

## ORIGINAL RESEARCH

# Genomic, transcriptomic, and viral integration profiles associated with recurrent/metastatic progression in high-risk human papillomavirus cervical carcinomas

Jing Jing Liu<sup>1,2,3</sup>  | Jung Yoon Ho<sup>1,3</sup> | Jung Eum Lee<sup>1</sup> | Soo Young Hur<sup>1,3</sup> | Jinseon Yoo<sup>3,4</sup> | Kyu Ryung Kim<sup>3,4</sup> | Daeun Ryu<sup>3,4</sup> | Tae Min Kim<sup>3,4</sup> | Youn Jin Choi<sup>1,3</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

<sup>2</sup>Department of Obstetrics and Gynecology, Yantai Affiliated Hospital of Bin Zhou Medical University, College of Medicine, Bin Zhou Medical University, Yantai, China

<sup>3</sup>Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

<sup>4</sup>Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

## Correspondence

Tae Min Kim, Department of Medical Informatics, College of Medicine, The Catholic University of Korea, 222, Banpo-daero, Seocho-gu, Seoul, 06591, Republic of Korea.  
Email: tmkim@catholic.ac.kr

Youn Jin Choi, Department of Obstetrics and Gynecology, Seoul Strz. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222, Banpo-daero, Seocho-gu, Seoul, 06591, Republic of Korea.  
Email: yunno@catholic.ac.kr

## Funding information

National Research Foundation of Korea, Grant/Award Number: 2016R1D1A1B04932862 and 2019R1A5A2027588

## Abstract

Acquisition of recurrent/metastatic potential by a tumor cell defines a critical step in malignant progression. However, understanding of metastatic progression at the molecular level is scarce for cervical carcinomas (CES). In this study, we performed genomic, transcriptomic, and viral profiling of five pairs of primary (CES-P) and matched recurrent/metastatic tumors (CES-R/M) with high risk human papillomavirus. Whole exome sequencing revealed mutation features of CES-R/M including elevated mutation burdens and prevalent copy number alterations compared to their matched CES-P. A relative deficit of APOBEC-related mutation signatures accompanying the transcriptional downregulation of *APOBEC3A* was observed for CES-R/M. Mutations in genes encoding epigenetic regulators were commonly observed as CES-R/M-specific alterations. Immunoprofiling and gene set analysis revealed CES-Ps were enriched with transcripts representing activated anticancer immunity such as interferon-gamma pathway, while CES-R/M exhibited upregulation of genes involved in epithelial-mesenchymal transition and angiogenesis. Viral capture sequencing revealed that integration sites remained enriched in viral E1 protein domain during malignant progression. Moreover, we found transcriptional upregulation of *POSTN* and downregulation of *APOBEC3A* were associated with unfavorable clinical outcomes in CES. Comprehensive genomic and transcriptomic profiling of a rare cohort including CES-R/M identified metastases-specific features to advance the molecular understanding into CES metastatic progression with potential clinical implications.

## KEYWORDS

APOBEC, cervical cancer, human papillomavirus, immunoprofiling, metastasis, recurrence

Tae Min Kim and Youn Jin Choi are equal contributors to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd

## 1 | INTRODUCTION

Cervical cancer (CES) is the fourth most common cause of morbidity and mortality among women. Disease recurrence often accompanying locoregional or distant metastases is responsible for ~90% of cancer deaths.<sup>1</sup> The majority of CES recurrence occurs locally (pelvis) and distant metastases occur in organs such as the lung (21%), bone (16%), para-aortic lymph nodes (11%), abdominal cavity (8%), supraclavicular nodes (7%), liver (4%), and heart (<1%).<sup>2</sup> Approximately 10% of patients at stages IB-IIA experience relapse or metastasis following surgical treatment including radical hysterectomy. Median survival of patients with recurrent CES is only 7-17 months and the majority (77%) of cases relapses within 3 years.<sup>3</sup> The incidence of recurrence and metastasis of cervical cancer at stages IA, IB, IIA, IIB, III, and IVA within 10 years from the diagnosis is 3%, 16%, 31%, 26%, 39%, and 75% respectively. A substantial challenge for public health remains, given the relatively low 5-year survival rate (<20%) for metastatic CES compared to that of localized cervical cancer (91.5%).<sup>4</sup>

Carcinogenesis is an evolutionary process whereby cells or clones in the cancer mass continuously acquire genetic diversity in the form of somatic mutations, rendering them subject to clonal selection. Consequently, tumor subclones often display distinct genetic heterogeneity within a tumor mass. Heterogeneity can be sub-classified into spatial and temporal genetic heterogeneity and is proposed to be the main cause of treatment failure, drug resistance, and tumor recurrence and metastasis.<sup>5,6</sup> Recently, advanced high-throughput genomic analyses have enabled the genomic characterization of different types of human malignancies as reported in a large-scale study from the Cancer Genome Atlas consortium.<sup>7</sup> In addition to identifying clinically relevant genomic features, high resolution genomic profiling can be employed to study cancer evolution in that whole-genome or whole-exome sequencing-derived mutations can serve as evolutionary markers to trace genomic evolution and spatial-temporal heterogeneity. The comparison of multiple lesions obtained from a single patient including primary and metastatic lesions may also be valuable in identifying genomic features unique to metastatic lesions. Using multiregion biopsies from the same individual, investigators have evaluated spatial heterogeneity<sup>8,9</sup> as well as the genomic changes accompanying benign-to-malignant or primary-to-metastatic progression.<sup>10</sup>

The cancer transcriptome has also been reported to include a substantial number of transcripts originating from nontumor cells that can be exploited to investigate tumor microenvironments. For example, the relative abundance of tumor-infiltrating immune cells and stromal cells can be estimated from bulk-level transcriptome sequencing data.<sup>11</sup> Moreover, such inferences can be further refined using previously acquired information such as the gene expression profiles of immune cells, as detailed in reports using the CIBERSORT algorithm.<sup>12</sup> These approaches have become more important as immune checkpoint inhibitors

have recently shown considerable clinical success.<sup>13</sup> Moreover, viral genome-capture sequencing has been used to identify viral integration sites as genomic hallmarks of human papillomavirus (HPV)-positive CES genomes.<sup>14</sup>

In the present study, we conducted integrative molecular profiling on multiple pairs of primary and recurrent/metastatic cervical cancers (CES-P and CES-R/M, respectively) from the same individuals, using three sequencing-based methods: whole-exome sequencing (WES), RNA sequencing (RNA-seq), and high-throughput viral integration detection (HIVID). We aimed to investigate: i) genomic and transcriptomic features of CES-P and CES-R/M and ii) CES-R/M-specific genomic and transcriptomic markers that may be associated with malignant progression of the disease.

## 2 | MATERIALS AND METHODS

### 2.1 | Patient samples

Paired normal, primary, and recurrent/metastatic tumor tissues were obtained from five cervical cancer patients (IB1-IVB). All subjects underwent radical hysterectomy and adjuvant treatment after their first diagnosis and received either debulking surgery or biopsy upon recurrence or metastasis. HPV genotyping was performed on tumor samples; all patients were found to be infected with high risk HPV types (HPV16, 18, and 31). Ethical approval for the study was obtained from the relevant ethics committees. Formalin-fixed paraffin-embedded (FFPE) tissue blocks were micro-dissected to obtain high-purity (>70%) primary and recurrent/metastatic tumor tissues. A DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) was used to extract the genomic DNA according to the manufacturer's recommendations.

### 2.2 | Whole-exome sequencing and DNA library preparation and qualification

A QIAamp DNA FFPE Tissue Kit (catalog number 56404; Qiagen) was used to extract the genomic DNA from tissue archives according to the manufacturer's recommendations. Exonic DNA capture was performed using an Agilent SureSelect Human All Exon 50 Mb kit (Agilent Technologies). A total of 500 ng DNA libraries were prepared and 101-bp paired-end sequencing reads were generated on an Illumina HiSeq2000 sequencer (Illumina). The Burrows-Wheeler aligner (BWA) was used to align sequencing reads against a human reference genome (UCSC hg19, NCBI build 37.1).<sup>15</sup> The Genome Analysis ToolKit (GATK) was used for local realignment and score recalibration of sequencing reads. Samtools and Picard were used to manage sequencing data.<sup>16</sup> MultiQC v1.1 (<http://multiqc.info/>) was used for raw

read- and alignment-level quality control by aggregating QC metrics from FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) into an interactive report HTML and then was converted into PDF file (**Extended Data 1 and 2**). We used Mutect and Indelocator with the default settings to detect somatic mutations and indels by comparing tumor and matched normal sequencing reads (ie CES-P and CES-R/M sequencing reads were compared with their matched normal data, Table S1).<sup>17</sup> Indels were filtered out with a stringent cut-off (mutant allele frequency or MAF > 0.2). Mutations arising due to sequencing artifacts including FFPE-related DNA fragmentation, were further filtered out using a software of FIREVAT (FInding RELiable Variants without ArTifacts).<sup>17</sup> The ANNOVAR package was applied to annotate the functional consequence of somatic mutations on the encoded amino acid residues.<sup>18</sup> Tier1 mutations are those with strong evidence of their oncogenic or tumor suppressive roles as presented in the Cancer Gene Census (CGC, <https://cancer.sanger.ac.uk/census>). We also classified somatic mutations into three regional categories, with common mutations including those observed in both CES-P and CES-R/M in given cases and CES-P- and CES-R/M-private mutations representing those observed only in the corresponding lesions.

### 2.3 | Mutational signature analysis

A total of 65 mutation signatures as normalized frequencies of 96 trinucleotide contexts representing six substitution subtypes (ie C>A, C>G, C>T, T>A, T>C, and T>G) along with bases immediately 5' and 3', were obtained from COSMIC database.<sup>19</sup> ([https://cancer.sanger.ac.uk/cosmic/signatures/SBS/mutation\\_signatures\\_version\\_3](https://cancer.sanger.ac.uk/cosmic/signatures/SBS/mutation_signatures_version_3)). The mutation signature levels or weights from a mutation profile (ie the frequency of mutations in the context of 96 trinucleotides) were estimated using the deconstructSigs R extension.<sup>20</sup> We performed mutation signature analysis for three mutation categories, ie, the common and CES-P-/CES-R/M-private mutations for each patient. Among the mutation signatures, we selected 16 signatures with known causal relationships, ie, SBS1/2/3/4/6/7a/7b/7d/10a/10b/13/14/15/20/26/44, and others. The mutation signature weights were combined when possible (eg SBS 2/13 associated with APOBEC activity and SBS6/14/15/20/26/44 associated with a deficiency of DNA mismatch repair, respectively).

### 2.4 | Clonal mutational structures

The number of mutation clones of given cases were inferred by SciClone (version 1.1.0)<sup>21</sup> (Extended Data Figure 1). SciClone applies a Bayesian clustering to the joint mutation profiles of paired samples such as longitudinal or multiregion biopsies (spatially or temporally separate) samples to infer

the sub-clonal mutation architecture and their evolutionary dynamics of a given cancer genome.

### 2.5 | Copy number profiling

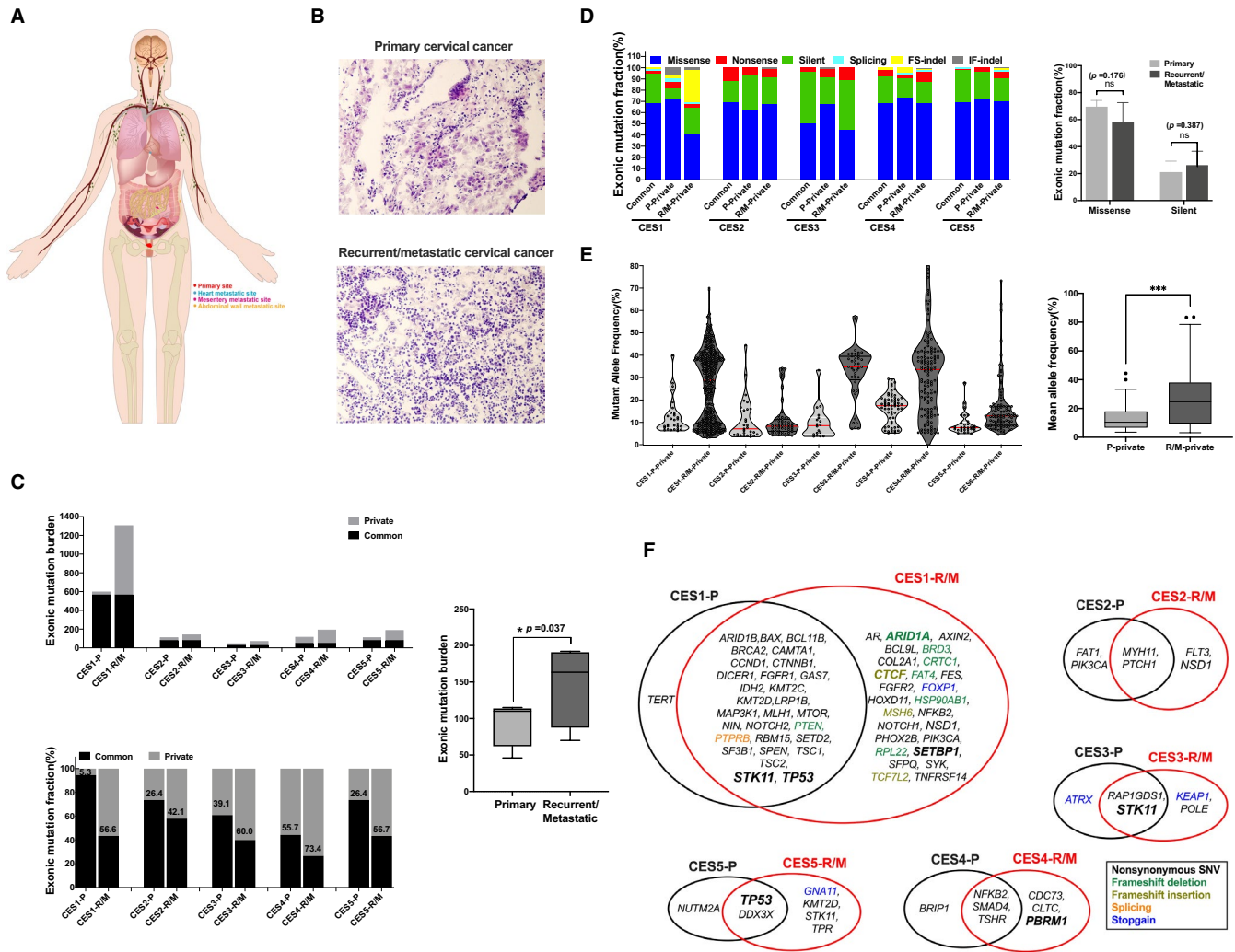
For copy number profiling, we used the Excavator algorithm based on read depth differences between the tumor and matched normal sequencing data.<sup>22</sup> To define amplifications and deletions, we set thresholds for the log<sub>2</sub> ratios (eg, amplification and deletion for >0.2 and < -0.2 log<sub>2</sub> ratio, respectively). According to the overlap between CES-P and CES-RM genomes, individual segments were classified as common or private to CES-P or CES-R/M. The visualization of copy number profiles was prepared using the IGV browser.

### 2.6 | RNA sequencing

RNA extraction was performed with the FFPE RNeasy kit (Qiagen, catalog 73504) according to the manufacturer's instructions under sterile RNase/DNase-free conditions. RNA concentration was determined with the NanoDrop (ND-1000) spectrophotometer (Thermo Fisher, Scientific, Inc). Transcriptome analysis of CES-P and CES-R/M was performed on an Illumina HiSeq2500 system (Illumina). A TruSeq RNA Access Library Prep Kit (Illumina) with 50 ng RNA was used to prepare cDNA libraries according to the manufacturer's instructions. Sequencing read quality was assessed for pre-mapped sequencing reads of R1 and R2 FASTQ files as well as aligned reads in BAM files using FastQC (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>). Low quality reads were filtered out and the remaining reads were hard trimmed using Trimmomatic. The cleaned-up reads were aligned using the Tophat2 splice-aware aligner.<sup>23</sup> Pre/post-mapped sequencing reads were visualized for QC metrics using MultiQC v1.1 (<http://multiqc.info/>) (**Extended Data 3 and 4**). FPKM (fragments per kilobase million) values were obtained from CuffLinks<sup>24</sup> and raw count values were calculated with the featureCounts tool using the Galaxy web-based platform (<https://usegalaxy.org/>). For transcripts, we used annotations derived from hg19 annotations (UCSC, TCGA.hg19.June2011.gaf). *P*-values were obtained by a paired *t* test between the primary and matched recurrent/metastatic tumor groups and fold changes were calculated. Differentially expressed genes (DEGs) were identified as those with log<sub>2</sub> |(fold change)| > 1 and *P*-values < .05 using DESeq2.<sup>25</sup>

### 2.7 | Immunoprofiling

The CIBERSORT deconvolution algorithm<sup>26</sup> was applied to the CES-P and CES-R/M RNA-seq data to estimate the relative



**FIGURE 1** Mutation abundance and features of primary and matched recurrent/metastatic cervical cancers. A, Schematic diagram of cervical cancer with different metastatic sites recruited in this study. B, Microscopic features of primary cervical cancer (top), and metastatic tumor on mesentery (bottom) by hematoxylin and eosin (H&E) staining. C, Abundance of exonic mutations (upper panel) are shown along the y-axis across five cases with each primary (P) and recurrent/metastatic (R/M) sample. Mutation fractions (lower panel) are further classified into common and primary mutations (black and grey, respectively). \*\* $P$ -value < .01 D, The relative abundance of six categories representing the functional consequences of somatic mutations are shown and the relative abundance of missense and silent mutations are compared between CES-P and CES-R/M tumors. E, The mutant allele frequencies (MAFs) are shown for CES-P and CES-R/M genomes with comparison of matched genomes. F, Cancer Census Genes (Tier-1) are listed for three regional categories of common, P-, and M-specific driver mutations for five cases

abundance of 22 immune cell subtypes. CIBERSORT algorithm is based on support vector regression and known to provide reliable estimates on the relative abundance of immune cells.

## 2.8 | Gene set enrichment analysis (GSEA)

To identify the functionally coordinated gene expression changes between CES-P and CES-R/M, we applied GSEA (version 2.0, <http://software.broadinstitute.org/gsea>) using three types of functional gene sets as available in the MSigDB database (<http://software.broadinstitute.org/gsea/msigdb/index.jsp>). The gene set categories used in the analysis were H (Hallmark), C2cp (curated pathways), and C5

(Gene Ontology) gene sets. The enrichment score (ES) or normalized ES along with permutation-based  $p$ -values were calculated for individual gene sets with multiple test adjustments. A conventional threshold  $P$ -value < .05 and false discovery rate (FDR) < 0.25 were considered significant.

## 2.9 | Human papillomavirus integration analysis (HIVID)

Genomic DNA was extracted using a QIAamp DNA FFPE Tissue Kit (catalog number 56404; Qiagen) according to the manufacturer's protocol. gDNA fragmentation, library preparation and sequencing were performed as described previously.<sup>27</sup>



HPV typing was done for five pairs of CES-P and CES-R/M cases (Table 1) using HPV probes<sup>28</sup> designed by MyGenostics (MyGenostics, Baltimore, MD, USA). Structural variation detection algorithms of SVDetect<sup>29</sup> and CREST<sup>30</sup> were used to identify spanning reads or soft-clipped reads supporting the HPV integration sites.<sup>27</sup> The paired end sequencing read, with one end uniquely aligned to the human genome (hg19 with chrY excluded) and another to the HPV genome, is defined as a discordant read pair.<sup>31</sup> A potential HPV integration locus requires one or more discordant read pairs<sup>32</sup> and at least one split read.<sup>33</sup> Based on the experimental threshold from MyGenostics GenCap Technology, the local-assembled sequence of more than three, soft-clipped reads consistent with the integration site reads, is considered to be a high-confidence site. HPV integration sites supported by either of the algorithms (SVDetect or CREST only) are low-confidence sites.

## 2.10 | Statistics analysis

Differential mutation abundance and gene expression between the primary and matched recurrent/metastatic tumor groups were analyzed using paired Student's *t* test. Overall survival (OS) and disease-free survival (DFS) were evaluated using the Kaplan-Meier method and the significance level was estimated using the log-rank test. Statistical analyses were performed with GraphPad Prism version 8.0 software (GraphPad Software, Inc). Error bars are shown, all data are represented by median  $\pm$  standard deviation (SD). Significance was defined as a *P*-value less than .05. All statistical tests are two-tailed, unless otherwise specified.

## 3 | RESULTS

### 3.1 | Exonic mutations of CES-P and CES-R/M

We noted a clear bimodal distribution of GC contents for the forward and the reverse reads across the cases examined. No other discernible systematic problems were noted for

the WES data (Extended Data 1 and 2). The median age of the five patients was 52 years (with a range of 34-58 years) (Table 1). Recurrent/metastatic CES occurred in the abdominal wall (3/5 patients), mesentery (1 patient), and heart (1 patient) (Table 1, Figure 1A). Representative images for primary and matched recurrent/metastatic tumor from CES1 mesentery site disclosed that cell morphology was more poorly differentiated in a large heterotypical change after recurrence (Figure 1B). We first identified two types of somatic mutations—single nucleotide substitutions (SNV or point mutations) and short insertions/deletions (indels)—from the WES data (hereafter, collectively referred to as somatic mutations). We compared the number of exonic mutations or mutation burdens between five pairs of CES-P and CES-R/M (Figure 1C). A total of 979 and 1894 somatic mutations identified in CES-P and CES-R/M, respectively, are available in Table S2. One of five cases (CES1) exhibited hypermutated genomes, with a high level of mutation burdens for CES1-P and CES1-R/M harboring 598 and 1305 somatic mutations respectively (Figure 1C). The mutation burdens of four recurrent/metastatic tumor samples (CES2-R/M–CES5-R/M, median 103.5; 70-192) were statistically significantly higher than those of matched primary tissues (median of 110 mutations, 46-115) (Figure 1C, \*\**P*-value = .037; paired *t* test). We observed that 55.1% (26.4-94.6%) of all SNVs were shared between CES-P and matched CES-R/M, representing common ancestral mutations (common mutations). For lesion-specific mutations, CES-P- or CES-R/M-specific mutations (as primary- or metastasis-private mutations) comprised 30.6% (5.4-55.7%) or 59.2% (49.0-73.4%) of mutations in given genomes respectively (Figure 1C, Table S2). The dominance of CES-R/M-specific mutations over CES-P-specific mutations suggests that CES-R/M genomes acquired more mutations than the CES-P genomes after metastatic dissemination; this also explains the higher mutation burdens of CES-R/M compared to those of CES-P.

For three regional mutation categories (common, P-, and R/M-private mutations), the consequences of mutations on encoded amino acids and mutant allele frequencies (MAFs) were further investigated (Figure 1D,E respectively). For CES1-R/M, frameshifting indels were prevalent compared to

**TABLE 1** Clinico-pathologic information of the five cervical patients

Patient ID	Age	Diagnosis	FIGO stage	Histologic differentiation	Metastasis site	HPV types	HPV typing by HIVID	Concurrent cancer
CES1	52	SCC	IVA	Grade 3	Mesentery	16	N/A	Colon cancer
CES2	44	SCC	IIB	Grade 2	Abdominal wall	18	18	—
CES3	58	ASC	IB1	Grade 3	Abdominal wall	16	16	Breast cancer
CES4	34	SCC	IIA2	Grade 3	Heart	16	16	—
CES5	49	SCC	IB1	Grade 3	Abdominal wall	31	31	—

Abbreviations: ASC, Adenosquamous cell carcinoma; N/A, Not available; SCC, Squamous cell carcinoma.

other genomes. The predominance of mutations and indels in CES1 suggests that this case may carry DNA repair deficiency, elevating the accumulation of both SNVs and indels. No significant differences between the ratio of missense and silent mutations were observed (Figure 1D). For MAFs, CES-R/M genomes showed significantly elevated levels of MAFs ( $12.88\% \pm 7.69\%$ ) compared to matched CES-P ( $25.13\% \pm 16.08\%$ ) (Figure 1E, \*\*\* $P$ -value  $< .001$ ). Lower MAFs in CES-P compared to CES-R/M suggested that the mutational subclonal architecture of primary tumors was more heterogeneous than that of metastatic tumors that may have undergone clone-level selection during metastatic dissemination. The initial seeding of metastatic clones originating from a single cell or oligoclonal cells from primary tumors must be clonal or oligoclonal, representing likely events of clonal selection associated with metastasis. In addition, the growth of metastatic tumors in lineage-unrelated microenvironments may allow for the acquisition of new subclonal mutations leading to increased mutation burdens of CES-R/M respectively. The newly acquired mutations—most of which remained subclonal without becoming fixed in the population—may constitute the subclonal mutations present in CES-R/M. In addition, the evolutionary dynamics associated with the metastatic progression was monitored by estimating the number of mutation clones between matched primary and metastatic genomes. The number of mutation clones were inferred using SciClone algorithm, which indicates the extent of heterogeneity of given cancer genomes. For two cases (CES1 and CES4), CES-R/M genomes showed more number of mutation subclones compared to CES-P genomes (eg, 2 and 4 mutation clones in primary and metastasis of CES1; 2 and 5 mutation clones in primary and metastasis of CES4, respectively, Extended Data Figure 1). For these tumors, metastatic progression accompanies the emergence of new sub-clones that might drive tumor growth with interactions between different sub-clones.<sup>34</sup> Moreover, for other tumors, the clone numbers were largely concordant between primary and metastasis (eg, 2/2 for CES1, 2/2 for CES2 and 2/5 for CES4) suggesting the evolutionary bottleneck during metastatic progression is not supported for these cases.

The number of cancer-related genes occurring in both primary and metastatic cervical cancer was also variable among patients (Figure 1F). Among the mutations involving cancer-related genes, frequent somatic mutation events were observed in *STK11* and *TP53*, as common mutations shared by CES-P and CES-R/M in three cases (Figure 1F), with similar frequencies of alterations as previously reported for cervical cancers.<sup>35</sup> For lesion-specific mutations, missense mutations on *NSDI* encoding histone H3 lysine 36 methyltransferase were observed as CES-R/M-specific mutations in two cases (CES1 and CES2). *NSDI* mutations are frequently observed in head-and-neck cancers<sup>36</sup> and are likely loss-of-function mutations leading to genome-wide hypomethylation.<sup>37</sup>

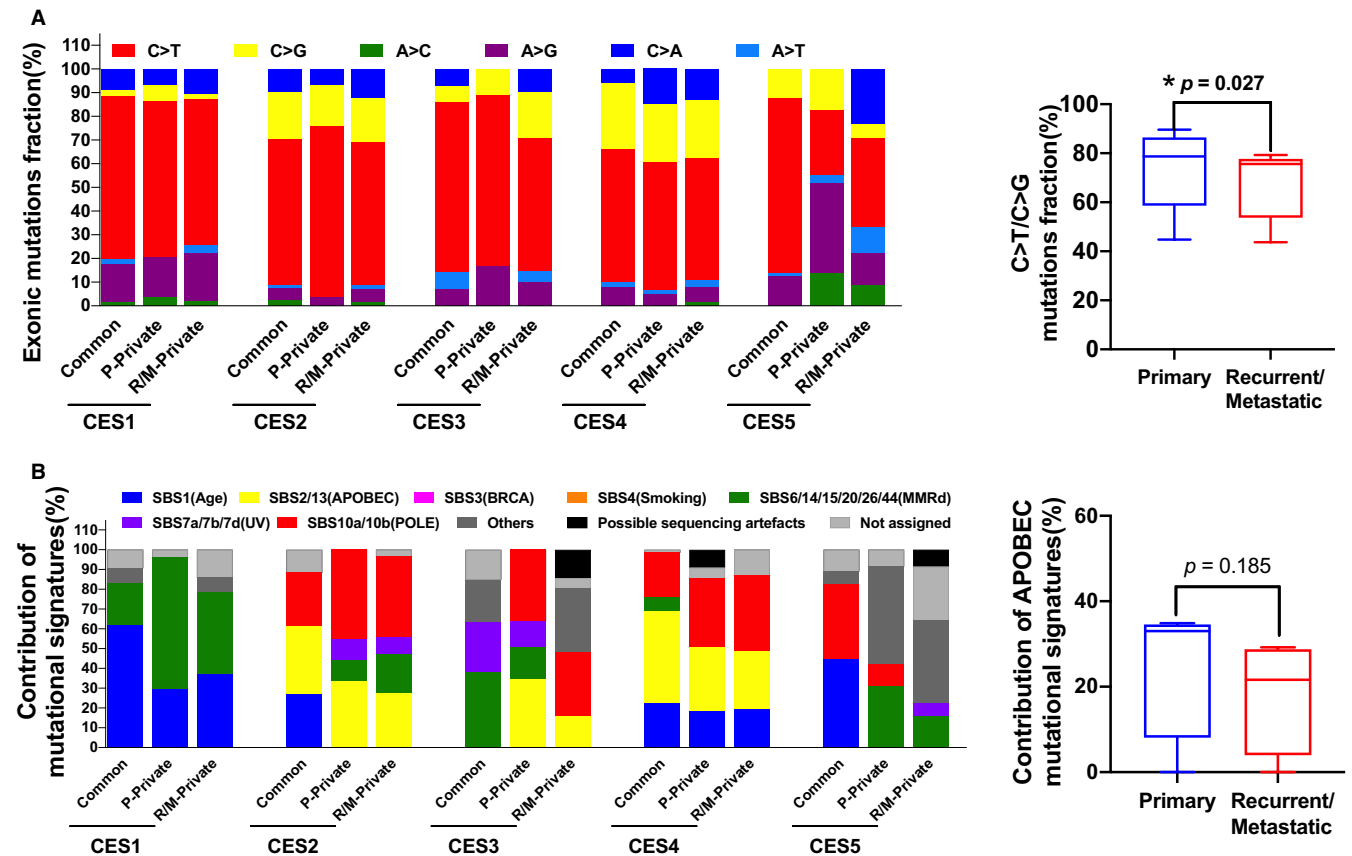
In addition, we observed a number of CES-R/M-specific mutations in epigenetic regulators such as *ARID1A*, *ARID1B*, *CTCF*, *KMT2C*, *KMT2D*, *PBRM1*, and *SETBP1*, suggesting that metastatic progression of CES may be associated with loss of epigenetic regulation. Among these epigenetic modifiers, *ARID1A* mutations may serve as therapeutic targets using synthetic lethality. For example, the inhibition of *EZH2*—the synthetic lethal partner of *ARID1A*—caused tumor regression in *ARID1A*-mutated ovarian cancer cells<sup>38</sup> and associated with hepatocellular carcinoma metastases.<sup>39</sup>

### 3.2 | Mutation signatures

Based on mutation spectra, C>T and C>G substitutions were the most common types of mutations in cervical cancers (in common, P-private, and R/M-private) (Figure 2A). We further performed further mutation signature analysis to deconvolute the mutation profiles of individual cancer genomes into a set of relative weights or contributions of distinct signatures. We used 65 SBS (single base substitution) mutation signatures as available in the database (COSMIC ver. 3 mutation signatures).<sup>12</sup> Seven major signatures are shown for their relative contribution (Figure 2B). Except for CES1, SBS2-SBS13 representing APOBEC overactivity yielding both C>T transitions and C>G transversions in the context of dipyrimidines were prevalent across the cases as a previous paper reported.<sup>21</sup> A significant reduction in the levels of C>T transitions and C>G transversions (Figure 2A,  $P$ -value = .027) associated with a relative deficit of APOBEC-associated signature (SBS2-SBS13) was found in three CES-R/Ms, compared to the matched CES-Ps (Figure 2B). It has been reported that *APOBEC* mRNA levels can serve as a proxy to APOBEC activity level as well as the abundance of the associated mutations.<sup>24,25</sup> Our transcriptomic analysis also demonstrated that the *APOBEC3A* mRNA levels—but not other APOBEC family members—were significantly higher in CES-P than in CES-R/M ( $P$ -value = .047; Table S3, Figure S1), suggesting that decreased *APOBEC3A* expression in CES-R/M may be responsible for the reduced levels of SBS2-SBS13. In the case of CES1, Sig6—representing mutations associated with a deficiency of DNA mismatch repair (MMRd)—was dominant along with age-related mutation signatures, suggesting that MMRd may be responsible for hypermutations along with a predominance of indels in CES1.

### 3.3 | Transcriptomic analysis

The base quality control was performed for transcriptomic sequencing data (Extended Data 3 and 4). To identify functional changes associated with the metastatic progression of CES, we conducted GSEA analysis using functional gene sets available in the MSigDB database (<http://software.broadinstitute.org/>



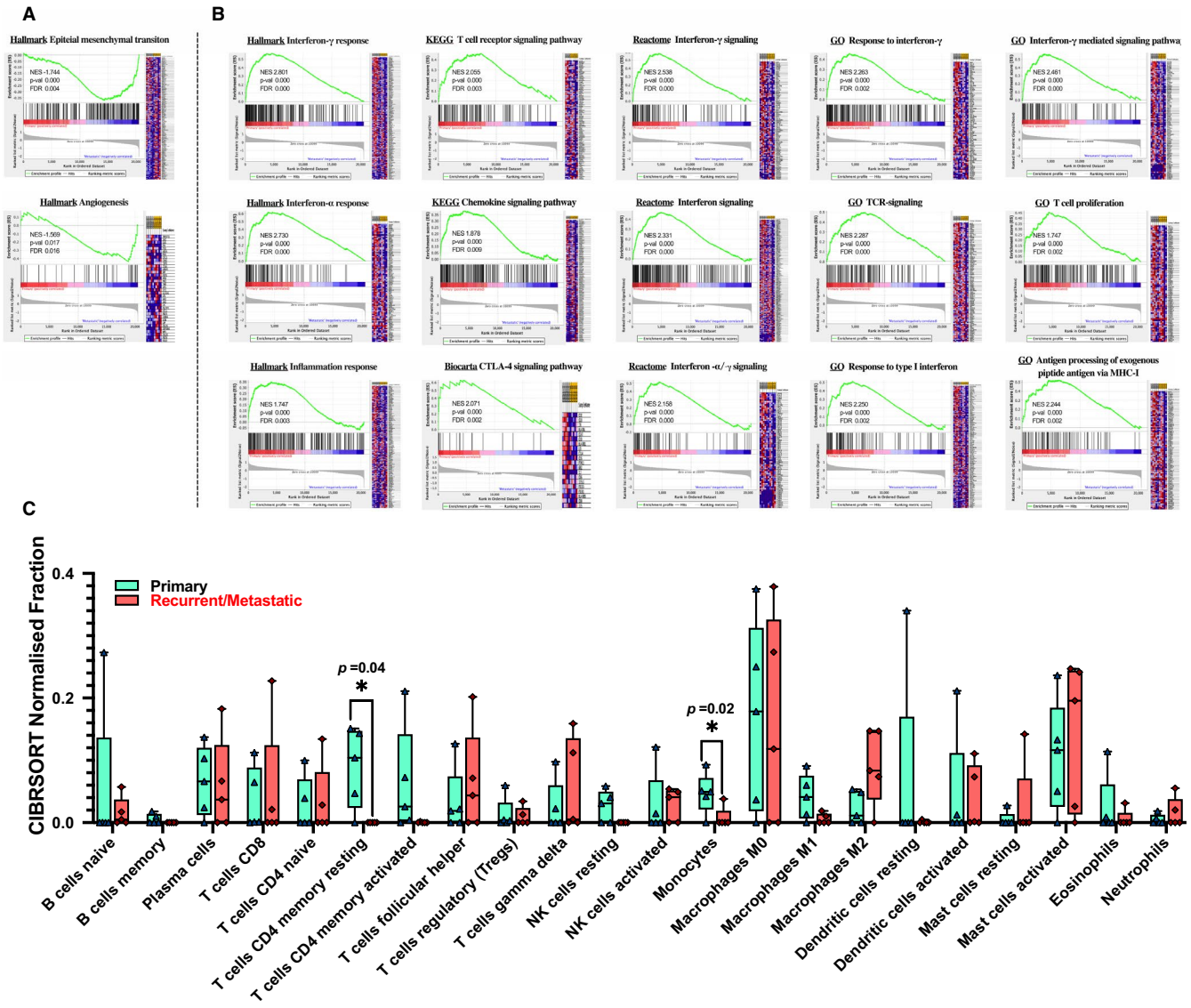
**FIGURE 2** Mutation signatures between primary and matched recurrent/metastatic genomes. A, The six mutation spectra are shown for three regional categories. The frequencies of C>T/C>G substitutions are compared between CES-P and CES-R/M. B, Relative proportion of eight representative mutation signatures are shown for three regional categories. For the top 16 most frequent mutation signatures, those with similar causal relationships are combined (eg, SBS2 and SBS13) and the remaining, less-frequent mutation signatures are collected as others. For APOBEC-related mutation signatures (SBS2 and SBS13), the relative contributions are compared between CES-P and CES-R/M

gsea/msigdb/). For hallmark gene sets, top significantly enriched gene sets for genes upregulated in CES-R/M compared to CES-P are those representing epithelial-to-mesenchymal transitions (EMT) and angiogenesis (Figure 3A; FDR < 0.01). In addition, immune- and inflammation-related gene sets were upregulated in CES-P compared to CES-R/M (eg T cell receptor signaling pathway and chemokine signaling, CTLA-4 signaling pathway and interferon signaling) (Figure 3B and Tables S4 and S5). To further estimate the level of individual tumor-infiltrating immune cells in CES, we performed CIBERSORT. As shown in Figure 3C, the relative abundance of immune cells is variable across samples. We observed that resting memory CD4 T cells as well as monocytes were enriched in CES-P in comparison with CES-R/M (Figure 3C,  $P$ -value < .01).

### 3.4 | HPV integration in the human genome

It has been reported that the persistent presence of HPV DNA in distant metastases of cervical carcinoma could play

a crucial role in the acquisition and maintenance of malignancy-related features.<sup>40</sup> New cases of ovarian neoplasia confirmed the cervical origin of the ovarian metastasis which was supported by the same HPV genotypes and their integration sites in the paired cervical and ovarian tumors.<sup>41</sup> To investigate whether HPV integration contributes to cervical cancer recurrence and metastasis, we used the HIVID technique to identify the integration sites of HPV for four pairs of CES-P and CES-R/M (CES2-CES4). A total of 633 HPV integration breakpoints were detected in four CES-P and 614 integration events in CES-R/M. An average of 214 and 234 high confidence breakpoint counts were detected in CES-P and CES-R/M respectively (Figure 4A). HPV integration breakpoints were highly enriched in the intergenic and intronic regions in all samples, with a comparable integration frequency in each region between CES-P and CES-R/M (Figure 4B). The observed breakpoints were scattered throughout the entire HPV genome, but were enriched in E1 in both CES-P and CES-R/M genomes (Figure 4C)—consistent with previous reports.<sup>42</sup> Persistent HPV infection that leads to carcinoma is characterized by overexpression



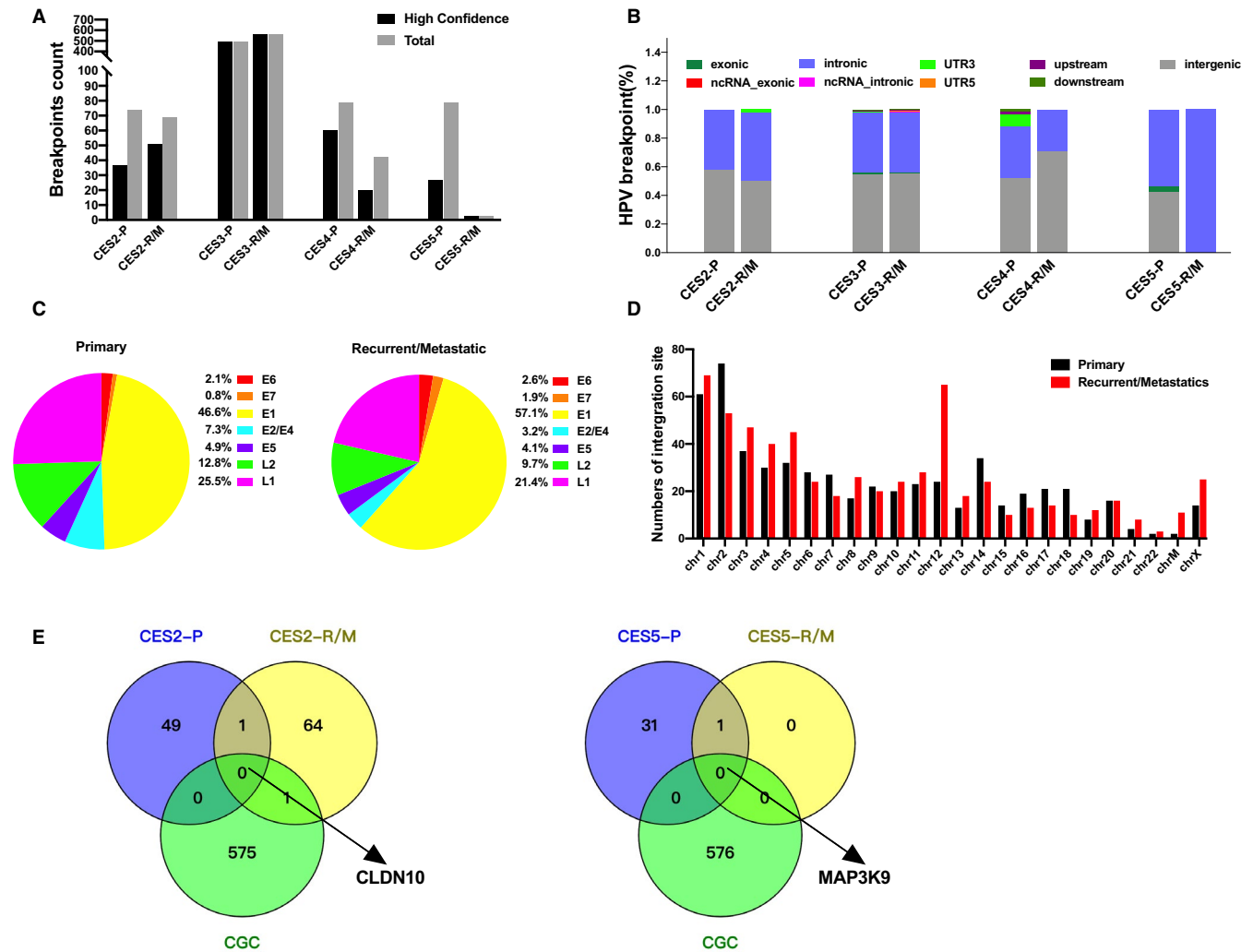
**FIGURE 3** Transcriptome-based functional gene set and immunological analysis of CES-P and CES-R/M. A, Enrichment plots of epithelial-to-mesenchymal transition and angiogenesis hallmark genetic sets whose gene members are relatively upregulated in CES-R/M compared to CES-P are shown. B, Immune-related gene and inflammation-related sets whose gene members are relatively upregulated in CES-P compared to CES-R/M are shown. Gene sets are as available in MSigDB, Hallmark (H), KEGG, Biocarta, Reactome (all in c2), and Gene Ontology (GO; c5 categories). C, Immune cell abundance ( $x$ -axis) estimated by the CIBERSORT algorithm for 22 immune cell subsets are shown. \* $P$ -value < .05

of E6/E7 and loss of E2 functions to induce cellular proliferation and viral immortalization.<sup>42,43</sup> We also detected a higher number of breakpoints in E6, E7, and E1 and a lower number of breakpoints in E2/E4, and L1/L2 in CES-R/M compared to those in CES-P (Figure 4C). Regarding viral integration sites in host genomes, 10.5% (68/673) of integration sites discovered in CES-R/M genomes were localized on chromosome 12, compared to 3.7% (27/722) in CES-P (Figure 4D and Table S6). HPV integration in intronic regions of *CLDN10* and *MAP3K9* with roles in the tight junction and MAPK signaling pathways, respectively, were also observed in both CES-P and CES-R/M samples (CES1 and CES5, respectively) (Figure 4E).

### 3.5 | Copy number profiling

It has been proposed that well-recognized SCNAs (somatic copy number alterations) in CES, including 3q gains, represent markers of stepwise disease progression and adverse clinical outcomes.<sup>44</sup> Genome-wide SCNA profiles of five CES pairs are shown in Figure 5A. A lack of SCNA was observed for CES1, consistent with a relative deficit of SCNA for hypermutated genomes such as MSI-H genomes. The extent of concordance of SCNAs between CES-P and CES-R/M were variable across the cases. For example, CES2 harbors three SCNAs (loss of 6p/q and 13p along with gain of Xp/q) as common SCNAs that emerged before dissemination of metastatic



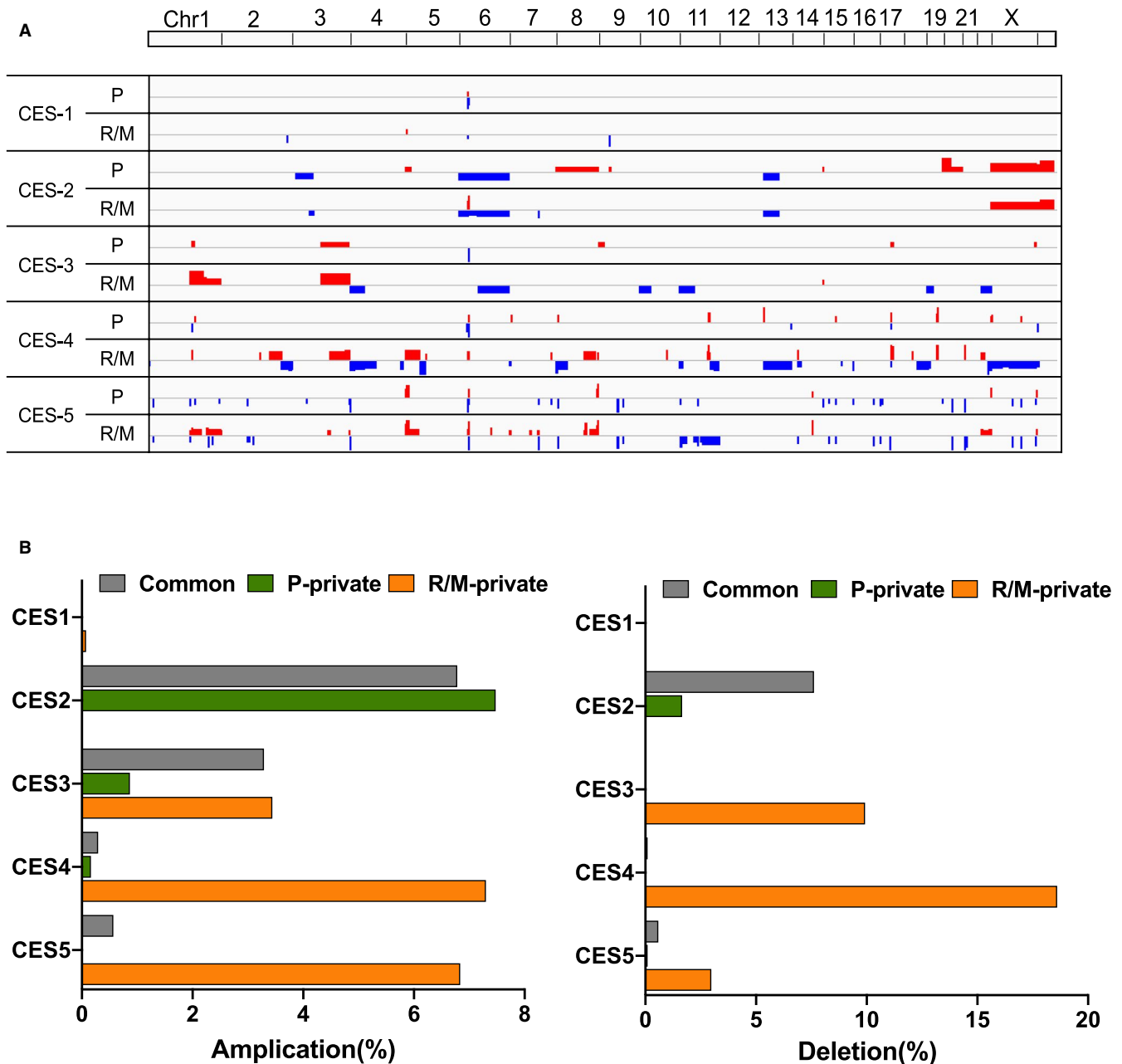


**FIGURE 4** HPV integration on CES genomes. A, The abundance of HPV integration is shown in two measures, total and high-confidence integrations. B, Ratio of HPV breakpoints in variants classified according to genomic region with untranslated region promoter, enhancer, exon, intron, and intergenic region included in primary and recurrent/metastatic types. C, Pie chart of breakpoints occurring in the annotated regions of the HPV genome. D, Genome-wide distribution of HPV integration breakpoints across the human genome. E, Venn diagrams show the number of HPV integration sites common or private to the CES-P and CES-R/M. Genes corresponding to the HPV integration sites are shown with gene symbols

tumors. For CES2, loss of 3p along with gain of 8p/q and 20p/q were CES-P specific and not observed in CES-R/M, suggesting that the CES-R/M cases did not acquire additional SCNAs after dissemination compared to parental, primary genomes. To the contrary, for CES3, a single SCNA (3q gain) was common while a majority of lesion-private SCNAs were observed as CES-R/M specific (e.g. gain of 1q and loss of 4p, 6q, 10p, 11p, 19p, and 22q). The dominance of CES-R/M specific SCNAs was also observed for CES4 and CES5 (Figure 5B). Thus, the concordance level between CES-P and CES-R/M was variable, but the dominance of CES-R/M-specific SCNAs was often observed, suggesting that aneuploidy level is elevated in CES metastatic genomes, probably with genomic instability. The number of genes and cancer-related genes belonging to common, CES-P-private, and CES-R/M-private SCNAs in each case are available in Table S7.

### 3.6 | Predictive biomarkers based on gene expression profiles

To identify expression-based biomarkers with potential clinical implications, we first identified 36 differentially expressed genes (DEGs) using the DESeq2 algorithm (ie, > 2-fold changes and  $P$ -values < .05, Table S3) between CES-P and CES-R/M. DEGs include *APOBEC3A*, *CXCL9*, *MUC21*, and *SPEG*, which were significantly downregulated in CES-R/M, while the expression of *POSTN* was significantly upregulated in CES-R/M compared to CES-P (Figure 6A). We then examined whether the selected DEGs have prognostic implications using GEPIA online tools (<http://gepia.cancer-pku.cn/>). For clinical evaluation, we used the expression and patient survival information for cervical cancer available from

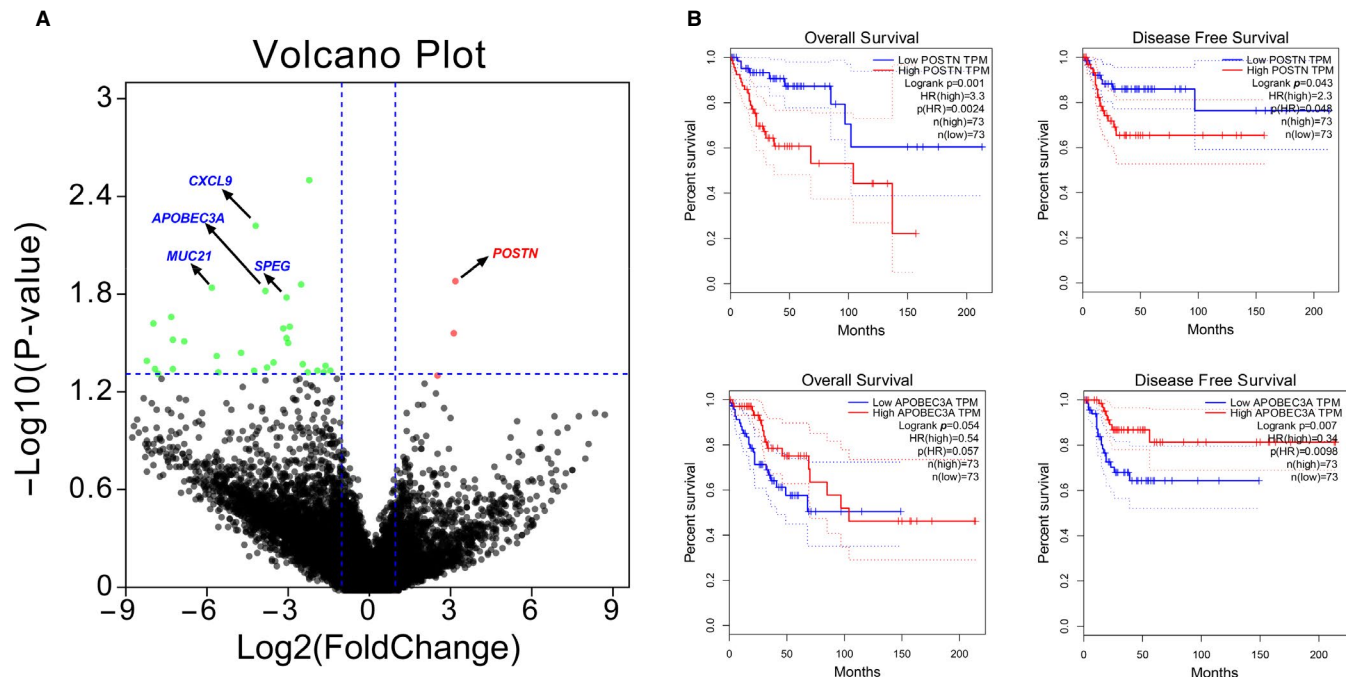


**FIGURE 5** Copy number profiling. A, Genome-wide heatmap representing the somatic copy number alterations (SCNAs) in the discovery cohort is shown. Red and blue represent chromosomal amplifications and deletions, respectively. B, Chromosomal segments are further distinguished into three regional categories of common, P-, and R/M-private SCNAs. The genomic percentages of three categories of SCNAs are shown for five cases separately for chromosomal amplifications and deletions (left and right, respectively)

the TCGA consortium. The analysis revealed that the transcriptional downregulation of *APOBEC3A* expression was associated with poor DFS and OS in human cervical cancers ( $P$ -value = .007 and 0.054, log-rank test). Transcriptional upregulation of *POSTN* expression was significantly associated with poor OS rate and DFS rate ( $P$ -value = .001 and .043, respectively, log-rank test, Figure 6B), indicating that *POSTN* may serve as a prognostic biomarker for metastatic cervical cancer.

## 4 | DISCUSSION

In this study, we report a comprehensive genomic and transcriptomic landscape of five cervical cancer patients by examining the matched pairs of primary and recurrent/metastatic specimens. Despite the availability of a large public database for primary cervical cancer, genomic datasets for metastatic cervical cancer have been lacking, given the inherent difficulties of performing biopsies or secondary debulking surgeries



**FIGURE 6** Relationship of overall survival rate and disease-free survival rate with candidate potential biomarkers. A, Differentially expressed genes based on the DESeq2 algorithm driven by volcano plot; red dots represent upregulated genes, green dots represent downregulated genes with a threshold  $\log_2$  (fold change) > 1 and  $P$ -value < .05. The decreased expression levels of *APOBEC3A*, *CXCL9*, *MUC21*, and *SPEG* and increased *POSTN* expression level marked the top genes when the threshold cutoff was set at  $\log_2$  (fold change) > 3 and  $P$ -value < .02. B, Overall survival and disease-free survival curves of potential DEGs were assessed by Gene Expression Profiling Interactive Analysis (GEPIA) based on primary cervical cancer profiles from The Cancer Genome Atlas database (TCGA); high/low gene expression groups were cutoff at the quartile,  $P$ -value < .05

when recurrence occurs. To the best of our knowledge, this is the first study to characterize somatic alterations, copy number variations, gene expression changes, immune profiles, and the genetic variability of HPV types in primary cervical cancers and their matching recurrent/metastatic cancers. The metastasis-specific genomic and transcriptomic features may uncover molecular insights into the mechanisms likely driving the mutational, transcriptomic, and viral landscape of cervical cancer recurrence.

APOBEC cytidine deaminases plays a crucial role in innate immune response and viral restriction by removing the amino group from a cytosine, converting it to uracil. Prolonged APOBEC activation during HPV infection can induce somatic mutations in the host genomes, thereby contributing to the initiation and progression of the disease. However, it remains unclear whether APOBEC enzymes cause a significant fraction of the driver mutations or genome rearrangements in cervical cancer recurrence/metastases. Recent reports revealed that the overactivity of APOBEC was associated with the generation of late-occurring subclonal mutations in the evolution of primary tumors, serving as source for intratumoral heterogeneity.<sup>45</sup> In our study, APOBEC mutagenesis was the most prominent signature in the majority of samples, consistent with a recent report.<sup>46</sup> Moreover, a substantial

depletion of C>T transitions and C>G transversions representing APOBEC mutagenesis (SBS 2/13) accompanying a decreased expression of APOBEC3 genes, was observed in metastatic samples when compared to primary samples. Our results are supported by a recent study that reported APOBEC signatures in stock cervical cancer cell lines but no discernible continuing activity in their daughter lineages.<sup>47</sup> In addition, the study reported that tumors with enriched interferon signaling pathways (eg, IFN- $\gamma$ ) were associated with high APOBEC3 expression and that increased expression of *STAT1*, a downstream effector of IFN- $\gamma$ , was accompanied by an increased expression of *APOBEC3A*; these results are consistent with our data (Figures 2A and 3B).

In our study, one patient (CES1) lacked the APOBEC signature in both primary and recurrent/metastatic tumors, instead showing the prevalence of mutation signatures associated with MMRd. This patient also showed concurrent colon cancer with an *MLH1* missense mutation, implying Lynch syndrome.<sup>48</sup> We previously showed that 2.3% of CES cases show MSI-H genotypes,<sup>49</sup> the proportion of which is relatively low compared to colorectal and uterine cancers. Our results suggest that concurrent cancers with clinical implications of Lynch syndrome—including CES—should be considered for the presence of hypermutation.

In cervical cancer, the inflammation induced by HPV leads to a complex immune milieu with various types of tumor-infiltrating immune cells that play an important role in tumorigenesis. Recently, Chen et al revealed that immune cell composition can be substantially different between tissues with inflammation and cancer. For example, resting memory CD4<sup>+</sup> T cells, activated mast cells, monocytes, and M1 macrophages are more abundant in inflammatory stages, while CD8<sup>+</sup> T cells, follicular helper T cells, and M0 macrophages are more abundant in cancer tissues.<sup>50</sup> Our CIBERSORT analysis showed that monocyte fractions were relatively lower in metastatic samples along with a deficit of immune-related and inflammation-related pathway activation. This finding is in accordance with previous research reporting that classical monocytes are recruited at higher rates by inflamed tissues and are able to attract other immune cells by secreting cytokines and antimicrobial factors.<sup>51,52</sup> Lack of memory CD4<sup>+</sup> T cells in metastatic samples is associated with impaired IFN-related signaling, given their close correlation with IFN- $\gamma$  production and inflammation activity. Moreover, our GSEA pathway enrichment data reveals that recurrent/metastatic cervical cancer samples exhibit impaired inflammation, chemokines, immune activity, interferon, T-cell signaling, and antigen processing signaling pathway, all of which are regarded as characteristics of “cold/non-inflamed” tumors with poor response to immunotherapy.<sup>53</sup>

Viral integration into intergenic and chromosome fragile sites frequently occurs in cervical cancers, resulting in the impairment of E2 and increased expression of E6 and E7.<sup>43,54</sup> Relatively enriched breakpoints in E6 and E7 and the disruption in E1 genes of CES-M were observed in comparison with CES-P. Although no recurrent high confidence HPV integration sites were observed among the metastatic samples, some of the singleton events were observed in genes with potential tumorigenic effects. For instance, in CES2, HPV integration was observed in *CLDN10*,<sup>55</sup> the loss-of-function of which may lead to the loss of cellular adherens junctions in EMT pathways,<sup>56</sup> contributing to tumor metastasis.<sup>55</sup> HPV integration into the *MAP3K9* gene was also observed in CES5.

Tumor mutational burden (TMB) has been proposed as a biomarker to predict the therapeutic response to immunotherapies.<sup>57,58</sup> In our study, significantly higher TMB was found in CES-R/M compared with CES-P. Higher mutation burdens in metastatic tumors compared to primary tumors were also reported in a large-scale genomic study on metastatic breast cancer.<sup>59</sup> Higher MAFs of mutations in CES-R/M genomes also suggest the presence of clonal selection events that are relatively unique to CES-R/M genomes. We did not observe any significant change in nonsynonymous mutations, which was reported to be an improved objective response rate factor between CES-P and CES-R/M according to our study.

Common mutations found to be altered in both the primary lesions and their metastases represent early cancer drivers, while the lesion-specific or private mutations are those acquired after metastatic dissemination. In addition, metastatic-specific mutations may serve as potential markers involved in metastatic progression. Wingo et al reported 20% of cervical cancers harbor somatically-acquired *STK11* mutations<sup>35</sup> and can suppress the tumor immune surveillance response by directly impacting NF- $\kappa$ B activity and other pathways in lung adenocarcinoma.<sup>60</sup> *STK11* mutation confers resistance to immune checkpoint inhibitors.<sup>61</sup> In our study, *STK11* harbored recurrent mutations for two patients along with recurrent chromosomal losses in two metastatic samples, suggesting that *STK11* alterations confer the ability of cervical cancers to metastasize.

Using transcriptome analysis, we found two potential predictive biomarkers (*APOBEC3A* and *POSTN*) for recurrent/metastatic cervical cancer. We determined that the transcriptional downregulation of *APOBEC3A* dictated a poor clinical outcome including abbreviated disease-free survival. The association of APOBEC expression and activity with disease progression remains contradictory, especially with regard to HPV infection. For example, higher expression levels of *APOBEC3A* can be induced by HPV16 infection in oropharyngeal cancer,<sup>62</sup> while APOBEC3A protein levels were downregulated during invasive progression in HPV-negative but not HPV-positive penile squamous cell carcinomas, suggesting that HPV-associated APOBEC3A activity is variable and context-dependent.<sup>63</sup> Given that decreased *APOBEC3A* levels were observed as CES-R/M group-specific features and as unfavorable clinical outcome-predicting features, it is assumed the APOBEC activity is repressed in metastatic tumors and may predict the disease progression and metastasis in CES with HPV infection. The second potential predictive biomarker we identified was *POSTN*. It is thought to be a crucial player in cancer development and metastasis,<sup>64</sup> as well as chemotherapy resistance<sup>65</sup> across different tumor types. High *POSTN* levels in tumors are associated with aggressive tumor behavior as well as advanced stage and/or poor prognosis in colorectal cancer.<sup>66</sup> All of these reports support our finding that upregulated expression of *POSTN* is associated with poor overall survival and disease-free survival rates in TCGA primary cervical cancer, identifying upregulated *POSTN* level as a novel potential predictive biomarker for cervical cancer recurrence/metastasis as studies in many other cancers.<sup>66-68</sup> An extended cohort will be required to validate the potential biomarkers that we identified in our analysis.

This study has several limitations. First, our study had a small sample size precluding definite conclusions which normally require a larger sample size to ensure a representative distribution of the population and to be considered representative of groups of people by statistical tests. Second, the use



of FFPE tissue archives instead of fresh-frozen (FF) tissues may be associated with known artifacts with impacts on the sequencing and the fidelity of expression profiles as previously proposed with limited quantity of DNA from FFPE samples.<sup>69-71</sup> Lastly, the validation of targeted RNA-seq by reverse transcriptase polymerase chain reaction (RT-PCR) and HPV integration sites by fluorescent in situ hybridization (FISH) or PCR should be considered in the further study.

In summary, due to difficulties in the acquisition of matched tissue biopsies from cervical cancer patients, hopefully, our results can be of help to pathologists and clinicians alike to identify both genome- and transcriptome-level features and markers associated with cervical cancer recurrence and metastasis, offering new potential biomarker targets and a basis for further validation studies.

## ACKNOWLEDGMENTS

We thank Editage (www.editage.com) for English language editing. This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B04932862) and Ministry of Science and ICT (2019R1A5A2027588).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Conceptualization, SYH, TMK, and YJC; Data curation, JLL, JYH, JEL, JSY, KRK, TMK, and YJC; Formal analysis, JLL and JEL; Funding acquisition, TMK and YJC; Investigation, JLL, JYH, and JEL; Methodology, JLL, DR, and JSY; Project administration, SYH, TMK, and YJC; Resources, JEL, KRK, and SYH; Software, JLL, JSY, and DR; Supervision, SYH, TMK, and YJC; Writing – original draft, JLL; Writing – review & editing, TMK and YJC

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## ORCID

Jing Jing Liu  <https://orcid.org/0000-0002-3828-1608>

## REFERENCES

- Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science*. 2011;331:1559-1564.
- Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. *Nat Med*. 2006;12:895-904.
- Elit L, Fyles AW, Devries MC, Oliver TK, Fung-Kee-Fung M. Gynecology Cancer Disease Site G. Follow-up for women after treatment for cervical cancer: a systematic review. *Gynecol Oncol*. 2009;114:528-535.
- Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49:1374-1403.
- Swanton C. Intratumor heterogeneity: evolution through space and time. *Cancer Res*. 2012;72:4875-4882.
- Burrell RA, Swanton C. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol*. 2014;8:1095-1111.
- Tomczak K, Czerwinska P, Wiznerowicz M. The cancer genome atlas (tcga): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;19:A68-77.
- Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. *Nature*. 2013;501:355-364.
- Cyll K, Ersvar E, Vlatkovic L, et al. Tumour heterogeneity poses a significant challenge to cancer biomarker research. *Br J Cancer*. 2017;117:367-375.
- Yates LR, Knappskog S, Wedge D, et al. Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell*. 2017;32(169-184):e167.
- Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
- Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453-457.
- Vitiello GA, Bowler TG, Liu M, et al. Differential immune profiles distinguish the mutational subtypes of gastrointestinal stromal tumor. *J Clin Invest*. 2019;129:1863-1877.
- Lace MJ, Anson JR, Klusmann JP, et al. Human papillomavirus type 16 (hpv-16) genomes integrated in head and neck cancers and in hpv-16-immortalized human keratinocyte clones express chimeric virus-cell mRNAs similar to those found in cervical cancers. *J Virol*. 2011;85:1645-1654.
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
- DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491-498.
- Kim H, Lee AJ, Lee J, Chun H, Ju YS, Hong D. FIREVAT: finding reliable variants without artifacts in human cancer samples using etiologically relevant mutational signatures. *Genome Med*. 2019;11:81.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
- Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578:94-101.
- Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016;17:31.
- Burns MB, Lackey L, Carpenter MA, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. 2013;494:366-370.
- Magi A, Tattini L, Cifola I, et al. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. *Genome Biol*. 2013;14:R120.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.

24. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511-515.
25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
26. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243-259.
27. Li W, Zeng XI, Lee NP, et al. HIVID: an efficient method to detect HBV integration using low coverage sequencing. *Genomics*. 2013;102:338-344.
28. Liu Y, Zhang C, Gao W, et al. Genome-wide profiling of the human papillomavirus DNA integration in cervical intraepithelial neoplasia and normal cervical epithelium by HPV capture technology. *Sci Rep*. 2016;6:35427.
29. Zeitouni B, Boeva V, Janoueix-Lerosey I, et al. SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics*. 2010;26:1895-1896.
30. Wang J, Mullighan CG, Easton J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat Methods*. 2011;8:652-654.
31. Liu Y, Lu Z, Xu R, Ke Y. Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology. *Oncotarget*. 2016;7:5852-5864.
32. Samuels DC, Han L, Li J, et al. Finding the lost treasures in exome sequencing data. *Trends Genet*. 2013;29:593-599.
33. Parfenov M, Peadarallu CS, Gehlenborg N, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A*. 2014;111:15544-15549.
34. Marusyk A, Tabassum DP, Altmann PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*. 2014;514:54-58.
35. Wingo SN, Gallardo TD, Akbay EA, et al. Somatic LKB1 mutations promote cervical cancer progression. *PLoS One*. 2009;4:e5137.
36. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517:576-582.
37. Brennan K, Koenig JL, Gentles AJ, Sunwoo JB, Gevaert O. Identification of an atypical etiological head and neck squamous carcinoma subtype featuring the CpG island methylator phenotype. *EBioMedicine*. 2017;17:223-236.
38. Bitler BG, Aird KM, Garipov A, et al. Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nat Med*. 2015;21:231-238.
39. He F, Li J, Xu J, et al. Decreased expression of ARID1A associates with poor prognosis and promotes metastases of hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2015;34:47.
40. Ikenberg H, Teufel G, Schmitt B, Kommos F, Stanimirovic B, Pfeleiderer A. Human papillomavirus DNA in distant metastases of cervical cancer. *Gynecol Oncol*. 1993;48:56-60.
41. Arfi A, Hequet D, Bataillon G, et al. HPV DNA integration site as proof of the origin of ovarian metastasis from endocervical adenocarcinoma: three case reports. *BMC Cancer*. 2019;19:375.
42. Hu Z, Zhu DA, Wang W, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet*. 2015;47:158-163.
43. Romanczuk H, Howley PM. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc Natl Acad Sci U S A*. 1992;89:3159-3163.
44. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500:415-421.
45. Roper N, Gao S, Maity TK, et al. APOBEC mutagenesis and copy-number alterations are drivers of proteogenomic tumor evolution and heterogeneity in metastatic thoracic tumors. *Cell Rep*. 2019;26(2651-2666):e2656.
46. Cancer Genome Atlas Research N, Albert Einstein College of M, Analytical Biological S, Barretos Cancer H, Baylor College of M, Beckman Research Institute of City of H, et al. Integrated genomic and molecular characterization of cervical cancer. *Nature*. 2017;543:378-384.
47. Petljak M, Alexandrov LB, Brummel JS, et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. *Cell*. 2019;176(1282-1294):e1220.
48. Antill YC, Dowty JG, Win AK, et al. Lynch syndrome and cervical cancer. *Int J Cancer*. 2015;137:2757-2761.
49. Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun*. 2017;8:15180.
50. Chen X, Xu C, Hong S, et al. Immune cell types and secreted factors contributing to inflammation-to-cancer transition and immune therapy response. *Cell Rep*. 2019;26(1965-1977):e1964.
51. Serbina NV, Pamer EG. Monocyte emigration from bone marrow during bacterial infection requires signals mediated by chemokine receptor CCR2. *Nat Immunol*. 2006;7:311-317.
52. Jakubczik CV, Randolph GJ, Henson PM. Monocyte differentiation and antigen-presenting functions. *Nat Rev Immunol*. 2017;17:349-362.
53. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol*. 2013;14:1014-1022.
54. Akagi K, Li J, Broutian TR, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res*. 2014;24:185-199.
55. Zhou Y, Xiang J, Bhandari A, et al. CLDN10 is associated with papillary thyroid cancer progression. *J Cancer*. 2018;9:4712-4717.
56. Wilson MR, Reske JJ, Holladay J, et al. ARID1A and PI3-kinase pathway mutations in the endometrium drive epithelial transdifferentiation and collective invasion. *Nat Commun*. 2019;10:3554.
57. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348:124-128.
58. Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. 2014;371:2189-2199.
59. Wentzensen N, Vinokurova S, von Knebel DM. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res*. 2004;64:3878-3884.
60. Schabath MB, Welsh EA, Fulp WJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene*. 2016;35:3209-3216.
61. Jure-Kunkel M, Wu S, Xiao F, et al. Somatic STK11/LKB1 mutations to confer resistance to immune checkpoint inhibitors

- as monotherapy or in combination in advanced NSCLC. *J Clin Oncol*. 2018;36:3028.
62. Kondo S, Wakae K, Wakisaka N, et al. APOBEC3A associates with human papillomavirus genome integration in oropharyngeal cancers. *Oncogene*. 2017;36:1687-1697.
63. Heller M, Prigge ES, Kaczorowski A, von Knebel DM, Hohenfellner M, Duensing S. APOBEC3A expression in penile squamous cell carcinoma. *Pathobiology*. 2018;85:169-178.
64. Kyutoku M, Taniyama Y, Katsuragi N, et al. Role of periostin in cancer progression and metastasis: inhibition of breast cancer progression and metastasis by anti-periostin antibody in a murine model. *Int J Mol Med*. 2011;28:181-186.
65. Sung P-L, Jan Y-H, Lin S-C, et al. Periostin in tumor micro-environment is associated with poor prognosis and platinum resistance in epithelial ovarian carcinoma. *Oncotarget*. 2016;7:4036-4047.
66. Oh HJ, Bae JM, Wen XY, Cho NY, Kim JH, Kang GH. Overexpression of POSTN in tumor stroma is a poor prognostic indicator of colorectal cancer. *J Pathol Transl Med*. 2017;51:306-313.
67. Nakazawa Y, Taniyama Y, Sanada F, et al. Periostin blockade overcomes chemoresistance via restricting the expansion of mesenchymal tumor subpopulations in breast cancer. *Sci Rep*. 2018;8:4013.
68. Park SY, Piao Y, Jeong KJ, Dong J, de Groot JF. Periostin (POSTN) regulates tumor resistance to antiangiogenic therapy in glioma models. *Mol Cancer Ther*. 2016;15:2187-2197.
69. Liu YU, Noon AP, Aguiar Cabeza E, et al. Next-generation RNA sequencing of archival formalin-fixed paraffin-embedded urothelial bladder cancer. *Eur Urol*. 2014;66:982-986.
70. Li J, Fu C, Speed TP, Wang W, Symmans WF. Accurate RNA sequencing from formalin-fixed cancer tissue to represent high-quality transcriptome from frozen tissue. *JCO Precis Oncol*. 2018;2:1-9.
71. Kissel HD, Paulson TG, Liu K, et al. Feasibility of RNA and DNA extraction from fresh pipelle and archival endometrial tissues for use in gene expression and SNP arrays. *Obstet Gynecol Int*. 2013;2013:576842.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Liu JJ, Ho JY, Lee JE, et al. Genomic, transcriptomic, and viral integration profiles associated with recurrent/metastatic progression in high-risk human papillomavirus cervical carcinomas. *Cancer Med*. 2020;9:8243–8257. <https://doi.org/10.1002/cam4.3426>