment from the HeLa cell genome. Next, the PCR product was sequenced, to confirm that the dual sgRNAs had precisely guided site-specific cleavage by the CRISPR/Cas9 system (Fig. 4*a*-i). To obtain information on cleavage efficiency, primers located in the viral and host genome were designed, which guaranteed the specific amplification of the viral–host junction sequence (Fig. 4*a*-i). As expected, PCR assay revealed a marked decrease of the viral–host junction fragment in sgRNA-transfected cells as compared with its control vector-transfected counterpart (Fig. 4*a*-ii, shown as "CTRL"). Consistently, a cleavage efficacy of approximately 70% was confirmed by real-time PCR (Fig. 4*a*-iii). Subsequently, the change of *MYC* expression brought about by the CRISPR/Cas9-mediated removal of the integrated HPV fragment was evaluated by real-time RT-PCR, and the expression of *MYC* was decreased by approximately 30% in the sgRNA-transfected cells (Fig. 4*a*-iv).

MPPED2 may act as a tumor suppressor gene in cervical cancer

To investigate whether there were novel candidate cervical cancer-related genes from those ITGs identified in this study, the expression of the six ITGs (*THSD4*, *FANCC*, *LEPREL1*, *MPPED2*, *TMEM49*, and *SLC256A36*) was detected and compared by real-time PCR in normal cervix tissue, cervical cancer tissues and cell lines. Of them, *LEPREL1*, *TMEM49*, *SLC25A36* and *FANCC* were repeatedly targeted by HPV integration, and so we speculated that these genes may have much more chance to play roles in cervical cancer. Because most of the ITGs were affected by HPV integration only for once, the role of these genes in cervical cancers should also attract our attention. To choose candidate genes from them, microarray-based expression profiles data were used to preliminary evaluate the aberrant expression of those ITGs[20]; *MPPED2* and *THSD4* were two genes, which were abnormally expressed and have never been reported in cervical cancer. Of them, *THSD4, LEPREL1* and *MPPED2* were significantly decreased in tumor tissues and cell lines compared with that in normal tissues (Fig. 4*b*). Among the aberrant expressing genes, *MPPED2* showed the greatest down-regulation in cervical cancers as well as in cell lines compared with its expression level in normal tissue. *MPPED2* has also been suggested as a tumor suppressor in neuroblastoma, because it caused both tumor cell apoptosis and cell cycle retardation *in vitro*.[21] However, the biological role of *MPPED2* as a tumor suppressor in cervical cancer has remained elusive. Therefore, its function was studied *in vitro* using two cervical cell lines (HeLa and SiHa). Analysis of the cell cycle progression by flow cytometry showed that ectopic expression of *MPPED2* in either of these cell lines caused a significant block in the G1/S phase transition. As shown in Figures 4*c*-i and 4*d*-i, cell populations were increased in G1 phase and decreased in S phase, compared with control cells. The cell growth retardation was further confirmed using an EdU incorporation assay, which showed that ectopic *MPPED2* expressing cells incorporated less EdU when compared with control cells (Figs. 4*c*-ii–iii and 4*d*-ii–iii). Consistent with these data, an MTT assay revealed that restoration of *MPPED2* expression suppressed cell proliferation (Figs. 4*c*-iv and 4*d*-iv). Taken together, these results implicated *MPPED2* as a possible tumor suppressor in cervical cancer. The cause of *MPPED2* down-regulation in cervical cancer needs further study. It is worthwhile to emphasize that the role of HPV integration should not be over interpret.

Discussion

Following the initial discovery of HPV DNA in the human genome, various studies have demonstrated its role in cervical carcinogenesis. In recent years, viral genome DNA integration-induced aberrant expression or functional change of host cellular genes has also been suggested as playing a role in a number of virus-related human malignancies.[13,18] To provide evidence that the same mechanism was also involved in HPV-related cervical cancer, the pattern of HPV integration into the human genome was analyzed. The large-scale analysis undertaken in this study involved 499 integration events and allowed the conclusion that HPV DNAs prefer to integrate into intragenic and gene-dense regions similar to what has been observed in HBV[13] and human immunodeficiency virus integration.[18] Importantly, viral integration into such regions increases the potential influence that the HPV integration events may have on host cellular genes. Besides, we searched the NCBI database and found that the insights that frequency of intragenic integration and integration into transcriptionally active chromosomal regions were firstly put forward. For the HPV integration hotspots, through the comprehensive analysis, we confirmed current hotspots such as 2q22, 2q33, 3q28, 8q24 9q34, 13q21-22 and 17q21 reported by previous articles and found out some new hotspots such as 1p36, 3q26, 9q22, 17q12 and 20p12.

Analysis of the chromosomal hotspot ITGs in this cohort revealed that the majority of cellular genes identified were

tumor related. Supporting Information Figure S2 illustrates in detail the physical map of the genes affected by HPV integration at several of the chromosomal hotspots. Among those identified, *MYC* and *TMEM49* are illustrative. The cellular oncogene *MYC*, located in the chromosomal hotspot 8q24, was the gene most recurrently targeted by viral integration. Moreover, a recent study has strongly suggested that through mechanism called long distance chromatin interaction, the HPV18 genome integrated ~500 kb upstream of *MYC* markedly activated its transcription, which resulted in the robust growth characteristics of the HeLa cell line.[19] Consistent with this, as shown in this study, CRISPR/Cas9-mediated "knock out" of the integrated HPV fragments resulted in significantly decreased expression of *MYC*. This result indicates that integration of HPV in human genome indeed exhibited influence on ITGs expression. *TMEM49* is located in the second commonest HPV integration site 17q23.1. This gene encodes a protein that is located in the plasma membrane and plays an important role in cell–cell connections and tight junction formation. It has been reported to act as a tumor suppressor in a variety of tumors, including breast,[22] pancreas cancer,[23] as well as in hepatocellular cancer.[24] Because all five integration events categorized in this cohort were located in the intragenic region of *TMEM49*, it will be interesting to detect in the future whether this gene were commonly mutant in cervical cancers.

The analysis of the RTGs identified in this large-scale review study also supports our hypothesis. The fact that the majority of RTGs seem to be functionally cancer related strongly suggests that integration-induced dysregulation of host cellular genes involved in cervical carcinogenesis represents a common oncogenic mechanism of HPV integration. The characteristics of the RTGs prompted the functional annotation of all the ITGs. As expected, functional annotation revealed that most of the ITGs were enriched in tumor-related terms and pathways, including GO terms of "regulation of proliferation" and "regulation of apoptosis" and the KEGG terms of "ErbB signaling pathway," "mTOR signaling pathway" and "pathway in cancer" term. In consistent with our results in the current study, a recent report also showed that ITGs were enriched in tumor-related KEGG terms.[25]

The analysis undertaken in this study strongly indicates that dysregulation of ITGs contributes to cervical carcinogenesis. Four patterns of how this occurs can be put forward. First, in the case of tumor suppressor genes such as *TMEM49* and *CASZ1* (castor zinc finger 1),[15] irrespective of whether the integration event occurred in introns or exons, a decrease in expression and/or a loss of function of the genes would occur (Supporting Information Figs. S2*a* and S2*b*). Second, integration events that result in the formation of viral–cellular fusion transcripts could be expected to promote oncogenicity. Third, in the case of oncogenes, the majority of integration events were located within promoter or upstream regions, and this would be expected to trigger the transcriptional activation of the gene (Supporting Information Fig. S2*d*). Finally, vicinal amplification or deletion close to integration sites may also lead to de-regulation the host genes. The first two were termed as dITGs, whereas the last two, to some extent, could be termed as cITGs in this study. Perfectively in concordance, different enriched terms of dITGs and cITGs were revealed by GO and KEGG analysis, with dITGs enriched in tumor suppressor related terms, whereas cITGs mostly enriched in oncogene-related terms. In addition, functional restoration experiments in this study implicated that *MPPED2*, one of the ITGs, acted as a potential new tumor suppressor in cervical cancer.

In conclusion, through a systematic comprehensive review and experimental investigation, this study has revealed that the viral integrations were prone to occur into transcriptionally active regions of the cellular genome. The integration-targeted host cellular genes were found to be enriched in tumor-related terms, indicating that alteration of human genome and functional aberration of ITGs by HPV integration might be a second mechanism that is relevant to HPV's contribution to cervical oncogenesis.

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011; 61:69–90.
2. Munoz N, Bosch FX, de Sanjose S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003; 348:518–27.
3. Castellsague X. Natural history and epidemiology of HPV infection and cervical cancer. *Gynecol Oncol* 2008; 110:S4–7.
4. de Sanjose S, Quint WG, Alemany L, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol* 2010; 11: 1048–56.
5. Li N, Franceschi S, Howell-Jones R, et al. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: variation by geographical region, histological type and year of publication. *Int J Cancer* 2011; 128:927–35.
6. de Miguel FJ, Sharma RD, Pajares MJ, et al. Identification of alternative splicing events regulated by the oncogenic factor SRSF1 in lung cancer. *Cancer Res* 2014; 74:1105–15.
7. Werness BA, Levine AJ, Howley PM. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* 1990; 248:76–9.
8. Ghittoni R, Accardi R, Hasan U, et al. The biological properties of E6 and E7 oncoproteins from human papillomaviruses. *Virus Genes* 2010; 40:1–13.
9. Howley PM, Munger K, Werness BA, et al. Molecular mechanisms of transformation by the human papillomaviruses. *Princess Takamatsu Symp* 1989; 20:199–206.
10. Corden SA, Sant-Cassia LJ, Easton AJ, et al. The integration of HPV-18 DNA in cervical carcinoma. *Mol Pathol* 1999; 52:275–82.
11. Xu B, Chotewutmontri S, Wolf S, et al. Multiplex Identification of human papillomavirus 16 DNA integration sites in cervical carcinomas. *PLoS One* 2013; 8:e66693.
12. Dall KL, Scarpini CG, Roberts I, et al. Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. *Cancer Res* 2008; 68: 8249–59.
13. Li X, Zhang J, Yang Z, et al. The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. *J Hepatol* 2014; 60:975–84.
14. Reuter S, Bartelmann M, Vogt M, et al. APM-1, a novel human gene, identified by aberrant co-transcription with papillomavirus oncogenes in a cervical carcinoma cell line, encodes a BTB/POZ-zinc finger protein with growth inhibitory activity. *EMBO J* 1998; 17:215–22.
15. Schmitz M, Driesch C, Beer-Grondke K, et al. Loss of gene function as a consequence of human

papillomavirus DNA integration. *Int J Cancer* 2012; 131:E593–602.
16. Wentzensen N, Vinokurova S, von Knebel DM. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res* 2004; 64: 3878–84.
17. Lv J, Zhu P, Yang Z, et al. PCDH20 functions as a tumour-suppressor gene through antagonizing the Wnt/beta-catenin signalling pathway in hepatocellular carcinoma. *J Viral Hepat* 2015;22: 201–11.
18. Wagner TA, McLaughlin S, Garg K, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* 2014; 345:570–3.
19. Adey A, Burton JN, Kitzman JO, et al. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* 2013; 500:207–11.
20. Yali Z, Rork K, Bin N, et al. Gene expression analysis of preinvasive and invasive cervical squamous cell carcinomas identifies HOXC10 as a key mediator of invasion. *Cancer Res* 2007; 67:10163–72.
21. Liguori L, Andolfo I, de Antonellis P, et al. The metallophosphodiesterase Mpped2 impairs tumorigenesis in neuroblastoma. *Cell Cycle* 2012;11:569–81.
22. Sauermann M, Sahin O, Sultmann H, et al. Reduced expression of vacuole membrane protein 1 affects the invasion capacity of tumor cells. *Oncogene* 2008; 27:1320–6.
23. Qian Q, Zhou H, Chen Y, et al. VMP1 related autophagy and apoptosis in colorectal cancer cells: VMP1 regulates cell death. *Biochem Biophys Res Commun* 2014; 443:1041–7.
24. Guo L, Yang LY, Fan C, et al. Novel roles of Vmp1: inhibition metastasis and proliferation of hepatocellular carcinoma. *Cancer Sci* 2012; 103: 2110–9.
25. Hu Z, Zhu D, Wang W, 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* 2015; 47:158–63.