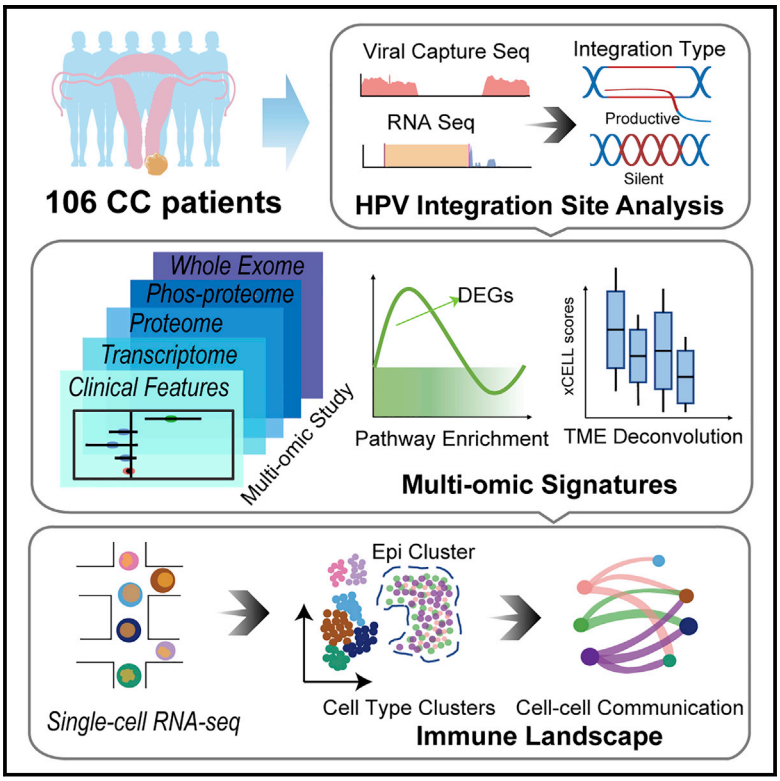
Cell Genomics

Article

Multi-omics characterization of silent and productive HPV integration in cervical cancer

Graphical abstract



Highlights

- HPV integration in cervical cancer shows silent or productive integration status
- Productive HPV integration is under selection and likely contributes to CC pathophysiology
- Productive HPV integration associates with tumor aggressiveness and immunoevasion

Authors

Junpeng Fan, Yu Fu, Wenju Peng, ..., Gang Chen, Ding Ma, Chaoyang Sun

Correspondence

cqzl_zdl@163.com (D.Z.), xiabairong@ustc.edu.cn (B.X.), songkun2001226@sdu.edu.cn (K.S.), tjchengang@hust.edu.cn (G.C.), dingma424@126.com (D.M.), suncydoctor@gmail.com (C.S.)

In brief

Using multi-omics and single-cell data, Fan et al. comprehensively demonstrate the characteristics of silent and productive HPV integration in a large-scale cervical cancer cohort. They reveal that productive HPV integrations are under selection and associated with enhanced tumor aggressiveness, immunoevasion, and tumor progression. The multidimensional integrative analyses deepen our knowledge of HPV integration and have the potential to inform patient management.

Fan et al., 2023, Cell Genomics 3, 100211
January 11, 2023 © 2022 The Authors.
https://doi.org/10.1016/j.xgen.2022.100211

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Multi-omics characterization of silent and productive HPV integration in cervical cancer

Junpeng Fan,[1,13] Yu Fu,[1,13] Wenju Peng,[1,13] Xiong Li,[2,13] Yuanming Shen,[3,13] Ensong Guo,[1] Funian Lu,[1] Shengtao Zhou,[4] Si Liu,[1] Bin Yang,[1] Xu Qin,[1] Dianxing Hu,[1] Rourou Xiao,[1] Xi Li,[1] Siqi Yang,[3] Cunzhong Yuan,[5,6,7] Yao Shu,[5,6,7] He Huang,[8] Ting Wan,[8] Yanan Pi,[9] Shuxiang Wang,[9] Wenjuan Chen,[10] Haixia Wang,[10] Lin Zhong,[10] Li Yuan,[10] Baogang Wen,[10] Beihua Kong,[5,6,7] Gordon B. Mills,[11,12] Dongling Zou,[10,*] Bairong Xia,[9,*] Kun Song,[5,6,7,*] Gang Chen,[1,*] Ding Ma,[1,*] and Chaoyang Sun[1,14,*]

[1]Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China
[2]Department of Gynecology & Obstetrics, Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China
[3]Department of Gynecologic Oncology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310000, China
[4]Department of Obstetrics and Gynecology, Key Laboratory of Birth Defects and Related Diseases of Women and Children of MOE and State Key Laboratory of Biotherapy, West China Second Hospital, Sichuan University and Collaborative Innovation Center, Chengdu 610000, China
[5]Department of Obstetrics and Gynecology, Qilu Hospital of Shandong University, Jinan 250000, China
[6]Gynecology Oncology Key Laboratory, Qilu Hospital of Shandong University, Jinan 250000, China
[7]Division of Gynecology Oncology, Qilu Hospital of Shandong University, Jinan 250000, China
[8]Department of Gynecologic Oncology, Sun Yat-sen University Cancer Center, Guangzhou 510000, China
[9]Department of Gynecology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230000, China
[10]Chongqing Key Laboratory of Translational Research for Cancer Metastasis and Individualized Treatment, Chongqing University Cancer Hospital, Chongqing 404100, China
[11]Department of Cell, Developmental, and Cancer Biology, Oregon Health and Sciences University, Portland, OR 97201, USA
[12]Knight Cancer Institute, Portland, OR 97201, USA
[13]These authors contributed equally
[14]Lead contact
*Correspondence: cqzl_zdl@163.com (D.Z.), xiabairong@ustc.edu.cn (B.X.), songkun2001226@sdu.edu.cn (K.S.), tjchengang@hust.edu.cn (G.C.), dingma424@126.com (D.M.), suncydoctor@gmail.com (C.S.)
https://doi.org/10.1016/j.xgen.2022.100211

SUMMARY

Cervical cancer (CC) that is caused by high-risk human papillomavirus (HPV) remains a significant public health problem worldwide. HPV integration sites can be silent or actively transcribed, leading to the production of viral-host fusion transcripts. Herein, we demonstrate that only productive HPV integration sites were nonrandomly distributed across both viral and host genomes, suggesting that productive integration sites are under selection and likely to contribute to CC pathophysiology. Furthermore, using large-scale, multi-omics (clinical, genomic, transcriptional, proteomic, phosphoproteomic, and single-cell) data, we demonstrate that tumors with productive HPV integration are associated with higher E6/E7 proteins and enhanced tumor aggressiveness and immunoevasion. Importantly, productive HPV integration increases from carcinoma *in situ* to advanced disease. This study improves our understanding of the functional consequences of HPV fusion transcripts on the biology and pathophysiology of HPV-driven CCs, suggesting that productive HPV integration should be evaluated as an indicator of high risk for progression to aggressive cancers.

INTRODUCTION

Cervical cancer (CC), which is often caused by high-risk human papillomavirus (HPV), remains the second most common malignancy of the female reproductive system worldwide and the eighth most frequent cause of death.[1] Although mortality rates have declined significantly with the implementation of prophylactic vaccination and successful screening strategies in First World countries,[2] CC continues to be a significant public health problem, particularly in low-income countries with limited access to high-quality health care.[3]

HPV integration into the host genome is a critical etiological event in cervical carcinogenesis and progression.[4,5] HPV integration sites have been reported to be randomly distributed throughout the host genome.[6] However, conflicting results exist with regard to a potential role for nonrandom integration in

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

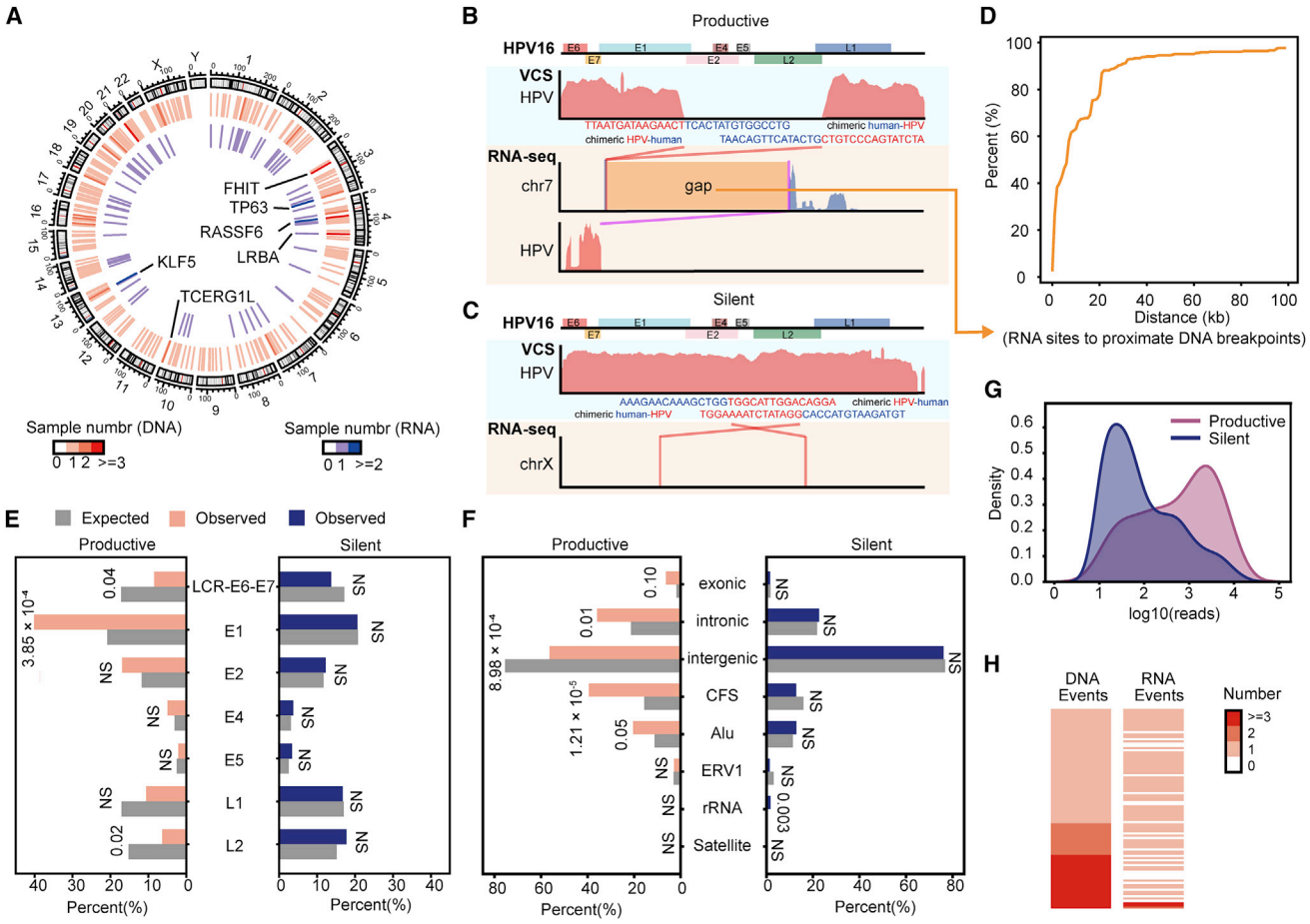


Figure 1. Characteristics of HPV integration at genomic and transcriptional levels

(A) Distribution of genomic and transcriptional HPV integration in the human genome in 98 cervical cancers. The outer circle represents the human genome, with chromosomes labeled 1–22, X, and Y. The middle and inner circles represent the distribution of genomic (DNA) and transcriptional (RNA) HPV integration across the human genome. The human genome is separated into 500 kb bins, and the number of samples that had an integration site in each bin was counted and colored red (DNA) or blue (RNA). Recurrent protein-coding genes closest to integration sites are labeled.

(B and C) Representative diagrams for productive (B) and silent (C) HPV integration. (Top) HPV genomic regions colored by gene. (Middle) Read depth of HPV genome (colored red) derived from VCS (virus capture sequencing) data. The bases under the graph are the chimeric sequence supporting the HPV DNA integration site and colored by source (red for HPV and blue for human). (Bottom) Read depth of HPV-human fusion transcripts derived from RNA sequencing (RNA-seq). Read depth of HPV is colored red and that of human is colored blue. The loci of the genomic integration site on HPV and human are linked by red lines and that of the transcriptional integration site is linked by a pink line. The gap between genomic and transcriptional integration sites on the human genome is colored orange.

(D) Cumulative distribution plot for the gaps between RNA fusion sites and their proximate DNA breakpoints within samples.

(E and F) Comparisons of the observed and expected percentages of breakpoints in various regions of the HPV16 genome (E) and the human genome (F) in silent and productive HPV integration. The expected values for the indicated HPV regions were calculated by random selection of breakpoints across the HPV genome. The p values were calculated by two-sided chi-square test. CFS, common fragile site; Alu, Alu element; ERV1, endogenous retroviral sequence 1; rRNA, ribosomal RNA.

(G) Density plot of the supporting reads from VCS for productive or silent HPV integration sites.

(H) Comparison of DNA and RNA HPV integration events within samples. HPV integration events are clusters of integration sites made by grouping integration sites within 500 kb in each sample. The bars are colored according to the number of integration events in each patient.

transcriptionally active regions or near to cancer-related genes.[4,7] These contradictory results are at least partially attributable to discordant results from genomic sequencing (whole-genome sequencing, HPV captured sequencing), RNA sequencing (RNA-seq), and molecular (PCR-based MassARRAY assay) approaches. Moreover, whether HPV integration and production of fusion transcripts (productive integration) is associated with a selective growth advantage during progression of CC remains unclear.

By assessing the consequences of HPV genomic integration in a cohort of 106 CC patients, we demonstrate that HPV integration leading to expression of fusion transcripts is associated with CC progression. Further, multidimensional integrative analyses (clinical, pathological, genomic, transcriptional, proteomic, and phosphoproteomic data) demonstrate that differences between silent and productive HPV integration have the potential to inform patient management. Single-cell RNA-seq (scRNA-seq) reveals

CellPress
OPEN ACCESS

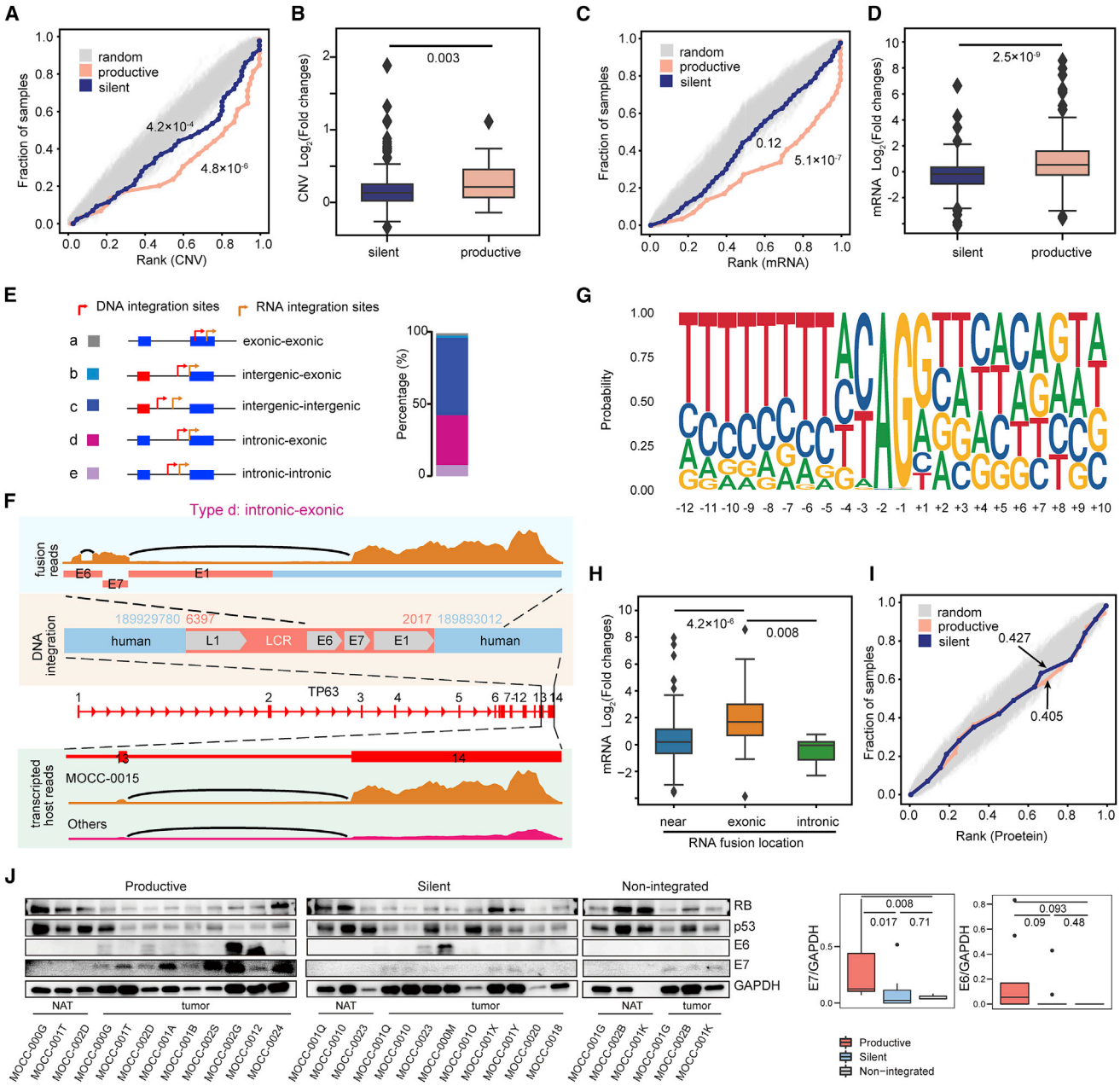


Figure 2. Impacts of various HPV integrations on local transcriptome and HPV oncoproteins

(A) Cumulative distribution curves for somatic copy number alteration ranks for genes near productive or silent HPV integration sites.

(B) Fold change in somatic copy number for genes near productive or silent HPV integration sites.

(C) Cumulative distribution curves for mRNA abundance ranks for genes near productive or silent HPV integration sites.

(D) Fold change in mRNA expression for genes near productive or silent HPV integration sites.

(E) Frequency of five types of DNA-RNA integration models based on DNA and RNA breakpoints of the most highly expressed fusion transcripts in individual patients. DNA breakpoints are labeled by red and RNA breakpoints are labeled by orange. Boxes of the same color represent exons of the same gene, and different colors represent different genes.

(F) An example of the type d DNA-RNA integration model with HPV integration at the *TP63* gene region in patient MOCC-0015. (Top) Read depth of HPV-human fusion transcripts spliced from HPV to the 14th exon of *TP63* derived from RNA-seq data. (Middle) HPV DNA integration pattern inferred from VCS data. (Bottom) The transcriptional read depth of the *TP63* gene region in patient MOCC-0015 and average read depth in other patients.

(G) Consensus motif of virus-host fusion transcripts around the breakpoints in the host genome. Bases located within ±12 bp of the breakpoints were included for analysis.

(H) Fold change in mRNA expression of genes near productive HPV integration sites according to the location of the integration site relative to the gene (exonic, intronic, or near).

(legend continued on next page)

the heterogeneity of the effects of HPV integration on transcriptional reprogramming in tumor cells and cell-immune interactions that increase our understanding of the effects of silent and productive HPV integration on CC pathophysiology.

RESULTS

Silent and productive HPV integration is associated with different functional consequences

To investigate the effects of HPV integration on the progression of CC, we performed multi-omics analyses, including high-throughput HPV capture sequencing, whole-exome sequencing, transcriptome sequencing, proteomics, and phosphoproteomics in paired tumors and normal adjacent tissues from 98 CCs. An additional 8 CC tumors were profiled by scRNA-seq to examine the heterogeneity of HPV integration status and functional effects on transcriptional reprogramming. Among the 98 CCs, 72 were HPV16 positive, 6 were HPV18 positive, and 2 were HPV negative, with the rest comprising other viral strains such as HPV58 and HPV31. Detailed clinical parameters and molecular assays performed on each patient are shown in Figures S1A and S1B and Table S1.

Genomic HPV integration was detected with our virus capture sequencing (VCS)[4] technology using an algorithm that decreases false positives.[8] In total, 762 virus-host DNA fusion breakpoints were found in 90 of 98 CCs (Table S1). Six HPV(+) tumors did not demonstrate HPV integration, and 2 CCs were HPV negative. Genomic HPV integrations were further subclassified as human-HPV or HPV-human, based on the direction of the HPV sequence fusion to the host genome (see STAR Methods and Figure S1C).

To study the distribution of HPV integration sites across the human genome, we separated the genome into 500 kb bins and counted the samples that had an integration site in each bin. As our study was performed only in XX women, we found that genomic HPV integration breakpoints were broadly distributed across all chromosomes, excepting the Y chromosome (Figure 1A), and only 21 recurrent integration regions were identified in two or more CCs, including genomic regions near *FHIT* (in 3 of 90 patients), *TP63* (in 2 of 90 patients), and *KLF5* (in 2 of 90 patients), which is consistent with our previous study[4] (Figure 1A).

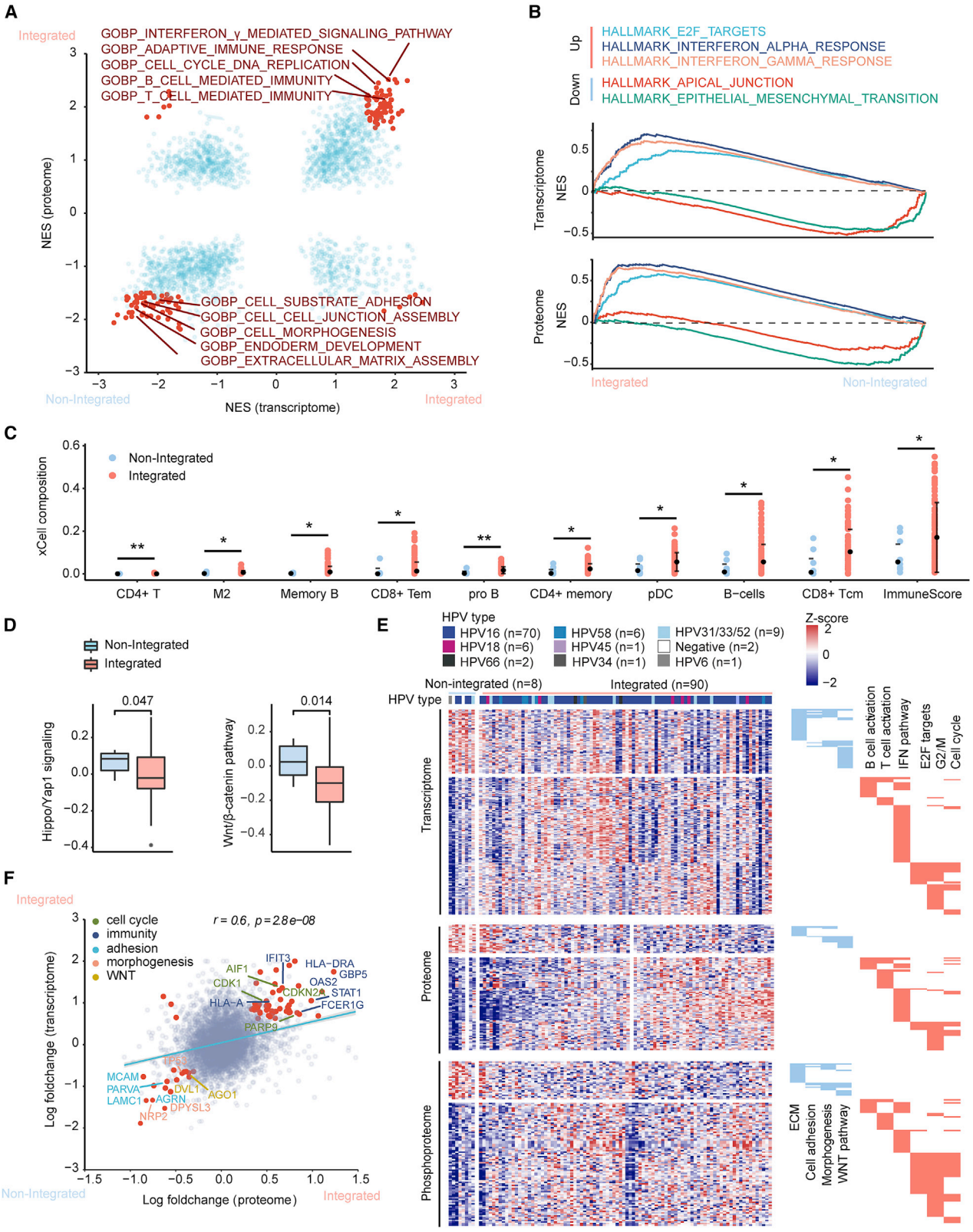
In contrast, the assessment of fusion transcripts in RNA-seq data revealed only 388 virus-host RNA fusion breakpoints, suggesting that not all genomic HPV integration sites were transcribed. We subsequently assessed the relationships between HPV DNA and RNA integration sites. Because of RNA splicing, there are gaps between DNA integration sites and their RNA fusion breakpoints (Figures 1B and 1C). We first evaluated the distances between RNA integration sites and their nearest genomic integration sites. Interestingly, most RNA integration sites were within 40 kb of the proximal DNA breakpoint (Figure 1D), while 379 (97.7%) of the RNA integration sites were less than 100 kb from their associated genomic site. Thus, DNA integration sites were defined as productive if there existed an RNA integration site <100 kb from its associated site in the human genome with identical integration type in the same tumor (see STAR Methods). Overall, 142 (18.6%) of the genomic integration sites were defined as productive DNA integrative sites (Figure S1D). Furthermore, we found that the most highly expressed fusion transcripts (Figures S1E) were initiated from the promoter in the viral upstream regulatory region (URR), which resulted in fusion transcripts including the *E6*/*E7* genes, with the *E1* and *E2* genes commonly being disrupted (Figure 1B), suggesting that fusion transcription is primarily driven by the URR within the viral genome.

We subsequently explored genomic features differentiating productive and silent HPV integration sites. First, silent breakpoints were distributed randomly across the viral genome (Figure 1E, right). In contrast, productive integration was nonrandom, with a statistically significant increase in breakpoints in the *E1* region and a tendency for an increase in the *E2* region of HPV (Figure 1E, left). This was associated with a statistically significant decrease in breakpoints located in the *L2* region and LCR-*E6*-*E7* viral regions (Figure 1E, left). Secord, silent integration sites were similarly distributed randomly across the human genome (Figure 1F, right). In contrast, there was a statistically significant increase in productive integration breakpoints within introns, common fragile sites (CFS), and Alu regions and a tendency for increases in exonic regions (Figure 1F, left). This was associated with a statistically significant decrease in breakpoints located in intergenic regions (Figure 1F, left). The differences in the randomness of silent and productive integration may contribute to previous discrepant reports related to randomness of HPV integration.[7,9,10] Furthermore, the nonrandomness of productive integration sites is consistent with a contention that productive integration sites are under selective pressure and likely contribute to CC pathophysiology. Third, productive integration sites had more reads supporting DNA integration than silent integration sites (Figure 1G). Notably, the featured characteristics of both productive and silent integration sites in the total CC population are recapitulated in HPV16-positive tumors (Figure S1F).

Different nearby HPV integration sites may be from the same integration event, due to viral sequence amplification in DNA and alternative splicing in RNA. So, we grouped HPV integration sites near one another (<500 kb apart based on a previous study[11]) within the same samples into a DNA or RNA "integration event." Interestingly, while nearly half of the tumors had two or more DNA integration events, only one HPV RNA integration

(I) Cumulative distribution curves for protein abundance ranks for genes near productive or silent HPV integration sites.
(J) Western blot (left) and relative quantification (right) of HPV16 oncoproteins and the indicated host proteins in HPV16-positive patients. NAT, normal adjacent tissue. In (A), (C), and (I), the ranks of the genes near HPV integration sites (<40 kb) were determined by ranking the corresponding value of each gene in the sample harboring the integration site with all other samples lacking the integration site. The p values were calculated by comparing the curve with that of 500 random sampling sites (gray) using a two-sided Kolmogorov-Smirnov test and shown as the median value of 500 replicates. In (B), (D), and (G), the fold changes in the genes (<40 kb from integration sites) were calculated by comparing the corresponding value of genes in the sample harboring the integration site with the same gene in all other samples lacking the integration site. The p values were calculated by two-sided Wilcoxon rank-sum test if not specified.

CellPress
OPEN ACCESS



(legend on next page)

event was found across the majority of patients (Figure 1H), further supporting selection of the host cells with productive integration. Moreover, there were a number of samples with an observed DNA integration event but no RNA integration event, supporting the existence of silent HPV integration Figure 1H).

Impacts of various HPV integrations on the local transcriptome

We next determined whether productive or silent integration sites were associated with changes in focal genome structure and gene transcription. Consistent with a previous study,[12] DNA copy number variation (CNV) increases were colocalized with HPV integration sites (Figures 2A and 2B). However, productive integration sites were statistically significantly more likely to be associated with focal CNV amplification than silent integration sites (Figures 2A and 2B). Strikingly, using a 40 kb region, productive but not silent HPV integration was associated with an increase in the expression of host genes (Figures 2C and 2D).

Polyadenylation is important for mRNA stability and translational efficiency.[13] The disruption of the *E1* or *E2* region (Figure 1E) renders the viral early polyadenylation signal (PAS) unavailable for polyadenylation of viral *E6/E7* bicistronic transcripts, and thus, the expression of *E6/E7* transcripts is dependent on a host PAS in close proximity to the insertion site. As indicated in Figures 1B and 1C, expression of fusion genes is under control of the viral URR,[14] and alternative splicing between virus and host contributes to formation of virus-host fusion transcripts by hijacking a host 3′ splice site and the host RNA PAS (Figures 1B and 1C). We thus explored the location of the most highly expressed breakpoints in DNA and RNA for productive expression in each sample and found that the expression of fusion RNAs could be separated into five models based on whether the DNA and RNA breakpoints were exonic, intronic, or intergenic (Figure 2E). The majority of the productive integration sites had both the DNA and the RNA breakpoint in the intergenic region, with the next most common being the DNA breakpoint in the intronic region and the RNA breakpoint in an exonic region. Where RNA breakpoints are in exons (type d, Figure 2E), the 3′ splice sites in the exon and the polyadenylation sites of the corresponding gene are involved in the facilitated production of the fusion transcripts. Surprisingly, the more common fusion transcripts with both DNA and RNA breakpoints in an intergenic region (type c, Figures 2E and 2F) indicate that functional polyadenylation sites exist in the intergenic region that were used by virus-host fusion transcripts (Figures 2E, S2A, and S2B). We also found splice motifs located within ±12 bp of the breakpoints in the host genome for viral-host fusion transcripts that were identical to canonical alternative splice sites of coding genes (Figure 2G).[15] Strikingly, only cases in which the RNA breakpoint was in an exonic region, and not in an intron or located upstream or downstream of the gene within 40 kb (near), demonstrated an increase in transcript levels (Figure 2H). However, even for fusion transcripts with the RNA breakpoint in exonic regions, the resultant transcripts were unlikely to produce endogenous host proteins (Figure 2F). As an example, for patient MOCC-0015, the DNA integration site was in the 13th intron of *TP63* with the corresponding RNA fusion integration site in the 14th exon based on an alternative splice site (Figure 2F, middle). Thus, the viral-host fusion transcripts undergo alternative RNA splicing with the 14th exon of gene *TP63* (Figure 2F, top). Importantly, only fusion transcripts including the 14th exon were substantially increased, whereas there was no detectable increase in the transcription of the 13th exon, which is upstream of the DNA integration site, compared with samples without HPV integration in the same genomic region (Figure 2F, bottom). Moreover, in tumors with type c integration, the productive HPV integrations induced aberrant viral-host fusion transcripts (Figures S2A and S2B). Notably, these intergenic regions were commonly transcriptionally silent in all other tumors lacking the integration breakpoint (Figures S2A and S2B, bottom).

Moreover, no significant changes in host protein (<40 kb from the integration site) levels were found in cases with productive integration compared with tumors without HPV integration in the same genomic regions by analyzing mass spectrometry proteomics data (Figures 2I and S2C). Altogether, it appears that productive integration results in an aberrant viral-host fusion transcript that does not produce an endogenous human protein. This is consistent with previous studies showing that viral integration does not usually result in production of normal host transcripts or, thus, proteins from genes located at or near the integration site.[16]

Association of HPV oncoprotein expression with different integration patterns

As indicated above, CC tumors can be classified into three types: nonintegrated (no detected integration site), productive integrated (at least one productive DNA or RNA integration

Figure 3. Multi-omics characterization demonstrates differences between HPV-integrated and -nonintegrated cervical cancers

(A) Scatterplot of normalized enrichment scores (NES) of enriched terms in both transcriptome and proteome data. The red dots represent significantly enriched (|NES| > 1.5, p < 0.05) terms of GO biological processes (GO BP) in both transcriptome and proteome. The p values were adjusted for the Benjamini-Hochberg procedure.

(B) Enrichment plots of significantly enriched terms from the Molecular Signatures Database (MSigDB) in transcriptome (top) and proteome data (bottom). The p values were adjusted for the Benjamini-Hochberg procedure.

(C) Plots of immune cells inferred by xCell. Bars indicate the standard deviation (SD), with black dots depicting the mean. The p values were derived by two-sided Wilcoxon rank-sum tests. *p < 0.05, **p < 0.01.

(D) Boxplots of indicated protein pathway scores. Boxes indicate the SD with bold horizontal lines representing the mean. The p values were derived by two-sided Wilcoxon rank-sum tests.

(E) Heatmaps for significantly differentially expressed molecular features between HPV-integrated and -nonintegrated CCs. Molecular features annotated to pathways are shown on the right.

(F) Scatterplot of differentially expressed genes between HPV-integrated and -nonintegrated CCs at mRNA (y axis) and protein levels (x axis). Red dots represent significantly differentially expressed genes from both transcriptome and proteome. The blue line demonstrates linear fitting for cross-omics correlation. The r and p values were computed by Spearman's correlation.

CellPress
OPEN ACCESS

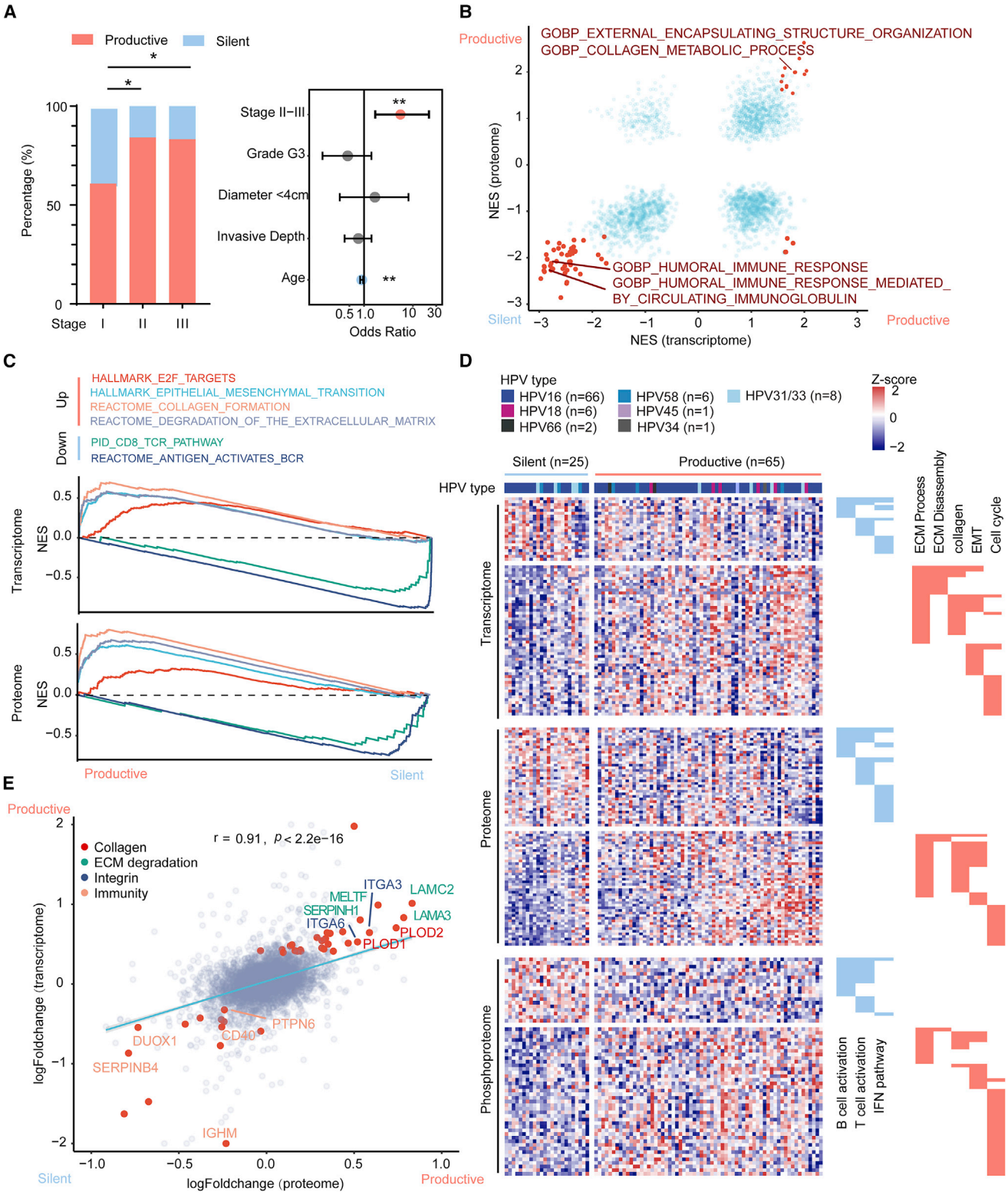


Figure 4. Multi-omics characterization demonstrates differences between productive and silent integrated cervical cancers

(A) The percentage of silent or productive HPV integration status in cervical cancers at various clinical stages (left). The forest plot depicts the associations between clinical characteristics and HPV integration status analyzed by multivariate linear regression. Red and green dots refer to clinical characteristics favoring productive or silent HPV integration, respectively (right).

(legend continued on next page)

site), and silent integrated (DNA integration with no RNA integration site) (Figure S2D). Notably, HPV A7 clades were more likely to be associated with productive HPV integration, confirming previous observations that HPV18 is strongly associated with the production of fusion transcripts (Figure S2D).[12] Interestingly, productive integrated tumors had lower levels of *E1*, *E2*, *E4*, and *E5* transcripts than nonintegrated or silent tumors. Unexpectedly, *E6* and *E7* transcripts were similar across the integration types (Figure S2E). *E6*/*E7* transcripts may be derived from nonintegrated HPV or integrated HPV fusion transcripts. Consistent with the reported higher stability of viral-cellular chimeric transcripts compared with those derived from viral DNA,[16] we found that E6 and especially E7 proteins were remarkably higher in productive integrated tumors than in silent integrated or nonintegrated tumors as detected by western blot (Figure 2J). p53 and RB proteins, the targets of E6 and E7, respectively, were lower in productive integrated CCs. Furthermore, the percentage of *E6* splicing was markedly higher in productive integrated tumors than in silent integrated or nonintegrated tumors (Figure S2F).

Multi-omics characterization demonstrates differences between HPV-integrated and -nonintegrated CCs

HPV integration into host DNA is regarded as important but not as a prerequisite for CC development.[5,17] Consistent with this concept, we found eight nonintegrated CCs in our cohort. To determine whether HPV-integrated and -nonintegrated CCs have different molecular characteristics, we performed gene set enrichment analysis (GSEA) using transcriptome and proteomics data. The results of GSEA using mRNA and proteome data were well correlated (92% of significantly enriched items matched). Functional inference using protein and transcriptome data both converged on biological networks of extracellular matrix (ECM) (cell adhesion, cell development, and morphogenesis terms) and epithelial-mesenchymal transition (EMT) in nonintegrated tumors compared with integrated tumors (Figures 3A, 3B, and S3A–S3E). Proliferation-related processes (cell cycle, DNA replication, E2F1 signaling), which can be dysregulated by HPV *E6* and *E7*, were elevated in integrated tumors compared with nonintegrated tumors (Figures 3A, 3B, and S3A–S3E). HPV-integrated tumors also had marked immune-response-related processes (T and B cell-mediated immunity and interferon-related pathways) compared with nonintegrated tumors (Figures 3A, 3B, and S3A–S3E). This was supported by xCell[18] cell-type-specific deconvolution of transcriptional data that demonstrated nonintegrated tumors with lower immuneScore, T cells, B cells, and plasmacytoid dendritic cells (pDC) cells, suggesting that nonintegrated tumors are more likely to be immune "cold" (Figure 3C). Consistent with these observations, protein pathway score analysis[19] suggested that Hippo/YAP1 and Wnt/β-catenin pathway activities, which have been associated with defects in antiviral immunity,[20] were elevated in nonintegrated CCs Figure 3D).

Notably, the differential pathway data in integrated and nonintegrated tumors was evidenced at individual gene levels, which were also highly consistent among transcriptomics, proteomics, and phosphoproteomics data (Figures 3E and 3F). And the differences between them can also be recapitulated in HPV16-positive tumors (Figures S4A–S4D). Collectively, nonintegrated CCs displayed distinct transcriptional and proteomic expression profiles and patterns, supporting the contention that the mechanisms by which tumors with integrated and nonintegrated HPV arise are distinct.

Multi-omics characterization demonstrates differences between productive and silent integrated CCs

To clarify the significance of silent and productive integrated CCs, we first correlated HPV integration status with clinical features, which showed that productive integration was associated with advanced clinical stage (Figures 4A and S5A). Considering that clinical stage is an important prognostic indicator in CC, this further supports the concept that productive integration is associated with more aggressive disease.

GSEA of transcriptome data demonstrated that EMT, proliferation (E2F1 signaling, DNA replication, and cell cycle), and ECM-associated pathways were enriched in productive integrated tumors, whereas silent integrated tumors exhibited enrichment of immune-related terms (T cell activation, lymphocyte differentiation, B cell activation) (Figures 4B–4D and S5B–S5C). Although both nonintegrated and productive integrated tumors were enriched for ECM pathway signatures (Figures 2C and 2D), the ECM genes upregulated in productive integrated tumors were enriched for ECM degradation (e.g., *SERPINH1*, *MELTF*, *LAMC2*, and *LAMA3*), collagen modification, and integrin proteins (e.g., *PLOD1*, *PLOD2*, *ITGA3*, *ITGA6*) (Figure 4E), with activation altering cell migration and promoting invasion.[21–25] In contrast, the ECM gene signatures in nonintegrated tumors were enriched for proteins mediating cell adhesion and morphogenesis (e.g., *LAMC1*, *AGRN*, *MCAM*, *PARVA*, and *NRP2*) (Figure 3F).

Protein pathway score analyses showed that cell cycle, ECM disassembly, and collagen formation tended to be elevated, while BCR and TCR signaling were lower in productive integrated tumors than in silent integrated cases (Figure S5D), similar to the transcriptome analysis. Network models constructed for productive integrated tumors using transcriptomic, proteomics, and phosphoproteomics pointed to both tumor-intrinsic (EMT, proliferation) and tumor-extrinsic (ECM, collagens) features that are associated with immune evasion and

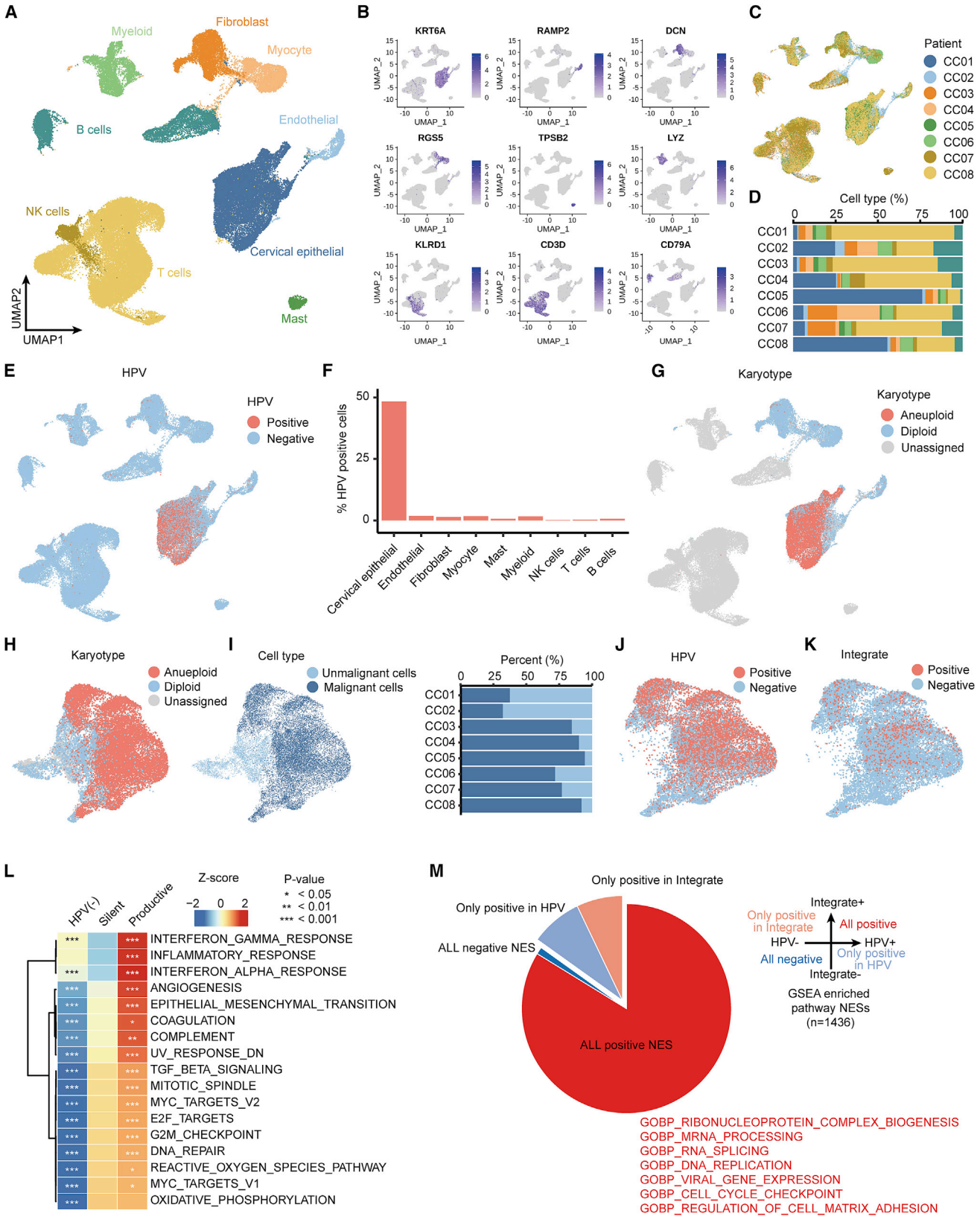
(B) Scatterplot of NES of enriched terms in both transcriptome and proteome data. The red dots represent significantly enriched ($|NES| > 1.5$, $p < 0.05$) terms of GO BP in both transcriptome and proteome data. The p values were adjusted for the Benjamini-Hochberg procedure.

(C) Enrichment plots of significantly enriched terms from MSigDB in transcriptome (top) and proteome data (bottom). The p values were adjusted for the Benjamini-Hochberg procedure.

(D) Heatmaps for significantly differentially expressed molecular features between productive and silent integrated tumors. Molecular features annotated to pathways are shown on the right. EMT, epithelial-mesenchymal transformation.

(E) Scatterplot of differentially expressed genes between productive and silent integrated tumors at mRNA (y axis) and protein abundance (x axis) levels. Red dots represent significantly differentially expressed genes in both transcriptome and proteome. The blue line demonstrates the linear fitting for cross-omics correlation. The r and p values were computed by Spearman's correlation.

CellPress
OPEN ACCESS



(legend on next page)

more aggressive tumors (Figures 4A and S5E). Importantly, analysis of the productive and silent integrated gene signatures of samples in TCGA data showed that tumors with productive integrated signature were associated with poor prognosis (Figure S5F). Moreover, the tumor-intrinsic and -extrinsic features were recapitulated in HPV16-positive tumors (Figures S6A–S6E).

scRNA-seq reveals differential effects of HPV negative and silent and productive integrated HPV

To elucidate intratumoral heterogeneity of HPV integration status and transcriptional properties in CC cells at the single-cell resolution, eight productive integrated tumors were subjected to scRNA-seq. After quality filtering (Figure S7A), up to 82,704 cells were retained (Figure 5A). Nine distinct clusters were annotated (see STAR Methods and Figures 5B and S7B), which were represented across all patients (Figures 5C and 5D). Interestingly, HPV transcripts (see STAR Methods) were readily detected in 10,033/20,672 (48.5%) epithelial cells. The extremely low HPV RNA positivity in other cell types could be due to phagocytosis of infected cells or contamination during 10× Genomics processing steps (Figures 5E, 5F, and S7C). We subsequently divided epithelial cells into five cell clusters (Figure S7D) and calculated the CNVs of each cell by CopyKAT using myocytes as normal cells.[26] The results revealed patient-unique CNVs commonly seen in CCs (Figures 5G, S7E, and S7F), which defined clusters 0, 2, 3, and 4 as malignant cells (Figures 5H and 5I). By detection of HPV transcripts and HPV integration fusion transcripts (see STAR Methods), we designated malignant cells as HPV(−) and silent and productive HPV(+) integrated cells (Figures 5J, 5K, and S7G). Notably, in five patients with paired bulk and scRNA-seq data, we found concordant HPV RNA integration events in both bulk and scRNA-seq data (Figure S5H). Interestingly, four of five patients had only one integration event detected in both scRNA-seq and bulk data, supporting clonal expansion likely due to positive selection during tumor progression. Only patient CC05 had five integration events detected in scRNA-seq and two integration events in bulk data, with the frequent events observed in bulk data readily detected in scRNA-seq data (Figure S7H).

Next, we investigated whether different HPV statuses had an impact on tumor-intrinsic features. By calculating the activity of tumor-intrinsic features using ssGSEA, we found that enrichment scores of EMT, E2F, G2M checkpoint, and DNA repair, which are HPV oncoprotein-driven pathways, showed a continuum of expression levels, from low in HPV(−) tumor cells, to intermediate in silent HPV(+) cells, to high expression in productive HPV(+) integrated cells (Figures 5L and S8A). Transcription factor (TF) analysis by SCENIC[27] showed that *TFDP2*, *E2F*, and *TWIST1*, the TFs implicated in proliferation and EMT, were activated in productive HPV(+) integrated cells, further supporting the observations from bulk RNA and protein analysis (Figure S8B). Interferon and immune response pathways were downregulated in silent HPV(+) integrated cells but increased in productive HPV(+) integrated cells (Figure 5L). To further validate transcriptional reprogramming from HPV(−) to silent and then productive HPV(+) CC cells, we computed differential gene expression and performed GSEA with GO functional terms. Transition from HPV nonintegration to silent and productive HPV integration was associated with upregulation of genes converging onto cell proliferation (mRNA processing, ribosome biogenesis, DNA repair, cell cycle)-, EMT-, viral gene expression-, and ECM-related pathways (Figures 5M and S8C).

Both class I and class II human leukocyte antigens (HLAs) were markedly decreased in HPV-positive cells, with class I HLA being lowest in productive HPV(+) integrated cells (Figure S8D). Expression of immune checkpoint ligands, including *CD276* (B7-H3) and *PVR* (ligand for *TIGIT*), was elevated in both silent and productive HPV(+) integrated cells. Moreover, *LGALS9* (ligand for *TIM-3*), *TNFRSF14* (ligand for *BTLA*), and *TNFSF9* displayed progressive upregulation from HPV(−) to silent and then productive HPV(+) CC cells (Figure S8E).

Overall, these results indicated that HPV status regulates transcriptional reprogramming of tumor cells that likely contributes to immune evasion.

HPV status correlates with cervical tumor cell-immune interactions

To examine the immune milieu and reveal cell-cell interactions, we reclustered immune cells to identify T cells, natural killer (NK) cells, B cells, and myeloid cells separately. This identified seven T cell subsets, two NK cell subsets, four B cell subsets (Figures 6A–6C), and seven myeloid subsets (Figures 6D–6F, S9A, and S9B). In the stromal compartment, we identified three fibroblast subsets, which were consistent with published gene signatures (Figures S9C–S9E); two endothelial cell subsets; two myocyte subsets; and two stromal cell subsets (Figures S9C

Figure 5. scRNA-seq reveals differential effects of HPV-negative and silent and productive integrated HPV

(A) Uniform manifold approximation and projection (UMAP) plot of cells colored by cell type.
(B) UMAP plot of cells colored by expression of selected canonical markers.
(C) UMAP plot of cells colored by patient.
(D) Proportions of each cell type in eight productive HPV-integrated cervical cancers.
(E) UMAP plot of HPV-positive cells across all cell types.
(F) Proportions of HPV-positive cells across all cell types.
(G) UMAP plot of cells colored by karyotype. The karyotypes of cells were inferred by CopyKAT.
(H) UMAP plot of cervical epithelial cells colored by karyotype.
(I) UMAP plot (left) and proportions (right) of malignant and nonmalignant cells by patient in cervical epithelial cells.
(J) UMAP plot of malignant cells colored by HPV expression.
(K) UMAP plot of malignant cells colored by HPV integration.
(L) Heatmap of single-cell GSEA enrichment scores of selected hallmark pathways in malignant cells across various HPV integration statuses. The p values were calculated by two-sided Wilcoxon rank-sum test.
(M) Pie chart distribution of significantly enriched pathways computed between HPV(+) vs. HPV(−) and integrated (+) vs. nonintegrated (−).

CellPress
OPEN ACCESS

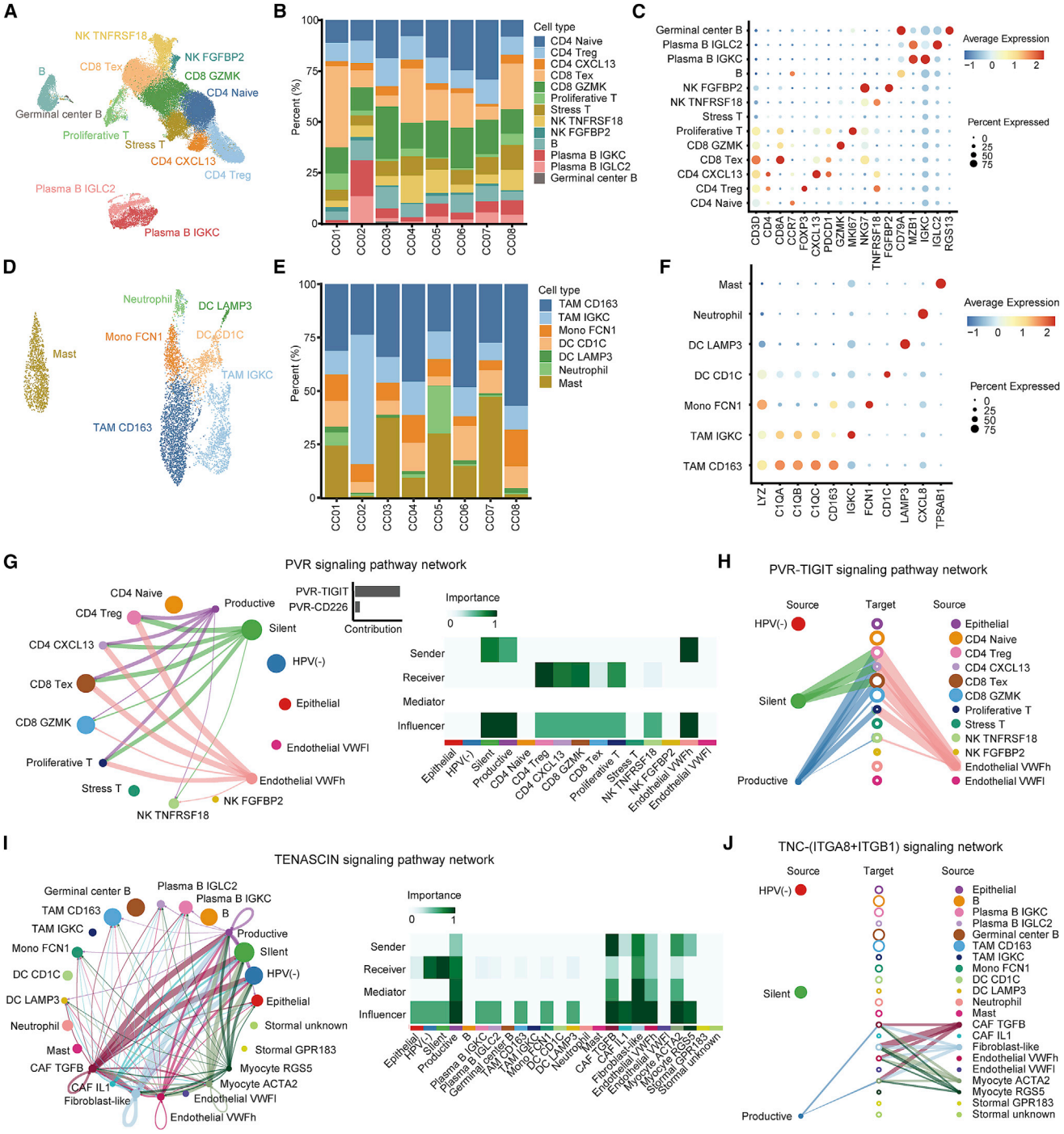


Figure 6. HPV status correlates with cervical tumor cell-immune interactions

(A) UMAP plot of lymphoid cells colored by cell type.

(B) Proportions of cell types of lymphoid cells in each tumor.

(C) Expression of selected marker genes in cell types of lymphoid cells.

(D) UMAP plot of myeloid cells colored by cell type.

(E) Proportions of cell types of myeloid cells in each tumor.

(F) Expression of selected marker genes in cell types of myeloid cells.

(legend continued on next page)

and S9F). All nonepithelial cell types could be identified in each of the CC tumors at varying frequencies (Figures 6B and 6E).

We then conducted cellular communication analysis using CellChat[28] to explore potential interactions between malignant and immune cell populations. Compared with HPV(−) malignant cells, PVR communication signaling was predicted to be increased between HPV(+) malignant cells and multiple exhausted and dysfunctional T cell subsets (CD4-CXCL13, CD4-FOXP3, CD8-PDCD1, CD8-MKI67) (Figures 6G and 6H). *PVR* was overexpressed in HPV(+) malignant cells (Figure S8E), and its receptor *TIGIT* was mainly expressed in CD4-CXCL13, CD4-FOXP3, CD8-PDCD1, and CD8-MKI67 T cells (Figure S9G). The PVR-TIGIT interaction has been proposed to represent immune inhibitory checkpoint signaling limiting adaptive and innate immunity. Thus PVR-TIGIT signaling between HPV(+) malignant cells and T cells could contribute to the immune evasion of HPV(+) malignant cells. Furthermore, HPV(+) malignant cells showed exclusive interactions with various stromal cells, including communication with CAF-TGFB, fibroblast-like, myocyte-ACTA2, and myocyte-RGS5 clusters via the TENASCIN pathway based on the tenascin C (TNC)-ITGA8/ITGB1and TNC-ITGA9/ITGB1 axes (Figures 6I, 6J, and S9H–S9J) and communication with all nonepithelial cells via the collagen signaling network (Figure S9K). TNC plays essential roles in tumor neovascularization, EMT, and metastasis.[29–31] Moreover, cancer-cell-derived TNC can orchestrate an immune-suppressive microenvironment by interacting with stroma cells or directly inhibiting T cell activation, proliferation, and cytokine production.[29,31] In addition, collagen activation accelerates cell proliferation and promotes invasion and metastasis and inhibitor immune response.[32] Thus, enhanced TNC and collagen interactions could play important roles in CC progression and immune escape.

Genomic HPV integration features in precancerous cervical lesions

A clearer understanding of the molecular changes from precancerous lesion to advanced CC could provide more accurate biomarkers to improve prevention and therapy strategies. Only a portion of early precancerous lesions (cervical intraepithelial neoplasia) with HPV integration will progress to CCs. However, which precancerous lesions progress and which do not remains unclear. As revealed above, productive HPV integration, which is tightly associated with advanced clinical CC stage, was nonrandom and harbored a predilection for specific types of genomic positions (i.e., introns, CFSs, Alu regions, and exonic regions) (Figures 1E–1G). We thus developed a least absolute shrinkage and selection operator (Lasso) regression model to predict productive HPV integration sites based on genomic features of HPV integrations revealed by VCS in advanced tumors (Table S6). Impressively, the predictive performance of this model demonstrated considerable accuracy (>85%) when calculating the area under the curve (AUC) for the ROC curve with 5-fold cross-validation (Figure 7A).

We then analyzed HPV integration status with VCS data using Pap smears of cervicovaginal samples from 213 additional cervical preneoplastic lesions. Only 26 of 213 (12.2%) preneoplastic lesions were HPV integration positive, which is much lower than in CC. Interestingly, among the 26 HPV integration-positive cases, 12 (46.2%) cases were predicted to have productive HPV integration, which is remarkably lower than CC of different clinical stages (Figure 7B). Given the aggressive features associated with productive HPV integration, these preneoplastic lesions with productive HPV integration may have a higher risk for progression to invasive carcinoma.

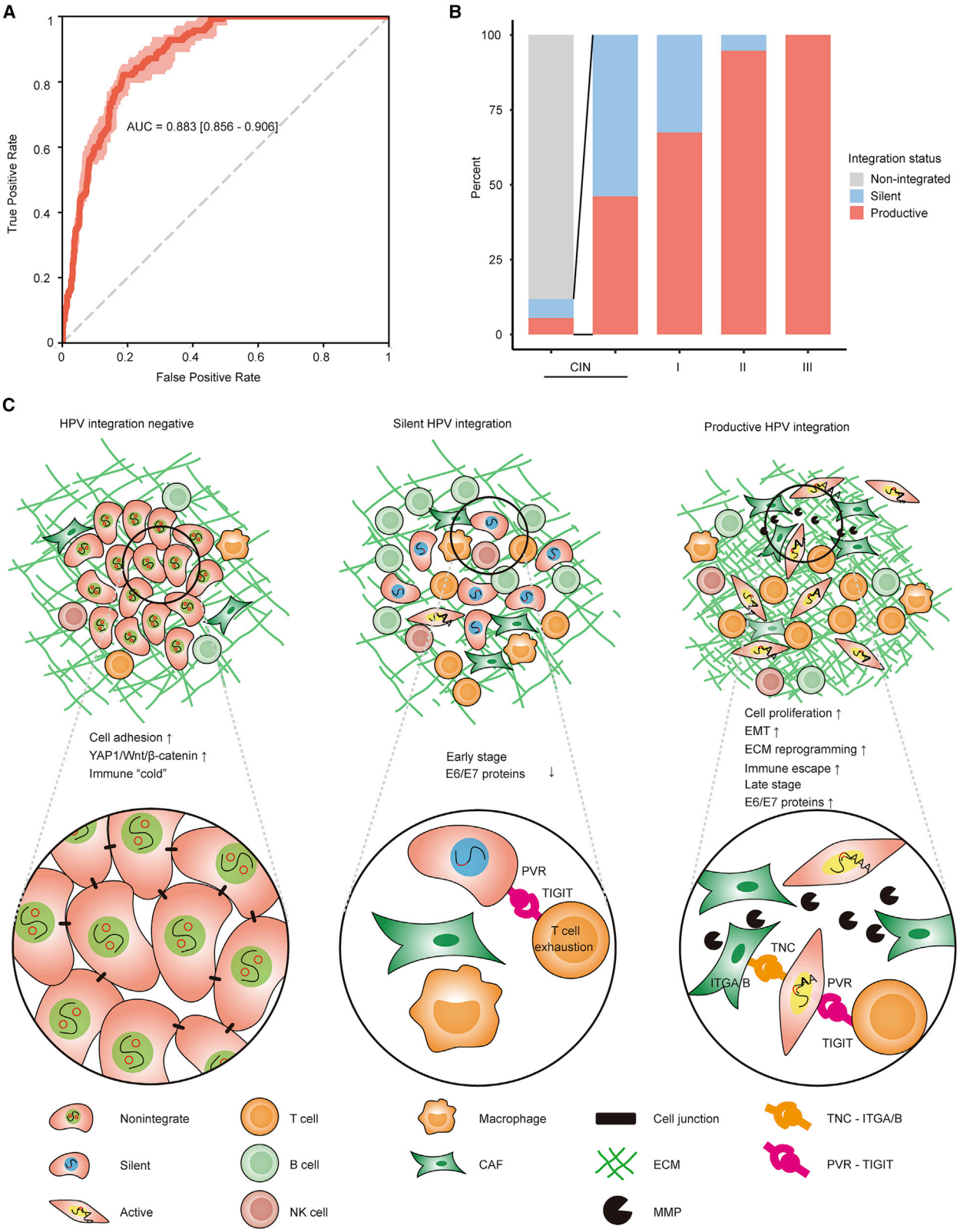
DISCUSSION

Here, by comparing DNA and RNA integration status in a cohort of CCs, we, for the first time to our knowledge, demonstrated that only productive HPV integration sites were nonrandomly distributed across both viral and host genomes, consistent with productive HPV integration sites being under selective pressure and contributing to CC pathophysiology. Furthermore, multi-omics data revealed that the tumors with productive HPV integration are associated with higher E6/E7 protein levels and enhanced tumor aggressiveness and immunoevasion at the tissue and single-cell levels (Figure 7C). Our study improves the understanding of the functional consequences of HPV fusion transcripts on the biology and pathophysiology of HPV-driven CCs and supports the potential for productive HPV integration to be used as an indication of a high risk for CC progression.

HPV integration is a common event during CC carcinogenesis. Our study supports that there are three HPV integration patterns in CCs associated with different degrees of aggressiveness and clinical stage and, thus, prognosis. Initially, the viral genome exists in an episomal form. Antivirus immune responses have been proposed to enhance HPV clearance, with potential selection for cells with HPV integration.[33] Initial HPV integration is prone to be randomly distributed across chromosomes, most of which are likely dead-end silent integrations that fail to produce sufficient E6/E7 oncoproteins to drive clonal expansion. Occasionally, integration occurs near a cellular splicing acceptor and PAS, generating a viral-host fusion transcript with intact *E6*/*E7* and a host PAS that we have classified as productive. Productive integration enhances viral oncogene expression not only by producing more stable viral-host chimeras but also by disrupting the *E2* gene and alleviating *E2*-mediated transcriptional repression of *E6* and *E7*. Cells with productive HPV integration appear to acquire selective advantages and drive tumor progression as indicated by selection of integration sites.

(G) Overview of intercellular communication network for PVR signaling pathway. Intercellular interactions (left), relative contribution of each ligand-receptor pair (middle), and relative importance of each cell group based on four network centrality measures (right) are shown.
(H) Hierarchical plot showing the communication network of the PVR-TIGIT signaling pathway.
(I) Overview of intercellular communication network for TENASCIN signaling pathway. Intercellular interactions (left) and relative importance of each cell group based on four network centrality measures (right) are shown.
(J) Hierarchical plot showing the communication network of the TNC-(ITGA8 + ITGB1) signaling pathway.

CellPress
OPEN ACCESS



(legend on next page)

Consistent with previous studies demonstrating that integration is not an essential event in cervical carcinogenesis,[17] we found eight tumors that were nonintegrated. High *E6*/*E7* expression is required for oncogenic transformation and phenotype maintenance in CC, and in theory, episomal expression of the viral oncogenes *E6* and *E7* is tightly regulated.[6] However, *E6* and *E7* mRNA expression levels in nonintegrated samples harboring purely episomal HPV were comparable to cases with integrated forms (Figures S2D and S2E). Hong et al. reported that the overexpression of *E6* and *E7* in tumors with episomal HPV might be mediated through a higher viral load.[17] Host antivirus immune responses are proposed to determine episomal HPV clearance and selection of cells with integrated HPV.[34,35] Given the immune "cold" microenvironment we revealed in integration-negative tumors (Figures 3A–3C), it is reasonable to hypothesize that an impaired immune response contributes to persistent HPV infection (failure of virus clearance) and the ability of cells with this phenotype to persist and become CCs. Similar to the effects of HPV integration on *E6* and *E7* production, persistent high-level expression of viral oncogenes in cells where episomal HPV is not cleared leads to cell transformation. In addition, we found that YAP1/Wnt/β-catenin were activated in nonintegrated tumors (Figure 3D). Hyperactivation of YAP1/Wnt/β-catenin accelerates HPV oncoprotein-induced tumorigenesis of the cervix even in the background of episomal HPV.[36–38] Consequently, while high HPV oncogene expression may be the initial hit in multistep tumorigenesis, hyperactivation of the Wnt/β-catenin pathway may facilitate development of CC in the absence of HPV integration.

HPV integration leading to the production of fusion transcripts may promote tumor progression in multiple ways. First, HPV integration often disrupts the viral *E2* gene that encodes a transcriptional inhibitor of *E6* and *E7* expression.[39] Second, productive integration results in viral-cellular chimeric transcripts, which are more stable than those derived from episomal viral DNA.[16] The derepression of transcription and the stabilization of fusion transcripts containing *E6* and *E7* synergistically enhance HPV oncogene expression (Figure 2J). Third, higher *E6* splicing in productive integrated tumors (Figure S2F) not only produces an E6* truncated protein that participates in multiple stages of HPV transformation and cervical carcinogenesis,[40] but also facilitates E7 oncoprotein translation by increasing the space between the *E6* stop codon and the *E7* start codon.[41] Given that CC is dependent on the expression of HPV oncogenes for continued proliferation and survival,[39] the dysregulation of *E6* and *E7* expression in productive integrated tumors likely provides a selective growth advantage and promotes oncogenic progression. Fourth, although HPV integration has been hypothesized to activate adjacent cellular oncogenes to promote cervical carcinogenesis,[39] our data suggest that productive fusion transcripts do not usually produce an endogenous human protein, nor do they alter the expression of nearby genes.

Evasion of immune surveillance is a necessary step in tumor evolution. However, our knowledge about immune escape in CC is rather limited, especially whether tumors with different HPV statuses evade immune surveillance via different mechanisms. Here, we highlighted the "PVR-TIGIT axis" as a potential player in immune evasion by HPV-positive cells during CC development (Figure 5). *PVR* expression in tumor cells has been associated with an unfavorable prognosis and anti-PD-1 immunotherapy resistance in multiple malignancies.[42–44] We showed that *PVR* was selectively upregulated in HPV-positive CC cells (Figure S8E). Moreover, *PVR* preferentially binds *TIGIT*, an inhibitory receptor expressed in lymphocytes,[45] and therefore induces an immunosuppressive profile in TIGIT-expressing cells. PVR-TIGIT interactions can skew the immune response toward an immunosuppressed phenotype and thus abrogate T cell cytotoxicity.[42,43,46] Interestingly, unlike *PDCD1*, which is highly expressed only in exhausted CD4 and CD8 T cells, *TIGIT* is also expressed on T regulatory (Treg) cells (Figure S9G), where it enhances their suppressive function and correlates with a poor clinical outcome following PD-1 and/or CTLA-4 pathway blockade.[47,48] Since targeting the PVR-TIGIT axis is being explored clinically, our findings provide potential therapeutic options for HPV-induced CCs. Furthermore, productive HPV integration may allow tumor cells to evade immune surveillance by a number of additional mechanisms. Multiple immune checkpoint ligands and HLAs were specifically up- and downregulated, respectively, in cells with productive HPV integration. In addition, the milieu of the advanced CC tumor microenvironment is rich in ECM, and with ECM remodeling likely contributing to tumor progression and immune escape. We found that cells with productive HPV integration demonstrate the potential for enhanced cell-ECM communication (Figure 6), which could also facilitate evasion of immune surveillance. Several recent examples of ECM-targeted immunotherapies, including targeting TNC with tenatumomab, have shown promising results in preclinical models,[49] which could lead to the development of more effective ECM-targeted immune therapies in CCs.

Limitations of the study

There are several limitations in our study. First, this study focused on a transcription-dependent process. It did not explore transcription-independent effects of HPV integration, such as altering the 3D genome architecture or transcriptional regulatory element activity around integration sites.[50,51] Second, here we focused on the impacts of HPV integration on the expression of coding genes by applied poly(A) mRNA-seq. So, whether HPV integration leads to aberrant expression of noncoding human genome elements or eRNAs needs further investigation. Third, although 379 (97.7%) of RNA integration sites were associated with DNA integration sites, the small number of discrepancies may be due to the limitations of virus capture DNA sequencing. Additional methods, such as whole-genome sequencing and long-read sequencing, may provide additional data on HPV integration sites and status. Finally,

Figure 7. Genomic HPV integration features in precancerous cervical lesions

(A) Receiver-operating characteristic (ROC) curve and corresponding area under the curve (AUC) statistics for predicting productive HPV integration sites using Lasso regression model. The confidence intervals are in red.

(B) Comparison of the percentages of various predicted HPV integration status in preneoplastic lesions (CIN) and different stages of CC.

(C) Schema of molecular landscapes of CCs with various HPV integration statuses.

considering that only eight CC cases were HPV-nonintegrated in our study, the finding that HPV-integrated CCs had elevated cell-cycle and immune-response-related processes may need further study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient cohort
- METHOD DETAILS
  - Sample quality
  - Sample processing for DNA and RNA extraction
  - Whole-exome sequencing
  - Bulk RNA sequencing
  - Single-cell RNA sequencing
  - Virus-capture sequencing
  - Mass spectrometry analysis of protein and phosphoprotein
  - Western blots
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Copy number analysis
  - RNA quantification
  - Protein and phosphoprotein quantification
  - HPV analysis
  - Single-cell RNA-seq data processing
  - Identification of subpopulations
  - Differential expressed gene and pathway enrichment
  - Pathway score calculation
  - Transcription factor analysis
  - Cell-cell communication
  - Generalized linear models
  - Survival analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2022.100211.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (81630060, 81902653, 82141106, 81874106, 82073259, 81974408), the Health and Family Planning Commission of Wuhan Municipality (WX18Q16), the Key R&D Program of Hubei Province (2020BCA067), the Hubei Province Science Fund for Distinguished Young Scholars (2020CFA066), and the Beijing Kanghua Foundation for the Development of Traditional Chinese and Western Medicine (KH-2021-LQJJ-006).

AUTHOR CONTRIBUTIONS

C.S., D.M., and G.C. conceived and supervised the project. Y. Shen., E.G., and F.L. collected the samples and performed IHC staining. J.F., Y.F., W.P., and Xiong Li performed bioinformatics analyses. J.F. and B.Y. performed experiments. C.S., J.F., and Y.F. wrote the manuscript. S.L., X.Q., D.H., R.X., Xi Li, S.Z., S.Y., C.Y., Y. Shu., H.H., T.W., Y.P., S.W., W.C., and H.W. collected the patients' clinical information and contributed to data processing and analyses. L.Z., L.Y., B.W., and G.B.M. provided expertise and feedback. D.Z., G.B.M., B.X., K.S., G.C., D.M., and C.S. contributed to project management and provided valuable critical discussion.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 23, 2022
Revised: June 21, 2022
Accepted: October 12, 2022
Published: January 11, 2023

REFERENCES

1. Guo, M., Xu, J., and Du, J. (2021). Trends in cervical cancer mortality in China from 1989 to 2018: an age-period-cohort study and joinpoint analysis. BMC Publ. Health *21*, 1329. https://doi.org/10.1186/s12889-021-11401-8.
2. Brisson, M., Kim, J.J., Canfell, K., Drolet, M., Gingras, G., Burger, E.A., Martin, D., Simms, K.T., Bénard, É., Boily, M.-C., et al. (2020). Impact of HPV vaccination and cervical screening on cervical cancer elimination: a comparative modelling analysis in 78 low-income and lower-middle-income countries. Lancet *395*, 575–590. https://doi.org/10.1016/S0140-6736(20)30068-4.
3. Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosé, S., Saraiya, M., Ferlay, J., and Bray, F. (2020). Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob. Health *8*, e191–e203. https://doi.org/10.1016/S2214-109X(19)30482-6.
4. Hu, Z., Zhu, D., Wang, W., Li, W., Jia, W., Zeng, X., Ding, W., Yu, L., Wang, X., Wang, L., et al. (2015). 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. https://doi.org/10.1038/ng.3178.
5. Vinokurova, S., Wentzensen, N., Kraus, I., Klaes, R., Driesch, C., Melsheimer, P., Kisseljov, F., Dürst, M., Schneider, A., and von Knebel Doeberitz, M. (2008). Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. Cancer Res. *68*, 307–313. https://doi.org/10.1158/0008-5472.CAN-07-2754.
6. Wentzensen, N., Vinokurova, S., and von Knebel Doeberitz, M. (2004). Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res. *64*, 3878–3884. https://doi.org/10.1158/0008-5472.CAN-04-0009.
7. Christiansen, I.K., Sandve, G.K., Schmitz, M., Dürst, M., and Hovig, E. (2015). Transcriptionally active regions are the preferred targets for chromosomal HPV integration in cervical carcinogenesis. PLoS One *10*, e0119566. https://doi.org/10.1371/journal.pone.0119566.
8. Rajaby, R., Zhou, Y., Meng, Y., Zeng, X., Li, G., Wu, P., and Sung, W.-K. (2021). SurVirus: a repeat-aware virus integration caller. Nucleic Acids Res. *49*, e33. https://doi.org/10.1093/nar/gkaa1237.
9. Bodelon, C., Untereiner, M.E., Machiela, M.J., Vinokurova, S., and Wentzensen, N. (2016). Genomic characterization of viral integration sites in HPV-related cancers. Int. J. Cancer *139*, 2001–2011. https://doi.org/10.1002/ijc.30243.
10. Thorland, E.C., Myers, S.L., Persing, D.H., Sarkar, G., McGovern, R.M., Gostout, B.S., and Smith, D.I. (2000). Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. Cancer Res. *60*, 5916–5921.
11. Gagliardi, A., Porter, V.L., Zong, Z., Bowlby, R., Titmuss, E., Namirembe, C., Griner, N.B., Petrello, H., Bowen, J., Chan, S.K., et al. (2020). Analysis of Ugandan cervical carcinomas identifies human papillomavirus clade–

specific epigenome and transcriptome landscapes. Nat. Genet. *52*, 800–810. https://doi.org/10.1038/s41588-020-0673-7.

12. Cancer Genome Atlas Research Network; Albert Einstein College of Medicine; Analytical Biological Services; Barretos Cancer Hospital; Baylor College of Medicine; Beckman Research Institute of City of Hope; Buck Institute for Research on Aging; Canada's Michael Smith Genome Sciences Centre; Harvard Medical School; Helen F. Graham Cancer Center &Research Institute at Christiana Care Health Services (2017). integrated genomic and molecular characterization of cervical cancer. Nature *543*, 378–384. https://doi.org/10.1038/nature21386.
13. Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev. *11*, 2755–2766. https://doi.org/10.1101/gad.11.21.2755.
14. Karimzadeh, M., Arlidge, C., Rostami, A., Lupien, M., Bratman, S.V., and Hoffman, M.M. (2020). Human papillomavirus integration transforms chromatin to drive oncogenesis. Preprint at biorxiv. https://doi.org/10.1101/2020.02.12.942755.
15. Sibley, C.R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. Nat. Rev. Genet. *17*, 407–421. https://doi.org/10.1038/nrg.2016.46.
16. Ehrig, F., Häfner, N., Driesch, C., Kraus Christiansen, I., Beer, K., Schmitz, M., Runnebaum, I.B., and Dürst, M. (2019). Differences in stability of viral and viral-cellular fusion transcripts in HPV-induced cervical cancers. Int. J. Mol. Sci. *21*, E112. https://doi.org/10.3390/ijms21010112.
17. Hong, D., Liu, J., Hu, Y., Lu, X., Li, B., Li, Y., Hu, D., Lu, W., Xie, X., and Cheng, X. (2017). Viral E6 is overexpressed via high viral load in invasive cervical cancer with episomal HPV16. BMC Cancer *17*, 136. https://doi.org/10.1186/s12885-017-3124-9.
18. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220. https://doi.org/10.1186/s13059-017-1349-1.
19. Krug, K., Jaehnig, E.J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L.C., et al. (2020). Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. Cell *183*, 1436–1456.e31. https://doi.org/10.1016/j.cell.2020.10.036.
20. Ljungberg, J.K., Kling, J.C., Tran, T.T., and Blumenthal, A. (2019). Functions of the WNT signaling network in shaping host responses to infection. Front. Immunol. *10*, 2521. https://doi.org/10.3389/fimmu.2019.02521.
21. Brooks, D.L.P., Schwab, L.P., Krutilina, R., Parke, D.N., Sethuraman, A., Hoogewijs, D., Schörg, A., Gotwald, L., Fan, M., Wenger, R.H., et al. (2016). ITGA6 is directly regulated by hypoxia-inducible factors and enriches for cancer stem cell activity and invasion in metastatic breast cancer models. Mol. Cancer *15*, 26. https://doi.org/10.1186/s12943-016-0510-x.
22. Liu, M., Zhang, Y., Yang, J., Zhan, H., Zhou, Z., Jiang, Y., Shi, X., Fan, X., Zhang, J., Luo, W., et al. (2021). Zinc-dependent regulation of ZEB1 and YAP1 coactivation promotes epithelial-mesenchymal transition plasticity and metastasis in pancreatic cancer. Gastroenterology *160*, 1771–1783.e1. https://doi.org/10.1053/j.gastro.2020.12.077.
23. Moon, Y.W., Rao, G., Kim, J.J., Shim, H.-S., Park, K.-S., An, S.S., Kim, B., Steeg, P.S., Sarfaraz, S., Changwoo Lee, L., et al. (2015). LAMC2 enhances the metastatic potential of lung adenocarcinoma. Cell Death Differ. *22*, 1341–1352. https://doi.org/10.1038/cdd.2014.228.
24. Sathyanarayana, U.G., Padar, A., Huang, C.X., Suzuki, M., Shigematsu, H., Bekele, B.N., and Gazdar, A.F. (2003). Aberrant promoter methylation and silencing of laminin-5-encoding genes in breast carcinoma. Clin. Cancer Res. *9*, 6389–6394.
25. Xiong, G., Chen, J., Zhang, G., Wang, S., Kawasaki, K., Zhu, J., Zhang, Y., Nagata, K., Li, Z., Zhou, B.P., et al. (2020). Hsp47 promotes cancer metastasis by enhancing collagen-dependent cancer cell-platelet interaction. Proc. Natl. Acad. Sci. *117*, 3748–3758. https://doi.org/10.1073/pnas.1911951117.
26. Gao, R., Bai, S., Henderson, Y.C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A., et al. (2021). Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat. Biotechnol. *39*, 599–608. https://doi.org/10.1038/s41587-020-00795-2.
27. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
28. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088. https://doi.org/10.1038/s41467-021-21246-9.
29. Jachetti, E., Caputo, S., Mazzoleni, S., Brambillasca, C.S., Parigi, S.M., Grioni, M., Piras, I.S., Restuccia, U., Calcinotto, A., Freschi, M., et al. (2015). Tenascin-C protects cancer stem–like cells from immune surveillance by arresting T-cell activation. Cancer Res. *75*, 2095–2108. https://doi.org/10.1158/0008-5472.CAN-14-2346.
30. Miroshnikova, Y.A., Mouw, J.K., Barnes, J.M., Pickup, M.W., Lakins, J.N., Kim, Y., Lobo, K., Persson, A.I., Reis, G.F., McKnight, T.R., et al. (2016). Tissue mechanics promote IDH1-dependent HIF1α–tenascin C feedback to regulate glioblastoma aggression. Nat. Cell Biol. *18*, 1336–1345. https://doi.org/10.1038/ncb3429.
31. Oskarsson, T., Acharyya, S., Zhang, X.H.-F., Vanharanta, S., Tavazoie, S.F., Morris, P.G., Downey, R.J., Manova-Todorova, K., Brogi, E., and Massagué, J. (2011). Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. Nat. Med. *17*, 867–874. https://doi.org/10.1038/nm.2379.
32. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K.J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat. Commun. *11*, 5120. https://doi.org/10.1038/s41467-020-18794-x.
33. Pett, M.R., Herdman, M.T., Palmer, R.D., Yeo, G.S.H., Shivji, M.K., Stanley, M.A., and Coleman, N. (2006). Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. Proc. Natl. Acad. Sci. *103*, 3822–3827. https://doi.org/10.1073/pnas.0600078103.
34. Boccardo, E., Lepique, A.P., and Villa, L.L. (2010). The role of inflammation in HPV carcinogenesis. Carcinogenesis *31*, 1905–1912. https://doi.org/10.1093/carcin/bgq176.
35. Hopman, A.H.N., Smedts, F., Dignef, W., Ummelen, M., Sonke, G., Mravunac, M., Vooijs, G.P., Speel, E.-J.M., and Ramaekers, F.C.S. (2004). Transition of high-grade cervical intraepithelial neoplasia to micro-invasive carcinoma is characterized by integration of HPV 16/18 and numerical chromosome abnormalities. J. Pathol. *202*, 23–33. https://doi.org/10.1002/path.1490.
36. Bulut, G., Fallen, S., Beauchamp, E.M., Drebing, L.E., Sun, J., Berry, D.L., Kallakury, B., Crum, C.P., Toretsky, J.A., Schlegel, R., et al. (2011). Beta-catenin accelerates human papilloma virus type-16 mediated cervical carcinogenesis in transgenic mice. PLoS One *6*, e27243. https://doi.org/10.1371/journal.pone.0027243.
37. He, C., Lv, X., Huang, C., Angeletti, P.C., Hua, G., Dong, J., Zhou, J., Wang, Z., Ma, B., Chen, X., et al. (2019). A human papillomavirus-independent cervical cancer animal model reveals unconventional mechanisms of cervical carcinogenesis. Cell Rep. *26*, 2636–2650.e5. https://doi.org/10.1016/j.celrep.2019.02.004.
38. Uren, A., Fallen, S., Yuan, H., Usubütün, A., Küçükali, T., Schlegel, R., and Toretsky, J.A. (2005). Activation of the canonical Wnt pathway during genital keratinocyte transformation: a model for cervical cancer progression. Cancer Res. *65*, 6199–6206. https://doi.org/10.1158/0008-5472.CAN-05-0455.
39. McBride, A.A., and Warburton, A. (2017). The role of integration in oncogenic progression of HPV-associated cancers. PLoS Pathog. *13*, e1006211. https://doi.org/10.1371/journal.ppat.1006211.
40. Chen, J., Xue, Y., Poidinger, M., Lim, T., Chew, S.H., Pang, C.L., Abastado, J.-P., and Thierry, F. (2014). Mapping of HPV transcripts in four human cervical lesions using RNAseq suggests quantitative rearrangements

during carcinogenic progression. Virology *462–463*, 14–24. https://doi.org/10.1016/j.virol.2014.05.026.

41. Olmedo-Nieva, L., Muñoz-Bello, J., Contreras-Paredes, A., and Lizano, M. (2018). The role of E6 spliced isoforms (E6*) in human papillomavirus-induced carcinogenesis. Viruses *10*, 45. https://doi.org/10.3390/v10010045.
42. Carlsten, M., Björkström, N.K., Norell, H., Bryceson, Y., van Hall, T., Baumann, B.C., Hanson, M., Schedvins, K., Kiessling, R., Ljunggren, H.-G., et al. (2007). DNAX accessory molecule-1 mediated recognition of freshly isolated ovarian carcinoma by resting natural killer cells. Cancer Res. *67*, 1317–1325. https://doi.org/10.1158/0008-5472.CAN-06-2264.
43. Gorvel, L., and Olive, D. (2020). Targeting the "PVR-TIGIT axis" with immune checkpoint therapies. F1000Research *9*, F1000, Faculty Rev-354. https://doi.org/10.12688/f1000research.22877.1.
44. Yang, Z.-Z., Kim, H.J., Wu, H., Jalali, S., Tang, X., Krull, J.E., Ding, W., Novak, A.J., and Ansell, S.M. (2020). TIGIT expression is associated with T-cell suppression and exhaustion and predicts clinical outcome and anti-PD-1 response in follicular lymphoma. Clin. Cancer Res. *26*, 5217–5231. https://doi.org/10.1158/1078-0432.CCR-20-0558.
45. Manieri, N.A., Chiang, E.Y., and Grogan, J.L. (2017). TIGIT: a key inhibitor of the cancer immunity cycle. Trends Immunol. *38*, 20–28. https://doi.org/10.1016/j.it.2016.10.002.
46. Johnston, R.J., Comps-Agrar, L., Hackney, J., Yu, X., Huseni, M., Yang, Y., Park, S., Javinal, V., Chiu, H., Irving, B., et al. (2014). The immunoreceptor TIGIT regulates antitumor and antiviral CD8(+) T cell effector function. Cancer Cell *26*, 923–937. https://doi.org/10.1016/j.ccell.2014.10.018.
47. Joller, N., Lozano, E., Burkett, P.R., Patel, B., Xiao, S., Zhu, C., Xia, J., Tan, T.G., Sefik, E., Yajnik, V., et al. (2014). Treg cells expressing the coinhibitory molecule TIGIT selectively inhibit proinflammatory Th1 and Th17 cell responses. Immunity *40*, 569–581. https://doi.org/10.1016/j.immuni.2014.02.012.
48. Kurtulus, S., Sakuishi, K., Ngiow, S.-F., Joller, N., Tan, D.J., Teng, M.W.L., Smyth, M.J., Kuchroo, V.K., and Anderson, A.C. (2015). TIGIT predominantly regulates the immune response via regulatory T cells. J. Clin. Invest. *125*, 4053–4062. https://doi.org/10.1172/JCI81187.
49. Cox, T.R. (2021). The matrix in cancer. Nat. Rev. Cancer *21*, 217–238. https://doi.org/10.1038/s41568-020-00329-7.
50. Cao, C., Hong, P., Huang, X., Lin, D., Cao, G., Wang, L., Feng, B., Wu, P., Shen, H., Xu, Q., et al. (2020). HPV-CCDC106 integration alters local chromosome architecture and hijacks an enhancer by three-dimensional genome structure remodeling in cervical cancer. J. Genet. Genomics *47*, 437–450. https://doi.org/10.1016/j.jgg.2020.05.006.
51. Warburton, A., Redmond, C.J., Dooley, K.E., Fu, H., Gillison, M.L., Akagi, K., Symer, D.E., Aladjem, M.I., and McBride, A.A. (2018). HPV integration hijacks and multimerizes a cellular enhancer to generate a viral-cellular super-enhancer that drives high viral oncogene expression. PLoS Genet. *14*, e1007179. https://doi.org/10.1371/journal.pgen.1007179.
52. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
53. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
54. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.
55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
56. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108. https://doi.org/10.1093/nar/gkt214.
57. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.
58. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319. https://doi.org/10.1038/nprot.2016.136.
59. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499. https://doi.org/10.1101/gr.209601.116.
60. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.
61. Yang, B., Fan, J., Huang, J., Guo, E., Fu, Y., Liu, S., Xiao, R., Liu, C., Lu, F., Qin, T., et al. (2021). Clinical and molecular characteristics of COVID-19 patients with persistent SARS-CoV-2 infection. Nat. Commun. *12*, 3501. https://doi.org/10.1038/s41467-021-23621-y.
62. Hornburg, M., Desbois, M., Lu, S., Guan, Y., Lo, A.A., Kaufman, S., Elrod, A., Lotstein, A., DesRochers, T.M., Munoz-Rodriguez, J.L., et al. (2021). Single-cell dissection of cellular components and interactions shaping the tumor immune phenotypes in ovarian cancer. Cancer Cell *39*, 928–944.e6. https://doi.org/10.1016/j.ccell.2021.04.004.
63. Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., et al. (2019). Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell *179*, 829–845.e20. https://doi.org/10.1016/j.cell.2019.10.003.
64. Wang, R., Dang, M., Harada, K., Han, G., Wang, F., Pool Pizzi, M., Zhao, M., Tatlonghari, G., Zhang, S., Hao, D., et al. (2021). Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. Nat. Med. *27*, 141–151. https://doi.org/10.1038/s41591-020-1125-8.
65. Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H.D., Krishnan, I., Maroni, G., Meyerovitz, C.V., Kerwin, C.M., et al. (2019). Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. Immunity *50*, 1317–1334.e10. https://doi.org/10.1016/j.immuni.2019.03.009.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
RB Rabbit Polyclonal antibody	Abclonal	Cat# A11409; RRID:AB_2758545
P53 Rabbit Polyclonal antibody	Abclonal	Cat#A16989; RRID:AB_2772689
HPV16-E6 Rabbit Polyclonal antibody	GeneTex	Cat# GTX132686; RRID:AB_2886710
HPV16-E7 Rabbit Polyclonal antibody	GeneTex	Cat# GTX133411; RRID:AB_2886971
GAPDH Rabbit Monoclonal antibody	Abclonal	Cat#A19056; RRID:AB_2862549
HRP Goat Anti-Rabbit IgG (H + L)	Abclonal	Cat#AS014; RRID:AB_2769854
Biological samples		
Primary tumor and normal tissue samples	This paper	See STAR Methods: Experimental model and subject details
Critical commercial assays		
Chromium Next GEM Single Cell 3′ Reagent Kits v3.1	10x Genomics	Cat#CG000204
BCA Protein Assay Kit	ThermoFisher	Cat#23225
SureSelect Human All Exon V6 kit	Agilent	Cat#51908864
VAHTS mRNA-seq v2 Library Prep Kit	Illumina	Cat#NR602-02
Deposited data		
Raw sequencing files	This paper	NGDC: PRJCA009817
Processed and deidentified figure data and code	This paper	Zendo: https://doi.org/10.5281/zenodo.7033608
Software and algorithms		
Bwa	Li and Durbin, 2009[52]	http://bio-bwa.sourceforge.net/
GATK 4.1.9.0	McKenna et al. 2010[53]	https://github.com/broadinstitute/gatk
CNVkit 0.9.7	Talevich et al. 2016[54]	http://cnvkit.readthedocs.io
STAR 2.7.6a	Dobin et al. 2013[55]	https://github.com/alexdobin/STAR
Subread 2.0.1	Liao et al. 2014[56]	http://subread.sourceforge.net/
RSEM 1.3.1	Li and Dewey, 2011[57]	https://github.com/deweylab/RSEM
MaxQuant 1.6.1.0	Tyanova et al. 2016[58]	http://www.maxquant.org
Scikit-learn 0.22.1	PyPI	https://pypi.org/project/scikit-learn
Survirus	Rajaby et al., 2021[8]	https://github.com/kensung-lab/SurVirus
UMI-tools	Smith et al., 2017[59]	https://umi-tools.readthedocs.io/en/latest/
Cell ranger	10X genomics	https://github.com/10XGenomics/cellranger
Seurat	Stuart et al. 2019[60]	https://satijalab.org/seurat
Copykat	Gao et al. 2021[26]	https://github.com/navinlabcode/copykat/#step-5-define-subpopulations-of-aneuploid-tumor-cells
SCENIC	Aibar et al. 2017[27]	https://scenic.aertslab.org/
AUCell	Aibar et al. 2017[27]	https://bioconductor.org/packages/release/bioc/vignettes/AUCell/inst/doc/AUCell.html
CellChat	Jin et al. 2021[28]	http://www.cellchat.org/
R 4.0.1	CRAN	https://www.r-project.org/
ggplot2	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
DESeq2	Bioconductor	https://bioconductor.org/packages/release/bioc/html/DESeq2.html

(*Continued on next page*)

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ClusterProfiler	Bioconductor	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Fgsea	Bioconductor	https://bioconductor.org/packages/release/bioc/html/fgsea.html
GSVA	Bioconductor	https://bioconductor.org/packages/release/bioc/html/GSVA.html
ESCAPE	Bioconductor	https://bioconductor.org/packages/release/bioc/html/escape.html
Survcomp	Bioconductor	https://bioconductor.org/packages/release/bioc/html/survcomp.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chaoyang Sun (suncydoctor@gmail.com).

Materials availability

Due to the limiting nature of primary samples, human tissues used in this study are not available upon request. This study did not involve any other unique materials.

Data and code availability

Whole exome sequencing, RNA-seq, scRNA-seq data from this study have been deposited in the National Genomics Data Center (NGDC) under the accession number NGDC: PRJCA009817. Code used for multi-omics data analyses is available at Zenodo: https://doi.org/10.5281/zenodo.7033608. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient cohort

This study was reviewed and approved by the Institutional Review Board of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (TJ-IRB20210535). All patients signed an informed consent form, and all samples collected were the rest after meeting to standard diagnostic tests, with no burden to the patients. All clinical data for the cohort is included in Table S1.

METHOD DETAILS

Sample quality

For quality control of the samples, the following criteria were used for sample selection, 1) tumor cellularity analysis, three frozen sections (top, middle, and bottom) of each tissue block were resected and subjected to H & E staining. All H & E-stained slides were scanned and submitted for pathological analysis. 2) Only tumors with at least the average of 60% tumor cell nuclei and less than 10% necrosis were reserved. 3)The histological assessment of all tumor samples was accomplished by two experienced pathologists separately.

Sample processing for DNA and RNA extraction

DNA and RNA were extracted from tumor and adjacent normal samples using QIAGEN's AllPrep DNA/RNA Mini Kit. DNA concentrations were measured with Qubit® DNA Assay Kit in Qubit® 2.0 Fluorimeter. Any sample that produced sufficient DNA yield and passed quality control was subjected to sequencing. RNA concentration was quantified using the Qubit and quality was assessed using Agilent 2100. A sample that passed RNA quality control and had a minimum RNA integrity number (RIN) score of 7 was sent for sequencing.

Whole-exome sequencing

Library construction was performed with 1μg DNA per sample using Agilent SureSelect Human All Exon V6 kit (Agilent Technologies, CA, USA) following the manufacturer's recommendations. Briefly, DNA was fragmented to 180-280 bp by hydrodynamic shearing system (Covaris, Massachusetts, USA) and processed with end repair/A-tailing, adaptor ligation, and library enrichment PCR using

Kapa Hyper-Prep reagents. The resulting libraries were hybridized with liquid phase with biotin labeled probe, then processed using magnetic beads with streptomycin to capture the exons of genes. Captured libraries were enriched in a PCR reaction to add index tags to prepare for sequencing. Products were purified using AMPure XP system (Beckman Coulter, Beverly, USA) and quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system.

The clustering of the index-coded samples was performed according to the manufacturer's instructions. After cluster generation, the DNA libraries were sequenced on Illumina NovaSeq platform and 150 bp paired-end reads were generated.

Bulk RNA sequencing

Library construction was performed using 2 μg RNA per sample using VAHTS mRNA-seq v2 Library Prep Kit for Illumina following manufacturer's recommendations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using fragmentation buffer. First strand cDNA was synthesized and second strand cDNA synthesis was subsequently performed. Remaining overhangs were converted into blunt ends. After adenylation of 3′ ends of DNA fragments, adaptors with hairpin loop structure were ligated. Then PCR was performed. At last, Qubit HS quantification, Agilent 2100 Bioanalyzer/Fragment Analyzer 5300 quality control, the final library size was about 350bp.

The libraries were sequenced on an Illumina NovaSeq platform to generate 150 bp paired-end reads, according to the manufacturer's instructions.

Single-cell RNA sequencing

Tumors were collected into heparin tubes (Becton, Dickinson and Co.) and processed using the Chromium Next GEM Single Cell 3 Reagent Kits v3.1 (10x Genomics; Pleasanton, CA, USA) according to the manufacturer's instructions as described in our previous study.[61]

Virus-capture sequencing

Libraries were prepared as described in our previous work.[4] In brief, The DNA extracted was sheared to 150-200 bp DNA fragments, which were then purified, end blunted, A-tailed and adaptor-ligated. Libraries were hybridized with HPV probes (HPV6, 16, 18, 26, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73 and 82) at 65°C for 24h and washed to remove uncaptured fragments. The products were amplified and proceeded to 151 cycles of paired-end index sequencing in the Illumina NovaSeq sequencer according to manufacturer's instructions.

Mass spectrometry analysis of protein and phosphoprotein

Protein extraction

The sample was grinded by liquid nitrogen into cell powder and then transferred to a 5-mL centrifuge tube. After that, four volumes of lysis buffer (8 M urea, 1% Protease Inhibitor Cocktail, 1% Phosphatase Inhibitor Cocktail) was added to the cell powder, followed by sonication three times on ice using a high intensity ultrasonic processor (Scientz). The remaining debris was removed by centrifugation at 12,000 g at 4°C for 10 min. Finally, the supernatant was collected and the protein concentration was determined with BCA kit according to the manufacturer's instructions.

For digestion, the protein sample was precipitated with TCA at 4°C for 2h. The protein precipitate was washed 2-3 times with acetone, and suspend with 200 mM TEAB. For trypsin digestion, trypsin was added at 1:50 trypsin-to-protein mass ratio for the first digestion overnight, and the protein solution was reduced with 10 mM DTT for 1 h at 37°C and alkylated with 20 mM IAA for 45 min at room temperature in darkness. Finally, trypsin was added 1:100 trypsin-to-protein mass ratio for a second 4 h-digestion.

TMT labeling

After trypsin digestion, peptide was desalted by Strata X C18 SPE column (Phenomenex) and vacuum-dried. Peptide was reconstituted in 0.5 M TEAB and processed according to the manufacturer's protocol for 11-PLEX TMT kit. Briefly, one unit of TMT reagent (defined as the amount of reagent required to label 100 μg of protein) was thawed and reconstituted in 24 μL buffer. The peptide mixtures were then incubated for 2 h at room temperature and pooled, desalted and dried by vacuum centrifugation. To facilitate quantitative comparison between all samples across experiments, a mix reference sample was included in each 11-plex (Table Labeling information).

HPLC fractionation

To reduce sample complexity, peptides samples was then fractionated into fractions by high pH reverse-phase HPLC using Agilent 300Extend C18 column (5 μm particles, 4.6 mm ID, 250 mm length). Briefly, 3.3 mg peptides were first separated with a gradient of 2%–60% acetonitrile in 10 mM ammonium bicarbonate pH 10 over 80 min into 80 fractions. The fractions were combined into 12 fractions, 10% of the volume of each of the fractions was allocated for proteome analysis. The remaining 90% of 12 concatenated fractions were further combined into 6 fractions for phosphopeptide enrichment.

Affinity enrichment

Peptide mixtures were first incubated with IMAC microsphere suspension with vibration in loading buffer (50% acetonitrile/6% trifluoroacetic acid). The IMAC microspheres with enriched phosphopeptides were collected by centrifugation, and the supernatant was removed. To remove nonspecifically adsorbed peptides, the IMAC microspheres were washed with 50% acetonitrile/6% trifluoroacetic acid and 30% acetonitrile/0.1% trifluoroacetic acid, sequentially. To elute enriched phosphopeptides from the IMAC

microspheres, elution buffer containing 10% NH4OH was added and the enriched phosphopeptides were eluted with vibration. The supernatant containing phosphopeptides was collected and lyophilized for LC-MS/MS analysis.

LC-MS/MS analysis

Peptides were subjected to NSI source followed by tandem mass spectrometry (MS/MS) in Q Exactive HF-X (Thermo) coupled with EASY-nLC 1200 UPLC system. The gradient was comprised of an increase from 6% to 22% solvent B (0.1% formic acid in 90% acetonitrile) over 38 min, 22%–32% in 14 min and climbing to 80% in 4 min then holding at 80% for the last 4 min. The phosphoproteome gradient was comprised of an increase from 3% to 20% solvent B (0.1% formic acid in 90% acetonitrile) over 88 min, 20%–32% in 24 min and climbing to 80% in 4 min then holding at 80% for the last 4 min.

The electrospray voltage applied was 2.2 kV. Intact peptides were detected in the Orbitrap at a resolution of 120,000. Peptides were then selected for MS/MS with NCE setting as 28 and the fragments were detected in the Orbitrap at a resolution of 45,000. A data-dependent procedure was alternated between one MS scan followed by 15 MS/MS scans with 30.0s dynamic exclusion. Automatic gain control (AGC) was set at 5×10^4, with an intensity threshold of 5000 charges and a maximum injection time of 90 ms. Fixed first mass was set as 100 *m/z*.

Western blots

Protein extracts from randomly selected samples were quantified using Coomassie (Beyotime, Cat. no: ST1119) and 20 μg of total protein was used for western blots. Proteins were separated on 10% SDS-PAGE gels (BioSci, Cat. no: 8012011) and transferred onto polyvinylidene difluoride membranes (Cytiva Life Sciences, Cat. no: 10600023). Membranes were blocked in 5% BSA for 1 h and incubated overnight at 4°C with antibodies specific for: RB (1:500, Abclonal, Cat. no: A11409), p53 (1:500, Abclonal, Cat. no: A16989), HPV16-E6 (1:1,000, Genetex, Cat. no: 132686), HPV16-E7 (1:1,000, Genetex, Cat. no: 133411;), GAPDH (1:1,000, Abclonal, Cat. no: A19056), followed by HRP-conjugated secondary antibodies (1:5,000, Abclonal, Cat. no: AS014 and Cat. no: AS003) for 1 h at room temperature. Bands were visualized using a Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific; Cat. no: 32209).

QUANTIFICATION AND STATISTICAL ANALYSIS

Copy number analysis

Copy number variation (CNV) analysis was performed on WES data generated from tumor and paired NAT. After adapter trimming, clean data was aligned to GRCh38 genome using BWA.[52] Following de-duplication and base recalibration using GATK[53] following the tool's best practice, Somatic CNVs were called using CNVkit[54] and the CNV of genes were extracted from the result using custom scripts.

RNA quantification

After adapter trimming, clean RNA sequencing data were aligned to GRCh38 reference using STAR.[55] The read count is generated by subread[56] with default parameters and the transcripts per million (TPMs) of genes were quantified using RSEM.[57] Gene annotation used in this study is Gencode release 33.

Protein and phosphoprotein quantification

Proteomics database search analysis

The MaxQuant search engine[58] was used for the standard database search of all MS/MS raw data. The MS/MS spectra were searched against the human Swiss-Prot database containing 20,431 sequences (downloaded in August, 2019) plus 475 Swiss-Prot HBV protein sequences. The cleavage enzyme was set as trypsin/P with allowing up to 2 missing cleavages. The mass tolerance for precursor ions was set as 20 ppm in first search and 4.5 ppm in main search, and the mass tolerance for fragment ions was set as 0.02 Da. The minimum peptide was set as 7. The fixed modification was set as Carbamidomethyl on Cys, oxidation on Met, acetylation on protein N-term and deamidation (NQ) were the variable modifications, and the Phospho (STY) was set for searching phosphoproteomes only. FDRs of PSM and protein were set to <1%, and the minimum score for modified peptides was set to >40. The TMT 11-plex was selected in the quantification method, and all other parameters of MaxQuant were set as default values.

Data normalization

The TMT-based intensity matrix of proteins or phosphorylation sites was extracted from the MaxQuant output files for all batches. For each protein or phosphorylation site, the TMT ratio value was calculated by dividing the intensity in the internal reference cell line of the same batch. The proteomic and phospholomic data of each sample was further normalized using the median centering method. After normalization, the normalized abundance values (NAVs) of proteins or phosphorylation sites followed a logarithmic normal distribution (log2) centered at zero.

Missing value imputation

To ensure the data quality and maximally use the proteomic or phospholomic data, proteins or phosphorylation sites quantified in >50% samples were reserved. Using the normal distribution imputation, the missing values were imputed with values representing a normal distribution around the detection limit of the mass spectrometer. For each sample, the mean and SD(SD) of the distribution of the raw protein or phosphorylation site were calculated. Then a new distribution with a downshift of 1.8 SD and a width of 0.3 SD

was automatically modeled. The total dataset was imputed before statistical analysis. The data imputation was conducted using Perseus 1.6.14.

Batch effect and data quality analysis

The batch effect due to TMT multiplexes was assessed by performed performing PCA was performed using scikit-learn, a useful toolkit for data mining and analysis. The leading PCs of the global proteomic or phospholomic data clearly separated the tumor from normal samples, and no obvious batch effect was observed among the 23 TMT batches.

Using a single-shot liquid chromatography–tandem mass spectrometry approach, we finally quantified a total of 6,739 unique protein groups and 17,760 phosphorylation sites quantified in >50% samples in our cohort.

HPV analysis

HPV typing and integration analysis were performed at DNA, RNA and scRNA levels using a reference GRCh38 contained HPV genome.

HPV typing and integration analysis

For VCS data, reads were cleaned and aligned to the reference using BWA-MEM. The aligned reads were de-duplicated using Picard and the ones that paired-end aligned perfectly to reference were counted and normalized to reads per base (read depth). HPV types were determined as the one who's mean read depth $\geq$ 10. HPV integration breakpoints were called using Survirus[8] and the breakpoints with support reads $\geq$ 10 were retained (Table S2).

For bulk RNAseq data, cleaned RNAseq reads were aligned to the reference using STAR and the read count of HPV was generated by subread with an HPV bed file. The HPV counts were normalized to count per million reads (CPM) and the HPV types were determined as the ones with CPM $\geq$1. HPV integration fusion transcripts were detected with STAR with params "–outSAMstrandField intronMotif –chimSegmentMin 12 –chimJunctionOverhangMin 8 –chimOutJunctionFormat 1 –alignSJDBoverhangMin 10 –alignMatesGapMax 100000 –alignIntronMax 100000 –limitSjdbInsertNsj 1500000 –alignSJstitchMismatchNmax 5 -1 5 5 –outSAMattrRGline ID:GRPundef –chimMultimapScoreRange 3 –chimScoreJunctionNonGTAG −4 –chimMultimapNmax 20 –chimNonchimScoreDropMin 10 –peOverlapNbasesMin 12 –peOverlapMMp 0.1 –alignInsertionFlush Right –alignSplicedMateMapLminOverLmate 0 –alignSplicedMateMapLmin 30" and HPV integration sites were calculated from chimeric output using custom scripts and retained if support reads $\geq$ 10 (Table S3).

For scRNA data, Unique molecular identifiers (UMIs) and cell barcodes were removed from sequencing data using UMI-tools[59] and the clean data was processed as described in RNAseq data.

HPV integration type and status analysis

The integration types of HPV integration sites were determined by the direction of the HPV sequence fusion to the host genome, while human-HPV indicates that human genomic sequences are 5′ to the HPV sequence, and HPV-human indicates that human sequences 3′ to the HPV sequence.

For integration status, the DNA integration sites were defined as productive if there exists RNA integration site <100 kb apart from with identical integration type in the same tumor., otherwise, DNA integration sites were defined as silent.

For integration events, integration sites within 500 kb apart were merged into a single integration event (Tables S4 and S5) as mentioned previously.[11]

HPV gene quantification

To determine the expression of HPV viral genes (Figures S3B–S3D), the aligned reads from RNAseq were performed using subread with a bed file containing HPV genes and the resulting counts were normalized to fragments per kilobase of exons per million reads mapped to genome (FPKM). Unspliced and spliced E6 (Figure S3D) were quantified by counting the reads that support unspliced E6 and spliced transcripts at the E6 splice donor site (position 226 for HPV16 and position 233 for HPV18) using custom script.[12]

Single-cell RNA-seq data processing

The sequencing data were processed using cell ranger against the GRCh38 reference contain 3 HPV genomes (HPV16, HPV56, HPV52). Cells were filtered for identification of more than 200 genes, unique molecular identifier between 500 and 30,000, and a mitochondrial gene proportion less than 15% using Seurat.[60] After that, we use Seurat method for data normalization, batch effect removal, integration, UMAP dimensionality reduction and clustering.

The major cell types were defined through per cluster mean expression of the following gene markers: 1) cervical epithelial: KRT6A, KRT17; 2) endothelial cells: RAMP2 A2M; 3) fibroblasts: DCN, LUM; 4) myocyte: ACTA2, RGS5; 5) mast cells: TPSB2, TPSAB1; 6) myeloid cells: LYZ, CD14; 7) NK cells, NKG7, KLRD1; 8) T cells: CD2, CD3E; 9) B cells: CD79A, MS4A1. For further analysis of the cervical epithelial cell, lymph cell (T, NK and B cells), myeloid cell (myeloid and mast cell) and stromal cell (endothelial, fibroblasts and myocyte) populations, the data of these cells was extracted, principal component analysis and visualized as described above. Cluster gene markers were identified using the Seurat FindAllMarkers function with default params.

To identifying neoplastic from epithelial cells, we use myocytes to define the reference cell-inferred copy number profiles and the predicted karyotype for individual cells was estimated using Copykat method.[26]

Identification of subpopulations

The fibroblast subclusters were identified by calculating gene signature scores using Seurat AddModuleScore of previously reported fibroblast phenotypes: panCAF, IL1-driven and TGFB-driven CAF, and fibroblast-like.[62] The myeloid subclusters were identified by

comparing the expression of marker gene of previously reported clusters: TAM-like, C4-LAMP3+, FCN1 mono, DC1, DC2, mast cell and neutrophils.[63–65]

Differential expressed gene and pathway enrichment

For RNAseq data, differential expressed gene (DEG) analysis was performed with gene counts using DESeq2. For proteomics data, differential expressed genes between groups were performed by comparing the fold changes of mean gene expression in the group.

Pathways from the Reactome, KEGG, GO and Hallmark databases were extracted from Molecular Signatures Database (MSigDB) for pathway enrichment analysis. The enriched pathways were obtained using ClusterProfiler or fgsea package depends on the data types.

Pathway score calculation

To obtain the pathway scores, single-sample gene set enrichment analysis (ssGSEA) was utilized based on single-cell normalized counts using R package ESCAPE, while the mean expression of genes in pathways was used for proteomics data.[19]

Transcription factor analysis

To carry out transcription factor network inference, data from malignant cells were performed using the SCENIC, which corresponds to RcisTar-get hg38 motif databases. Activity of the regulatory networks was evaluated on the full dataset in the scoring step with AUCell.[27]

Cell-cell communication

Cell-cell interactions were performed using CellChat.[28] We load the normalized counts into CellChat and run the official workflow with default parameters. To compute the contribution of each ligand-receptor pair to the overall signaling pathway, the function netAnalysis_contribution was applied. To determine the senders and receivers in the network, the function netAnalysis_signallingRole was applied. To visualize the inferred communication networks, the function netVisual_aggregate, was applied.

Generalized linear models

Demographic and clinical characteristics, including age, clinical stages, tumor grades, tumor diameter, and invasive depth, from 90 CCs were fitted with generalized linear model using R 4.0.1. R package ggplot2 was used to show the odds ratios, p values, and 95% CI between silent and productive integration group.

Survival analysis

We performed a multistep analysis to get gene expression signatures for productive or silent integration tumors as previously reported.[64] Briefly, first, the DE-Gs from RNAseq with adjust p value < 0.05 were selected. Next, C-index of each gene was calculated using a univariate Cox proportional-hazards regression model with the R package survcomp. Finally, 8 genes were obtained with C-index > 0.5 and an absolute $\log_2$ fold change >1.

To explore HPV integration status on the survival in CCs, we selected tumor samples from 263 patients with RNAseq data and follow-upinformation in the TCGA CESC cohort. The samples were defined as silent or productive based on the following equation:

For gene g.

$$L\,g,j = \begin{cases} 1 * I(g), E\,g,j > median(E\,g) \\ -1 * I(g), E\,g,j \leq median(E\,g) \end{cases}$$

I(g) = 1 if gene g is highly expressed in our productive group; otherwise, I(g) = −1.

L, g, j and E present label, gene, jth sample and gene expression, respectively.

The scores of 8 genes for each TCGA samples were summed and the sample was defined as "productive type" while the total scores >0. For survival analyses, including overall survival (OS) and survival time, we used the log rank test to calculate p values between groups, and the Kaplan-Meier method to plot survival curves using R package survival and survminer.