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

Yi-Fan Zhang

Fudan University

Qiang Huang

Fudan University

Hui-Ying Huang

Fudan University

Yu-jie Shen

Fudan University

Hui-Ching Lau

Fudan University

Yang Guo

Fudan University

Chi-Yao Hsueh

Fudan University

Heng-Lei Ren

Fudan University

Liang Zhou (✉ zhoulent@126.com)

Fudan University

Research Article

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

Posted Date: January 27th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1199714/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Head and neck squamous cell carcinoma (HNSCC) was the seventh-most-common cancer worldwide in 2018. Moreover, lymphatic metastasis (LM) closely relates to HNSCC's prognosis and recurrence. However, mechanism driving LM remains unclear. Therefore, this study aimed to identify 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. Random Forest model, the Search Tool for the Retrieval of Interacting Genes (STRING), Cytoscape, and cytoHubba were used to identify hub genes in DEGs, including KRT20 (Cytokeratins 20). We analyzed survival and pan-cancer of KRT20 in the TCGA. After overexpressing KRT20 in HNSCC cell lines, migration and invasion were investigated. We analyzed KRT20's expression in HNSCC tissue microarrays, as well as the survival and clinicopathological features.

Results

In total, we identified 243 DEGs—143 upregulated genes and 100 downregulated genes. Further analysis revealed KRT20 as a potential key gene associated with lymphatic metastasis and overall survival rates among patients with HNSCC. Overexpression of KRT20 increased migration and invasion ability of Tu686 and FD cells. Our analysis of tissue microarrays revealed an overexpression of KRT20 among N1+ patients. Survival analysis results and HNSCC tissue microarrays' clinicopathological features were consistent with our analysis of the TCGA. Moreover, high level of KRT20 expression might suggest an adverse HNSCC prognosis. Additionally, our GSEA analysis revealed that KRT20 participates in many compounds' metabolic pathways—including pathways related to tumorigenesis and development.

Conclusion

We proposed that KRT20 may be a key gene in HNSCC with LM.

Research Highlights

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

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) [1]. Cervical lymphatic metastasis (LM) is among the leading causes of HNSCC's recurrence and fatality [2]. HNSCC has often been associated with lymphatic—rather than hematogenous—metastasis, and patients with LM require neck dissection and prophylactic neck radiotherapy [3]. 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 necessarily destroys the association between adjacent tumors and stromal cells, the survival rate of blood and lymphatic vessels, regional and systemic circulation, and growth after implantation or deposition into regional or distant tissues [4]. However, the mechanism driving LM remains unclear. Therefore, an exploration of LM's key genes and the potential biomarkers of HNSCC are urgently needed.

Bioinformatics analysis has become a powerful tool with which to identify potential key genes in cancers. 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 [5]. Meanwhile, The Cancer Genome Atlas (TCGA) is a landmark cancer genome project that has identified the molecular features of more than 20,000 primary cancers [6][7]. The TCGA's creation—and the ubiquity of high-throughput sequencing data—means that genome data sets related to cancer research have become more and more widely shared. In the current study, we aimed to screen LM and non-LM HNSCC data for differentially expressed genes (DEGs). More importantly, we identified the hub genes that affect LM among patients with HNSCC using the Random Forest model and protein-protein interaction (PPI) network.

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 the TCGA (<https://portal.gdc.cancer.gov/>)—contained 298 HNSCC tissue samples including 135 N0 samples and 163 N1+ samples. All of this study's sequencing procedures were performed using the Illumina HiSeq-Counts platform. Moreover, all data used were obtained from the TCGA. Because the data's original deposition adhered to the TCGA's ethical guidelines, no additional ethical clearance was needed from our local research ethics committee.

DEG identification

We identified differentially expressed genes (DEGs) using the DESeq2 package [8]. DEGs were identified using the following criteria: an absolute log2fold change (FC) > 1.0 and $p < 0.05$.

Functional and pathway enrichment analysis

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 $p < 0.05$.

PPI network construction and the Random Forest model

Next, we used the Search Tool for the Retrieval of Interacting Genes (STRING) database to analyze DEGs' PPI network. Then, these results were 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 [10]. At this point, a degree algorithm was applied, and the screening criterion was $degree \geq 6$. We used a Random Forest model (ntree = 500) for DEGs with a Random Forest package. Next, we ranked DEGs by their MeanDecreaseGini scores and selected the top 30 genes. We then repeated this process three times to verify its accuracy.

Survival analysis of hub genes

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

The hub gene's expression in pan-cancer contexts

We used the TIMER tool (<https://cistrome.shinyapps.io/timer/>), which systematically analyzes expression and immune infiltration in more than 30 cancer types [11]. Specifically, we used "Cancer Exploration" in this study.

Tissue microarrays' immunofluorescence

Referring to our previous articles [12][13], we conducted an immunofluorescence analysis of paired tissues from a total of 68 HNSCC patients from our cohort. We used cytokeratin 20 (ab76126, abcam, 1:200 dilution) for this analysis. We also conducted survival analysis of the hub gene in human tissue microarrays. Patients were divided into two 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 two groups. All the participants signed a written informed consent form. The Clinical Research Ethics Committee of Fudan University's Eye & ENT Hospital approved the protocols (NO. KJ2008-01).

Statistical analysis of tissue microarrays' clinicopathological features

We analyzed and visually represented statistical data using GraphPad Prism 5.0 (GraphPad Software, La Jolla, CA, USA) and SPSS 22.0 software. An χ^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 two HNSCC cell lines (Tu686 and FD-LSC-1). Bronchial Epithelial Cell Growth Medium (BEGM, Lonza, Walkersville, MD, USA) was used to cultivate FD-LSC-1. At 37°C in a humidified 5% CO₂ atmosphere, HNSCC cell line Tu686 was cultivated in RPMI-1640 (HyClone, Logan, UT) supplemented with 1% (v/v) penicillin-streptomycin (Genom Biotechnology, China) and 10% fetal bovine serum (FBS). Mycoplasma contamination was tested in all HNSCC cell lines.

qPCR and western blotting

Total RNA was isolated with TRIzol reagent (Invitrogen, Thermo Fisher Scientific) and then reversed-transcribed using an Evo M-MLV RT Kit (AG, China). The ABI 7500 Real-Time PCR System was used to perform qRT-PCR with the SYBR Green Premix Pro Taq HS qPCR Kit (AG, China) (Life Technologies, Shanghai, China). Sangon Biotech synthesized the primers (Shanghai). The sequences of KRT20 as follows: forward, 5'-GGTGAAGTATGGGAGCGATCT-3', and reverse, 5'-CTAGACGGTCATTTAGGTTCTGC-3'. The sequences of GAPDH as follows: forward, 5'-TGTAAGTGGAGTCAATGAAGGG-3', and reverse, 5'-ACATCGCTCAGACACCATG-3'.

The protein concentrations were determined using the BCA assay kit (Beyotime, China). Cytokeratin 20 (ab76126, abcam, 1:1000) was mixed with 5% BSA. Using standard techniques, primary and secondary antibodies were incubated with polyvinylidene difluoride membranes. Immunodetection was accomplished through the use of enhanced chemiluminescence. Blots were digitally imaged for chemiluminescence using ECL solution (Bio-Rad) according to manufacturer's instructions (Bio-Rad, Hercules, CA, USA), allowing saturation to be checked. Overall emitted photons were quantified for each band, particularly for homogeneously loaded loading controls. The protocol was similar to our published articles[13].

Cell transfection

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

Transwell Migration/Invasion Assays

2×10⁴ cells were seeded in the upper chamber of the 8µm transwell inserts (BD Biosciences, Franklin Lakes, NJ, USA) for the transwell/migration assay with 100µl of serum-free medium. The lower chamber was filled with medium (50µl) containing 10% bovine serum albumin. After 48 hours of incubation, cells in the upper chamber were carefully removed. The cells that adhered to the membrane were fixed in paraformaldehyde for 15 minutes before being stained with 1% crystal violet (KeyGEN Biotech, Nanjing City, China) for 20 minutes. In invasion assays, the upper chamber was coated with 50µl of Matrigel (BD Biosciences, Bedford, MD, USA) diluted 1:8 with serum-free medium before seeding 2×10⁵ cells in 100µl

of serum-free medium. The rest of the procedure followed the same pattern as the transwell migration in our published articles[12].

Gene set enrichment analysis of whole-genome dataset and KRT20 expression levels

To further examine the potential mechanism underlying diverse KRT20 expression groups in HNSCC tumor tissues, we then performed a gene set enrichment analysis (GSEA). Results meeting the following criteria were regarded as statistically significant: nominal $p < 0.05$, normalized enrichment score > 1 , and FDR < 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—figures 1A and 1B show. Next, we conducted KEGG and GO pathway enrichment analyses using the clusterProfiler R packages. To examine their cellular components, we enriched the DEGs in cornified envelopes, intermediate filaments, and extracellular space. Additionally, to assess their molecular function, we enriched the DEGs with peptide crossing-linking, keratinocyte differentiation, and keratinization. KEGG pathway clustering revealed that the DEGs were mostly enriched by serotonergic synapse, 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 Forests analysis of differentially expressed genes

Using STRING analysis, we input the DEGs into the PPI network, including 58 nodes and 128 sides (Figure 1E). Then, we used cytoHubba—a plugin to rank nodes by their network capabilities. We also used the MCC (maximal clique centrality) method to identify the top 20 genes (Table 1). Our findings may indicate that these 20 genes, including LOR, LCE3E and KRT2, 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

Next, we used Venn plots to select intersect genes. KRT20, KRT3, and ALOX12B were identified as key genes (Figure 2A). We then conducted survival analysis of KRT20, KRT3, and ALOX12B (figures 2B–2D). For all HNSCC cases in the TCGA, the results of this analysis revealed that high expression levels of

KRT20 were associated with poorer overall survival rates among patients with HNSCC—rather than high expression levels of KRT3 or ALOX12B. We also used TIMER 2 to explore KRT20 expression in various tumors. The results of this exploration revealed that KRT20 expression can vary between normal tissue and tumor tissue, based on the TCGA (Figure 2E). Among patients with HNSCC, KRT20 was highly expressed in tumor tissue compared to normal tissue. Our analysis also revealed a similar tendency to highly express KRT20 across the LIHC, LUAD, LUSC, and PRAD samples.

KRT20 expression is higher among HNSCC patients with N1+, indicating poorer overall survival rates

Total 68 patients were selected from our database and were analyzed. KRT20 immunofluorescence staining showed higher KRT20 expression among patients with N1+ HNSCC compared to patients with N0 HNSCC ($p = 0.0373$) (Figures 3A and 3B). Additionally, our survival analysis showed that the overexpression of KRT20 and the N1+ stage were related to poorer overall survival rates among patients with HNSCC ($p = 0.0203$ and $p = 0.0359$, respectively) (Figure 3C and 3D). Further analysis revealed that KRT20 expression levels were related to cervical lymph node metastasis (Table 3).

Overexpression of KRT20 promotes migration and invasion in HNSCC cells.

Based on bioinformatic analysis of KRT20, we investigated the ability of KRT20 in HNSCC cell lines. KRT20 expressing lentivirus were successfully transfected into HNSCC cell lines (Figures 4A-B). We performed transwell migration/invasion assays in Tu686 and FD-LSC-1 cell lines. 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 vs. the vector group (KRT20 Vector group) (Figures 4C-F).

KRT20 is related to important biological processes

To further explore KRT20's molecular mechanism 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 screen a total of 243 DEGs, comprising 143 upregulated genes and 100 downregulated genes. After this screening with the PPI network and a Random Forest model, we selected KRT20 as a potentially key gene related to HNSCC's LM and prognosis. 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.

Cytokeratins are filament cytoskeleton proteins that are 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 made of keratin filaments braid the nucleus, span the cytoplasm, and adhere to the cytoplasmic plaques of epithelial cell-cell junctions (desmosomes) in epithelia[14]. Keratins are an integral part of the stability continuum from single cells through tissue formation, and they play a key role in the integrity and mechanical stability of both single epithelial cells and epithelial tissues. [15][16]. Several types of cytokeratins have been described in both normal urothelium and neoplastic urothelium [17][18].

KRT20, also known as *keratin 20*, is a low-molecular-weight cytokeratin encoded by the KRT20 gene, and it is located on chromosome 17q21.2. KRT20 is expressed differently in tumors. In the context of urothelial bladder cancer, strong evidence has suggested KRT20's diagnostic and prognostic value. Moreover, molecular techniques have demonstrated higher KRT20 expression levels in bladder cancer samples compared to non-neoplastic bladder tissue [19]. In 2010, Ye et al. found that KRT20's staining intensity in cancer tissues significantly correlated with bladder carcinoma grades, distant metastasis, and TNM grades [20]. Moreover, a previous study suggested similar results for colorectal tumors. KRT20 expression levels were significantly higher in colorectal tumors than in normal mucosa. Furthermore, KRT20 expression was associated with colorectal cancer recurrence and median survival rates among that study's patient population [21]. Yet, there are also several studies indicated different roles of KRT20 in tumor progression. Researchers found that tumors characterized by the complete absence of KRT20 expression were very poorly differentiated and contained high percentages of Ki67+ cells[22]. More research was required to learn about KRT20.

KRT20's role in HNSCC remains unclear—especially in the LM context. Through genome-wide co-expression analysis, we found that KRT20 may participate in drug metabolism cytochrome P450, glycerophospholipid metabolism, glutathione metabolism, retinol metabolism, lysine degradation, valine leucine and isoleucine degradation, and steroid hormone biosynthesis. Cytochrome P450s' major function is the metabolism of foreign compounds. Previous studies have demonstrated cytochromes involvement in different tumors' development and drug resistance. Furthermore, CYP1A1, a member of the cytochromes family, is regarded as a typical lung cancer biomarker [23]. CYP1A1 is also thought to play a role in the pathophysiology of inflammatory illnesses by acting as an oxygenase in eicosanoid metabolism, according to recent research [24]. The relationship between inflammation and malignancies has been well-studied. Chemokines, prostaglandins, and cytokines, all of which have been demonstrated to inhibit CYP enzyme function, are components of cancer inflammation [25]. The proinflammatory cytokines IL-6, INF-, TNF-, and IL-1 are thought to be the most powerful mediators of cytochrome activity and expression reduction. Eicosanoids generated from cytochromes are affected by cancer-related inflammation [26].

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 glycerophospholipids supply.

New fatty acids are needed to continuously provide glycerophospholipids for membrane production. Therefore, unlike normal tissues, endogenously synthesized fatty acids are primarily esterified to phospholipids—rather than triacylglycerols—from phosphatidic acid in membrane production [27]. Glutathione (GSH) is the most common antioxidant in living creatures, and it has a variety of roles, the majority of which are related to maintaining the redox equilibrium of cells. 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. A considerable glutathione route and dipeptide metabolites have been linked to stage IV clear-cell renal cell carcinoma, according to metabolomic analysis [28][29].

The results of the current study indicated that KRT20 is highly expressed among patients with N1+ HNSCC. Moreover, KRT20 overexpression was found to relate to poorer overall survival rates among patients with HNSCC. These results were verified in our clinical samples database. The research also showed the 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 entailed several limitations. First, it was based on the TCGA. 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. Secondly, our study was verified in cell experiments but not in *in vivo* experiments. Besides, proposed molecular mechanism and its genotype–phenotype correlation were not investigated in corresponding *in vitro* and *in vivo* experiments. KRT20's molecular mechanism in the LM of HNSCC remains unclear, and further studies are needed to elucidate this mechanism.

Conclusion

Our study has shown that high KRT20 expression levels among patients with HNSCC are associated with both LM and unfavorable prognoses. KRT20 could 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 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.

Abbreviations

HNSCC

head and neck squamous cell carcinoma

RT-qPCR

real-time quantitative polymerase chain reaction

miRNAs
microRNAs
UTR
untranslated region
LSCC
laryngeal squamous cell carcinoma
EMT
epithelial–mesenchymal transition
CRC
colorectal cancer

Declarations

Ethics approval and consent to participate

All participants provided written informed consent. The protocols were authorized by the 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 a legally authorized representatives for anonymized patient information to be published in this article.

Consent for publication

Not applicable.

Availability of data and material

All the source data supporting the findings of this study are available from the corresponding author upon reasonable request. The datasets analysed during the current study are downloaded from TCGA database (<https://portal.gdc.cancer.gov/projects/TCGA-HNSC>). The data contains 298 HNSCC tissue samples as of October 29, 2020. AJCC clinical staging Nx, M1 or Mx cases were excluded. All TCGA cases data can be downloaded in supplementary materials.

Disclosure Statement

The authors report no conflict of interest.

Funding

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].

Authors' contributions

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

Acknowledgements

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 The Cancer Genome Atlas (TCGA).

References

1. Chow, L. Q. M. Head and Neck Cancer. *N Engl J Med* 382, 60–72 (2020).
2. Petti, S., Masood, M. & Scully, C. The Magnitude of Tobacco Smoking-Betel Quid Chewing-Alcohol Drinking Interaction Effect on Oral Cancer in South-East Asia. A Meta-Analysis of Observational Studies. *PLoS ONE* 8, e78999 (2013).
3. Colevas, A. D. et al. NCCN Guidelines Insights: Head and Neck Cancers, Version 1.2018. *J Natl Compr Canc Netw* 16, 479–490 (2018).
4. Wirtz, D., Konstantopoulos, K. & Searson, P. C. The physics of cancer: the role of physical interactions and mechanical forces in metastasis. *Nat Rev Cancer* 11, 512–522 (2011).
5. Nevin, L. & on behalf of the PLOS Medicine Editors. Advancing the beneficial use of machine learning in health care and medicine: Toward a community understanding. *PLoS Med* 15, e1002708 (2018).
6. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *wo* 1A, 68–77 (2015).
7. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* 578, 82–93 (2020).
8. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550 (2014).
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 Research* 47, W556–W560 (2019).
10. Chin, C.-H. et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8, S11 (2014).
11. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Research* 48, W509–W514 (2020).
12. Huang, Q. 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 Letters* 517, 1–13 (2021).

13. Huang, Q. et al. Lack of miR-1246 in small extracellular vesicle blunts tumorigenesis of laryngeal carcinoma cells by regulating Cyclin G2. *IUBMB Life* 72, 1491–1503 (2020).
14. Haines, R. L. & Lane, E. B. Keratins and disease at a glance. *Journal of Cell Science* 125, 3923–3928 (2012).
15. Shavandi, A., Silva, T. H., Bekhit, A. A. & Bekhit, A. E.-D. A. Keratin: dissolution, extraction and biomedical application. *Biomater. Sci.* 5, 1699–1735 (2017).
16. Schweizer, J. et al. New consensus nomenclature for mammalian keratins. *Journal of Cell Biology* 174, 169–174 (2006).
17. Moll, R., Divo, M. & Langbein, L. The human keratins: biology and pathology. *Histochem Cell Biol* 129, 705–733 (2008).
18. Karantza, V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene* 30, 127–138 (2011).
19. Schmidt, J. et al. Diagnostic and prognostic value of bladder cancer-related transcript markers in urine. *J Cancer Res Clin Oncol* 142, 401–414 (2016).
20. Ye, Y. et al. KRT20 and Ki-67 as significant prognostic factors in human bladder carcinoma. *Clin Exp Med* 10, 153–158 (2010).
21. Tunca, B. et al. Overexpression of KRT20, MAP3K8 and EIF5A correlates with poor prognosis in early-onset colorectal cancer patients. *J Cancer Res Clin Oncol* 139, 691–702 (2013).
22. Dalerba, P. et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* 29, 1120–1127 (2011).
23. Kwon, Y.-J., Shin, S. & Chun, Y.-J. Biological roles of cytochrome P450 1A1, 1A2, and 1B1 enzymes. *Arch. Pharm. Res.* 44, 63–83 (2021).
24. Elfaki, I., Mir, R., Almutairi, F. M. & Duhier, F. M. A. Cytochrome P450: Polymorphisms and Roles in Cancer, Diabetes and Atherosclerosis. *Asian Pac J Cancer Prev* 19, (2018).
25. Harvey, R. D. & Morgan, E. T. Cancer, Inflammation, and Therapy: Effects on Cytochrome P450–Mediated Drug Metabolism and Implications for Novel Immunotherapeutic Agents. *Clin Pharmacol Ther* 96, 449–457 (2014).
26. Theken, K. N. et al. Activation of the Acute Inflammatory Response Alters Cytochrome P450 Expression and Eicosanoid Metabolism. *Drug Metab Dispos* 39, 22–29 (2011).
27. Dolce, V., Cappello, A. R., Lappano, R. & Maggiolini, M. Glycerophospholipid Synthesis as a Novel Drug Target Against Cancer. 9.
28. Bansal, A. & Simon, M. C. Glutathione metabolism in cancer progression and treatment resistance. *Journal of Cell Biology* 217, 2291–2298 (2018).
29. Hakimi, A. A. et al. An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* 29, 104–116 (2016).

Tables

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

GENE	MeanDecreaseGini
ST8SIA6-AS1	0.75624077
CHCHD2P7	0.72569256
KRT20	0.72414612
KRT3	0.70645376

Table 3. Relationship between KRT20 and clinicopathological features of HNSCC patients.

Variable	n	KRT20 LOW	KRT20 HIGH	P-value
Age(years)				
<=55	8	2	6	0.26
>55	60	31	29	
Gender				
Female	18	0	0	0.26
Male	50	32	36	
Smoking				
Yes	38	9	9	1.00
No	30	24	26	
Drinking				
Yes	38	20	18	0.474
No	30	13	17	
Hypertension				
Yes	49	20	29	0.059
No	19	13	6	
Diabetes				
Yes	58	29	29	0.406
No	10	4	6	
T-stage				
1/2	18	10	8	0.337
3/4	50	23	27	
N-stage				
N0	34	22	12	0.015
N1+	34	11	23	
M-stage				

M0	68	32	36	
M1	0	0	0	1.00

Figures

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.

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. (E) Pan-cancer analysis of KRT20 expression.

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 ($p < 0.05$). (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.

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. $p < 0.05$ was considered significant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TCGAN0.csv](#)
- [TCGAN1.csv](#)