Original Research

Identification of differentially expressed genes and pathways for risk stratification in HPV-associated cancers governing different anatomical sites

Eun Jung Kwon[1,†], Hye Ran Lee[2,†], Ju Ho Lee[2], Chorong Seo[2], Mihyang Ha[1], Jin Roh[3], Yun Hak Kim[4,5,*], Jeon Yeob Jang[2,6,*]

[1]Interdisciplinary Program of Genomic Science, Pusan National University, 50612 Yangsan, Republic of Korea
[2]Department of Otolaryngology, Ajou University School of Medicine, 16499 Suwon, Republic of Korea
[3]Department of Pathology, Ajou University School of Medicine, 16499 Suwon, Republic of Korea
[4]Department of Anatomy, School of Medicine, Pusan National University, 50612 Yangsan, Republic of Korea
[5]Department of Biomedical Informatics, School of Medicine, Pusan National University, 50612 Yangsan, Republic of Korea
[6]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, 16499 Suwon, Republic of Korea
*Correspondence: yunhak10510@pusan.ac.kr (Yun Hak Kim); manup1377@gmail.com (Jeon Yeob Jang)
[†]These authors contributed equally.

Abstract

Background: Human papillomavirus (HPV) is the major cause of cervical cancer (CC) etiology; its contribution to head and neck cancer (HNC) incidence is steadily increasing. As individual patients' response to the treatment of HPV-associated cancer is variable, there is a pressing need for the identification of biomarkers for risk stratification that can help determine the intensity of treatment. **Methods**: We have previously reported a novel prognostic and predictive indicator (HPPI) scoring system in HPV-associated cancers regardless of anatomical location by analyzing The Cancer Genome Atlas and Gene Expression Omnibus databases. In the present study, we comprehensively investigated the association of group-specific expression patterns of common differentially expressed genes (DEGs) between high- and low-risk groups in HPV-associated CC and HNC, identifying molecular biomarkers and pathways for risk stratification. **Results**: Among the 174 identified DEGs, the expression of genes associated with extracellular matrix (ECM)-receptor interaction pathway (*ITGA5, ITGB1, LAMB1,* and *LAMC1*) was increased in high-risk groups in both HPV-associated CC and HNC, while the expression of genes associated with T-cell immunity (*CD3D, CD3E, CD8B, LCK,* and *ZAP70*) was decreased and *vice versa*. The individual genes showed significant prognostic impact on HPV-associated cancers but not on HPV-negative cancers. The expression levels of identified genes were similar between HPV-negative and HPV-associated high-risk groups with distinct expression patterns only in HPV-associated low-risk groups. Each group of genes showed negative correlations and distinct patterns of immune cell infiltration in tumor microenvironments. **Conclusions**: These results allowed us to identify molecular biomarkers and pathways for risk stratification in HPV-associated cancers regardless of anatomical location. The identified targets were found to be selectively working in only HPV-associated cancers and not in HPV-negative cancers, indicating the possibility of selective targets governing HPV-infective tumor microenvironments.

Keywords: Human papillomavirus; Risk stratification; Head and neck cancer; Cervical cancer; Prognostic biomarker

1. Introduction

Human papillomavirus (HPV) is a small circular virus with approximately 8 kb of double-stranded DNA genome comprising the following three major functional regions: (i) an upstream regulatory region (URR) with transcription factor-binding sites controlling gene expression; (ii) an early region encoding six genes (E1, E2, E4, E5, E6, E7) having multiple functions including viral replication, and (iii) a late region encoding capsid proteins L1 and L2 that produce the virion by self-assembly [1,2]. HPV induces a productive infection by targeting the epithelial cells in the basal layer of skin and mucosa [1]. More than 150 HPV genotypes have been discovered to date, and the majority show subclinical manifestation or cause the growth of benign lesions ranging in severity from self-limiting to

debilitating [3]. However, the chronic infection of some oncogenic high-risk types of HPV, especially types 16 and 18, results in cancer progression [4,5]. In more than 95% of invasive cervical cancers (CC) and most cases of HPV-associated head and neck cancers (HNC), the incorporation of high-risk HPV genomes into the host genome was detected [6]. Among HNCs, the rate of HPV association in oropharyngeal squamous cell carcinoma (OPSCC) has increased and is reported to have reached 80% in recent years [2]. HPV-associated OPSCC is considered a unique disease entity as its clinical and molecular characteristics are distinct from those of non-HPV-related OPSCC; it also presents a good overall prognosis with better response to treatment, including radiation [7,8]. Persistent infection of HPV also acts as an important etiology of other anogenital
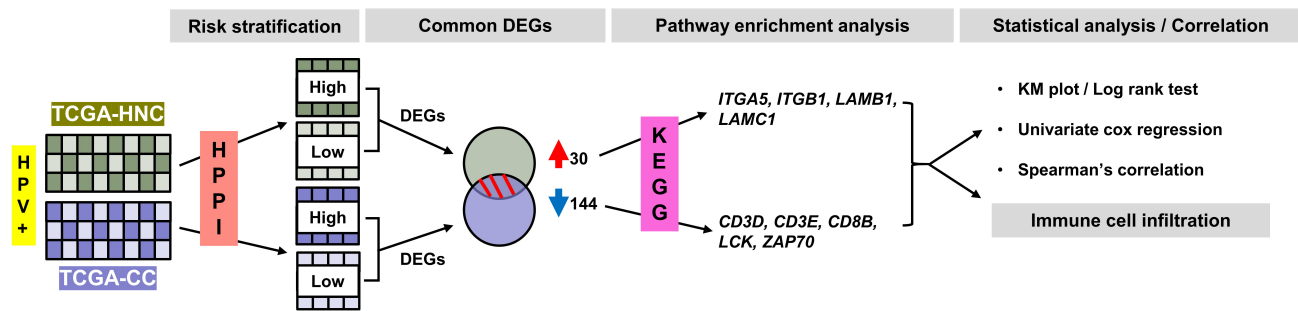
Fig. 1. Overall flow chart of the study. HNC, head and neck cancer; CC, cervical cancer; HPPI, HPV-related prognostic and predictive indicator; DEGs, differentially expressed genes.

tract malignancies such as anal, vaginal, and penile, making it a notable public health issue [9]. Therefore, more advanced treatment strategies that consider the additional prognostic factors for each patient require urgent development.

In HPV-related cancers, high-risk type HPV continuously expresses representative oncoproteins E6 and E7 that primarily induce cellular transformation through the decomposition of p53 and pRb, respectively [10,11]. The accumulation of cellular mutations by which E6 and E7 promote cell-cycle progression, chromosomal instability, and integration of foreign DNA, which ultimately contribute to carcinogenesis have also been elucidated [1,10,12]. However, the precise molecular mechanisms involved in oncogenesis and clinical expression, or the interaction of diverse genomes at each stage, are still not clearly understood. These processes are also dynamically influenced by additional factors, such as the location of tumor occurrence including the peritumoral microenvironment and the intrinsic factors of each patient, which intensifies gene expression heterogeneity and makes it difficult to accurately identify disease progression patterns and prognostic stratification [13,14]. In recent years, cancer bioinformatics has been actively developing as a new strategy that combines molecular biology and *in silico* information engineering based on gene sequencing data to discover clinically meaningful biomarkers related to treatment response and prognosis in each tumor and apply it to clinical practice [15,16]. More appropriate personalized treatment will become possible as our knowledge of the gene expression profiles and molecular behavior heterogeneity in HPV-related cancers improves through such analyses.

Genomic changes that can be important causes in the development of cancer (such as an abnormal gene expression) stimulate the production of tumor antigens by the immune system and trigger a cellular immune response [17]. At this moment, immune cells, such as T cells infiltrate into the tumor microenvironment and control tumor progression [17]. This strongly indicates that gene expression can have a significant effect on immune cell infiltration levels. Immune cell infiltration is known to be important for the sur-

vival of cancer patients and is regulated by chemokines, but the mechanism is not yet fully elucidated [18,19]. Therefore, we expect that the combination of immune cell infiltration and genetic biomarkers will help us understand the development of cancer.

Previously, we analyzed common HPV-specific prognostic gene signatures from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, according to HPV-related cancers in different anatomic sites of the head and neck, and cervix [20]. Using these results, we developed an HPV-related prognostic and predictive indicator (HPPI) scoring system that can perform risk stratification using the 34 selected genes, and demonstrated that there was a significant difference in the prognosis. In the present study, we investigated differentially expressed genes (DEGs) in the high and low-risk groups of HPV-related HNC and CC according to HPPI scoring, and aimed to identify genes that influence clinical manifestations, including survival and their expression profiles. The study workflow is illustrated in Fig. 1.

2. Materials and methods

2.1 Patient data collection and risk stratification

Clinical data and RNA-seq gene expression profiles of TCGA HPV-related HNC and CC were downloaded from Broad GDAC Firehose (http://gdac.broadinstitute.org/). We chose the sample type of primary solid tumor and gene expression data generated an Illumina HiSeq 2000 RNA-Seq platform (level 3) using the preprocessed RNAseqV2 normalized count expression values based on RNA-Seq by expectation-maximization (RSEM). Using the HPPI risk scoring system, we divided patients into high-risk (HNC = 27, CC = 43) and low-risk groups (HNC = 70, CC = 226) [20]. Patient characteristics are presented in Table 1.

2.2 Identification of common DEGs

We identified DEGs between the high-risk and low-risk groups for each cohort using "*siggenes 1.60.0*" in the R package using the Significance Analysis of Microarrays (SAM) method. We then identified the common DEGs that exist in both cohorts. We used the cutoff value for False

Table 1. Patient characteristics.

		TCGA-HNC		TCGA-CC	
		(N = 518)		(N = 290)	
		HPV (+)	HPV (−)	HPV (+)	HPV (−)
		97	421	269	21
Age	<55	37	112	196	11
	≥55	60	309	73	10
	mean ± sd	57.8 ± 9.6	61.6 ± 12.2	47.5 ± 13.7	55.5 ± 13.5
Clinical	I	3	24	148	11
Stage	II	11	62	57	6
	III	10	71	40	1
	IV	42	223	18	3
Clinical	N0	30	213	-	-
N stage	N1	10	72	-	-
	N2	52	110	-	-
	N3	3	6	-	-
Clinical	M0	91	388	-	-
M stage	M1	1	5	-	-
Gender	Female	11	124	-	-
	Male	86	297	-	-
Risk group	High	27	-	43	-
	Low	70	-	226	-

Discovery Rate (FDR) = 0.01 and set 1.7 fold change (FC) values to define upregulated and downregulated genes. We used the "*gplots 3.1.1*" and "*ggplot2 3.3.3*" R (version 4.0.2) packages to visualize the results.

2.3 KEGG pathway enrichment analysis and selection of target genes

Pathway enrichment analysis based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to confirm the biological function of common DEGs using NetworkAnalyst 3.0 (v2019) (https://www.networkanalyst.ca/). We selected genes that clustered at least three times in the top four enrichment pathways with the most significant *p*-values.

2.4 Analysis of tumor-infiltrating immune cells

To better understand the selected genes and tumor immune microenvironment, we investigated the correlation of the selected gene expression with tumor purity and tumor-infiltrating immune cells using the Tumor Immune Estimation Resource (TIMER) version 2.0 (http://timer.cistrome.org/). The TIMER database provides several analyses of immune infiltration of tumors from the TCGA database. Of the 32 TCGA tumor data included in the TIMER database, HPV (+) HNC was used for analysis. Although we analyzed six immune cell types (CD8+ T cells, CD4+ T cells, B cells, macrophages, neutrophils, dendritic cells), only significant results with Rho >0.5 were selected. The correlation analysis was performed by the purity-adjusted partial Spearman's correlation method.

2.5 Immunohistochemical staining

Ten pathologic slides from oropharyngeal cancer surgical specimens were evaluated by immunohistochemical staining for Intergrin alpha 5 (ITGA5). Immunohistochemical staining was performed using the Bench Mark XT (Ventana Medical Systems, AZ, USA) according to the manufacturer's instructions. Briefly, 4-μm-thick sections were deparaffinized in xylene and dehydrated in graded ethanol solutions. Antigen retrieval with the cell conditioning buffer 1 (CC1) was performed for 40 min. The sections were incubated with the primary antibody, Intergrin alpha 5 (rabbit monoclonal, Abcam, Cambridge, UK), for 1 h and with the secondary anti-HRP antibody (Abcam, Cambridge, UK) for 32 min. Counterstaining was performed with hematoxylin II (Sigma-Aldrich, St. Louis, MO, USA) for 10 min and subsequent bluing for 6 min. The whole-tissue section slides were examined under a light microscope.

2.6 Statistical analysis

After setting the cutoff as the median gene expression value for each selected gene, we performed the Kaplan-Meier (KM) analysis for computing survival, log rank test, and univariate cox regression analysis to predict overall survival (OS) and estimate the accuracy using the R packages: "*survival 3.1-12*" and "*survminer 0.4.9*". The cor-
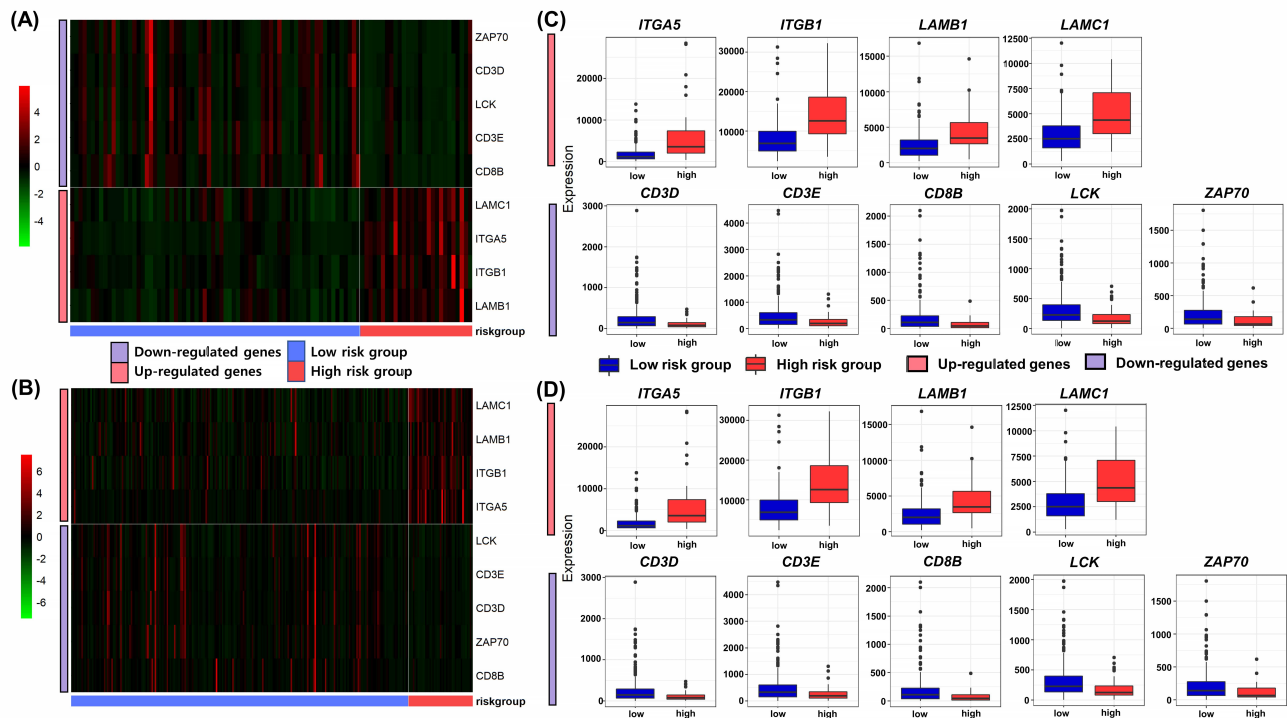
Fig. 2. Analysis of differentially expressed genes (DEGs) between HPPI high-risk and low-risk groups. Heatmaps visualizing DEGs between two groups: (A) HPV (+) HNC, and (B) HPV (+) CC. Boxplots visualizing DEGs between two groups: (C) HPV (+) HNC, and (D) HPV (+) CC.

relation between the selected target genes was evaluated using Spearman's correlation analysis. The threshold for significant correlation was set at $p < 0.05$ and Spearman's correlation coefficient was >0.3. Comparisons of ITGA5 expression was performed by Mann-Whitney tests. All statistical analyses were performed using R (version 4.0.2) or SPSS (version 18) (Chicago, IL, USA).

3. Results

3.1 Common DEGs between high and low-risk groups in HNC and CC

According to a previous study [20], we divided HPV (+) HNC and HPV (+) CC into high-risk and low-risk groups, respectively and identified DEGs in each, as well as DEGs that were common in both cohorts and defined them as common DEGs. We identified 174 common DEGs, of which 30 were upregulated and 144 were downregulated (**Supplementary Table 1**). KEGG pathway enrichment analysis was performed to select target genes that explain the molecular mechanisms among common DEGs, and four upregulated genes (*ITGA5, ITGB1, LAMB1,* and *LAMC1*), and five downregulated genes (*CD3D, CD3E, CD8B, LCK,* and *ZAP70*) were selected (Fig. 2A–D, Tables 2 and 3).

3.2 Prognostic significance of common DEGs in HNC and CC

To confirm the association between the OS and expression levels of target genes, we performed KM analysis

and log rank test on the HPV (+) HNC and HPV (+) CC cohorts (Fig. 3A,B). High levels of gene expression of *ITGA5, ITGB1, LAMB1,* and *LAMC1* were associated with a worse OS, while low levels of gene expression of *CD3D, CD3E, CD8B, LCK,* and *ZAP70* were associated with a worse OS of patients with HPV-related cancers (Fig. 3A,B). Univariate cox regression analysis revealed nine target genes as independent prognostic factors for HPV-related cancers (Table 4). However, these target genes had no prognostic effects on HPV (–) cancers (**Supplementary Table 2**).

3.3 Expression patterns of common DEGs in HNC and CC

To determine the expression patterns of the target genes of normal, HPV (–), and risk-stratified HPV (+) cancers, we compared the relative expression levels of the target genes between the following four groups: normal, HPV (–), HPV (+) low-risk group, and HPV (+) high-risk group (Fig. 4). Interestingly, the expression patterns of the target genes were similar in the HPV (–) and HPV (+) high-risk groups in both cancer cohorts. We found that four upregulated genes (*ITGA5, ITGB1, LAMB1,* and *LAMC1*) among the nine target genes showed high expression in HPV (–) and HPV (+) high-risk groups but relatively low expression in HPV (+) low-risk and normal groups. In contrast, the remaining five downregulated genes (*CD3D, CD3E, CD8B, LCK,* and *ZAP70*) showed low expression in HPV (–) and HPV (+) high-risk groups, but relatively high expression in HPV (+) low-risk group (Fig. 4A,B). To validate the dif-
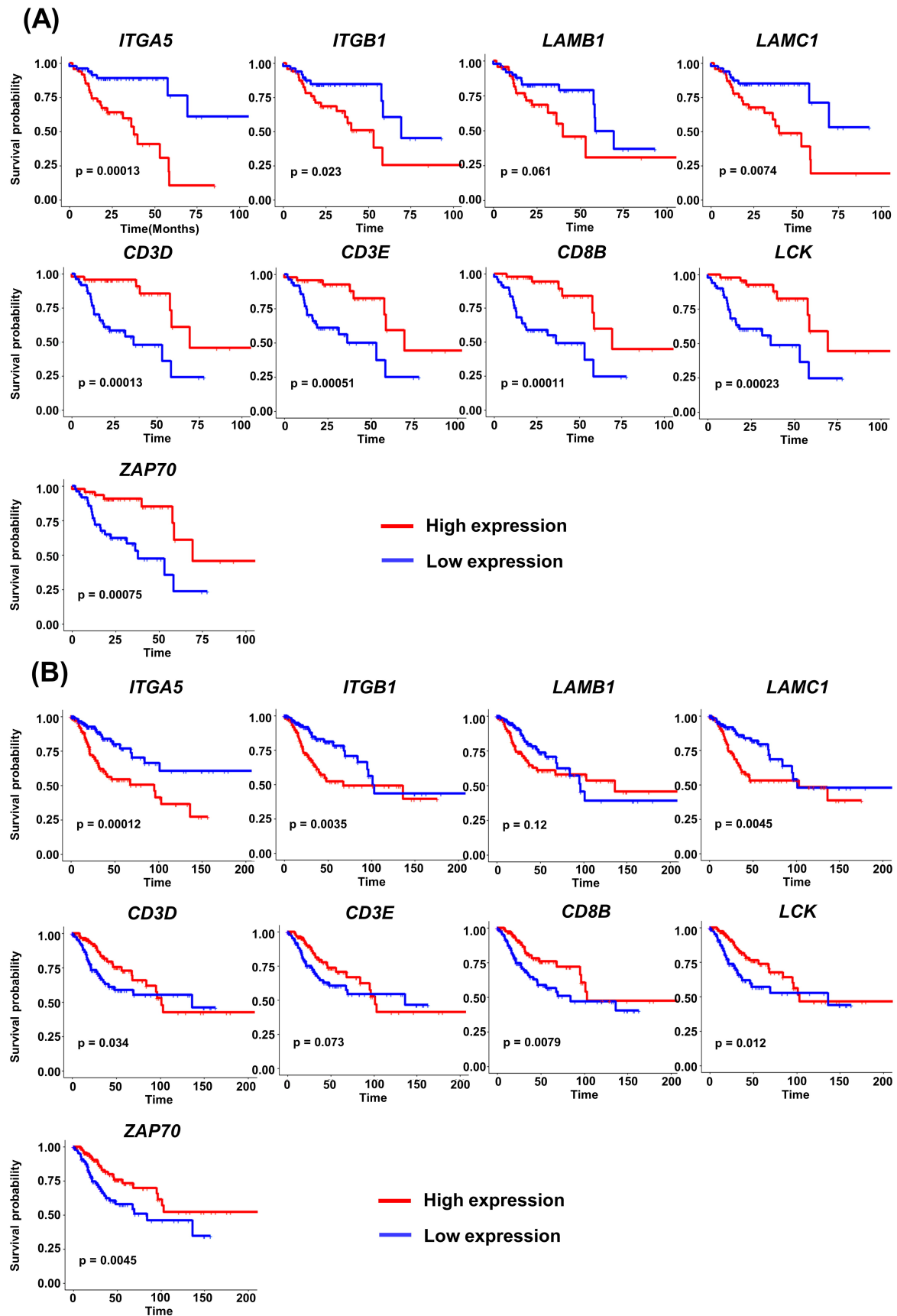
Fig. 3. Kaplan-Meier survival analysis based on the expression of target genes. (A) HPV (+) HNC, and (B) HPV (+) CC. P, log rank test *p*-value.
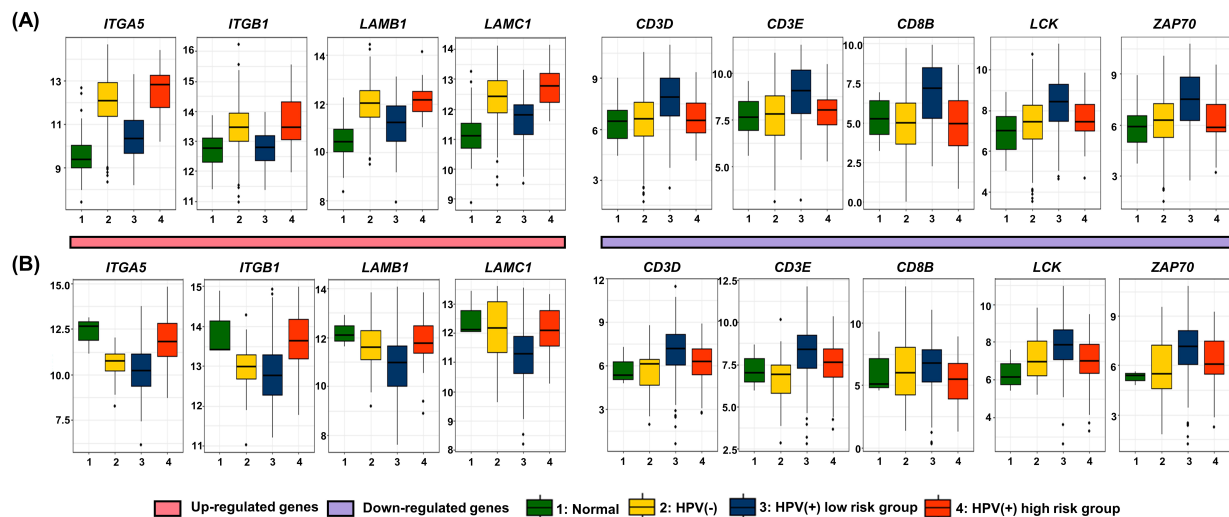
Fig. 4. Comparison of the relative expression levels of target genes between the following four groups: normal, HPV (–), HPV (+) low-risk group, and HPV (+) high-risk group. (A) HNC, and (B) CC. The boxplot shows the distribution of the data and the median (middle line). The dot means outlier, which is more or less than 3/2 times of upper or lower quartiles. Log2 transformation is used to draw the plot for the gene expression level of the data.

Table 2. Results of KEGG pathway enrichment analysis.

	Pathway	*p*-value	FDR	Genes
Upregulated genes	ECM-receptor interaction	1.94×10^{-5}	0.00379	*ITGA5, ITGB1, LAMB1, LAMC1*
	Small cell lung cancer	3.19×10^{-5}	0.00379	*ITGB1, LAMB1, LAMC1, CDK6*
	Focal adhesion	3.71×10^{-5}	0.00379	*ITGA5, ITGB1, LAMB1, LAMC1, SHC1*
	PI3K-Akt signaling pathway	4.77×10^{-5}	0.00379	*ITGA5, ITGB1, LAMB1, LAMC1, CDK6, OSMR*
Downregulated genes	Primary immunodeficiency	1.25×10^{-6}	0.000398	*CD3D, CD3E, CD8B, LCK, ZAP70*
	Th1 and Th2 cell differentiation	0.000114	0.0141	*CD3D, CD3E, LCK, ZAP70, IL12A*
	Hematopoietic cell lineage	0.000146	0.0141	*CD3D, CD3E, CD8B, CD2, FLT3LG*
	T-cell-receptor signaling pathway	0.000177	0.0141	*CD3D, CD3E, CD8B, LCK, ZAP70*

Table 3. Pathway enrichment analysis–Gene list.

	Gene symbol	Full name
Upregulated genes	*ITGA5*	*Integrin Subunit Alpha 5*
	ITGB1	*Integrin Subunit Beta 1*
	LAMB1	*Laminin Subunit Beta 1*
	LAMC1	*Laminin Subunit Gamma 1*
Downregulated genes	*CD3D*	*CD3d Molecule*
	CD3E	*CD3e Molecule*
	CD8B	*CD8b Molecule*
	LCK	*LCK Proto-Oncogene, Src Family Tyrosine Kinase*
	ZAP70	*Zeta Chain Of T Cell Receptor Associated Protein Kinase 70*

ferential expression of the ECM-related gene in different tumor microenvironments, we further performed immuno-histochemical staining of ITGA5 in 10 available oropha-ryngeal cancer specimens. The ITGA5 was mainly ex-pressed in tumor stromal regions rather than tumor cell nests (Fig. 5). Importantly, ITGA5 expression was significantly increased in tumor stromal regions of HPV (–) cancers com-

pared with HPV (+) low-risk cancers, indicating the evi-dences for the differential expression patterns of ITGA5.

3.4 Correlation analysis of common DEGs in HNC and CC

In order to determine if there were any correlations between the target genes, we conducted a Spearman's cor-relation analysis. We found a high correlation between

Table 4. Univariate cox regression in HPV (+) cancers.

Dataset	Genes		p-value	Correlation coefficient	HR	95% CI	
HPV (+) HNC	Upregulated	ITGA5	0.000465 ***	1.5377	4.6537	1.967	11.01
		ITGB1	0.0272 *	0.8657	2.3767	1.103	5.123
		LAMB1	0.0653	0.6954	2.0044	0.9569	4.199
		LAMC1	0.0103 *	1.0245	2.7857	1.273	6.095
	Downregulated	CD3D	0.000453 ***	−1.533	0.216	0.09171	0.5086
		CD3E	0.00118 **	−1.3572	0.2574	0.1134	0.5844
		CD8B	0.000417 ***	−1.5371	0.215	0.09156	0.5049
		LCK	0.000638 ***	−1.4263	0.2402	0.1059	0.5446
	ZAP70		0.00161 **	−1.3149	0.2685	0.1186	0.6078
HPV (+) CC	Upregulated	ITGA5	0.000219 ***	0.9798	2.6639	1.584	4.479
		ITGB1	0.00424 **	0.7334	2.0822	1.259	3.442
		LAMB1	0.12	0.391	1.4785	0.9033	2.42
		LAMC1	0.00537 **	0.722	2.0585	1.238	3.422
	Downregulated	CD3D	0.0365 *	−0.5273	0.5902	0.3601	0.9674
		CD3E	0.0755	−0.4455	0.6405	0.3919	1.047
		CD8B	0.00906 **	−0.6748	0.5092	0.3068	0.8453
		LCK	0.0132 *	−0.6225	0.5366	0.328	0.878
		ZAP70	0.00534 **	−0.7096	0.4919	0.2986	0.8103

genes with the same expression patterns but no high correlation between genes with different expression patterns (Fig. 6A,B). All the downregulated genes had a high correlation and among the upregulated genes, *ITGA5* and *ITGB1*, *ITGB1* and *LAMC1* had high correlations (|coefficient| >0.6) (Fig. 6A,B).

3.5 Correlation analysis between common DEGs and immune cell infiltration in HPV (+) HNC

As the pattern of immune infiltrate is an important aspect of tumor survival and progression, we explored the correlation between the expression levels of the target genes and immune cell infiltration using data for several immune cell subtypes and The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma (TCGA-HNSC) HPV (+) from TIMER 2.0 (Fig. 7A–E). The expression levels of the downregulated target genes were positively correlated with CD4+ T cells and neutrophils (Fig. 7A–E), while immune cells correlated with upregulated target genes were not found.

4. Discussion

Investigation of genomic and molecular biomarkers that possess prognostic importance, including for the prognosis of HPV-related cancer, is still underway. The need for the identification of specific prognostic factors has emerged in HPV-related OPSCC, which includes a subgroup of patients with aggressive behavior, although the majority show favorable treatment responses [21]. In this study, risk stratification of HPV-related HNC and CC was performed using the HPPI scoring system developed in our previous study

[20]. Of the identified DEGs, the expression of the four upregulated and five downregulated target genes was compared with the normal and HPV (−) cohort. A particularly noteworthy finding was that both HPV (−) and HPV (+) high-risk groups in HNC showed similar patterns of expression of the target genes, and different expression patterns only appeared in the HPV (+) low-risk group (Fig. 4). In the case of de-escalating treatment trials in HPV-related OPSCC, as part of an effort to reduce the toxicity caused by chemoradiation and to perform less invasive surgery for improving the quality of life [21,22], it is important to accurately select the low-risk group by risk stratification of HPV (+) HNC. Our study can be instrumental in these instances as it can provide genomic data for the development of patient-specific treatment, which are important in both cancer biology and other clinical applications.

The extracellular matrix (ECM) is an essential milieu of tissues, consisting of a dynamic network of biochemical components [23,24]. ECM contributes to the interaction of cells with their microenvironments, which plays a crucial role in regulating cellular behavior and maintaining organ homeostasis. It is known that the disruption of these control processes can cause diseases such as cancer; thus, the expression of ECM-related genes has been studied [24,25]. It has been recently reported that the expression of ECM genes differs according to HPV association in OPSCC and with high-grade dysplasia and invasive cancer [26]. In this current study, we focused on laminin genes *LAMB1* and *LAMC1*, and their receptors, integrin genes *ITGA5* and *ITGB1* among 30 upregulated DEGs that participate in ECM-receptor interaction in the KEGG path-
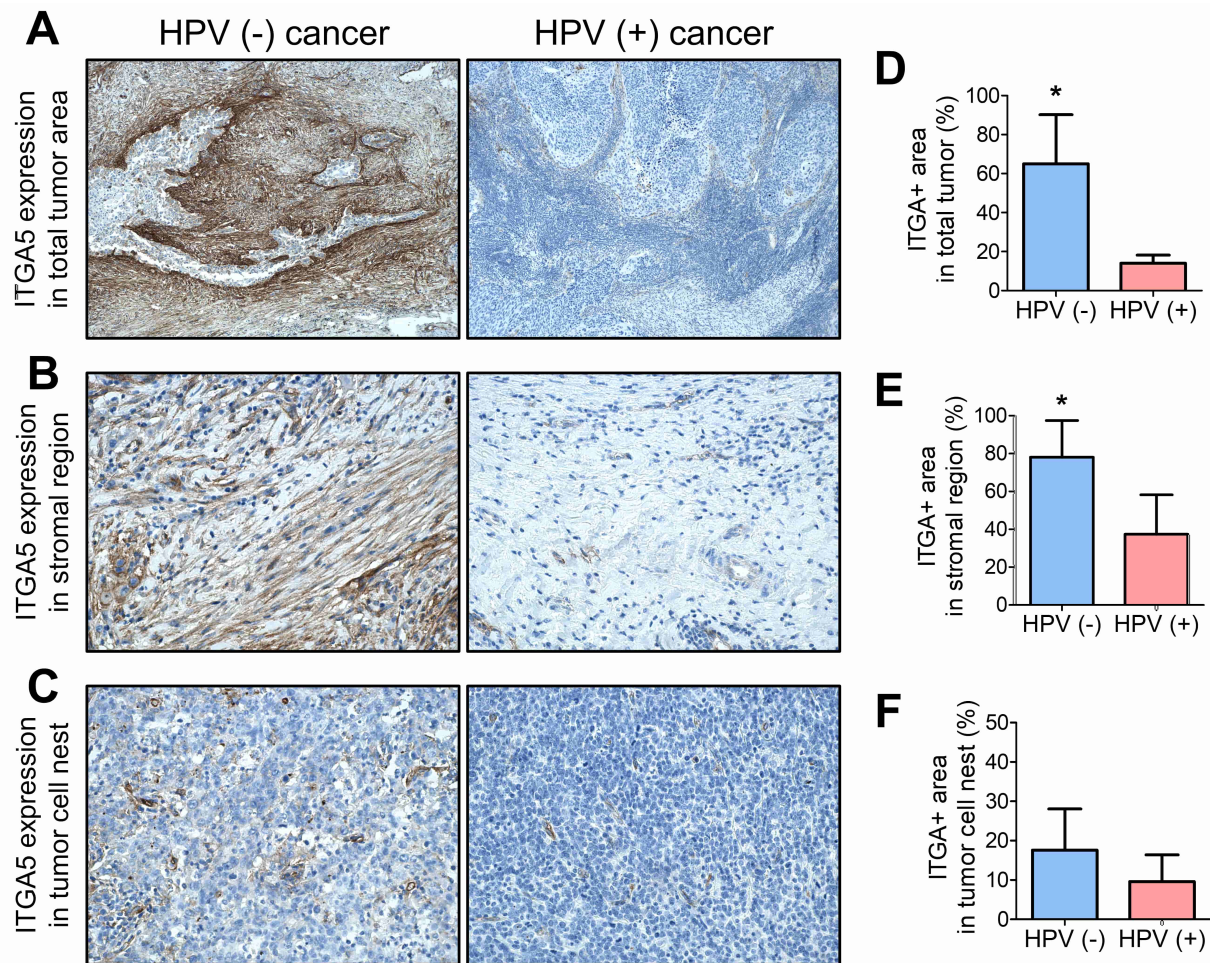
Fig. 5. ITGA5 expression analysis in patient samples of HPV (–) and HPV (+) oropharyngeal cancers. (A–C) Representative images of immunohistochemical staining showing ITGA5 expression patterns in total tumor area, stromal region, and tumor cell nests, respectively. (D–F) Comparisons of ITGA5+ area (%) between HPV (–) and HPV (+) oropharyngeal cancers in total tumor area, stromal region, and tumor cell nests, respectively. Each group, n = 5. *$p < 0.05$ versus HPV-positive cancers.

way enrichment analysis (Table 2) [25]. Laminin is a major component of the basement membrane and is essential for the adhesive interaction between the cell and the basement membrane; laminins bind laminin-binding integrin heterodimers [27]. LAMB combines with LAMC; together, they initiate cell assembly to promote cancer invasion and gene expression in various invasive cancers, including *LAMB1* gene expression in head and neck, liver, and uterine endometrial cancers, and *LAMC1* gene expression in gastric cancer [28–30]. Integrin subunits $\alpha 5$ and $\beta 1$ primarily conjoin to form $\alpha 5 \beta 1$ heterodimers, and *ITGA5* and *ITGB1,* the genes encoding $\alpha 5$ and $\beta 1$, respectively have been suggested to be associated with tumor aggressive behavior, such as cancer progression and treatment resistance [31,32]. In our study, an increase in *ITGA5* and *ITGB1* expression was characteristically identified in high-risk HPV (+) and HPV (–) groups in HNC, and was associated with the worsening OS in both HPV-related HNC and CC. Therefore, we postulate that *LAMB1, LAMC1, ITGA5,*

and *ITGB1* genes may be potential biomarkers for high risk and poor prognosis in HNC and CC, especially in HPV-related cancers.

Other components of the tumor microenvironment, including the expression of genetic markers, and distribution of immune cells and molecules that regulate cellular function, have been studied for targeted immunotherapy [33,34]. Of 144 downregulated DEGs in our study, five were *CD3D* and *CD3E* as T cell markers, CD8+ T cell-specific marker *CD8B*, *LCK*, and *ZAP70,* which participates in primary immunodeficiency and the T cell-receptor signaling pathway in the KEGG pathway analysis (Table 2) [33]. *LCK* and *ZAP70* are protein tyrosine kinases, and *ZAP70* activated by *LCK* plays an important role in initiating intracellular signaling pathways related to T cell antigenic receptor in the early stages of T cell activation [35,36]. A meta-analysis of several solid tumors revealed that greater infiltration of CD3+ and CD8+ T cells had a positive effect on survival, and a recent report on HNC re-
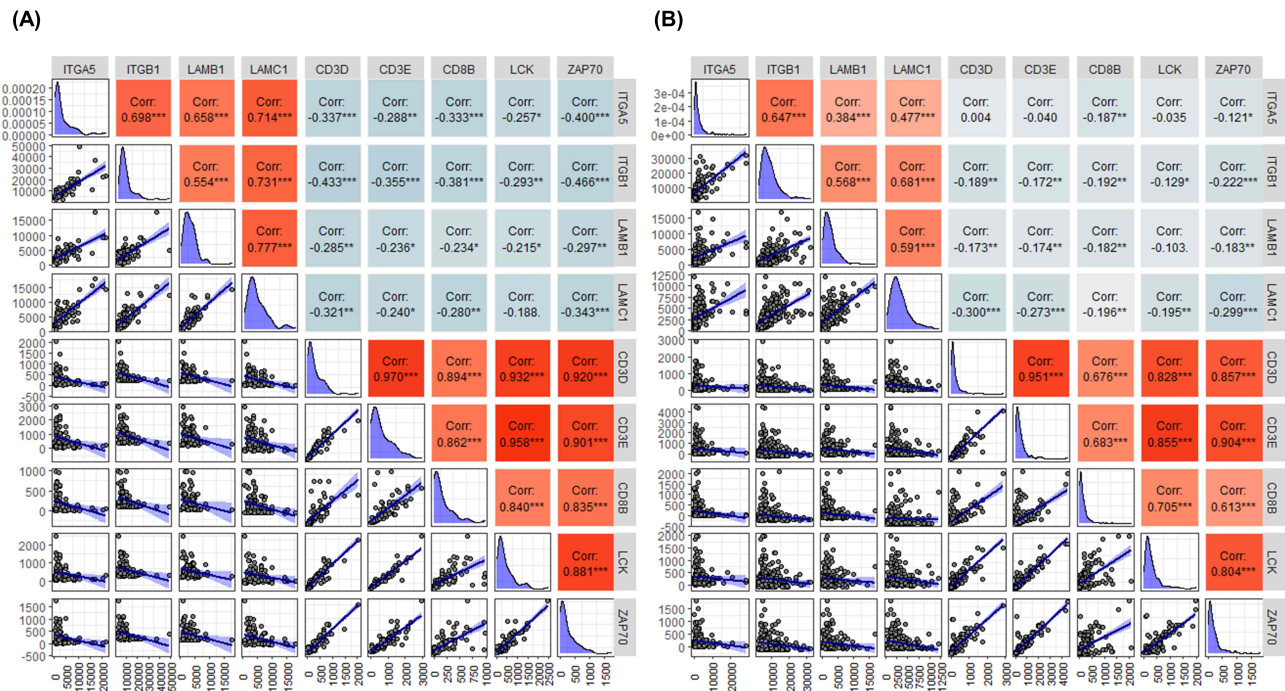
8

Fig. 6. Spearman's correlation analysis is shown by multi-correlation plots. The distribution of each gene expression is displayed on the diagonal. At the top of the diagonal, corr denotes correlation coefficient. The greater the absolute value of the number, the greater the relevance. Blue color indicates a negative correlation, and red color indicates a positive correlation. The darker the color, the higher the correlation. Asterisks indicate statistically significant p-value (***, $p < 0.001$: **, $p < 0.01$: *, $p < 0.05$). The bottom of the diagonal shows scatterplots with a fitted line. (A) HPV (+) HNC, and (B) HPV (+) CC.
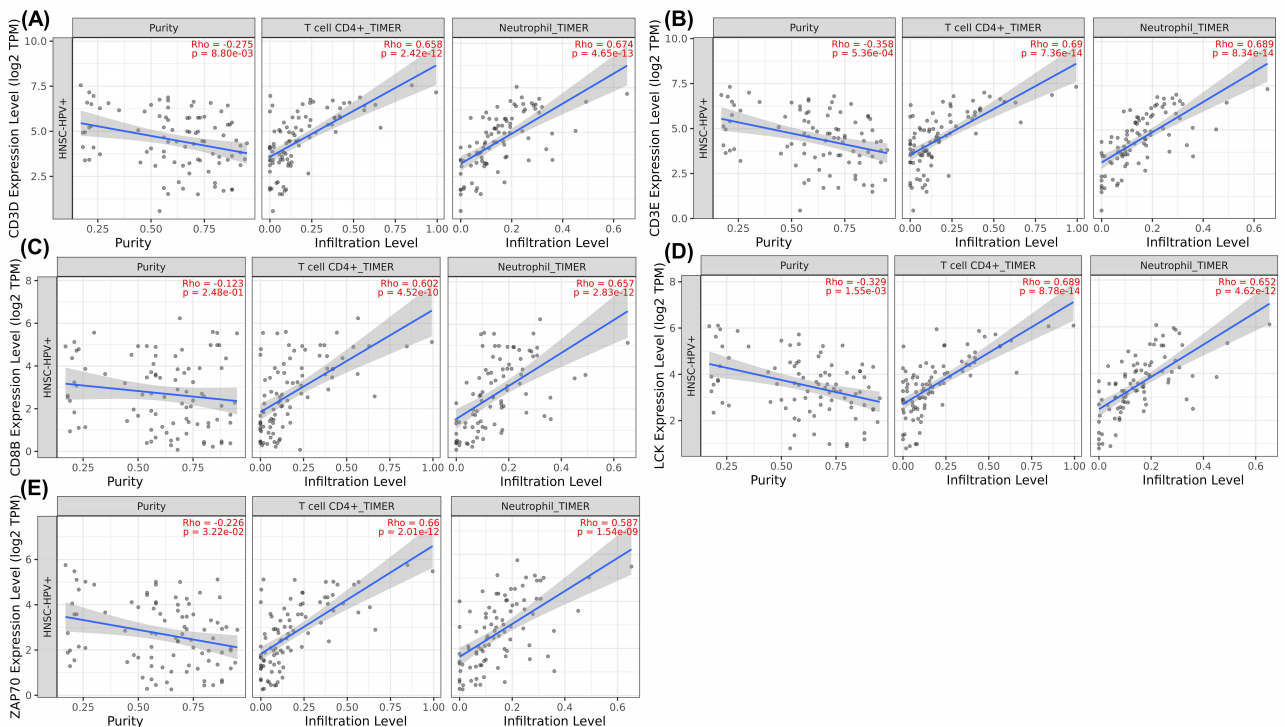

Fig. 7. Correlation of gene expression with levels of immune cell infiltration in HPV (+) HNC. Of the various molecular subtypes, only those that are commonly significant in the downregulated group are shown (p-value < 0.05, Rho >0.5). The y-axis represents the level of gene expression, and the x-axis represents the tumor purity and the infiltration level of immune cells. (A) CD3D, (B) CD3E, (C) CD8B, (D) LCK, and (E) ZAP70.

vealed that more tumor-infiltrating immune cells were associated with better prognosis [34,37]. An increase in T cell infiltrates, which emphasizes the role of immune system, was suggested as an important prognostic determinant contributing to tumor suppression, especially in HPV-related OPSCC [38]. In our study, contrary to ECM genes, the expression of T cell-related genes was significantly higher in the low-risk HPV (+) group and lower in the high-risk HPV (+) and HPV (−) groups in HNC, and higher expression improved OS in HPV-related HNC and CC. Therefore, in addition to ECM genes, T cell-related genes *CD3D*, *CD3E*, *CD8B*, *LCK*, and *ZAP70* could be useful biomarkers for risk stratification of HPV-related cancer.

We also performed correlation analyses between the selected target genes in the HPV-related HNC and CC groups. A positive correlation was observed between DEGs with the same expression pattern, while a negative correlation was observed between upregulated ECM genes and downregulated T cell-related genes. It has been proposed that cellular components, including immune cells in the tumor microenvironment, contribute to the regulation of ECM dynamics [24]. A study on HNC that investigated the tumor ecosystem with a single-cell transcriptome proved that heterogeneous cellular components, including T cells, influence the expression of ECM markers related to tumor migration, such as laminin, by active interaction [39]. This study also suggests the possibility that different ECM pathways and immunity within the tumor microenvironment are complexly intertwined with cancer biology and thus require further study.

p16 is widely used as an important biomarker suggesting the presence or absence of HPV in cancer [40]. The HPV E7 gene inactivated the retinoblastoma gene family (Rb), a tumor suppressor protein, and as a result, overexpression of p16 is induced in HPV-positive cancers [41]. For this reason, p16 immunohistochemistry (IHC) is used to discriminate HPV-positive cancers [41]. However, overexpression of p16 is infrequently observed in some HPV-negative patients [42]. To overcome this limitation, a multimodal approach combined with p16 IHC is being considered, which reduces false-positive and false-negative cases in HPV detection [43,44]. We divided the high-risk group and the low-risk group in HPV-positive cancers using the HPPI risk scoring system we previously developed, which is different from the existing methods to detect high-risk HPV. This new method is not dependent on p16 overexpression and has an analytical mechanism that is completely distinct from currently used HPV detection methods. Therefore, we suggest an excellent clinical-diagnostic approach for HPV-positive cancers by using this new method in parallel with the existing methods.

5. Conclusions

In summary, this study contributes to our understanding of the tumor biology of virus-related cancers and suggests biomarkers for risk stratification in the HPV (+) group by conducting an integrated bioinformatics analysis on the association between DEGs and prognosis of HPV infection in CC and HNC. However, we only analyzed the upregulation and downregulation of genes and their association in selected pathways using the cohort in the TCGA database, and the lack of experimental verification of the expression patterns of the target genes in cancer tissues requires future investigation.

Abbreviations

HPV, Human papillomavirus; CC, Cervical cancer; HNC, Head and neck cancer; OPSCC, Oropharyngeal squamous cell carcinoma; DEGs, Differentially expressed genes; HPPI, HPV-related prognostic and predictive indicator; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; ECM, Extracellular matrix; URR, upstream regulatory region; SAM, Significance Analysis of Microarrays; FDR, False Discovery Rate; FC, fold change; KEGG, Kyoto Encyclopedia of Genes and Genomes; TIMER, Tumor Immune; Estimation Resource; KM, Kaplan-Meier; OS, Overall survival.

Author contributions

YHK and JYJ initiated the study and guided the work. EJK, HRL, JHL, CS, MH, and JR with support from all coauthors, analyzed and interpreted the experimental data. YHK, JYJ, EJK, HRL, and MH wrote the manuscript with input from all coauthors, and all authors read and approved the final manuscript.

Ethics approval and consent to participate

The study protocol was approved by the Institutional Review Board of Ajou University Hospital (approval No. BMR-KSP-20-156) and was allowed to waive the requirement to obtain informed consent. All methods were performed in accordance with relevant guidelines and regulations.

Acknowledgment

Not applicable.

Funding

Conflict of interest

The authors declare no conflict of interest.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.31083/j.fbl2701002.

References

[1] Prati B, Marangoni B, Boccardo E. Human papillomavirus and genome instability: from productive infection to cancer. Clinics. 2018; 73: e539s.

[2] Taberna M, Mena M, Pavón MA, Alemany L, Gillison ML, Mesía R. Human papillomavirus-related oropharyngeal cancer. Annals of Oncology. 2017; 28: 2386–2398.

[3] Doorbar J, Egawa N, Griffin H, Kranjec C, Murakami I. Human papillomavirus molecular biology and disease association. Reviews in Medical Virology. 2015; 25: 2–23.

[4] Bravo IG, Félez-Sánchez M. Papillomaviruses: Viral evolution, cancer and evolutionary medicine. Evolution, Medicine, and Public Health. 2015; 2015: 32–51.

[5] Tornesello ML, Buonaguro FM. Human papillomavirus and cancers. Cancers. 2020; 12: 3772.

[6] Koneva LA, Zhang Y, Virani S, Hall PB, McHugh JB, Chepeha DB, et al. HPV Integration in HNSCC Correlates with Survival Outcomes, Immune Response Signatures, and Candidate Drivers. Molecular Cancer Research. 2018; 16: 90–102.

[7] Shenker RF, May NH, Waltonen JD, Yang JP, O'Neill SS, Frizzell BA, et al. Comparing Outcomes for Patients with Human Papillomavirus (HPV) Type 16 versus other High-Risk HPV Types in Oropharyngeal Squamous Cell Carcinoma. Head and Neck Pathology. 2021; 15: 866–874.

[8] You EL, Henry M, Zeitouni AG. Human papillomavirus-associated oropharyngeal cancer: review of current evidence and management. Current Oncology. 2019; 26: 119–123.

[9] de Sanjosé S, Serrano B, Tous S, Alejo M, Lloveras B, Quirós B, et al. Burden of Human Papillomavirus (HPV)-Related Cancers Attributable to HPVs 6/11/16/18/31/33/45/52 and 58. JNCI Cancer Spectrum. 2018; 2: pky045.

[10] Walline HM, Komarck CM, McHugh JB, Bellile EL, Brenner JC, Prince ME, et al. Genomic Integration of High-Risk HPV Alters Gene Expression in Oropharyngeal Squamous Cell Carcinoma. Molecular Cancer Research. 2016; 14: 941–952.

[11] Wan F, Miao X, Quraishi I, Kennedy V, Creek KE, Pirisi L. Gene expression changes during HPV-mediated carcinogenesis: a comparison between an in vitro cell model and cervical cancer. International Journal of Cancer. 2008; 123: 32–40.

[12] Martinez I, Wang J, Hobson KF, Ferris RL, Khan SA. Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. European Journal of Cancer. 2007; 43: 415–432.

[13] Kyritsis K, Akrivou M, Giassafaki L-P, Grigoriadis N, Viziriianakis I. Analysis of TCGA data of differentially expressed EMT-related genes and miRNAs across various malignancies to identify potential biomarkers. World Academy of Sciences Journal. 2020; 3: 2632–2919.

[14] Miller DL, Davis JW, Taylor KH, Johnson J, Shi Z, Williams R, et al. Identification of a human papillomavirus-associated oncogenic miRNA panel in human oropharyngeal squamous cell carcinoma validated by bioinformatics analysis of the Cancer Genome Atlas. The American Journal of Pathology. 2015; 185: 679–692.

[15] Shen Y, Liu J, Zhang L, Dong S, Zhang J, Liu Y, et al. Identification of Potential Biomarkers and Survival Analysis for Head and Neck Squamous Cell Carcinoma Using Bioinformatics Strategy: a Study Based on TCGA and GEO Datasets. BioMed Research International. 2019; 2019: 7376034.

[16] Wu D, Rice CM, Wang X. Cancer bioinformatics: a new approach to systems clinical medicine. BMC Bioinformatics. 2013; 13: 71.

[17] Wen L, Han Z, Du Y. Identification of gene biomarkers and immune cell infiltration characteristics in rectal cancer. Journal of Gastrointestinal Oncology. 2021; 12: 964–980.

[18] Kohli K, Pillarisetty VG, Kim TS. Key chemokines direct migration of immune cells in solid tumors. Cancer Gene Therapy. 2021. (in press)

[19] Melero I, Rouzaut A, Motz GT, Coukos G. T-cell and NK-cell infiltration into solid tumors: a key limiting factor for efficacious cancer immunotherapy. Cancer Discovery. 2014; 4: 522–526.

[20] Kwon EJ, Ha M, Jang JY, Kim YH. Identification and Complete Validation of Prognostic Gene Signatures for Human Papillomavirus-Associated Cancers: Integrated Approach Covering Different Anatomical Locations. Journal of Virology. 2021; 95: e02354-20.

[21] Kaka AS, Kumar B, Kumar P, Wakely PE, Kirsch CM, Old MO, et al. Highly aggressive human papillomavirus-related oropharyngeal cancer: clinical, radiologic, and pathologic characteristics. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology. 2013; 116: 327–335.

[22] Mirghani H, Blanchard P. Treatment de-escalation for HPV-driven oropharyngeal cancer: where do we stand? Clinical and Translational Radiation Oncology. 2018; 8: 4–11.

[23] Eble JA, Niland S. The extracellular matrix in tumor progression and metastasis. Clinical & Experimental Metastasis. 2019; 36: 171–198.

[24] Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. The Journal of Cellular Biology. 2012; 196: 395–406.

[25] Poltavets V, Kochetkova M, Pitson SM, Samuel MS. The Role of the Extracellular Matrix and its Molecular and Cellular Regulators in Cancer Cell Plasticity. Frontiers in Oncology. 2018; 8: 431.

[26] Haeggblom L, Ahrlund-Richter A, Mirzaie L, Farrajota Neves da Silva P, Ursu RG, Ramqvist T, et al. Differences in gene expression between high-grade dysplasia and invasive HPV$^+$ and HPV$^-$ tonsillar and base of tongue cancer. Cancer Medicine. 2019; 8: 6221–6232.

[27] Nishiuchi R, Takagi J, Hayashi M, Ido H, Yagi Y, Sanzen N, et al. Ligand-binding specificities of laminin-binding integrins: a comprehensive survey of laminin-integrin interactions using recombinant alpha3beta1, alpha6beta1, alpha7beta1 and alpha6beta4 integrins. Matrix Biology. 2006; 25: 189–197.

[28] Govaere O, Petz M, Wouters J, Vandewynckel YP, Scott EJ, Topal B, et al. The PDGFRalpha-laminin B1-keratin 19 cascade drives tumor progression at the invasive front of human hepatocellular carcinoma. Oncogene. 2017; 36: 6605–6616.

[29] Jiang P, He S, Li Y, Xu Z. Identification of therapeutic and prognostic biomarkers of lamin C (LAMC) family members in head and neck squamous cell carcinoma. Medical Science Monitor. 2020; 26: e925735.

[30] Lee H, Kim WJ, Kang HG, Jang JH, Choi IJ, Chun KH, et al. Upregulation of LAMB1 via ERK/c-Jun axis promotes gastric cancer growth and motility. International Journal of Molecular Sciences. 2021; 22: 626.

[31] Kawahara R, Niwa Y, Simizu S. Integrin beta1 is an essential factor in vasculogenic mimicry of human cancer cells. Cancer Science. 2018; 109: 2490–2496.

[32] Zhu H, Wang G, Zhu H, Xu A. ITGA5 is a prognostic biomarker and correlated with immune infiltration in gastrointestinal tumors. BMC Cancer. 2021; 21: 269.

[33] Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, et al. Gene expression markers of Tumor Infiltrating Leukocytes. Journal for Immunotherapy of Cancer. 2018; 5: 18.

[34] Wang Z, Yuan H, Huang J, Hu D, Qin X, Sun C, et al. Prognostic value of immune-related genes and immune cell infiltration analysis in the tumor microenvironment of head and neck squamous cell carcinoma. Head & Neck. 2021; 43: 182–197.

[35] Bommhardt U, Schraven B, Simeoni L. Beyond TCR signaling: Emerging functions of Lck in cancer and immunotherapy. International Journal of Molecular Sciences. 2019; 20: 3500.

[36] Fischer A, Picard C, Chemin K, Dogniaux S, le Deist F, Hivroz C. ZAP70: a master regulator of adaptive immunity. Seminars in Immunopathology. 2010; 32: 107–116.

[37] Gooden MJM, de Bock GH, Leffers N, Daemen T, Nijman HW. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. British Journal of Cancer. 2011; 105: 93–103.

[38] Ward MJ, Thirdborough SM, Mellows T, Riley C, Harris S, Suchak K, *et al*. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. British Journal of Cancer. 2014; 110: 489–500.

[39] Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, *et al*. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell. 2017; 171: 1611–1624.e24.

[40] Perez-Ordoñez B, Beauchemin M, Jordan RCK. Molecular biology of squamous cell carcinoma of the head and neck. Journal of Clinical Pathology. 2006; 59: 445–453.

[41] Klussmann JP, Gültekin E, Weissenborn SJ, Wieland U, Dries V, Dienes HP, *et al*. Expression of p16 protein identifies a distinct entity of tonsillar carcinomas associated with human papillomavirus. The American Journal of Pathology. 2003; 162: 747–753.

[42] Smeets SJ, Hesselink AT, Speel EM, Haesevoets A, Snijders PJF, Pawlita M, *et al*. A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. International Journal of Cancer. 2007; 121: 2465–2472.

[43] Zito Marino F, Ronchi A, Stilo M, Cozzolino I, La Mantia E, Colacurci N, *et al*. Multiplex HPV RNA in situ hybridization/p16 immunohistochemistry: a novel approach to detect papillomavirus in HPV-related cancers. a novel multiplex ISH/IHC assay to detect HPV. Infectious Agents and Cancer. 2020; 15: 46.

[44] Zito Marino F, Sabetta R, Pagliuca F, Brunelli M, Aquino G, Perdonà S, *et al*. Discrepancy of p16 immunohistochemical expression and HPV RNA in penile cancer. A multiplex in situ hybridization/immunohistochemistry approach study. Infectious Agents and Cancer. 2021; 16: 22.