

# LIM Homeobox Transcription Factors as Novel Prognosis Biomarkers in Head and Neck Squamous Cell Carcinoma

Chen Lu

Hospital of Stomatology, Jilin University

Cong-cong Duan

Hospital of Stomatology, Jilin University

Han Bing (✉ [hbing@jlu.edu.cn](mailto:hbing@jlu.edu.cn))

Hospital of Stomatology, Jilin University

---

## Research Article

**Keywords:** LIM homeobox genes, head and neck squamous cell carcinoma, transcription factors, prognosis biomarkers, immune infiltration, methylation, TCGA

**Posted Date:** December 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1204015/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** LIM class homeobox (LHX) genes, an important subfamily of the homeobox genes, encode transcription factors that have a fundamental role during embryonic development. However, knowledge regarding the function and mechanism of LHXs in head and neck squamous cell carcinoma (HNSCC) is still lacking.

**Methods:** We conduct a bioinformatic analysis to systematically explore the mRNA expression, clinical correlation, prognostic values, and underlying mechanisms of distinct LHXs in HNSCC. The differentially expressed mRNAs in the LIM homeobox gene family and their correlation with clinical variables were determined and verified with packages in software R. The prognosis values of LHXs expression levels were evaluated by Kaplan-Meier (KM) method and Cox proportional hazard model. Gene set enrichment analysis (GSEA) was conducted to understand the potential biological function of LHXs. The immune cell infiltration patterns were estimated through CIBERSORT and TIMER, and the methylation levels of LHXs in HNSCC were explored in UALCAN and MEXPRESS.

**Results:** We found that among 12 LHXs, 8 genes (ISL1, LHX1-3, 5, 9, LMX1A, and LMX1B) showed altered expression in HNSCC tissues and detected significant correlations between their expression and clinical variables. Survival analysis revealed that LHX1, LHX5, LMX1A, LMX1B can serve as unfavorable prognosis predictors, and ISL1, LHX2, LHX9 can serve as favorable predictors in all HNSCC patients. Gene sets enrichment analysis and immune infiltration analyses showed that the aberrant expression of LHXs was closely related to cancer-associated processes and immune cell infiltration patterns. Finally, we observed hyper-methylation in the promoters of ISL1, LHX2, 5, 9, LMX1A, LMX1B and hypo-methylation in LHX3 promoter, suggesting the regulatory mechanism of LHXs abnormal expression may be related to aberrant DNA methylation.

**Conclusions:** Our study found the oncogenic roles of LHX1,5 and LMX1B and the tumor-suppressor roles of ISL1 and LHX2 in patients with HNSCC, suggesting these LHXs as novel diagnostic and prognostic biomarkers for HNSCC.

## Background

Head and neck squamous cell carcinoma (HNSCC) is a malignant tumor derived from the mucosal epithelium in the oral cavity, larynx and pharynx (1). It is the sixth most common cancer by incidence worldwide and survivors of HNSCC have the second-highest suicide rates due to depression and depressed quality of life after cancer treatment (2, 3). Although continuous improvement of HNSCC treatment has been achieved by surgery, chemotherapy or radiotherapy, the survival rate of this cancer remains stagnant, especially for HPV-negative HNSCC (4). Thus, there is an urgent need to develop effective target therapeutic strategies and search for novel biomarkers to improve diagnosis, prognosis, and treatment for HNSCC.

Homeobox genes encode transcription factors that are critical for embryonic development. In human genomes, homeobox genes are divided into several subfamilies, one of which is the LIM class homeobox (LHX) genes. The proteins encoded by LHXs contain a centrally located homeodomain for binding DNA and two tandem cysteine-rich zinc-finger LIM domains for protein-protein interactions, which enable them to participate in a wide range of biological processes(5-8). To date, 12 LHXs have been identified in human genomes and their roles in neoplastic transformation and tumor progression have attracted great interest in recent years (9). However, the scientific role of distinct LHXs in HNSCC remained unknown. Thus, a comprehensive study of LHX transcription factors will help to understand the molecular mechanisms behind HNSCC and develop novel biomarkers for this cancer.

In this study, we conducted a bioinformatic analysis to investigate the role of LHXs in HNSCC and their underlying mechanisms in tumor formation. We examined 8 LHXs with altered expression in HNSCC and confirmed their discriminatory power between normal and tumor samples using Receiver Operating Characteristic (ROC) curve analysis. Then their connection with clinical attributes and patient survival was evaluated. Subsequently, we conducted the gene set enrichment analysis (GSEA) as well as immune infiltration analysis to understand the potential biological function of LHXs. Finally, methylation analysis was performed to explore the epigenetic mechanism behind the abnormal expression of LHXs in HNSCC.

## Methods

### Data collection

The information of 257 human homeobox genes was gathered from the homeobox database (<http://homeodb.zoo.ox.ac.uk/>) (10) (Additional file 1, Table S1). The mRNA expression data of 502 HNSCC samples and 44 normal adjacent tissue samples (dataset ID: TCGA.HNSC.sampleMap/HiSeqV2, version: 2017-10-13), and clinical information (dataset ID: survival/HNSC survival.txt, version: 2018-09-13; dataset ID: TCGA.HNSC.sampleMap/ HNSC clinicalMatrix, version: 2019-12-06) were extracted from The Cancer Genome Atlas (TCGA) database using USCS XenaBrowser (<http://xena.ucsc.edu/>) (11). The clinical features of HNSCC patients are shown in Table S2 (Additional file 2).

### Differential Expression Analysis

First, the differential expression of 251 homeobox genes between HNSCC tissues and normal samples in 546 patients was examined using the R package “limma” (12). 72 differentially expressed genes (DEGs) in homeobox gene family were selected with  $p$ -values  $< 0.05$  and  $|\log_2$  fold change (FC)|  $> 1.5$  as thresholds. The results were visualized using a volcano plot by R package “ggplot2”, and the classification of these differentially expressed homeobox genes was displayed as a heatmap using the R package “pheatmap”. According to the classification, 8 DEGs belong to the LIM homeobox gene family and were kept for subsequent analysis. We also performed a Student’s paired t-test to validate the differential expression of selected LHXs between tumor and corresponding normal samples. Moreover, to assess whether the

differentially expressed LHXs can distinguish between normal and tumor samples, we performed a receiver operating characteristic curve (ROC) analysis and calculated the area under the curve (AUC) using the R package “pROC” (13). In addition, the co-expression relationships of selected LHXs were analyzed and visualized using the R package “ggcorrplot” (14).

## Clinical and Survival Data Analysis

The relationships between LHXs expression and the following clinical characteristics were assessed: clinical stage (I, II, III, IV), pathological grade (G1, G2, G3-4), primary tumor site (oral cavity, pharynx, larynx), and HPV status (Positive, Negative) using the R package “ggstatsplot” (15). In survival analysis, patients were divided into high and low expression groups based on the cutoff value of selected LHXs calculated by the R Package “survminer” (16). After that, the overall survival (OS), disease-specific survival (DSS), and Disease-free survival (DFS) with a 5-year follow-up were estimated with the Kaplan-Meier (KM) method and the Cox proportional hazard model.

## Gene Set Enrichment Analysis and Immune Infiltration Analysis

To identify the potential pathways that are associated with selected LHXs, we first divided the HNSCC patients into two groups by the median expression value of LHXs. DESeq2 was then applied to find the significant differential gene sets between these two groups. After that, we used “fgsea” (32) and “clusterProfiler” (34) in R software to run gene set enrichment analysis (GSEA) and visualize the results. The KEGG gene sets (c2.cp.kegg.v7.4.entrez) and hallmark gene sets (h.all.v7.4.entrez) was downloaded from the Molecular Signature Database v7.4 (<https://www.gsea-msigdb.org/gsea/msigdb>) (17).

The infiltrating levels of 22 immune cell types in each HNSCC sample were calculated using the CIBERSORT algorithm (<https://cibersort.stanford.edu/>) (18). Samples with CIBERSORT  $p < 0.05$  were considered eligible for subsequent analysis and divided into high and low groups according to the median expression value of LHXs. Then, the tumor-infiltrating immune cell levels between these two groups were compared. Moreover, another immune infiltration data was analyzed and visualized by the TIMER database (Tumor Immune Estimation Resource, <https://cistrome.shinyapps.io/timer/>) (19).

## Analysis of DNA Methylation and mRNA Expression

In order to investigate the epigenetic mechanism behind the abnormal expression of LHXs in HNSCC, the methylation levels of LHXs promoters between HNSCC tissues and normal samples were assessed using the UALCAN database (<http://ualcan.path.uab.edu/>) (20). After that, we utilized the MEXPRESS database (<https://mexpress.be>) (21) to investigate the association between LHXs expression and their DNA methylation status.

# Statistical Analyses

All statistical analyses were performed using R software 4.1.1(2021-08-10). Normality distribution was analyzed using the Shapiro–Wilk test and the equality of variances was calculated with Levene's test. Expression correlation between selected LHXs was tested by Pearson's correlation analysis. To compare the expression levels of LHXs according to clinical variables, a Welch's unequal variances t-test (for 2 groups) or a Welch's ANOVA (3 or more groups) was used based on the number of levels in the grouping parameters. The post-hoc test was performed using the Games-Howell test and the p-value was adjusted using the False Discovery Rate (FDR) method. For the KM curve, we used the log-rank test to evaluate the differences between groups. The Cox proportional hazards regression was applied with the Wald test and the tied event times were handled with the Efron approximation method. For GSEA analysis, p-values were calculated based on the adaptive multilevel splitting Monte Carlo approach (22). |Normalized enrichment scores (NES)| > 1, adjusted p-value < 0.05 and FDR q-value < 0.25 were considered significant.

## Results

### Differentially Expression of LHXs in HNSCC Tumor Samples

The flow diagram of this study is shown in Figure 1. We first detected 72 differentially expressed genes (DEGs) in the homeobox gene superfamily between HNSCC tumor and normal tissues using  $p < 0.05$  and  $|\log_2FC| > 1.5$  as cutoffs (Additional file 2, Table S2). These DEGs belong to 6 subclasses (LIM, ANTP, CUT, POU, PRD, TALE, Figure 2). Among 12 LIM homeobox genes (LHXs), 6 LHXs (LHX1, LHX2, LHX3, LHX5, LHX9, LMX1B) were indicated as upregulated, while ISL1 and LMX1A were found to be downregulated in HNSCC patients. For ISL2, LHX4, LHX6, and LHX8, no significant changes were observed (Figure 3A). The correlation of their expression between each other was calculated using Pearson's correlation analysis and shown in Figure 3B. The gene with the most co-expression genes was LHX1, which was positively correlated with LHX5, LMX1B, LHX3, and negatively correlated with LHX2, ISL1, and LMX1A (Pearson's correlation coefficient = 0.24, 0.23, 0.11, -0.26, -0.18, -1, respectively).

Next, we validated the altered expression of 8 LHXs in tumor and corresponding normal tissue using paired t-test. Similarly, significant up-regulation of LHX1, LHX2, LHX3, LHX5, LHX9, LMX1B and downregulated of ISL1 and LMX1A were examined in HNSCC tissues (all  $p < 0.0001$ , Figure 3C). Furthermore, the ability to distinguish between normal and tumor samples was tested using the ROC curve. As shown in Figure 3D, all the selected LHXs showed favorable accuracy (LHX5: AUC = 0.956, LHX1: AUC = 0.919, LHX2: AUC = 0.913, LHX9: AUC = 0.815, LHX3: AUC = 0.757, LMX1B: AUC = 0.776, LMX1A: AUC = 0.750, ISL1: AUC = 0.735).

### Association Between LHXs Expression and Clinicopathological Features

To understand the clinical significance of the selected LHXs, we analyzed the relationships between clinical variables and LHXs levels in patients with HNSCC. As shown in Figure 4, the expression of LHX1 and LHX2 was correlated with clinical stages ( $p= 0.041$ ,  $p= 0.019$ , respectively) and patients with more advanced clinical stages have higher LHX2 expression levels (Figure 4A). As for LHX2 and LMX1A, high mRNA expression was correlated with advanced tumor grade ( $p= 7.23e-5$ ,  $p= 0.01$ , respectively), in contrast to LHX1 which was highly expressed in G1 tumors ( $p= 0.015$ ) (Figure 4B). Figure 4C displayed the significant correlations between LHX1, LHX2, LHX5 expression and primary tumor sites (all  $p < 0.0001$ ). The highest expression of LHX1, LHX2, and LHX5 was evaluated in the oral cavity, pharynx and larynx, respectively. For HPV status, Figure 4D showed a remarkable higher expression for LHX1 and LHX5 in HPV-negative patients ( $p = 8.98e-13$  and  $1.57e-5$ ). Conversely, LHX2 was found to express highly in HPV-positive patients ( $p=1.22e-10$ ).

## Prognostic Values of LHXs Expression in Patients With HNSCC

Figure 5 showed the association between patient survival (OS, DSS, DFS) and LHXs expression levels using the KM curve, LHX1 tended to have significantly adverse prognosis of OS, DSS, and DFS when expressed highly (OS and DSS,  $p < 0.0001$ ; DFS,  $p = 0.0055$ ). High expression of ISL1 was correlated with good prognosis for OS and DSS ( $p = 0.017$  and  $0.011$ , respectively), but not for DFS ( $p = 0.4$ ). High LHX2 expression was associated with favorable OS and DFS ( $p = 0.048$  and  $0.02$ , respectively), but not DSS ( $p = 0.11$ ). LHX5, as well as LMX1B, was correlated with unfavorable OS and DSS (LHX5: OS,  $p = 0.0053$ ; DSS:  $p = 0.0014$ ; LMX1B: OS,  $p = 0.00016$ ; DSS,  $p = 0.0087$ ), and showed no statistical association with DFS (LHX5:  $p = 0.31$ ; LMX1B:  $p = 0.16$ ). High level of LHX9 was related to good DSS and DFS ( $p = 0.032$  and  $0.038$ , respectively), but not OS ( $p = 0.17$ ). On the contrary, high expression level of LMX1A was statistically associated with adverse prognosis in patients with HNSCC (DSS,  $p = 0.089$ ; DFS,  $p = 0.03$ ), but independent of OS ( $p = 0.21$ ). There is no significant relationship between the expression of LHX3 and OS, DSS or DFS in patients with HNSCC.

The findings were further verified using Cox regression analysis. As shown in Figure 6, LHX1, LHX5, and LMX1B were risk factors for OS and DSS (OS: LHX1: HR, 2.15; 95% CI, 1.55-2.98; LHX5: HR, 1.58; 95% CI, 1.16-2.15; LMX1B: HR, 1.92; 95% CI, 1.36-2.71; DSS: LHX1: HR, 2.56; 95% CI, 1.73-3.78; LHX5: HR, 1.75; 95% CI, 1.24-2.57; LMX1B: HR, 1.75; 95% CI, 1.15-2.66; all  $p < 0.05$ ), while ISL1 and LHX2 were related to favorable OS (ISL1: HR, 0.51; 95% CI, 0.3-0.86; LHX2: HR, 0.68; 95% CI, 0.46-1; all  $p < 0.05$ ). ISL1 and LHX9 were predictors of favorable DSS (HR, 0.42; 95% CI, 0.21-0.84; HR, 0.68; 95% CI, 0.48-0.97; respectively, all  $p < 0.05$ ). In terms of DFS, LHX1 and LMX1A were indicated as unfavorable predictors (HR, 4.7; 95% CI, 1.41-15.67; HR, 2.31; 95% CI, 1.06-5.05; respectively, all  $p < 0.05$ ), while LHX2 and LHX9 were indicated as favorable factors (HR, 0.13; 95% CI, 0.02-0.97; HR, 0.41; 95% CI, 0.17-0.98; respectively, all  $p < 0.05$ ).

# LHXs Correlated Genes Are Involved in Numerous Biological Processes

After indicated selected LHXs as prognosis factors for patient survival, we conducted the GSEA analysis to assess these genes at a functional level (Additional file 2, Table 3). We found that high-LHX1 expression exhibited enrichment of genes associated with environmental information processing pathway (ECM-receptor interaction and ) and immune system pathway (RIG-I-like receptor signaling pathway, cytosolic DNA-sensing pathway ). Gene sets that are involved in metabolism processes like metabolism of xenobiotics by cytochrome P450 were associated with low expression of LHX1 and LMX1B. As for LHX5, GSEA analysis revealed that pathways like focal adhesion and ribosome were up-regulated and pathways related to immune response (T cell receptor signaling pathway, primary immunodeficiency, and Natural killer cell mediated cytotoxicity) were down-regulated. Pathways implicated in cancer invasion and metastases like focal adhesion and tight junction were activated in the high-expression groups of LHX5, LMX1A, LMX1B and low-expression groups of LHX2 and LHX9. The enrichment of post-transcriptional control related pathways like ribosome, spliceosome, proteasome, and RNA degradation was correlated to high LHX3 and LHX5 expression while low ISL1, LHX2 and LMX1A expression. Moreover, LHXs have a close relationship with immune-related pathways, which were enriched among the low-expression groups of ISL1 (cytosolic DNA sensing pathway, antigen processing and presentation, RIG-I-like receptor signaling pathway), LHX2 (cytosolic DNA sensing pathway), LHX9(intestinal immune network for IgA production, T-cell receptor signaling pathway), and LMX1B (primary immunodeficiency), Figure 7A.

Next, differential gene sets associated with high and low LHXs expression phenotypes with regard to hallmark gene sets (h.all.v7.4) were analyzed (Additional file 2, Table S4). High LHX1 expression was related to the process of IL6 JAK STAT3 signaling, TGF beta signaling and angiogenesis. High expression of LHX3 was a connection with G2M checkpoint, MYC family targets, DNA repair, TGF alpha response, and E2F targets. LHX5 and LMX1B high expression was associated with angiogenesis and epithelial-mesenchymal transition (EMT). Low ISL1 expression corresponded with the process of reactive oxygen species (ROS) pathway, TGF beta signaling and MYC targets v2. Low LHX2 expression corresponded with angiogenesis, epithelial-mesenchymal transition, P53 pathway, TNF signaling via NF-kB, stress reaction pathways like unfolded protein response and UV response, immune response pathways like inflammatory response, interferon-alpha response, and interferon-gamma response. Low LHX9 expression corresponded with the process of IL2 STAT5 signaling, KRAS signaling and immune response pathways like allograft rejection and inflammatory response, Figure 7B.

## Immune Infiltration Is associated with LHXs Expression

Figure 8 displayed the CIBERSORT analysis of the relationship between tumor-infiltration immune cells and LHXs expression. ISL1 expression level had negative associations with the ratio of resting memory CD4+ T cells, M0 macrophages, M1 macrophages, and positive relationships with the ratio of plasma

cells. The expression levels of LHX1 and LMX1B showed positive correlations with the proportion of resting memory CD4+ T cells, resting NK cells, CD4+ T cells, M0 macrophages, as well as M1 macrophages, and negative association with the ratio of plasma cells and CD8+ T cells. LHX2 expression was positively correlated with the infiltration of plasma and CD4+ T cells, but negatively correlated with M0 macrophages and activated dendritic cells. A higher proportion of resting NK cells and M1 macrophages, while a lower proportion of activated dendritic cells was observed in the LHX3 high-expression group. LHX5, as well as LMX1A, had high M0 macrophages frequency and low plasma cells, CD4+ T cells, M1 macrophages frequencies in their high-expression group. As for LHX9, the CIBERSORT output revealed that the M1 macrophages and activated dendritic cells proportion were significantly higher in LHX9 high-expression group.

The tumor-infiltration immune cells according to LHXs expression evaluated in Timer database was shown in Figure 9. The increased immune cells infiltration in total HNSCC patients was found to correlated with the expression of ISL1 (B cell: cor = 0.209, p = 4.07e-6; CD8+ T cell: cor = 0.097, p = 3.51e-2; CD4+ T cell: cor = 0.143, p = 1.71e-3), LHX2 (B cell: cor = 0.181, p = 6.96e-5; CD8+T cell: cor = 0.157, p = 6.02e-4; CD4+ T cell: cor = 0.202, p = 8.18e-6; Macrophage: cor = 0.111, p = 1.43e-2; Neutrophil: cor = 0.138, p = 2.47e-3; Dendritic cell: cor = 0.191, p = 2.41e-5), and LMX1A (B cell: cor = 0.158, p = 5.46e-4; CD8+ T cell: cor = 0.137, p = 2.78e-3; CD4+ T cell: cor = 0.253, p = 1.99e-8; Macrophage: cor = 0.31, p = 3.30e-12; Neutrophil: cor = 0.11, p = 1.59e-2; Dendritic cell: cor = 0.198, p = 1.16e-5). The decreased immune cells infiltration was evaluated to be related to the expression of LHX5 (B cell: cor = -0.139, p = 2.36e-3; CD8+ T cell: cor = -0.19, p = 3.15e-5; Neutrophil: cor = -0.119, p = 9.35e-3; dendritic cell: cor = -0.108, p = 1.72e-2) and LHX9 (B cells: cor = -0.119, p = 1.77e-2; Macrophages: cor = -0.188, p = 3.32e-5; Dendritic cell: cor = -0.079, p = 8.41e-2). LHX1 was positively related to CD4+ T cell, neutrophil, and dendritic cell infiltration (cor = 0.117, p = 1.04e-2; cor = 0.177, p = 9.42e-5; cor = 0.094, p = 3.81e-2; respectively), but negatively related to B cell and CD8+ T cell infiltration (cor = -0.194, p = 2.04e-5; cor = -0.159, p = 5.14e-4; respectively). LMX1B was examined with positive association a decrease of B cell (cor = -0.145, p = 1.52e-3) infiltration, as well as an increase of Macrophage (cor = 0.098, p = 3.08e-2) infiltration.

## Relationship between DNA Methylation and LHXs Dysregulation

To understand the epigenetic regulatory mechanisms underlying the abnormal expression of LHXs in HNSCC, we explored the methylation level of LHXs promoters in the UALCAN database. The significant hyper-methylation of ISL1 (p = 1.624e-12), LHX2 (p = 5.384e-8), LHX5 (p < 1e-12), LHX9 (p < 1e-12), LMX1A (p < 1e-12), and LMX1B (p < 5.754e-4) promoters and hypo-methylation of LHX3 (p = 5.384e-8) promoter was displayed in Figure 10A. There is no statistically significant change in the methylation level of LHX1 promoters between HNSCC tissues and normal samples (p = 0.145).

We further explored the relationship between DNA methylation and the mRNA expression levels of LHXs that are down-regulated in HNSCC tissues using MEXPRESS. As shown in Figure 10B, low ISL1 expression was found to be significantly associated with the hyper-methylation for all promoter regions probes in HNSCC tissues (Pearson correlation coefficients:  $-0.329 \sim -0.537$ , all  $p < 0.001$ ). As for LMX1A, we found the 3 significant methylation sites (cg09248345, cg04351049, and cg09069138) in the DNA sequences of LMX1A that were negatively associated with its expression levels (Figure S1).

## Discussion

HNSCC is a common malignancy associated with high aggressiveness and many difficulties in successful treatment. Given the significant morbidities and high recurrence rates associated with conventional treatments such as surgery, radiation therapy, and chemotherapy, innovative approaches are urgently needed to effectively treat HNSCC(4). Although targeted therapy and immunotherapy have shown great promise in recent years, the long-term efficacy is still not ideal due to the lack of specific targets(23, 24). Homeobox genes comprise one of the largest groups of annotated oncogenes and have been reported to participate in the development and progression of various cancers. However, as an important group in homeobox genes, there are only sporadic studies on the roles of LHX genes in HNSCC.

Here, we conduct a bioinformatic analysis of TCGA data to determine the roles of LHXs in patients with HNSCC. Results from our study demonstrated that among 12 LHXs, 8 genes had altered expression levels in HNSCC tumor tissues compared to normal samples and their expression levels were significantly correlated with clinic-pathological characters such as tumor grade, tumor localizations as well as HPV status. Besides, LHX1, LHX5, LMX1A, and LMX1B were indicated as unfavorable prognosis predictors and ISL1, LHX2, as well as LHX9 were indicated as favorable predictors through KM analysis and Cox regression analysis. Moreover, the GSEA and immune infiltration analysis confirmed their anticipation in various biological processes and association with immune cell infiltration patterns. Finally, we determined the hyper-methylation and hypo-methylation of ISL1, LHX2, LHX5, LHX9, LMX1A, LMX1B promoter and hypo-methylation of LHX3 promoter in HNSCC tissues, then verified the relationship between the abnormal DNA methylation and aberrant mRNA expression of ISL1 as well as LMX1A in HNSCC. To our knowledge, this is the first comprehensive analysis to explore the mRNA expression, clinical correlation, prognostic values and underlying mechanisms of distinct LHXs in HNSCC.

By reviewing the literature, we observed that only LMX1B and LHX1 have been previously described in HNSCC. Smith et al. (25) found that LMX1B was significantly over-expressed in laryngeal squamous cell carcinoma (LSCC) tissues and was an independent prognostic indicator in terms of OS and (Recurrence Free Survival) RFS. Besides, they suggested that elevated gene body methylation could be responsible for its upregulation. Consistent with the literature, our current research found that LMX1B mRNA was more highly expressed in HNSCC tissues, and was correlated with poor OS and DSS. Besides, we found its expression to be significantly related to tumor sites and HPV states. LMX1B is highly expressed in the oral cavity and HPV-negative patients. Interestingly, in terms of tumor grade, the expression level of LMX1B tended to be lower in higher tumor grades.

Previous studies have found LHX1 to be highly expressed in gastric, pancreatic cancers, chronic leukemia, human clear cell renal cell carcinoma (ccRCC), both primary and metastasis medulloblastoma, as well as hepatitis B virus-related hepatocellular carcinoma (26-30). Dormoy et al. found that LHX1 knockdown decreased ccRCC cell motility, migration, invasion, and clonogenicity. Functional Genomics identified LHX1 as one of the drivers of Medulloblastoma Dissemination. In terms of HNSCC, studies showed over-expression of LHX1 was found in oral tongue squamous cell carcinoma, and KM analysis indicated high expression of LHX1 with poor OS in OTSCC (31). In our study, LHX1 was significantly up-regulated in HNSCC tissues compared to normal samples, and the ROC curve confirmed that LHX1 was capable of distinguishing between normal and tumor samples (with AUC = 0.919). In addition, patients with high LHX1 expression had poor OS, DSS, and DFS, which indicates LHX1 as a diagnosis and prognostic biomarker for HNSCC. Furthermore, GSEA analysis in HNSCC by up-regulating ECM-receptor interaction and genes participated in IL6 JAK STAT3 signaling, TGF beta signaling and angiogenesis. Notably, the effects of LHX1 on cell movements and metastasis development are consistent with earlier findings in the literature, although in other tumor types. Dormoy et al. (32) showed that LHX1 knockdown decreased ccRCC cell motility, migration, invasion, and clonogenicity, and Mumert's (33) functional genomics experiment identified LHX1 as one of the drivers of Medulloblastoma Dissemination.

ISL1 is a LIM homeodomain transcription factor that plays a critical role in the development of function in the development of the nervous system, heart, and pancreas (34-36). Previous researches have indicated ISL1 as a good marker for well-differentiated pancreatic neuroendocrine neoplasms (37, 38). Zhang et al. (39) demonstrated that ISL1 was significantly over-expressed in non-Hodgkin lymphoma(NHL), and could contribute to NHL proliferation and progression by promoting the transcription of c-Myc. The in vitro and in vivo experiments conducted by Guo et al. (40) revealed that ISL1 was frequently overexpressed in gastric cancer (GC) samples and associated with worse clinical outcomes. Patients with positive ISL1 staining tended to have poor prognosis and Cox regression analysis proved ISL1 expression as an independent prognostic factor. In our present study, the expression level of ISL1 was lower in HNSCC tissues compared to normal tissues, and low ISL1 level was correlated with poor OS and DSS. The KEGG pathway analysis and hallmark analysis for ISL1 confirmed its role as an anti-tumor gene in HNSCC by down-regulating post-transcriptional pathways such as ribosome and proteasome and genes participated in reactive oxygen species pathway, TGF-beta signaling, as well as MYC targets. Using UALCAN, we found hyper-methylation of ISL1 promoter in HNSCC tissues, and confirmed that ISL1 expression and promoter methylation level are negatively correlated using MEXPRESS, suggesting that DNA methylation may be one of the main mechanisms of downregulation of ISL1 gene expression in HNSCC.

LHX2 is an evolutionarily conserved, versatile and powerful regulator of multiple developmental phenomena such as tissue patterning, control of cell differentiation, and axon pathfinding (41). Recently, the role of LHX2 in carcinogenesis has gained increasing attention. The high expression level of LHX2 was observed in non-small cell lung cancer (42), cervical cancer (43), chronic myelogenous leukemia (44) pancreatic ductal adenocarcinoma (PDA) (45) and osteosarcoma (46). Besides, previous studies have discovered the functions of LHX2 as an oncogenic factor enhancing proliferation, metastasis and

chemoresistance in many cancers (42, 43). Zhou et al. (45) reported that LHX2 promoted PDA cells proliferation in vitro/vivo and knockdown of LHX2 reduced cell viability. Functional experiments in vivo and in vitro demonstrated that LHX2 plays an important role during EMT and tumor progression in breast carcinogenesis (47). Moreover, LHX2 has been identified as an autophagy inhibitor via mTOR signaling in osteosarcoma (46). On the other hand, LHX2 exhibited low expression in extranodal diffuse large B-cell lymphoma and can be used as independent predictors for the prognosis of glioblastoma (48, 49). In this study, LHX2 was highly expressed in HNSCC tissues and capable of discriminating between normal and tumor samples (AUC = 0.913). Besides, LHX2 was highly expressed in advanced stages and higher tumor grades. Interestingly, KM curve analysis found that patients in LHX2 high-expression group had more favorable OS and DFS. In contrast to LHX1, KEGG analysis found that LHX2 was associated with the upregulation of pathways related to cytochrome P450 metabolism and downregulation of pathways like focal adhesion and ribosome. The further hallmarks analysis for LHX2 confirmed its function in deregulating multiple biological processes associated with tumorigenesis such as angiogenesis, epithelial-mesenchymal transition, P53 pathway, TNF signaling via NF- $\kappa$ B, unfolded protein response and UV response, interferon-alpha and gamma response.

LHX3, located at 9q34.3, plays a fundamental role in spinal cord motor neurons and pituitary development (50-52). Only recently, the relationship between LHX3 and cancer has been reported. Dietrich et al (53) detected improved methylation of LHX3 in ductal carcinoma compared with non-tumor tissue. One study by Lin (54) examined the expression level, prognostic value and biological role of LHX3 in lung adenocarcinoma (ADC), the results showed that LHX3 is highly expressed in tumor tissues and significantly associated with clinical factors of ADC patients. Besides, survival analyses indicate that up-regulation of LHX3 is related to adverse survival and multivariate regression analysis verified LHX3 as an independent prognostic factor. In our current study, high LHX3 expression was found in HNSCC tissues, and KEGG analysis found LHX3 to be associated with the upregulation of post-transcriptional pathways such as ribosome and spliceosome. Hallmarks analysis revealed that genes involved in the G2M checkpoint, MYC targets, E2F targets, and interferon-alpha response are up-regulated in LHX3 high-expression group. Moreover, we found that promoter methylation of LHX3 was lower in HNSCC tissues compared to normal tissues, suggesting that the hypo-methylation of LHX3 promoter may be responsible for the upregulation of LHX3 in HNSCC.

As an important paralog of LHX1, LHX5 functions as a transcriptional regulator in the control of forebrain differentiation and development (55, 56). So far, LHX5 was only found to be associated with Breast cancer. A genome-wide methylation profiling by Kim et al. (57) demonstrated that LHX5 showed significantly higher frequencies of hypermethylation in breast cancer tumor tissues, and had an independent prognostic value for this cancer. In our study, LHX5 was highly expressed in HNSCC tissues, and the ROC curve confirmed its discriminability between normal and tumor samples (AUC = 0.956). In survival analysis, we found patients with high LHX5 expression tend to have unfavorable OS and DSS. KEGG analysis for LHX5 confirmed its role as an oncogene gene in HNSCC by up-regulating focal adhesion, glutathione metabolism as well as the ribosome, and down-regulating immune response pathways like primary immunodeficiency and T cell receptor signaling. Further hallmarks analysis

revealed that gene sets implicated in EMT, apical junction formation, hedgehog signaling and angiogenesis were activated in LHX5 high-expression group.

LHX9 is a LIM homeodomain transcription factor functions in the control of neural cell definition (58). It has been demonstrated that hypermethylation of LHX9 leads to gene silence in high-grade gliomas of childhood (59) and follicular lymphoma (60). Moreover, an in vitro functional study by Vladimirova et al. (59) found the epigenetic alteration of LHX9 may be responsible for glioma cell migration and invasiveness. Our results showed that LHX9 was up-regulated in HNSCC tissues, and the expression levels of LHX3, LHX5, LHX2 and LHX1 were positively correlated with LHX9. GSEA analysis found that LHX9 was related to the downregulation of cytochrome P450 metabolism as well as Cytokine-cytokine receptor interaction and genes positively regulated by LHX9 are involved in MYC targets and interferon-alpha response.

LMX1A was found to play an important role in the development of the nervous system(60). Immunohistochemical analyses found LMX1A had a higher expression in meningiomas and gliomas tissues, and LMX1A high-expression was correlated with more advanced tumor grade(61). Moreover, the methylation level of LMX1A was also found to be closely related to various cancers. Dong et al (62) found that LMX1A was silenced due to DNA hypermethylation in gastric cancer and restoration of LMX1A can improve cell apoptosis and inhibit anchorage-independent growth. Here, we found that mRNA expression of LMX1A was lower in HNSCC tumor tissues, while patients with lower LMX1A expression tended to have better DFS. The KEGG pathway analysis and hallmark analysis found LMX1A was related to the down-regulation of post-transcriptional pathways such as ribosome and spliceosome and genes that participated in DNA repair, interferon-alpha signaling, as well as MYC targets. Using UALCAN, we found that the promoter methylation level of LMX1A was significantly improved in HNSCC tissues, suggesting that DNA methylation may be one of the mechanisms of downregulation of LMX1A gene expression in HNSCC.

## Conclusions

In the present study, the exact roles of LIM homeobox genes in HNSCC were systematically investigated for the first time. The findings suggested the oncogenic roles of LHX1,5 and LMX1B as well as the tumor-suppressor roles of ISL1 and LHX2 in patients with HNSCC, indicating these LHXs as potential diagnostic and prognostic biomarkers for the improvement of HNSCC survival and prognostic accuracy.

## Abbreviations

LHX: Lim homeobox gene; HNSCC: head and neck squamous cell carcinoma; KM: Kaplan-Meier; GSEA: gene set enrichment analysis; ROC: receiver operating characteristic; TCGA: The Cancer Genome Atlas; DEG: differentially expressed genes; FC: fold change; AUC: area under the curve; OS: overall survival; DSS: disease-specific survival; DFS: disease-free survival; TIMER: Tumor Immune Estimation Resource; FDR: false discovery rate; NES: normalized enrichment score; EMT: epithelial-mesenchymal transition; LSCC:

laryngeal squamous cell carcinoma; RFS: recurrence free survival; ccRCC: clear cell renal cell carcinoma; OTSCC: oral tongue squamous cell carcinoma; NHL: non-Hodgkin lymphoma; PDA: pancreatic ductal adenocarcinoma; ADC: adenocarcinoma;

## **Declarations**

## **Ethics approval and consent to participate**

Not applicable.

## **Consent for publication**

Not applicable.

## **Availability of data and materials**

The datasets generated and analysed during the current study are available in the XenaBrowser repository (<http://xena.ucsc.edu/>) (11), HomeoDB repository (<http://homeodb.zoo.ox.ac.uk/>) (10), Ualcan repository (<http://ualcan.path.uab.edu/index.html>) (20), MsigDB repository (<https://www.gsea-msigdb.org/gsea/msigdb>) (17), and MEXPRESS repository (<https://mexpress.be/>) (21).

## **Competing interests**

All financial and non-financial competing interests must be declared in this section.

## **Funding**

This work is supported by National Natural Science Foundation of China (Grant No: 3A417E443431).

## **Authors' contributions**

CL analyzed and interpreted the relevant data of TCGA, and was a major contributor in writing the manuscript. CCD assisted in organizing data and editing. BH assisted in reviewing and editing the manuscript. All authors read and approved the final manuscript.

## **Acknowledgements**

Not applicable.

## References

1. Johnson DE, Burtneß B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers*. 2020;6(1).
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-49.
3. Osazuwa-Peters N, Simpson MC, Zhao L, Boakye EA, Olomukoro SI, Deshields T, et al. Suicide risk among cancer survivors: Head and neck versus other cancers. *Cancer*. 2018;124(20):4072-9.
4. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao WH, Kim E, et al. Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States. *J Clin Oncol*. 2011;29(32):4294-301.
5. Bach I. The LIM domain: regulation by association. *Mech Dev*. 2000;91(1-2).
6. Glenn DJ, Maurer RA. MRG1 binds to the LIM domain of Lhx2 and may function as a coactivator to stimulate glycoprotein hormone alpha-subunit gene expression. *The Journal of biological chemistry*. 1999;274(51):36159-67.
7. Howard PW, Maurer RA. Identification of a conserved protein that interacts with specific LIM homeodomain transcription factors. *The Journal of biological chemistry*. 2000;275(18):13336-42.
8. Ostendorff HP, Peirano RI, Peters MA, Schlüter A, Bossenz M, Scheffner M, et al. Ubiquitination-dependent cofactor exchange on LIM homeodomain transcription factors. *Nature*. 2002;416(6876):99-103.
9. Hobert O, Westphal H. Functions of LIM-homeobox genes. *Trends Genet*. 2000;16(2):75-83.
10. Zhong YF, Holland PW. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev*. 2011;13(6):567-8.
11. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675-8.
12. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
13. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77.
14. Kassambara A. ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'. R package version 0.1.3. 2019. <https://CRAN.R-project.org/package=ggcorrplot>. Accessed 9 Dec 2021
15. Patil I. Visualizations with statistical details: The 'ggstatsplot' approach. *Journal of Open Source Software*. 2021;6(61):3167.
16. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.9. . 2021. <https://CRAN.R-project.org/package=survminer>. Accessed 2 Dec 2021
17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.

- Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.
18. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019;37(7):773-82.
  19. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer research.* 2017;77(21):e108-e10.
  20. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia.* 2017;19(8):649-58.
  21. Koch A, De Meyer T, Jeschke J, Van Criekinge W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *Bmc Genomics.* 2015;16.
  22. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. 2016.
  23. Marur S, Forastiere AA. Head and Neck Squamous Cell Carcinoma: Update on Epidemiology, Diagnosis, and Treatment. *Mayo Clinic Proceedings.* 2016;91(3):386-96.
  24. Alsahafi E, Begg K, Amelio I, Raulf N, Lucarelli P, Sauter T, et al. Clinical update on head and neck cancer: molecular biology and ongoing challenges. *Cell Death & Disease.* 2019;10(8).
  25. Fan L, Zhang A, Deng P. LMX1B mRNA expression and its gene body CpG methylation are valuable prognostic biomarkers for laryngeal squamous cell carcinoma. *Biomed Pharmacother.* 2019;117:109174.
  26. Dormoy V, Béraud C, Lindner V, Thomas L, Coquard C, Barthelmebs M, et al. LIM-class homeobox gene *Lim1*, a novel oncogene in human renal cell carcinoma. *Oncogene.* 2011;30(15):1753-63.
  27. Guertl B, Senanayake U, Nussold E, Leuschner I, Mannweiler S, Ebner B, et al. *Lim1*, an Embryonal Transcription Factor, Is Absent in Multicystic Renal Dysplasia, but Reactivated in Nephroblastomas. *Pathobiology.* 2011;78(4):210-9.
  28. Gu S, Chen K, Yin M, Wu Z, Wu Y. Proteomic profiling of isogenic primary and metastatic medulloblastoma cell lines reveals differential expression of key metastatic factors. *J Proteomics.* 2017;160:55-63.
  29. Zhang L, Tao H, Li J, Zhang E, Liang H, Zhang B. Comprehensive analysis of the competing endogenous circRNA-lncRNA-miRNA-mRNA network and identification of a novel potential biomarker for hepatocellular carcinoma. *Aging (Albany NY).* 2021;13(12):15990-6008.
  30. Shawlot W, Behringer RR. Requirement for *Lim1* in head-organizer function. *Nature.* 1995;374(6521):425-30.
  31. Wang S, Fu Z, Wang Y, Sun Y, Cui L, Wang C, et al. Correlation of carbonic anhydrase 9 (CA9) with pathological T-stage and prognosis in patients with oral tongue squamous cell carcinoma. *Ann Transl Med.* 2020;8(22):1521.

32. Hamaidi I, Coquard C, Danilin S, Dormoy V, Béraud C, Rothhut S, et al. The Lim1 oncogene as a new therapeutic target for metastatic human renal cell carcinoma. *Oncogene*. 2019;38(1):60-72.
33. Mumert M, Dubuc A, Wu X, Northcott PA, Chin SS, Pedone CA, et al. Functional genomics identifies drivers of medulloblastoma dissemination. *Cancer Res*. 2012;72(19):4944-53.
34. Bu L, Jiang X, Martin-Puig S, Caron L, Zhu S, Shao Y, et al. Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. *Nature*. 2009;460(7251):113-7.
35. Roy A, Francius C, Rousso DL, Seuntjens E, Debruyjn J, Luxenhofer G, et al. Onecut transcription factors act upstream of Isl1 to regulate spinal motoneuron diversification. *Development*. 2012;139(17):3109-19.
36. Wilfinger A, Arkhipova V, Meyer D. Cell type and tissue specific function of islet genes in zebrafish pancreas development. *Dev Biol*. 2013;378(1):25-37.
37. Schmitt AM, Riniker F, Anlauf M, Schmid S, Soltermann A, Moch H, et al. Islet 1 (Isl1) expression is a reliable marker for pancreatic endocrine tumors and their metastases. *The American journal of surgical pathology*. 2008;32(3):420-5.
38. Hermann G, Konukiewitz B, Schmitt A, Perren A, Klöppel G. Hormonally defined pancreatic and duodenal neuroendocrine tumors differ in their transcription factor signatures: expression of ISL1, PDX1, NGN3, and CDX2. *Virchows Archiv : an international journal of pathology*. 2011;459(2):147-54.
39. Zhang Q, Yang Z, Jia Z, Liu C, Guo C, Lu H, et al. ISL-1 is overexpressed in non-Hodgkin lymphoma and promotes lymphoma cell proliferation by forming a p-STAT3/p-c-Jun/ISL-1 complex. *Molecular Cancer*. 2014;13(1):181.
40. Guo T, Wen X-Z, Li Z-Y, Han H-B, Zhang C-G, Bai Y-H, et al. ISL1 predicts poor outcomes for patients with gastric cancer and drives tumor progression through binding to the ZEB1 promoter together with SETD7. *Cell Death & Disease*. 2019;10(2).
41. Chou S-J, Tole S. Lhx2, an evolutionarily conserved, multifunctional regulator of forebrain development. *Brain Research*. 2019;1705:1-14.
42. Shi X, Zhan L, Xiao C, Lei Z, Yang H, Wang L, et al. miR-1238 inhibits cell proliferation by targeting LHX2 in non-small cell lung cancer. *Oncotarget*. 2015;6(22):19043-54.
43. Shi F, Su J, Liu Z, Wang J, Wang T. miR-144 reverses cisplatin resistance in cervical cancer via targeting LHX2. *Journal of cellular biochemistry*. 2019;120(9):15018-26.
44. Wu HK, Minden MD. Transcriptional activation of human LIM-HOX gene hLH-2 in chronic myelogenous leukemia is due to a cis-acting effect of Bcr-Abl. *Biochemical and biophysical research communications*. 1997;234(3):742-7.
45. Zhou F, Gou S, Xiong J, Wu H, Wang C, Liu T. Oncogenicity of LHX2 in pancreatic ductal adenocarcinoma. *Mol Biol Rep*. 2014;41(12):8163-7.
46. Song H, Liu J, Wu X, Zhou Y, Chen X, Chen J, et al. LHX2 promotes malignancy and inhibits autophagy via mTOR in osteosarcoma and is negatively regulated by miR-129-5p. *Aging (Albany NY)*. 2019;11(21):9794-810.

47. Kuzmanov A, Hopfer U, Marti P, Meyer-Schaller N, Yilmaz M, Christofori G. LIM-homeobox gene 2 promotes tumor growth and metastasis by inducing autocrine and paracrine PDGF-B signaling. *Mol Oncol.* 2014;8(2):401-16.
48. Zhao C, Huang R, Zeng Z, Yang S, Lu W, Liu J, et al. Downregulation of USP18 reduces tumor-infiltrating activated dendritic cells in extranodal diffuse large B cell lymphoma patients. *Aging (Albany NY).* 2021;13(10):14131-58.
49. Cheng Q, Huang C, Cao H, Lin J, Gong X, Li J, et al. A Novel Prognostic Signature of Transcription Factors for the Prediction in Patients With GBM. *Frontiers in genetics.* 2019;10:906.
50. Sharma K, Sheng HZ, Lettieri K, Li H, Karavanov A, Potter S, et al. LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. *Cell.* 1998;95(6):817-28.
51. Sheng HZ, Zhadanov AB, Mosinger B, Fujii T, Bertuzzi S, Grinberg A, et al. Specification of pituitary cell lineages by the LIM homeobox gene Lhx3. *Science (New York, NY).* 1996;272(5264):1004-7.
52. Thaler JP, Lee S-K, Jurata LW, Gill GN, Pfaff SL. LIM Factor Lhx3 Contributes to the Specification of Motor Neuron and Interneuron Identity through Cell-Type-Specific Protein-Protein Interactions. *Cell.* 2002;110(2):237-49.
53. Dietrich D, Lesche R, Tetzner R, Krispin M, Dietrich J, Haedicke W, et al. Analysis of DNA Methylation of Multiple Genes in Microdissected Cells From Formalin-fixed and Paraffin-embedded Tissues. *Journal of Histochemistry & Cytochemistry.* 2009;57(5):477-89.
54. Lin X, Li Y, Wang J, Han F, Lu S, Wang Y, et al. LHX3 is an early stage and radiosensitivity prognostic biomarker in lung adenocarcinoma. *Oncology Reports.* 2017;38(3):1482-90.
55. Sheng HZ, Bertuzzi S, Chiang C, Shawlot W, Taira M, Dawid I, et al. Expression of murine Lhx5 suggests a role in specifying the forebrain. *Developmental Dynamics.* 1997;208(2):266-77.
56. Zhao Y, Hermes E, Yarolin MC, Westphal H. Genomic structure, chromosomal localization and expression of the human LIM-homeobox gene LHX5. *Gene.* 2000;260(1-2).
57. Kim MS, Lee J, Oh T, Moon Y, Chang E, Seo KS, et al. Genome-wide identification of OTP gene as a novel methylation marker of breast cancer. *Oncology Reports.* 2012;27(5):1681-8.
58. Failli V, Rogard M, Mattei MG, Vernier P, Rétaux S. Lhx9 and Lhx9alpha LIM-homeodomain factors: genomic structure, expression patterns, chromosomal localization, and phylogenetic analysis. *Genomics.* 2000;64(3):307-17.
59. Vladimirova V, Mikeska T, Waha A, Soerensen N, Xu J, Reynolds PC, et al. Aberrant methylation and reduced expression of LHX9 in malignant gliomas of childhood. *Neoplasia.* 2009;11(7):700-11.
60. Bennett LB, Schnabel JL, Kelchen JM, Taylor KH, Guo J, Arthur GL, et al. DNA hypermethylation accompanied by transcriptional repression in follicular lymphoma. *Genes, Chromosomes and Cancer.* 2009;48(9):828-41.
61. Tsai W-C, Lin C-K, Yang Y-S, Chan D-C, Gao H-W, Chang F-N, et al. The Correlations of LMX1A and Osteopontin Expression to the Clinicopathologic Stages in Pancreatic Adenocarcinoma. *Applied Immunohistochemistry & Molecular Morphology.* 2013;21(5):395-400.

62. Dong W, Feng L, Xie Y, Zhang H, Wu Y. Hypermethylation-mediated reduction of LMX1A expression in gastric cancer. *Cancer Sci.* 2011;102(2):361-6.

## Figures

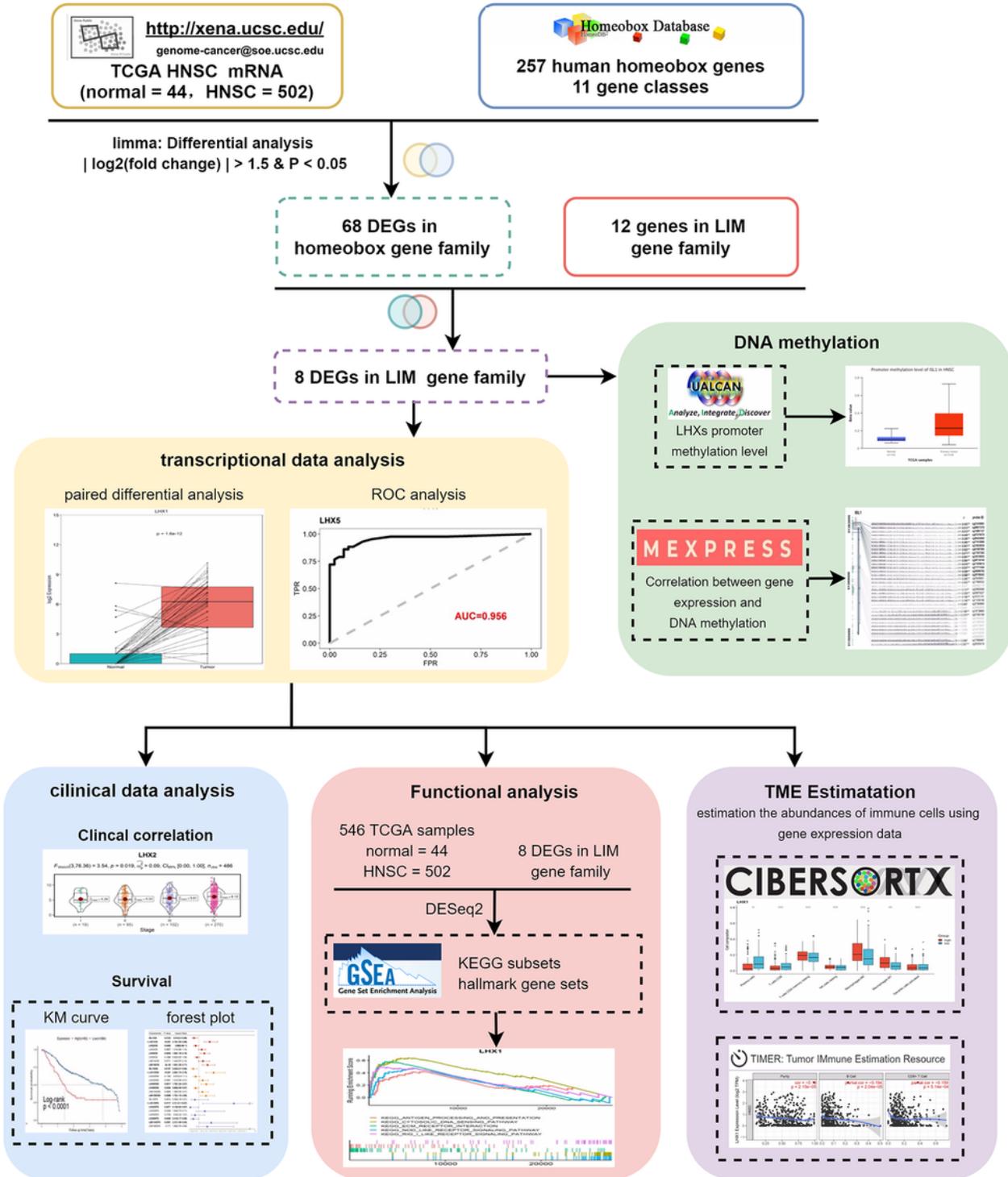
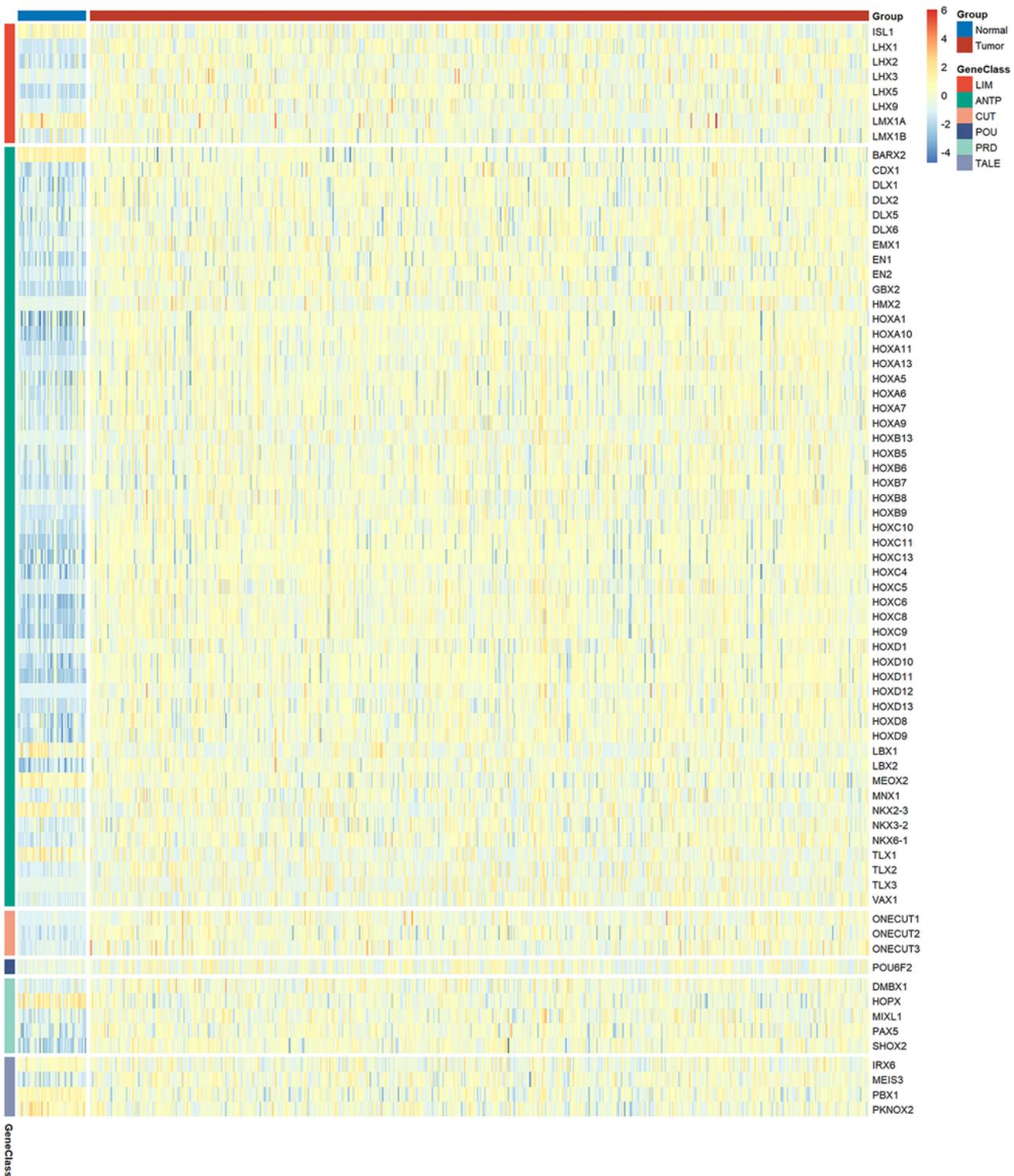


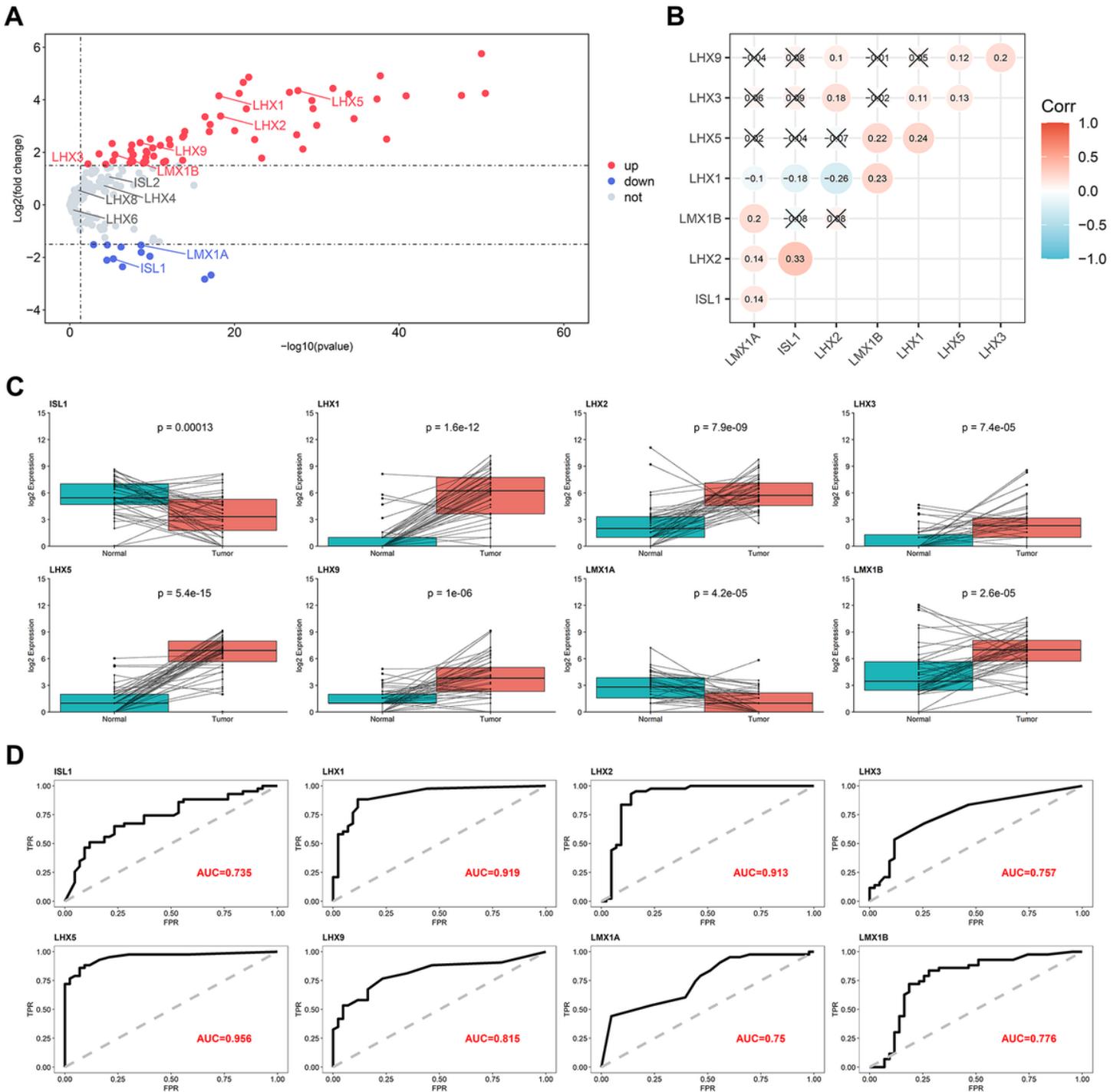
Figure 1

Flow diagram of our study on the transcriptional expression, clinical correlation, prognostic value, and underlying mechanisms of LHXs in HNSCC.



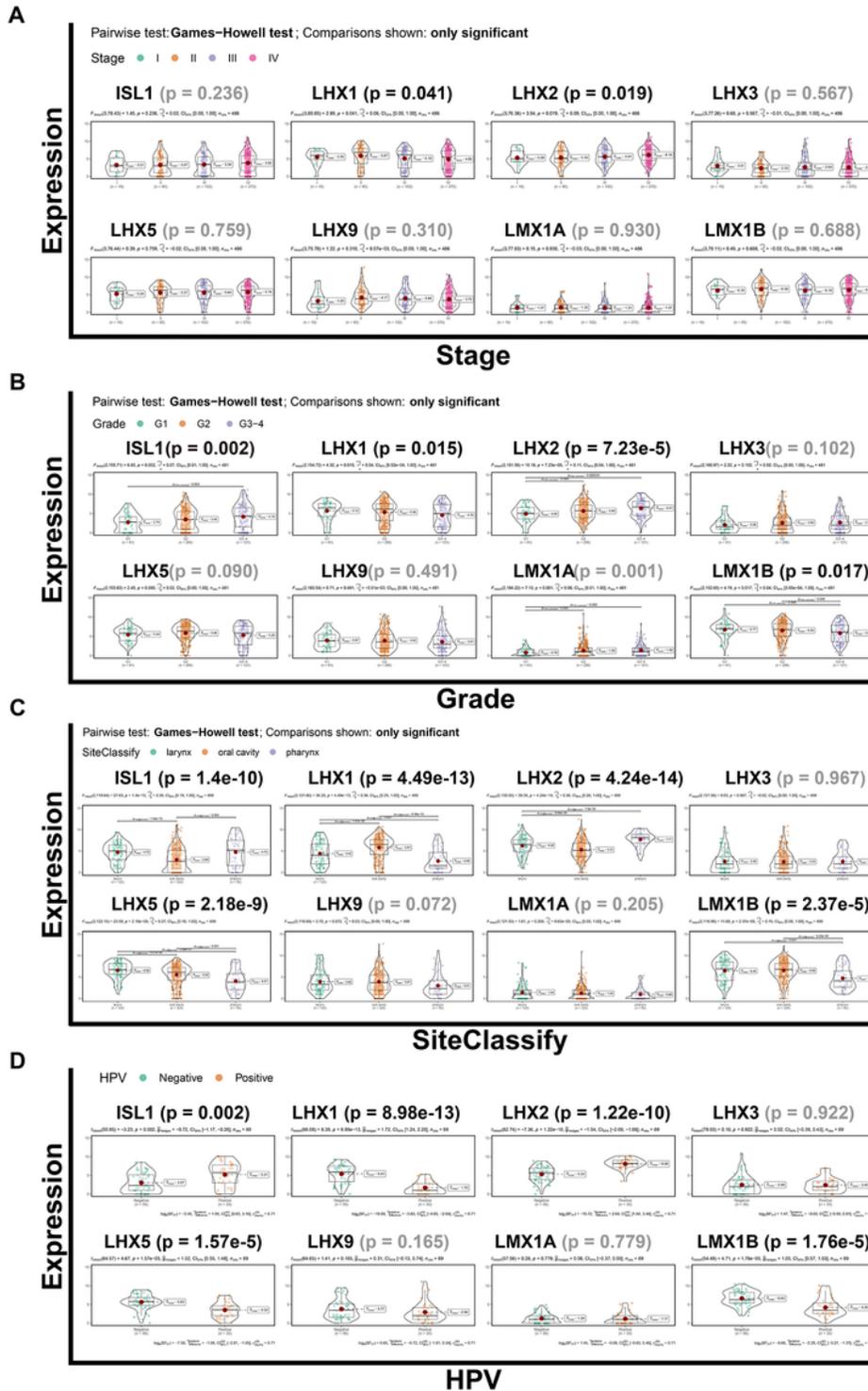
**Figure 2**

Heatmap of the expression and classification of 72 DEGs ( $p < 0.05$  and  $|\log_2FC| > 1.5$ ) in homeobox genes between 502 HNSCC samples and 44 normal samples.



**Figure 3**

The differential expression and discriminatory power of LHXs in HNSCC. (A) Volcano plot of 72 DEGs ( $p < 0.05$  and  $|\log_2FC| > 1.5$ ) in homeobox genes between 502 HNSCC samples and 44 normal samples; genes that are labeled out belong to the LIM homeobox gene family; (B) Heatmap of the gene expression correlation between 8 selected LHXs in 502 HNSCC patients; the color of each circle and the given number represents the correlation coefficients; circles with a cross mark indicate no significant correlation (Pearson's  $p$  value  $> 0.05$ ); (C) Paired box plots and (D) ROC analysis of 8 selected LHXs between 43 HNSCC samples and 43 corresponding normal samples.



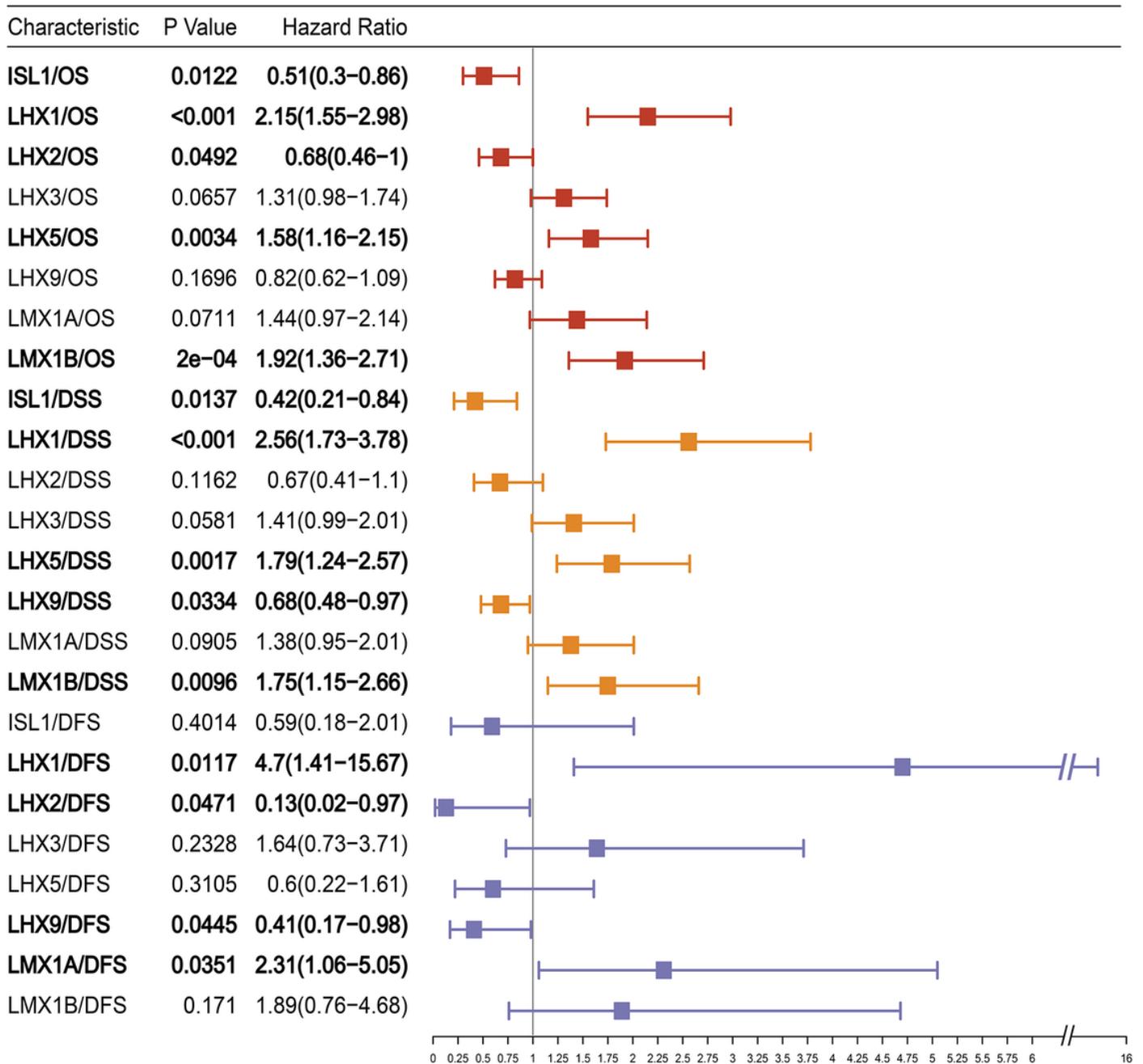
**Figure 4**

Correlation between 8 selected LHXs transcriptional expression and clinical variables in HNSCC. (A) Relationship between LHXs expression and clinical stages (I: n = 19, II: n = 95, III: n = 102, IV: n = 270); (B) Relationship between LHXs expression and tumor grades (G1: n = 61, G2: n = 299, G3-4: n = 121); (C) Relationship between LHXs expression and primary tumor sites (larynx: n = 120, oral cavity: n = 329,

pharynx: n = 50); (D) Relationship between LHXs expression and HPV status (Positive: n = 56, Negative: n = 33).

**Figure 5**

KM curves displaying the prognosis value of LHXs expression levels for patient survival (OS, DSS, DFS with a 5-year follow-up).

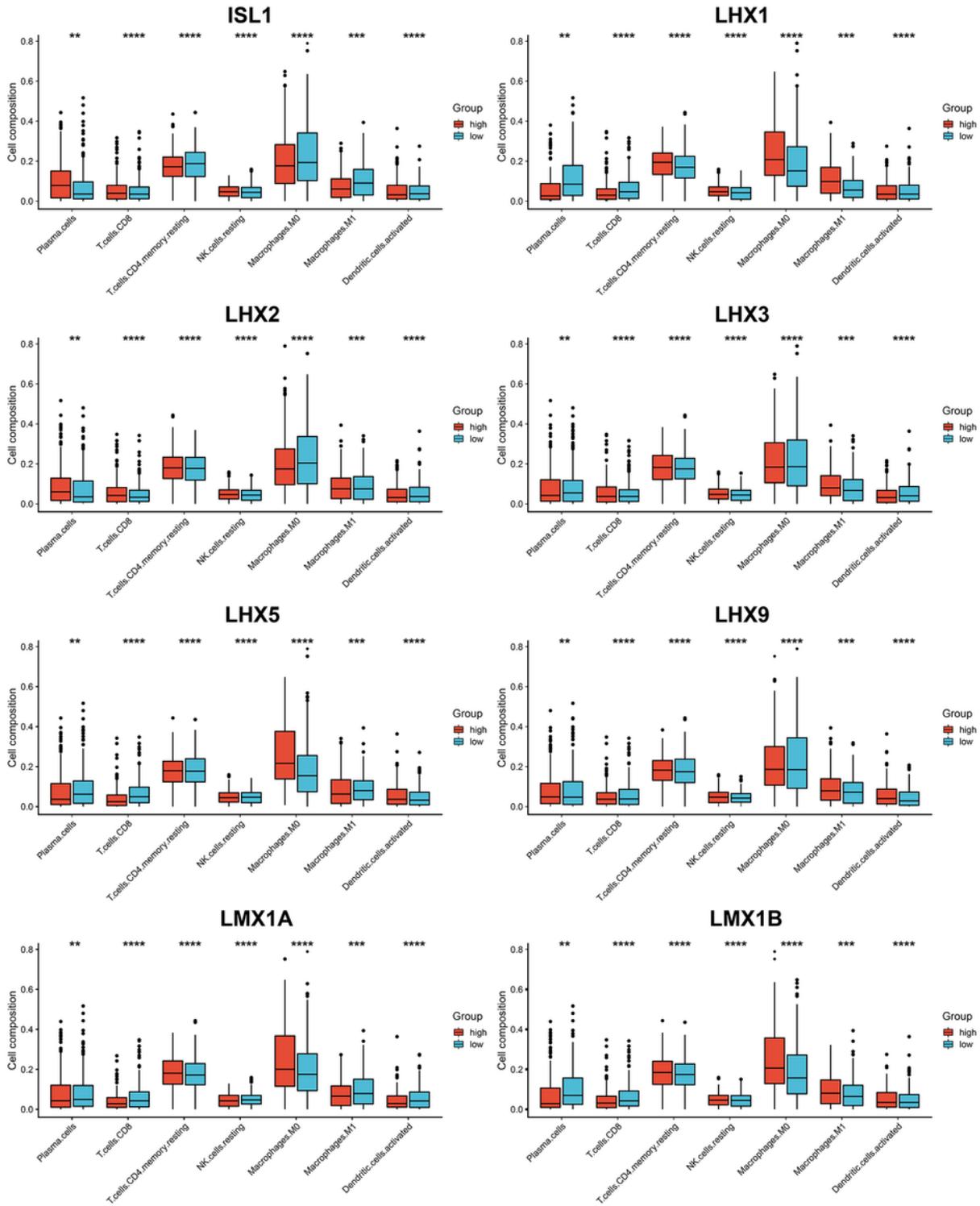


**Figure 6**

Forest plot of the Cox regression analysis for determining the prognosis value of LHXs expression levels for patient survival (OS, DSS, DFS with a 5-year follow-up). Bold font indicates statistically significant values ( $P < 0.05$ ).

## Figure 7

GSEA analysis for examining the biological processes associated with LHXs aberrant expression. (A) GSEA results showing the KEGG gene sets significantly ( $p < 0.05$ ,  $FDR < 0.25$ , and  $|NES| > 1$ ) enriched in LHXs high-expression phenotypes; (B) Hallmark Gene sets significantly ( $p < 0.05$  and  $FDR < 0.25$ ) enriched in high and low LHXs expression phenotypes;



**Figure 8**

CIBERSORT analysis of the correlation between statistically significant immune cells infiltration and 8 selected LHXs expression. Note: \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001.

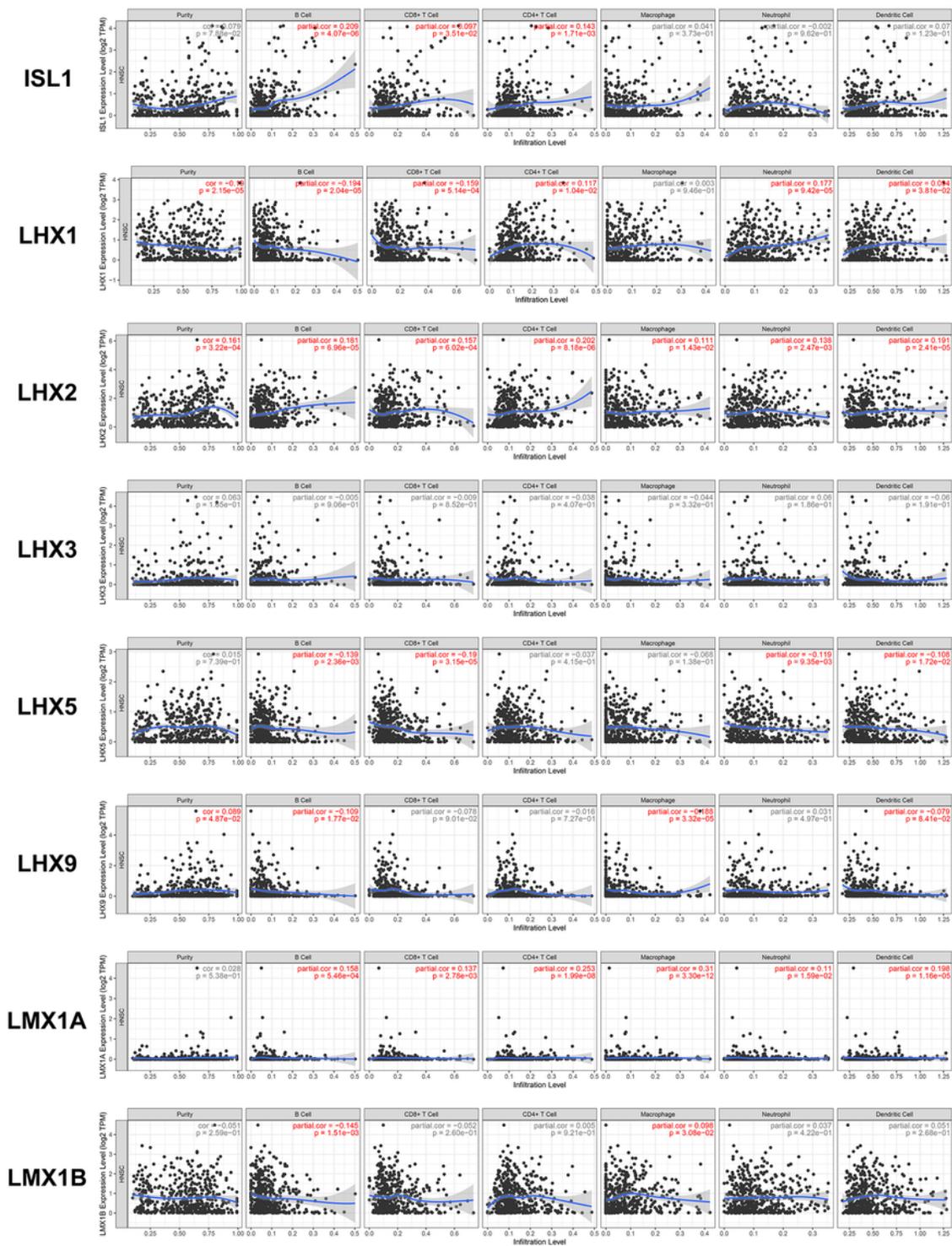


Figure 9

Timer visualization of the correlation between immune cells infiltration and LHXs expression levels (log2TPM).



## Supplementary Files

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

- [Additionalfile1.xlsx](#)
- [Additionalfile2.docx](#)
- [Additionalfile3FigureS1.tif](#)