Journal of Translational Medicine

RESEARCH

Open Access

Dual-dimensional profiling of host genomic variations and HPV integration in PD-L1-stratified cervical cancer via Oxford Nanopore Technology

Ruijiao Lu[1†], Jie Zhang[2†], Xiumin Ma[1*] and Yangchun Feng[3*]

Abstract

Background The integration of human papillomavirus (HPV) DNA into the host genome is a key step in the development of HPV-associated cervical cancer (CC). However, the genomic characteristics of host genomic variations and HPV integration within the context of programmed death-ligand 1 (PD-L1) expression stratification have not been systematically investigated.

Methods Whole-genome sequencing was performed using Oxford Nanopore Technology (ONT) on six samples (three from the high PD-L1 expression group and three from the low PD-L1 expression group). The characteristics of host genomic variations under different PD-L1 expression stratifications were explored, including structural variations (SV), copy number variations (CNV), single nucleotide polymorphisms (SNP), and insertion-deletions (Indel). Subsequently, the distribution features of HPV integration sites were analyzed, different integration types were identified, and pathway analysis was conducted.

Results Whole-genome SV analysis revealed that the total number of SVs and the composition of mutation types were similar between the high and low PD-L1 expression groups, with insertions (INS) and deletions (DEL) predominating in both. These variations were primarily enriched in intergenic regions and introns. In the low PD-L1 expression group, integration events were observed at multiple chromosomal loci, with the most frequent integration occurring in the KLF5 gene region on chromosome 13. No frequently integrated loci were identified in the high PD-L1 expression group. Additionally, four distinct HPV integration breakpoint patterns were preliminarily identified and analyzed.

Conclusion PD-L1 expression stratification did not significantly alter the overall genomic instability of the host. However, differences were observed in the distribution patterns of HPV integration sites. These findings provide

†Ruijiao Lu and Jie Zhang contributed equally to this work.

*Correspondence:
Xiumin Ma
maxiumin1210@sohu.com
Yangchun Feng
paopao1987123@163.com

Full list of author information is available at the end of the article

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

new insights into the genomic heterogeneity of CC under different PD-L1 expression backgrounds and may lay the groundwork for future research exploring stratified immunotherapy based on HPV integration features.

Keywords HPV, PD-L1, Cervical cancer, Oxford Nanopore Technology

Introduction

As a populous country, China bears a heavy burden of CC, accounting for approximately 18 and 17% of global incidence and mortality, respectively. Although China's age-standardized incidence and mortality rates are below the global average, its age-standardized mortality rate reaches approximately twice that of the United States and the United Kingdom [1]. The integration of HPV DNA into the CC cells genome is a crucial step in HPV infection-mediated carcinogenesis. One primary mechanism is the disruption and inactivation of the E2 gene during integration, leading to the overexpression of the oncoproteins E6 and E7. E6 mediates the inactivation of the tumor suppressor p53, blocking apoptotic pathways, while E7 impairs the function of the pRb family, causing cell cycle dysregulation. This results in uncontrolled epithelial cell apoptosis, unlimited proliferation, malignant transformation, and exacerbates genomic instability [2]. A second potential mechanism is the possible regulation of immune evasion pathways. For instance, transcriptionally active integration can induce tumor cells to evade immune surveillance via the PVR-TIGIT pathway, weakening T-cell killing effects and interfering with immune responses [3]. A third mechanism involves the inherent fragility of integration sites, prone to breakage and rearrangement, causing structural and numerical chromosomal abnormalities in the host (e.g., translocations, DEL, amplifications), thereby inducing genomic instability. If integration occurs within functional gene regions, it can also lead to alterations or inactivation of relevant genes, driving tumorigenesis and progression [4, 5]. Studies have reported the identification of numerous HPV integration sites within the human genome, revealing that they are not randomly distributed but tend to cluster in specific chromosomal regions, such as 2q22.3, 3p14.2, 3q28, 8q24.22, 9q22, 13q22.1, 14q24.1, 17p11.1, and 17q23.1–17q23.2 [6]. Many integration events occur near or within known cancer driver genes or tumor-associated genes, including POU5F1B, FHIT, KLF12, LRP1B, LEPREL1, HMGA2, DLG2, SEMA3D, CASC8, KLF5, RAD51B, CASC21, ERBB2, TP63, TEX41, RAP2B, CCDC106, MACROD2, and MYC [7–9].

High expression of PD-L1 in cervical lesion cells is a crucial molecular marker for immune evasion and progression towards malignancy [10]. It also serves as a key detection target for the clinical application of PD-1/PD-L1 immune checkpoint inhibitors in CC [11]. Furthermore, PD-L1 expression has been reported to correlate with poor prognosis and immunotherapy response in CC patients [12]. Systematic analyses of host genomic variation characteristics across different PD-L1 expression stratifications have not yet been reported. A few studies have suggested that HPV DNA integration into the CC cell genome might upregulate PD-L1 expression [13], but the specific impact of HPV integration sites on PD-L1 expression remains unclear. ONT is a groundbreaking single-molecule, real-time DNA/RNA sequencing technology. Compared to previous HPV integration detection techniques, ONT offers significant advantages, such as long-read length and real-time sequencing, which provide substantial benefits in characterizing the integration patterns of the HPV genome with host DNA [14]. This study focused on the characteristics of host genomic variations and HPV integration under PD-L1 expression stratification. By systematically analyzing the distribution patterns and high-frequency variation regions of host SV, CNV, SNP, and Indel in groups stratified by PD-L1 expression levels, we aimed to reveal potential associations between PD-L1 expression differences and host genomic variation patterns. Building upon this, the study further explored HPV integration hotspots related to PD-L1 expression. This provides experimental evidence for elucidating the genomic regulatory mechanisms of PD-L1 expression in HPV-associated CC, while also offering potential targets and theoretical references for developing personalized immunotherapy strategies based on PD-L1 expression and genomic features.

Materials and methods

Sample collection

Fresh frozen tumor tissue specimens were collected from 30 patients diagnosed with CC by histopathological examination and positive for HPV DNA detection. These specimens were obtained from the Specimen Bank of the Institute of Cancer Prevention and Research at the Affiliated Tumor Hospital of Xinjiang Medical University between January 2024 and December 2024. Corresponding formalin-fixed paraffin-embedded specimens were acquired from the Clinical Pathology Center, along with relevant clinicopathological data from the patients. All patients had not received any other treatments, such as radiotherapy or chemotherapy, prior to surgery. This study was approved by the Ethics Committee of the Affiliated Tumor Hospital of Xinjiang Medical University (G-2023001).

Sample screening and quality control

This study initially included fresh frozen tumor tissue specimens from 30 histopathologically confirmed, HPV-positive CC patients. During the pre-processing stage, two specimens (Sample IDs: S9, S10) were excluded due to insufficient tissue quantity to meet subsequent experimental requirements. The remaining 28 specimens underwent Western Blot (WB) protein extraction and internal reference protein detection. Among these, seven specimens (Sample IDs: S1, S2, S3, S4, S11, S21, S22) were excluded because the internal reference protein expression was undetectable and did not meet the data analysis standards. Ultimately, 21 samples met the quality control standards and proceeded to subsequent experiments.

PD-L1 expression detection

WB

Fresh frozen tumor tissues were lysed in pre-cooled radioimmunoprecipitation assay buffer (PR20001, Proteintech, China) supplemented with protease inhibitors (4693116001, Sigma-Aldrich, USA) on ice for 30 minutes. After adding protein loading buffer (P1040, Solarbio, China), samples were boiled for 15 minutes. Proteins were then separated on 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis gels and transferred onto 0.45 μm polyvinylidene fluoride membranes (ISEQ00010, Millipore, USA). Membranes were blocked with 5% non-fat milk at room temperature for 1 hour. Primary antibody incubation was performed overnight at 4 °C using anti-PD-L1 (1:1000 dilution, mouse monoclonal antibody, 66248-1-Ig, Proteintech, China) and GAPDH (1:10000 dilution, mouse monoclonal antibody, 60004-1-Ig, Proteintech, China). Subsequently, membranes were incubated with horseradish peroxidase-conjugated goat anti-mouse secondary antibody (1:10000 dilution, SA00001-1, Proteintech, China) at room temperature for 45 minutes. Finally, protein bands were visualized using enhanced chemiluminescence reagent (P0018FM, Beyotime, China).

Immunohistochemistry (IHC)

After deparaffinization and rehydration, sections underwent high-temperature and high-pressure antigen retrieval in citrate buffer (pH 6.0). Endogenous peroxidase activity was blocked with 3% hydrogen peroxide (10011218, Sinopharm, China) for 30 minutes, followed by non-specific blocking with 10% goat serum (AR1009, Boster Biological Technology, China) for 30 minutes. Sections were incubated overnight at 4 °C with anti-PD-L1 rabbit monoclonal primary antibody (1:100 dilution, ab213524, Abcam, UK). After washing, sections were incubated with HRP-labeled goat anti-rabbit secondary antibody (1:2000 dilution, ab205718, Abcam, UK) at 37 °C for 45 minutes. Color development was performed using 3,3'-diaminobenzidine (1:20 dilution, DAB-4033, Fuzhou Maixin Biotech, China), followed by counterstaining with hematoxylin (BA4041, Baso Diagnostics, China) for 1 minute. Sections were then dehydrated, cleared, and mounted with an eco-friendly mounting medium. Whole-slide images were acquired using an Olympus C×31 microscope (Olympus, Japan). PD-L1 expression was assessed using the Tumor Proportion Score (TPS), defined as the percentage of viable tumor cells exhibiting partial or complete membrane staining of any intensity. TPS values ranged from 1% to 100%. The entire section (containing ≥100 viable tumor cells) was independently assessed by two pathologists blinded to each other's results. Only tumor cells with unequivocal linear membrane staining were counted (irrespective of staining intensity). The final result was the average of both evaluations.

PD-L1 expression grouping and sequencing sample selection

Based on the quantitative WB results (grayscale ratio of PD-L1/GAPDH protein), the median value of PD-L1 expression within the cohort was first determined. Sample 15, whose expression level corresponded to this median value, served as the reference sample. According to the criteria of WB PD-L1/GAPDH ratio > median value and IHC TPS ≥ 50%, three samples were selected for inclusion in the high PD-L1 expression group. Similarly, according to the criteria of WB PD-L1/GAPDH ratio < median value and IHC TPS ≤ 1%, three samples were selected for inclusion in the low PD-L1 expression group.

Nanopore library construction and sequencing

ONT was performed by Wuhan Ruixing Biotechnology Co., Ltd (Wuhan, China). Genomic DNA was extracted from tissues using the TIANamp Genomic DNA Kit (DP431, Tiangen Biotech, China). DNA concentration and purity were assessed by measuring absorbance at 260 nm/280 nm (A260/A280) using an Ultrafine spectrophotometer (N50 touch, IMPLEN, Germany). DNA integrity was further verified by 1.0% agarose gel electrophoresis. For each sample, DNA fragments larger than 1.5–2 kb were purified using 0.7X Ampure XP beads (self-prepared). Purified DNA samples were used to construct sequencing libraries with the third-generation library preparation kit (VAHTS TGS DNA Library Prep Kit For ONT V2) (TS202, Vazyme Biotech, China). Finally, 300 fmol of adapter-ligated library was loaded onto a PromethION 2 Solo (P2S-02184, Nanopore) sequencing instrument chip (vR10.4.4) for sequencing.

Bioinformatics analysis

Raw ONT sequencing data were processed using SeqKit v2.8.2 for initial data statistics. Data filtering was performed with chopper v0.8.0 using parameters: minimum quality score (Q) of 10 and minimum read length of 50 bp. Filtered data statistics and visualization were generated using NanoPlot v1.42.0. Filtered full-length reads were aligned to the National Center for Biotechnology Information (NCBI) reference genome (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_45/GRCh38.primary_assembly.genome.fa.gz) using minimap2 v2.28-r1209. Alignment results were summarized using samtools v1.7 (command: stats). Subsequent analysis of genomic SV, CNV, and SNP/Indel mutations was conducted using the following software: Sniffles2 v2.6.2 for SV detection, cnvkit v0.9.10 for CNV analysis, DeepVariant v1.9.0 and bcftools v1.2 for SNP/Indel calling, and ANNOVAR v20200607 for variant annotation and filtering. We intersected our called SNVs, indels, and SVs with gnomAD (v4.1.0, https://gnomad.broadinstitute.org/) to remove likely germline variants. Variants were retained for downstream analysis only if they were novel (not present in gnomAD) or had a genome allele frequency (gnomAD41_genome_AF) below 0.1%. Finally, Gene Ontology (GO) enrichment analysis (covering Molecular Function, Cellular Component, and Biological Process) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed. Protein-protein interaction (PPI) network analysis was constructed using the online Search Tool for the Retrieval of Interacting Genes (STRING) database.

Statistical analysis

Statistical analyses were performed using GraphPad Prism software (version 10.0) and IBM SPSS Statistics software (version 27.0). Quantitative analysis of WB results was conducted using AlphaEaseFC software. Comparisons between groups for continuous variables (normally distributed) were conducted using the independent samples t-test. Comparisons between groups for continuous variables (non-normally distributed) were conducted using the Mann-Whitney U test. Comparisons of categorical variables were performed using Fisher's exact test. The significance threshold was set at $p < 0.05$.

Results

Detection of PD-L1 expression and grouping

This study collected and analyzed fresh frozen tumor tissues and corresponding paraffin-embedded specimens from 21 HPV-positive CC patients. WB analysis revealed that samples with a PD-L1/GAPDH ratio above the median value included numbers 5, 6, 8, 14, 23, 27, etc., while those below the median included numbers 7, 12, 19, 26, 28, etc. (Fig. 1A,B). IHC results are shown in Fig. 1C. Based on predefined grouping criteria, three PD-L1 high-expression samples (Sample IDs: S14, S23, and S27) and three low-expression samples (Sample IDs: S19, S26, and S28) were selected for ONT. The baseline clinicopathological characteristics of the six patients are summarized in Table 1.

Sequencing data quality

The quality of extracted DNA was validated using a spectrophotometer (concentration: 105.27–333.62 ng/μL; A260/A280: 1.87–1.96) and Qubit Fluorometric Quantitation (concentration: 42.2–66.4 ng/μL) (Supplementary Table 1). The integrity of the extracted genomic DNA was further confirmed by 1.0% agarose gel electrophoresis (Supplementary Fig. 1), indicates that the sample quality was high and can proceed to the subsequent sequencing. Sequencing generated a total of 485.8 Gb of data, with an average read length of 5404.33 bp per sample (Table 2). The N50 ranged from 6826 bp to 9642 bp, Q20 ranged from 92.84% to 93.58%, and Q30 ranged from 87.63% to 88.59%. Base composition analysis showed an average GC content of 40.72% in the sequencing data. These metrics confirmed the high reliability of the sequencing data (Supplementary Table 2). Contour plots depicting the Q-score versus read length values for each sample were generated using NanoPlot (Fig. 2A–F), where the X-axis represents sequence length, the Y-axis represents the average Q-value of the sequence, and the color or line density indicates the density of data points. The results demonstrated good overall sequencing data quality and high reliability. Filtered full-length reads were aligned to the reference genome (GRCh38.p14) using Minimap2, achieving alignment rates exceeding 98% for all samples (Supplementary Table 3). Supplementary Fig. 2 illustrates the genome-wide coverage for each sample, revealing variations in coverage across chromosomes among different samples.

Host genomic variants

Distinguishing somatic from germline variants is crucial. Unfortunately, matched normal tissue or blood samples were not available for the patients in this retrospective cohort. To reduce the noise of germline variants, we have performed a rigorous filtering step by intersecting our called SNVs, indels, and SVs with gnomAD (v4.1.0, https://gnomad.broadinstitute.org/). A total of 221,867 SVs were detected across the six samples. The total number of SVs was comparable between the PD-L1 high expression group and the low expression group (high: 111,195; low: 110,672). INS (high: 54.28%, low: 54.05%) and DEL (high: 44.35%, low: 44.56%) were the predominant types (Fig. 3A). We compared the counts of each SV type between the two groups. However, limited by the sample size, the differences in the distribution of INS, DEL,

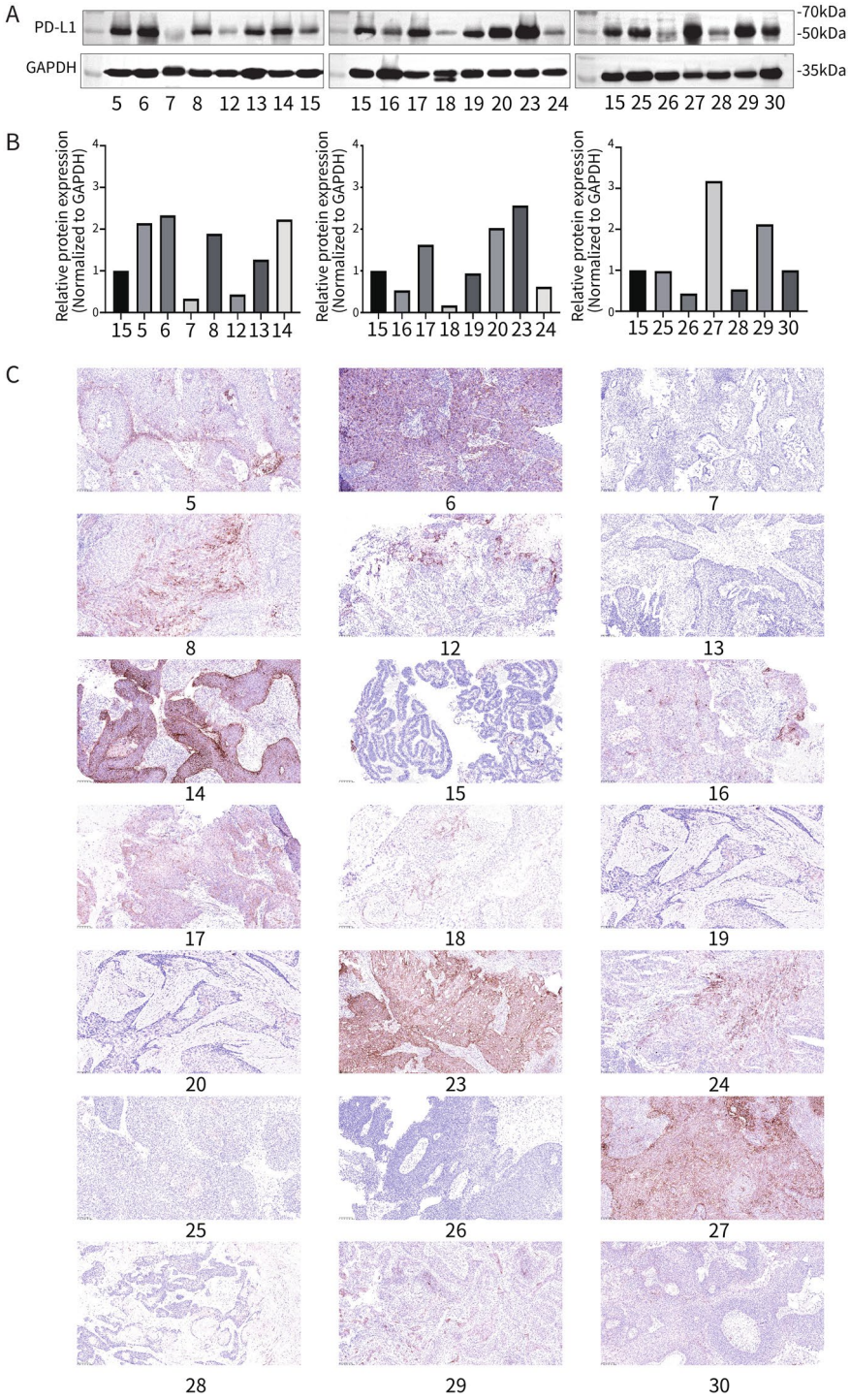

Fig. 1 PD-L1 expression levels. (**A**) PD-L1 expression status was evaluated by Western blot. (**B**) Statistical chart of PD-L1 protein expression determined by Western blot analysis. (**C**) PD-L1 expression status was evaluated by immunohistochemical staining

inversions (INV), duplications (DUP), and breakends (BND) did not reach statistical significance ($p > 0.05$). The lengths of the SVs were statistically analyzed according to different SV types; detailed results are provided in Supplementary Table 4. The SVs from both groups were categorized into seven length intervals: 0–50 bp, 50–100 bp, 100–500 bp, 500–1000 bp, 1000– 10,000 bp, 10,000–50,000 bp, and > 50000 bp. The results showed that BND events in all samples were concentrated in the 0–50 bp interval, with no detection in other intervals. In both groups, INS events were most abundant in the 100–500 bp interval. DUP and INV events were concentrated in

Table 1 Baseline clinicopathological characteristics

Sample	Group	Age	Stage FIGO	HPV type	Histologic Subtype	Lymph Node Metastasis	Degree of Differentiation
14	PD-L1 high expression group	50	IIIB	16	Squamous cell carcinoma	Yes	Moderately differentiated
23	PD-L1 high expression group	41	IIIC1	18	Squamous cell carcinoma	Yes	Moderately to poorly differentiated
27	PD-L1 high expression group	49	IIB	45	Squamous cell carcinoma	Yes	Poorly differentiated
19	PD-L1 low expression group	62	IV	16	Squamous cell carcinoma	Yes	Moderately differentiated
26	PD-L1 low expression group	60	IIIB	16	Squamous cell carcinoma	Yes	Moderately differentiated
28	PD-L1 low expression group	55	IIIC1	16	Squamous cell carcinoma	Yes	Poorly differentiated

the 1000– 10,000 bp interval, with almost none detected below 50 bp. DEL events were concentrated in the 0–50 bp interval. No statistically significant difference in SV length distribution was observed between the two groups ($p > 0.05$). Further analysis of the subgroup infected solely with HPV16 showed that within this subgroup, the distribution proportions of SV types and the SV length distribution between the PD-L1 high- and low-expression groups remained similar, with no significant differences. We used the ANNOVAR tool for functional annotation of the SVs, the specific genomic locations of the variants were determined, including upstream/downstream regions, intergenic regions, exons, and introns. It was found that SVs in both groups were significantly enriched in intergenic regions (Fig. 3B-G). Variants in exonic regions accounted for only 0.61% and 0.57% in the high and low expression groups, respectively. Functional variant annotation revealed that PTPRN2 and DLGAP2 exhibited high variant frequencies in both groups. Cumulative variants in PTPRN2 accounted for 526 in the high expression group and 508 in the low expression group. Cumulative variants in DLGAP2 accounted for 411 in the high expression group and 425 in the low expression group. These findings were validated in the subgroup of samples infected solely with HPV16, where the trends remained consistent (Supplementary Table 5). SVs in exonic regions were further classified by their functional impact (e.g., frameshift, non-frameshift, nonsynonymous, stop-loss, stop-gain, and synonymous) as detailed in Fig. 3H-M. Functional annotation results revealed that the profiles of functional impacts from exonic SVs were similar between the PD-L1 high and low expression groups. Both groups were predominantly characterized by frameshift mutations and non-frameshift mutations, with the vast majority of non-frameshift mutations being non-frameshift deletions. A total of 321 frameshift mutations and 242 non-frameshift mutations were identified in the high expression group, compared to 311 and 205, respectively, in the low expression group (Supplementary Table 6).

Next, we visualized the CNV profiles of the six samples (Fig. 4). Average coverage depth and log2 ratio data indicated differences in genomic stability among the samples. Sample 19 showed copy number increase on chromosomes 1 and 5 (Fig. 4D), while sample 28 showed copy number deletion on chromosomes 3 and 6 (Fig. 4F). The PD-L1 high expression group exhibited consistent copy number increases across samples at the CTNND2 gene locus on chromosome 5 and copy number deletions at the CCNYL3 gene locus on chromosome 16 and the DMD gene locus on the X chromosome. The PD-L1 low expression group exhibited copy number deletions at the CSMD1 gene locus on chromosome 8 and the ERBB4 gene locus on chromosome 2 (Supplementary Table 7). Given that large CNVs are more indicative of somatic genomic instability than germline polymorphisms, we extended our analysis to characterize CNVs by size. The CNVs were classified into four length-based categories: 0–10 kb, 10–50 kb, 50–100 kb, and >100 kb. The majority of CNVs exceeded 100 kb in length (Fig. 5B). A detailed comparison of the proportions of CNVs in these size categories between PDL1 high and low expression groups revealed that CNVs > 100 kb were more prevalent in the PDL1 low expression group, whereas those in the 0–10 kb and 10–50 kb ranges were more prevalent in the PDL1 high expression group (Fig. 5C). In contrast, the number of chromosomal breakend sites did not differ significantly between the two groups (Fig. 5A).

SNPs and Indels were identified using DeepVariant software and filtered using internationally accepted standards. Figure 6 illustrates the number of SNPs and Indels in different genomic and coding regions. Whole-genome variant calling identified a total of 225,405 SNPs and 20,256 Indels across both groups, with no significant difference between groups ($p > 0.05$). SNPs and Indels in intronic and intergenic regions predominated (Fig. 6A-F). In the coding sequence regions of the PD-L1 high expression group, there were 417 synonymous single nucleotide polymorphisms, 671 non-synonymous single nucleotide polymorphisms, 7 stop-gain mutations, 1 stop-loss mutation, 3 start-loss mutations, and 9 mutations of unknown function. In the PD-L1 low expression group coding sequence regions, there were 407 synonymous single nucleotide polymorphism, 621 non-synonymous single nucleotide polymorphism, 11 stop-gain mutations, 1 start-loss mutation, and 13 unknown function. The

Table 2 NanoPlot software converts and filters raw fastq data from ONT sequencing

Sample	Group	Mean Read Length	Mean Read Quality	Median Read Length	Median Read Quality	Number of Reads	Read Length N50	Stdev Read Length	Total Bases	Q20 Number	Q20%	Q20 Base	Q30 Number	Q30%	Q30 Base
14	PD-L1 high expression group	5601	20	4836	26	14351131	9044	5330	80380545879	11973182	83.40%	66401.4Mb	3848153	26.80%	15098.7Mb
23	PD-L1 high expression group	5417	20	4426	27	15524574	7941	5503	84103970764	13265355	85.40%	71849.8Mb	4187085	27.00%	17421.4Mb
27	PD-L1 high expression group	5334	19	4772	25	12770623	7245	3985	68126812739	10061047	78.80%	51972.3Mb	2434557	19.10%	10195.0Mb
19	PD-L1 low expression group	3265	20	1565	27	24606096	6826	4276	80339820602	20455098	83.10%	66445.3Mb	8029839	32.60%	17499.6Mb
26	PD-L1 low expression group	5529	19	4987	25	15706727	7054	3738	86852465747	12232670	77.90%	66222.9Mb	2592310	16.50%	12635.8Mb
28	PD-L1 low expression group	7280	19	6588	26	11815828	9642	5060	86029353033	9588192	81.10%	69193.6Mb	2163415	18.30%	13159.7Mb

Note: Q20 Number: The number of reads in which the quality score meets or exceeds Q20 (accuracy ≥ 99%); Q20%: The percentage of reads with a quality score of Q20 or higher out of the total number of reads; Q20 Base: The total number of bases across all reads with a quality score of Q20 or higher (accuracy ≥ 99%); Q30 Number: The number of reads in which the quality score meets or exceeds Q30 (accuracy ≥ 99.9%); Q30%: The percentage of reads with a quality score of Q30 or higher out of the total number of reads; Q30 Base: The total number of bases across all reads with a quality score of Q30 or higher (accuracy ≥ 99.9%)

proportions of these mutation categories showed no statistically significant difference between the PD-L1 high expression and low expression groups ($p > 0.05$) (Supplementary Table 8). After characterizing the genomic distribution of the SNPs, we next analyzed their molecular characteristics and functional impacts. The ratio of transitions to transversions (Ti/Tv) across the whole genome was 1.70 on average for the PD-L1 high expression group and 1.63 on average for the low expression group. Genome-wide SNP detection revealed heterozygous genotypes accounting for 67.85% and 62.34% in the high and low groups, respectively, and homozygous genotypes accounting for 32.15% and 37.66%, respectively. Genome-wide Indel detection revealed heterozygous genotypes accounting for 65.06% and 59.95% in the high and low groups, respectively, and homozygous genotypes accounting for 34.94% and 40.05%, respectively (Table 3). The length distribution of Indels is shown in Supplementary Fig. 3, revealing a peak at 1 bp across all samples. In the subgroup of samples infected solely with HPV16, the distribution patterns of SNPs and Indels across various categories and the results of intergroup comparisons remained consistent.

HPV integration characteristics stratified by PD-L1 expression

HPV integration status was analyzed using ONT sequencing in the six samples, revealing significant heterogeneity (Table 4). HPV integration breakpoints were determined based on the following criteria: (1) Raw reads aligned to the HPV genome with > 98% identity and length > 100 bp; (2) Chimeric reads were filtered, where one end aligned to the HPV genome and the other end to the human genome; (3) The human ends of the chimeric reads mapped to the same chromosome. Detailed information on the 90 identified insertion sites is provided in Table 5. Sample 19 exhibited 81 breakpoints (supporting reads = 256), sample 14 had 3 breakpoints (supporting reads = 8), and sample 28 had 6 breakpoints (supporting reads = 12). No valid breakpoints were detected in samples 23, 27, and 26. There was no significant difference in the total number of integration breakpoints between groups ($U = 1.159$, $p = 0.246$). Within the low expression group, integration breakpoints were highly enriched on chromosome 13 (67/87, 77.01%, $p < 0.001$), with the KLF5 gene region identified as an integration hotspot (supporting reads = 219). Due to the extremely low number of breakpoints in the high expression group, inter-group comparison of chromosomal distribution was not performed. The distribution of HPV integration breakpoints across the human genome is illustrated in Fig. 7, where red bars represent the PD-L1 low expression group, blue bars represent the PD-L1 high expression group, and bar length corresponds to the number of supporting reads at

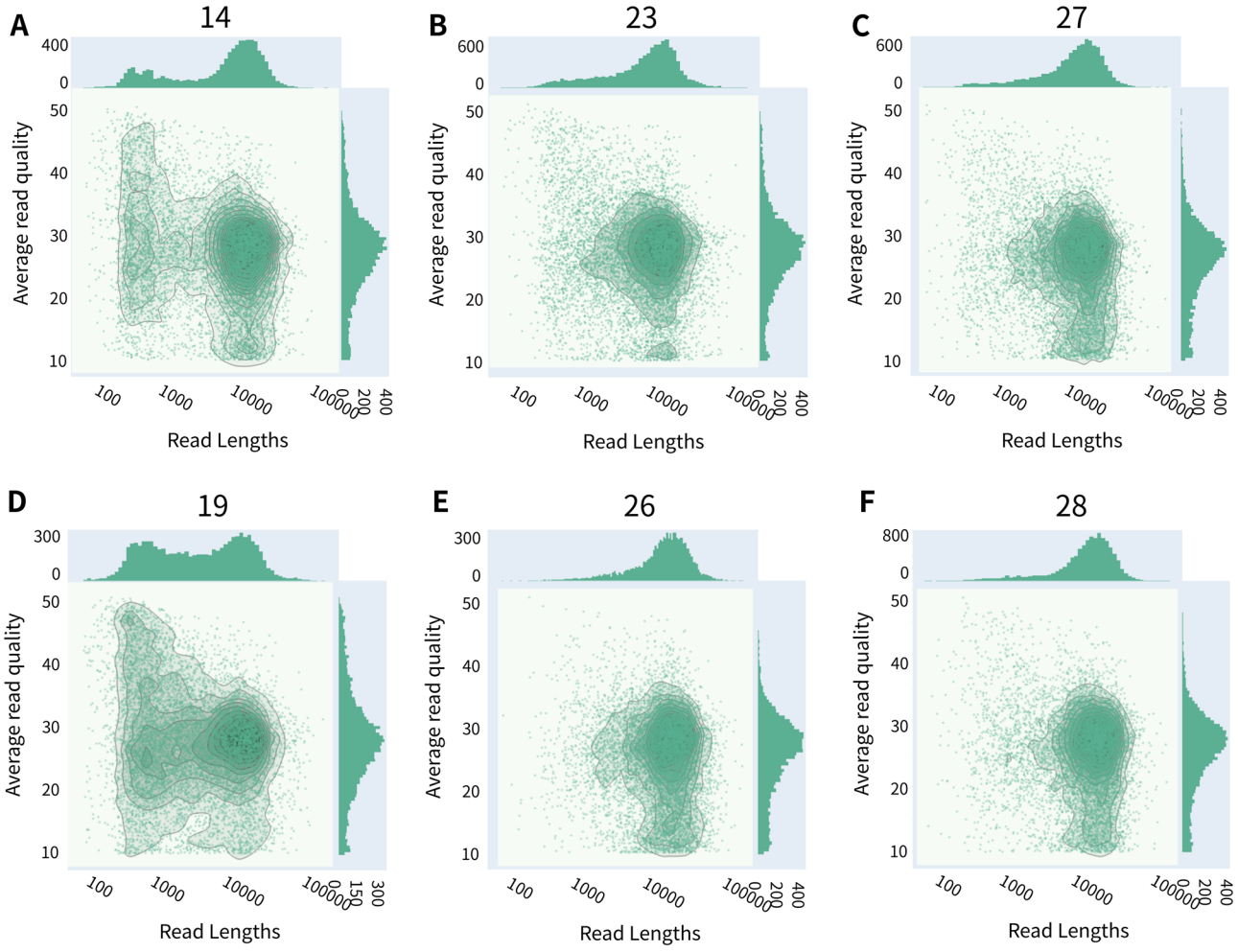

Fig. 2 Visualization of sequencing data read length and sequencing Q value. (**A**) Sample 14. (**B**) Sample 23. (**C**) Sample 27. (**D**) Sample 19. (**E**) Sample 26. (**F**) Sample 28. (Contour density plot showing the distribution of nanopore reads by length (x-axis) and average read quality score (y-axis))

each locus. In summary, the PD-L1 high expression group had a total of 3 insertion sites involving 5 genes, while the PD-L1 low expression group had 87 insertion sites involving 26 genes. Among these, KLF5 and PTPN13 are previously reported HPV integration sites. Specific analysis targeting the PD-L1 gene locus (chromosome 9:5,450,503–5,470,566) for HPV integration revealed no detectable HPV insertion events within the PD-L1 gene in any of the samples. To evaluate the potential influence of HPV subtypes, we analyzed the subgroup of samples infected solely with HPV16, and found that the high integration burden in the PD-L1 low-expression group and the trend of KLF5 hotspot integration persisted.

Schematic illustration of four different types of integration events

Based on breakpoint junction structure analysis, this study categorized HPV integration events into four types: Type I: Intact E6/E7 Integration. Characterized by partial deletion of the HPV genome while retaining the intact E6/E7 oncogenes, with both ends connected to the host genome (Fig. 8A). Type II: E6/E7-Disrupted Integration. Characterized by HPV genomic integration that does not contain the complete E6/E7 open reading frame (Fig. 8B). Type III: Multi-copy Tandem Integration. Characterized by integration containing at least one complete HPV genome tandem repeat unit, accompanied by partial viral fragments (Fig. 8C). Type IV: The variants of Type I with multiple orientationally inconsistent human genomic fragments flanking both sides of the HPV DNA integration site (Fig. 8D,E). Additionally, the study identified clonal integration events where tumor cells shared identical integration breakpoints. For instance, integration events such as 025f4e53-271f-42bc-b360-e09b6aaae55d, 09667a7a-7a26-4303-bbc4-9b4dba6e6ad2, and 1367ce64-14b9-4ac0-9edb-979762c43e43 all involved independent HPV16 integration events occurring on chromosome 13. These events shared an identical HPV16 fragment but had distinct breakpoint/junction sites (Supplementary Table 9).

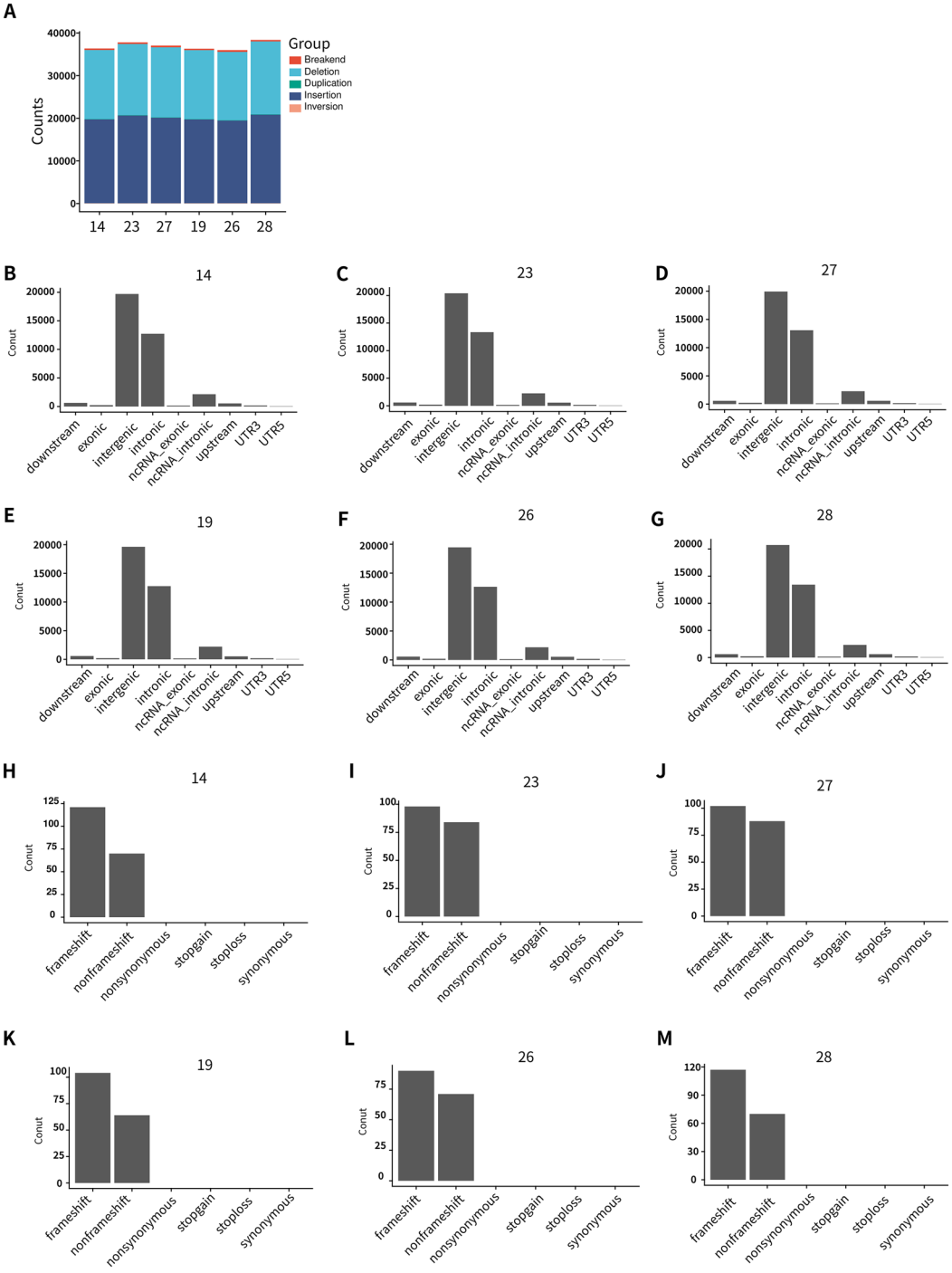

Fig. 3 Landscape of structural variations (SVs) in PD-L1 high and low expression groups. (**A**) The distribution of SV types across all samples (samples are grouped by PD-L1 expression level (high: samples 14, 23, 27; low: samples 19, 26, 28)). (**B–G**) genomic distribution of SV for each individual sample: (**B**) sample 14 (high), (**C**) sample 23 (high), (**D**) sample 27 (high), (**E**) sample 19 (low), (**F**) sample 26 (low), and (**G**) sample 28 (low). The bar charts show the number of SV located in intergenic, intronic, exonic, upstream/downstream, UTR, ncRNA_exonic, and ncRNA_intronic. (**H–M**) functional impact classification of exonic structural variants for each individual sample: (**H**) sample 14 (high), (**I**) sample 23 (high), (**J**) sample 27 (high), (**K**) sample 19 (low), (**L**) sample 26 (low), and (**M**) sample 28 (low). Categories include frameshift, non-frameshift, nonsynonymous, stopgain, stoploss, synonymous mutations, among others

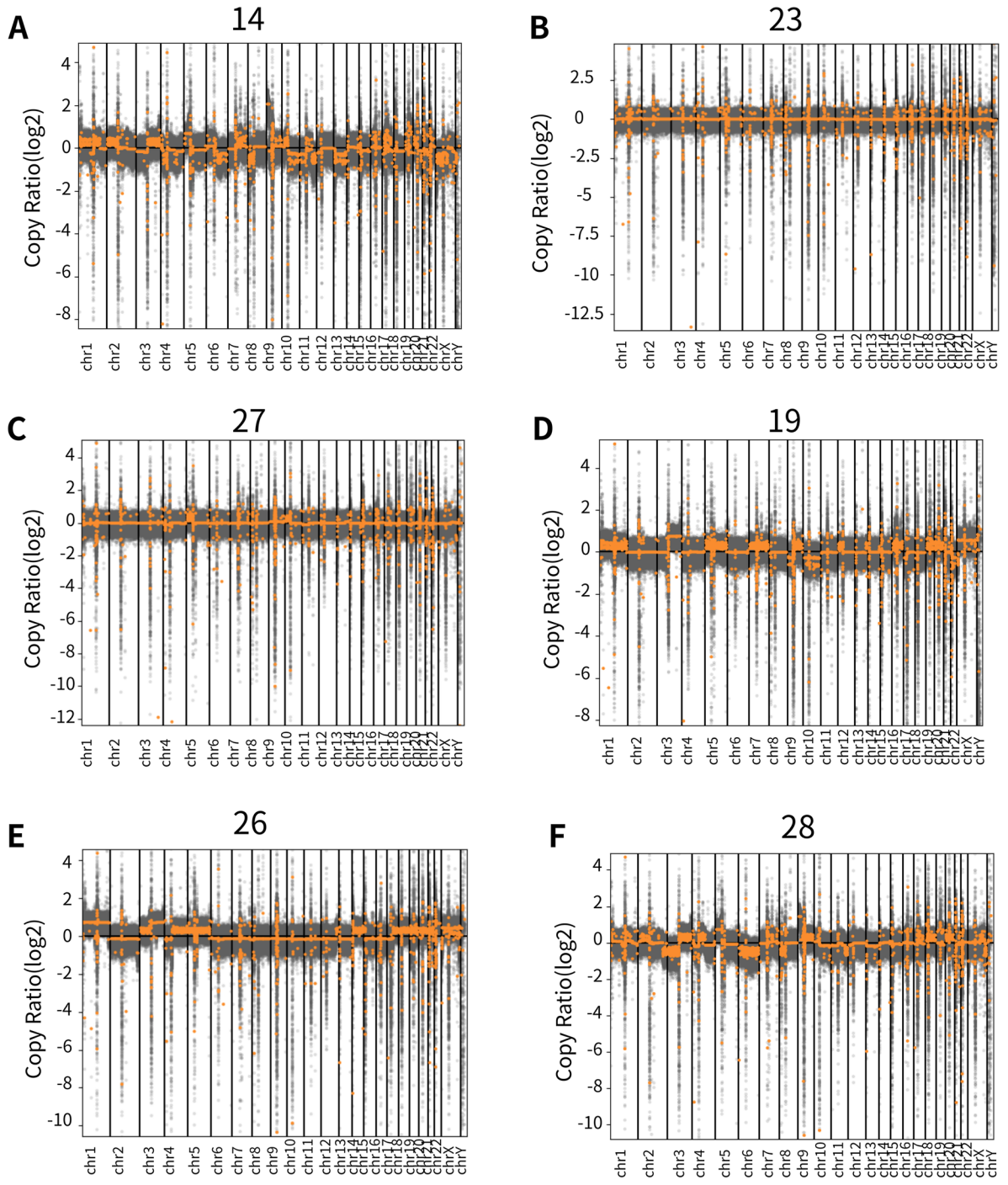

Fig. 4 Identification of gene copy number variations. (**A**) Sample 14. (**B**) Sample 23. (**C**) Sample 27. (**D**) Sample 19. (**E**) Sample 26. (**F**) Sample 28. (This plot displays the distribution of copy-number variants across the entire genome. The x-axis represents chromosomal positions. The y-axis shows the log2-transformed copy ratio)

Enrichment analysis

Given that the PD-L1 high expression group involved only 5 genes, enrichment analysis was not performed for this group. GO enrichment (Supplementary Table 10) and KEGG pathway analyses (Supplementary Table 11) were conducted for the PD-L1 low expression group. The results for Biological Processes revealed that genes affected by HPV integration in the PD-L1 low expression group were significantly enriched in "regulation of transcription by RNA polymerase II" (Fig. 9A). For Cellular Components, significant enrichment was observed in "membrane", "extracellular exosome", and "nucleoplasm"

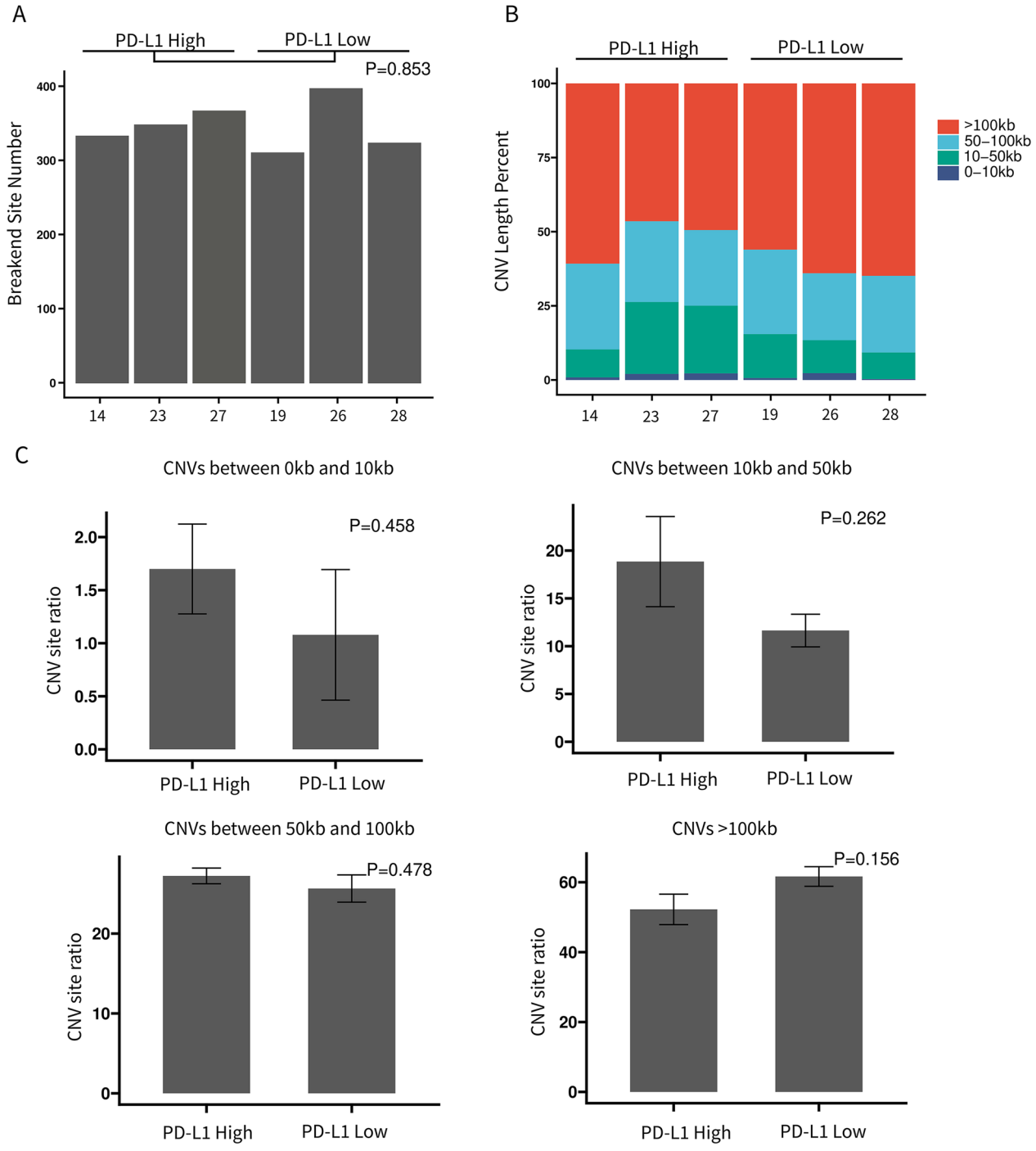

Fig. 5 Statistic analysis of translocations and copy number alterations stratified by PD-L1 expression. (**A**) Bar plot of the number of the detected chromosomal breakend sites in each sample. (**B**) Bar plot displaying the fractions of different lengths of CNVs (0–10 kb, 10–50 kb, 50–100 kb, and >100 kb) in each sample. (**C**) Differential burden of CNV size fractions. For each CNV size range defined in panel **B**, the fractions were compared between PD-L1-high and PD-L1-low groups

Fig. 6 The number of SNPs and Indels in different regions of the genome and coding regions. (**A–F**) Pie chart showing the proportion of SNPs and Indels across different genomic regions ((**A**) sample 14 (high), (**B**) sample 23 (high), (**C**) sample 27 (high), (**D**) sample 19 (low), (**E**) sample 26 (low), and (**F**) sample 28 (low))

(Fig. 9B). Regarding Molecular Functions, the predominant enrichment was in "protein binding" (Fig. 9C). KEGG pathway analysis identified significant enrichment in "progesterone-mediated oocyte maturation" and "synthesis, secretion, and action of parathyroid hormone". To understand the interactions among the corresponding genes, a PPI network was constructed. This network comprised 58 genes (Fig. 9D).

Discussion

HPV integration is a complex process resulting from the interplay between the virus, host genes, and environmental factors. The multifaceted nature of this process underscores the importance of dynamic assessment to understand this event. Since the integration of HPV DNA into the host genome disrupts its structure, previous short-read sequencing technologies have been unable

Table 3 Characteristics of SNPs and Indels on the genome

Sample	Group	Single Nucleotide Polymorphism												Insertion-Deletions					
		Total	Heterozygous Genotype	Homozygous Genotype	Multialleic Site	Transition	Transversion	Transition/Transversion Ratio	Percentage of SNPs in dbSNP	Novel SNP	Novel Transition	Novel Transversion	Novel Transition/Transversion Ratio	Total	Heterozygous Indel	Homozygous Indel	Multialleic Indel Site	Percentage of Indels in dbSNP	Novel Indel
14	PD-L1 high expression group	23082	12216	10866	100	13920	9062	1.54	12680(54.93%)	10402	5563	4778	1.16	2120	1253	867	338	772(36.42%)	1348
23	PD-L1 high expression group	50083	35221	14862	143	31923	18017	1.77	30430(60.76%)	19653	11245	8322	1.35	3911	2600	1311	450	1459(37.31%)	2452
27	PD-L1 high expression group	46312	33628	12684	131	29618	16563	1.79	28164(60.81%)	18148	10519	7539	1.4	3347	2248	1099	410	1251(37.38%)	2096
19	PD-L1 low expression group	40801	27784	13017	137	25940	14724	1.76	23959(58.72%)	16842	9556	7185	1.33	3125	1974	1151	369	1142(36.54%)	1983
26	PD-L1 low expression group	44001	28572	15429	143	27724	16134	1.72	24372(55.39%)	19629	11075	8461	1.31	3430	2075	1355	426	1260(36.73%)	2170
28	PD-L1 low expression group	20424	9244	11180	111	11936	8377	1.42	10311(50.48%)	10113	5259	4781	1.1	1976	1065	911	354	686(34.72%)	1290

Table 4 HPV integration site information

Sample	Group	HPV Genotype	Number of Reads Aligned to HPV	Number of Reads with ident > 98% and Length > 100 bp	Number of Reads with Sequences on Both Ends	Number of Reads Aligned to the Same Chromosome	Number of Insertion Sites Detected
14	PD-L1 high expression group	HPV16	3281	491	23	8	3
23	PD-L1 high expression group	HPV18	285	52	0	-	-
27	PD-L1 high expression group	HPV45	76	14	1	0	-
19	PD-L1 low expression group	HPV16	49633	8644	677	256	81
26	PD-L1 low expression group	HPV16	631	101	0	-	-
28	PD-L1 low expression group	HPV16	2596	338	29	12	6

to effectively span these complex regions to identify HPV integration breakpoints. Therefore, in this study, we employed ONT long-read sequencing to perform whole-genome analysis on six CC specimens. This single-molecule detection approach, coupled with its long-read length capability, significantly enhances the precision of locus identification, providing support for revealing host genomic variations and HPV integration characteristics stratified by PD-L1 expression. Analysis revealed no statistically significant differences between the PD-L1 high expression and low expression groups regarding the total number of SVs, the distribution of SV types (INS, DEL, INV, DUP, BND), the distribution of SV length intervals, the total number of SNPs, the total number of Indels, or the proportional distribution of functional categories of SNPs/Indels in coding regions. This suggests that the overall burden and primary characteristics of host genomic variations are not the main drivers of PD-L1 expression differences in the samples studied here. The ratio of Ti/Tv and the proportions of heterozygous to homozygous genotypes were also highly similar between the two groups, further supporting the consistency in overall genomic variation patterns between groups. Analysis of HPV integration revealed a striking contrast between the PD-L1 expression groups. In the high-expression group, despite only 3 HPV16 insertion sites being identified, these events impacted 5 genes. This suggests a pattern of precision integration, where the functional impact may far exceed their numerical scarcity. The integration events in the high-expression group were enriched for genes that play pivotal roles in key biological processes such as signaling pathway regulation, immune regulation, and cell proliferation/migration [15–17]. We speculate that the oncogenic model in the PD-L1 high-expression group may not rely on large-scale genomic instability, but rather on a few HPV16-driven integration events that precisely disrupt the immune regulatory network. In contrast to the precision integration pattern of the high-expression group, the PD-L1 low-expression group exhibited a classic integration-driven trait. Its high integration load implies extensive genomic instability. Furthermore, we identified recurrent integration in the KLF5 gene and a preference for integration on chromosome 13 within the low-expression group. No HPV insertion events within the PD-L1 gene were detected in any sample. Given that the average sequencing depth across all samples exceeded 28X and 10X coverage depth was achieved for over 90% of the genome, we consider the absence of detectable HPV INS within the PD-L1 gene unlikely to be a false negative result due to insufficient sequencing depth or uneven coverage. Therefore, we hypothesize a causal relationship between HPV integration patterns and PD-L1 expression levels: extensive HPV integration-driven genomic instability on a large scale may shape a tumor microenvironment with low PD-L1 expression, whereas a limited number of precise integration events may directly lead to a high PD-L1 expression state by impacting key immune regulatory networks.

This study, through whole-genome analysis, revealed the association between PD-L1 expression stratification and host genomic variation patterns. The analysis found that the total number of SVs was comparable between the two groups. Although no significant inter-group differences were observed, high-frequency cumulative SV variations were noted in both groups for the PTPRN2 and DLGAP2 genes. This finding resonates significantly with recent deep profiling of early lung cancer genomes by Cui et al [18]. Their study, utilizing ONT, confirmed not only high-frequency tandem repeat amplifications of DLGAP2 in lung cancer but also concomitant abnormal hypermethylation in its promoter region, leading to epigenetic silencing of the gene and driving tumor progression. This suggests that the cumulative SVs in DLGAP2 observed in our study may directly disrupt its tumor-suppressive function, potentially synergizing with methylation-mediated silencing to promote tumor immune escape. Our study also revealed a significant regional preference in the distribution of SVs, SNPs, and Indels, with predominant enrichment in intergenic regions and introns. This pattern aligns with the characteristic high proportion of non-coding regions in the genome [19]. CNVs were detected in multiple samples. Notably, in the PD-L1 high expression group, we observed copy number increases at the CTNND2 locus on chromosome 5 and copy number deletions at the CCNYL3 locus on

Table 5 Detailed HPV insertion site information

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
14	PD-L1 high expression group	chr21	10325058	10325058	1	Distal Intergenic	-	BAGE2	Distal Intergenic	-	BAGE2
14	PD-L1 high expression group	chr3	190013033	189867554	4	Intron (ENST00000319332.10/55214, intron 1 of 14)	ENSG00000090530	P3H2	Intron (ENST00000264731.8/8626, intron 6 of 13)	ENSG00000216058	MIR944
14	PD-L1 high expression group	chr4	86440115	86712964	3	Intron (ENST00000642035.1/5602, intron 1 of 12)	ENSG00000109339	MAPK10	Intron (ENST00000427191.6/5783, intron 7 of 46)	ENSG00000163629	PTPN13
19	PD-L1 low expression group	chr1	29515048	29515041	1	Distal Intergenic	ENSG00000230523	LINC01756	Distal Intergenic	ENSG00000230523	LINC01756
19	PD-L1 low expression group	chr10	74381520	74381529	1	Intron (ENST00000672429.1/132, intron 4 of 9)	ENSG00000156110	ADK	Intron (ENST00000672429.1/132, intron 4 of 9)	ENSG00000156110	ADK
19	PD-L1 low expression group	chr11	76277829	76277829	1	Distal Intergenic	ENSG00000255362	LINC02761	Distal Intergenic	ENSG00000255362	LINC02761

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr11	82469617	29276131	1	Intron (ENST00000668951.1/101928989, intron 1 of 3)	ENSG00000245832	MIR4300HG	Exon (ENST00000525759.1/ ENST00000525759.1, exon 1 of 1)	ENSG00000249867	LINC02742
19	PD-L1 low expression group	chr12	125299442	4386540	1	Intron (ENST00000682704.1/114795, intron 1 of 8)	ENSG00000139364	TMEM132B	Intron (ENST00000648100.1/57103, intron 11 of 11)	ENSG00000118972	FGF23
19	PD-L1 low expression group	chr12	45931345	45902805	1	Exon (ENST00000546534.1/9169, exon 2 of 2)	ENSG00000139218	SCAF11	Intron (ENST00000334344.11/196528, intron 20 of 20)	ENSG00000189079	ARID2
19	PD-L1 low expression group	chr13	109081881	73066830	1	Intron (ENST00000356711.7/23026, intron 27 of 34)	-	MYO16–AS2	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	24724399	73041576	1	Intron (ENST00000649394.1/105370118, intron 2 of 2)	-	LOC105370118	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	56898016	73066148	1	Distal Intergenic	-	LOC124903235	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73041559	73041576	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73041576	73041559	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73041576	73041576	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73056059	73041576	1	Intron (ENST00000477333.5/688, intron 1 of 1)	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73056059	73041577	1	Intron (ENST00000477333.5/688, intron 1 of 1)	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73066574	36509873	1	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000133101	CCNA1

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73067388	73066148	1	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73076072	73066148	1	3′UTR	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73077109	73066545	1	3′UTR	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73077888	73041576	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73078630	73068213	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73079854	73041576	2	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73079904	73086068	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73081059	73066148	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73085730	73066545	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086055	73041568	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086055	73068856	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086055	73070861	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086056	73066148	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	66430213	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000377865.7/5101, intron 4 of 4)	ENSG00000184226	PCDH9
19	PD-L1 low expression group	chr13	73086068	73041559	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73041576	32	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73041577	3	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73041587	2	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086068	73042323	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73062488	35	Distal Intergenic	ENSG00000102554	KLF5	Exon (ENST00000477333.5/688, exon 2 of 2)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73062503	3	Distal Intergenic	ENSG00000102554	KLF5	Exon (ENST00000477333.5/688, exon 2 of 2)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73066148	61	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73066154	4	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73066180	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086068	73066195	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73066545	5	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73066830	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73068213	7	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73068323	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73068481	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086068	73068606	3	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73068717	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73068856	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73069466	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73069565	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73069598	3	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086068	73069892	2	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73070017	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73070136	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73070296	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73070431	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73070829	2	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086068	73070848	2	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73070943	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73071281	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73071769	3	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73071907	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73072280	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086068	73074319	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73074777	1	Distal Intergenic	ENSG00000102554	KLF5	Intron (ENST00000539231.5/688, intron 3 of 3)	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73075091	1	Distal Intergenic	ENSG00000102554	KLF5	Promoter	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73079637	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73079904	3	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	73086015	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr13	73086068	73086068	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73086068	80348039	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000136158	SPRY2
19	PD-L1 low expression group	chr13	73087279	73041576	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr13	73108246	73041576	1	Distal Intergenic	ENSG00000102554	KLF5	Distal Intergenic	ENSG00000102554	KLF5
19	PD-L1 low expression group	chr16	86496945	86495506	1	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR
19	PD-L1 low expression group	chr16	86496958	86489024	1	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR	Exon (ENST00000598996.3/400550, exon 3 of 3)	ENSG00000268388	FENDRR

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
19	PD-L1 low expression group	chr16	86496958	86495506	1	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR
19	PD-L1 low expression group	chr16	86496958	86496968	20	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR
19	PD-L1 low expression group	chr16	86496959	86496968	2	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR
19	PD-L1 low expression group	chr16	86565764	86496968	1	Intron (ENST00000563280.3/103752587, intron 1 of 1)	ENSG00000176692	FOXC2	Intron (ENST00000595886.1/400550, intron 1 of 3)	ENSG00000268388	FENDRR
19	PD-L1 low expression group	chr6	71668285	92438065	1	Distal Intergenic	-	LINC01626	Distal Intergenic	ENSG00000200492	LOC124901505
19	PD-L1 low expression group	chrX	44327664	44327692	1	Intron (ENST00000420999.2/80258, intron 1 of 14)	ENSG00000183690	EFHC2	Intron (ENST00000420999.2/80258, intron 1 of 14)	ENSG00000183690	EFHC2

Table 5 (continued)

Sample	Group	Chromosome	Front Partner Genomic Position	Back Partner Genomic Position	Number of Supporting Reads	Front Partner Gene Annotation	Front Partner ENSEMBL Gene ID	Front Partner Gene Symbol	Back Partner Gene Annotation	Back Partner ENSEMBL Gene ID	Back Partner Gene Symbol
28	PD-L1 low expression group	chr3	190013033	189867554	2	Intron (ENST00000319332.10/55214, intron 1 of 14)	ENSG00000090530	P3H2	Intron (ENST00000264731.8/8626, intron 6 of 13)	ENSG00000216058	MIR944
28	PD-L1 low expression group	chr4	86440111	86712964	1	Intron (ENST00000642035.1/5602, intron 1 of 12)	ENSG00000109339	MAPK10	Intron (ENST00000427191.6/5783, intron 7 of 46)	ENSG00000163629	PTPN13
28	PD-L1 low expression group	chr4	86440115	72804956	1	Intron (ENST00000642035.1/5602, intron 1 of 12)	ENSG00000109339	MAPK10	Distal Intergenic	ENSG00000156140	ADAMTS3
28	PD-L1 low expression group	chr4	86440115	86712954	1	Intron (ENST00000642035.1/5602, intron 1 of 12)	ENSG00000109339	MAPK10	Intron (ENST00000427191.6/5783, intron 7 of 46)	ENSG00000163629	PTPN13
28	PD-L1 low expression group	chr4	86440115	86712964	6	Intron (ENST00000642035.1/5602, intron 1 of 12)	ENSG00000109339	MAPK10	Intron (ENST00000427191.6/5783, intron 7 of 46)	ENSG00000163629	PTPN13
28	PD-L1 low expression group	chr4	86440115	86712973	1	Intron (ENST00000642035.1/5602, intron 1 of 12)	ENSG00000109339	MAPK10	Intron (ENST00000427191.6/5783, intron 7 of 46)	ENSG00000163629	PTPN13

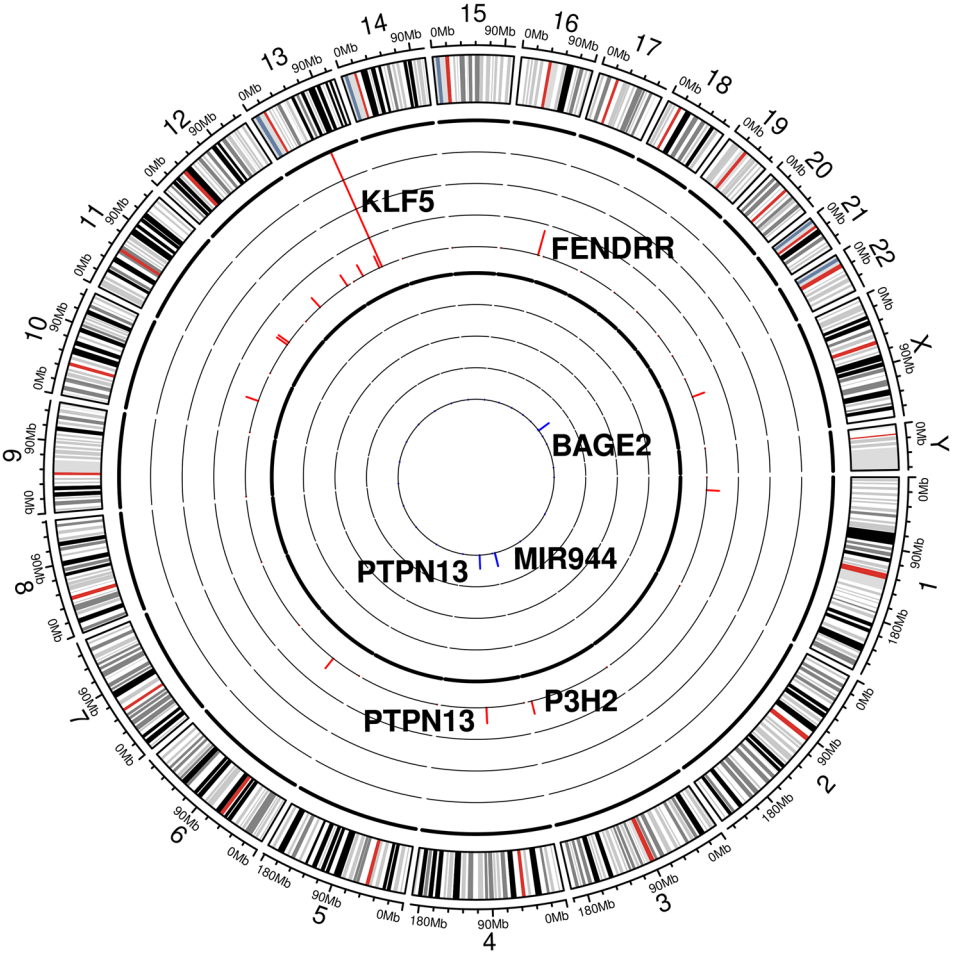

Fig. 7 Distribution of integration breakpoints in the human genome. (A circos plot was generated to illustrate the genomic locations of the detected HPV integration sites. The inner circle, annotated in blue, represents the PD-L1 high group, while the outer circle, annotated in red, represents the PD-L1 low group. The bars positioned along the genome axis indicate the genomic locations, with their heights corresponding to the number of detected sites)

chromosome 16 and the DMD locus on the X chromosome. The low expression group exhibited copy number deletions at the CSMD1 locus on chromosome 8 and the ERBB4 locus on chromosome 2. Such CNVs can promote cancer susceptibility by altering the dosage of genes involved in DNA damage repair, tumor suppression, or oncogenesis, including RB1, PTEN, CDKN2A/B, ARID1A, BRCA1/2, MSH2/6, ERBB2, EGFR, MYC, PIK3CA, FGFR1/2, CCND1, MET, and CDK6 [20]. Furthermore, a study analyzing genetic aberrations in 9p24.1 and PD-L1 protein expression in 27 cases of Hodgkin lymphoma found a significant correlation between PD-L1 protein expression and 9p24.1 copy number alterations [21]. Similarly, copy number changes in CD274 (encoding PD-L1) and increased PD-L1 expression are highly correlated in multiple tumor types, including CC [22]. Our study detected a region of CD274 copy number increase (log2 ratio = 2.04911, probes = 3158, weight = 2768.57) on chromosome 9 in sample 14 (PD-L1 high expression group). Conversely, regions of CD274 copy number deletion were detected on chromosome 9 in samples 19 and 26 (PD-L1 low expression group) (log2 ratio = −0.292229 and −0.128926, probes = 1960 and 7734, weight = 1801.19 and 7130.4). This suggests a potential association between CD274 copy number increase and high PD-L1 expression, and between CD274 copy number deletion and low PD-L1 expression. These findings indicate that further analysis of CNV signatures across different samples could help identify regions harboring potential tumor suppressor genes or oncogenes, providing candidate targets for subsequent targeted therapies, and facilitate the exploration of their interactions with PD-L1 expression levels and their impact on the tumor immune microenvironment.

Previous studies have indicated that HPV integration events tend to occur at chromosomal fragile sites, Alu sequences, exonic and intronic regions of genes, and have identified recurrent integration hotspots in genes including CCDC106, MACROD2, FHIT, POU5F1B, KLF12, KLF5, LRP1B, LEPREL1, HMGA2, DLG2, and

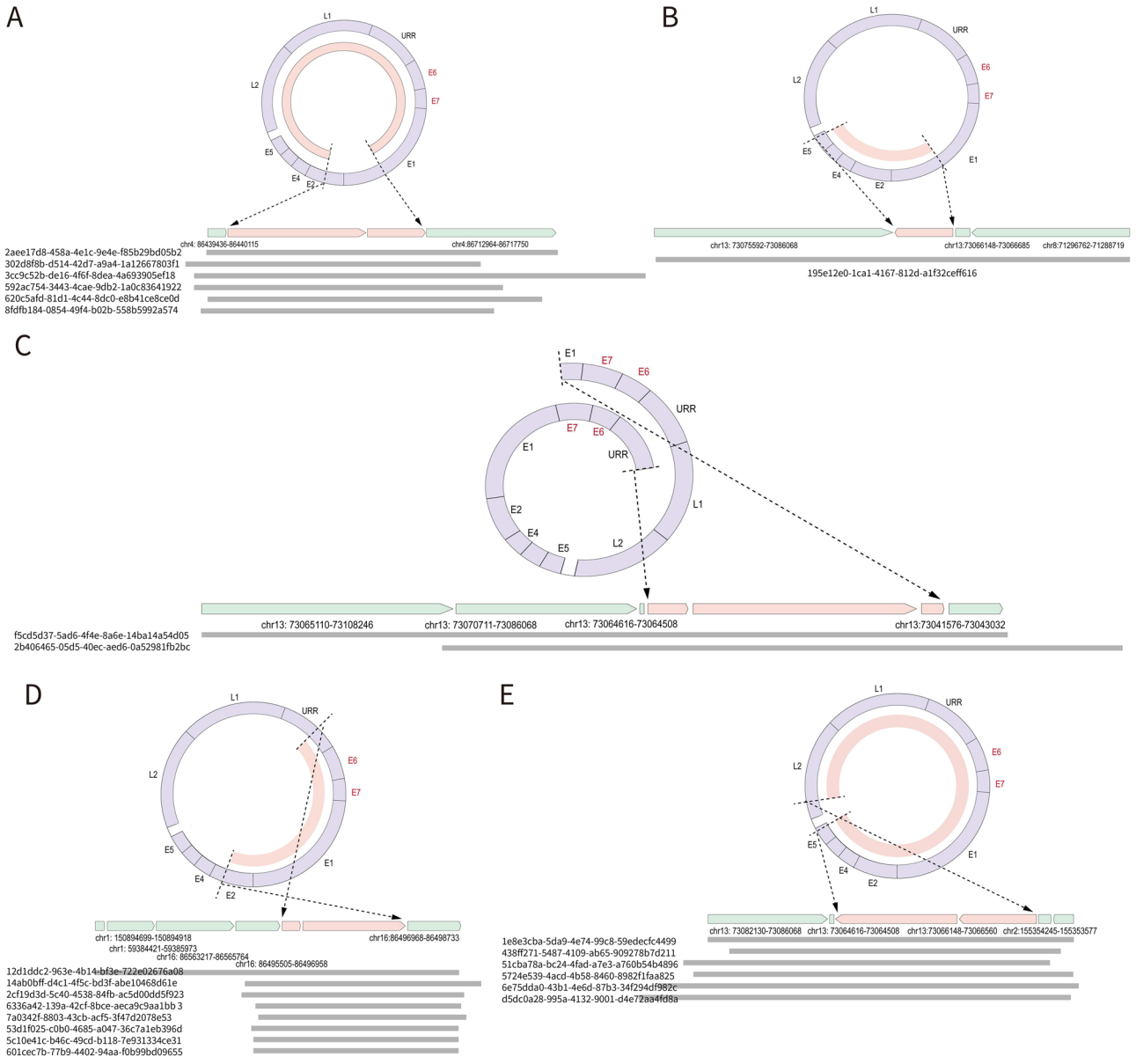

Fig. 8 Four types of integrated HPV DNA segments in clonal integration events. (**A**) Type I: E6/E7 complete integration. (**B**) Type II: E6/E7 deletion-integration. (**C**) Type III: multi-copy tandem integration. (**D**, **E**) type IV: the variants of type I with multiple orientationally inconsistent human genomic fragments flanking both sides of the HPV DNA integration site. (In each panel, the bottom grey lines indicate the sequencing reads supporting the corresponding HPV integration type, with the reads IDs being included. In panels **A**, **B**, **D** and **E**, the outer circles on the top indicate the HPV genome and the inner indicates DNA fragments integrated into the human genome; while in panel **C**, the light purple circle indicates the integrated HPV genome)

ERAS [23]. This study found that in CC patients with low PD-L1 expression, HPV integration events were significantly enriched at the KLF5 gene locus. KLF5 belongs to the Krüppel-like factor family and is widely expressed in various solid tumors [24]. Its high expression in CC promotes the proliferation, migration, and invasion of CC cells [25]. Studies suggest that HPV integration may regulate KLF5 expression and potentially modulate the function of nearby long non-coding RNAs, indicating that HPV integration could influence disease progression by interfering with KLF5 gene function. Notably, the function of KLF5 extends beyond the tumor cells themselves. In basal-like breast tumor, KLF5 has been demonstrated to be closely associated with the tumor immune microenvironment. Specifically, inhibiting KLF5 promotes CD8+ T-cell infiltration, induces the formation of memory T cells, and enhances the efficacy of anti-PD1 monoclonal antibody therapy [26]. Although its role in the CC immune microenvironment is not yet fully elucidated, we hypothesize that HPV integration may lead to aberrant KLF5 expression. Given that KLF5 has been shown to possess potent immunosuppressive properties,

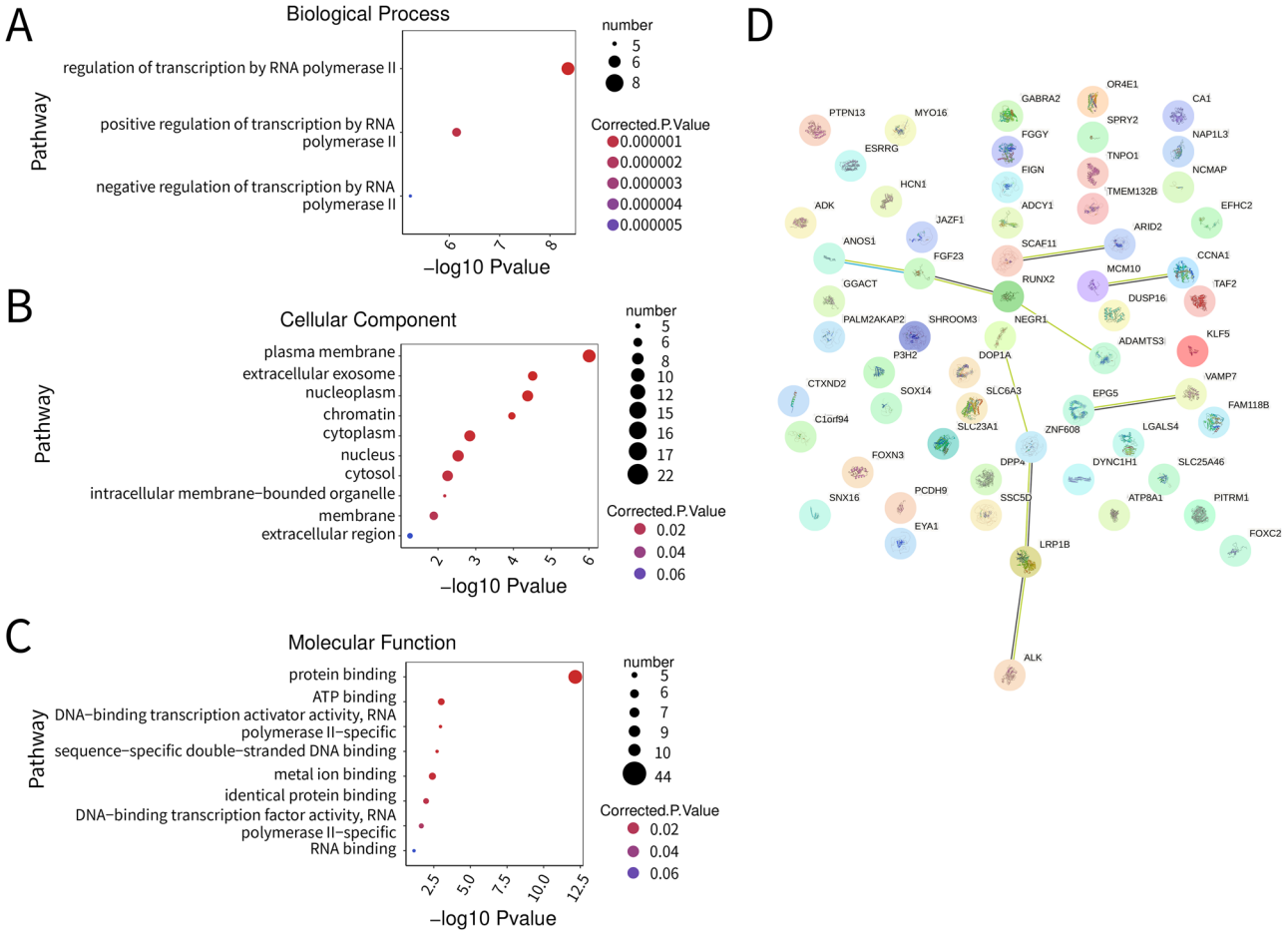

Fig. 9 Functional enrichment and interaction network analysis. (**A**) Biological pathway. (**B**) Cellular component. (**C**) Molecular function. (**D**) PPI analysis

this HPV integration-driven dysregulation of KLF5 is highly likely to be one of the key mechanisms contributing to the immunosuppressive state within the tumor microenvironment of PD-L1 low expression CC. Furthermore, the therapeutic targetability of KLF5 is supported by multiple studies. Eight classes of compounds with distinct mechanisms of action—including metformin, mifepristone, BRD4 inhibitors, CDK7 inhibitors, PRMT5 inhibitors, RSK2 inhibitors, and HDAC inhibitors have been shown to significantly inhibit tumor growth by downregulating KLF5 protein expression [27–31]. These findings provide a reference basis for the feasibility of targeting KLF5 in CC treatment and lay the groundwork for precision therapy in the PD-L1 low expression subtype.

Fine characterization of HPV integrated fragments within the host genome is crucial for deciphering viral oncogenic mechanisms. Previous studies utilizing Southern blotting and short-read sequencing technologies have revealed patterns of single integration, concatemeric integration, and looping integration [32]. However, constrained by their resolution, these methods struggled to systematically resolve the complete structural features of integration events. Subsequently, researchers employed long-read sequencing technologies to characterize four different types of integrated HPV DNA fragments: truncated HPV genomes containing E6/E7, truncated HPV genomes lacking E6/E7, overflown continuous fragments containing the full HPV genome, and combinations of the above three types [14]. Furthermore, other researchers performed HPV integration detection in three HPV-positive cell lines using ONT. They found that in the SiHa cell line, the viral integration pattern mainly manifested as partial insertion of viral sequences. In the HeLa cell line, integration was identified through the formation of tandem repeats between the viral genome and the human genome. In the CaSki cell line, features included both multiple viral tandem repeats and simple insertion patterns [33]. Utilizing long-read sequencing technology, we characterized four types of integration events observed in clonal integration, aligning with previous findings. Notably, we identified multiple independent HPV integration events within the same sample. These events all utilized an identical HPV16 viral fragment (HPV16REF: positions 2746–1 bp, 7906–7326 bp, and 3538–4002 bp). Despite

being independent events, they were highly concentrated within a relatively small interval on chromosome 13 (between 42,753,969 and 73,068,082 bp), suggesting that the genomic structure at this locus may be particularly susceptible to HPV integration.

Limitations

Our study has the following limitations. First, while the clinical sample size was sufficient for initial exploration, it limited the statistical power for subgroup analyses. Although ONT captured long-read integration reads in the six representative samples, the detection of rare events requires validation in larger cohorts. Future work will expand the sample size through multi-center collaborations. Second, the matched normal tissue or blood samples were not available for the patients in this retrospective cohort, which prevents direct distinguishing somatic from germline variants. Third, the causal relationship between the identified integration sites and PD-L1 expression has not been validated in gene-editing models. This constitutes a key objective of our next research phase.

Conclusion

This study, utilizing ONT, has revealed the characteristics of host genomic variations and HPV integration stratified by PD-L1 expression. These findings not only enhance our understanding of the genomic complexity in CC but also provide new insights into the relationship between PD-L1 expression and tumor genomic features. Furthermore, this work lays a solid foundation for advancing personalized therapy and precision medicine research in CC.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-07674-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Acknowledgements
We sincerely thank all the authors and study participants for their support in this study. We appreciate technical support from Wuhan Ruixing Biotechnology Co., Ltd.

Author contributions
RJL, JZ, XMM, and YCF designed the study. RJL wrote the manuscript. RJL,and JZ drew the diagrams and analyzed the data. XMM and YCF proofread the manuscript. All authors contributed to the article and approved the submitted version.

Funding
The research was founded by the National Natural Science Foundation of China (32360041), the Xinjiang Uygur Autonomous Region Tianchi Talent-Young Doctor (2025TCYCZJ) and the Innovation and Entrepreneurship Project of Postgraduates in Xinjiang Medical University (CXCY2025015).

Data availability
For data availability please email corresponding author.

Declarations

Ethics approval
This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Affiliated Tumor Hospital of Xinjiang Medical University (Approval No. G-2023001).

Consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that there is no conflict of interests.

Author details
[1]Department of Medical Laboratory Center, Tumor Hospital Affiliated to Xinjiang Medical University, Urumqi, China
[2]Gynecological Tumor Radiotherapy Second Ward, Tumor Hospital Affiliated to Xinjiang Medical University, Urumqi, China
[3]Department of Medical Laboratory Center, The First Huizhou Affiliated Hospital of Guangdong Medical University, Guangdong, China

Received: 21 July 2025 / Accepted: 30 December 2025
Published online: 07 January 2026

References

1. Yan H, Wang Q, Qiao Y. Cervical cancer prevention in China: Where are we now, and what's next? Cancer Biol Med. 2024;21(3):213–17.
2. Molina MA, Steenbergen RDM, Pumpe A, Kenyon AN, Melchers WJG. HPV integration and cervical cancer: a failed evolutionary viral trait. Trends Mol Med. 2024;30(9):890–902.
3. Fan J, Fu Y, Peng W, et al. Multi-omics characterization of silent and productive HPV integration in cervical cancer. Cell Genom. 2023;3(1).
4. Fan J, Lu F, Qin T, Peng W, Zhuang X, Li Y, et al. Multiomic analysis of cervical squamous cell carcinoma identifies cellular ecosystems with biological and clinical relevance. Nat Genet. 2023;55(12):2175–88.
5. Mallick S, Choi Y, Taylor AM, Cosper PF. Human papillomavirus-induced chromosomal instability and aneuploidy in squamous cell cancers. Viruses. 2024;16(4):501.
6. Balaji H, Demers I, Wuerdemann N, Schrijnder J, Kremer B, Klussmann JP, et al. Causes and consequences of HPV integration in head and neck squamous cell carcinomas: state of the art. Cancers (Basel). 2021;13(16):4089.
7. Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, 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(2):158–63.
8. Zhi W, Wei Y, Lazare C, Meng Y, Wu P, Gao P, et al. HPV-CCDC106 integration promotes cervical cancer progression by facilitating the high expression of CCDC106 after HPV E6 splicing. J Med Virol. 2023;95(1):e28009.
9. Kamal M, Lameiras S, Deloger M, Morel A, Vacher S, Lecerf C, et al. Human papilloma virus (HPV) integration signature in cervical cancer: identification of MACROD2 gene as HPV hot spot integration site. Br J Cancer. 2021;124(4):777–85. https://doi.org/10.1038/s41416-020-01153-4. Epub 2020 Nov 16. Erratum in: Br J Cancer. 2023;128(9):1790.
10. Ling J, Sun Q, Tian Q, Shi H, Yang H, Ren J. Human papillomavirus 16 E6/E7 contributes to immune escape and progression of cervical cancer by regulating miR-142-5p/PD-L1 axis. Arch Biochem Biophys. 2022;731:109449.

11. Bräutigam K, Schmidt T, Baur M, Tauber N, Kontomanolis EN, Hemptenmacher F, et al. Prevalence of PD-L1 in cervical cancer patients and the potential for combining an immune checkpoint inhibitor with lenvatinib. Anticancer Res. 2024;44(2):503–10.
12. Baek MH, Chen L, Tekin C, Cristescu R, Jin XY, Shao C, et al. Prevalence and prognostic value of PD-L1 expression and tumor mutational burden in persistent, recurrent, or metastatic cervical cancer. J Gynecol Oncol. 2024;35(6):e105.
13. Yang-Chun F, Sen-Yu W, Yuan Z, Yan-Chun H. Genome-wide profiling of human papillomavirus DNA integration into human genome and its influence on PD-L1 expression in Chinese Uygur cervical cancer women. J Immunol Res. 2020;2020:6284960.
14. Zhou L, Qiu Q, Zhou Q, Li J, Yu M, Li K, et al. Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer. Nat Commun. 2022;13(1):2563.
15. Tang X, Qi C, Zhou H, et al. Critical roles of PTPN family members regulated by non-coding RNAs in tumorigenesis and immunotherapy. Front Oncol. 2022;12:972906.
16. Cheng C, Zhang Z, Wang J, et al. CircPGM5 regulates Foxo3a phosphorylation via MiR-21-5p/MAPK10 axis to inhibit bladder cancer progression. Cell Signal. 2024;121:111297.
17. Shen J, Wang Q, Liang C, et al. Novel insights into miR-944 in cancer. Cancers. 2022;14(17):4232.
18. Cui X, Lin Q, Chen M, Wang Y, Wang Y, Wang Y, et al. Long-read sequencing unveils novel somatic variants and methylation patterns in the genetic information system of early lung cancer. Comput Biol Med. 2024;171:108174.
19. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023;24(6):430–47.
20. Dugo E, Piva F, Giulietti M, Giannella L, Ciavattini A. Copy number variations in endometrial cancer: from biological significance to clinical utility. Int J Gynecol Cancer. 2024;34(7):1089–97.
21. Ohsawa K, Momose S, Nishikori A, Nishimura MF, Gion Y, Sawada K, et al. Copy number analysis of 9p24.1 in classic Hodgkin lymphoma arising in immune deficiency/dysregulation. Cancers (Basel). 2024;16(7):1298.
22. Huang RSP, Murugesan K, Montesion M, Pavlick DC, Mata DA, Hiemenz MC, et al. Pan-cancer landscape of *CD274* (PD-L1) copy number changes in 244 584 patient samples and the correlation with PD-L1 protein expression. J Immunother Cancer. 2021;9(5):e002680.
23. Bi Y, Hu J, Zeng L, Chen G, Cai H, Cao H, et al. Characteristics of HPV integration in cervical adenocarcinoma and squamous carcinoma. J Cancer Res Clin Oncol. 2023;149(20):17973–86.
24. Xiao H, Cheng G, Zhang H, Liu Y, Chen Z, Gao Y, et al. Role of KLF5 in enhancing ovarian cancer stemness and PARPi resistance: mechanisms and therapeutic targeting. J Transl Med. 2025;23(1):492.
25. Zeng X, Wang Y, Liu B, Rao X, Cao C, Peng F, et al. Multi-omics data reveals novel impacts of human papillomavirus integration on the epigenomic and transcriptomic signatures of cervical tumorigenesis. J Med Virol. 2023;95(5):e28789.
26. Wu Q, Liu Z, Gao Z, Luo Y, Li F, Yang C, et al. KLF5 inhibition potentiates anti-PD1 efficacy by enhancing CD8 T-cell-dependent antitumor immunity. Theranostics. 2023;13(4):1381–400.
27. Chen CH, Yang N, Zhang Y, Ding J, Zhang W, Liu R, et al. Inhibition of super enhancer downregulates the expression of KLF5 in basal-like breast cancers. Int J Biol Sci. 2019;15(8):1733–42.
28. Liu R, Zhi X, Zhou Z, Zhang H, Yang R, Zou T, et al. Mithramycin a suppresses basal triple-negative breast cancer cell survival partially via down-regulating Krüppel-like factor 5 transcription by Sp1. Sci Rep. 2018;8(1):1138.
29. Wang X, Qiu T, Wu Y, Yang C, Li Y, Du G, et al. Arginine methyltransferase PRMT5 methylates and stabilizes KLF5 via decreasing its phosphorylation and ubiquitination to promote basal-like breast cancer. Cell Death Differ. 2021;28(10):2931–45.
30. Jiang D, Qiu T, Peng J, Li S, Tala, Ren W, et al. YB-1 is a positive regulator of KLF5 transcription factor in basal-like breast cancer. Cell Death Differ. 2022;29(6):1283–95.
31. Kong Y, Ren W, Fang H, Shah NA, Shi Y, You D, et al. Histone deacetylase inhibitors (HDACi) promote KLF5 ubiquitination and degradation in basal-like breast cancer. Int J Biol Sci. 2022;18(5):2104–15.
32. Zhao Q, Yang S, Hao S, Chen Z, Tang L, Wu Z, et al. Identification of transcriptionally-active human papillomavirus integrants through nanopore sequencing reveals viable targets for gene therapy against cervical cancer. J Med Virol. 2024;96(6):e29769.
33. Cui X, Li Y, Zhang C, Qi Y, Sun Y, Li W. Multiple HPV integration mode in the cell lines based on long-reads sequencing. Front Microbiol. 2023;14:1294146.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.