Serum, Cell-Free, HPV-Human DNA Junction Detection and HPV Typing for Predicting and Monitoring Cervical Cancer Recurrence

Anne Van Arsdale[1,2], Olga Mescheryakova[1], Sonia Gallego[1], Elaine C. Maggi[1], Bryan Harmon[3], Dennis Y.S. Kuo[2], Koenraad Van Doorslaer[4], Mark H. Einstein[2], Brian J. Haas[5], Cristina Montagna[1,6]*, and Jack Lenz[1]*

[1]Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, 10461, USA,
[2]Department of Obstetrics Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, NY, 10461, USA,
[3]Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461, USA,
[4]School of Animal and Comparative Biomedical Sciences, College of Agriculture and Life Sciences BIO5 Institute, University of Arizona, Tucson, AZ, 85721, USA,
[5]Broad Institute, Cambridge, MA 02142, USA,
[6]Rutgers Cancer Institute, 195 Little Albany St., New Brunswick, NJ, 08901, USA

*communicating authors

Keywords – cervical cancer, human papilloma virus, HPV, cell-free DNA, cfDNA, DNA integration, cancer recurrence.

Abstract

Almost all cervical cancers are caused by human papillomaviruses (HPVs). In most cases, HPV DNA is integrated into the human genome. We found that tumor-specific, HPV-human DNA junctions are detectable in serum cell-free DNA of a fraction of cervical cancer patients at the time of initial treatment and/or at six months following treatment. Retrospective analysis revealed these junctions were more frequently detectable in women in whom the cancer later recurred. We also found that cervical cancers caused by HPV types outside of phylogenetic clade $\alpha 9$ had a higher recurrence frequency than those caused by $\alpha 9$ types in both our study and The Cancer Genome Atlas cervical cancer database, despite the higher prevalence of $\alpha 9$ types including HPV16 in cervical cancer. Thus, HPV-human DNA junction detection in serum cell-free DNA and HPV type determination in tumor tissue may help predict recurrence risk. Screening serum cell-free DNA for junctions may also offer an unambiguous, non-invasive means to monitor absence of recurrence following treatment.

Introduction

The vast majority of cervical cancers are caused by human papillomaviruses (HPVs), and they remain a significant cause of cancer deaths. Despite the availability of HPV vaccines and effective procedures for detecting and treating pre-invasive cervical lesions, they rank as the fourth most common cancer in women globally[1]. In the United States, an estimated 13,820 new cases of cervical carcinoma are projected for 2024, resulting in 4,360 deaths. The 5-year overall survival (OS) rate across all stages is 66%, with rates dropping to 58% for loco-regional spread and 17% for distant metastasis[2]. Early-stage invasive cervical cancers (FIGO IA1-IB2) are typically treated with surgical excision as a standard of care. However, for patients with advanced stages (FIGO IB3-IVB), surgery is not beneficial, and first line therapy entails cisplatin-radiosensitization combined with ionizing radiation plus brachytherapy, with recurrence risk increasing with stage[3]. An enhanced ability to stratify women during initial care by identifying those at highest risk of treatment failure may improve patient care.

HPV genomes replicate as circular, extra-chromosomal, double-stranded DNAs of about 8 kbp. However, in the vast majority of cervical cancers, HPV DNA is integrated into human DNA, either within a chromosome or as an HPV-human heterocatemeric, extra-chromosomal DNA[4,5,6,7]. Each junction between HPV and human DNAs is an unambiguous, patient tumor-specific identifier of clonally expanded tumor cells derived from the cell that initially underwent the integration event.

Therefore, we investigated whether HPV-human DNA junctions could be detected circulating in serum cell-free DNA (cfDNA), and whether they could be used to assess residual disease following initial cervical cancer treatment.

Results

Detection of HPV-human DNA junctions in cervical cancer patient tumors

HPV-human DNA junctions were retrospectively assessed using archived samples from 16 cervical cancer patients (12 squamous carcinomas, 4 adenocarcinomas) with available frozen tissue and at least one serum sample obtained during initial care. (**Figure 1A, 1B)**. Eight patients had locally advanced cervical cancer (FIGO IB3-IIIC1r), while eight had surgically resectable, early-stage cancers (FIGO IB1-IB2). HPV-human DNA junctions were identified using DNA isolated from tissue biopsies or surgical excisions by hybridization capture for full genomes of 225 HPV types followed by Illumina sequencing (HC+SEQ) and computational identification of junctions using CTAT-Virus Integration Finder[8] (**Table 1**). This approach also enabled the unambiguous identification of HPV types and genome coverage of each specific type, thus allowing inference of the structures of the integrated viral DNA segments in the tumor genomes (**Figure 1C, 1D, 1E**). This included identification of the lower prevalence types HPV70 and HPV73 in one patient each. In most of the tumors, the inserted HPV DNA comprised a subgenomic fragment of the viral genome (**Figure 1C, 1D**). Complete viral segment sequences determined by HC+SEQ also allows identification of individual, tumor-specific differences compared to HPV type reference sequences at single nucleotide resolution. Junction-PCRs using tumor-specific primers, one near the end of the integrated HPV DNA segment, the other in the flanking human genome sequences (**Methods**) confirmed the HPV-human junctions in the cognate tumor tissue DNAs (data not shown).

Five of the sixteen patients had tumor recurrences, all occurring at distant metastatic sites outside standard-of-care radiation fields or excised organs. Metastasis biopsy samples were available from all five recurrences, and junction-PCR analysis on DNAs from them confirmed identical junctions to those found in each cognate primary tumor (**Figure 2)**, thus confirming clonal derivation from the primary tumors.

Tumor-derived, HPV-human DNA junctions in serum cell-free DNA

The usefulness of junction detection in serum cfDNAs was evaluated using samples obtained during initial care from fifteen of the patients and at six months post-treatment from seven of the patients (**Figure 1A**). Since levels of cfDNA in serum from cancer patients are generally elevated over levels in control individuals[9], serum DNA concentrations were measured in the samples obtained at initial care. They were significantly higher in women diagnosed with locally advanced cancers stage compared to those with early-stage cancers (Wilcoxon rank sum test, $p<0.05$) (**Figure 3A**), but no significant difference was observed between those with or without a recurrence (**Figure 3B**). HPV-human DNA junctions in cfDNAs were assessed by HC+SEQ directly on cfDNA and by junction PCR methods, which yielded concordant results except for Patient 2 (**Methods**). Junctions were detected in cfDNA from seven patients, including all cases of recurrences except no baseline sample was available for Patient 4 (**Figure 1A**).

HPV-human DNA junction detection in cfDNA for prediction of recurrence-free survival

In light of HPV-human DNA junctions being unambiguous markers for tumor cell clones containing integrated HPV DNA, the significance of junction detection for predicting recurrence at baseline and six months was assessed by Kaplan-Meier analyses of recurrence-free survival (**Figures 3C, 3D**). Stratified log-rank tests adjusted for stage at diagnosis and left-censored to the dates of last-follow-up showed that detection at 6 months was statistically significant and prognostic for recurrence (p=0.035). Detection at initial treatment was of borderline statistical significance (p=0.06). Due to the small sample size, accurate determination of effect size of detection at either time point with appropriate 95% confidence intervals could not be accomplished using either univariable or multivariable Cox regression modeling. Nonetheless, junction detection in serum cfDNA at 0 and 6 months showed promise as a predictor of tumor recurrence.

HPV DNA detection alone in cfDNA of cervical cancer patients has demonstrated value for helping predict tumor recurrence risk[10]. The E7 gene is crucial for HPV tumorigenesis[11], and the URR-E6-E7 segment was present in all the tumors (**Figure 1D and 1E**). However, while circulating tumor DNA (ctDNA) is the predominant component of cfDNA in cancer patients, HPV DNA from non-tumor tissues could also be present in cfDNA[12,13]. Therefore, HPV type-specific E7 DNA detection in cfDNA was also evaluated (**Figures 1B**). Among available samples, 14/15 (93.3%) baseline serum cfDNAs and 6/7 post-treatment sera, including all five from patients with recurrences, tested positive for HPV type-specific E7 DNA by HC+SEQ and/or E7 gene PCR (**Table 1**). No HPV DNA was detected by PCR or HC+SEQ in serum cfDNA from Patient 13, where HPV18 DNA was integrated in the cognate tumor. Kaplan-Meier analysis at baseline and six months post-treatment using stratified log-rank tests for HPV E7 positivity, adjusted for stage, did not significantly associate with recurrence at either time point (p=0.52 and p=0.31, respectively) (**Figure 3E and 3F**). Due to the small sample size, accurate effect size determination with 95% confidence intervals using univariable or multivariable Cox regression modeling was not feasible.

Interestingly, Patient 16 serum cfDNA tested positive for both HPV16 and HPV18 DNA by HC+SEQ with genome coverage fractions of 0.982 and 0.919, respectively, and by E7 PCRs (**Figure 4, Table 1**). In contrast, the primary tumor tested positive only for HPV16 DNA by HC+SEQ and E7 PCR (**Table 1 and Figure 4A, 4B**). Moreover, the HPV16 DNA detected in serum cfDNA had different mutations relative to a reference HPV16 genome than the HPV16 DNA in the tumor (**Figure 4C**). These results implied that the cervical cancer in this patient was caused by the HPV16 variant detected in the tumor tissue, and that neither the HPV16 nor the HPV18 that were detected in the serum cfDNA were present in the tumor at the time of initial treatment and presumably derived from an alternative source. Moreover, no junctions with human DNA were detected with either of the HPV DNAs in serum cfDNA, and HPV16 sequence reads in cfDNA were continuous across the viral genome at the positions of the junctions with human DNA in the cancer. These observations suggested that the patient was coinfected with both the HPV16 and HPV18 found in serum cfDNA at the time that the baseline sample was

obtained, and that there was no evidence for a causative, tumorigenesis role by either those HPVs. In summary, these results indicate that HPV DNA in serum cfDNA may arise from unclear tissue sources, and that HPV-human DNA junctions provide a more rigorous, tumor-specific means for detecting the presence of tumor cells than detection of just HPV DNA, at least in some patients.

HPV phylogenetic clade type as a risk factor for cancer recurrence

Unexpectedly, all five recurrences in this cohort involved HPV types from phylogenetic clade α7 (HPV18, HPV39, HPV45, HPV70) or α5 (HPV69), with none involving clade α9 types (HPV16, HPV35) (**Figure 1A**). Patients with cervical cancers containing clade α7 HPV types have inferior survival versus those with clade α9[14,15]. Therefore, HPV clade types were also assessed as an independent prognostic factor for recurrence for the larger cervical cancer dataset in The Cancer Genome Atlas (TCGA) using Kaplan-Meier analysis of clinical information and tumor tissue sequences (**Tables 2 and 3, Figure 5**). HPV types were determined using CTAT-Virus Integration Finder[8] analysis of TCGA sequencing data, revealing 214 tumors (69.5%) with clade α9 types, 83 (26.9%) with other clade types (α5, α6, α7, or α11) , and 11 (3.6%) being HPV-negative. Clinicopathologic variables known to be *a priori* prognostic factors were extracted from cBioPortal[16 17], including age, histology, and stage at diagnosis, with N=297 patients having complete data. Comparisons of HPV clade α9 with non-clade-α9 (α5/α6/α7/α11) showed no significant baseline differences in these variables (**Table 2**). However, patients with non-α9 clades exhibited inferior recurrence-free survival compared to those with HPV clade-α9 tumors based on a log-rank test stratified for histology and stage (p=0.03). A multivariable Cox regression model including age (years), histology (squamous/adenocarcinoma) and stage (I, II, III, IV) (**Table 3**) was performed to account for these *a priori* variables, and women with HPV clades other than α9 showed inferior recurrence-free survival in comparison to HPV clade-α9 tumors (Hazard Ratio (HR)=2.48, 95% CI: 1.34-4.49, p=0.004) after adjusting for all variables (**Figure 5**). The effect of HPV clade on recurrence-free survival was second only to stage IV disease (HR: 4.40, 95% CI: 1.89-10.18, p=0.001) in this adjusted analysis (**Table 3**). In summary, these results show that non-α9 clade HPV types are a risk factor for tumor recurrence following current standards of care.

Discussion

The results presented here show that HPV-human junction detection in serum cfDNA and that HPV type determination in tumor tissue at the time of initial cervical cancer treatment might be useful to help predict the likelihood of cervical cancer recurrence. In particular, they also show that junction detection at six-months post-treatment may be useful for unambiguous detection of residual tumor tissue. Detection of HPV DNA alone in cfDNA is useful for predicting relapse[10 18]. However, the detection of HPV's different than the one integrated in the cfDNA of Patient 16 here (**Figure 4**) illustrates the potential for occasional, misleading results, and HPV DNA has been reported in blood of healthy donors and women with documented preinvasive cervical dysplasia up to 50% of the time[12,13]. These are consistent with the possibility that HPV DNA in

cfDNA may occasionally arise from non-tumor tissue sources, thus emphasizing the potential advantage of HPV-human DNA junctions for assessing residual disease. Our results also suggest that HPV-human DNA junctions could be tested in other types of HPV-induced cancers.

HPV-human DNA junctions in cfDNA offer the advantage of unequivocal origin from the cancer cell clones in patients. Long-term prospective studies are required to confirm the findings here, and the potential for improved recurrence risk prediction by HPV-human DNA junction detection in serum emphasizes the need to develop improved upfront therapies for women at risk for cervical cancer relapse. In addition, junction detection in serum cfDNA potentially offers a non-invasive means for rigorous, longitudinal monitoring of disease recurrence in patients following cervical cancer treatment.

Identification of non-α9 clade HPV types (at least α5, α6, α7, and α11) as a risk factor for recurrence in the TCGA/cBioPortal data[16 17] and in the small retrospective study presented here establish that HPV clade type specifically affects cervical cancer recurrence *per se* following treatment as well as survival, and may reflect the same underlying molecular genetic properties of these HPV types. These effects contrast with the high prevalence of α9 HPV16 (~50%) in cervical cancer and the correlation between HPV16 and progression of preinvasive, cervical intraepithelial neoplasia[19 20]. They also highlight a potential role for precise HPV typing in cervical cancer management using multiplexed HPV-typing by hybridization capture plus sequencing methods like HC+SEQ as used here to facilitate clade assignment. These approaches are technically comparable to clinically-utilized assays like exon capture, and inclusion of all known HPV types also enables broad diagnostic specificity plus tumorigenicity assessment of rare HPV types. Again, the potential for improved recurrence risk prediction emphasizes the need to develop improved upfront therapies for women at risk for cervical cancer relapse.

In summary, HPV-human DNA junction detection in cfDNA and HPV type determination at the time of initial treatment offer promise for predicting and monitoring cervical cancer recurrence.

Methods

Patients, sample collection and nucleic acid preparation

All procedures were approved by the Internal Review Board of the Albert Einstein College of Medicine at Einstein-Montefiore. Tumors from women diagnosed with invasive cervical carcinoma were collected as part of an ongoing tissue collection protocol of the Department of Obstetrics & Gynecology and Women's Health at the Albert Einstein College of Medicine (IRB# 2009-265). 16 cervical cancer patients were chosen retrospectively based on availability of frozen tumor tissue plus an archived serum sample from time of initial treatment and/or 6 months post-treatment. All patients receiving adjuvant or curative radiotherapy completed their course within 8 weeks. Diagnosis was by standard-of-care histopathology on biopsies and included 12 (75%) squamous cell carcinomas and 4 (25%) adenocarcinomas (**Figure 1, Table 1**), which is reflective of published epidemiological prevalence data.

From all samples, 10 µM sections of the frozen tissue were cut for H&E confirmation of the presence of viable tumor tissue and presence of the diagnosed histologic subtype as reviewed by a board-certified gynecologic pathologist (Bry.H.). DNA was extracted from sections of the frozen tissue using the QIAamp DNA Mini Kit (Qiagen) stored at −80°C until use. Serum cfDNA was extracted from samples banked at date of diagnosis (baseline) and/or at 6 months post-treatment completion. Serum was also collected the five patients that suffered a recurrence at the time of the recurrence. Qiagen QIAamp Circulating Nucleic Acid kits were used to extract serum cfDNA. Extracted DNAs were subjected to two separate serial dilutions, and DNA concentrations were quantified using the Qubit Fluorometric Quantification method (Thermo Fisher Scientific) and DNA integrity was assessed using the Bioanalyzer capillary electrophoresis system (Agilent).

HPV DNA hybridization capture with short read Illumina sequencing (HC+SEQ) and data analysis

HPV DNA including fragments containing HPV-human DNA junctions were enriched using a DNA hybridization-capture panel containing the genomes of 225 HPV types, followed by Illumina sequencing. Sequence reads were aligned to a custom-assembled sequence file comprising the human and 225 HPV genomes to identify HPV types, assess viral genome coverage, and identify HPV-human DNA junctions.

Targeted enrichment of HPV DNA was performed using custom hybridization capture probes homologous to the full-length genomes of 225 different HPV types as previously described[5,21]. Briefly, 1 µg of tumor genomic DNA was mechanically fragmented to 200 bp for short-range sequencing (Covaris, Woburn, MA), and Illumina adaptors were ligated at each end using the KAPA Hyper Prep kit per manufacturer instructions. Libraries were then hybridized to the biotinylated HPV oligonucleotide probes specific for each of the 225 HPV types for 72 h using the Roche target enrichment protocol (Roche Nimblegen SeqCap EZ System, Basel, Switzerland) following manufacturer instructions. After library purification, the DNA was sequenced on one Illumina Novaseq lane (Illumina, San Diego, CA) using the paired end, 150 bp sequencing mode. The sequencing reads were next adaptor-cleaned, and those that passed QC were de-duplicated, and paired end reads were aligned to a custom human (GRCh38/hg38) plus HPV reference genome containing 225 HPV types from the Papillomavirus Episteme[22] using the STAR Aligner[23]. Average sequencing coverage of the sample typed HPV genome after de-duplication was calculated based on the mapped, on-target, 150 bp reads. HPV-human junction fragments were computationally identified using CTAT-Virus Integration Finder that specifies a chimeric read analysis to define integration sites[8]. When integrated HPV DNAs were of subgenomic length, the edges of the HPV genome coverage plots (**Figure 1C**) uniformly corresponded to junctions with human DNA (**Figure 1D**). In some tumors, more than two junctions were identified during HC+SEQ that each clustered within a single segment of the human genome, a previously described phenomenon[24-28] suggesting that integrated HPV DNA segments undergo rearrangements to some extent at least in some lesions. The junctions illustrated and used for PCR analyses here represent those with the highest numbers of HC+SEQ read counts. The other

junctions presumably represent subclonal variants and may be similarly useful for the types of junction analyses described in this study. HC+SEQ using serum cfDNA was performed similarly except that the amounts of starting material were lower (5 to 100 ng). No additional amplification was performed beyond the KAPA Hyper Prep step and the final capture libraries.

HC+SEQ sequence data availability

HC+SEQ sequence reads are available (SUB14649435) at the Sequence Read Archive (SRA) of the National Library of Medicine, National Institute of Health.

HPV type and human DNA junction analysis by PCR and Sanger sequencing

PCR was also used to confirm the human genome junctions and HPV types detected by HC+SEQ. PCR primers to detect the HPV-human DNA segments unique to each individual tumor/cfDNA sample (junction PCRs) were designed to flank each side of the junctions obtained from HC+SEQ split reads with one primer in the HPV sequences and the other in the flanking human sequences. We were unable to design primers that successfully amplified the junctions for patient 2, even though they could be detected in the cfDNA by HC+SEQ. HPV type was also assessed using PCR primers specific for the E7 genes of each HPV type, as it was retained in all tumor samples from the cohort (**Figure 1**). Primers and amplicon size for junctions positive by PCR are in **(Supplementary Table 1)**. PCR reactions were performed using GoTaq Green (Promega) using a standard of 35 cycles of 94°C for 30 sec, 60°C for 30 sec, and 72°C for 45 sec and adjusted if needed based on individual primer Tm. PCR products were resolved using a 2% agarose gel and extracted using the QIAquick® Gel Extraction Kit (Qiagen) for Sanger sequencing.

Statistical Analysis and TCGA dataset

Detailed medical records were abstracted to obtain data regarding age at diagnosis, histological subtype, stage, and treatment related variables (**Supplementary Table 2**). Histologic classification was designated either squamous cell carcinoma or adenocarcinoma of the cervix. Stage was determined by the 2018 International Federation of Gynecology and Obstetrics (FIGO) staging criteria[29]. Treatment related variables included: 1) primary surgical intervention which consisted of radical hysterectomy, bilateral salpingo-oophorectomy, and pelvic lymph node dissection, 2) radiation therapy delivered as external beam and/or brachytherapy as either for primary curative intent or as adjuvant therapy based on Sedlis criteria[30] and 3) systemic cisplatin chemotherapy administered as a radiosensitizer in either primary curative intent or adjuvant setting. Based on biopsy, Patient 9 was diagnosed with stage 1B2, but based on histopathological analysis of the surgically excised tissue, additional high-risk features were noted including positive parametria and positive pelvis lymph nodes, and thus also received chemo-radiation therapy. HPV type was determined based on HC+SEQ HPV typing and PCR confirmation. Patients were stratified into HPV phylogenetic clades for the viral types in this

cohort as follows: HPV clade α7 (types 18, 39, 45 and 70), HPV clade α9 (HPV types 16 and 35), and HPV clades "other" (HPV types 69 and 73, clades α5 and α11, respectively).

The primary exposures of interest were 1) HPV-human DNA junction positivity by either HC+SEQ or PCR and 2) HPV E7 positivity by either HC+SEQ or PCR. The primary outcome of interest was time to recurrence. Recurrence-free survival (RFS) was calculated from the date of diagnosis to the date of first recurrence. Patients were left-censored at the date of last follow up. This included patients that were lost to follow up, which was assumed non-informative, and patients that did not experience a recurrence event.

Demographic and disease characteristics by recurrence status were described using means and standard deviation for normally distributed continuous variables, medians and interquartile ranges (IQR) for non-normally distributed variables, and frequencies and percentages for categorical variables. Differences in these variables were assessed using the Student's t-test, Wilcoxon test, and Fisher's Exact test, respectively. Recurrence-free survival was summarized by HPV-human DNA junction or HPV E7 baseline positivity, and by HPV-human junction or HPV E7 at 6 months post-treatment using Kaplan-Meier survival curves. Differences were assessed using the stratified log-rank test, stratifying on the *a priori* variables histology and stage, which are both independently associated with disease recurrence.

To evaluate whether HPV type was associated with recurrence in the cervical cancer dataset of The Cancer Genome Atlas, we downloaded the RNA-sequencing, whole genome sequencing (WGS), and/or whole exome sequencing (WES) data for 302 cervical cancer samples from TCGA[16]. HPV typing was determined using CTAT-Virus Integration Finder developed by our group[8]. Sample HPV type was determined based on support with at least 50 mapped reads, >0.10 fraction genome coverage. Based on the inferred HPV type, samples were categorized into HPV clades α9, and "α7 plus other" as described above. RFS and demographic variables including age, histology, and stage at diagnosis were extracted from CBioportal[17]. Survival probabilities for RFS were estimated using the Kaplan-Meier method, and the log-rank test was used to compare RFS by HPV clade. Univariable and multivariable Cox proportional hazard regression analyses were also performed to assess further whether HPV clade was an independent risk factor for RFS. Previously reported risk factors related to RFS included in all models were age, stage, and histology. All analyses were performed using Stata version 18.0 (StataCorp, College Station, TX).

Acknowledgements

This work was supported by NIH grants CA240580 (CM and JL), CA180922 (Bri.H.), and Montefiore Einstein Cancer Center grant 3A3216. AVA was supported by NIH K08 CA273527 and an American Association of Obstetricians and Gynecologists (AAOFG) Scholar Award.

Figure Legends

Figure 1. Time course characterizations of the 16 patients in the cohort with individual patients (P), type of cancer (C) either squamous carcinoma (S) or adenocarcinoma (A), diagnostic stage (stage), and HPV type identified in the tumor tissue (HPV). **Panel A** shows analysis of HPV-human DNA junctions in serum cfDNA at initial examination and at the 6-month post-treatment follow-up. **Panel B** shows detection of the cognate HPV E7 DNA in serum cfDNA at the same time points. Descriptions of the symbols are in the box included in each panel. **Panel C** shows the Integrated Genome Viewer (IGV) [31] plots of sequence reads obtained by HPV DNA hybridization capture for each sample as a function of the standard HPV genome numbering system[22]. **Panel D** shows the integrated HPV genomic structures for the ten tumors that contained only subgenomic segments of the viral genome. Viral DNAs were centered on the URR-E6-E7 stretch of the viral genomes. HPV open reading frame structure is shown at the top. The 5' splice site (5'ss) shown in the viral genome is just inside the 5' end of the E1 open reading frame. As ORF sizes and precise positions vary slightly among HPV types, integrated DNA segments may be slightly off-scale. Some of the HPV DNA insertions had more than two junctions with human DNA clustered within stretches up to tens of kilobase pairs of the human genome, a well-established phenomenon most likely due to instability of inserted HPV DNA including extra-chromosomal circularization of HPV-human DNA heterocatemer segments, and nearby reintegration of such structures[4-6,24-28]. In those instances, the junctions shown were evident as read gap borders in Panel C, and they also had the largest number of PCR-confirmed, hybridization capture + sequencing reads. **Panel E** shows the presumed genomic structures of integrated HPV DNAs in six tumors containing viral segments that were longer than unit length. All were from phylogenetic clade α9, five HPV16 and one HPV35. Arrows show the viral 5' to 3' transcriptional orientation of the HPV protein coding strands.

Figure 2. PCR detection of HPV-human DNA junctions in human genomic DNAs isolated from biopsy tissue of the cancer recurrences that occurred in the five indicated patients. PCRs for each patient included a sample of DNA from the recurrence and a positive control using the primary tumor tissue DNA from the individual patients. The junctions in the five recurrences were also detected by hybridization capture plus Illumina sequencing on the recurrence biopsies, which likewise confirmed clonal derivation from the cognate primary tumors.

Figure 3. DNA concentration determinations and Kaplan-Meier analyses performed using cfDNA from patients. **Panel A** shows total cfDNA concentrations in early- vs. late-stage cancers. **Panel B** shows the same DNA concentration determinations plotted from patients who did not have a recurrence vs. those that did. **Panels C and D** show Kaplan-Meier analyses of recurrence-free survival in patients where HPV-human DNA junctions were detected in serum vs. those where it was not detected at initial examination (**Panel C**) or at 6-months post-treatment (**Panel D**). **Panels E and F** show Kaplan-Meier analyses of recurrence-free survival in patients where HPV E7 DNA was detected in serum vs. those where it was not detected at initial examination (**Panel E**) or at 6-months post-treatment (**Panel F**). In all four plots, squares show left-censoring at the last patient visits.

Figure 4. Analyses of HPV DNAs in Patient 16. **Panel A** shows PCR analysis for HPV16 E7 DNA in serum cfDNA plus DNA from the primary tumor as a positive control. **Panel B** shows PCR analysis for HPV18 E7 DNA in serum cfDNA plus DNA from the primary tumor as a positive control. **Panel C** shows the IGV plots of sequence reads following HPV DNA hybridization capture from the primary tumor and cfDNA. Sequence reads for HPV16 and HPV18 in serum cfDNA were plotted separately.

Figure 5. Kaplan-Meier recurrence-free survival plot of 297 TCGA cervical cancer patients.

References

1. Sung, H.*, et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249 (2021).
2. Siegel, R.L., Giaquinto, A.N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J Clin* **74**, 12-49 (2024).
3. Girda, E.*, et al.* Cervical cancer treatment update: A Society of Gynecologic Oncology clinical practice statement. *Gynecol Oncol* **179**, 115-122 (2023).
4. Rossi, N.M.*, et al.* Extrachromosomal Amplification of Human Papillomavirus Episomes Is a Mechanism of Cervical Carcinogenesis. *Cancer Res* **83**, 1768-1781 (2023).
5. Van Arsdale, A.*, et al.* Structure and transcription of integrated HPV DNA in vulvar carcinomas. *NPJ Genom Med* **9**, 35 (2024).
6. Akagi, K.*, et al.* Intratumoral Heterogeneity and Clonal Evolution Induced by HPV Integration. *Cancer Discov* **13**, 910-927 (2023).
7. Morgan, I.M., DiNardo, L.J. & Windle, B. Integration of Human Papillomavirus Genomes in Head and Neck Cancer: Is It Time to Consider a Paradigm Shift? *Viruses* **9**(2017).
8. CTAT Virus Integration Finder.
9. Lu, L., Bi, J. & Bao, L. Genetic profiling of cancer with circulating tumor DNA analysis. *J Genet Genomics* **45**, 79-85 (2018).
10. Han, K.*, et al.* Clinical Validation of Human Papilloma Virus Circulating Tumor DNA for Early Detection of Residual Disease After Chemoradiation in Cervical Cancer. *J Clin Oncol* **42**, 431-440 (2024).
11. Hoppe-Seyler, K., Bossler, F., Braun, J.A., Herrmann, A.L. & Hoppe-Seyler, F. The HPV E6/E7 Oncogenes: Key Factors for Viral Carcinogenesis and Therapeutic Targets. *Trends Microbiol* **26**, 158-168 (2018).
12. Vergara, N.*, et al.* Detection and quantitation of human papillomavirus DNA in peripheral blood mononuclear cells from blood donors. *J Med Virol* **91**, 2009-2015 (2019).
13. Conceicao Gomes Nascimento, K.*, et al.* Detection of Human Papillomavirus DNA in Paired Peripheral Blood and Cervix Samples in Patients with Cervical Lesions and Healthy Individuals. *J Clin Med* **10**(2021).

14. Gagliardi, A., *et al.* Analysis of Ugandan cervical carcinomas identifies human papillomavirus clade-specific epigenome and transcriptome landscapes. *Nat Genet* **52**, 800-810 (2020).
15. Rader, J.S., *et al.* Genetic variations in human papillomavirus and cervical cancer outcomes. *Int J Cancer* **144**, 2206-2214 (2019).
16. Cerami, E., *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).
17. Gao, J., *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).
18. Leung, E., *et al.* HPV Sequencing Facilitates Ultrasensitive Detection of HPV Circulating Tumor DNA. *Clin Cancer Res* **27**, 5857-5868 (2021).
19. Sykes, P.H., *et al.* Predicting regression of cervical intraepithelial neoplasia grade 2 in women under 25 years. *Am J Obstet Gynecol* **226**, 222 e221-222 e213 (2022).
20. Tainio, K., *et al.* Clinical course of untreated cervical intraepithelial neoplasia grade 2 under active surveillance: systematic review and meta-analysis. *BMJ* **360**, k499 (2018).
21. Van Arsdale, A., *et al.* Insertional oncogenesis by HPV70 revealed by multiple genomic analyses in a clinically HPV-negative cervical cancer. *Genes Chromosomes Cancer* **59**, 84-95 (2020).
22. Papillomavirus Episteme.
23. Dobin, A., *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
24. Akagi, K., *et al.* Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res* **24**, 185-199 (2014).
25. Holmes, A., *et al.* Mechanistic signatures of HPV insertions in cervical carcinomas. *NPJ Genom Med* **1**, 16004 (2016).
26. 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 (2015).
27. Xu, B., *et al.* Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS One* **8**, e66693 (2013).
28. 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 (2016).
29. Bhatla, N. & Denny, L. FIGO Cancer Report 2018. *Int J Gynaecol Obstet* **143 Suppl 2**, 2-3 (2018).
30. Sedlis, A., *et al.* A randomized trial of pelvic radiation therapy versus no further therapy in selected patients with stage IB carcinoma of the cervix after radical hysterectomy and pelvic lymphadenectomy: A Gynecologic Oncology Group Study. *Gynecol Oncol* **73**, 177-183 (1999).
31. Deleage, C., *et al.* Defining HIV and SIV Reservoirs in Lymphoid Tissues. *Pathog Immun* **1**, 68-106 (2016).

Figure 1

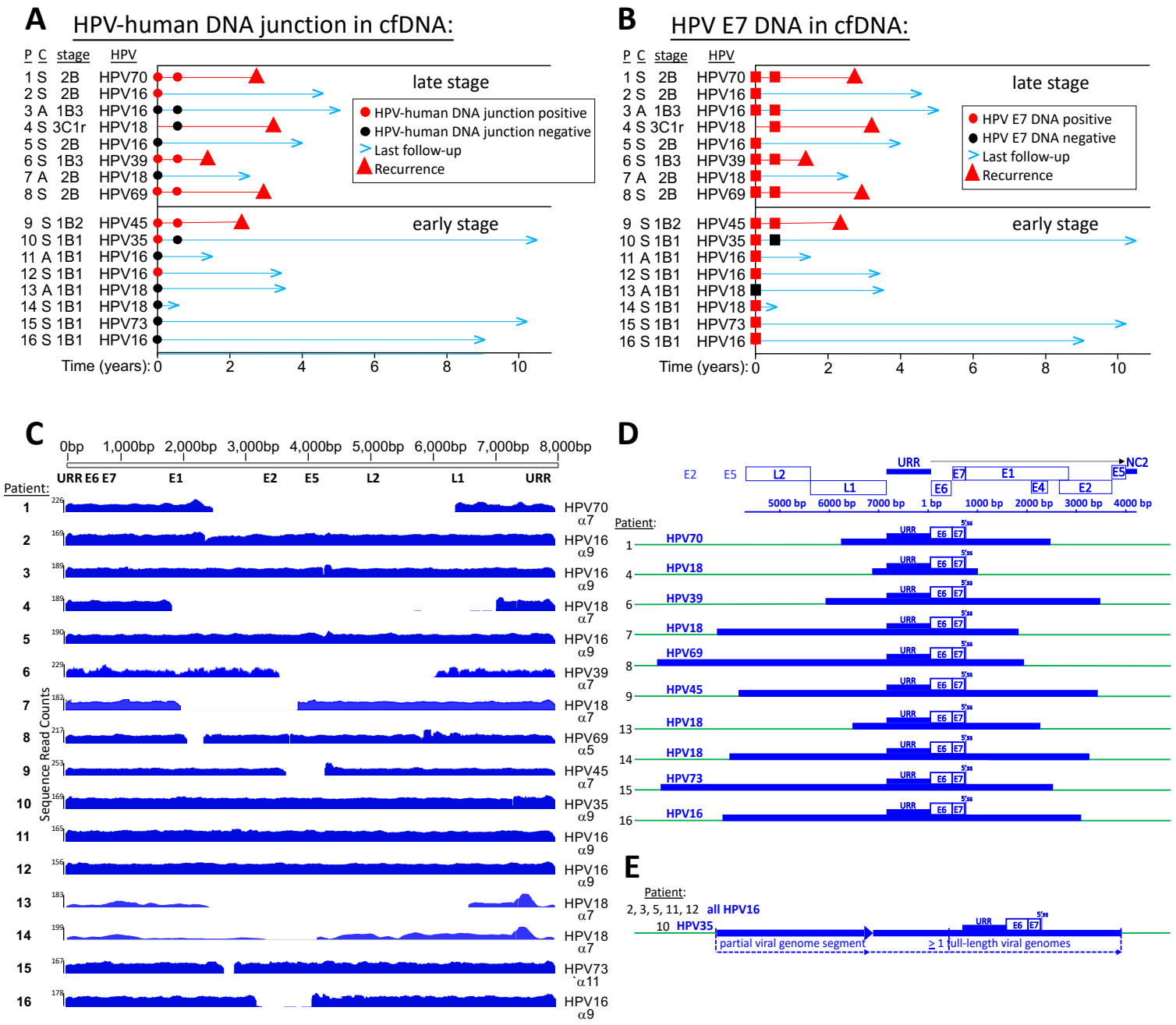

A
HPV-human DNA junction in cfDNA:
P C stage HPV
1 S 2B HPV70
2 S 2B HPV16
3 A 1B3 HPV16
4 S 3C1r HPV18
5 S 2B HPV16
6 S 1B3 HPV39
7 A 2B HPV18
8 S 2B HPV69
9 S 1B2 HPV45
10 S 1B1 HPV35
11 A 1B1 HPV16
12 S 1B1 HPV16
13 A 1B1 HPV18
14 S 1B1 HPV18
15 S 1B1 HPV73
16 S 1B1 HPV16
late stage
early stage
HPV-human DNA junction positive
HPV-human DNA junction negative
Last follow-up
Recurrence
Time (years): 0 2 4 6 8 10
B
HPV E7 DNA in cfDNA:
P C stage HPV
1 S 2B HPV70
2 S 2B HPV16
3 A 1B3 HPV16
4 S 3C1r HPV18
5 S 2B HPV16
6 S 1B3 HPV39
7 A 2B HPV18
8 S 2B HPV69
9 S 1B2 HPV45
10 S 1B1 HPV35
11 A 1B1 HPV16
12 S 1B1 HPV16
13 A 1B1 HPV18
14 S 1B1 HPV18
15 S 1B1 HPV73
16 S 1B1 HPV16
late stage
early stage
HPV E7 DNA positive
HPV E7 DNA negative
Last follow-up
Recurrence
Time (years): 0 2 4 6 8 10
C
0bp 1,000bp 2,000bp 3,000bp 4,000bp 5,000bp 6,000bp 7,000bp 8,000bp
URR E6 E7 E1 E2 E5 L2 L1 URR
Patient:
Sequence Read Counts
1 HPV70 α7
2 HPV16 α9
3 HPV16 α9
4 HPV18 α7
5 HPV16 α9
6 HPV39 α7
7 HPV18 α7
8 HPV69 α5
9 HPV45 α7
10 HPV35 α9
11 HPV16 α9
12 HPV16 α9
13 HPV18 α7
14 HPV18 α7
15 HPV73 α11
16 HPV16 α9
D
E2 E5 L2 L1 URR E6 E7 E1 E4 E2 E5 NC2
5000 bp 6000 bp 7000 bp 1 bp 1000 bp 2000 bp 3000 bp 4000 bp
Patient:
1 HPV70
4 HPV18
6 HPV39
7 HPV18
8 HPV69
9 HPV45
13 HPV18
14 HPV18
15 HPV73
16 HPV16
URR E6 E7 5'ss
E
Patient:
2, 3, 5, 11, 12 all HPV16
10 HPV35
partial viral genome segment
≥ 1 full-length viral genomes

Table 1. HPV types, viral phylogenetic clades, human genome integration sites, serum cfDNA coverage by HC+SEQ, and serum cfDNA type-specific E7 PCR

Patient	Stage	Cancer[a]	HPV type[b]	Clade	Human genome HPV DNA position	Serum cfDNA HPV genome coverage	Type-specific HPV E7 PCR
1	2B	SCC	HPV70	α7	Chr14:99,216,916-Chr14:99,216,952	0.506	positive
2	2B	SCC	HPV16	α9	Chr11:119,699,940-119,699,962	0.238	positive
3	1B3	AC	HPV16	α9	ChrX:137,024,646[c]	0.237	positive
4	3C1r	SCC	HPV18	α7	Chr8:126,992,703-127,083,002	0	positive
5	2B	SCC	HPV16	α9	Chr5:1,346,104-1,386,461[d] Chr6:106,847,095-106,849,161	1	positive
6	1B3	SCC	HPV39	α7	Chr14:102,789,751-102,790,038	0.673	positive
7	2B	AC	HPV18	α7	Chr16:71,705,032-71,732,359	0.33	positive
8	2B	SCC	HPV69	α5	Chr9:87,266,100-87,266,133	0.903	positive
9	1B2	SCC	HPV45	α7	ChrX:13,185,650-13,391,674	ND[e]	positive
10	1B1	SCC	HPV35	α9	Chr8:59,436,141-59,436,178	ND	positive
11	1B1	AC	HPV16	α9	Chr7:117,590,654-117,605,110	0.726	negative
12	1B1	SCC	HPV16	α9	Chr2:63,048,331-63,048,344	0	positive
13	1B1	AC	HPV18	α7	Chr6:134,482,206-134,527,589	0.818	negative
14	1B1	SCC	HPV18	α7	Chr15:89,108,581-89,144,328	0.049	positive
15	1B1	SCC	HPV73	α11	Chr4:67,867,259-68,002,635	0.087	negative
16	1B1	SCC	HPV16	α9	ChrX:35,045,995-35,046,002	0.982/0.919[f]	positive

[a]SCC, squamous cell carcinoma; AC, adenocarcinoma.
[b]HPV type determined by HC+SEQ on tumor tissue.
[c]Only one junction between HPV16 and human DNA was detected by hybridization capture for which effective PCR primers were developed for Patient 3.
[d]Tumor 5 had two DNA insertions of the identical HPV16 variant, one on chromosome 5 and the other on chromosome 6 at the indicated positions.
[e]ND, not done.
[f]Both HPV16 and HPV18 were present in serum cfDNA of Patient 16 with respective genome coverages shown.

Figure 2

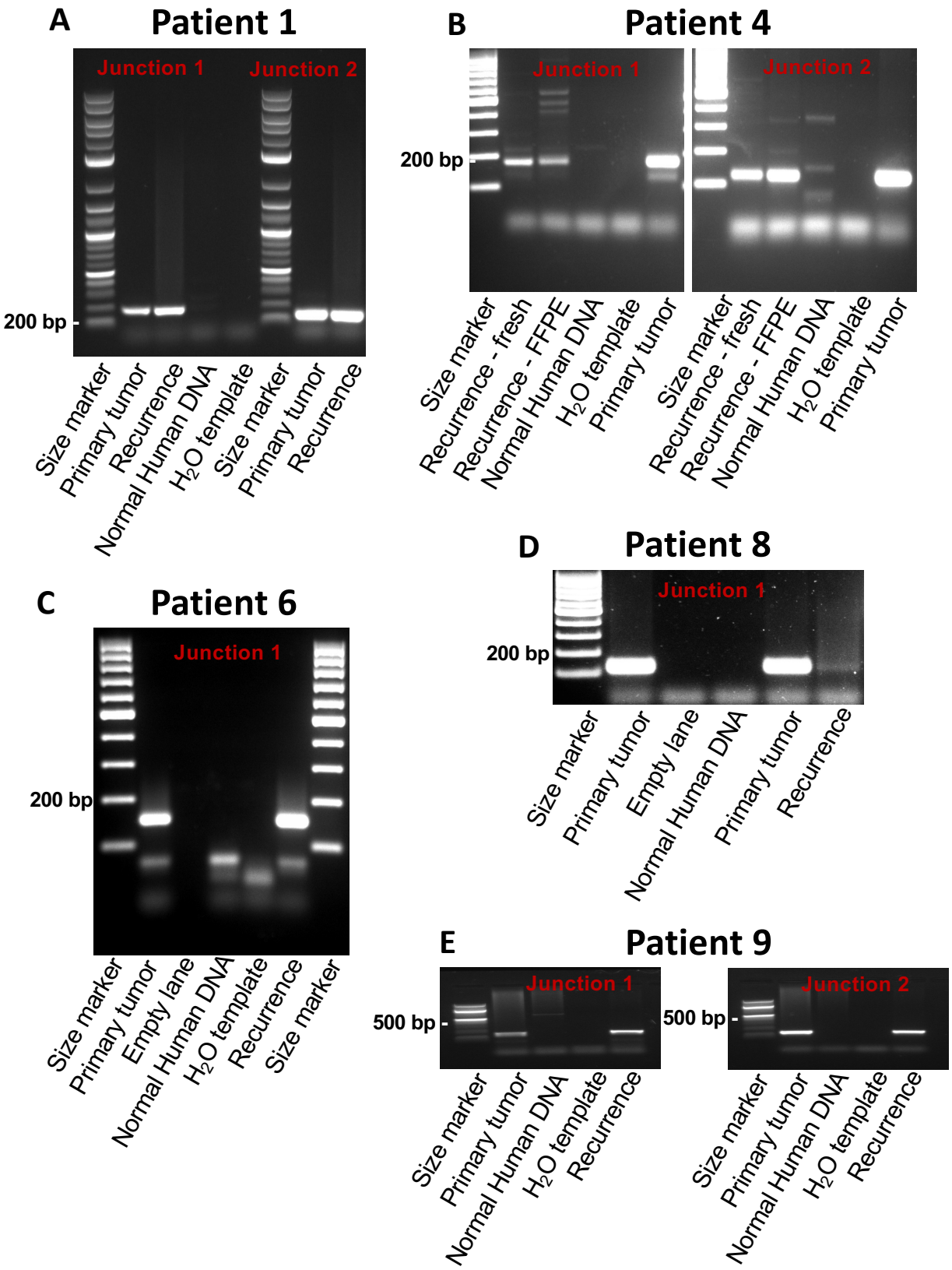

A
Patient 1
Junction 1
Junction 2
200 bp
Size marker
Primary tumor
Recurrence
Normal Human DNA
H2O template
Size marker
Primary tumor
Recurrence
B
Patient 4
Junction 1
Junction 2
200 bp
Size marker
Recurrence - fresh
Recurrence - FFPE
Normal Human DNA
H2O template
Primary tumor
Size marker
Recurrence - fresh
Recurrence - FFPE
Normal Human DNA
H2O template
Primary tumor
C
Patient 6
Junction 1
200 bp
Size marker
Primary tumor
Empty lane
Normal Human DNA
H2O template
Recurrence
Size marker
D
Patient 8
Junction 1
200 bp
Size marker
Primary tumor
Empty lane
Normal Human DNA
Primary tumor
Recurrence
E
Patient 9
Junction 1
500 bp
Size marker
Primary tumor
Normal Human DNA
H2O template
Recurrence
Junction 2
500 bp
Size marker
Primary tumor
Normal Human DNA
H2O template
Recurrence

Figure 3

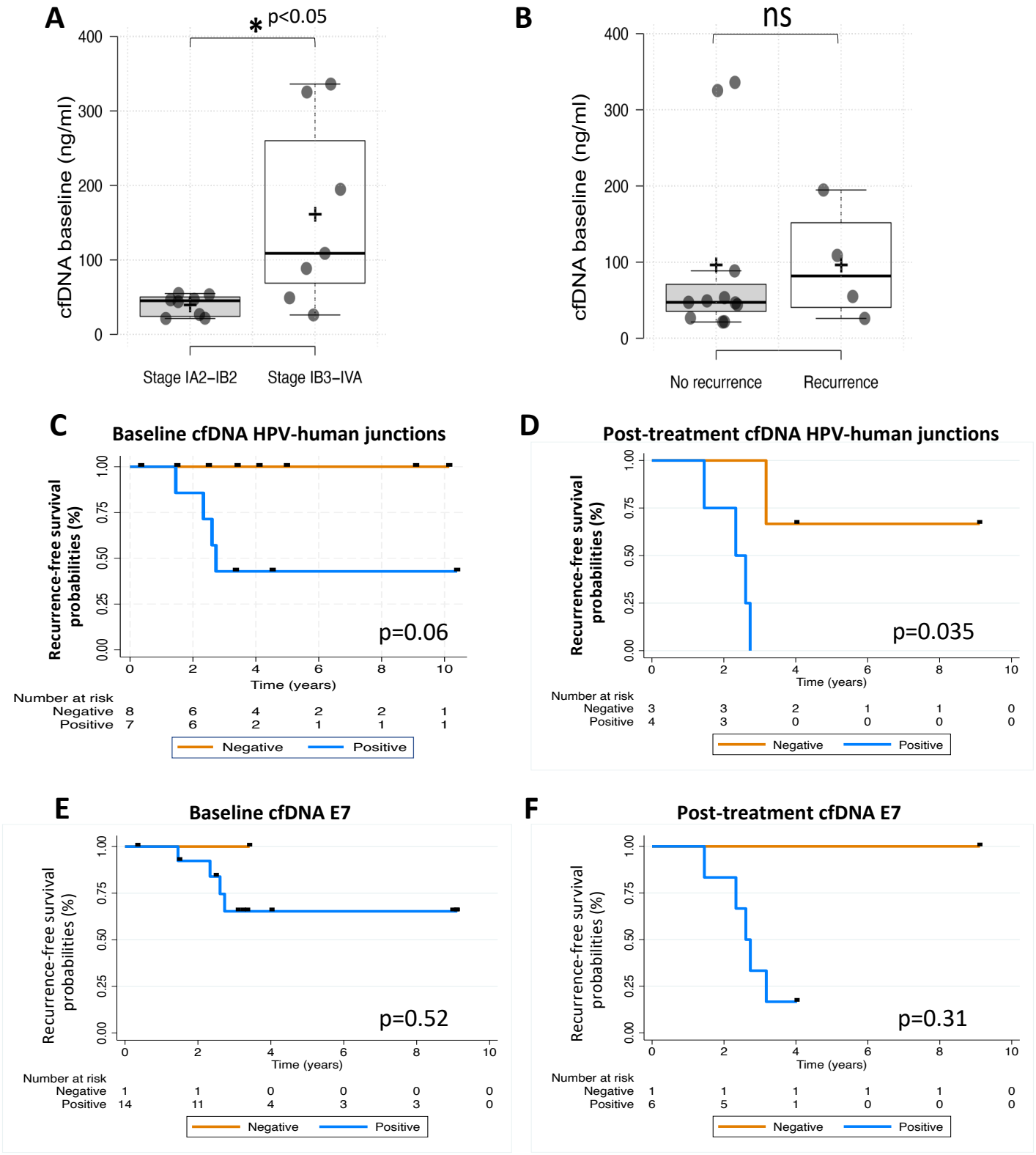
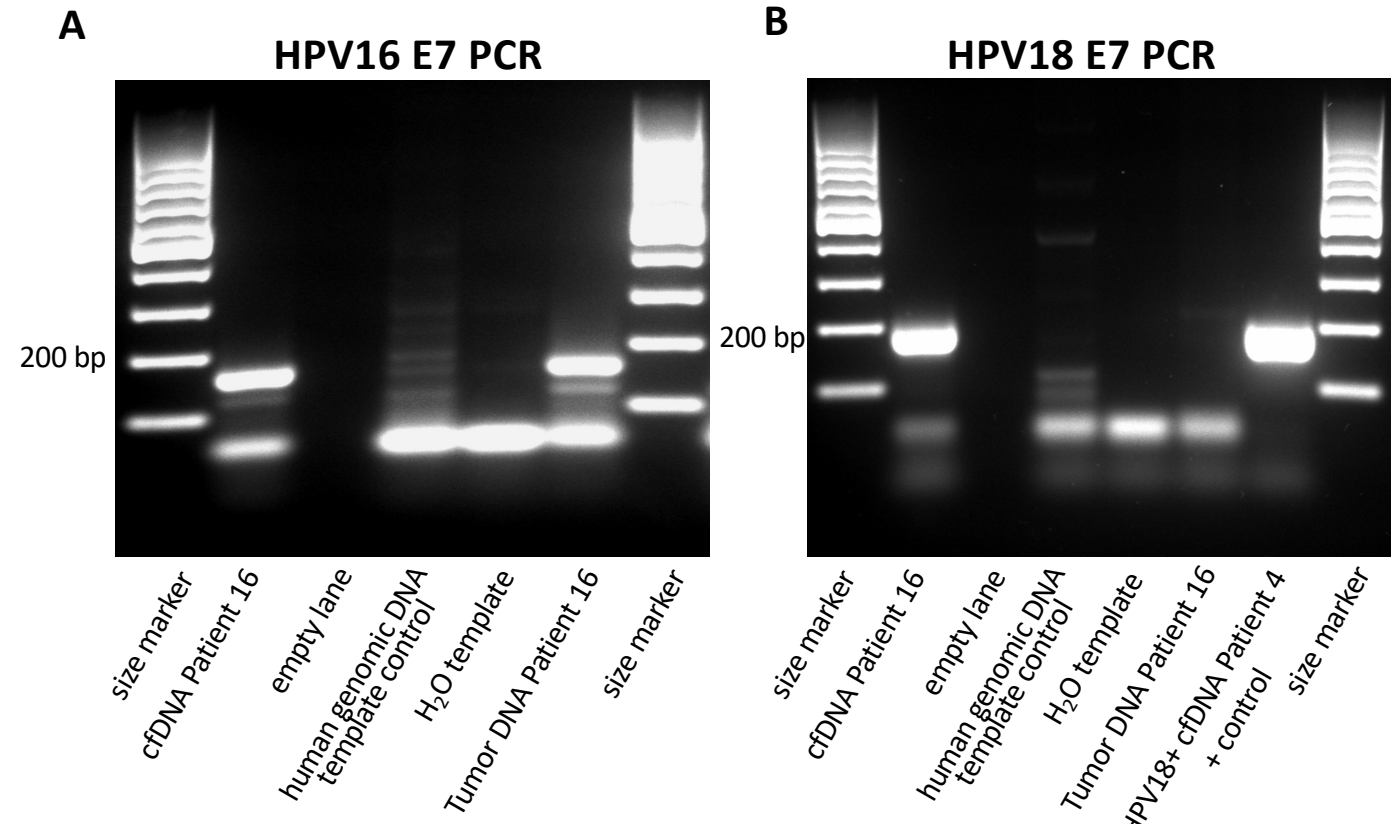

A
* p<0.05
cfDNA baseline (ng/ml)
400
300
200
100
0
Stage IA2–IB2
Stage IB3–IVA
B
ns
cfDNA baseline (ng/ml)
No recurrence
Recurrence
C
Baseline cfDNA HPV-human junctions
Recurrence-free survival probabilities (%)
1.00
0.75
0.50
0.25
0.00
p=0.06
Time (years)
Number at risk
Negative 8 6 4 2 2 1
Positive 7 6 2 1 1 1
Negative
Positive
D
Post-treatment cfDNA HPV-human junctions
Recurrence-free survival probabilities (%)
p=0.035
Time (years)
Number at risk
Negative 3 3 2 1 1 0
Positive 4 3 0 0 0 0
Negative
Positive
E
Baseline cfDNA E7
Recurrence-free survival probabilities (%)
p=0.52
Time (years)
Number at risk
Negative 1 1 0 0 0 0
Positive 14 11 4 3 3 0
Negative
Positive
F
Post-treatment cfDNA E7
Recurrence-free survival probabilities (%)
p=0.31
Time (years)
Number at risk
Negative 1 1 1 1 1 0
Positive 6 5 1 0 0 0
Negative
Positive

Figure 4


A
HPV16 E7 PCR
200 bp
size marker
cfDNA Patient 16
empty lane
human genomic DNA template control
H2O template
Tumor DNA Patient 16
size marker
B
HPV18 E7 PCR
200 bp
size marker
cfDNA Patient 16
empty lane
human genomic DNA template control
H2O template
Tumor DNA Patient 16
HPV18+ cfDNA Patient 4 + control
size marker

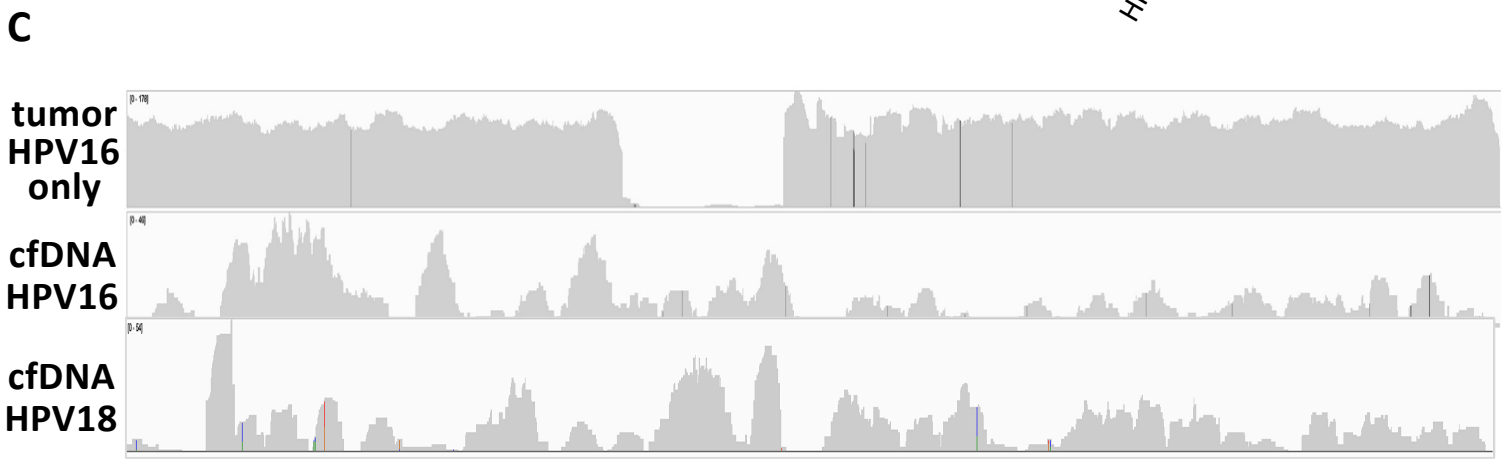

C
tumor HPV16 only
cfDNA HPV16
cfDNA HPV18

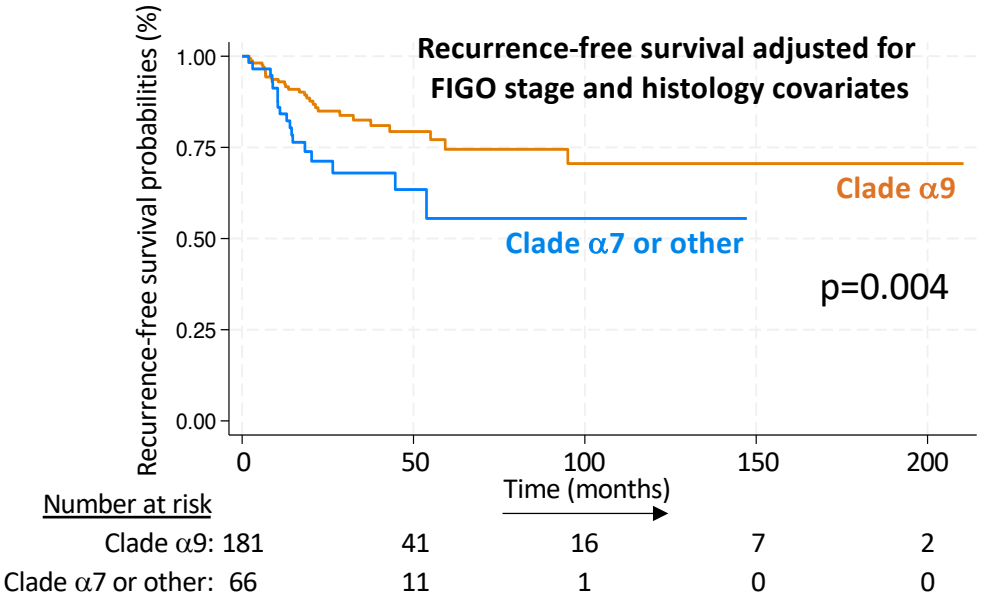
Table 2. *A priori* clinicopathologic variables of the TCGA cohort by HPV phylogenetic clade

Variable	**Cohort (n=297)**	**Clade 9 (n=214)**	**Clade 7 or other (n=83)**	**p-value**
Age (years)	47.8 (13.6)	48.6 (13.8)	45.8 (13.0)	0.12
Histology				
Squamous Cell	254 (84.4)	183 (85.9)	67 (83.8)	0.71
Adenocarcinoma	47 (15.6)	30 (14.1)	13 (16.2)	
Stage at Dx				0.57
Stage I	156 (54.6)	113 (54.6)	43 (54.4)	
Stage II	65 (22.7)	44 (21.3)	21 (26.6)	
Stage III	46 (16.1)	34 (16.4)	12 (15.2)	
Stage IV	19 (6.6)	16 (7.7)	3 (3.8)	

Table 3. Multivariable Cox proportional hazard regression model for recurrence-free survival

Variable	Multivariable Hazard Ratio	95% CI	p-value
Clade			
Clade 9	1.00	--	--
Clade 7	2.48	1.34-4.49	0.004
Age (years)	1.01	0.99-1.04	0.28
Histology			
Squamous	1.00	--	--
Adenocarcinoma	1.08	0.51-2.28	0.84
Stage at Dx			
Stage I	1.00	--	--
Stage II	0.63	0.28-1.41	0.26
Stage III	0.43	0.13-1.46	0.18
Stage IV	4.40	1.89-10.18	0.001

Figure 5


Recurrence-free survival adjusted for
FIGO stage and histology covariates
Recurrence-free survival probabilities (%)
1.00
0.75
0.50
0.25
0.00
Clade α9
Clade α7 or other
p=0.004
0
50
100
150
200
Time (months)
Number at risk
Clade α9: 181 41 16 7 2
Clade α7 or other: 66 11 1 0 0

Supplementary Table 2. Patient clinical parameters

Variable	Cohort (n=16)	No Recurrence (n=11)	Recurrence (n=5)	p-value[a]
Age (years)	57.1 (13.7)	52.6 (12.4)	67 (12.0)	0.048
Histology				
Squamous Cell	12 (75.0)	7 (63.6)	5 (100.0)	0.25
Adenocarcinoma	4 (25.0)	4 (36.4)	0 (0.0)	
Stage				0.28
Early (IA-IB2)	8 (50.0)	7 (63.6)	1 (20.0)	
LACC (IB3-IVA)	8 (50.0)	4 (36.4)	4 (80.0)	
HPV by Clade				0.013
Clade 7	6 (37.5)	2 (18.2)	4 (80.0)	
Clade 9	8 (50.0)	8 (72.7)	0 (0.0)	
Other	2 (12.5)	1 (9.1)	1 (20.0)	
Baseline cfDNA concentration (ng/mL)	49.1 [21.4,336]	47.2 [21.4, 336]	81.9 [26.1, 194.7]	0.49
Surgery				0.30
No	6 (37.5)	3 (27.3)	3 (60.0)	
Yes	10 (62.5)	8 (72.7)	2 (40.0)	
EBRT + brachy[b]				0.25
No	4 (25.0)	4 (36.4)	0 (0.0)	
Yes	12 (75.0)	7 (63.6)	5 (100.0)	
Total Dose (Gy)	69 [60,84]	60 [60,78]	81 [60,84]	0.07
Chemotherapy				0.31
No	7 (43.8)	6 (54.6)	1 (20.0)	
Yes	9 (56.2)	5 (45.4)	4 (80.0)	

[a]Students t-test, Wilcoxon test, or Fisher's Exact test were utilized for normally distributed variables, non-normal variables and categorical variables respectively. $p<0.05$ was considered statistically significant.

[b]External beam radiotherapy (EBRT) and brachytherapy (brachy)