Identification of Genes Associated With Laryngeal Squamous Cell Carcinoma by Weighted Gene Co-Expression Network Analysis

Chao Huang
Second Xiangya Hospital
Jun He
Second Xiangya Hospital
Yi Dong
Central South University
Yichao Chen
Second Xiangya Hospital
Anquan Peng
Second Xiangya Hospital
Hao Huang (xyskhuanghao@csu.edu.cn)
Central South University First Hospital: Xiangya Hospital Central South University
https://orcid.org/0000-0001-9051-4483

Primary research

Keywords: Laryngeal Squamous Cell Carcinoma, WGCNA, Differentially expressed genes, Key gene, Gene modules

Posted Date: May 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-531467/v1

Abstract

Background: Laryngeal squamous cell carcinoma (LSCC) is one of the leading malignant cancers of the head and neck. Patients with LSCC infiltration and metastasis have a poor prognosis. There is an urgent need to identify more potential targets for drugs and biomarkers for early diagnosis.

Methods: RNA sequence data from LSCC and patient clinic traits were obtained from the Gene Expression Omnibus (GEO) (GSE142083) and The Cancer Genome Atlas (TCGA) database. Differentially expressed genes (DEGs) analysis and weighted gene co-expression network analysis (WGCNA) were performed to identify the hub genes. Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, prognostic values analysis and the tumor-infiltrating immune cell (TIC) abundance profiles estimation were performed.

Results: In present study, a total of 432 DEGs, including 199 up-regulated genes and 233 down-regulated genes were screened in GSE142083 dataset. After intersecting with DEGs in TCGA, 211 common DEGs were screened. Using WGCNA, five modules were identified to be closely related to LSCC. After compared with common DEGs and performed with univariate Cox regression analysis, only eight genes, including *CEACAM6*, *FSCN1*, *INHBA*, *MYO1B*, *PLAU*, *SERPINH1*, *TJP3* and *TNFRSF12A*, were screened out to be significantly related to the prognosis of patients diagnosed with LSCC.

Conclusions: The results in present study shows that *CEACAM6*, *FSCN1*, *INHBA*, *MYO1B*, *PLAU*, *SERPINH1*, *TJP3* and *TNFRSF12A* have the potential to become new therapeutic targets and biomarkers for LSCC.

1 Introduction

Laryngeal squamous cell carcinoma (LSCC) is a common kind of head and neck squamous cell carcinomas (HNSCC) accounting for approximately 20% of all cancer patients and 2.4% of new malignancies worldwide each year [1–3]. Patients with LSCC infiltration and metastasis have a poor prognosis, and their 5-year survival rate is about 60% [4]. In recent years, despite the latest advances in comprehensive surgical, chemotherapy and radiotherapy treatment strategies, the global mortality rate associated with LSCC has not decreased [5]. There are almost no typical symptoms in the early stages of LSCC which is always being ignored by patients [6]. Therefore, identification of abnormally expressed genes in LSCC and early intervention are important strategies to prolong the survival time of LSCC patients.

In recent years, with the development of gene chip technology, cancer diagnosis methods have become more efficient and much simpler, which allow researchers to grasp cancer diagnosis information in a relatively short period of time and find the correct treatment measures [7–9]. It is also because of these advantages that the research of molecular markers of cancer has been a hot spot in recent years [10–12]. Weighted gene co-expression network analysis (WGCNA) is a powerful approach in identifying gene co-expression modules, exploring the correlation of the modules and phenotypes and discovering hub genes

that regulate critical biological processes [13–15]. However, there are few reports in the literature describing the hub genes and biomarkers in LSCC patients revealed by WGCNA.

In this study, a transcriptomes dataset of LSCC and adjacent normal tissue of patients from the Gene Expression Omnibus (GEO) was downloaded to identify the patterns and differences of expression profiles between LSCC groups and control groups using WGCNA and differentially expressed gene (DEG) screening, with the aim to understand the LSCC pathogenesis, pinpoint the molecular mechanism and provide insights into novel therapeutic targets for drugs.

2 Methods

2.1 Data collection

The workflow for the current study is presented in (**Figure 1)**. The gene expression matrix of 106 LSCC samples (53 LSCC tissues and 53 paired adjacent normal mucosa tissues) was downloaded from the GSE142083 dataset which was obtained from the GEO database [16]. The GEO expression matrix was annotated with gene symbols using the information form the GPL20301 Illumina HiSeq 4000 platform file and TPM and $\log_2$ transformed in R (version 4.0.4) if necessary. Principal component analysis (PCA) was performed and the outliers were excluded (GSM4219698 and GSM4219730) (**Figure S1**), only 52 LSCC tissues and 52 paired normal tissues data were enrolled for later analysis.

2.2 DEGs identification

After removing the outliers, R language was utilized for standardization of all expression data. The limma software package (http://www.bioconductor.org/packages/release/bioc/html/limma.html) was used to complete the DEG analysis between LSSC and adjacent normal control from the GSE142083 dataset. Genes with an expression adjusted-$p < 0.05$ and $\log_2$ (FC) >1.5 were regarded as DEGs. The volcano and heatmap plots were generated by R packages. The DEGs of Head and Neck squamous cell carcinoma (HNSC) database from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) were obtained via GEPIA (Gene Expression Profiling Interactive Analysis) (http://gepia.cancer-pku.cn/) [17]. Extracellular matrix-associated genes were annotated with Matrisome Project (http://matrisomeproject.mit.edu) [18].

2.3 Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of DEGs

GO and KEGG enrichment analyses of DEG and hub modules were executed using clusterProfiler R package (http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html) [19]. Three main processes in GO analysis as follows: biological processes (BP), molecular functions (MF), and cellular components (CC). The *p*-value was conventionally set at 0.05. Plots were generated by R packages.

2.4 Weighted gene co-expression network construction

The WGCNA package of R [13] was used to construct the co-expression networks of LSCC and adjacent normal samples. Genes with mean FPKM over 0.5 were chosen. The adjacency matrices which stored the information of the whole co-expression network were created based on Pearson's correlation matrices. A topological overlap measure (TOM) matrix was created from the adjacency matrix to estimate the network's connectivity property. Average linkage hierarchical clustering was used to construct a clustering dendrogram of the TOM matrix with a minimum module size of 30. Finally, similar gene modules were merged, with a threshold of 0.25.

2.5 Validation of hub genes

The hub genes were identified as intersecting between the Midnightblue, darkgrey, Blue, Salmon and Greenyellow modules from the WGCNA and common DEGs. Then, these genes were validated for prognostic signature. R package 'survival' was utilized to conduct overall survival analysis. The Univariate Cox regression analysis was performed to investigate the correlation between gene expression and overall survival. GSVA package was performed to analyze the tumor-infiltrating immune cell (TIC) abundance profiles.

3 Results

3.1 DEG screening

We analyzed the DEG of GSE142083 by using the limma package with the threshold of $|\log_2(\text{fold-change})| > 1.5$ and adjust $p < 0.05$. A total of 432 DEGs, including 199 up-regulated genes and 233 down-regulated genes were screened in LSCC samples and adjacent normal samples. The heatmap and the volcano plot showed the expression pattern of those DEGs (**Figure 2A-B**).

The clusterprofiler package was then applied to uncover the role of DEGs in the pathogenesis of LSCC with a cutoff criterion of $p < 0.05$. For biological process group, epidermis development, cornification, skin development, keratinization and keratinocyte differentiation were significantly enriched. The genes in the cellular component group were significantly enriched in cornified envelope, collagen-containing extracellular matrix (ECM) and ECM component. In the molecular function group, the DEGs were mainly enriched in ECM structural constituent, chemokine and cytokine activity. For KEGG analysis, IL-17 signaling pathway, salivary secretion, amoebiasis, cytokine-cytokine receptor interaction and ECM-receptor interaction were significantly enriched (**Figure 2C**). Interestingly, except for immune-related genes, we also noticed that ECM-associated genes were enriched in the DEGs and most of them were high-expressed in LSCC samples (**Figure S2**). Moreover, hsa04970:Salivary secretion is one of few terms which have decreasing Z-scores. That may indicate a significantly decrease of function of salivary secretion in LSCC patients.

We then obtained 848 DEGs using a cohort of Head and Neck squamous cell carcinoma (HNSC) from TCGA using GEPIA with a same threshold. After compared with DEGs in GSE142083, 211 common DEGs were identified both in DEGs of TCGA-HNSC and GSE142083 (**Figure 2D**).

3.2 Weighted co-expression network construction and analysis

In order to detect the functional module in LSCC, we applied WGCNA package based on the GSE142083 dataset to establish the gene co-expression networks. In order to ensure the network was scale-free, empirical analysis was run to choose an optimal parameter β. Both the scale-free topology model fit index (R^2) and mean connectivity reach steady status when β is equal to 9 (**Figure 3A-B**).

A total of 28 modules were identified via average linkage hierarchical clustering and each module is represented in different color (**Figure 3C**). Among the modules, module midnightblue, module darkgrey, module blue, module salmon and module greenyellow have high correlation with cancer traits (**Figure 3D**). Accordingly, they were selected as the clinically significant modules for further analysis. A set of 400 selected genes were identified for the network heatmap (**Figure S3**).

GO and KEGG analysis were used to assess the biological function, molecular function, cellular component and KEGG pathways of genes for modules. The midnightblue module was mainly related to glycosylation events and glycosphingolipid biosynthesis, while the blue module was mainly associated with separation of sister chromatids and cell cycle pathway. The darkgrey module is associated with thermogenesis. The salmon module was mainly associated with extrinsic apoptotic signaling and ECM-receptor interaction. The greenyellow module was for biogenesis of ribosomes and mitochondrial gene expression (**Figure 4, Table S1**).

3.3 Identification of hub genes

Candidate genes were screened using the cut-off criteria: the absolute value of module membership (MM) score > 0.8 ($|MM| > 0.8$) and the absolute value of gene significance (GS) score > 0.2 ($|GS| > 0.2$). Based on this cut-off criteria, we subsequently sorted the genes according to their connectivity to select candidate genes (**Figure 3E-I**). 41 genes in midnightblue module, 14 genes in darkgrey module, 120 genes in blue module, 31 genes in salmon module and 99 genes in greenyellow module were screened out.

In order to identify the “real” key genes, we then compared those hub genes with the 211 common DEGs. A total of 41 “real” key genes from midnightblue module, blue module, salmon module and common DEGs were screened out (**Figure 5**).

3.4 Validation of key genes

In order to validate and explore the prognostic values of these “real” key genes in LSCC, we conducted univariate Cox regression analysis of overall survival using the TCGA-HNSC database. Only six genes with high expression, including *FSCN1*, *INHBA*, *MYO1B*, *PLAU*, *SERPINH1* and *TNFRSF12A*, were found to be significantly associated with shorter overall survival time among patients with cancer, while two genes (*CEACAM6*, *TJP3*) with lower expression were detected to be significantly associated with shorter overall survival time (**Figure 6A**). Furthermore, we also observed that higher expression of *FSCN1*, *INHBA*, *MYO1B*, *PLAU*, *SERPINH1* and *TNFRSF12A*, as well as lower expression of *CEACAM6* and *TJP3*, were significantly correlated with worse disease specific survival among HNSC patients (**Figure 6B**). The

scatter plot showed the same expression pattern of those eight key genes in TCGA-HNSC and GSE142083 databases (**Figure 6C**).

The distribution of TICs is an important indicator of patients' lymph node status and prognosis. In order to explore the relationship of the eight key genes expressed differently expression with the tumor immune microenvironment, we analyzed the correlation of genes and tumor-infiltrating immune subsets in TCGA-HNSC database based on GSVA package. The results showed that 24 types of TIC in tumor samples are illustrated (**Figure 7**). For the expression of *FSCN1*, *INHBA*, *MYO1B*, *PLAU*, *SERPINH1* and *TNFRSF12A*, most of them have a high correlation with infiltration degree of T helper cells 1 (Th1), T gamma delta cells (Tgd), neutrophils, natural killer (NK) cells and a high negatively correlation with NK CD56bright cells, B cells, Th17 cells. While most of these TICs were inversed with the expression of *CEACAM6* and *TJP3*.

4 Discussion

In present study, 211 common DEGs both in GSE142083 and TCGA-HNSC database were screened. Weighted co-expression networks were constructed using the WGCNA algorithm based on GSE142083 database. Five modules in the LSCC condition were detected based on the co-expression network. After intersecting with common DEGs from GSE142083 and TCGA-HNSC database, 41 "real" key genes were screened out. Analysis of functions and pathways were performed and only eight genes of *CEACAM6*, *FSCN1*, *INHBA*, *MYO1B*, *PLAU*, *SERPINH1*, *TJP3* and *TNFRSF12A* were found to be significantly related to the prognosis of patients diagnosed with LSCC.

LSCC is one of the leading malignant cancers of the head and neck. There is an urgent need to identify more potential targets for drugs and biomarkers for early diagnosis [20–22]. ECM is a fundamental and important component of all tissue organs and has been found to interact with tumor cells and regulate tumor growth, proliferation, differentiation, adhesion and metastasis [23–25]. Here, we confirmed that most of the Core-ECM genes like Collagens, Proteoglycans and Glycoproteins were high expressed in LSCC compared with adjacent normal mucosa tissues (**Figure S2**). Meanwhile, ECM-associated biological processes were also enriched in DEGs of GSE142083 (Fig. 2C). We also identified *CEACAM6*, *INHBA*, *PLAU* and *SERPINH1* as LSCC significant key genes associated with ECM (Fig. 6). Those results further confirmed the importance of ECM accumulation in LSCC. Due to the deposition of ECM, tumors in this region may be regarded as difficult to treat [26–28]. Therefore, exploring new methods to reduce the deposition of ECM may be a strategy for LSCC treatment.

Inhibin subunit beta A (*INHBA*) encodes a member of the TGF-β superfamily of proteins [29]. TGF-β signaling is critical with the processes of ECM deposition and pro-invasive oof cancer cells by epithelial-mesenchymal-transition (EMT) [30, 31]. Wu et al. also evidenced that *INHBA* as a novel biomarker for HNSCC based on data mining [32]. But the detailed mechanism of *INHBA* remains unclear.

Carcinoembryonic antigen-related cell adhesion molecule 6 (*CEACAM6*) is an important member of glycosyl phosphatidyl inositol (GPI) anchored cell surface glycoproteins and plays a role in cell adhesion [33, 34]. A lot of researches have shown that *CEACAM6* was highly associated with cancer progression. In

accordance with our data in GSE142083 and TCGA-HNSC, Bednarek et al. also reported that *CEACAM6* was downregulated in a set of 16 LSCC samples and this downregulation was dependent on the gene's promoter DNA methylation [35]. Tian et al. silenced the expression of *CEACAM6* and found an inhibition of proliferation, migration and invasion in gallbladder cancer cell lines [36]. However, the role of *CEACAM6* is likely to be multifunctional. Iwabuchi et al. has showed that only co-expression of *CEACAM6* and *CEACAM8* can lead to cell adhesion [37]. Therefore, the role of *CEACAM6* in LSCC still need more investigation.

In comparion with the other seven key genes, Tight Junction Protein 3 (*TJP3*) has not been intensively investigated in malignant cancers of the head and neck. *TJP3* is a member of the membrane-associated guanylate kinase-like (MAGUK) protein family and plays a role in the linkage between the actin cytoskeleton and tight-junctions [38, 39]. A study has shown that *TJP3* was up-regulated in ovarian cancer and could play an important role in EMT, migratory and invasive potential of ovarian cancer cells [40]. But the role of *TJP3* in LSCC is still not understood.

5 Conclusion

After the differentially expressed genes screening and weighted gene co-expression network analysis of laryngeal squamous cell carcinoma and normal samples, we find *CEACAM6*, *FSCN1*, *INHBA*, *MYO1B*, *PLAU*, *SERPINH1*, *TJP3* and *TNFRSF12A* can be defined as key genes associated with LSCC and may play a significant role in the prediction of the LSCC prognosis. The above-mentioned key genes have the potential to become the therapeutic target and biomarkers for LSCC.

6 Abbreviations

LSCC, Laryngeal squamous cell carcinoma; HNSCC, head and neck squamous cell carcinomas; WGCNA, Weighted gene co-expression network analysis; GEO, Gene Expression Omnibus; DEG, differentially expressed gene; HNSC, Head and Neck squamous cell carcinoma; TCGA, The Cancer Genome Atlas; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TIC, tumor-infiltrating immune cell; ECM, extracellular matrix; EMT, epithelial-esenchymal-ransition.

Declarations

Acknowledgement

Gene expression profiling dataset GSE142083 was downloaded from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/).

Authors' contributions

HH and AP conceived and directed the project. CH and YD collected the data and information. CH, JH and YC analysed and interpreted the data. CH and HH wrote the manuscript with the help of all the other

authors. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Project funded by China Postdoctoral Science Foundation (2020TQ0363 and 2020M682598); the National Natural Science Foundation of China (81570928).

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Ethics approval and consent to participate

The Review Board of the Second Xiangya Hospital of Central South University approved the present study.

Competing interests

The authors declare that they have no conflict of interest.

Consent for publication

All authors have read and approved the final manuscript for publication.

References

1. Liang J, Zhu X, Zeng W, Yu T, Fang F, Zhao Y. Which risk factors are associated with stomal recurrence after total laryngectomy for laryngeal cancer? A meta-analysis of the last 30 years. Braz J Otorhinolaryngol. 2020;86(4):502–12.
2. Lin HW, Bhattacharyya N. Staging and survival analysis for nonsquamous cell carcinomas of the larynx. LARYNGOSCOPE. 2008;118(6):1003–13.
3. Chen X, Gao L, Sturgis EM, Liang Z, Zhu Y, Xia X, Zhu X, Chen X, Li G, Gao Z. HPV16 DNA and integration in normal and malignant epithelium: implications for the etiology of laryngeal squamous cell carcinoma. ANN ONCOL. 2017;28(5):1105–10.
4. Marioni G, Marchese-Ragona R, Cartei G, Marchese F, Staffieri A. Current opinion in diagnosis and treatment of laryngeal carcinoma. CANCER TREAT REV. 2006;32(7):504–15.
5. Haddad RI, Posner M, Hitt R, Cohen E, Schulten J, Lefebvre JL, Vermorken JB. Induction chemotherapy in locally advanced squamous cell carcinoma of the head and neck: role, controversy, and future directions. ANN ONCOL. 2018;29(5):1130–40.
6. Gao C, Hu S. miR-506 is a YAP1-dependent tumor suppressor in laryngeal squamous cell carcinoma. CANCER BIOL THER. 2019;20(6):826–36.
7. Grosselin K, Durand A, Marsolier J, Poitou A, Marangoni E, Nemati F, Dahmani A, Lameiras S, Reyal F, Frenoy O, et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in

breast cancer. NAT GENET. 2019;51(6):1060–6.

8. Zhang P, Xia JH, Zhu J, Gao P, Tian YJ, Du M, Guo YC, Suleman S, Zhang Q, Kohli M, et al. High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. NAT COMMUN. 2018;9(1):2022.
9. Braun S, Enculescu M, Setty ST, Cortes-Lopez M, de Almeida BP, Sutandy F, Schulz L, Busch A, Seiler M, Ebersberger S, et al. Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis. NAT COMMUN. 2018;9(1):3315.
10. Teufel M, Seidel H, Kochert K, Meinhardt G, Finn RS, Llovet JM, Bruix J: **Biomarkers Associated With Response to Regorafenib in Patients With Hepatocellular Carcinoma**. *GASTROENTEROLOGY* 2019, **156**(6):1731–1741.
11. Guo B, Wu S, Zhu X, Zhang L, Deng J, Li F, Wang Y, Zhang S, Wu R, Lu J, et al. Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. EMBO J. 2020;39(1):e102190.
12. Barros-Filho MC, Dos RM, Beltrami CM, de Mello J, Marchi FA, Kuasne H, Drigo SA, de Andrade VP, Saieg MA, Pinto C, et al: **DNA Methylation-Based Method to Differentiate Malignant from Benign Thyroid Lesions**. *THYROID* 2019, **29**(9):1244–1254.
13. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC BIOINFORMATICS. 2008;9:559.
14. Ding M, Li F, Wang B, Chi G, Liu H. A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. J CELL BIOCHEM. 2019;120(6):10855–63.
15. Liu J, Zhou S, Li S, Jiang Y, Wan Y, Ma X, Cheng W. Eleven genes associated with progression and prognosis of endometrial cancer (EC) identified by comprehensive bioinformatics analysis. CANCER CELL INT. 2019;19:136.
16. Gao W, Zhang Y, Luo H, Niu M, Zheng X, Hu W, Cui J, Xue X, Bo Y, Dai F, et al. Targeting SKA3 suppresses the proliferation and chemoresistance of laryngeal squamous cell carcinoma via impairing PLK1-AKT axis-mediated glycolysis. CELL DEATH DIS. 2020;11(10):919.
17. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. NUCLEIC ACIDS RES. 2017;45(W1):W98–102.
18. Naba A, Clauser KR, Ding H, Whittaker CA, Carr SA, Hynes RO. The extracellular matrix: Tools and insights for the "omics" era. MATRIX BIOL. 2016;49:10–24.
19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
20. Hui L, Zhang J, Ding X, Guo X, Jiang X. Matrix stiffness regulates the proliferation, stemness and chemoresistance of laryngeal squamous cancer cells. INT J ONCOL. 2017;50(4):1439–47.
21. Ma J, Hu X, Dai B, Wang Q, Wang H. Bioinformatics analysis of laryngeal squamous cell carcinoma: seeking key candidate genes and pathways. PEERJ. 2021;9:e11259.

22. Wang P, Li QY, Sun YN, Wang JT, Liu M. **Long Noncoding RNA NEAT1: A Potential Biomarker in the Progression of Laryngeal Squamous Cell Carcinoma**. *ORL J Otorhinolaryngol Relat Spec* 2021:1–7.
23. Yang X, Wu Q, Wu F, Zhong Y. Differential expression of COL4A3 and collagen in upward and downward progressing types of nasopharyngeal carcinoma. ONCOL LETT. 2021;21(3):223.
24. He Y, Liu R, Yang M, Bi W, Zhou L, Zhang S, Jin J, Liang X, Zhang P. Identification of VWF as a Novel Biomarker in Lung Adenocarcinoma by Comprehensive Analysis. FRONT ONCOL. 2021;11:639600.
25. Liot S, Balas J, Aubert A, Prigent L, Mercier-Gouy P, Verrier B, Bertolino P, Hennino A, Valcourt U, Lambert E. Stroma Involvement in Pancreatic Ductal Adenocarcinoma: An Overview Focusing on Extracellular Matrix Proteins. FRONT IMMUNOL. 2021;12:612271.
26. Holle AW, Young JL, Spatz JP. In vitro cancer cell-ECM interactions inform in vivo cancer treatment. Adv Drug Deliv Rev. 2016;97:270–9.
27. Rahbari NN, Kedrin D, Incio J, Liu H, Ho WW, Nia HT, Edrich CM, Jung K, Daubriac J, Chen I, et al. Anti-VEGF therapy induces ECM remodeling and mechanical barriers to therapy in colorectal cancer liver metastases. SCI TRANSL MED. 2016;8(360):135r–360r.
28. Hosein AN, Brekken RA, Maitra A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. Nat Rev Gastroenterol Hepatol. 2020;17(8):487–505.
29. Lyu S, Jiang C, Xu R, Huang Y, Yan S. INHBA upregulation correlates with poorer prognosis in patients with esophageal squamous cell carcinoma. CANCER MANAG RES. 2018;10:1585–96.
30. Pallasch FB, Schumacher U. **Angiotensin Inhibition, TGF-beta and EMT in Cancer**. *Cancers (Basel)* 2020, 12(10).
31. Ou M, Xu X, Chen Y, Li L, Zhang L, Liao Y, Sun W, Quach C, Feng J, Tang L. **MDM2 induces EMT via the BRaf signaling pathway through 1433**. *ONCOL REP* 2021, 46(1).
32. Wu ZH, Tang Y, Niu X, Cheng Q. Expression and gene regulation network of INHBA in Head and neck squamous cell carcinoma based on data mining. Sci Rep. 2019;9(1):14341.
33. Pandey R, Zhou M, Islam S, Chen B, Barker NK, Langlais P, Srivastava A, Luo M, Cooke LS, Weterings E, et al. Carcinoembryonic antigen cell adhesion molecule 6 (CEACAM6) in Pancreatic Ductal Adenocarcinoma (PDA): An integrative analysis of a novel therapeutic target. Sci Rep. 2019;9(1):18347.
34. Steiner N, Hajek R, Nachbaur D, Borjan B, Sevcikova S, Gobel G, Gunsilius E: **Levels of CEACAM6 in Peripheral Blood Are Elevated in Patients with Plasma Cell Disorders**: **A Potential New Diagnostic Marker and a New Therapeutic Target?** *DIS MARKERS* 2019, **2019**:1806034.
35. Bednarek K, Kostrzewska-Poczekaj M, Szaumkessel M, Kiwerska K, Paczkowska J, Byzia E, Ustaszewski A, Janiszewska J, Bartochowska A, Grenman R, et al. Downregulation of CEACAM6 gene expression in laryngeal squamous cell carcinoma is an effect of DNA hypermethylation and correlates with disease progression. AM J CANCER RES. 2018;8(7):1249–61.
36. Tian C, Zhang B, Ge C. Effect of CEACAM6 silencing on the biological behavior of human gallbladder cancer cells. ONCOL LETT. 2020;20(3):2677–88.

37. Iwabuchi E, Miki Y, Onodera Y, Shibahara Y, Takagi K, Suzuki T, Ishida T, Sasano H. Co-expression of carcinoembryonic antigen-related cell adhesion molecule 6 and 8 inhibits proliferation and invasiveness of breast carcinoma cells. Clin Exp Metastasis. 2019;36(5):423–32.
38. Kiener TK, Selptsova-Friedrich I, Hunziker W. Tjp3/zo-3 is critical for epidermal barrier function in zebrafish embryos. DEV BIOL. 2008;316(1):36–49.
39. Kiener TK, Sleptsova-Friedrich I, Hunziker W: **Identification, tissue distribution and developmental expression of tjp1/zo-1, tjp2/zo-2 and tjp3/zo-3 in the zebrafish, Danio rerio**. *GENE EXPR PATTERNS* 2007, **7**(7):767–776.
40. Luo M, Zhang L, Yang H, Luo K, Qing C. Long noncoding RNA NEAT1 promotes ovarian cancer cell invasion and migration by interacting with miR1321 and regulating tight junction protein 3 expression. MOL MED REP. 2020;22(4):3429–39.

Figures

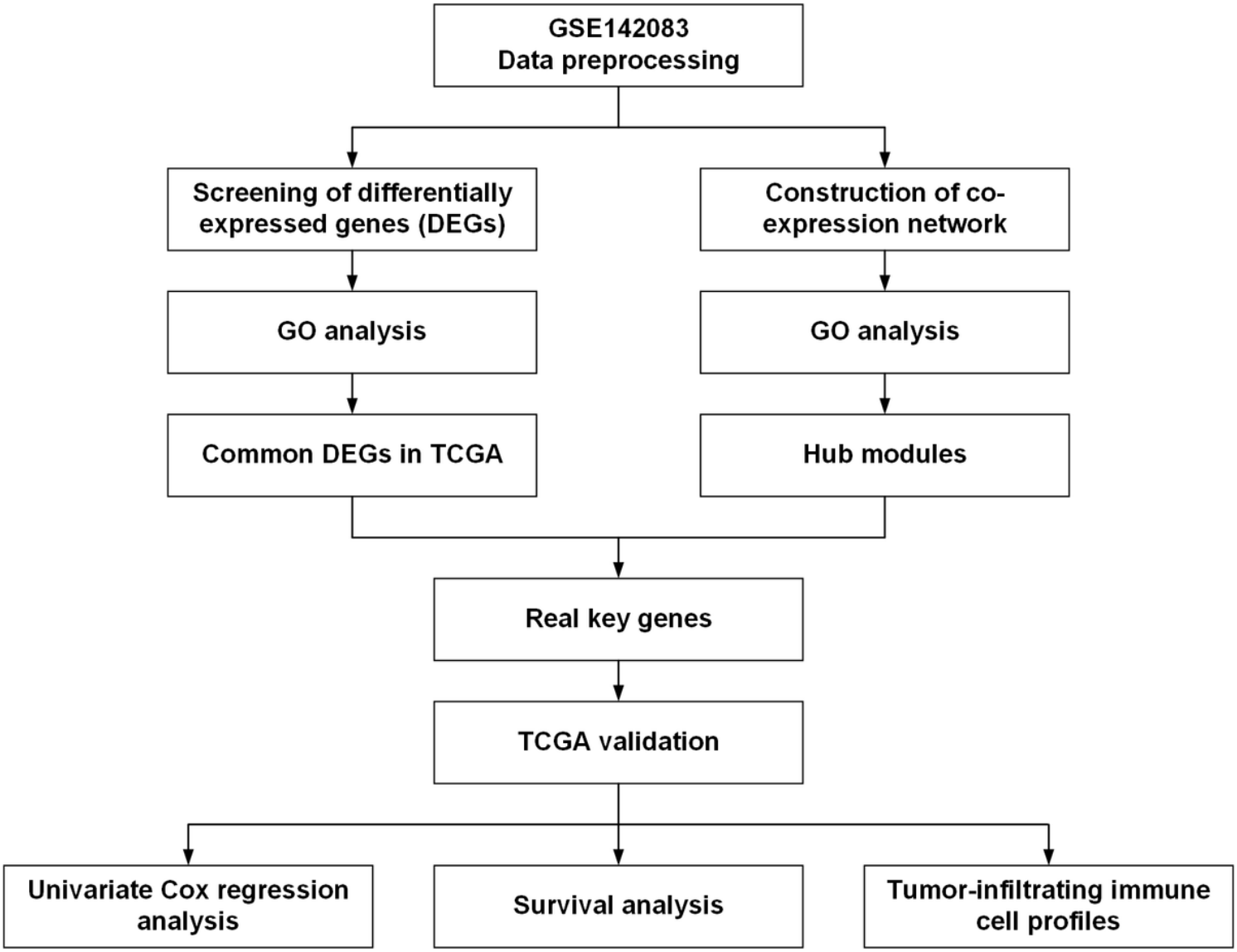
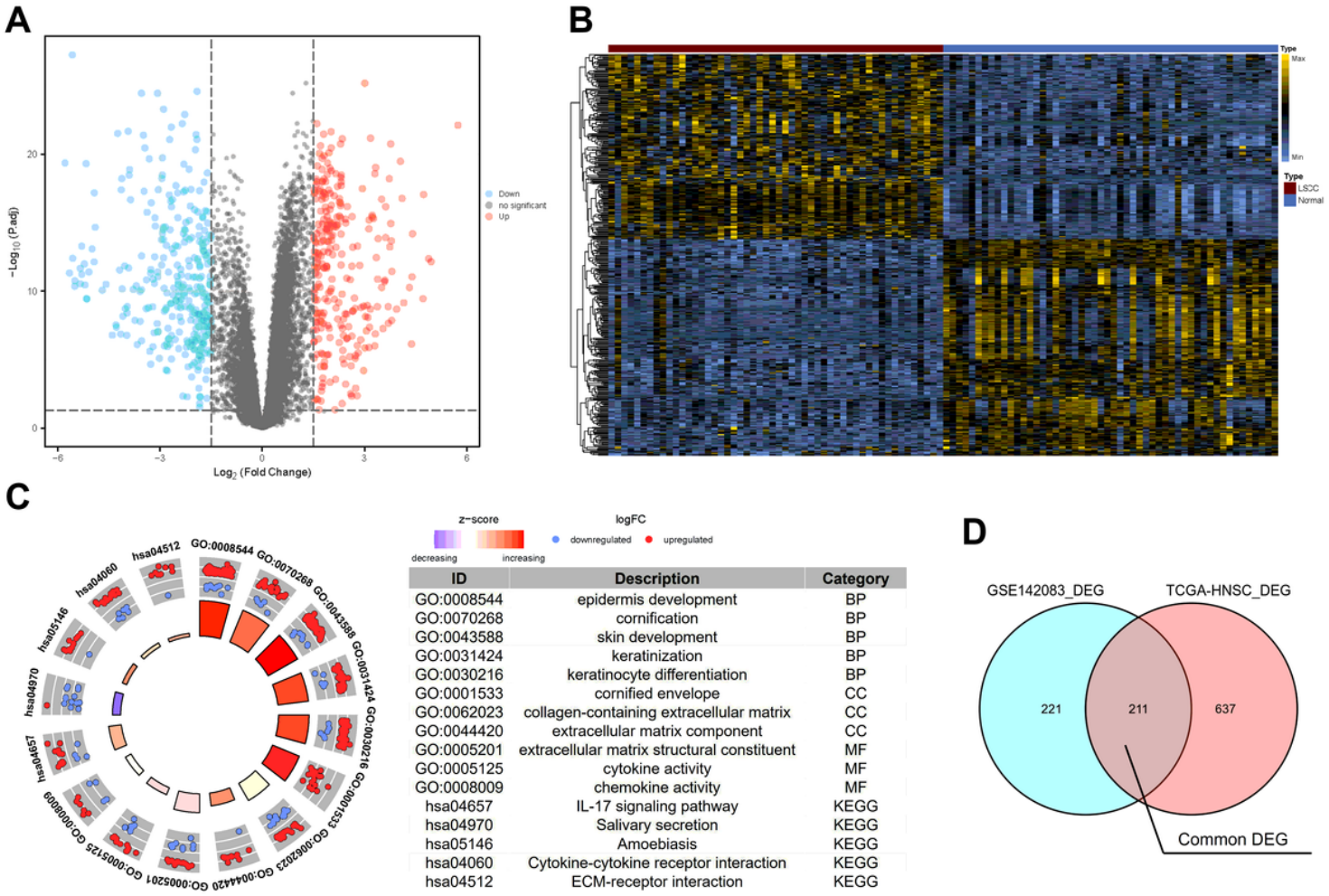


Figure 1

The workflow for the current study.



ID	Description	Category
GO:0008544	epidermis development	BP
GO:0070268	cornification	BP
GO:0043588	skin development	BP
GO:0031424	keratinization	BP
GO:0030216	keratinocyte differentiation	BP
GO:0001533	cornified envelope	CC
GO:0062023	collagen-containing extracellular matrix	CC
GO:0044420	extracellular matrix component	CC
GO:0005201	extracellular matrix structural constituent	MF
GO:0005125	cytokine activity	MF
GO:0008009	chemokine activity	MF
hsa04657	IL-17 signaling pathway	KEGG
hsa04970	Salivary secretion	KEGG
hsa05146	Amoebiasis	KEGG
hsa04060	Cytokine-cytokine receptor interaction	KEGG
hsa04512	ECM-receptor interaction	KEGG

Figure 2

Differentially expressed genes screening. (A) Volcano map of DEGs between LSCC and normal samples in GSE142083. The red plots in the volcano represent up-regulation and the blue points represent down-regulation genes. (B) Heatmap of the all DEGs. The color in heatmaps from blue to yellow shows the progression from low expression to high expression. (C) GO and KEGG analysis of DEGs. The outer circle presents the scatter plot of assigned gene log2FC for all terms, the red ones stand for increased expression, whereas the blue ones represent decreased expression. The inner circle indicates the Z-score value and the number of genes. Red means the higher z-score value and purple means lower Z-score value. BP, biological processes; CC, cell components; MF, molecular functions. (D) A Venn diagram was utilized to screen the common DEGs between the GSE142083_DEGs and the TCGA-HNSC_DEGs.

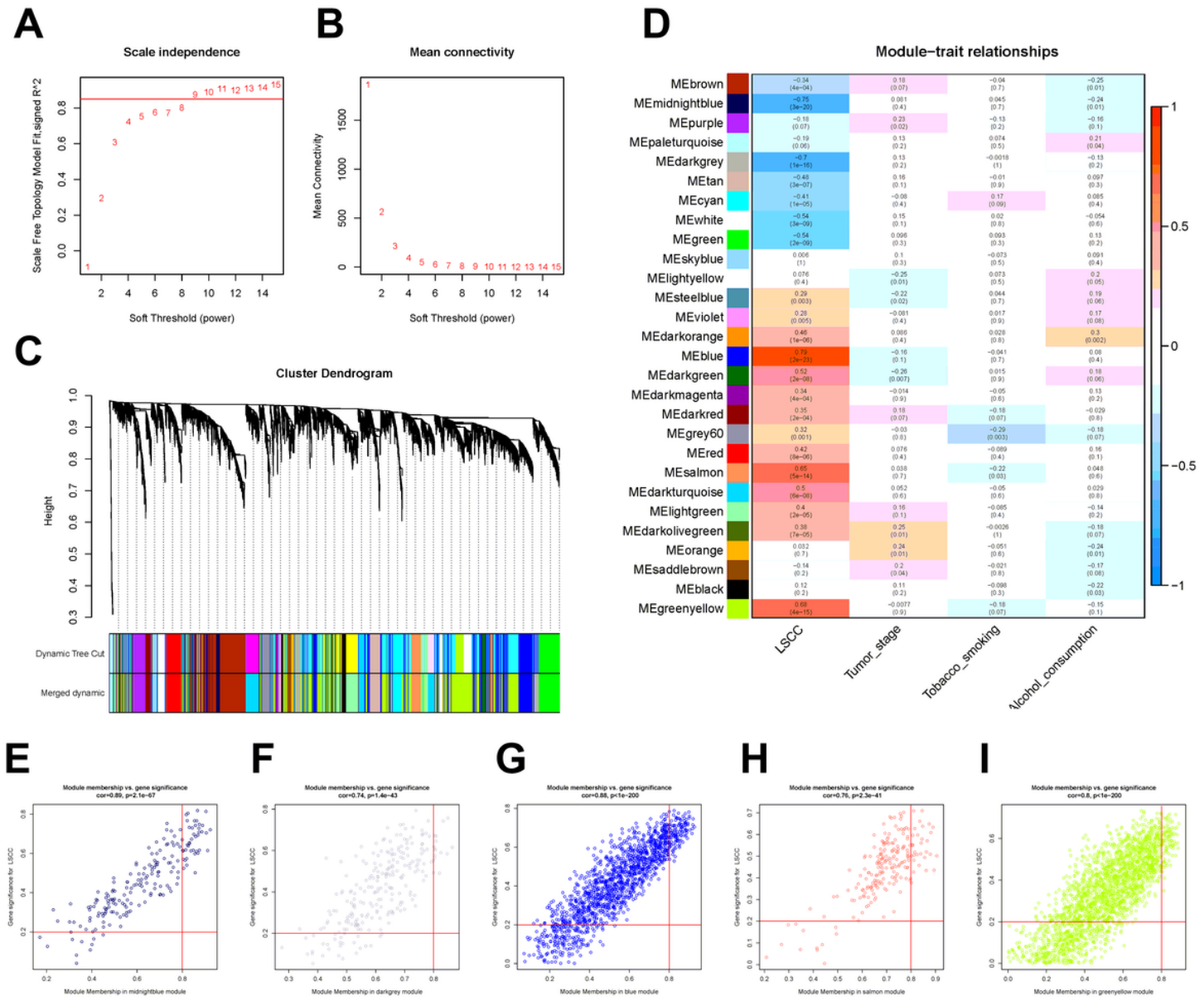


Figure 3

WGCNA of the LSCC samples. (A, B) Analysis of the network topology for various soft thresholding powers. (A) shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). Horizontal red line shows x-axis = 0.85. (B) displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis). The power was set as 9 for further analysis. (C) Hierarchical cluster analysis was conducted to detect co-expression clusters with corresponding color assignments. Each color represents a module in the constructed gene co-expression network by WGCNA. (D) Module–trait relationships. Each row represents a color module and every column represents a clinical trait. Each cell contains the corresponding correlation and p-value. (E-I) A scatter plot of gene significance (GS) for LSCC versus the module membership (MM) in the (E) midnightblue module, (F) darkgrey module, (G) blue module, (H) salmon module and (I) greenyellow module. Vertical red line shows the |MM| = 0.8, horizontal red line shows the |GS| = 0.2.

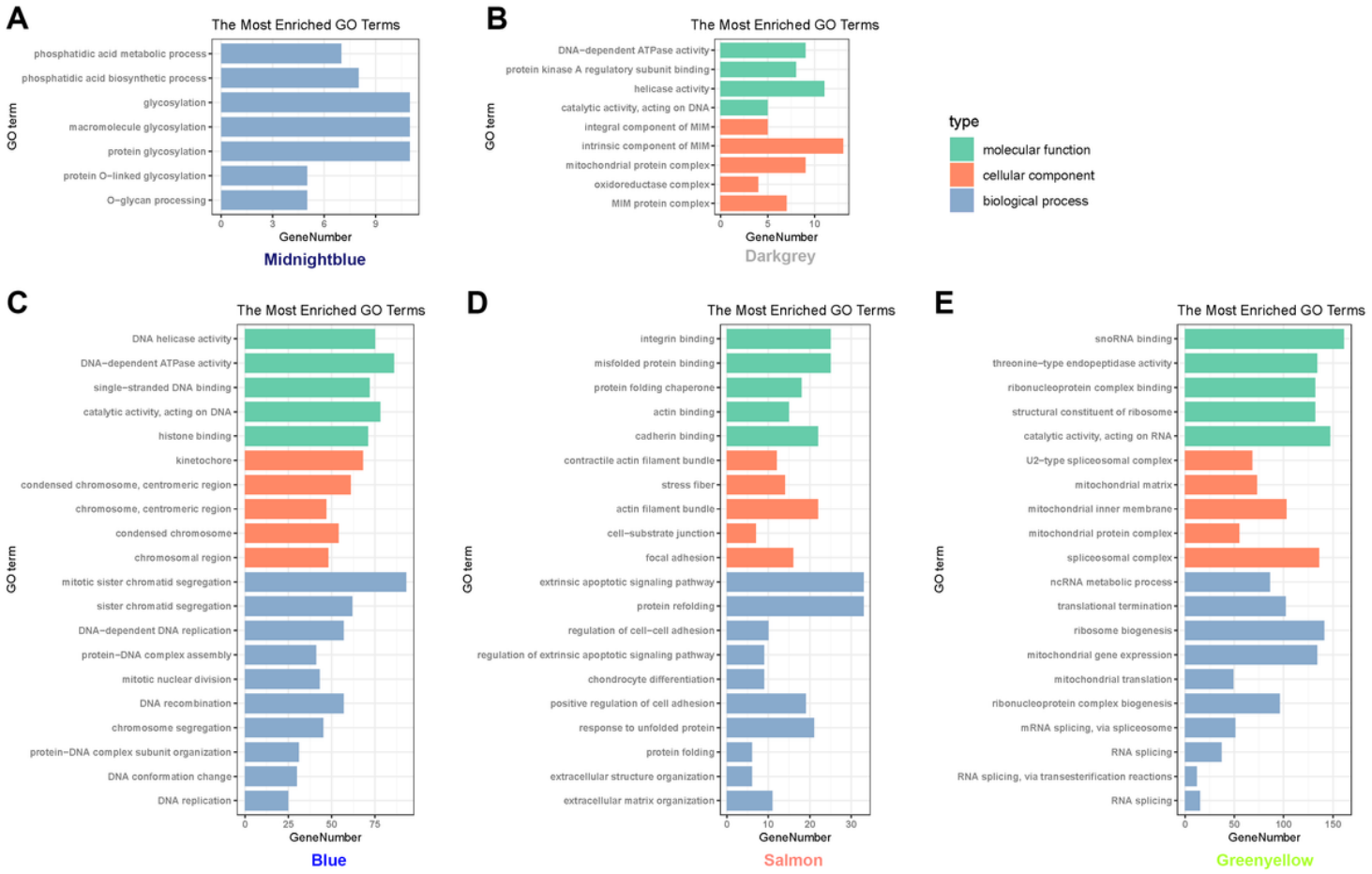


Figure 4

GO analysis of midnightblue module, darkgrey module, blue module, salmon module and greenyellow module.

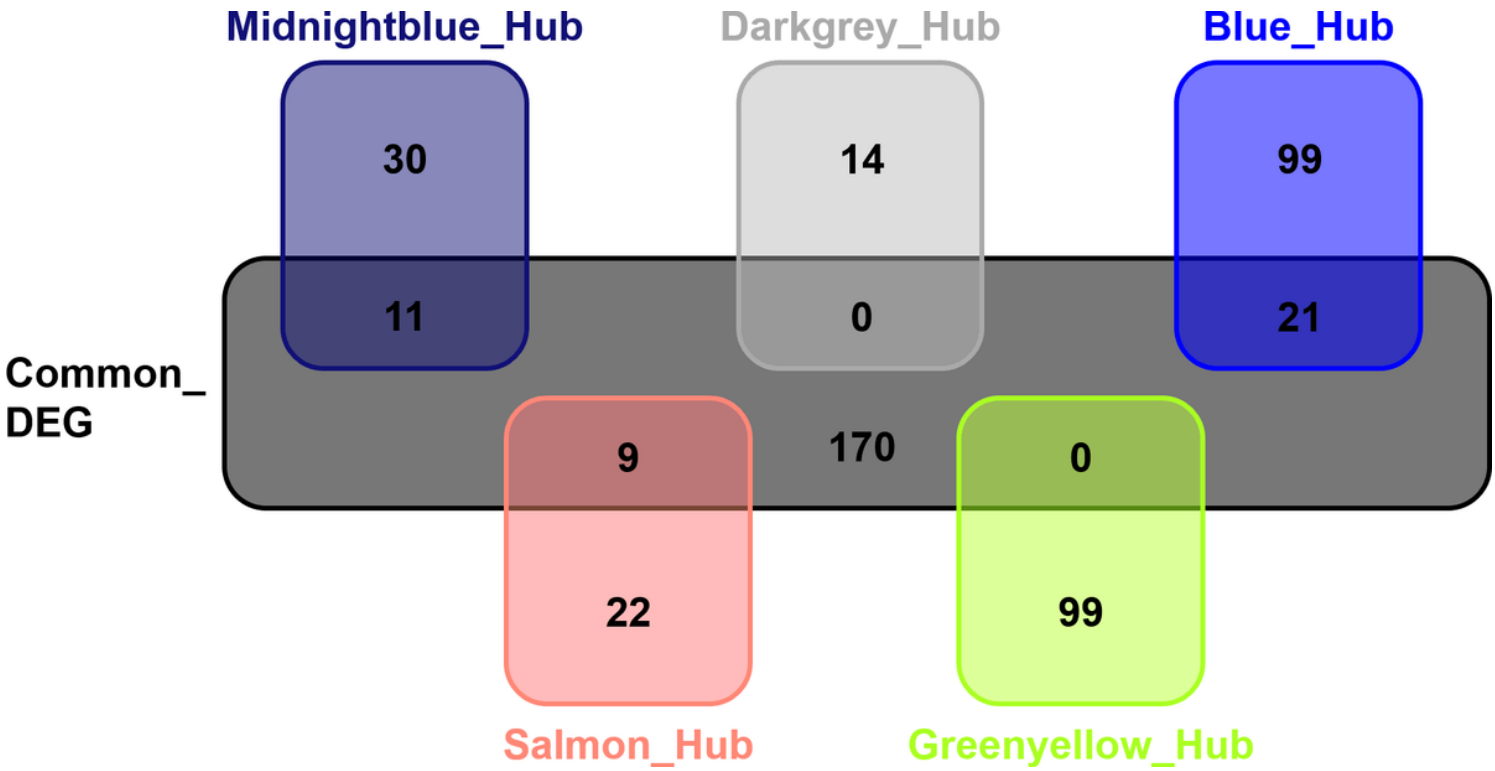


Figure 5

Real key genes belonging to both the five hub modules and the common DEGs.

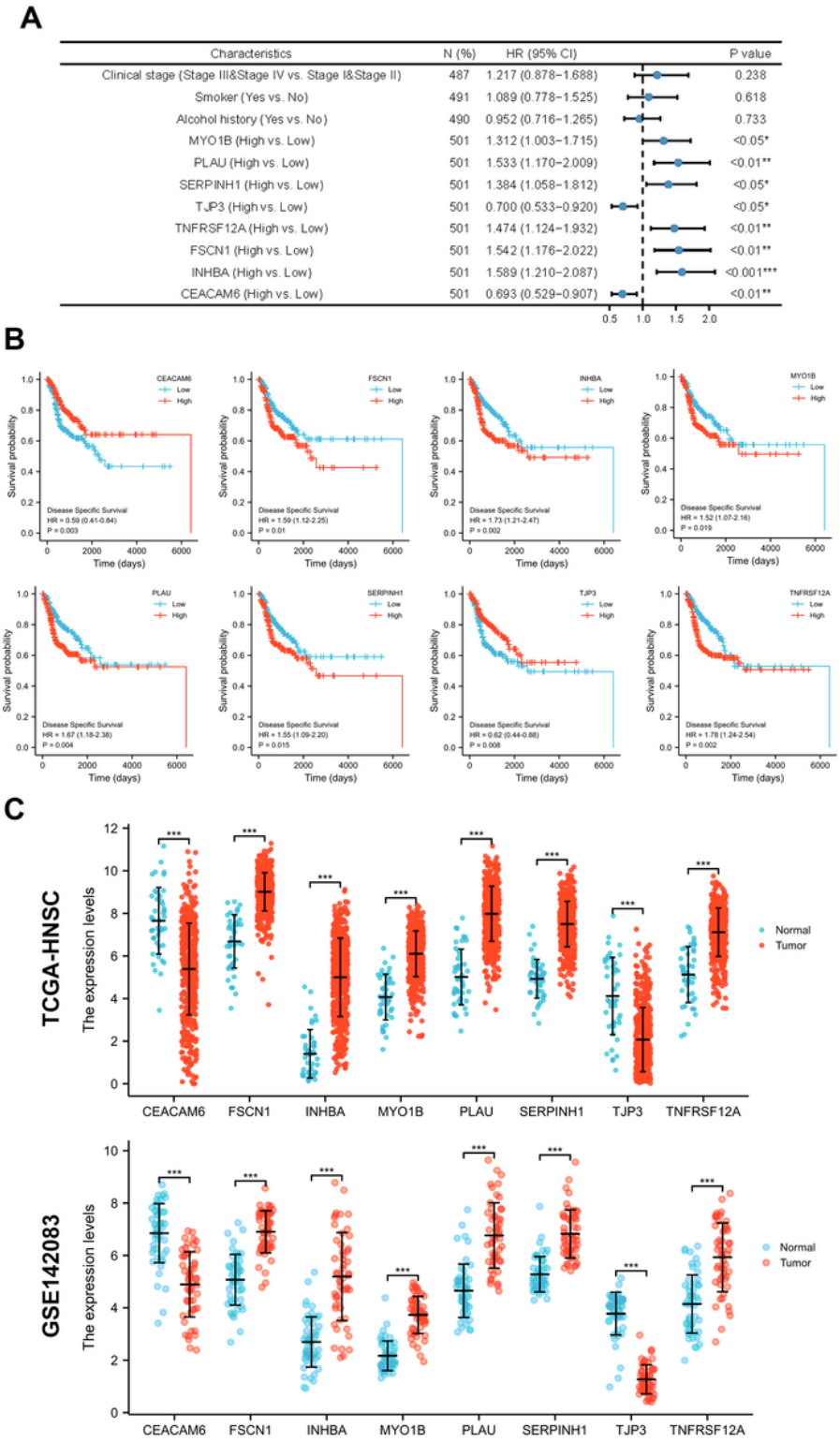


Figure 6

Validation of key genes. (A) Univariate Cox regression analysis to predict prognostic factors associated with patient overall survival. (B) Disease specific survival analysis of CEACAM6, FSCN1, INHBA, MYO1B, PLAU, SERPINH1, TJP3 and TNFRSF12A. (C) Expression scatter diagram of CEACAM6, FSCN1, INHBA, MYO1B, PLAU, SERPINH1, TJP3 and TNFRSF12A in TCGA-HNSC and GSE142083. Wilcoxon rank sum test is applied in the significance test. ***, $p < 0.001$.

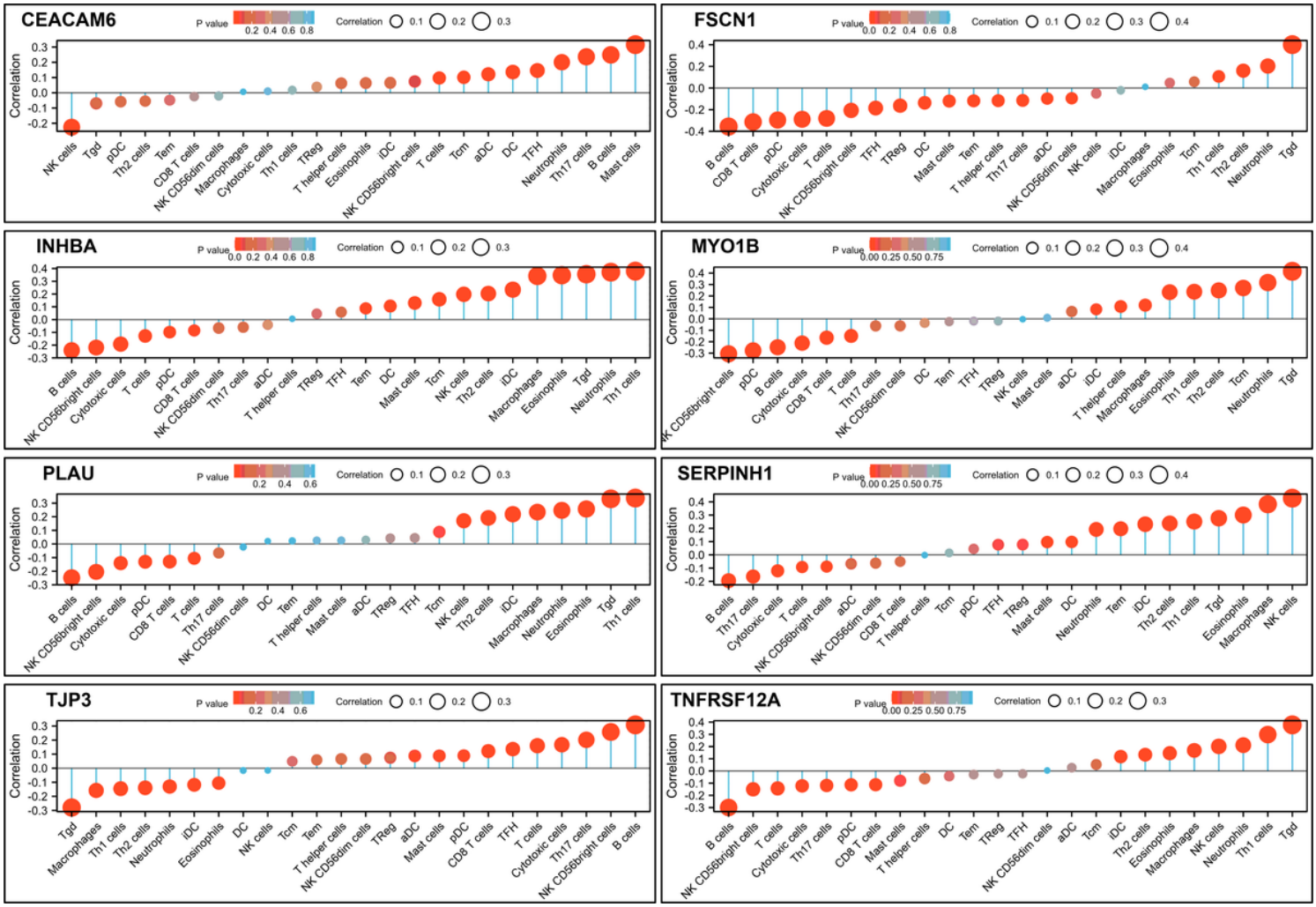


Figure 7

24 kinds of tumor-infiltrating immune cell correlation analysis based on TCGA-HNSC database. Spearman coefficient is applied in the significance test. NK, natural killer; Tgd, T gamma delta; pDC, plasmacytoid dendritic cell; Th2, T helper cells 2; Tem, T central memory; Treg, regulatory T; iDC, immature DC; Tcm, T central memory; aDC, activated DC; TFH, T follicular helper.

Supplementary Files

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

- FigureS1.tif
- FigureS2.jpg
- FigureS3.tif
- TableS1.docx