



# Identifying KRT20 as a Potential Key Gene in Lymphatic Metastasis of Head and Neck Squamous Cell Carcinoma

Technology in Cancer Research & Treatment  
 Volume 21: 1-11  
 © The Author(s) 2022  
 Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
 DOI: 10.1177/15330338221107710  
[journals.sagepub.com/home/tct](https://journals.sagepub.com/home/tct)  


Yi-Fan Zhang, PhD, MD\* , Qiang Huang, PhD, MD\*,  
 Hui-Ying Huang, PhD, MD\*, Heng-Lei Ren, PhD, MD,  
 and Liang Zhou, PhD, MD

## Abstract

**Background:** Head and neck squamous cell carcinoma (HNSCC) was the seventh most common cancer worldwide in 2018. Lymphatic metastasis (LM) is closely related to HNSCC prognosis and recurrence. However, the underlying mechanism of LM remains unclear. Therefore, this study aimed to identify the key genes in the LM of HNSCC. **Methods:** We used The Cancer Genome Atlas (TCGA) to identify differentially expressed genes (DEGs) between LM and non-LM cases. A random forest model, the Search Tool for the Retrieval of Interacting Genes, Cytoscape, and cytoHubba were used to identify hub genes among DEGs, including KRT20 (Cytokeratins 20). We analyzed the survival of KRT20 in TCGA, and we overexpressed KRT20 in HNSCC cell lines to investigate its effects on migration and invasion. We also correlated the expression of KRT20 in HNSCC tissue microarrays with survival and clinicopathological features. **Results:** We identified 243 DEGs—143 upregulated genes and 100 downregulated genes. Further analysis revealed that KRT20 is a potential key gene associated with LM and overall survival rates among patients with HNSCC. Overexpression of KRT20 increased the migration and invasion ability of HNSCC cell lines Tu686 and FD-LSC-1. Tissue microarray studies demonstrated an overexpression of KRT20 among N1+ patients (including N1-N3 patients). Survival analysis results and the clinicopathological features of HNSCC tissue microarrays were consistent with our analysis of TCGA. Thus, a high KRT20 expression level might suggest an adverse HNSCC prognosis. Our gene set enrichment analysis showed that KRT20 participates in many metabolic pathways, including those related to tumorigenesis and cancer development. **Conclusions:** We propose that KRT20 may be a key gene in HNSCC with LM.

## Keywords

KRT20, head, and neck carcinoma, lymphatic metastasis, The Cancer Genome Atlas

## Abbreviations

HNSCC, head and neck squamous cell carcinoma; RT-qPCR, real-time quantitative polymerase chain reaction; LSCC, laryngeal squamous cell carcinoma; EVI1, ecotropic viral integration site 1.

Received: March 2, 2022; Revised: May 22, 2022; Accepted: May 31, 2022.

## Research Highlights

1. KRT20 is a potential key gene associated with lymphatic metastasis and overall survival rates among HNSCC patients.
2. The overexpression of KRT20 increased the migration and invasion ability in HNSCC cells.
3. GSEA analysis revealed that KRT20 participates in many compounds' metabolic pathways.

Department of Otorhinolaryngology, Eye & ENT Hospital, Fudan University, Shanghai, China

\*These authors contributed to this study equally.

### Corresponding Authors:

Heng-Lei Ren, Department of Otorhinolaryngology, Eye & ENT Hospital, Fudan University, Shanghai 200031, China.  
 Email: rhl1987@163.com

Liang Zhou, Department of Otorhinolaryngology, Eye & ENT Hospital, Fudan University, Shanghai 200031, China.  
 Email: zhoulent@126.com



## Introduction

Head and neck squamous cell carcinoma (HNSCC) was the seventh most common cancer worldwide in 2018 (with 890 000 new cases and 450 000 related deaths).<sup>1</sup> Cervical lymphatic metastasis (LM) is among the leading causes of HNSCC recurrence and fatality. HNSCC has often been associated with lymphatic—rather than hematogenous—metastasis, and patients with LM require neck lymph node dissection and radiotherapy.<sup>2</sup> Although multi-therapy has significantly improved the prognosis of patients with locally advanced HNSCC, the emergence of local LM and distant metastases remains a prognostic challenge. Previous studies have shown that the gradual metastasis of primary tumor cells destroys the association between adjacent tumors and stromal cells. Metastasis also destroys blood and lymphatic vessels, leading to tumor growth after implantation or deposition into regional or distant tissues.<sup>3</sup> However, the mechanism driving LM remains unclear. Therefore, an exploration of the key LM genes and the potential biomarkers of HNSCC is urgently needed.

Bioinformatics analysis has become a powerful tool for identifying potential key genes in cancer. Machine learning, such as the random forest method, can improve the classification performance of transcriptome data screening. One of this method's primary merits is its feature selection. Random forest has shown great potential to identify key therapeutic targets.<sup>4</sup> Further, The Cancer Genome Atlas (TCGA) is a landmark cancer genome project that has identified the molecular features of more than 20 000 primary cancers.<sup>5,6</sup> The creation of TCGA and the ubiquity of high-throughput sequencing data have enhanced the sharing of genome datasets related to cancer research. In the current study, we aimed to screen LM and non-LM HNSCC data for differentially expressed genes (DEGs) and used a random forest model and a protein–protein interaction (PPI) network to identify the hub genes that affect LM among patients with HNSCC.

## Materials and Methods

### Sources of Expression Profiles and Clinical Data

We obtained a dataset containing the mRNA expression profiles of paired samples of HNSCC tissue. As of October 29, 2020, this dataset—which we acquired from TCGA (<https://portal.gdc.cancer.gov/>)—contained 298 HNSCC tissue samples, including 135 N0 samples and 163 N1+ samples (including N1–N3 patients). All of this study's sequencing procedures were performed using the Illumina HiSeq-Counts platform. All data used were obtained from TCGA. Given that the data's original deposition adhered to TCGA's ethical guidelines.

### DEG Identification

We identified DEGs using the DESeq2 package.<sup>7</sup> DEGs were identified using the following criteria: an absolute log<sub>2</sub>fold change (FC) > 1.0 and  $P < .05$ .

### Functional and Pathway Enrichment Analyses

We performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the DAVID 6.8 tool (<https://david.ncifcrf.gov/tools.jsp>). This study's cutoff for significance was set to  $P < .05$ .

### PPI Network Construction and the Random Forest Model

We used the Search Tool for the Retrieval of Interacting Genes (STRING) database to analyze the PPI network of the DEGs. These results were then visually represented using Cytoscape software. The cutoff criterion for the combined score was >0.4. Subsequently, we used a cytoHubba plug-in in Cytoscape to further screen candidate genes,<sup>8</sup> applying a degree algorithm with a screening criterion degree  $\geq 6$ . We used a random forest model (ntree=500) for DEGs with a random forest package, ranked the DEGs by their MeanDecreaseGini scores, and selected the top 30 genes. This process was repeated 3 times to verify its accuracy.

### Survival Analysis of Hub Genes

We used GEPIA 2 (<http://gepia2.cancer-pku.cn/#survival>), a tool based on TCGA,<sup>9,10</sup> to analyze survival. The top 50% of the samples with higher expression levels were classified as having high expression, while the remaining 50% of the samples were classified as having low expression. We then used the Kaplan–Meier estimator with a log rank test to compare the prognostic differences between the 2 groups. The hazard ratios (HRs) are shown, along with their 95% confidence intervals and log-rank  $P$  values.

### Tissue Microarray Immunofluorescence

Referring to our previous studies,<sup>11,12</sup> we conducted an immunofluorescence analysis of paired tissues from 68 HNSCC patients in our cohort. A tissue microarray was obtained from 80 patients (160 cores) diagnosed with HNSCC after partial laryngectomy or total laryngectomy at Eye & ENT Hospital, Fudan University. None of these patients had taken any treatment before surgery. Several cores were damaged during preparation process. Therefore, this research included 68 matched HNSCC patients. All N1+ diseases were determined by postoperative pathological examination. We used cytokeratin 20 (ab76126, abcam, 1:200 dilution) for this analysis. We also conducted a survival analysis of the hub gene in human tissue microarrays. The patients were divided into 2 groups according to their median KRT20 expression. A Kaplan–Meier estimator with a log-rank test was used to examine the prognostic difference between the 2 groups.

### Statistical Analysis of Clinicopathological Features of Tissue Microarrays

We analyzed and visually represented the statistical data using GraphPad Prism 5.0 (GraphPad Software) and SPSS 22.0

software. An  $\chi^2$  test was used to analyze the correlations between KRT20 expression levels and the clinicopathological features of patients with HNSCC.

### Cell Culture

This study used 2 HNSCC cell lines (Tu686 and FD-LSC-1). Tu686 was provided by the China Academy of Sciences Cell Bank. FD-LSC-1 was used in our previous study.<sup>12</sup> Both were laryngeal squamous cancer cell lines. Bronchial epithelial cell growth medium (BEGM, Lonza) was used to culture FD-LSC-1, at 37 °C in a humidified 5% CO<sub>2</sub> atmosphere. Tu686 was cultured in RPMI-1640 (HyClone) supplemented with 1% (v/v) penicillin-streptomycin (Genom Biotechnology) and 10% fetal bovine serum. All cells were tested for mycoplasma contamination.

### qPCR and Western Blotting

Total RNA was isolated with TRIzol reagent (Invitrogen, Thermo Fisher Scientific) and reversed-transcribed using an Evo M-MLV RT Kit (AG11728, Accurate Biology, China). The ABI 7500 Real-Time PCR System was used to perform qRT-PCR with the SYBR Green Premix Pro Taq HS qPCR Kit (AG11717, Accurate Biology, China) using primer sets synthesized by Sangon Biotech (Shanghai). The sequences of KRT20 are as follows: forward, 5'-GGTGAACATATGGGAGCGATCT-3', and reverse, 5'-CTAGACGGTCATTTAGGTTCTGC-3'. The sequences of GAPDH are as follows: forward, 5'-TGTAGTTGAGGTCAATGAAGGG-3', and reverse, 5'-ACATCGCTCAGACACCATG-3'.

The protein concentrations were determined using a BCA assay kit (Beyotime). Cytokeratin 20 (ab76126, abcam, 1:1000) was mixed with 5% BSA. Using standard techniques, polyvinylidene difluoride membranes were incubated with primary and secondary antibodies. Immunodetection was accomplished through enhanced chemiluminescence. Blots were digitally imaged for chemiluminescence using ECL solution (Bio-Rad) according to the manufacturer's instructions (Bio-Rad), allowing saturation to be checked. Overall, the emitted photons were quantified for each band, particularly for homogeneously loaded loading controls. The protocol was similar to that described in our previous studies.<sup>12</sup>

### Cell Transfection

For stable cell transfection, the KRT20 overexpression (OE) lentivirus (PGMLV-CMV-KRT20-eGFP-3Flag-PGK, Puromycin Lentivirus, Genomeditech) was mixed with serum-free RPMI and BEGM containing 1 µg/mL polybrene (Yeasen Biotechnology). After 48 h of incubation, 5 µg/mL puromycin (Sigma) was added to the culture media, and the concentration was gradually increased up to 10 µg/mL. The process lasted for 1 week to select the stably transfected cells.

### Transwell Migration/Invasion Assays

We seeded  $2 \times 10^4$  cells in the upper chamber of the 8 µm transwell inserts (BD Biosciences) for the transwell/migration assay with 100 µL of serum-free medium. The lower chamber was filled with a medium (50 µL) containing 10% bovine serum albumin. After 36 h of incubation, cells in the upper chamber were carefully removed. The cells that adhered to the membrane were fixed in paraformaldehyde for 15 min before being stained with 1% crystal violet (KeyGEN Biotech) for 20 min. In invasion assays, the upper chamber was coated with 50 µL of Matrigel (BD Biosciences) diluted 1:8 with serum-free medium before seeding  $2 \times 10^5$  cells in 100 µL of serum-free medium. The rest of the procedure was similar to the transwell migration protocol used in our previous studies.<sup>11</sup>

### Gene Set Enrichment Analysis of the Whole-genome Dataset and KRT20 Expression Levels

To further examine the potential mechanism underlying diverse KRT20 expression groups in HNSCC tumor tissues, we performed a gene set enrichment analysis (GSEA). Results that met the following criteria were regarded as statistically significant: nominal  $P < .05$ , normalized enrichment score  $> 1$ , and false discovery rate  $< 0.3$ .

## Results

### Differentially Expressed Gene Identification and Bioinformatics Analysis

After data preprocessing and normalization using the DESeq package, we identified 243 DEGs—comprising 143 upregulated genes and 100 downregulated genes, as shown in Figure 1A and B. We conducted KEGG and GO pathway enrichment analyses using the clusterProfiler R package. To examine their cellular components, we enriched the DEGs in cornified envelopes, intermediate filaments, and extracellular space. To assess their molecular function, we also enriched the DEGs with peptide cross-linking, keratinocyte differentiation, and keratinization. KEGG pathway clustering revealed that the DEGs were mostly enriched by serotonergic synapses, tyrosine metabolism, and salivary secretion (Figure 1C). The DEGs' top 30 GO terms are shown in Figure 1D. Our cluster analysis of biological processes suggested that the DEGs were significantly enriched by “structural molecule activity,” “serine-type endopeptidase inhibitor activity,” and “structural constituent of epidermis.”

### PPI Network Construction and Random Forest Analysis of Differentially Expressed Genes

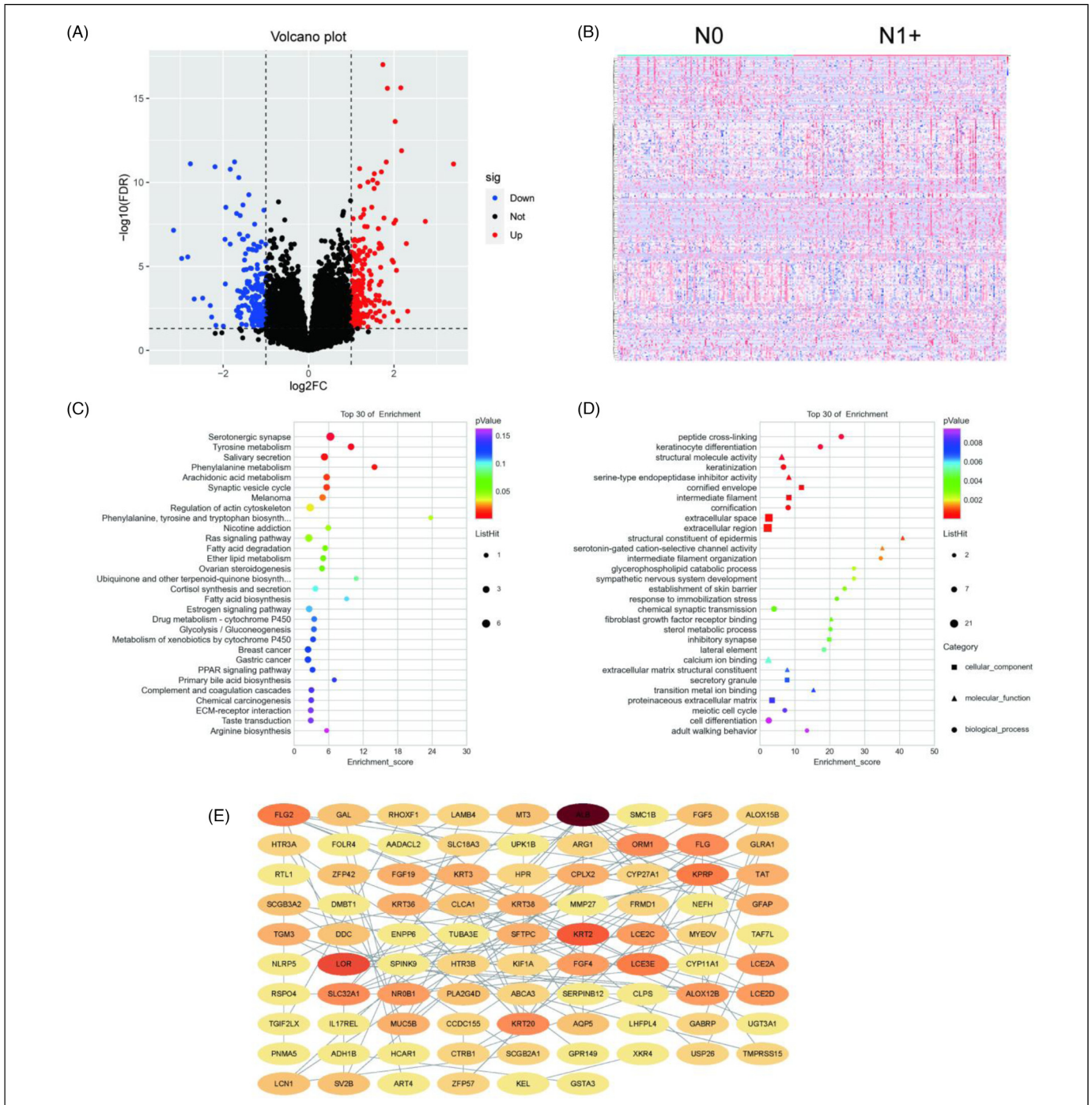
Using STRING analysis, we input the DEGs into the PPI network, including 87 nodes and 128 sides (Figure 1E). The node indicates the protein involved in the string network, and the side indicates the direct interaction between different

proteins. The details of the 87 nodes and 128 sides can be seen in Figure 1E. We then used the plugin cytoHubba to rank the nodes by their network capabilities. We also used the maximal clique centrality method to identify the top 20 genes (Table 1), including LOR, LCE3E, and KRT2, which may play important roles in HNSCC. To further identify hub genes, we used random forest packages in R to

identify hub genes. The top 30 hub genes are shown in Table 2.

### Identifying KRT20 as the Hub Gene

We used Venn plots to select intersect genes. KRT20, KRT3, and ALOX12B were identified as key genes (Figure 2A). We



**Figure 1.** Bioinformatic analysis of differentially expressed genes. (A) Volcano plot. (B) Heatmap. (C) *Kyoto Encyclopedia of Genes and Genomes* analysis. (D) Gene ontology analysis. (E) PPI network.

**Table 1.** Hub Genes Identified by cytoHubba.

Rank	Name	Score
1	LOR	70
2	LCE3E	48
3	KRT2	42
4	KPRP	31
5	LCE2C	30
5	LCE2A	30
7	KRT20	26
8	LCE2D	25
9	KRT38	24
9	KRT36	24
9	KRT3	24
12	FLG2	21
12	ALB	21
14	FLG	20
15	ORM1	9
16	SLC32A1	8
16	ALOX12B	8
18	NR0B1	6
18	FGF4	6
18	FGF19	6

**Table 2.** Hub Genes Identified by Random Forest Model.

GENE	MeanDecreaseGini
OTOP3	1.69517091
ACSBG1	1.62439616
OTOP2	1.5067028
GAPDHP27	1.44413024
LAMB4	1.23335332
TGM3	1.21570619
MUCL3	1.21060149
RPL7P38	1.17705076
ADH1B	1.13261751
MASP1	1.12515731
MYEOV	1.04466863
NWD2	1.04155006
DNAJB6P3	1.04024268
FRG2B	0.97080214
HTR3A	0.96627733
ROR1-AS1	0.94997745
SPINK7	0.89296173
ALOX12B	0.85848232
FRMD1	0.83678373
DMBT1	0.82331013
ALOX15B	0.81460052
LINC00452	0.80704212
PLA2G4D	0.78506928
TBC1D3P1	0.78415146
COL19A1	0.77486406
H2AC5P	0.7650745
ST8SIA6-AS1	0.75624077
CHCHD2P7	0.72569256
KRT20	0.72414612
KRT3	0.70645376

then conducted survival analysis of KRT20, KRT3, and ALOX12B (Figure 2B-D). For all HNSCC cases in TCGA, the results of this analysis revealed that high expression levels of KRT20 were associated with poorer overall survival rates among patients with HNSCC; this was not the case for the high expression levels of KRT3 or ALOX12B.

### *KRT20 Expression is Higher Among HNSCC Patients with N1+, Indicating Poorer Overall Survival Rates*

A total of 68 patients were randomly selected from our database and analyzed. KRT20 immunofluorescence staining showed higher KRT20 expression among patients with N1+ HNSCC compared to patients with N0 HNSCC ( $P = .0373$ ) (Figure 3A and B). Our survival analysis showed that the overexpression of KRT20 and the N1+ stage were related to poorer overall survival rates among patients with HNSCC ( $P = .0203$  and  $P = .0359$ , respectively) (Figure 3C and D). Further analysis revealed that KRT20 expression levels were related to cervical lymph node metastasis (Table 3).

We separated the 68 HNSCC patients into 2 groups—KRT20 high expression and KRT20 low expression—according to median KRT20 expression (the KRT20 high expression group was defined as relative fluorescence intensity  $>1.25$ ). The analysis of KRT20 expression and clinical characters of HNSCC patients revealed that KRT20 expression was related to N stage ( $P = .015$ ) and extracapsular spread ( $P = .046$ ). In 2015, Lewis et al defined extracapsular spread as tumor cells growing outside the lymph node envelope, perilymph node fatty infiltration, or tumor cells located outside the lymph node.<sup>13</sup> All patients with extracapsular spread in our study were identified by pathological examination.

### *Overexpression of KRT20 Promotes Migration and Invasion in HNSCC Cells*

Based on a bioinformatics analysis of KRT20, we investigated the ability of KRT20 in HNSCC cell lines. KRT20 expressing lentiviruses were successfully transfected into HNSCC cell lines (Figure 4A and B). We performed transwell migration/invasion assays using Tu686 and FD-LSC-1 cells. The results of the transwell migration/invasion assays showed that more cells migrated or invaded through the chamber in the KRT20 overexpression (KRT20 OE group) group compared to the vector group (KRT20 vector group) (Figure 4C-F).

### *KRT20 is Related to Important Biological Processes*

To further explore the molecular mechanism of KRT20 in HNSCC, we used the GSEA method to analyze a whole-genome dataset of HNSCC tumor tissues across different KRT20 expression levels. This GSEA analysis revealed that drug metabolism cytochrome P450, glycerophospholipid metabolism, glutathione metabolism, retinol metabolism, lysine degradation, valine leucine and isoleucine degradation,

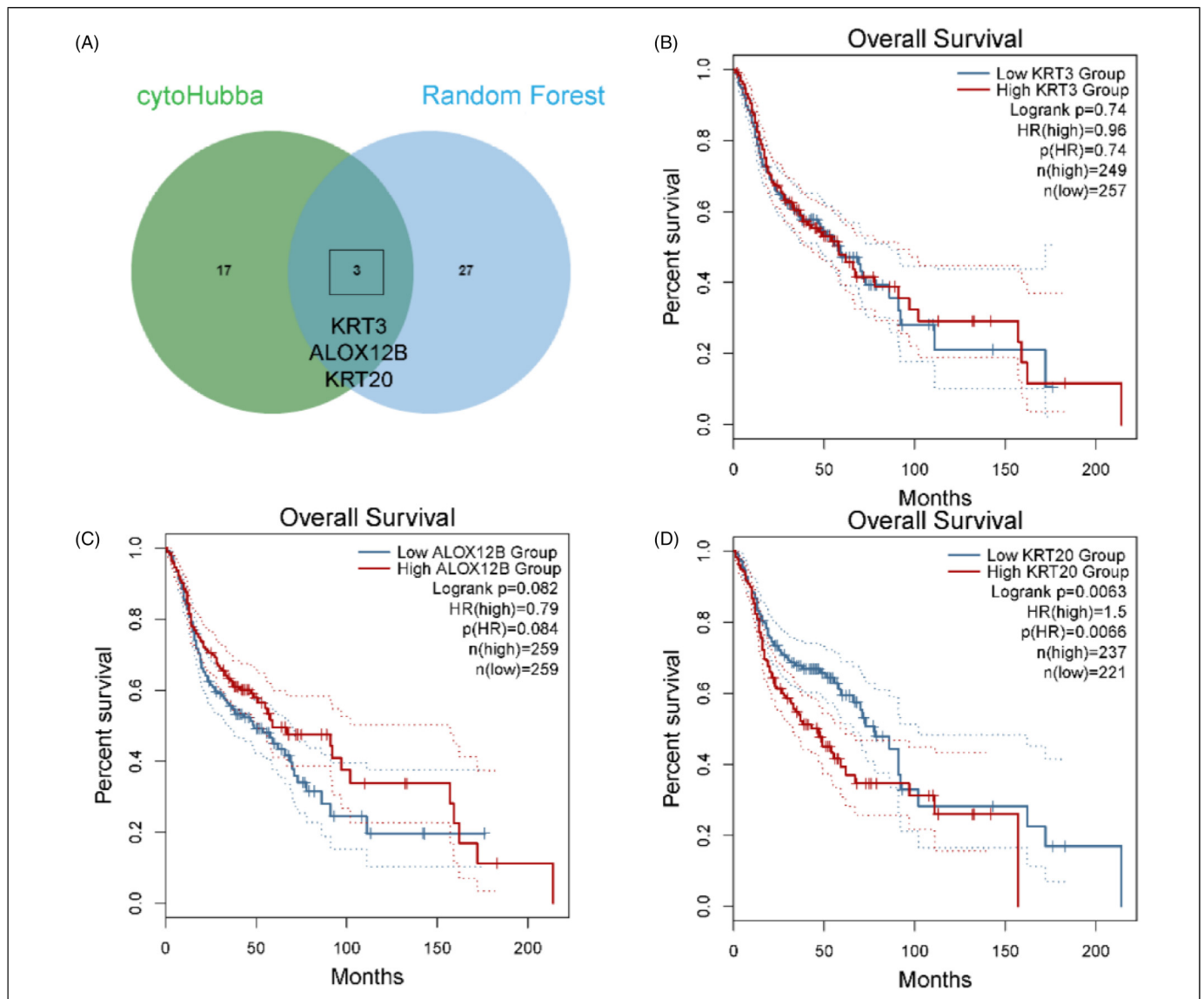
and steroid hormone biosynthesis were significantly associated with high KRT20 expression levels among patients with HNSCC (Figure 5).

## Discussion

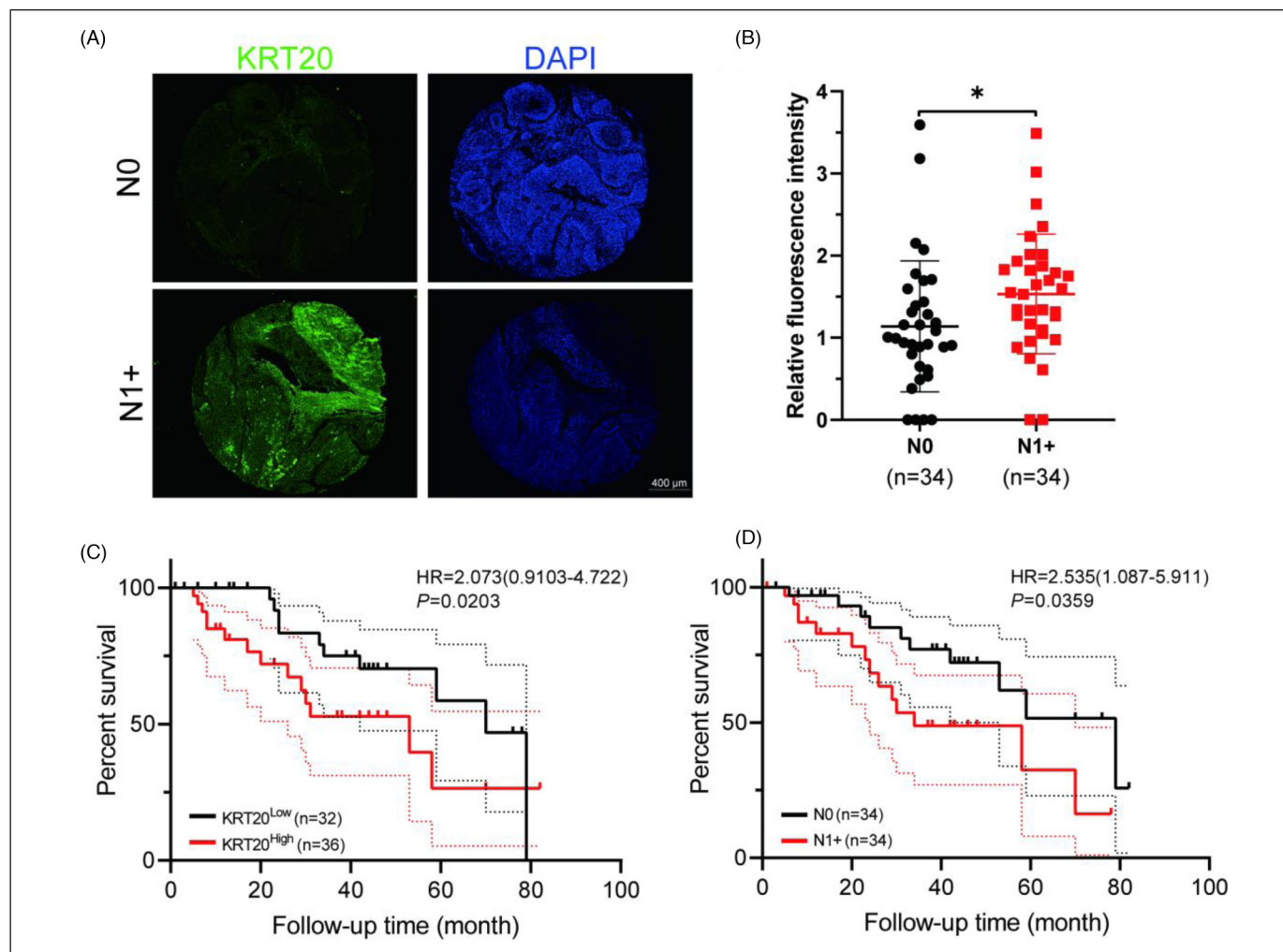
In this study, we screened a total of 243 DEGs, comprising 143 upregulated genes and 100 downregulated genes, using a PPI network, and constructed a random forest model. We selected KRT20 as a potentially key gene related to LM and the prognosis of HNSCC. Our analysis of HNSCC tissue microarrays was consistent with the results of our bioinformatics analysis, both indicating that KRT20 is highly expressed in N1+HNSCC patients and predicts a poor prognosis. Interestingly, the overall survival curve of TCGA was significantly different around 50

months, but the 2 survival curves overlapped around 100 months and seemed similar around this time point. HNSCC is a cancer of adults, with a median age at diagnosis of 66 years for HPV-negative HNSCC, 53 years for HPV-positive HNSCC, and 50 years for EBV-positive HNSCC.<sup>1</sup> For elderly patients, many factors affect their overall survival, such as cardiovascular disease, and brain disease, which may lead to this phenomenon.

Cytokeratins are filament cytoskeleton proteins found in all cells. Keratins are the most prevalent intermediate filament category found in epithelial cells, and they are expressed differently depending on the cell type. Tonofilaments are keratin filaments that braid the nucleus, span the cytoplasm, and adhere to the cytoplasmic plaques of epithelial cell–cell junctions (desmosomes) in epithelia.<sup>14</sup> Keratins are an integral part of the stability continuum from single cells through



**Figure 2.** Analysis of head and neck squamous cell carcinoma tissue microarrays. (A) Venn plot of hub genes from cytoHubba and the Random Forest Model. (B) Kaplan–Meier survival analysis of KRT3 in The Cancer Genome Atlas. (C) Kaplan–Meier survival analysis of ALOX12B in The Cancer Genome Atlas. (D) Kaplan–Meier survival analysis of KRT20 in The Cancer Genome Atlas.



**Figure 3.** KRT20 expression is higher among N1+ head and neck squamous cell carcinoma (HNSCC) patients, indicating poorer overall survival rates. (A) Immunofluorescence analysis of KRT20 (green) and DAPI (blue) among patients with N0 HNSCC and N1 + HNSCC tissues. (B) KRT20's relative expression level was significantly higher among patients with N1 + HNSCC compared to patients with N0 HNSCC ( $P < .05$ ). (C) Kaplan–Meier survival analysis of KRT20 among the tissue microarrays of patients with HNSCC. (D) Kaplan–Meier survival analysis of patients with N0 and N1 + HNSCC.

tissue formation, and they play a key role in the integrity and mechanical stability of both single epithelial cells and epithelial tissues.<sup>15</sup> Several studies have provided evidence supporting the role of active keratin in cancer cell invasion and metastasis. The transfection of CK8 and CK18 in mouse L cells, which leads to the formation of keratin filaments and is related to higher migration and invasion ability, suggests that keratin may affect the shape and migration of cells through interaction with the extracellular environment.<sup>16,17</sup>

KRT20, also known as *keratin 20*, is a low-molecular-weight cytokeratin encoded by the KRT20 gene and located on chromosome 17q21.2. KRT20 is differentially expressed in tumors. In the context of urothelial bladder cancer, strong evidence suggests the diagnostic and prognostic value of KRT20. Moreover, molecular techniques have demonstrated higher KRT20 expression levels in bladder cancer samples compared to non-neoplastic bladder tissue.<sup>18</sup> In 2010, Ye et al

found that the staining intensity of KRT20 in cancer tissues was significantly correlated with bladder carcinoma grades, distant metastasis, and TNM grades.<sup>19</sup> A previous study suggested similar results for colorectal tumors: KRT20 expression levels were significantly higher in colorectal tumors than in normal mucosa, and its expression was associated with colorectal cancer recurrence and median survival rates among the study's patient population.<sup>20</sup> However, several studies have also indicated different roles of KRT20 in tumor progression. Researchers have found that tumors characterized by the complete absence of KRT20 expression are very poorly differentiated and contain high percentages of Ki67 + cells,<sup>21</sup> contrary to our findings. According to the mentioned study, in colorectal cancer, KRT20 is expressed in enterocytes.<sup>21</sup> However, our immunofluorescence results showed that KRT20 was mainly expressed in the intercellular matrix (Figure 3A). We supposed that KRT20 is involved in different processes of tumor

**Table 3.** Relationship Between KRT20 and Clinicopathological Features of HNSCC Patients.

Variable	n	KRT20 LOW (n = 33)	KRT20 HIGH (n = 35)	P-value
Age(years)				.260
≤55	8	2	6	
>55	60	31	29	
Gender				1.000
Female	0	0	0	
Male	68	33	35	
Smoking				.884
Yes	18	9	9	
No	50	24	26	
Drinking				.474
Yes	38	20	18	
No	30	13	17	
Hypertension				.059
Yes	49	20	29	
No	19	13	6	
Diabetes				.735
Yes	58	29	29	
No	10	4	6	
T-stage				.586
T1-2	18	10	8	
T3-4	50	23	27	
N-stage				<b>.015</b>
N0	34	22	12	
N1+	34	11	23	
M-stage				1.00
M0	68	32	36	
M1	0	0	0	
Location				.393
Supraglottic	22	12	10	
Glottic	28	15	13	
Subglottic	1	1	0	
Pyriiform sinus	13	4	9	
Postcricoid region	2	1	1	
Posterior pharyngeal	2	0	2	
Lymphovascular Invasion				.590
Yes	3	1	2	
No	65	32	33	
Extracapsular spread				<b>.046</b>
Yes	11	31	26	
No	57	2	9	
Resection margins				.966
Positive	2	1	1	
Negative	66	32	34	
Perineural invasion				.258
Yes	1	1	0	
No	67	32	35	

Bold values represent  $p < 0.05$ .

development. More research is needed to elucidate the functions and roles of KRT20.

Several studies have also focused on biomarkers in the LM of HNSCC. In a study of a tissue cohort that included 389 primary HNSCCs, the researchers found that high ecotropic viral integration site 1 (EVII) expression in the primary tumor correlated with

lymph node metastasis, although EVII expression did not predict poor survival in HNSCC.<sup>22</sup> Proteomic analysis of 32 HNSCC tissue samples in a 2018 study showed that several proteins may be involved in the process of LM, especially annexin A1, fatty acid-binding protein E-FABP, keratin, and type I cytoskeletal 13. Similar to our results, they found that KRT13 and KRT6A expression levels were higher in HNSCC with LM. They also revealed low expression levels of KRT1 and KRT9 in HNSCC without LM.<sup>23</sup>

The role of KRT20 in HNSCC remains unclear, especially in the LM context. Through genome-wide co-expression analysis, we found that KRT20 may participate in drug metabolism-related cytochrome P450, glycerophospholipid metabolism, glutathione metabolism, retinol metabolism, lysine degradation, valine leucine and isoleucine degradation, and steroid hormone biosynthesis. The major function of cytochrome P450 is the metabolism of foreign compounds. Previous studies have demonstrated the involvement of cytochrome proteins in the development and drug resistance of various tumors. Furthermore, CYP1A1, a member of the cytochrome family, is regarded as a typical lung cancer biomarker.<sup>24</sup> Glycerophospholipid metabolism plays an important role in tumor cell survival, and it affects fundamental cellular processes, including signal transduction and gene expression. Highly proliferating cancer cells require the synthesis of new fatty acids to provide a constant glycerophospholipid supply. Glutamyl transferase (GGT) is a membrane-bound enzyme that catalyzes the breakdown of extracellular GSH, allowing the creation of the constituent glutamate and cysteine required for intracellular GSH synthesis to take place. Under oxidative stress, GGT levels are considerably increased, particularly in highly metabolic cancer cells.<sup>25,26</sup>

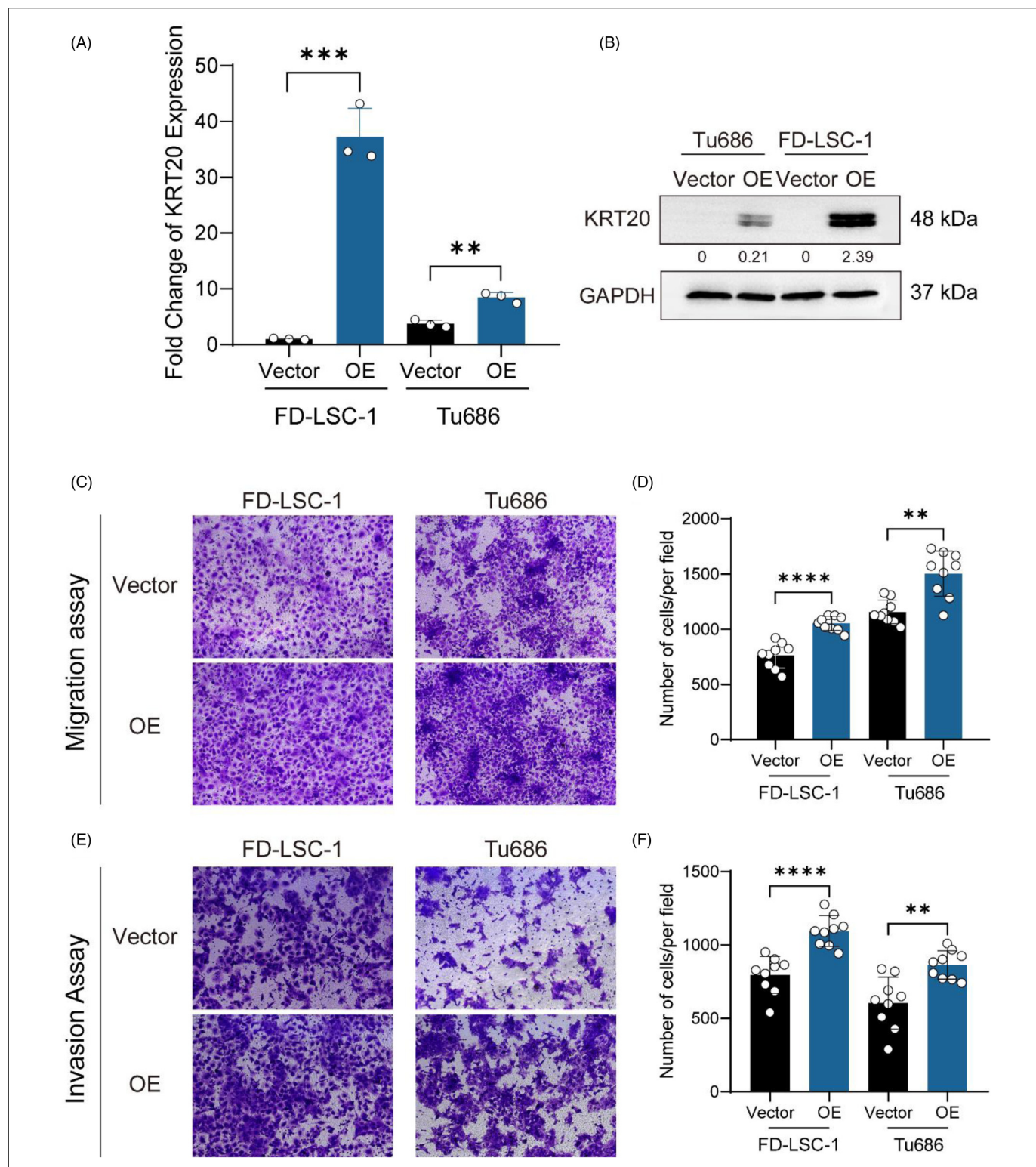
The results of the current study indicated that KRT20 was highly expressed among patients with N1+ HNSCC. Moreover, KRT20 overexpression was found to be related to poorer overall survival rates among patients with HNSCC. These results were verified in our clinical sample database. We also showed that overexpression of KRT20 promoted migration and invasion in HNSCC cell lines. Accordingly, we propose that KRT20 is a potential biomarker involved in the LM of HNSCC.

However, this study has several limitations. First, it was based on TCGA database. Although the results were verified in our clinical samples containing 68 HNSCC patients, the sample size was insufficient. Importantly, this study's results must be validated using external databases. Second, our study was verified in cell experiments but not with *in vivo* experiments. Further, the proposed molecular mechanism and its genotype-phenotype correlation were not investigated in corresponding *in vitro* and *in vivo* experiments. Thus, the molecular mechanism of KRT20 in the LM of HNSCC remains unclear, highlighting the need for further studies.

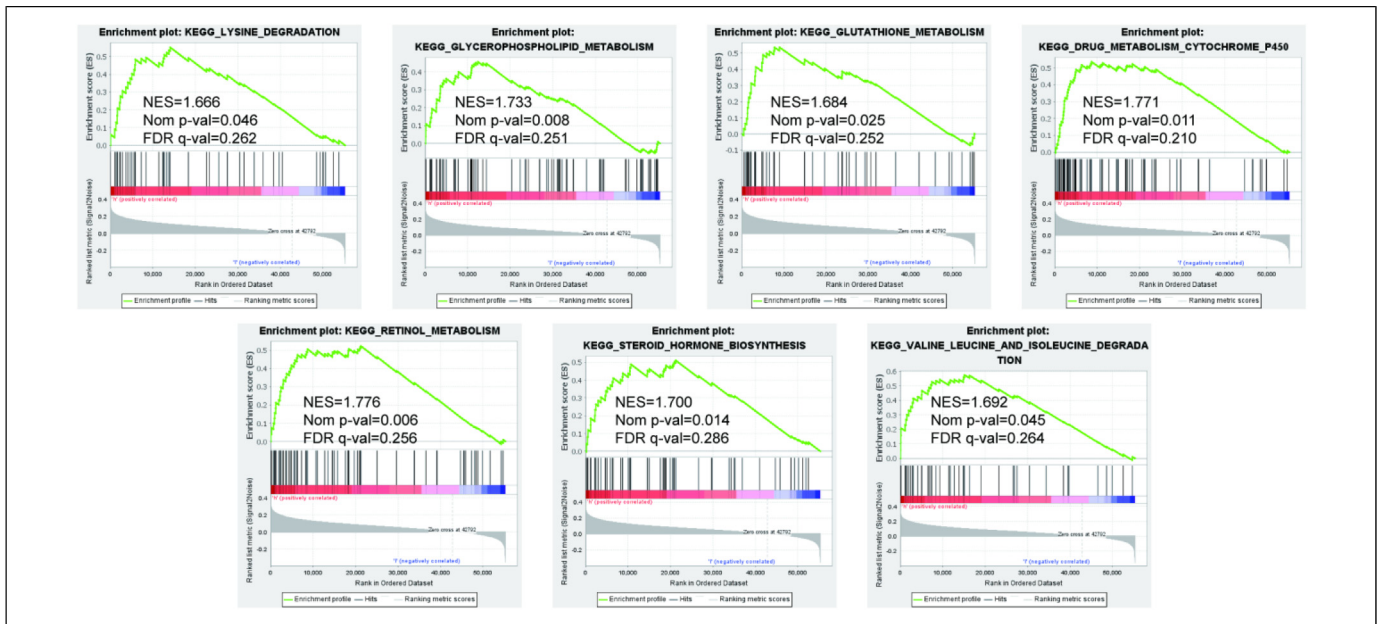
## Conclusion

Our study showed that high KRT20 expression levels among patients with HNSCC were associated with both LM and





**Figure 4.** KRT20 overexpression promotes migration and invasion in HNSCC cells. (A) The results of the quantitative real-time (qRT)-PCR analysis showed the mRNA levels of KRT20 expression in vector control (Vector) and overexpression (OE) groups. (B) western blotting showed the protein level of KRT20 in Vector and OE groups. (C) Representative fields in transwell migration assays of KRT20 Vector and OE groups. (D) Quantification results of the migration of cells per field. (E) Representative fields in transwell invasion assays of KRT20 Vector and OE groups. (F) Quantification results of the invasion of cells per field.



**Figure 5.** Gene set enrichment analysis was performed to further screen for the significant pathway between the different expression-level groups of KRT20 among patients with HNSCC.  $P < .05$  was considered significant.

unfavorable prognoses. KRT20 may also promote the ability of migration and invasion in HNSCC cells. Through genome-wide functional enrichment analysis, we found that KRT20 may be involved in the following biological processes and pathways in HNSCC: drug metabolism-related cytochrome P450, glycerophospholipid metabolism, glutathione metabolism, retinol metabolism, lysine degradation, valine leucine and isoleucine degradation, and steroid hormone biosynthesis. However, our results must be verified by future studies.

### Acknowledgments

We thank all the patients who participated in this study for providing the samples and clinical data. Data utilized in this publication include those generated by The Cancer Genome Atlas (TCGA).

### Author Contributions

YFZ, QH, HLR, and LZ conceptualized the study. YFZ and QH performed experiments. YFZ and QH performed bioinformatics data analysis. HYH contributed to the enrolment of patients, collection and processing of clinical samples, and collection and analysis of clinical data. YFZ and HYH wrote analysis scripts and drafted the manuscript. HLR and LZ revised the manuscript. HLR and LZ managed funding. All authors reviewed the manuscript.

### Availability of Data and Materials

All the source data supporting the findings of this study are available from the corresponding author upon reasonable request. The datasets analyzed during the current study were downloaded from TCGA database (<https://portal.gdc.cancer.gov/projects/TCGA-HNSC>). The data contained 298 HNSCC tissue samples as of February 1, 2021. AJCC clinical staging Nx, M1, or Mx cases were excluded.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


### Ethics Approval and Consent to Participate

All participants provided written informed consent. The experimental protocol was established, according to the ethical guidelines of the Helsinki Declaration and was approved by the Clinical Research Ethics Committee of the Eye & ENT Hospital of Fudan University (NO. KJ2008-01). Written informed consent was obtained from legally authorized representatives for anonymized patient information to be published in this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The present study was supported by grants from National Natural Science Foundation of China [No. 81972529; 82002874] and Science and Technology Commission of Shanghai Municipality [No. 19411961300].

### ORCID iD

Yi-Fan Zhang  <https://orcid.org/0000-0002-5110-4657>

### References

1. Chow LQM. Head and neck cancer. *N Engl J Med.* 2020;382(1):60-72.
2. Colevas AD, Yom SS, Pfister DG, et al. NCCN guidelines insights: head and neck cancers, version 1.2018. *J Natl Compr Canc Netw.* 2018;16(5):479-490.
3. Wirtz D, Konstantopoulos K, Searson PC. The physics of cancer: the role of physical interactions and mechanical forces in metastasis. *Nat Rev Cancer.* 2011;11(7):512-522.

4. Nevin L, PLOS Medicine Editors. Advancing the beneficial use of machine learning in health care and medicine: toward a community understanding. *PLoS Med.* 2018;15(11):e1002708.
5. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn).* 2015;19(1A):A68-A77.
6. CGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature.* 2020;578(7793):82-93.
7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
8. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. Cytohubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8(Suppl 4):S11.
9. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 2019;47(W1):W556-W560.
10. Li T, Fu J, Zeng Z, et al. TIMER2.0 For analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020;48(W1):W509-W514.
11. Huang Q, Hsueh CY, Shen YJ, et al. Small extracellular vesicle-packaged TGFβ1 promotes the reprogramming of normal fibroblasts into cancer-associated fibroblasts by regulating fibronectin in head and neck squamous cell carcinoma. *Cancer Lett.* 2021;517:1-13.
12. Huang Q, Hsueh CY, Guo Y, Wu XF, Li JY, Zhou L. Lack of miR-1246 in small extracellular vesicle blunts tumorigenesis of laryngeal carcinoma cells by regulating cyclin G2. *IUBMB Life.* 2020;72(7):1491-1503.
13. Lewis JSJr, Carpenter DH, Thorstad WL, Zhang Q, Haughey BH. Extracapsular extension is a poor predictor of disease recurrence in surgically treated oropharyngeal squamous cell carcinoma. *Mod Pathol.* 2011;24(11):1413-1420. doi:10.1038/modpathol.2011.105
14. Haines RL, Lane EB. Keratins and disease at a glance. *J Cell Sci.* 2012;125(Pt 17):3923-3928.
15. Shavandi A, Silva TH, Bekhit AA, Bekhit AEA. Keratin: dissolution, extraction and biomedical application. *Biomater Sci.* 2017;5(9):1699-1735.
16. Moll R, Divo M, Langbein L. The human keratins: biology and pathology. *Histochem Cell Biol.* 2008;129(6):705-733.
17. Karantza V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene.* 2011;30(2):127-138.
18. Schmidt J, Propping C, Siow WY, et al. Diagnostic and prognostic value of bladder cancer-related transcript markers in urine. *J Cancer Res Clin Oncol.* 2016;142(2):401-414.
19. Otto W, Denzinger S, Fritsche HM, et al. Introduction and first clinical application of a simplified immunohistochemical validation system confirms prognostic impact of KI-67 and CK20 for stage T1 urothelial bladder carcinoma: single-center analysis of eight biomarkers in a series of three hundred six patients. *Clin Genitourin Cancer.* 2013;11(4):537-544.
20. Tunca B, Tezcan G, Cecener G, et al. Overexpression of CK20, MAP3K8 and EIF5A correlates with poor prognosis in early-onset colorectal cancer patients. *J Cancer Res Clin Oncol.* 2013;139(4):691-702.
21. Dalerba P, Kalisky T, Sahoo D, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol.* 2011;29(12):1120-1127.
22. Idel C, Ribbat-Idel J, Kuppler P, et al. EVI1 As a marker for lymph node metastasis in HNSCC. *IJMS.* 2020;21(3):854.
23. Vidotto A, Polachini GM, de Paula-Silva M, et al. Differentially expressed proteins in positive versus negative HNSCC lymph nodes. *BMC Med Genomics.* 2018;11(1):73.
24. Kwon YJ, Shin S, Chun YJ. Biological roles of cytochrome P450 1A1, 1A2, and 1B1 enzymes. *Arch Pharm Res.* 2021;44(1):63-83.
25. Bansal A, Simon MC. Glutathione metabolism in cancer progression and treatment resistance. *J Cell Biol.* 2018;217(7):2291-2298.
26. Hakimi AA, Reznik E, Lee CH, et al. An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer Cell.* 2016;29(1):104-116.