To better understand the pathogenesis of cervical intraepithelial neoplasia and cervical cancer identifies potential targets for cervical cancer early warning

Jian Huang,1,2 Zhaoyang Qian,3,4 Yuhua Gong,5 Yanzhou Wang,6 Yanfang Guan,5 Yingxin Han,1 Xin Yi,3 Wanqiu Huang,1 Liyan Ji,5 Jiajia Xu,3,4 Mengyuan Su,3,4 Qing Yuan,1 Shujian Cui,2 Jinling Zhang,7 Chaohui Bao,1 Weilong Liu,8 Xi Chen,3,4 Ming Zhang3,4 Xiaohuan Gao,3,4 Renhua Wu,3,4 Yinxin Zhang,1 Huicheng Xu,6 Shida Zhu,4 Hongmei Zhu,3,4 Ling Yang,5 Xun Xu,4 Pingyu Zhou,9 Zhiqing Liang6

ABSTRACT

Background To better understand the pathogenesis of cervical cancer (CC), we systematically analysed the genomic variation and human papillomavirus (HPV) integration profiles of cervical intraepithelial neoplasia (CIN) and CC.

Methods We performed whole-genome sequencing or whole-exome sequencing of 102 tumour-normal pairs and human papillomavirus probe capture sequencing of 45 CCs, 44 CIN samples and 25 normal cervical samples, and constructed strict integrated workflow of genomic analysis.

Results Mutational analysis identified eight significantly mutated genes in CC including four genes (PATT, MLL3, MLL2 and FADD), which have not previously been reported in CC. Targetable alterations were identified in 55.9% of patients. In addition, HPV integration breakpoints occurred in 97.8% of the CC samples and 70.5% of the CIN samples and 42.8% of the normal cervical samples with HPV infection. Integrations of high-risk HPV strains in CCs, including HPV16, 18, 33 and 58, also occurred in the CIN samples. Moreover, gene mutations were detected in 52% of the CIN specimens, and 54.8% of these mutations occurred in genes that also mutated in CCs.

Conclusion Our results lay the foundation for a deep understanding of the molecular mechanisms and finding new diagnostic and therapeutic targets of CC.

INTRODUCTION

Human papillomavirus (HPV) infection can lead to a variety of human cancers, including cervical cancer (CC), which is the fourth most common cause of cancer and a leading cause of mortality in women worldwide. It has been estimated that approximately 989,000 Chinese women were diagnosed with CC and 305,000 died from it in 2015. Despite improved multidisciplinary treatments, the prognosis of advanced CC is still poor, with a 5-year survival rate of approximately 15% for late-stage patients.

CC screening as a secondary prevention is important in CC control. HPV DNA and liquid-based cytology have been widely used in CC screenings; however, there are limitations to both approaches. Many studies have reported the use of new approaches, such as assessing E6/E7 mRNA and protein expression or p16ink4a overexpression, in CC screening. The identification and validation of new molecular biomarkers play an important role in CC prevention. Previous studies have identified recurrent genetic mutations in PIK3CA, FBXW7, EP300, MAPK1, HLA-B, NFE2L2, TP53, ERBB2, ELF3 and CFBFB in CC. However, the patterns of cellular alterations and virus integrations in cervical lesions, especially in cervical intraepithelial neoplasia (CIN), are largely unknown.

Treatments for CC are mainly based on surgery and chemoradiotherapy. Recent advances in targeted therapies against specific somatic alterations have transformed the management of cancers in general. The discovery of new therapeutic targets could improve the current strategies to combat CC, such as agents targeting the PI3K pathway. Thus, it is necessary to evaluate genetic alterations in cancer and to subclassify cancers based on genetic alterations.

Here, we systematically analysed the genomic variation profiles of 102 tumour-normal pairs and 25 CIN specimens with matched germline DNA. Somatic mutations, copy number alterations (CNAs), genomic rearrangements and HPV integration events were identified and used to define the genomic landscape in CC and to find potential biomarkers for the screening and treatment of CC.

METHODS

Samples

This project and protocols were conducted according to the Declaration of Helsinki. Tumour tissues and matched adjacent normal tissues or peripheral blood were collected. Tissue cells of normal or CIN were collected from female patients attending a sexually transmitted disease clinic at Shanghai Skin Disease Hospital. Sample collection, HPV typing, liquid-based cytology and biopsy were performed under the guidelines or demands of the patients with appropriate informed consent.

Next-generation sequencing and data analysis

Whole-genome sequencing (WGS) or whole-exome sequencing (WES), HPV probe-capture sequencing (HPCS) and transcriptome using RNAseq were performed as described in the online supplementary materials. Read alignment and processing were performed using the BWA,11 Picard (http://broadinstitute.github.io/picard/) and GATK tools.12 We detected somatic single nucleotide variants (SNVs), insertions and deletions (InDels), CNAs and structural variations (SVs) using MuTect,13 Strelka,14 Varscan2,15 CLImmAT16 and Clipping Reveals Structure (CREST).17 MutSigCV and GISTIC V.2.0 were employed to analyse the genes significantly affected by mutations and CNAs.

Statistical analysis

P values were calculated using Fisher’s exact test, t-test, Mann-Whitney U test and Delong’s test as indicated. Receiver operating characteristic (ROC) curves with the area under the curve (AUC) were performed to evaluate the specificity and sensitivity of HPV testing for detecting CC using pROC package.18

Data availability

All the sequence data have been deposited in the database NCBI SRA: SRA315538 and the project number PRJNA305342.

See the online supplementary materials and methods for more detailed methods.

RESULTS

Profiles of somatic mutations in CCs

To explore the profiles of molecular variants in CC samples, we selected 102 paired samples, including tumours (95 HPV-positive cases and 7 HPV-negative cases) and matched-normal tissues (n=8) or peripheral blood (n=94) (see online supplementary table S1). We then performed WES analysis of 76 tumour-normal sample pairs using the Illumina HiSeq 2000 platform and the Ion Proton platform, WGS analysis of 27 tumour-normal samples using the HiSeq X-Ten platform and the complete genomics (CG) platform (one sample sequenced by WES and WGS), and transcriptome analysis of 6 tumour-normal samples (all six samples overlapped with WGS/WES) using the HiSeq 2000 platform (see online supplementary figure S1 and supplementary table S1). The average depth of coverage was −100× for WES data and −30× (HiSeq X-Ten) or 100× (CG) for WGS (see online supplementary table S2), respectively.

Using multiple tools, we identified 17034 somatic mutations in the exons and splice regions, including 8479 missense, 3755 synonymous, 691 nonsense, 166 splice, 40 in-frame InDels, 112 frameshift, 11 initiation-site loss, 10 read-through and 3770 UTR mutations (see online supplementary table S3). There was an average of 167 exonic and splicing mutations (ranging from 2 to 1361) per sample, suggesting high heterogeneity in the CC samples. The mean and median exonic mutation rates were 3.53/ Mb and 2.05/Mb, respectively (figure 1A). These rates are in the same range as the rates observed in the majority of adult solid tumours.19 In total, 91.9% (226/246) of the randomly selected somatic mutations were completely validated by mass spectrometry (MS) genotyping or Sanger sequencing (see online supplementary table S3).

Through mutation spectrum analysis, we observed a high C>T transition rate and a high C>G transversion rate in the CC samples, accounting for 55.0% and 18.9% of the total cases and 7 HPV-negative cases) and matched-normal tissues (n=8) or peripheral blood (n=94) (see online supplementary table S1). We then performed WES analysis of 76 tumour-normal sample pairs using the Illumina HiSeq 2000 platform and the Ion Proton platform, WGS analysis of 27 tumour-normal samples using the HiSeq X-Ten platform and the complete genomics (CG) platform (one sample sequenced by WES and WGS), and transcriptome analysis of 6 tumour-normal samples (all six samples overlapped with WGS/WES) using the HiSeq 2000 platform (see online supplementary figure S1 and supplementary table S1). The average depth of coverage was −100× for WES data and −30× (HiSeq X-Ten) or 100× (CG) for WGS (see online supplementary table S2), respectively.

Using multiple tools, we identified 17034 somatic mutations in the exons and splice regions, including 8479 missense, 3755 synonymous, 691 nonsense, 166 splice, 40 in-frame InDels, 112 frameshift, 11 initiation-site loss, 10 read-through and 3770 UTR mutations (see online supplementary table S3). There was an average of 167 exonic and splicing mutations (ranging from 2 to 1361) per sample, suggesting high heterogeneity in the CC samples. The mean and median exonic mutation rates were 3.53/ Mb and 2.05/Mb, respectively (figure 1A). These rates are in the same range as the rates observed in the majority of adult solid tumours.19 In total, 91.9% (226/246) of the randomly selected somatic mutations were completely validated by mass spectrometry (MS) genotyping or Sanger sequencing (see online supplementary table S3).

Through mutation spectrum analysis, we observed a high C>T transition rate and a high C>G transversion rate in the CC samples, accounting for 55.0% and 18.9% of the total
mutations, respectively (see online supplementary figure S2A-S2B). Context analysis revealed thatTp*CpN>(T/G) was the predominant mutation type, which is consistent with previous report.9 (Tp*CpW=W=A or T)>(T/G) mutations constituted 42.3% of all the mutations. It has been reported that the overexpression of the cytokine deaminases in the APOBEC family,20–22 may cause this mutation signature. Indeed, transcriptomic analysis demonstrated that the APOBEC family member APOBEC3H was expressed at higher levels in CC samples than those in the adjacent normal tissues (p = 0.02, online supplementary table S4), while other members of APOBEC family including APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G showed no significant dysregulation in the same samples (see online supplementary table S4, all p > 0.05). These data suggest that the mechanisms underlying the major APOBEC3H-characterised mutation types (Tp*CpW > (T/G)) play an important role in cervical carcinogenesis.

**Driver mutations in CCs**

To identify recurrently mutated genes, we analysed all somatic SNVs and InDels in 102 CC samples using MutsigCV.19 Five genes showed a statistically significant level of recurrent mutations (with false-discovery rates (FDRs)<0.1): PIK3CA (16.7%, 17/102), FBXW7 (12.8%, 13/102), MLL3 (7.8%, 8/102), CASP8 (3.9%, 4/102) and FADD (3.9%, 4/102) (figure 1B and online supplementary table S5). Although the results were not statistically significant, FAT1 (8.8%, 9/102), MLL2 (5.9%, 6/102) and EP300 (5.9%, 6/102), which had high mutation rates and were enriched in truncation mutations (see online supplementary table S6), were predicted to have roles in CC. Mutations in these genes were validated by MS genotyping or Sanger sequencing (see online supplementary table S3). Recurrent mutations in PIK3CA, FBXW7 and EP300, which have been previously reported in CC,8 were also identified in our cohort. The majority of PIK3CA mutations clustered in the helical domain (73.7%, including four E542K, one E545G and nine E545K) but not in the kinase domain (15.8%, 3/19). Moreover, several hotspot mutations were identified in FBXW7 (three R465C, two R465H, five R505G and one R505L), and nonsense mutations were frequently observed in EP300 (50%, 3/6; online supplementary table S3). We found recurrent mutations in CASP8, FADD, MLL2, ML3 and FAT1, which are, to our knowledge, reported here for the first time in CC (figure 1B; online supplementary table S5-S6). Mutations in these genes were more frequent than those previously reported in the COSMIC database (see online supplementary table S6).

In our cohort, CASP8 and FADD each harboured four mutations, most of which were inactivating mutations (75%, 6/8; figure 1B and online supplementary table S6). Interestingly, although the results were not statistically significant (p = 0.15), FADD mutations were found only in cases without CASP8 mutations (figure 1B). CASP8 is normally recruited to death receptors by binding to the apoptotic adaptor FADD, which subsequently initiates caspase-mediated apoptosis. Loss of function of these two genes may promote carcinogenesis by inducing resistance to cell death.

The MLL2 and MLL3 genes, which each encodes a histone 3-lysine 4 methyltransferase, contained truncation mutations in 85.7% (6/7) and 77.8% (7/9) of the samples, respectively (figure 1B and online supplementary table S6). Both genes are frequently mutated in various tumours, including bladder cancer23 and breast cancer.24 The enrichment in loss-of-function mutations in these genes suggested that the MLL2 and MLL3 genes play roles in tumour suppression in multiple cancer types, including CC.

Another frequently mutated gene was FAT1, which was mutated in 8.8% (9/102) of tumours (figure 1B and online supplementary table S6). Reportedly, recurrent somatic mutations in FAT1 lead to aberrant Wnt activation in multiple human cancers.25 Mutations in related family members, such as FAT2, FAT3 and FAT4, were also found in our cohort (see online supplementary table S3). FAT2 gene-silencing by small interfering RNAs promoted the growth of CC cells, suggesting that FAT2 may be a new tumour-suppressor gene in CC (see online supplementary figure S2C).

**Somatic CNAs and structure variations in CCs**

In this study, we first compared differences in CNAs between WGS at a low depth and WES in eight samples. The results showed a high correlation across platforms (average correlation coefficient=0.93, 0.87–0.97; online supplementary figure S3A), suggesting that the WES data were also suitable for CNA analysis.

We performed CNA analysis on the WES or WGS data from 102 sample pairs, using a modified CLImAT algorithm.16 The results revealed that 17.6% (18/102) of the samples presented polyloidy genome features (the average ploidy was >3.5), of these, 52.2% (14.8%–76.5%) of the chromosomal regions contained CNAs (online supplementary figures S3B-S3C; 82.4% (84/102) of the samples presented diploid genome features, and of these, 23.8% (0.0%–59.7%) of the chromosomal regions contained CNAs. This value was significantly lower than the proportion observed in the polyloidy group (p<0.05, Fisher’s exact test; online supplementary figure S3B).

Among the 84 diploid samples, genome-wide segmented copy number analysis showed that many chromosome arms had undergone large-scale gains or losses in copy number, with frequent gains observed on chromosomes 1q, 3q, 5p, 8q, 9q and 20q and frequent losses observed on chromosomes 3p, 4p, 4q, 6q, 8p, 11p, 11q, 17p, 18p and 19p (see online supplementary table S7). The overall CNA pattern was broadly consistent with the results of other published studies of CC.8 26 27

We identified 23 focal events using the GISTIC2.0 algorithm,28 including 10 amplification peaks and 13 deletion peaks that involved 2541 genes (see online supplementary figure S3D and table S8). These genes included well-known onco-genes and tumour-suppressor genes, such as BIRC3 (11q22.1), PIK3CA/SOX2 (3q29), EGFR (7p11.2), CCND1 (11q13.3), BAP1 (3p11.2) and ATM (11q23.3) (figures 1C and 2A).

Considering the limitations of the GISTIC tool, we used an in-house-designed algorithm to identify more potential CNA peaks. We identified seven peaks with a high ratio that involved 1461 genes (see online supplementary table S8). For example, MCL1, which is one of the most frequently amplified genes in human cancer,29 was amplified in 39.3% of the CC samples. Another frequently amplified gene, E2F1 (20q11.21), is a key regulator of cell cycle and was amplified in 25.0% of the CC samples.

CREST software30 was used to identify SVs in CCs. A total of 770 SVs were identified, including 109 interchromosomal translocations, 161 intrachromosomal translocations, 115 insertions and 385 deletions (see online supplementary figure S3E and table S9). Chromothripsis occurred in two samples: CC-X009 and CC-X019 (figure 2B,C). We therefore deemed it useful to further examine the relationship between SVs and cervical carcinogenesis.
Driver pathway analysis in CCs

Using the list of driver genes that we found to be altered by mutations or copy number changes, we searched for over-represented pathways with known roles.\textsuperscript{10} The results showed that well-defined cancer-related pathways, including the RTK/RAS/PI(3)K, cell cycle and apoptosis pathways, were altered in 88%, 74% and 73% of cases, respectively (figure 3A). The activation of EGFR, ERBB2, ERBB3, ERBB4 (mutations and amplifications) and NRAS (amplification) affected the RTK/RAS pathway, and the activation of PIK3CA (mutations and amplifications) predominately resulted in the dysregulation of the PI(3)K/akt/mTOR pathway. Altered genes in the cell-cycle pathway were mainly related to the G1/S transition, including activated CCND1, CDK4, E2F1 and E2F3 (amplifications) and inactivated CDKN1C, CDKN2D and RB1 (deletions). We also identified mutations that frequently affect the apoptosis pathway, including deletions and mutations of FAS, TNF, PADD and caspases 3, 6, 7, 8, 9 and 10. In addition, the amplification of the BIRC2/3 genes on 11q22 played a role in the deregulation of the cancer cell-death pathway.

Genes involved in the chromatin remodelling pathway were recurrently altered by mutations or CNAs. Potential driver gene analysis identified several histone modifiers (MLL2, MLL3 and EP300), and many other genes in this pathway were also mutated or deleted. In addition, 87% of cases had mutations or deletions in components of the SWI/SNF nucleosome remodelling complex, including ARID1A, ARID1B, SMARCA1, SMARCA3 and PBRM1. We also found alterations in genes involved in oxidative stress response in 59% of cases, including mutations and deletions in CUL3 and KEAP1, as well as amplifications and mutations in NFE2L2. Specifically, genes with known roles in squamous cell differentiation were also frequently altered.

Amplifications in SOX2 and TP63 were widely identified in CC samples as were mutations and deletions in NOTCH family members (NOTCH1,2,3). To identify genes that could potentially be targeted by drugs in the treatment of CC, we used the TARGET (tumour alterations relevant for genomics-driven therapy) database\textsuperscript{11} for the integrated analysis of mutations and CNAs. The results showed that 34 genes with targetable oncogenic mutations were revealed in 55.9% (57/102) of the samples. Of these 57 samples, each contained at least one mutation or one CNA (≥five copies) in a gene that predicts sensitivity or resistance to anticancer agents (ranging from one to five) (figure 3B). These results suggested that patients with CC may potentially benefit from next-generation sequencing (NGS)-based molecular genotyping and that newly identified genetic alterations may serve as potential candidate drug targets for CC.

Somatic mutations in both CINs and CCs

To study the relationship between somatic mutations and the initiation and progression of CC, we performed WES in 35 additional sample pairs, including 10 CIN1, 9 CIN2, 6 CIN3 and paired blood cell DNA, as well as 10 CC tumour-normal specimens. The average depth of coverage was 95×, and the average percentage of region covered at least 1x was 99.42% (see online supplementary table S2). We found 1747 SNVs and 27 InDels in all 35 samples, 1743 of which occurred in CCs and involved 1532 genes (see online supplementary table S10). Interestingly, 31 gene mutations were detected in 52% (13/25) of the CIN specimens, and 54.8% (17/31) of these mutations (in 32% (8/25) of the CIN specimens) occurred in genes that also mutated in CCs (figure 4). All mutations observed in CINs were validated.
Figure 3  Frequently altered pathways and potential therapeutic targets and drugs in cervical carcinomas. Somatic mutations and copy number alterations (CNAs) in components of the RTK/RAS/PI3K, apoptosis, chromatin remodelling, cell cycle, oxidative stress response and squamous differentiation pathways. Red, activating genetic alterations; blue, inactivating genetic alterations. The percentages shown indicate the activation or inactivation of at least one allele. (B) The left panel represents the alteration frequency for each potential drug-targeted gene in the different pathways. The middle panel represents the recurrent somatic mutations (which reoccurred in this cohort or were previously reported in the COSMIC database) and CNAs (≥5 copies). The right panel represents the potential drugs that can be used against the alterations shown in the middle panel.

manually by IGV or by Sanger sequencing (see online supplementary figure S4). Our current findings suggested that mutations could be detected in CIN samples. These results provide a solid foundation for further study of the CC pathogenesis and provide new molecular biomarkers for the early warning and screening of CC.

HPV DNA integrations in both CINs and CCs
To study the relationship between HPV integration in the human genome and the occurrence and development of CC, we performed an HPCS technology. We analysed HPV integration breakpoints in 25 normal cervical samples (including 18 HPV-negative and 7 HPV-positive samples), 44 CIN samples (including 24 CIN1 and 20 CIN2-3 samples) and 45 CC samples (see online supplementary table S1).

A total of 2466 HPV integration breakpoints were identified in 84.3% (75/89) of the abnormal samples, including 97.8% (44/45) of the CC samples and 70.5% (31/44) of the CIN samples (figure 5A, online supplementary table S2 and S11). To confirm the HPV integration events, we randomly chose 29 integration junctions for validation using Sanger sequencing. The results showed that 82.8% (24/29) of the integration junctions were validated (see online supplementary table S11), indicating that the results were reliable for further analysis.

As expected, no HPV integration breakpoint was detected in the normal samples with HPV negative. HPV integration events...
were found in 42.8% (3/7) of the HPV-positive normal cervical samples (figure 5A). Moreover, we found that HPV integration breakpoints had more supported reads in CC samples than those in the normal samples (p=2.88E-05, Mann-Whitney U test), CIN1 samples (p<2.2E-16, Mann-Whitney U test) and CIN2-3 samples (p=3.79E-09, Mann-Whitney U test), whereas the HPV integration breakpoints had more supported reads in the CIN specimens than those in the normal samples (p=8.78E-05 for CIN2-3 vs normal and p=1.94E-02 for CIN1 vs normal, Mann-Whitney U test; figure 5B). Interestingly, we found that the integrations of certain HPV strains in the CCs, including HPV16, 18, 33 and 58, had also occurred in the CIN1 and CIN2-3 samples (figure 5C). The integration occurrence of high-risk HPV was higher than that of low-risk HPV in CINs and CC as compared with normal tissues with HPV positive (p=0.004 for CIN1, p=0.0015 for CIN2-3 and p=1.34E-06 for CC, respectively; Fisher’s exact test; online supplementary figure 5B). These data suggest that oncogenic HPV integration events may occur early in preneoplastic lesions.

To validate the diagnostic performance using HPV integration, ROC curve analysis was conducted in CC and CIN compared with normal individuals (see online supplementary table S12 and figure S5). We found that HPV integration alone or combined with HPV DNA testing could distinguish CC and normal. We used multiple HPV integration as biomarkers to estimate its diagnostic value, the AUC values were 76.5% (95% CI 64.17% to 88.83%), 81.2% (95% CI 69% to 93.4%) and 95.5% (95% CI 88.93% to 100%) for CIN1, CIN2-3 and CC, respectively. Specifically, HPV16/18 integration exclusively showed an AUC value of 96.7% (95% CI 92.98% to 100%; DeLong’s test) for CC versus normal, suggesting an ideal biomarker for detecting CC. The combination of HPV integration and DNA testing had a trend towards higher AUC value than HPV DNA testing (DeLong’s test; online supplementary table S12) in CIN and CC, suggesting a better biomarker for cervical cancer screening.

Our results showed that HPV integration breakpoints are distributed over the whole HPV genome (see online supplementary figure S6A). Statistical analyses indicated that integration junctions were more likely to be located in the E1 (p=7.2E-4) and E2 genes (p=4.5E-4), whereas integration junctions were less likely to be located in the E6 gene (p=4.7E-2), the long control region (LCR, p=2.8E-3), the L1 gene (p=2.8E-2) or the L2 gene (p=1.2E-3; online supplementary figure S6B). Our data suggested that intact forms of the E6 gene and the LCR were more likely to be observed in the human genome.

Local algorithms predicted that the integrated HPV fragment sizes ranged from 43 to 7885 bp (see online supplementary figure S6C). To validate the HPV fragment sizes predicted by NGS, we randomly selected 20 HPV integration events and performed long PCR to amplify the full-length fragments. Junctions residing on both sides of the six successful amplicons were validated by Sanger sequencing (see online supplementary figure S6D). The results revealed a 6649 bp HPV integrant in the CC-H011 genome (see online supplementary figure S7A), and the other five HPV fragments, ranging from 1897 to 4648 bp, were inserted into the human genome of CC samples (see online supplementary figures S7B-7F). These results indicated that HPV fragments of varied fragment sizes randomly insert into the human genome. Transcriptomic data showed that integration of the E4, E5, E6 and E7 genes led to the expression of full length of viral genes (see online supplementary figure S7G), whereas integration of the E1, E2, L1 and L2 genes resulted in only partial transcription or no transcription (see online supplementary figure S7G). These data indicate that although all HPV genes can be integrated into the host genome, only the oncogenes are completely expressed.

Furthermore, HPV integration sites were found within the exons or introns of 886 human genes, including MYC, RAD51B, FHIT, KIF1B, ALK and PDE4D (see online supplementary figure S8A and table S11). We identified 53 integration breakpoints in the MYC gene and its flanking regions (the upstream or downstream 500K) in 24.4% (11/45) of the samples (see online supplementary table S11). Reportedly, integration breakpoints are closely associated with genomic instability. Indeed, we found (1) focal amplification of the MYC gene in CC-H011 and CC-H012, in which the HPV

Figure 4 Somatic mutations in cervical intraepithelial neoplasia (CINs) and cervical carcinomas. All somatic mutations that occurred in the 25 CINs and 112 cervical cancers (CCs) are shown in the left panel, with each row indicating a gene and each column representing a sample. Genes mutated in the CIN specimens are zoomed in and shown in the right panel.
integration breakpoints were located in regions both upstream and downstream of the MYC gene; (2) a 17 Mb gain in 8q that included integration into MYC in CC-H024, in which the HPV integration breakpoints were located ~240 kb upstream of MYC and (3) focal amplification of POU5F1B (317 kb upstream of MYC) in CC-H044, in which HPV integration breakpoints were located both upstream and downstream of POU5F1B (see online supplementary figure S8B). Focal amplification may be identified using a ‘looping’ model, as previously reported.32 Gene mutations were identified especially in CIN1 in this study and the study by Liu et al.19 The reason might be ascribed to the discrepancies in sample size and sequencing depth (~10× for samples by Liu et al., ~6427× for specimens by Hu et al. and ~61217× in our cohort, respectively). And the integration events were validated by Sanger with a ratio of 82.8%, indicating that high-depth HIVID could be a sensitive method to detect integrated HPV. The different types of virus integration spectra were observed in different stages of cervical samples. HPV16 and HPV18 are the dominant HPV types integrated into CC samples. Integration of multiple HPV types, including HPV16, HPV18, HPV33 and HPV58, has been identified in CIN samples. Big data of CC screening showed that HPV 16, 33 and 58 were the most common HPV types in high-grade squamous intraepithelial lesions in Chinese women.42 It is possible that different integrated HPV types indicate different outcomes of CIN samples, which may require further follow-up and validation for these patients. Novel disease-specific biomarkers, such as gene mutations and HPV integrations, may serve as secondary markers after positive HPV DNA tests to identify women with prevalent precancers who require immediate colposcopy or treatment.

In clinical practice, advanced patients with CC often choose platinum-containing chemotherapies, but their prognosis remains poor. Therefore, it is necessary to search for new therapeutic drugs, especially targeted drugs, via NGS-based molecular genotyping. Our results showed that 55.9% of the CC samples harboured at least one actionable alteration. PI(3)K/AKT/mTOR inhibitors,43 tyrosine-kinase inhibitors,44 receptor tyrosine kinase antibodies45 and cyclin-dependent kinase inhibitors46 have been approved by the Food and Drug Administration or are in clinical trials for several cancers but not for CCs. Patients with alterations in the targets of these drugs might benefit from targeted therapies. In addition, 50.9% (29/57) of these patients harboured more than one actionable alteration, indicating that combinational therapies may allow patients with CC to receive greater treatment benefits.

In summary, we systematically analysed the genomic variations and HPV integration profiles of CIN and CC, our findings provide the foundation for detecting precancerous lesions early and developing new biotechnology for the screening and therapy of CC.
Author affiliations

1 Key Laboratory of Systems Biomedicine (Ministry of Education) and Collaborative Innovation Center of Systems Biomedicine, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China
2 Shanghai-MOST Key Laboratory for Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, China
3 Bihai Genomics Institute, BGI-Tianjin, Tianjin, China
4 BGI-Shenzhen, Shenzhen, Guangdong, China
5 Genepius-Beijing, Beijing, China
6 Department of Obstetrics and Gynecology, Southwestern Hospital, Third Military Medical University, Chongqing, China
7 Shenzhen People’s Hospital, Second Clinical Medical College of Jinan University, Shenzhen, China
8 Shenzhen Key Laboratory of Infection and Immunity, Shenzhen Third People’s Hospital, Guangdong Medical University, Shenzhen, Guangdong, China
9 STD Institute, Shanghai Skin Disease Hospital, Tong Ji University, Shanghai, China

Correction notice

This article has been corrected since it was published Online First. Dr Pingyu Zhou’s corresponding address has been corrected.

Contributors

JH, PJ and ZL conceived the project and JH and XY designed the experiments. YG, XX, YH, SC, XG, YZ, S2, JX, RW, HZ, LF and XY performed sequencing and ZQ, YG, LJ, MC, CB and JH analysed the sequencing data. WJ and BZ performed the pathology experiments. QY performed the RNA interference experiments. MS performed the pathway analysis. JH, PJ and XY contributed reagents, materials and analysis tools. ZL, PJ, YW and HX contributed the samples. JH, YQ and ZZ integrated, analysed and interpreted all the data. JH contributed to the supervision of the work. JH, YH and ZZ wrote the manuscript with the assistance and final approval of all authors.

Funding

This work was supported by grants from the National Natural Science Foundation of China (81972274), the China National Key Projects for Infectious Disease (2017ZX10203207), the project of Precision Medicine of Southwestern Hospital (SWHZH0126DCX1013), the Clinical Research Plan of Shanghai Hospital Development Center (16CR1029B and 16CR3111B), the Chinese National Key Program on Basic Research (2014CB965002) and the Shanghai Commission for Science and Technology (15431902900).

Competing interests

None declared.

Patient consent for publication

Not required.

Ethics approval

This study was approved by the Research Ethics Committee of Southwestern Hospital (No. 2014–016) and EthicsCommittee of the Shanghai Skin Disease Hospital (No. SKIN2015–010).

Provenance and peer review

Not commissioned; externally peer reviewed.

Open access

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

Cancer genetics


