Am J Transl Res 2023;15(1):373-391
www.ajtr.org /ISSN:1943-8141/AJTR0146307

Original Article

Identification of metabolic signatures related to metastasis and immunotherapy resistance in oral squamous cell carcinoma

Haoran Chen[1,2*], Xin Liu[3*], Feng Yao[1*], Miao Yin[1], Bo Cheng[1], Sisi Yang[1]

[1]Department of Stomatology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, China; [2]Department of Immunology and National Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (CAMS) and Peking Union Medical College, Beijing, China; [3]Department of Bacteriology and Immunology, Beijing Chest Hospital, Capital Medical University/Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China. [*]Equal contributors.

Received September 3, 2022; Accepted December 19, 2022; Epub January 15, 2023; Published January 30, 2023

Abstract: Objectives: In this study, we aimed to identify the metabolic genes associated with the metastasis and immunotherapy resistance of oral squamous cell carcinoma (OSCC) and to construct a metabolic gene-related predictive model for the prognosis of OSCC. Methods: RNA-seq data were download from The Cancer Genome Atlas (TCGA). Weighted gene co-expression network analysis (WGCNA) was applied to identify the modules related to EMT, stemness, and checkpoint signatures in OSCC. Univariate Cox and the least absolute shrinkage and selection operator (LASSO) methods were used to construct the metabolic gene signature. Furthermore, the scRNA-seq data were obtained from Gene Expression Omnibus (GEO) database and analyzed using "Seurat" and "CopyKAT" packages. Results: The risk prediction model was constructed using the 12 metabolic-related gene signature. Based on this model, risk score of each sample was calculated and used to divide the samples into low- and high-risk groups. Our model was effective as the risk score was significantly associated with clinical features and genetic mutations. Meanwhile, we found that lipid metabolism, glycolysis, amino acid metabolism, and drug metabolism differed between high- and low-risk groups. Pathways associated with malignant tumor and immunosuppression were enriched in high-risk group. Furthermore, low-risk group showed a more activated immune status and was predicted to have better response to immunotherapy. Finally, through single-cell transcriptome analysis, we assessed the expression of these 12 genes in tumor and non-tumor cells and verified the existence of two clusters of tumor cells with different degrees of malignancy at the cellular level. Conclusions: Our study demonstrates the clinical significance of metabolic related gene signature for the treatment of OSCC and suggests potential therapeutic targets and pathways for OSCC.

Keywords: Oral squamous cell carcinoma, WGCNA, metabolism, metastasis, immunotherapy resistance

Introduction

Oral squamous cell carcinoma (OSCC) is the most fatal oral disease and the most common head and neck malignancy. Its incidence is increasing with approximately 400,000 new cases and 180,000 deaths annually [1]. OSCC originates from the mucosal epithelium of the oral cavity and pharynx, and the risk factors for OSCC include smoking, excessive alcohol consumption, betel nut chewing, and human papillomavirus (HPV) exposure [2, 3]. Although early OSCC can achieve a good prognosis through surgical resection, most OSCC patients do not exhibit symptoms at the early stage but in advanced stages at the time of diagnosis. Furthermore, OSCC also have high recurrence rates [4]; hence, the survival rate of OSCC is still not significantly improved even with the improved treatment strategies [5, 6].

Lymph node metastasis occurs in more than half of the patients with OSCC and is closely related to the prognosis of OSCC [7, 8]. An important factor attributing to lymph node and distant metastasis is epithelial-mesenchymal

transition (EMT), a genetic and epigenetically regulated biological process in which an epithelial cell transforms into a mesenchymal state [9-11]. Tumor cells with EMT are characterized by changes in cytoskeleton, decreased adhesion to their surroundings, enhanced migration, and apoptotic resistance [12, 13]. Many studies have demonstrated that EMT activation is closely interconnected with the generation of cancer stem cells (CSCs) in various types of carcinomas [14]. EMT activation also enables tumor cells to escape immune killing by increasing the expression of immune checkpoint genes and acquire resistance to chemotherapy and immunotherapy [15-18], which are also characteristics of CSCs. Anti-PD1 monoclonal antibodies, such as pembrolizumab and nivolumab, have been used as first-line treatment for malignant OSCC [19, 20], however, many patients do not respond to immunotherapy. Studies have shown that tumor cells resistant to chemotherapy, targeted therapy, or immunotherapy, commonly exhibit activated EMT [21].

EMT is a dynamic and reversible process that involves extensive crosstalk with metabolic reprogramming, which is regulated by EMT transcription factors (EMT-TFs) and epigenetic modifying enzymes that use metabolites as cofactors and substrates. Moreover, changes in the migration and invasion of tumor cells require metabolic reprogramming to provide energy [22]. Transforming Growth Factor β (TGF-β) and hypoxia can activate EMT in tumor cells as well as induce glycolysis and fatty acid oxidation (FAO) pathways [23, 24]. Glutamine metabolism and mitochondrial metabolism are also changed during EMT, and the expression levels of SLC2A1/3, HK2, CD36, and GLS1/2 involved in the above metabolic processes are altered [25-27]. These extensive changes in metabolic processes are essential for EMT and constitute potential metabolic targets for inhibiting EMT.

Although EMT activation has been known in the tumor cells of patients with OSCC, contributing to the poor prognosis of OSCC [28], the interaction between metabolic pathways and EMT in OSCC remains unclear. Investigating this link is essential for improving immunotherapy outcomes in patients with OSCC and providing metabolically targeted therapies. Therefore, in this study, we used TCGA database to identify the most relevant metabolic gene modules for EMT, stemness, and checkpoint signatures in OSCC. We constructed a metabolism-related prognostic signature containing 12 genes using least absolute shrinkage and selection operator (LASSO) regression. Furthermore, we showed the association between the risk score derived from the signature and the clinical features and immune landscape. Finally, we used single-cell RNA sequencing (scRNA-seq) data from OSCC patients to explore the expression of these 12 genes in tumor and non-tumor cells. We divided tumor cells into two clusters based on copy number variation (CNV) to explain the heterogeneity of tumor cells in terms of risk metabolic gene.

Materials and methods

Study design and data collection

The workflow diagram of this study was presented in Figure S1. First, RNA-seq data, clinical information, somatic mutations, and CNVs data were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). In total, 330 patients with OSCC were screened from the TCGA-HNSC cohort. Next, ScRNA-seq data were downloaded from the GEO database (GSE103322, http://www.ncbi.nlm.nih.gov/geo/) which was provided by Bradley et al. [29], and contained data of 5902 cells from 18 OSCC patients. R4.1.2 software was used for data preprocessing.

Co-expression network construction and identification of signature-related modules

We obtained 2119 genes related to "Metabolism" pathways (R-HSA-1430728) from the Reactome database using the "ReactomePA" package [30]. Among these genes, the EMT signature- [31], stemness signature- [32], and checkpoint signature- [33] related genes were downloaded and subjected to single-sample gene set enrichment analysis (ssGSEA) to calculate the scores of EMT, stemness, and checkpoint. The "WGCNA" package was applied to construct a co-expression network of metabolic genes. The matrix of adjacencies was built based on the Pearson's correlation value between paired genes. An unsigned scale-free co-expression network was constructed based on a soft thresholding power of $\beta = 4$. Then, "TOMsimilarity" and "cutreeDynamic" were

used to construct the metabolic gene co-expression module, and the merge height cut was set at 0.4 to merge modules. The correlation between module eigengenes (MEs) and signatures was tested using Pearson's correlation. The adjacency threshold was set at 0.05 to select hub genes from the MEs of interest. The "iRegulon" plug in Cytoscape was used to identify the transcription factors of hub genes (FDR < 0.001).

Construction of the prognosis signature

We used genes with an adjacency threshold greater than 0.05 in significant MEs for the univariate Cox regression. LASSO regression was used to screen out the genes most related to survival and to calculate their coefficients. Finally, the risk score was calculated for patients in TCGA data using the formula below, and validation was performed in the GEO dataset (GSE41613).

$$\text{Risk score} = \sum_{i=1}^{n} expr_{gene(i)} * coef_{gene(i)}$$

Functional and pathway enrichment analyses

Functional enrichment analysis of the hub genes was performed using Metascape (http://metascape.org) [34]. The software GSEA (http://www.gsea-msigdb.org/) and the C2, C6, and C7 gene sets were downloaded to analyze different pathways in the high- and low-risk groups. Differential expression gene (DEG) analysis was performed by "limma3.50.3" package using the expression matrix of patients in the top 25% and the bottom 25% of the risk score, and genes with |logFC| > 1, $P < 0.05$ were screened for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. Only the gene sets with $P < 0.05$ were considered significant.

Metabolic pathways, immune signatures, and cell types enrichment analysis

Metabolic gene sets, involving 90 human metabolic pathways, were downloaded from the KEGG pathway database (https://www.genome.jp/kegg/pathway.html) [35]. Immune signatures were identified based on previously published studies [33, 36, 37]. Cell type characterization and composition estimation were performed using CIBERSORT (http://cibersort.stanford.edu) [38]. Package "ggplot2" was used to display the immune cell composition and immune signature of different groups. Function "mantel_test" was used to identify the relationship between the 19 signatures and the risk score, and the results were shown by "ggcor" package. The relationship among different types of immune cells was analyzed by Spearman correlation and graphed by "corrplot" package. In addition, we calculated the hazard ratio (HR) of 22 types of immune cells by using "coxph" function in "survival" package.

Evaluation of immune-checkpoint blockade (ICB) response and drug sensitivity prediction

The ICB response signature was adopted from published research, and ssGSEA was used to estimate the ICB response score. Since the KMplot tool (http://kmplot.com/analysis) contains immunotherapy data for multiple tumors, we selected anti-PD1 treatment data and pre-treatment samples to obtain the survival curves of high- and low-risk patients. Furthermore, we used "pRRophetic" package to calculate the half-maximal inhibitory concentration (IC50) of 5-Fluorouracil, Cetuximab and Cisplatin according to "cgp2016" dataset.

Single cell RNA-seq analysis

ScRNA-seq data of 5902 cells were first subjected to quality control using the "Seurat" package, and the data with nFeature > 2000, nCount < 20000, and percent.mt < 15 were filtered for further analysis, yielding a total of 5746 cells. The data were normalized using log2 (TPM/10+1), and t-distributed stochastic neighbor embedding (tSNE) was used for dimensionality reduction and cluster classification. Furthermore, CopyKAT (Copynumber Karyotyping of Tumors), a package designed to identify genomic copy number for scRNA-seq data, was used to distinguish tumor cells by recognizing aneuploid cells, and we set win.size at 25 & KS.cut at 0.2. Then, tumor cells were divided into two clones by calculating Euclidean distance with CopyKAT [39].

Results

Identification of significant genes related to EMT, stemness, and checkpoints

We obtained RNA-seq data of 546 HNSC samples from the TCGA database and screened 330 oral cancer samples from the lip, gingiva,

tongue, oral cavity, and pharynx. The detailed patient information was provided in Table S1. Since EMT is a key factor contributing to the metastasis of cancers [40], and regulates not only tumor cell migration but also the generation of cancer stem cells as well as the expression of checkpoint genes [41], in this study, we calculated the EMT, stemness, and checkpoint signature scores of 330 HNSC samples using ssGSEA (Table S2). A total of 2119 metabolic-related genes were obtained from the Reactome database. We chose $\beta = 4$ (scale-free $R^2 = 0.92$) as the soft threshold power to build a scale-free network (Figure S2A and S2B). According to the expression of these genes, we divided them into seven modules by average linkage clustering (Figure S2C-E). We then displayed a heatmap of the correlation between modules and signatures, showing a significant positive relationship among the green module, EMT, and checkpoint signatures (**Figure 1A**). As invasive metastasis is an inefficient process in which only a small portion of EMT tumor cells can survive immune surveillance [42], to obtain genes with greater clinical value for immune therapy, we also selected from the brown module genes that negatively correlated with the checkpoint for subsequent analysis.

According to topological overlap above 0.05 in MEs, there were 165 hub genes in the green module and 94 hub genes in the brown module. We used "iRegulon" to predict the transcription factors (TFs) of genes in green and brown modules (FDR < 0.001). The top 5 TFs for green modules were *SPL1*, *SOX21*, *CTCF*, *CEBPA*, *ELF2*, while the top 5 TFs for brown modules were *FOXO1*, *TCF12*, *EP300*, *MSRB3*, *NR4A1* (Figure S3). Upon importing these genes into Metascape for analysis, hub genes in green module were significantly related to glycosaminoglycan metabolism and chondroitin sulfate/dermatan sulfate metabolism pathways (**Figure 1B**) which are mainly involved the composition of the extracellular matrix and likely to facilitate the activation of EMT process. Moreover, the enrichment of vitamin and lipid metabolic pathways in the green module support the rapid energy supply for the malignant tumor cells. The hub genes in brown module were enriched in the pathways associated with oxidative stress (**Figure 1B**), such as *NRF2* which is activated under oxidative stress conditions and can reduce inflammatory damage by stimulating the expression of antioxidants and anti-inflammatory cytokines. Furthermore, 259 genes expressed in TCGA OSCC cohort were analyzed for its association with overall survival (OS) by univariate Cox regression, and the results showed that 21 genes were closely related to the prognosis of patients ($P < 0.05$) (**Figure 1C**).

Construction of a metabolism-related prognostic signature in OSCC

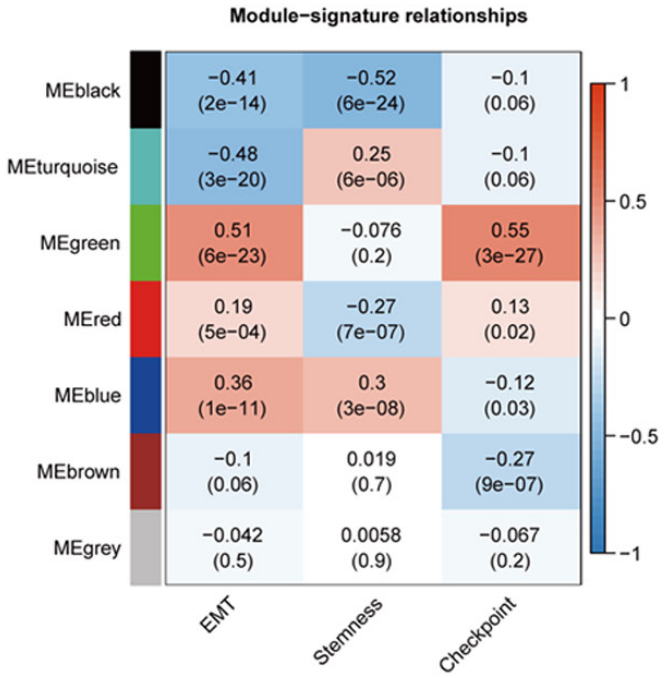
Using the LASSO regression algorithm, an immune-related metabolic signature including 12 genes for prognosis prediction was constructed (**Figure 2A**, **2B**), and the coefficients of the 12 genes were shown in Table S3. The 330 OSCC patients were divided into high- (n = 165) and low-risk groups (n = 165) using the median risk score as a threshold, and the mortality rate of patients increased with the risk score (**Figure 2C**) as the OSCC patients in low-risk group showed better survival rates than those in high-risk group (**Figure 2E**). Gene expression patterns were shown in **Figure 2D**. Furthermore, the area under the curve (AUC) of the prognostic signature for 1-, 3-, and 5-year survival was 0.66, 0.74, and 0.71, respectively, which was further validated in GSE41613 cohort (**Figure 2F**, **2G**), indicating the reliability of this model.

Association of the risk score with clinical features and somatic mutations in OSCC

Figure 3A showed the distribution of the relevant patient characteristics and the risk scores. To investigate the relationship between the risk score and the clinical features, we grouped patients according to TNM staging system and found that the risk score of the metabolic signature increased with increased T and N stages, suggesting that the signature was reliable (**Figure 3B**). However, because most patients were classified as M0, the risk score did not show a significant difference in the M classification.

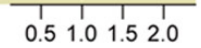
Moreover, we analyzed the somatic mutations and CNV data downloaded from TCGA database. Patients in high-risk group had higher tumor mutation burden (TMB) (**Figure 3C**) and exhibited a higher frequency of *TP53* mutations than the patients in the low-risk group (**Figure**

A



C

gene	pvalue	HR (95% CI)
PTGR1	0.0057165772	1.19 (1.05 to 1.35)
FUT7	0.0134199687	0.56 (0.35 to 0.89)
ALDH1B1	0.0094990885	1.25 (1.06 to 1.48)
DHCR7	0.0359309535	1.16 (1.01 to 1.33)
ADCY4	0.0477718265	0.59 (0.35 to 0.99)
NAMPT	0.0458623277	1.17 (1.00 to 1.37)
FMOD	0.0394247176	0.86 (0.75 to 0.99)
MCCC1	0.0451890767	1.23 (1.00 to 1.51)
SLC2A3	0.0013088961	1.27 (1.10 to 1.47)
NOSTRIN	0.0251622373	0.33 (0.13 to 0.87)
UGT1A10	0.0276166130	1.20 (1.02 to 1.42)
INPP5D	0.0235753228	0.77 (0.62 to 0.97)
ABCC3	0.0205869758	1.24 (1.03 to 1.49)
HK1	0.0004678158	1.75 (1.28 to 2.39)
G6PD	0.0218523046	1.20 (1.03 to 1.39)
PRKCA	0.0288990931	1.36 (1.03 to 1.80)
NQO1	0.0081406125	1.18 (1.04 to 1.33)
CSGALNACT1	0.0069372497	0.69 (0.53 to 0.90)
TKT	0.0421371107	1.25 (1.01 to 1.54)
NADSYN1	0.0435790117	1.25 (1.01 to 1.56)
TXNRD1	0.0078949171	1.19 (1.05 to 1.36)



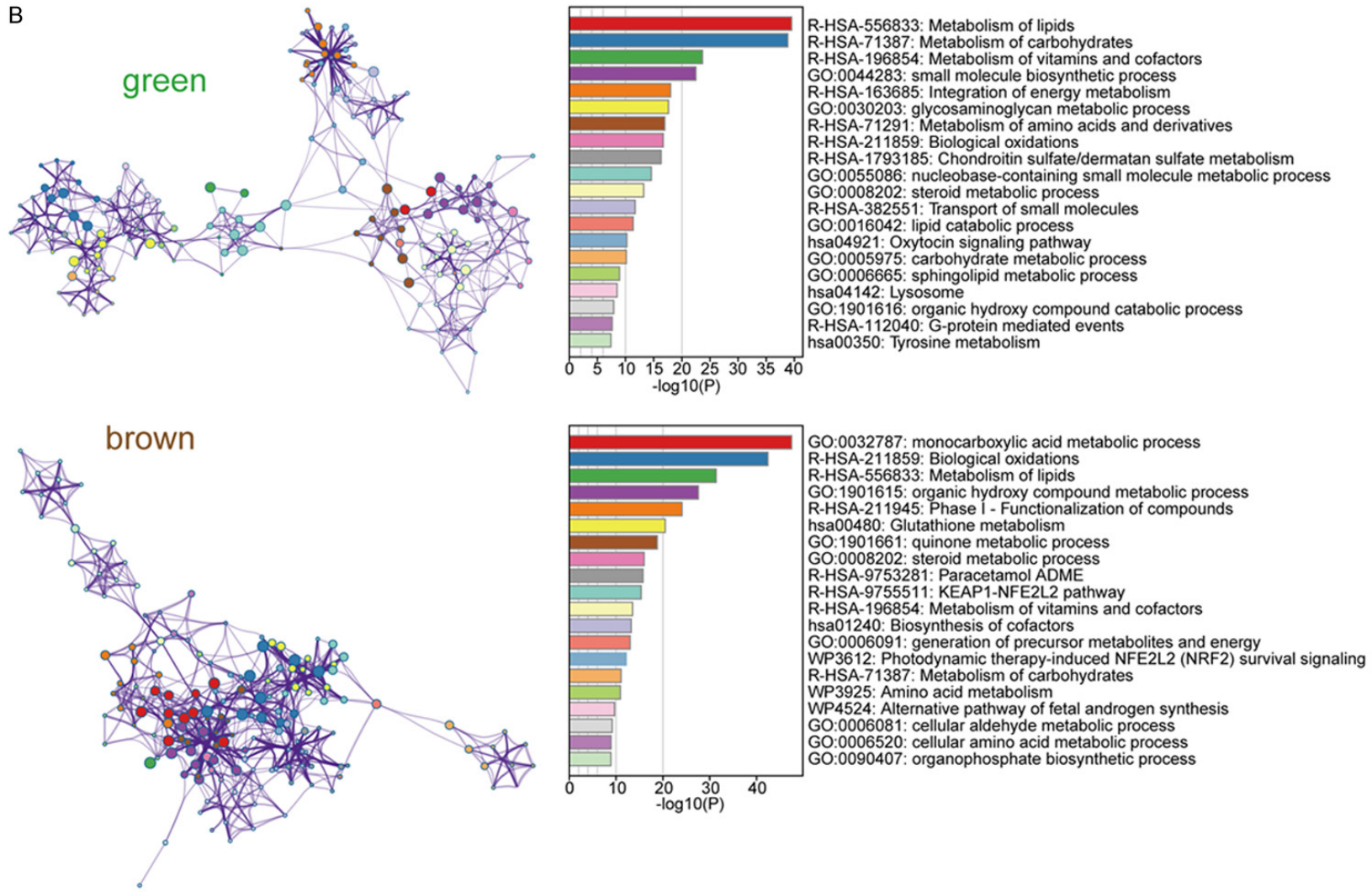


Figure 1. Module-trait associations and identification of significant genes in green and brown modules. A. The correlation between modules and signatures. Each cell contains the corresponding correlation and *P* value. B. Function enrichment of genes in green and brown module by Metascape. C. The results of 259 genes in green and brown modules and 21 genes were significantly associated with the survival of patients.

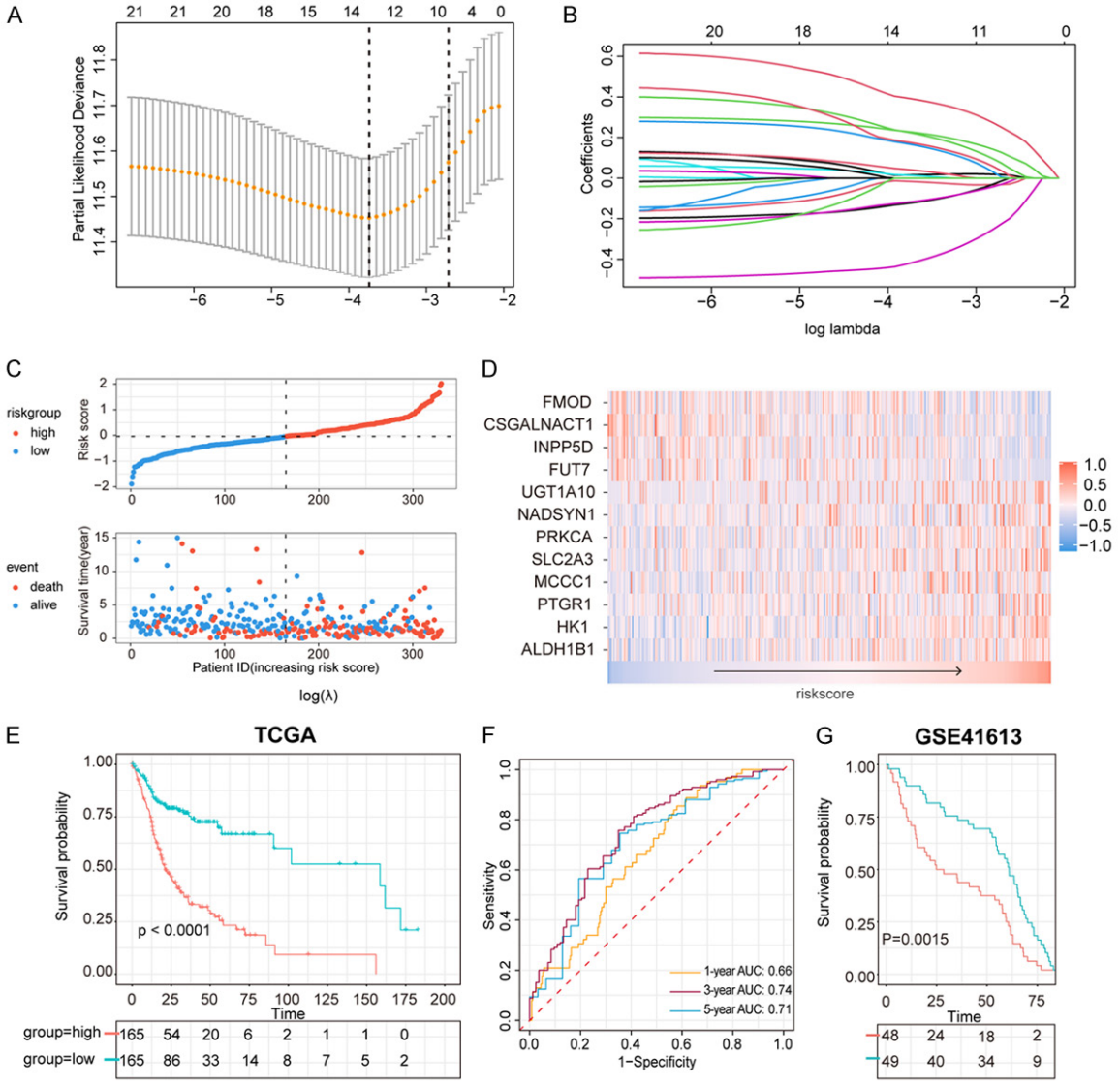


Figure 2. Construction of metabolic prognostic signature. A. Selection of the tuning parameter (lambda) in the LASSO model. B. LASSO coefficient profiles of 12 genes in metabolic signature. C. Distribution of risk score in patients, and scatterplot of OS events with risk score. D. Heatmap of 12 genes in metabolic signature with the increasing risk score. E. OS between high-risk group and low risk group in TCGA cohort. F. Survival-dependent ROC curve of metabolic signature at 1, 3, 5 years in TCGA cohort. G. Validation of metabolic signature in GEO41613.

3D). Similarly, the high- and low-risk groups harbored different CNV alteration sites, and more CNVs in both amplification (AMP) and deletion (DEL) were observed in the high-risk group (**Figure 3D**). For example, AP_9:3q28, a large amplification of *TP63*, was observed in the high-risk group, but not in the low-risk group.

Metabolic signature was an independent factor of a survival predictive nomogram

We used univariate and multivariate Cox regression analyses to explore the relationship between the clinical characteristics and the metabolism-related risk scores. We found that patient's age (hazard ratio [HR] = 1.02) and metabolism-related risk score (HR = 2.70) were two independent prognostic factors (P < 0.05) (**Figure 4A**, **4B**). Furthermore, the age and the risk score were used to construct a nomogram to estimate the patient's 1-, 3-, and 5-year survival (**Figure 4C**). The calibration curve for predicting patient survival indicated that the predicted OS was close to the actual OS (**Figure 4D**). Moreover, the ROC curve demonstrated that the three-index combined model was more

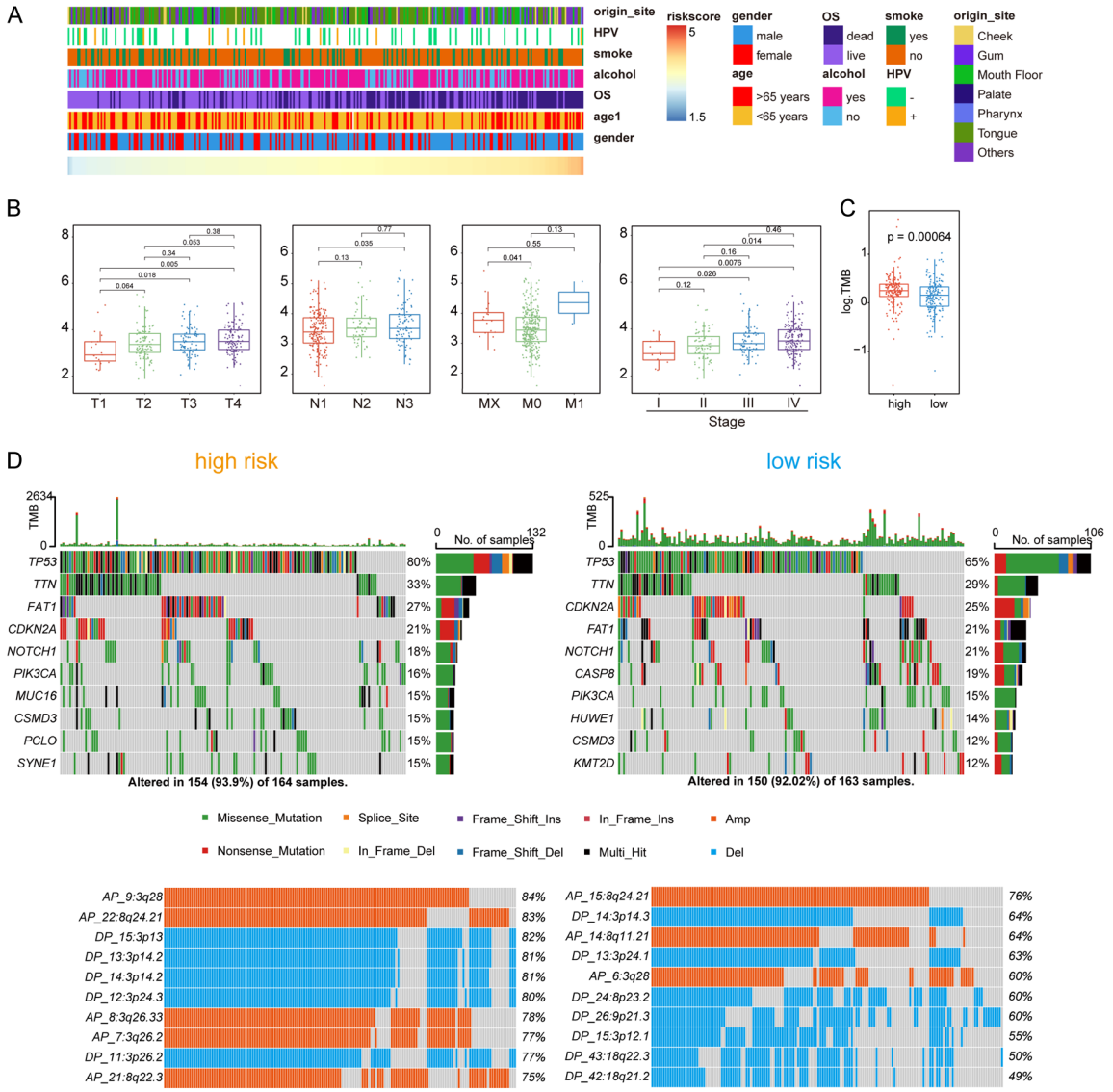


Figure 3. The relationship between risk score and clinical features and comparison of somatic mutations and CNV in different risk group. A. Distribution of clinical features with the increasing risk score. B. Comparison of risk score in TNM stage and tumor stage. C. The overall TMB in different risk group. D. The somatic mutations (upper) and CNV (lower) in high-risk group and low-risk group.

accurate than each index alone (**Figure 4E**). The C-index value of the model combining age and risk score was 0.721.

Identification of different metabolic pathways and functional enrichment analysis

To further investigate the changes in metabolism between the two risk groups, we calculated the scores of all samples in 90 KEGG metabolic pathways using ssGSEA and then performed a t-test to screen the pathways with *p*-values less than 0.01 (**Figure 5A**; Table S4).

We found that eight of the 31 metabolic pathways with significant difference between risk groups were related to lipid metabolism and were enriched in the low-risk group. The pathways associated with glucose metabolism were enriched in the high-risk group, including the glycolysis/gluconeogenesis pathway. Notably, many redox-related pathways were enriched in the high-risk group, including ascorbate and aldarate metabolism, glutathione metabolism, and cytochrome P450-mediated xenobiotic metabolism, which are related to the resistance of adverse environments. In addition, the

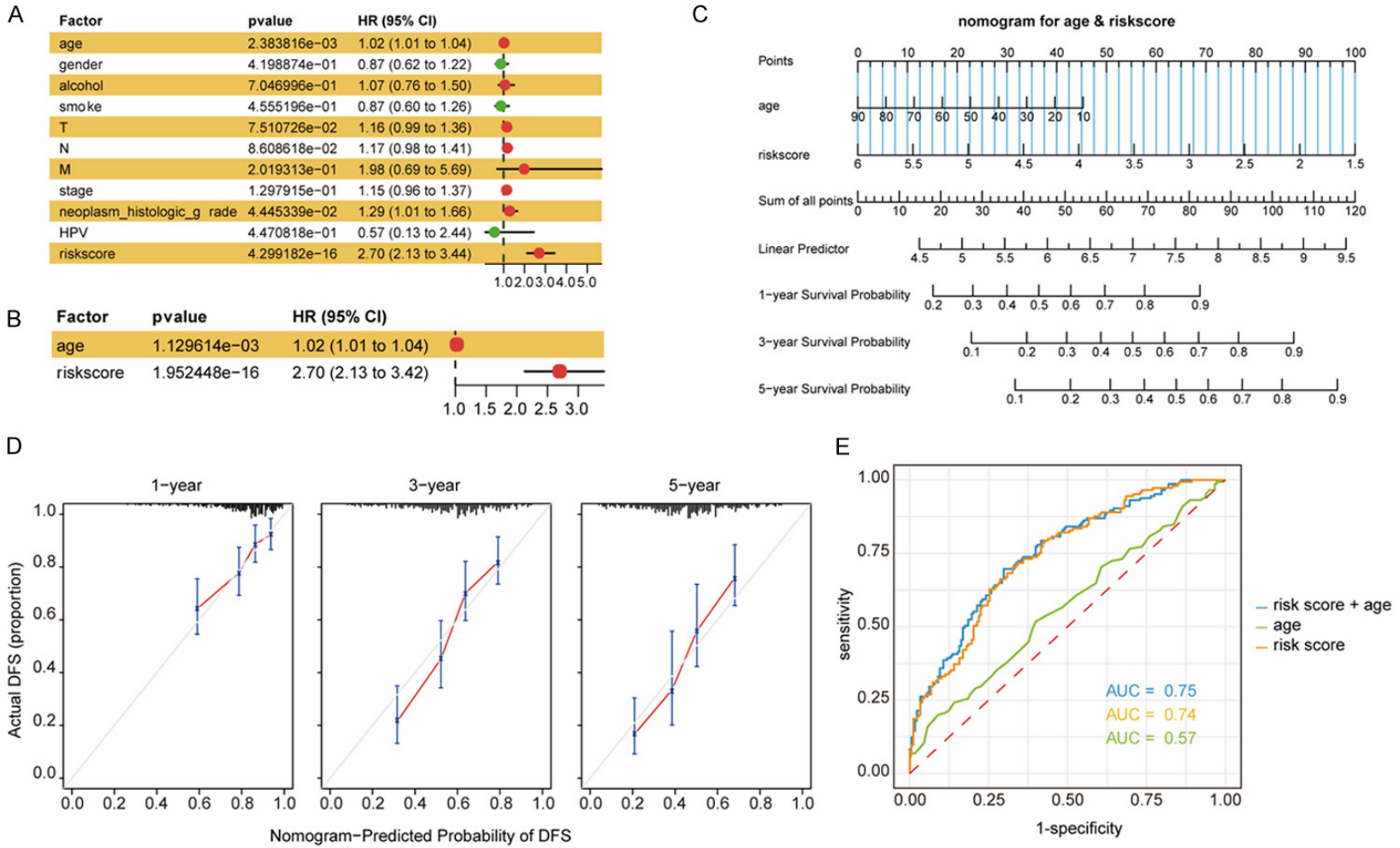


Figure 4. The metabolic signature was an independent factor for prognosis of the survival of OSCC patients. A. Univariate Cox regression analysis of clinical features and risk score with the OS of patients. B. Multivariate Cox regression analysis of risk score and age with the OS of patients. C. Nomogram predictions for survival predictions at 1, 3, 5 years. D. The calibration plot for the nomogram at 1, 3, 5 years. E. ROC curve for the predictions by age, risk score and the combine of them.

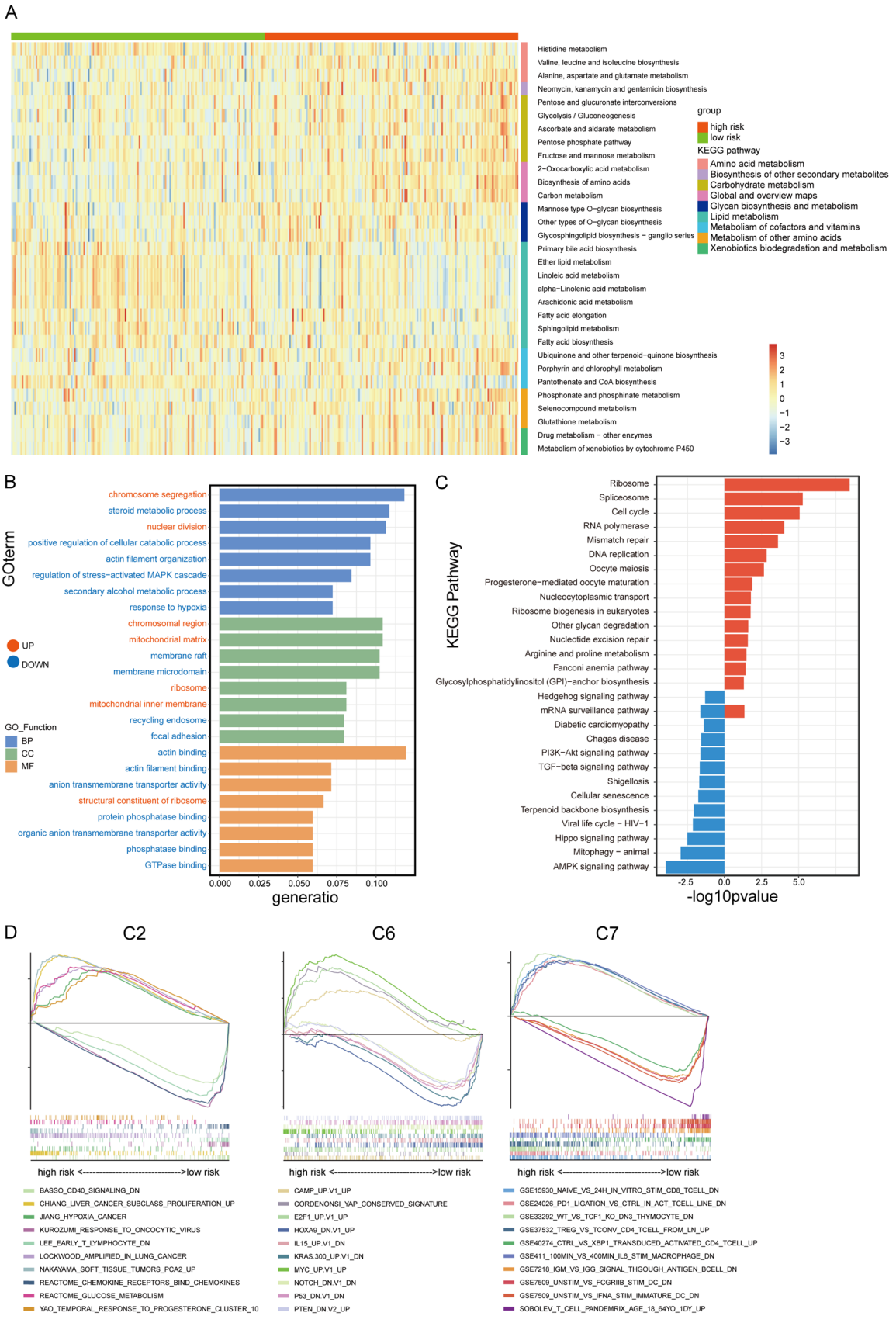

A
Histidine metabolism
Valine, leucine and isoleucine biosynthesis
Alanine, aspartate and glutamate metabolism
Neomycin, kanamycin and gentamicin biosynthesis
Pentose and glucuronate interconversions
Glycolysis / Gluconeogenesis
Ascorbate and aldarate metabolism
Pentose phosphate pathway
Fructose and mannose metabolism
2−Oxocarboxylic acid metabolism
Biosynthesis of amino acids
Carbon metabolism
Mannose type O−glycan biosynthesis
Other types of O−glycan biosynthesis
Glycosphingolipid biosynthesis − ganglio series
Primary bile acid biosynthesis
Ether lipid metabolism
Linoleic acid metabolism
alpha−Linolenic acid metabolism
Arachidonic acid metabolism
Fatty acid elongation
Sphingolipid metabolism
Fatty acid biosynthesis
Ubiquinone and other terpenoid−quinone biosynthesis
Porphyrin and chlorophyll metabolism
Pantothenate and CoA biosynthesis
Phosphonate and phosphinate metabolism
Selenocompound metabolism
Glutathione metabolism
Drug metabolism − other enzymes
Metabolism of xenobiotics by cytochrome P450
group
high risk
low risk
KEGG pathway
Amino acid metabolism
Biosynthesis of other secondary metabolites
Carbohydrate metabolism
Global and overview maps
Glycan biosynthesis and metabolism
Lipid metabolism
Metabolism of cofactors and vitamins
Metabolism of other amino acids
Xenobiotics biodegradation and metabolism
3
2
1
0
−1
−2
−3
B
GOterm
chromosome segregation
steroid metabolic process
nuclear division
positive regulation of cellular catabolic process
actin filament organization
regulation of stress−activated MAPK cascade
secondary alcohol metabolic process
response to hypoxia
chromosomal region
mitochondrial matrix
membrane raft
membrane microdomain
ribosome
mitochondrial inner membrane
recycling endosome
focal adhesion
actin binding
actin filament binding
anion transmembrane transporter activity
structural constituent of ribosome
protein phosphatase binding
organic anion transmembrane transporter activity
phosphatase binding
GTPase binding
UP
DOWN
GO_Function
BP
CC
MF
0.000
0.025
0.050
0.075
0.100
generatio
C
KEGG Pathway
Ribosome
Spliceosome
Cell cycle
RNA polymerase
Mismatch repair
DNA replication
Oocyte meiosis
Progesterone−mediated oocyte maturation
Nucleocytoplasmic transport
Ribosome biogenesis in eukaryotes
Other glycan degradation
Nucleotide excision repair
Arginine and proline metabolism
Fanconi anemia pathway
Glycosylphosphatidylinositol (GPI)−anchor biosynthesis
Hedgehog signaling pathway
mRNA surveillance pathway
Diabetic cardiomyopathy
Chagas disease
PI3K−Akt signaling pathway
TGF−beta signaling pathway
Shigellosis
Cellular senescence
Terpenoid backbone biosynthesis
Viral life cycle − HIV−1
Hippo signaling pathway
Mitophagy − animal
AMPK signaling pathway
-2.5
0.0
2.5
5.0
-log10pvalue
D
C2
C6
C7
high risk <----------->low risk
high risk <----------->low risk
high risk <----------->low risk
BASSO_CD40_SIGNALING_DN
CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_UP
JIANG_HYPOXIA_CANCER
KUROZUMI_RESPONSE_TO_ONCOCYTIC_VIRUS
LEE_EARLY_T_LYMPHOCYTE_DN
LOCKWOOD_AMPLIFIED_IN_LUNG_CANCER
NAKAYAMA_SOFT_TISSUE_TUMORS_PCA2_UP
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES
REACTOME_GLUCOSE_METABOLISM
YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_10
CAMP_UP.V1_UP
CORDENONSI_YAP_CONSERVED_SIGNATURE
E2F1_UP.V1_UP
HOXA9_DN.V1_UP
IL15_UP.V1_DN
KRAS.300_UP.V1_DN
MYC_UP.V1_UP
NOTCH_DN.V1_DN
P53_DN.V1_DN
PTEN_DN.V2_UP
GSE15930_NAIVE_VS_24H_IN_VITRO_STIM_CD8_TCELL_DN
GSE24026_PD1_LIGATION_VS_CTRL_IN_ACT_TCELL_LINE_DN
GSE33292_WT_VS_TCF1_KO_DN3_THYMOCYTE_DN
GSE37532_TREG_VS_TCONV_CD4_TCELL_FROM_LN_UP
GSE40274_CTRL_VS_XBP1_TRANSDUCED_ACTIVATED_CD4_TCELL_UP
GSE411_100MIN_VS_400MIN_IL6_STIM_MACROPHAGE_DN
GSE7218_IGM_VS_IGG_SIGNAL_THGOUGH_ANTIGEN_BCELL_DN
GSE7509_UNSTIM_VS_FCGRIIB_STIM_DC_DN
GSE7509_UNSTIM_VS_IFNA_STIM_IMMATURE_DC_DN
SOBOLEV_T_CELL_PANDEMRIX_AGE_18_64YO_1DY_UP

Figure 5. The ssGSEA score of metabolic pathways in different group and the DEGs analysis by GSEA, GO, KEGG pathway enrichment. A. Heatmap of 31 different metabolism pathways in high-risk and low-risk group ($P < 0.01$). B. GO enrichment of DEGs. C. KEGG enrichment of DEGs. D. GSEA enrichment of DEGs.

enrichment of amino acid synthesis and drug metabolism pathways in the high-risk group was in consistent with high malignancy.

Subsequently, we quadrupled the patients by risk score and took the highest and lowest quartiles for differential gene expression analysis. The "limma3.50.3" package was used to identify the DEGs among these two groups (high risk vs low risk), and 371 genes were obtained, with 281 genes upregulated and 90 genes downregulated in the high-risk group. GO pathway enrichment analysis showed that the upregulated genes were enriched in chromosome segregation and nuclear division, while genes with decreased expression were enriched in steroid metabolic process, positive regulation of cellular catabolic process, actin filament organization, regulation of stress-activated mitogen-activated protein kinase (MAPK) cascade, secondary alcohol metabolism process, and response to hypoxia (**Figure 5B**). KEGG enrichment analysis showed that DNA replication, cell cycle, and mismatch repair pathways were enriched in the high-risk group, while the AMPK and Hippo pathways were enriched in the low-risk group (**Figure 5C**). Taken together, we found that more genetic mutations, cell cycle and metabolic disorders existed in the high-risk group, whereases more stable cytoskeleton and stronger cell adhesion were in the low-risk group. In addition, we analyzed the DEGs by using GSEA software with three gene sets: C2 (curated gene sets), C6 (oncogenic signature gene sets), and C7 (immunologic signature gene sets) (**Figure 5D**). The results showed that pathways relevant to immune response, such as chemokine receptors, T lymphocytes, and CD40 signaling pathways, were enriched in the low-risk group, while pathways related to tumor malignancy and proliferation were enriched in the high-risk group. In the C6 gene sets, some tumor suppressor gene-related signaling pathways such as *P53* and *PTEN*, and some oncogenic pathways such as *NOTCH1* and *KRAS* were normal in the low-risk group but were mutated in the high-risk group. Higher *MYC* expression was also noted in the high-risk group, indicating that tumors in the high-risk group were more malignant. In addition, the C7 gene sets were associated with immune pathways, where the high-risk group was enriched with more immunosuppressive pathways, including PD1 ligation, Treg, and IL6-induced M2 polarization of macrophages, while the low-risk group was enriched with CD8+ T cells and interferon (IFN) pathways.

Distinct immune landscape and ICB efficiency between different risk groups

To further investigate the immune characteristics of the different risk groups, 22 types of immune cells were analyzed using CIBERSORT. We ranked the samples with increasing number of CD8+ T cells in the immune cell heatmap and demonstrated that the area of CD8+ T cells (blue) was larger in low-risk patients (**Figure 6A**). To better show the differences in cell type between the high- and low-risk groups, we used box plots to present the results. As shown in **Figure 6B**, the high-risk group was enriched with M2 macrophages which are recognized as immunosuppressive cells. M0 Macrophages were also enriched in the high-risk group although the difference from the low-risk group was not significant. Interestingly, **Figure 6B** also indicated that CD8+ T cells, dendritic cells, B cells, T follicular helper cells (Tfh), and Tγδ were enriched in the low-risk group. Some studies have shown that B cells and CD4+ T cells contribute to the formation of tertiary lymphoid structures (TLSs) in EMT. Crosstalk between immune cells in TLSs could promote a series of immune processes including antigen presentation, maintenance of T cell states, and the formation of memory T cells; the interplay network was shown in **Figure 6C**. Given that DEGs were enriched in the CD40 signaling pathway in GSEA enrichment analysis, we investigated CD40 and CD40LG expression in high- and low-risk groups and found that CD40LG expression was higher in the low-risk group, which might promote the activation of T cells and the killing of CD40+ tumor cells (**Figure 6D**). Furthermore, we collected gene sets that represented immune molecules and processes (Table S5) and analyzed them using ssGSEA. The data suggested

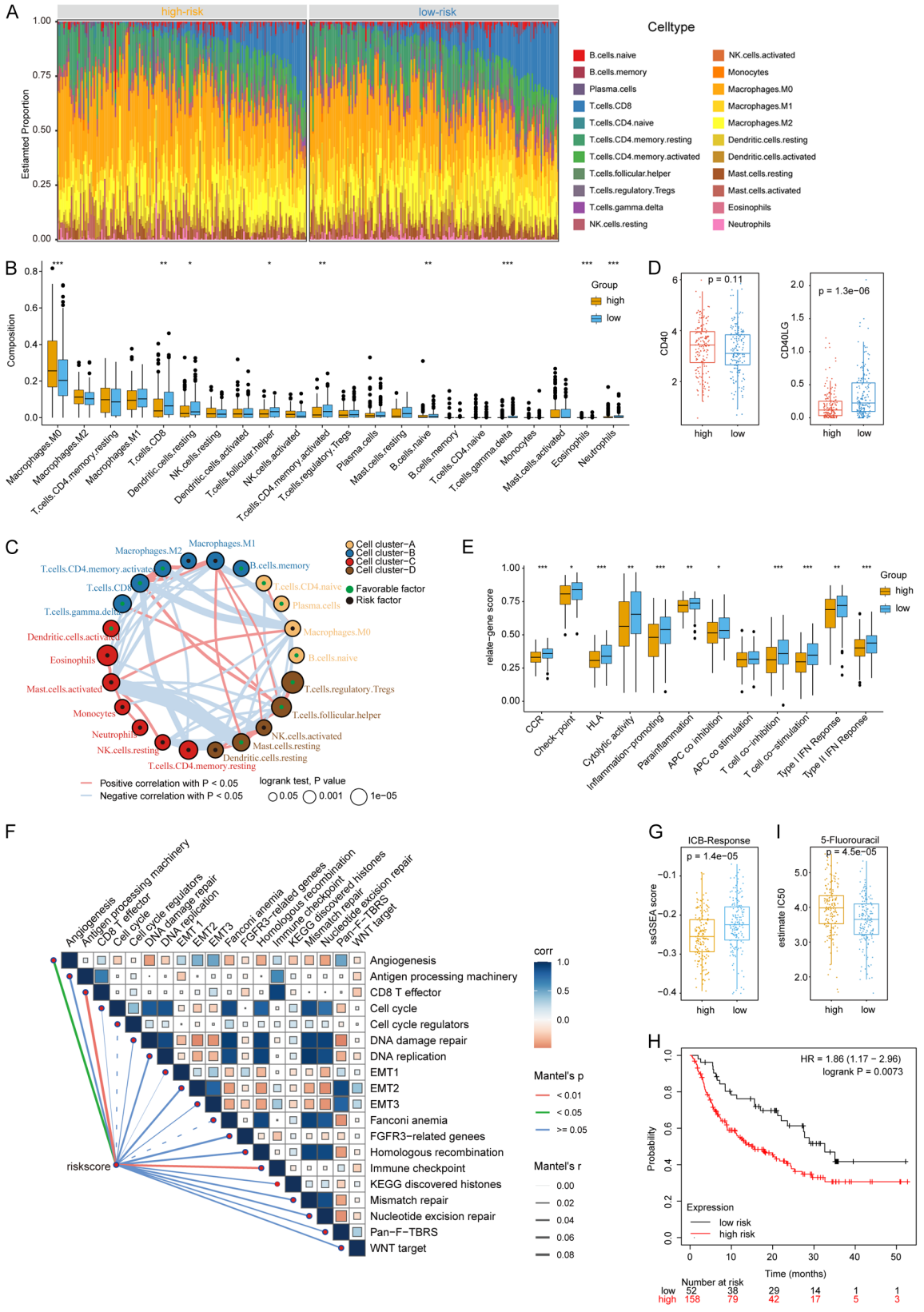


Figure 6. The immune landscape in different risk groups and the relationship between rick score and drug response. A. The heatmap of 22 types of immune cells estimated by CIBERDORT. B. The boxplots of 22 types of immune cells estimated by CIBERDORT. C. The crosstalk between 22 types of immune cells. D. CD40 and CD40LG expression of

different risk groups. E. The ssGASE score of immune processes in different risk groups. F. The correlation of risk score and 19 signatures provided by Thomas Powles et al. G. The ICB-Response score in high- and low-risk groups. H. Kaplan-Meier survival of patients grouped by metabolic signature in KM plot's anti-PD1 therapy array. I. The evaluated IC50 of 5-Fluorouracil. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

that the low-risk group provided a microenvironment more suitable for T cell co-stimulation; thus, cytolytic activity and type I and II IFN responses were stronger in the low-risk group than in the high-risk group, which further enhanced the antitumor effect and increased the MHC presentation on tumor cells (**Figure 6E**). Moreover, the low-risk group had higher checkpoint scores due to many factors, such as increased T-cell infiltration and better T-cell activation, suggesting that patients in this group might benefit from ICB treatment. The expression levels of the seven checkpoint genes were presented in Figure S4A. To further validate our ssGSEA results, we performed ssGSEA again using the dataset from Thomas Powles et al. [36] and conducted a correlation analysis with each signature (**Figure 6F**). The results showed that the risk score was associated with EMT, antigen presentation, CD8+ T effector cells, and DNA repair as well as angiogenesis. More importantly, the metabolic risk score could be used as an indicator to predict the efficacy of immunotherapy. We compared the therapeutic response to anti-PD-1 between the two risk groups via the ICB response score and revealed that the low-risk group had a better response to ICB therapy (**Figure 6G**). Consistently, data from KMplot dataset which contains information on patients with anti-PD1 mAb treatment showed that the low-risk group had higher survival rate than the high-risk group (**Figure 6H**). We also used pRRophetic package to evaluate the IC50 of three first-line drugs: 5-Fluorouracil (**Figure 6I**), Cetuximab (Figure S4B), and Cisplatin (Figure S4C), and the results suggested that 5-Fluorouracil might not have therapeutic effect in patients at the high-risk group.

Single-cell sequencing analysis reveals the expression of 12 genes on different cell clusters

The published scRNA sequencing dataset (GSE136103) from the GEO database contains 5902 cells from 18 patients. After filtering, we used 5746 cells for subsequent analysis (Figure S5A-C). Eighteen clusters were generated using the tSNE algorithm (Figure S5D). Based on the expression signatures of these cells (Table S6), we classified them into eight clusters: B cells, CD4+ T cells, CD8+ NK cells, dendritic cells, endothelial cells, tumor cells, fibroblasts, macrophages, and mast cells (Figures S5E, **7A**). Our results were consistent with the results from dataset provided by Bradley et al. [29]. We also used *EPCAM* to identify malignant cell clusters (**Figure 7B**). Furthermore, we performed single cell CNV analysis by using the "CopyKAT" package, which discriminated malignant cells from normal cells. The results showed that out of the 5746 cells, 2318 cells were malignant cells, while the other 3428 cells were normal cells (Figure S6). In addition, the malignant cells belonged to the same group as the tumor cell population we previously identified. Furthermore, we divided the tumor cells into two clusters (1654 cells in cluster 1 and 664 cells in cluster 2) based on CNV (Figures S7, **7C**) and found that *EPCAM* was expressed at significantly higher levels in cluster 2 compared to cluster 1. It is well known that *EPCAM* is a biomarker of cancer stem cells and is involved in EMT. In addition, cells in cluster 2 overexpressed *SOX2*, another marker of cancer stem cells (**Figure 7D**). Combined the CNV mutation data with the expression of *EPCAM* and *SOX2*, we speculated that tumors in cluster 2 were more malignant than that in cluster 1.

Among the 12 genes we used to construct metabolism-related signature, *HK1, PTGR1, MCCC1, UGT1A10*, and *NADSYN1* were highly expressed in tumor cells, while *SLC2A3*, *INPP5D*, and *CSGALNACT1* were highly expressed in immune cells and fibroblasts. In addition, *SLC2A3* was also expressed in cluster 2 tumor cells, while FOMD was expressed only in fibroblasts (**Figure 7E**). *SLC2A3*, also known as *GLUT3*, has the highest affinity for glucose among the four glucose transporters and tends to be highly expressed in malignant tumors to meet their metabolic requirements. Furthermore, *PTGR1* and *UGT1A10* expression was upregulated in tumors of cluster 2 compared to tumors in cluster 1. In contrast, *INPP5D*, *CSGALNACT1*, and *FOMD* were only

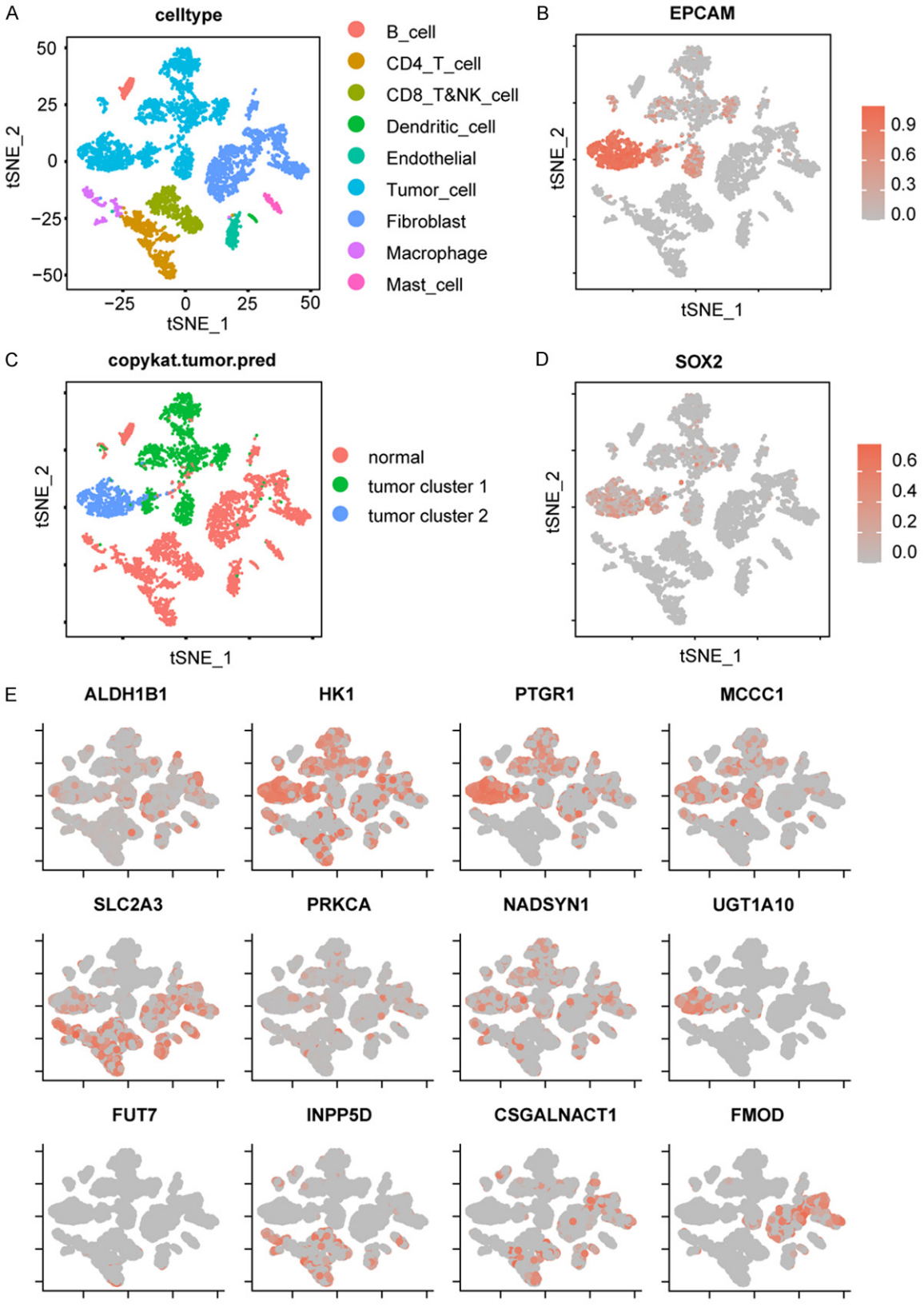


Figure 7. Single cell sequencing revealed the specific expression of the 12 genes in 8 cell types. A. Cell types were identified by marker genes. B. The expression of *EPCAM* in each cell type. C. Tumor cells were divided into 2 clusters by CopyKAT. D. The expression of *SOX2* in each cell type. E. Expression profile of the 12 genes in each cell type.

expressed in non-tumor cells and had negative coefficients with the risk score. Together, our analysis on the gene expression profile in each cell type provides a useful reference for the development of therapeutic drugs.

Discussion

Various studies have focused on the EMT process in OSCC to investigate the mechanisms by which OSCC metastasizes. EMT activation results in the acquisition of stem cell properties, aggressive mobility, and immune resistance. EMT is usually associated with metabolic reprogramming, which is widely implemented by malignant cells to meet their needs for rapid proliferation and mobility in harsh tumor microenvironments. In this study, we used WGCNA, a systematic biology method, to identify the metabolic genes related to EMT, stemness, and checkpoints. We identified a green module which had the most positive correlation with EMT and checkpoints, but not with stemness, which might be due to the existence of the multiple states of EMT, each exhibiting distinct cancer stem cells marker [43]. We also selected the brown module related to the checkpoint signature for subsequent analysis. The biological functions of the green module were related to extracellular matrix synthesis. It has been reported that glycosaminoglycan can suppress the activation of T cells and dendritic cells [44], and that abundant ECM can impede immune cell infiltration, leading to the immune escape of tumor cells in OSCC. In addition, the oxidative stress-related pathways were also enriched in the module, suggesting the important role in the induction of EMT. One important molecule in the oxidative stress-related pathways was NRF2 which can promote the transcription of several genes involved in the formation of NADPH to help scavenge ROS [45].

We also constructed a metabolic signature consisting of 12 genes for predicting OSCC prognosis by comprehensive bioinformatic analysis of the green and brown modules. We verified that the risk score and age could be used as independent prognostic factors. OSCC patients with high-risk scores had poor outcomes. Furthermore, the risk score also correlated with other common malignant features. For example, high-risk score was associated with a more severe TNM stage and a higher TMB. A high mutation rate of TP53 was found in OSCC patients, especially in the high-risk group. The CNV patterns differed between the two risk groups, with the high-risk group showing more genomic amplifications of several genes, including TP63. Nevertheless, whether TP63 acts as a tumor suppressive or tumor promoting factor is not clear [46-48], and our results showed that TP63 amplification was associated with more malignant OSCC.

The high- and low-risk groups presented different patterns of metabolic pathways. Glycolysis, amino acid synthesis, and glutamine metabolism pathways were upregulated in the high-risk group. Similarly, drug metabolism and CYP450 pathway were upregulated, suggesting that the high-risk group were more resistant to chemotherapy. Interestingly, the lipid metabolism and arachidonic acid metabolism pathways were enriched in the low-risk group, as these pathways are generally considered to be positively correlated with malignant tumors. Furthermore, GO pathway analysis revealed that the MAPK pathway was enriched in the low-risk group. Although activating mutations in MAPK are common in tumors and are generally considered to be oncogenic [49, 50], recent studies suggest that the role of MAPK pathway mutations may be different in HNSCC, which was supported by our results. Other evidence also supports this notion. For example, multiple MAPK pathway activating mutations found in HNSCC could significantly suppress p-ErbB3 expression and promote CD8+ T cell infiltration, which are associated with better survival [51]. Furthermore, previous studies have shown that arachidonic acid and other fatty acids can activate members of the MAPK superfamily [52, 53], which explains why lipid and arachidonic acid metabolism were active in the low-risk OSCC group. We also found the differential expression of some classical tumor-associated pathways, including *TP53*, *PTEN*, *MYC*, *KRAS*, and *E2F1*, between the high- and low-risk groups.

In this study, we showed that the risk score was associated with the immune landscape of OSCC patients. First, the population of CD8+ T cells was expanded in the low-risk group, and other factors, including HLA, cytolytic activity, and co-stimulation, also indicated that the low-risk group had a more active immune environ-

ment. In the low-risk group, T cells exhibited higher CD40L expression, but tumor cells exhibited lower CD40 expression, suggesting the possibility of applying CD40LG signal in the design of chimeric antigen receptor T cell (CAR-T) for an improved OSCC therapy. Another significant finding from our analysis was that the risk score was reliable in predicting the response to immunotherapy, with a lower risk score indicating a higher ICB-response score and a better survival outcome. The predicted IC50 results showed that the high-risk score predicted resistance to 5-fluorouracil but not to cetuximab and cisplatin, suggesting cetuximab and cisplatin as more efficient therapeutic drug in treating malignant OSCC.

Finally, scRNA-seq data were used to further explain the signature we constructed. The tumors in cluster 2 expressed high levels of *EPCAM* and *SOX2* which are markers of malignant tumor cells. Among the 12 signature genes, *PTGR1*, *UTG1A10* and *SLC2A3*, were differentially expressed between the two tumor clusters, indicating that these genes could be potential therapeutic targets in the treatment of refractory OSCC. Indeed, inhibitors of *PTGR1* and *UTG1A10*, including indomethacin (PTGR1 inhibitor) and sorafenib (UTG1A10 inhibitor), have been used clinical in other cancer types [54, 55]. As for *SLC2A3,* the molecular basis for its inhibition by exofacial inhibitors has been reported [56], which could serve as potential therapeutic target. On the other hand, Hexokinase 1 (HK1), the first rate-limiting enzyme in glycolysis and a closely related enzyme to cancer progression, is expressed in both tumor and non-tumor cells. It has been reported that HK1 is upregulated when T cells are activated [57], which would probably interfere with our predictions using the risk score. Therefore, the risk score would be more suitable for the assessment of patients prior to treatment. Notably, another two genes in the 12 gene signature, *CSGALNACT1* and *FMOD* were associated with the metabolism of the extracellular matrix, suggesting that the immune state of the EMT can be improved by regulating the extracellular matrix to achieve good therapeutic outcomes in OSCC.

In conclusion, we constructed a metabolic gene co-expression network for OSCC and identified metabolic genes most relevant to EMT, stemness, and checkpoint signatures. A prognostic metabolic signature was constructed and validated. The risk score derived from the signature was closely related to the clinical features and could be used to predict the response to immune therapy. Importantly, some potential targets identified from the analyzing metabolic characteristics in different risk groups may help develop clinical treatment plans.

Nevertheless, there were some limitations in this work. First, datasets of large immunotherapy, especially in OSCC, should be used to verify our signature. Second, in vitro and in vivo experiments should be carried out to validate these potential targets and pathways. Finally, our research only studied the relationship between metabolism and OSCC metastasis at the transcriptional level, and we look forward to a multi-omics study to reveal more details about the role of metabolism in OSCC metastasis.

Acknowledgements

We thank all researchers sharing the data used in this work. Supported by the projects named as Health Commission of Hubei Province scientific research project [grant number WJ2021M166], and Supporting project of medical science and technology innovation platform of Zhongnan Hospital of Wuhan University [grant number PTXM2021014].

Disclosure of conflict of interest

None.

Address correspondence to: Drs. Sisi Yang and Bo Cheng, Department of Stomatology, Zhongnan Hospital of Wuhan University, 169 Donghu Road, Wuhan 430071, Hubei, China. Tel: +86-1534-2261967; E-mail: 200632220036@whu.edu.cn (SSY); Tel: +86-027-87686110; E-mail: chengbo@znhospital.cn (BC)

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, Dal Maso L,

Daudt AW, Fabianova E, Fernandez L, Wünsch-Filho V, Franceschi S, Hayes RB, Herrero R, Koifman S, La Vecchia C, Lazarus P, Levi F, Mates D, Matos E, Menezes A, Muscat J, Eluf-Neto J, Olshan AF, Rudnai P, Schwartz SM, Smith E, Sturgis EM, Szeszenia-Dabrowska N, Talamini R, Wei Q, Winn DM, Zaridze D, Zatonski W, Zhang ZF, Berthiller J and Boffetta P. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst 2007; 99: 777-789.

[3] Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, Bernstein L, Schoenberg JB, Stemhagen A and Fraumeni JF Jr. Smoking and drinking in relation to oral and pharyngeal cancer. Cancer Res 1988; 48: 3282-3287.

[4] Zittel S, Moratin J, Horn D, Metzger K, Ristow O, Engel M, Mrosek J, Freier K, Hoffmann J and Freudlsperger C. Clinical outcome and prognostic factors in recurrent oral squamous cell carcinoma after primary surgical treatment: a retrospective study. Clin Oral Investig 2022; 26: 2055-2064.

[5] Geum DH, Roh YC, Yoon SY, Kim HG, Lee JH, Song JM, Lee JY, Hwang DS, Kim YD, Shin SH, Chung IK and Kim UK. The impact factors on 5-year survival rate in patients operated with oral cancer. J Korean Assoc Oral Maxillofac Surg 2013; 39: 207-216.

[6] D'Cruz AK, Vaish R and Dhar H. Oral cancers: current status. Oral Oncol 2018; 87: 64-69.

[7] Ozdek A, Sarac S, Akyol MU, Unal OF and Sungur A. Histopathological predictors of occult lymph node metastases in supraglottic squamous cell carcinomas. Eur Arch Otorhinolaryngol 2000; 257: 389-392.

[8] Roberts TJ, Colevas AD, Hara W, Holsinger FC, Oakley-Girvan I and Divi V. Number of positive nodes is superior to the lymph node ratio and American Joint Committee on Cancer N staging for the prognosis of surgically treated head and neck squamous cell carcinomas. Cancer 2016; 122: 1388-1397.

[9] Zeng YT, Liu XF, Yang WT and Zheng PS. REX1 promotes EMT-induced cell metastasis by activating the JAK2/STAT3-signaling pathway by targeting SOCS1 in cervical cancer. Oncogene 2019; 38: 6940-6957.

[10] Gumireddy K, Li A, Gimotty PA, Klein-Szanto AJ, Showe LC, Katsaros D, Coukos G, Zhang L and Huang Q. KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. Nat Cell Biol 2009; 11: 1297-1304.

[11] Nieto MA, Huang RY, Jackson RA and Thiery JP. EMT: 2016. Cell 2016; 166: 21-45.

[12] Saxena M, Stephens MA, Pathak H and Rangarajan A. Transcription factors that mediate epithelial-mesenchymal transition lead to multidrug resistance by upregulating ABC transporters. Cell Death Dis 2011; 2: e179.

[13] Dongre A and Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol 2019; 20: 69-84.

[14] Scheel C and Weinberg RA. Cancer stem cells and epithelial-mesenchymal transition: concepts and molecular links. Semin Cancer Biol 2012; 22: 396-403.

[15] Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, Campbell LL, Polyak K, Brisken C, Yang J and Weinberg RA. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 2008; 133: 704-715.

[16] Dongre A, Rashidian M, Reinhardt F, Bagnato A, Keckesova Z, Ploegh HL and Weinberg RA. Epithelial-to-mesenchymal transition contributes to immunosuppression in breast carcinomas. Cancer Res 2017; 77: 3982-3989.

[17] Chockley PJ, Chen J, Chen G, Beer DG, Standiford TJ and Keshamouni VG. Epithelial-mesenchymal transition leads to NK cell-mediated metastasis-specific immunosurveillance in lung cancer. J Clin Invest 2018; 128: 1384-1396.

[18] Nilsson MB, Sun H, Robichaux J, Pfeifer M, McDermott U, Travers J, Diao L, Xi Y, Tong P, Shen L, Hofstad M, Kawakami M, Le X, Liu X, Fan Y, Poteete A, Hu L, Negrao MV, Tran H, Dmitrovsky E, Peng D, Gibbons DL, Wang J and Heymach JV. A YAP/FOXM1 axis mediates EMT-associated EGFR inhibitor resistance and increased expression of spindle assembly checkpoint components. Sci Transl Med 2020; 12: eaaz4589.

[19] Paz-Ares L, Ciuleanu TE, Cobo M, Schenker M, Zurawski B, Menezes J, Richardet E, Bennouna J, Felip E, Juan-Vidal O, Alexandru A, Sakai H, Lingua A, Salman P, Souquet PJ, De Marchi P, Martin C, Pérol M, Scherpereel A, Lu S, John T, Carbone DP, Meadows-Shropshire S, Agrawal S, Oukessou A, Yan J and Reck M. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international, randomised, open-label, phase 3 trial. Lancet Oncol 2021; 22: 198-211.

[20] Ferris RL, Blumenschein G Jr, Fayette J, Guigay J, Colevas AD, Licitra L, Harrington KJ, Kasper S, Vokes EE, Even C, Worden F, Saba NF, Docampo LCI, Haddad R, Rordorf T, Kiyota N, Tahara M, Lynch M, Jayaprakash V, Li L and Gil-

lison ML. Nivolumab vs investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck: 2-year long-term survival update of CheckMate 141 with analyses by tumor PD-L1 expression. Oral Oncol 2018; 81: 45-51.
[21] Shibue T and Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat Rev Clin Oncol 2017; 14: 611-629.
[22] Sciacovelli M and Frezza C. Metabolic reprogramming and epithelial-to-mesenchymal transition in cancer. FEBS J 2017; 284: 3132-3144.
[23] Tam WL and Weinberg RA. The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat Med 2013; 19: 1438-1449.
[24] Corbet C, Bastien E, Santiago de Jesus JP, Dierge E, Martherus R, Vander Linden C, Doix B, Degavre C, Guilbaud C, Petit L, Michiels C, Dessy C, Larondelle Y and Feron O. TGFβ2-induced formation of lipid droplets supports acidosis-driven EMT and the metastatic spreading of cancer cells. Nat Commun 2020; 11: 454.
[25] Anderson M, Marayati R, Moffitt R and Yeh JJ. Hexokinase 2 promotes tumor growth and metastasis by regulating lactate production in pancreatic cancer. Oncotarget 2017; 8: 56081-56094.
[26] Park JH, Vithayathil S, Kumar S, Sung PL, Dobrolecki LE, Putluri V, Bhat VB, Bhowmik SK, Gupta V, Arora K, Wu D, Tsouko E, Zhang Y, Maity S, Donti TR, Graham BH, Frigo DE, Coarfa C, Yotnda P, Putluri N, Sreekumar A, Lewis MT, Creighton CJ, Wong LC and Kaipparettu BA. Fatty acid oxidation-driven Src links mitochondrial energy reprogramming and oncogenic properties in triple-negative breast cancer. Cell Rep 2016; 14: 2154-2165.
[27] Guerra F, Guaragnella N, Arbini AA, Bucci C, Giannattasio S and Moro L. Mitochondrial dysfunction: a novel potential driver of epithelial-to-mesenchymal transition in cancer. Front Oncol 2017; 7: 295.
[28] Wang J, Jiang C, Li N, Wang F, Xu Y, Shen Z, Yang L, Li Z and He C. The circEPSTI1/miR-942-5p/LTBP2 axis regulates the progression of OSCC in the background of OSF via EMT and the PI3K/Akt/mTOR pathway. Cell Death Dis 2020; 11: 682.
[29] Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, Deschler DG, Varvares MA, Mylvaganam R, Rozenblatt-Rosen O, Rocco JW, Faquin WC, Lin DT, Regev A and Bernstein BE. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell 2017; 171: 1611-1624, e24.
[30] Yu G and He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst 2016; 12: 477-479.
[31] Alonso SR, Tracey L, Ortiz P, Pérez-Gómez B, Palacios J, Pollán M, Linares J, Serrano S, Sáez-Castillo AI, Sánchez L, Pajares R, Sánchez-Aguilera A, Artiga MJ, Piris MA and Rodríguez-Peralto JL. A high-throughput study in melanoma identifies epithelial-mesenchymal transition as a major determinant of metastasis. Cancer Res 2007; 67: 3450-3460.
[32] Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyeuski A, Bruun J, Micke P, de Reynies A and Nelson BH. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. Proc Natl Acad Sci U S A 2019; 116: 9020-9029.
[33] Xiao B, Liu L, Li A, Xiang C, Wang P, Li H and Xiao T. Identification and verification of immune-related gene prognostic signature based on ssgsea for osteosarcoma. Front Oncol 2020; 10: 607622.
[34] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C and Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019; 10: 1523.
[35] Xing L, Guo M, Zhang X, Zhang X and Liu F. A transcriptional metabolic gene-set based prognostic signature is associated with clinical and mutational features in head and neck squamous cell carcinoma. J Cancer Res Clin Oncol 2020; 146: 621-630.
[36] Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE III, Koeppen H, Astarita JL, Cubas R, Jhunjhunwala S, Banchereau R, Yang Y, Guan Y, Chalouni C, Ziai J, Şenbabaoğlu Y, Santoro S, Sheinson D, Hung J, Giltnane JM, Pierce AA, Mesh K, Lianoglou S, Riegler J, Carano RAD, Eriksson P, Höglund M, Somarriba L, Halligan DL, van der Heijden MS, Loriot Y, Rosenberg JE, Fong L, Mellman I, Chen DS, Green M, Derleth C, Fine GD, Hegde PS, Bourgon R and Powles T. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 2018; 554: 544-548.
[37] Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, Hodi FS, Martín-Algarra S, Mandal R, Sharfman WH, Bhatia S, Hwu WJ, Gajewski TF, Slingluff CL Jr, Chowell D, Kendall SM, Chang H, Shah R, Kuo F, Morris LGT, Sidhom JW, Schneck JP, Horak CE, Weinhold N and Chan TA. Tumor and microenvironment evolution during immunotherapy with nivolumab. Cell 2017; 171: 934-949, e16.
[38] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M and Alizadeh

AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12: 453-457.
[39] Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, Kumar T, Hu M, Sei E, Davis A, Wang F, Shaitelman SF, Wang JR, Chen K, Moulder S, Lai SY and Navin NE. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat Biotechnol 2021; 39: 599-608.
[40] Mittal V. Epithelial mesenchymal transition in tumor metastasis. Annu Rev Pathol 2018; 13: 395-412.
[41] Jiang Y and Zhan H. Communication between EMT and PD-L1 signaling: new insights into tumor immune evasion. Cancer Lett 2020; 468: 72-81.
[42] Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF and Groom AC. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol 1998; 153: 865-873.
[43] Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, Zhou A, Eyob H, Balakrishnan S, Wang CY, Yaswen P, Goga A and Werb Z. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature 2015; 526: 131-135.
[44] du Souich P, García AG, Vergés J and Montell E. Immunomodulatory and anti-inflammatory effects of chondroitin sulphate. J Cell Mol Med 2009; 13: 1451-1463.
[45] Wu S, Lu H and Bai Y. Nrf2 in cancers: a double-edged sword. Cancer Med 2019; 8: 2252-2267.
[46] Melino G. p63 is a suppressor of tumorigenesis and metastasis interacting with mutant p53. Cell Death Differ 2011; 18: 1487-1499.
[47] Zhu D, Jiang XH, Jiang YH, Ding WC, Zhang CL, Shen H, Wang XL, Ma D, Hu Z and Wang H. Amplification and overexpression of TP63 and MYC as biomarkers for transition of cervical intraepithelial neoplasia to cervical cancer. Int J Gynecol Cancer 2014; 24: 643-648.
[48] Watanabe H, Ma Q, Peng S, Adelmant G, Swain D, Song W, Fox C, Francis JM, Pedamallu CS, DeLuca DS, Brooks AN, Wang S, Que J, Rustgi AK, Wong KK, Ligon KL, Liu XS, Marto JA, Meyerson M and Bass AJ. SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. J Clin Invest 2014; 124: 1636-1645.
[49] Hiraide T, Ikegami K, Sakaguchi T, Morita Y, Hayasaka T, Masaki N, Waki M, Sugiyama E, Shinriki S, Takeda M, Shibasaki Y, Miyazaki S, Kikuchi H, Okuyama H, Inoue M, Setou M and Konno H. Accumulation of arachidonic acid-containing phosphatidylinositol at the outer edge of colorectal cancer. Sci Rep 2016; 6: 29935.
[50] Hong SH, Avis I, Vos MD, Martínez A, Treston AM and Mulshine JL. Relationship of arachidonic acid metabolizing enzyme expression in epithelial cancer cell lines to the growth effect of selective biochemical inhibitors. Cancer Res 1999; 59: 2223-2228.
[51] Ngan HL, Liu Y, Fong AY, Poon PHY, Yeung CK, Chan SSM, Lau A, Piao W, Li H, Tse JSW, Lo KW, Chan SM, Su YX, Chan JYK, Lau CW, Mills GB, Grandis JR and Lui VWY. MAPK pathway mutations in head and neck cancer affect immune microenvironments and ErbB3 signaling. Life Sci Alliance 2020; 3: e201900545.
[52] Cussac D, Schaak S, Denis C and Paris H. alpha 2B-adrenergic receptor activates MAPK via a pathway involving arachidonic acid metabolism, matrix metalloproteinases, and epidermal growth factor receptor transactivation. J Biol Chem 2002; 277: 19882-19888.
[53] Gijón MA and Leslie CC. Regulation of arachidonic acid release and cytosolic phospholipase A2 activation. J Leukoc Biol 1999; 65: 330-336.
[54] Wang X, Yin G, Zhang W, Song K, Zhang L and Guo Z. Prostaglandin reductase 1 as a potential therapeutic target for cancer therapy. Front Pharmacol 2021; 12: 717730.
[55] Zhu YJ, Zheng B, Wang HY and Chen L. New knowledge of the mechanisms of sorafenib resistance in liver cancer. Acta Pharmacol Sin 2017; 38: 614-622.
[56] Wang N, Zhang S, Yuan Y, Xu H, Defossa E, Matter H, Besenius M, Derdau V, Dreyer M, Halland N, He KH, Petry S, Podeschwa M, Tennagels N, Jiang X and Yan N. Molecular basis for inhibiting human glucose transporters by exofacial inhibitors. Nat Commun 2022; 13: 2632.
[57] Araujo L, Khim P, Mkhikian H, Mortales CL and Demetriou M. Glycolysis and glutaminolysis cooperatively control T cell function by limiting metabolite supply to N-glycosylation. Elife 2017; 6: e21330.

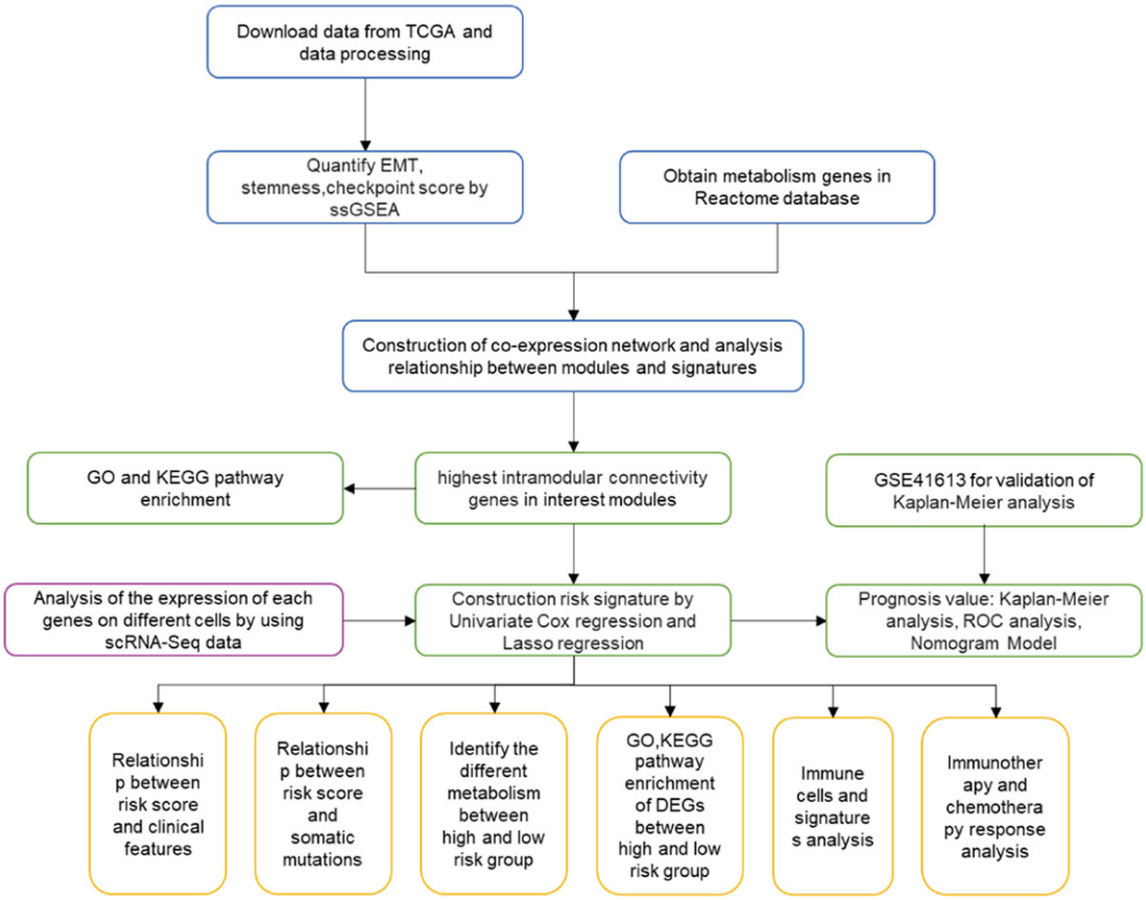


Figure S1. The flow chart of this study.

Table S1. 330 OSCC patients information in TCGA cohort

Clinical characteristics		Number	n (%)
Age	< 65	126	38.18
	> 65	204	61.82
Gender	male	229	69.39
	female	101	30.61
T	TX-T1	26	7.88
	T2	107	32.42
	T3	83	25.15
	T4	101	30.61
	Null	13	3.94
N	NX-N0	183	55.45
	N1	56	16.97
	N2	85	25.76
	N3	6	1.82
M	MX	19	5.76
	M0	309	93.64
	M1	2	0.61
	Null	3	0.91

Tumor Stage	I	12	3.64
	II	79	23.94
	III	67	20.30
	IV	162	49.09
	Null	10	3.03
HPV	+	10	3.03
	-	58	17.58
	Null	262	79.39
Origin tissue	Tongue	144	43.64
	Mouth Floor	72	21.82
	Cheek	19	5.76
	Pharynx	11	3.33
	Gum	10	3.03
	Plate	5	1.52
	others	69	20.91

Table S2. EMT, stemness, checkpoint signatures

signature	genes
EMT	ANLN, APLP2, CD63, CDH2, CLIC4, VCAN, CTSB, CX3CR1, DSG2, EDNRB, EMP1, ENC1, FGG, FZD1, TUBA4A, HMMR, ITGAV, LUM, L1CAM, MFAP1, MMP2, PFN1, PRKCA, RAB1A, RAN, RRAGA, SDCBP, SELENOP, SERPINA3, SMARCA1, SPA17, SPARC, TUBA4A, TUBA3C, TUBA1A, TUBB3
stemness	DNMT3B, PFAS, XRCC5, HAUS6, TET1, IGF2BP1, PLAA, TEX10, MSH6, DLGAP5, SKIV2L2, SOHLH2, RRAS2, PAICS, CPSF3, LIN28B, IPO5, BMPR1A, ZNF788, ASCC3, FANCB, HMGA2, TRIM24, ORC1, HDAC2, HESX1, INHBE, MIS18A, DCUN1D5, MRPL3, CENPH, MYCN, HAUS1, GDF3, TBCE, RIOK2, BCKDHB, RAD1, NREP, ADH5, PLRG1, ROR1, RAB3B, DIAPH3, GNL2, FGF2, NMNAT2, KIF20A, CENPI, DDX1, XXYLT1, GPR176, BBS9, C14orf166, BOD1, CDC123, SNRPD3, FAM118B, DPH3, EIF2B3, RPF2, APLP1, DACT1, PDHB, C14orf119, DTD1, SAMM50, CCL26, MED20, UTP6, RARS2, ARMCX2, RARS, MTHFD2, DHX15, HTR7, MTHFD1L, ARMC9, XPOT, IARS, HDX, ACTRT3, ERCC2, TBC1D16, GARS, KIF7, UBE2K, SLC25A3, ICMT, UGGT2, ATP11C, SLC24A1, EIF2AK4, GPX8, ALX1, OSTC, TRPC4, HAS2, FZD2, TRNT1, MMADHC, SNX8, CDH6, HAT1, SEC11A, DIMT1, TM2D2, FST, GBE1
checkpoint	IDO1, LAG3, CTLA4, TNFRSF9, ICOS, CD80, PDCD1LG2, TIGIT, CD70, TNFSF9, ICOSLG, KIR3DL1, CD86, PDCD1, LAIR1, TNFRSF8, TNFSF15, TNFRSF14, IDO2, CD276, CD40, TNFRSF4, TNFSF14, HHLA2, CD244, CD274, HAVCR2, CD27, BTLA, LGALS9, TMIGD2, CD28, CD48, TNFRSF25, CD40LG, ADORA2A, VTCN1, CD160, CD44, TNFSF18, TNFRSF18, BTNL2, C10orf54, CD200R1, TNFSF4, CD200, NRP1

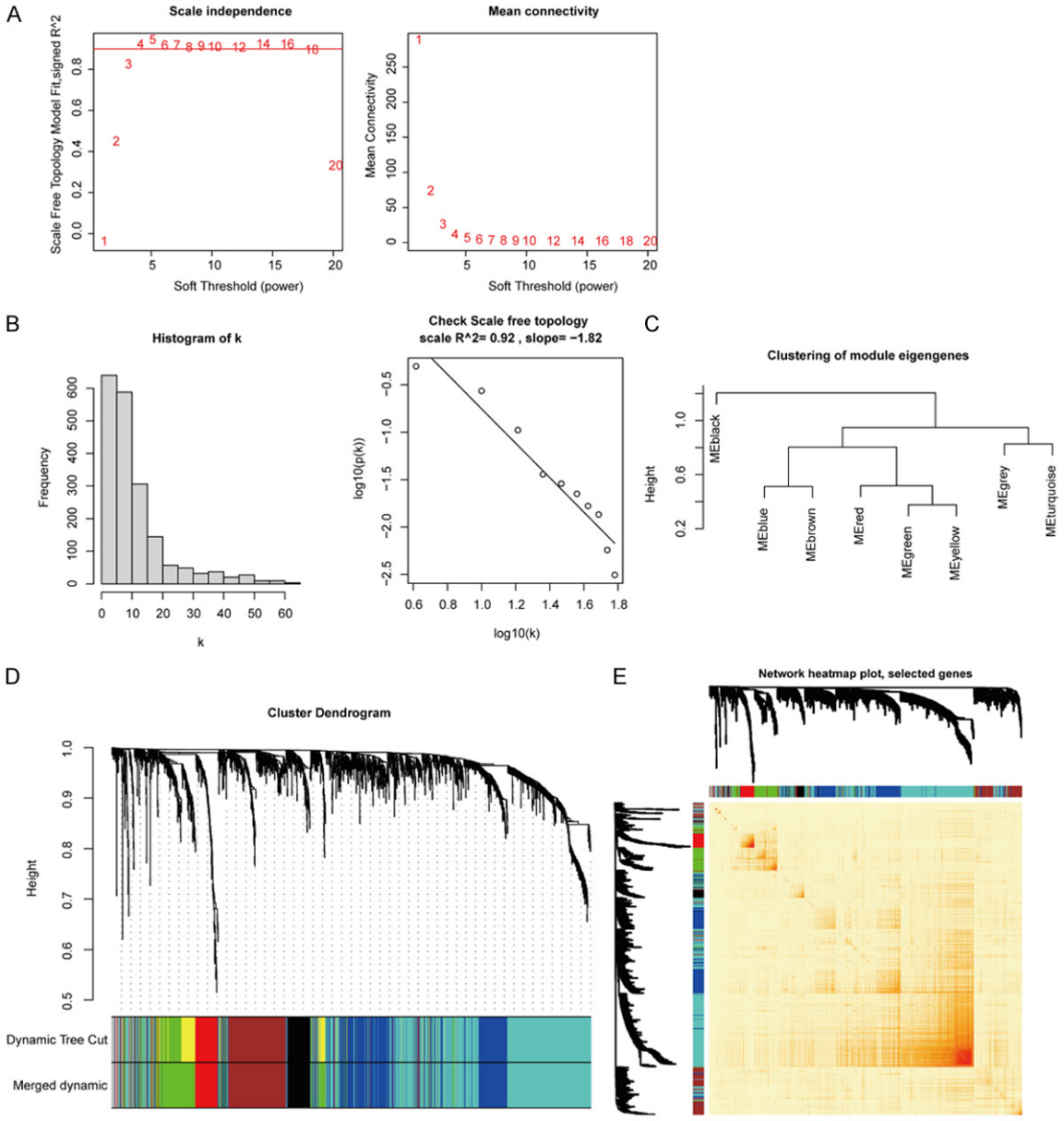


Figure S2. Construction of gene co-expression network. A. Analysis the scale-free fit index and mean connectivity of soft threshold of power. B. The distribution of k and liner model of fitting of R2 index. C. The dendrogram of gene modules. D. Gene dendrogram, the cut height was set as 0.4 to merge modules. E. Eigengene adjacency heatmap of 800 genes.

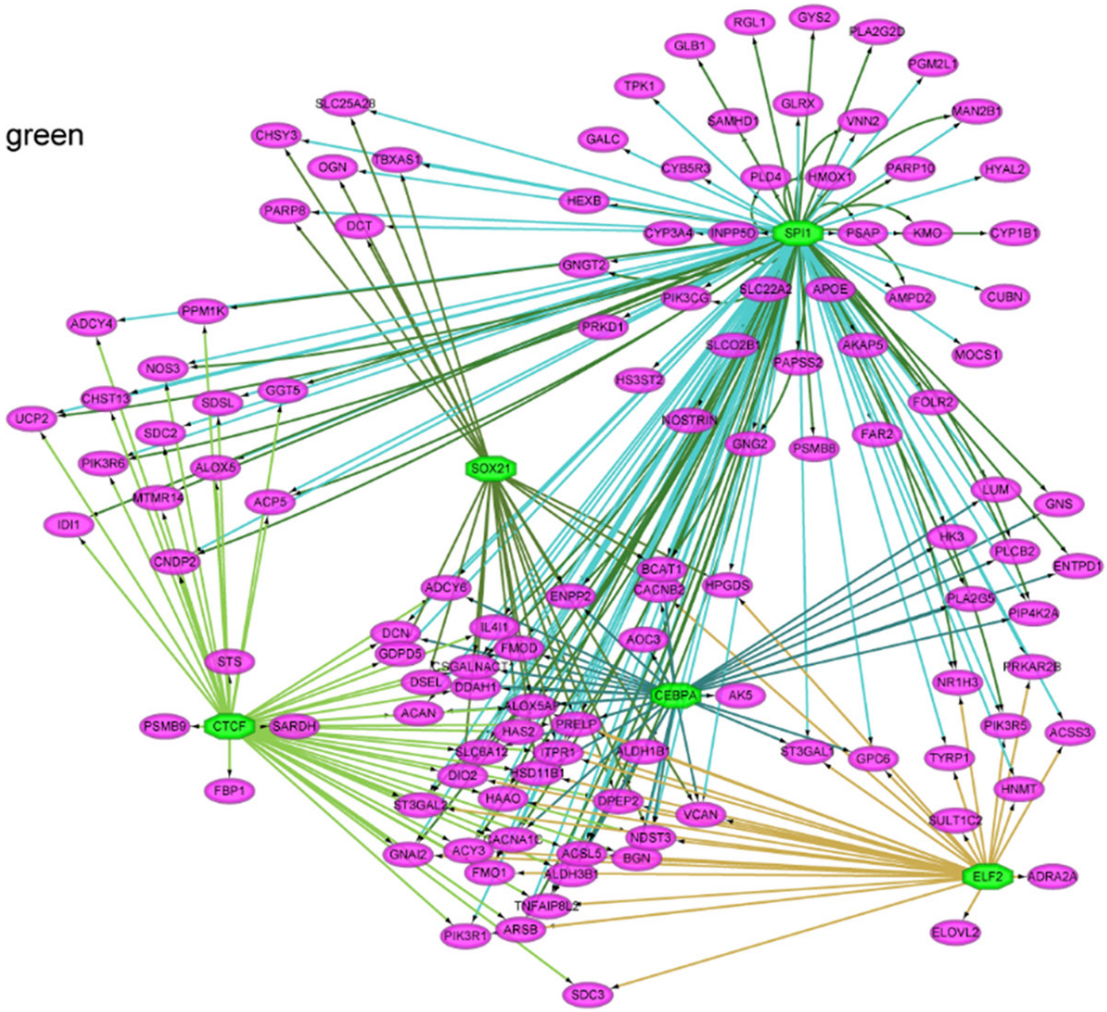

green
RGL1
GYS2
PLA2G2D
GLB1
PGM2L1
TPK1
SAMHD1
GLRX
VNN2
MAN2B1
SLC25A28
CHSY3
GALC
TBXAS1
OGN
CYB5R3
PLD4
HMOX1
PARP10
HYAL2
PARP8
DCT
HEXB
CYP3A4
INPP5D
SPI1
PSAP
KMO
CYP1B1
GNGT2
PIK3CG
SLC22A2
APOE
AMPD2
CUBN
ADCY4
PPM1K
PRKD1
SLCO2B1
AKAP5
MOCS1
NOS3
HS3ST2
PAPSS2
CHST13
SDSL
GGT5
NOSTRIN
FOLR2
UCP2
SDC2
GNG2
PSMB8
FAR2
PIK3R6
ALOX5
SOX21
MTMR14
ACP5
LUM
GNS
IDI1
HK3
CNDP2
PLCB2
ENTPD1
BCAT1
ADCY6
ENPP2
CACNB2
HPGDS
PLA2G5
PIP4K2A
IL4I1
DCN
FMOD
AOC3
GDPD5
PRKAR2B
STS
CHST14
NR1H3
DSEL
DDAH1
CEBPA
AK5
ALOX5AP
ACAN
PSMB9
CTCF
SARDH
HAS2
PRELP
PIK3R5
ACSS3
SLC8A12
ITPR1
ALDH1B1
ST3GAL1
GPC6
TYRP1
DIO2
HSD11B1
FBP1
HNMT
ST3GAL2
HAAO
DPEP2
VCAN
SULT1C2
CACNA1C
NDST3
GNAI2
ACY3
ACSL5
BGN
FMO1
ALDH3B1
ELF2
ADRA2A
TNFAIP8L2
PIK3R1
ARSB
ELOVL2
SDC3
brown
NQO1
G6PD
ABCC1
AKR1C2
AKR1B10
GSTA1
ALDH3A2
ABCC3
AKR1C3
TCF12
AKR1B15
AKR1B1
PTGR1
ME1
STAR
PGD
GSR
TXNRD1
GSTM4
PCCB
PARP1
ALDH3A1
AKR1C1
GPX2
GCLC
NFYB
ABCC2
CHPT1
GCLM
ALDH1A1
CACNA1D
ABCC5
ABHD4
POR
CYP2S1
EP300
UGT1A1
NAMPT
AKR1C4
CYB5A
FOXO1
CYP4F12
GSTA4
UGT8
MTMR7
EPHX1
HS3ST4
LRP6
CYP4F3
B3GALT2
CES1
ADH1C
MSRB3
TKT
UGDH
DDC
ENPP6
ACNA2D2
FADS1
CYP26A1
LIPE
NR4A1
CYP4F11
UGT1A7
UGT1A6
FADS2
PSMA8
ARNT2
NMNAT3

Figure S3. Transcription factors of genes in green and brown modules predicted by iRegulon.

Table S3. The 12 genes signature and coefficient

Gene	Coef
PTGR1	0.006695
FUT7	0.006255
ALDH1B1	0.307717
FMOD	-0.17457
MCCC1	0.084753
SLC2A3	0.276657
UGT1A10	0.032856
INPP5D	-0.1686
HK1	0.464965
PRKCA	0.245153
CSGALNACT1	-0.53195
NADSYN1	0.243603
NADSYN1	0.243603

Table S4. KEGG metabolic pathways

	Pathways	statistic	*p*.value
1	2-Oxocarboxylic acid metabolism - Homo sapiens (human)	5.110798	5.47E-07
2	Alanine, aspartate and glutamate metabolism - Homo sapiens (human)	4.916578	1.39E-06
3	alpha-Linolenic acid metabolism - Homo sapiens (human)	-6.06745	3.59E-09
4	Amino sugar and nucleotide sugar metabolism - Homo sapiens (human)	2.567315	0.010711
5	Arachidonic acid metabolism - Homo sapiens (human)	-5.05392	7.22E-07
6	Arginine and proline metabolism - Homo sapiens (human)	0.578506	0.563321
7	Arginine biosynthesis - Homo sapiens (human)	-0.23138	0.817162
8	Ascorbate and aldarate metabolism - Homo sapiens (human)	3.877937	0.000128
9	beta-Alanine metabolism - Homo sapiens (human)	-0.36899	0.712372
10	Biosynthesis of amino acids - Homo sapiens (human)	6.769541	5.97E-11
11	Biosynthesis of unsaturated fatty acids - Homo sapiens (human)	-1.04469	0.296937
12	Biotin metabolism - Homo sapiens (human)	-1.83431	0.067514
13	Butanoate metabolism - Homo sapiens (human)	0.351868	0.725168
14	Caffeine metabolism - Homo sapiens (human)	-1.66422	0.097038
15	Carbon metabolism - Homo sapiens (human)	4.715426	3.62E-06
16	Citrate cycle (TCA cycle) - Homo sapiens (human)	0.085758	0.931713
17	Cysteine and methionine metabolism - Homo sapiens (human)	2.376877	0.01804
18	D-Arginine and D-ornithine metabolism - Homo sapiens (human)	-1.85745	0.064143
19	D-Glutamine and D-glutamate metabolism - Homo sapiens (human)	1.691564	0.091706
20	Drug metabolism - cytochrome P450 - Homo sapiens (human)	1.332127	0.183755
21	Drug metabolism - other enzymes - Homo sapiens (human)	3.482113	0.000565
22	Ether lipid metabolism - Homo sapiens (human)	-5.70615	2.59E-08
23	Fatty acid biosynthesis - Homo sapiens (human)	-2.89668	0.004025
24	Fatty acid degradation - Homo sapiens (human)	0.338237	0.735408
25	Fatty acid elongation - Homo sapiens (human)	-2.79499	0.005496
26	Fatty acid metabolism - Homo sapiens (human)	-1.02522	0.306018
27	Folate biosynthesis - Homo sapiens (human)	-0.29924	0.764948
28	Fructose and mannose metabolism - Homo sapiens (human)	5.032899	8.03E-07

29	Galactose metabolism - Homo sapiens (human)	2.088841	0.037496
30	Glutathione metabolism - Homo sapiens (human)	3.517366	0.000498
31	Glycerolipid metabolism - Homo sapiens (human)	-1.42396	0.155408
32	Glycerophospholipid metabolism - Homo sapiens (human)	-1.26808	0.20567
33	Glycine, serine and threonine metabolism - Homo sapiens (human)	1.570814	0.117192
34	Glycolysis/Gluconeogenesis - Homo sapiens (human)	3.694357	0.000258
35	Glycosaminoglycan biosynthesis - chondroitin sulfate/dermatan sulfate - Homo sapiens (human)	1.517116	0.130202
36	Glycosaminoglycan biosynthesis - heparan sulfate/heparin - Homo sapiens (human)	1.350717	0.177728
37	Glycosaminoglycan biosynthesis - keratan sulfate - Homo sapiens (human)	1.484128	0.138735
38	Glycosaminoglycan degradation - Homo sapiens (human)	-0.71578	0.474636
39	Glycosphingolipid biosynthesis - ganglio series - Homo sapiens (human)	2.67079	0.007947
40	Glycosphingolipid biosynthesis - globo and isoglobo series - Homo sapiens (human)	-0.20665	0.836411
41	Glycosphingolipid biosynthesis - lacto and neolacto series - Homo sapiens (human)	-1.46715	0.143298
42	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis - Homo sapiens (human)	1.936882	0.053618
43	Glyoxylate and dicarboxylate metabolism - Homo sapiens (human)	2.579394	0.010345
44	Histidine metabolism - Homo sapiens (human)	-3.73655	0.00022
45	Inositol phosphate metabolism - Homo sapiens (human)	-2.509	0.012593
46	Linoleic acid metabolism - Homo sapiens (human)	-5.93368	7.59E-09
47	Lipoic acid metabolism - Homo sapiens (human)	1.082252	0.279937
48	Lysine degradation - Homo sapiens (human)	1.825642	0.068813
49	Mannose type O-glycan biosynthesis - Homo sapiens (human)	4.115939	4.90E-05
50	Metabolic pathways - Homo sapiens (human)	0.329685	0.741849
51	Metabolism of xenobiotics by cytochrome P450 - Homo sapiens (human)	2.643925	0.008599
52	Mucin type O-glycan biosynthesis - Homo sapiens (human)	-0.23394	0.81518
53	N-Glycan biosynthesis - Homo sapiens (human)	0.638976	0.523286
54	Neomycin, kanamycin and gentamicin biosynthesis - Homo sapiens (human)	4.040443	6.67E-05
55	Nicotinate and nicotinamide metabolism - Homo sapiens (human)	0.820551	0.412499
56	Nitrogen metabolism - Homo sapiens (human)	0.442635	0.658322
57	One carbon pool by folate - Homo sapiens (human)	2.410379	0.016486
58	Other glycan degradation - Homo sapiens (human)	0.884766	0.376937
59	Other types of O-glycan biosynthesis - Homo sapiens (human)	3.175186	0.00164
60	Oxidative phosphorylation - Homo sapiens (human)	0.190115	0.849337
61	Pantothenate and CoA biosynthesis - Homo sapiens (human)	-5.84823	1.23E-08
62	Pentose and glucuronate interconversions - Homo sapiens (human)	2.829573	0.004956
63	Pentose phosphate pathway - Homo sapiens (human)	4.044598	6.56E-05
64	Phenylalanine metabolism - Homo sapiens (human)	-0.29081	0.771388
65	Phenylalanine, tyrosine and tryptophan biosynthesis - Homo sapiens (human)	0.122598	0.902501
66	Phosphonate and phosphinate metabolism - Homo sapiens (human)	3.727206	0.000228
67	Porphyrin and chlorophyll metabolism - Homo sapiens (human)	3.366537	0.000854
68	Primary bile acid biosynthesis - Homo sapiens (human)	-3.37802	0.000818
69	Propanoate metabolism - Homo sapiens (human)	1.604057	0.1097
70	Purine metabolism - Homo sapiens (human)	0.847311	0.39745
71	Pyrimidine metabolism - Homo sapiens (human)	-1.04799	0.29542
72	Pyruvate metabolism - Homo sapiens (human)	2.582007	0.010257
73	Retinol metabolism - Homo sapiens (human)	-0.59925	0.549423
74	Riboflavin metabolism - Homo sapiens (human)	-0.95403	0.340774
75	Selenocompound metabolism - Homo sapiens (human)	3.339217	0.000938
76	Sphingolipid metabolism - Homo sapiens (human)	-3.30776	0.00105

77	Starch and sucrose metabolism - Homo sapiens (human)	0.724978	0.468983
78	Steroid biosynthesis - Homo sapiens (human)	0.555455	0.578964
79	Steroid hormone biosynthesis - Homo sapiens (human)	1.646613	0.100607
80	Sulfur metabolism - Homo sapiens (human)	-1.58328	0.114323
81	Synthesis and degradation of ketone bodies - Homo sapiens (human)	-0.49565	0.620471
82	Taurine and hypotaurine metabolism - Homo sapiens (human)	-1.06229	0.288888
83	Terpenoid backbone biosynthesis - Homo sapiens (human)	-1.99851	0.046487
84	Thiamine metabolism - Homo sapiens (human)	0.457475	0.647634
85	Tryptophan metabolism - Homo sapiens (human)	-1.53283	0.126328
86	Tyrosine metabolism - Homo sapiens (human)	0.226162	0.82122
87	Ubiquinone and other terpenoid-quinone biosynthesis - Homo sapiens (human)	2.923608	0.003705
88	Valine, leucine and isoleucine biosynthesis - Homo sapiens (human)	3.22197	0.001401
89	Valine, leucine and isoleucine degradation - Homo sapiens (human)	2.326173	0.020629
90	Vitamin B6 metabolism - Homo sapiens (human)	0.122998	0.902184

Table S5. Immune related process and ICB-response signatures

signature	genes
APC co inhibition	C10orf54, CD274, LGALS9, PDCD1LG2, PVRL3, CD40, CD58, CD70, ICOSLG, SLAMF1, TNFSF14, TNFSF15, TNFSF18, TNFSF4, TNFSF8, TNFSF9
CCR	CCL16, TPO, TGFBR2, CXCL2, CCL14, TGFBR3, IL11RA, CCL11, IL4I1, IL33, CXCL12, CXCL10, BMPER, BMP8A, CXCL11, IL21R, IL17B, TNFRSF9, ILF2, CX3CR1, CCR8, TNFSF12, CSF3, TNFSF4, BMP3, CX3CL1, BMP5, CXCR2, TNFRSF10D, BMP2, CXCL14, CCL28, CXCL3, BMP6, CCL21, CXCL9, CCL23, IL6, TNFRSF18, IL17RD, IL17D, IL27, CCL7, IL1R1, CXCR4, CXCR2P1, TGFB1I1, IFNGR1, IL9R, IL1RAPL1, IL11, CSF1, IL20RA, IL25, TNFRSF4, IL18, ILF3, CCL20, TNFRSF12A, IL6ST, CXCL13, IL12B, TNFRSF8, IL6R, BMPR2, IFNE, IL1RAPL2, IL3RA, BMP4, CCL24, TNFSF13B, CCR4, IL2RA, IL32, TNFRSF10C, IL22RA1, BMPR1A, CXCR5, CXCR3, IFNA8, IL17REL, IFNB1, IFNAR1, TNFRSF1B, CCL17, IFNL1, IL16, IL1RL1, ILK, CCL25, ILDR2, CXCR1, IL36RN, IL34, TGFB1, IFNG, IL19, ILKAP, BMP2K, CCR10, ILDR1, EPO, CCR7, IL17C, IL23A, CCR5, IL7, EPOR, CCL13, IL2RG, IL31RA, TNFAIP6, IFNL2, BMP1, IL12RB1, TNFAIP8, IL4R, TNFRSF6B, TNFAIP8L1, TNFRSF10B, IFNL3, CCL5, CXCL6, CXCL1, CCR3, TNFSF11, CSF1R, IL21, IL1RAP, IL12RB2, CCL1, IL17RA, CCR1, IL1RN, TNFRSF11B, TNFRSF14, IL13, IL2RB, BMP8B, CCL2, IL24, IL18RAP, TGFBI, TNFSF10, TNFRSF11A, CXCL5, IL5RA, TNFSF9, IL1RL2, TNFRSF13C, IL36G, IL15RA, TNFRSF21, CXCL8, IL22RA2, TNFAIP8L2, IL18R1, IFNLR1, CXCR6, CCL3L3, TNFRSF1A, IL17RE, IFNGR2, IL17RC, TNFAIP8L3, ILVBL, TGFBRAP1, CCL4L1, CSF2RA, CCRN4L, CCL26, TNFAIP1, CCRL2, IFNA10, TNFRSF17, IFNA13, IL20, IL18BP, CCL3L1, TNFSF12-TNFSF13, IL5, IL23R, IL26, TNF, TGFA, CSF2, IL1F10, CXCL17, TNFSF13, IFNA4, IL37, IL12A, IL7R, IFNA1, IL1A, IL4, IL2, CCL22, CSF3R, IL10, IFNK, TGFB2, IL1R2, IL1B, IL17F, IL27RA, IL15, TNFSF8, IL36B, XCL1, CXCL16, TNFRSF19, IL3, CCL3, IFNA2, BMPR1B, IFNA21, TNFSF18, CCL8, IL17RB, TNFRSF25, IL22, IL10RB, IFNAR2, CCL18, IFNA16, CSF2RB, IL36A, TNFAIP3, IL13RA2, IL13RA1, CCR9, TNFRSF10A, IFNA7, IFNW1, XCL2, TNFSF14, CCR2, BMP15, BMP10, CCL15-CCL14, TGFBR1, IFNA5, BMP7, IFNA14, IL20RB, IL10RA, IFNA17, CCR6, TGFB3, CCL15, CCL4, CCL27, TNFRSF13B, TNFAIP2, IL31, IL17A, TNFSF15, CCL19, IFNA6, IL9
checkpoint	IDO1, LAG3, CTLA4, TNFRSF9, ICOS, CD80, PDCD1LG2, TIGIT, CD70, TNFSF9, ICOSLG, KIR3DL1, CD86, PDCD1, LAIR1, TNFRSF8, TNFSF15, TNFRSF14, IDO2, CD276, CD40, TNFRSF4, TNFSF14, HHLA2, CD244, CD274, HAVCR2, CD27, BTLA, LGALS9, TMIGD2, CD28, CD48, TNFRSF25, CD40LG, ADORA2A, VTCN1, CD160, CD44, TNFSF18, TNFRSF18, BTNL2, C10orf54, CD200R1, TNFSF4, CD200, NRP1
Cytolytic activity	PRF1, GZMB, GZMA
HLA	HLA-E, HLA-DPB2, HLA-C, HLA-J, HLA-DQB1, HLA-DQB2, HLA-DQA2, HLA-DQA1, HLA-A, HLA-DMA, HLA-DOB, HLA-DRB1, HLA-H, HLA-B, HLA-DRB5, HLA-DOA, HLA-DPB1, HLA-DRA, HLA-DRB6, HLA-L, HLA-F, HLA-G, HLA-DMB, HLA-DPA1
Inflammation-promoting	CCL5, CD19, CD8B, CXCL10, CXCL13, CXCL9, GNLY, GZMB, IFNG, IL12A, IL12B, IRF1, PRF1, STAT1, TBX21
Parainflammation	CXCL10, PLAT, CCND1, LGMN, PLAUR, AIM2, MMP7, ICAM1, MX2, CXCL9, ANXA1, TLR2, PLA2G2D, ITGA2, MX1, HMOX1, CD276, TIRAP, IL33, PTGES, TNFRSF12A, SCARB1, CD14, BLNK, IFIT3, RETNLB, IFIT2, ISG15, OAS2, REL, OAS3, CD44, PPARG, BST2, OAS1, NOX1, PLA2G2A, IFIT1, IFITM3, IL1RN

T cell co-inhibition	BTLA, C10orf54, CD160, CD244, CD274, CTLA4, HAVCR2, LAG3, LAIR1, TIGIT
T cell co-stimulation	CD2, CD226, CD27, CD28, CD40LG, ICOS, SLAMF1, TNFRSF18, TNFRSF25, TNFRSF4, TNFRSF8, TNFRSF9, TNFSF14
Type I IFN Reponse	DDX4, IFIT1, IFIT2, IFIT3, IRF7, ISG20, MX1, MX2, RSAD2, TNFSF10
Type II IFN Reponse	GPR146, SELP, AHR
ICB-response	AKR1C4, KCNA1, CREB3L3, NEU4, PIGR, ACOD1, CPN2, CCL16, HGD, DEFA1, TAT, PDCD1, GLYATL1, MT1H, COLEC10, THRSP, HS3ST3B1, IL24, IFNG, SMCO2, TMEM155, DBH
CD8 T effector	CD8A, CXCL10, CXCL9, GZMA, GZMB, IFNG, PRF1, TBX21
DNA damage repair	ALKBH2, ALKBH3, APEX1, APEX2, APLF, ATM, ATR, ATRIP, BLM, BRCA1, BRCA2, BRIP1, CCNH, CDK7, CETN2, CHAF1A, CHEK1, CHEK2, CLK2, DCLRE1C, DDB1, DDB2, DUT, ENDOV, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, ERCC6, ERCC8, FAN1, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCL, FANCM, GTF2H1, GTF2H2, GTF2H3, GTF2H4, GTF2H5, H2AFX, HLTF, HUS1, LIG1, LIG3, LIG4, MBD4, MDC1, MGMT, MLH1, MLH3, MMS19, MNAT1, MPG, MSH2, MSH3, MSH4, MSH5, MSH6, MUTYH, NEIL1, NEIL2, NEIL3, NHEJ1, NTHL1, NUDT1, OGG1, PALB2, PARP1, PARP2, PARP3, PCNA, PER1, PMS1, PMS2, PNKP, POLB, POLD1, POLE, POLG, POLH, POLL, POLM, POLQ, PRKDC, RAD1, RAD17, RAD18, RAD23A, RAD23B, RAD51C, RAD9A, RECQL4, RECQL5, RIF1, RNF168, RNF4, RNF8, RPA1, RPA2, RPA3, RPA4, RRM2B, SETMAR, SHPRH, SMUG1, TDP1, TDP2, TOPBP1, TP53, TREX1, UBE2A, UBE2B, UBE2N, UBE2V2, UNG, UVSSA, WRN, XAB2, XPA, XPC, XRCC1, XRCC4, XRCC5, XRCC6
Pan-F-TBRS	ACTA2, ACTG2, ADAM12, ADAM19, CNN1, COL4A1, CTGF, CTPS1, FAM101B, FSTL3, HSPB1, IGFBP3, PXDC1, SEMA7A, SH3PXD2A, TAGLN, TGFBI, TNS1, TPM1
Antigen processing machinery	B2M, HLA-A, HLA-B, HLA-C, TAP1, TAP2
Immune checkpoint	CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG2, TIGIT
EMT (1)	CLDN3, CLDN7, CLDN4, CDH1, VIM, TWIST1, ZEB1, ZEB2
FGFR3-related genees	FGFR3, TP63, WNT7B
KEGG discovered histones	HIST1H2AG, HIST1H2AI, HIST1H2BL, HIST2H2BF
Angiogenesis	CDH5, SOX17, SOX18, TEK
Fanconi anemia	APITD1, ATR, ATRIP, BLM, BRCA1, BRCA2, BRIP1, C17orf70, C19orf40, EME1, EME2, ERCC1, ERCC4, FAN1, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, HES1, MLH1, MUS81, PALB2, PMS2, POLH, POLI, POLK, POLN, RAD51, RAD51C, REV1, REV3L, RMI1, RMI2, RPA1, RPA2, RPA3, RPA4, SLX4, STRA13, TELO2, TOP3A, TOP3B, UBE2T, USP1, WDR48
Cell cycle	ABL1, ANAPC1, ANAPC10, ANAPC11, ANAPC13, ANAPC2, ANAPC4, ANAPC5, ANAPC7, ATM, ATR, BUB1, BUB1B, BUB3, CCNA1, CCNA2, CCNB1, CCNB2, CCNB3, CCND1, CCND2, CCND3, CCNE1, CCNE2, CCNH, CDC14A, CDC14B, CDC16, CDC20, CDC23, CDC25A, CDC25B, CDC25C, CDC26, CDC27, CDC45, CDC6, CDC7, CDK1, CDK2, CDK4, CDK6, CDK7, CDKN1A, CDKN1B, CDKN1C, CDKN2A, CDKN2B, CDKN2C, CDKN2D, CHEK1, CHEK2, CREBBP, CUL1, DBF4, E2F1, E2F2, E2F3, E2F4, E2F5, EP300, ESPL1, FZR1, GADD45A, GADD45B, GADD45G, GSK3B, HDAC1, HDAC2, MAD1L1, MAD2L1, MAD2L2, MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MDM2, MYC, ORC1, ORC2, ORC3, ORC4, ORC5, ORC6, PCNA, PKMYT1, PLK1, PRKDC, PTTG1, PTTG2, RAD21, RB1, RBL1, RBL2, RBX1, SFN, SKP1, SKP2, SMAD2, SMAD3, SMAD4, SMC1A, SMC1B, SMC3, STAG1, STAG2, TFDP1, TFDP2, TGFB1, TGFB2, TGFB3, TP53, TTK, WEE1, YWHAB, YWHAE, YWHAG, YWHAH, YWHAQ, YWHAZ, ZBTB17
DNA replication	DNA2, FEN1, LIG1, MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, PCNA, POLA1, POLA2, POLD1, POLD2, POLD3, POLD4, POLE, POLE2, POLE3, POLE4, PRIM1, PRIM2, RFC1, RFC2, RFC3, RFC4, RFC5, RNASEH1, RNASEH2A, RNASEH2B, RNASEH2C, RPA1, RPA2, RPA3, RPA4, SSBP1
Nucleotide excision repair	CCNH, CDK7, CETN2, CUL4A, CUL4B, DDB1, DDB2, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, ERCC6, ERCC8, GTF2H1, GTF2H2, GTF2H3, GTF2H4, GTF2H5, LIG1, MNAT1, PCNA, POLD1, POLD2, POLD3, POLD4, POLE, POLE2, POLE3, POLE4, RAD23A, RAD23B, RBX1, RFC1, RFC2, RFC3, RFC4, RFC5, RPA1, RPA2, RPA3, RPA4, XPA, XPC
Homologous recombination	BLM, BRCA2, EME1, MRE11A, MUS81, NBN, POLD1, POLD2, POLD3, POLD4, RAD50, RAD51, RAD51B, RAD51C, RAD51D, RAD52, RAD54B, RAD54L, RPA1, RPA2, RPA3, RPA4, SHFM1, SSBP1, TOP3A, TOP3B, XRCC2, XRCC3
Mismatch repair	EXO1, LIG1, MLH1, MLH3, MSH2, MSH3, MSH6, PCNA, PMS2, POLD1, POLD2, POLD3, POLD4, RFC1, RFC2, RFC3, RFC4, RFC5, RPA1, RPA2, RPA3, RPA4, SSBP1
EMT2	AXL, FAP, LOXL2, ROR2, TAGLN, TWIST2, WNT5A
EMT3	FOXF1, GATA6, SOX9, TWIST1, ZEB1, ZEB2
WNT target	EFNB3, MYC, TCF12, VEGFA
Cell cycle regulators	ATM, CCND1, CCNE1, CDKN1A, CDKN2A, E2F3, FBXW7, MDM2, RB1, TP53

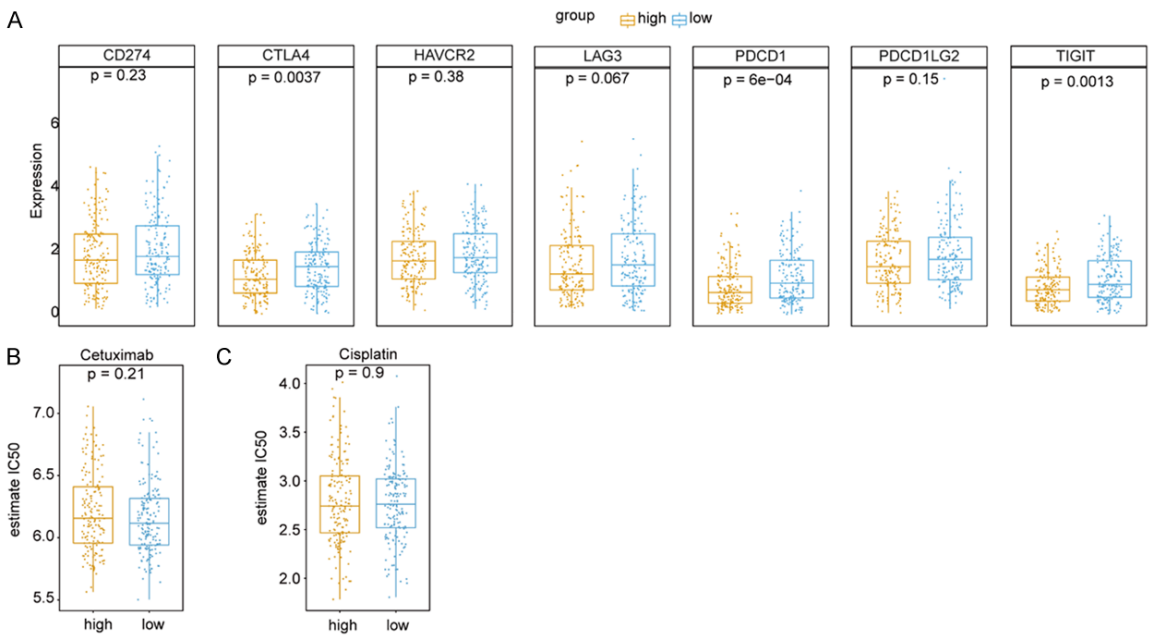


Figure S4. Checkpoints and IC50 of drugs. A. The 7 checkpoints expression in different risk group. B, C. IC50 of Cetuximab and Cisplatin estimated by "pRRophetic" package.

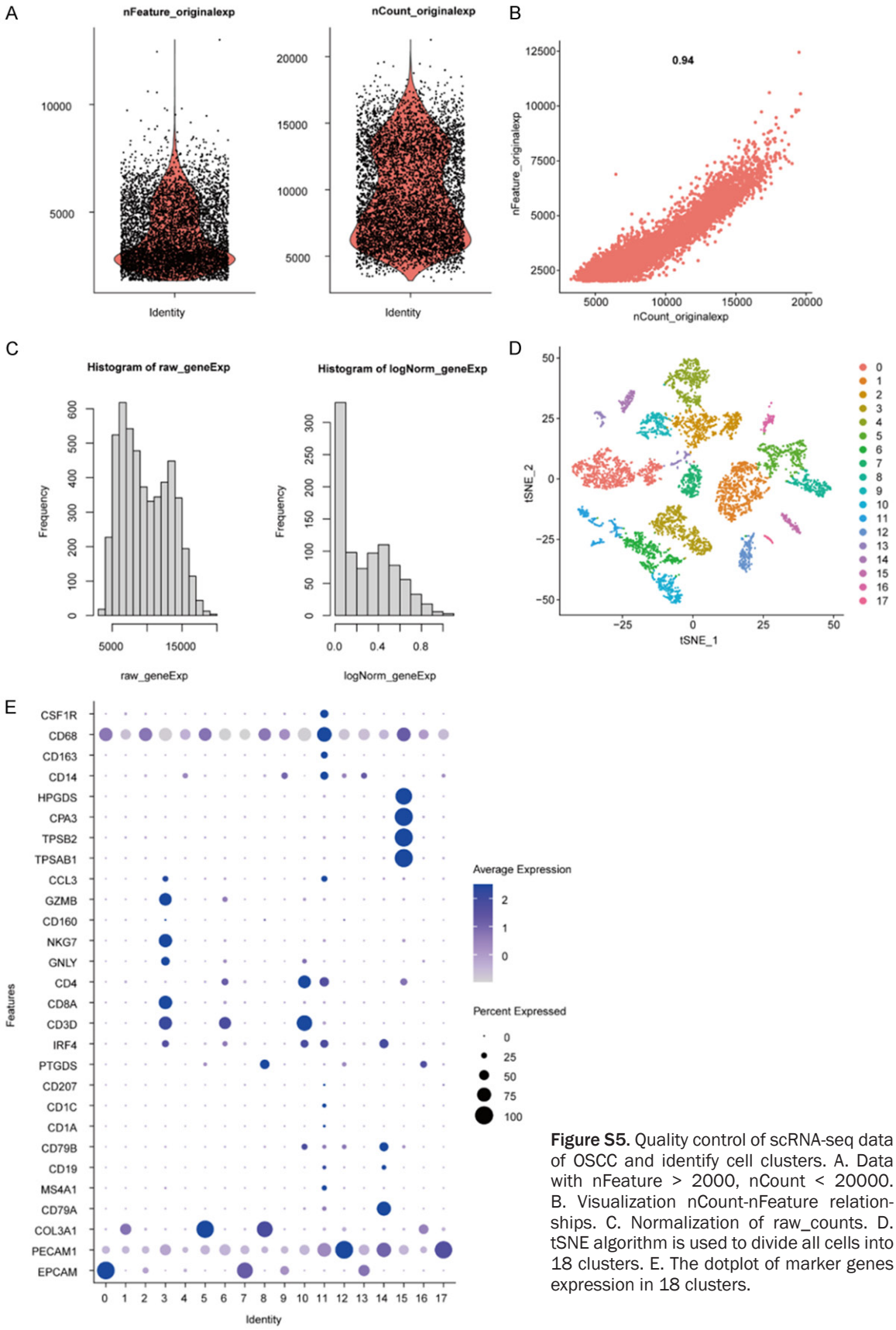


Figure S5. Quality control of scRNA-seq data of OSCC and identify cell clusters. A. Data with nFeature > 2000, nCount < 20000. B. Visualization nCount-nFeature relationships. C. Normalization of raw_counts. D. tSNE algorithm is used to divide all cells into 18 clusters. E. The dotplot of marker genes expression in 18 clusters.

Table S6. Markers used to identify cell types in 18 cell clusters

Cells	Maker genes
Tumor cells	EPCAM
Endothelial cells	PECAM1
Fibroblasts	COL3A1
B cells	CD79A, MS4A1, IGHM, CD19, CD79B
Dendritic cells	CD1A, CD1C, CD207, PTGDS, IRF4
T cell	CD3D, CD8A, CD4
NK cells	GNLY, NKG7, CD160, GZMB, CCL3
Mast cells	TPSAB1, TPSB2, CPA3, HPGDS
Macrophages	CD14, CD163, CD68, CSF1R

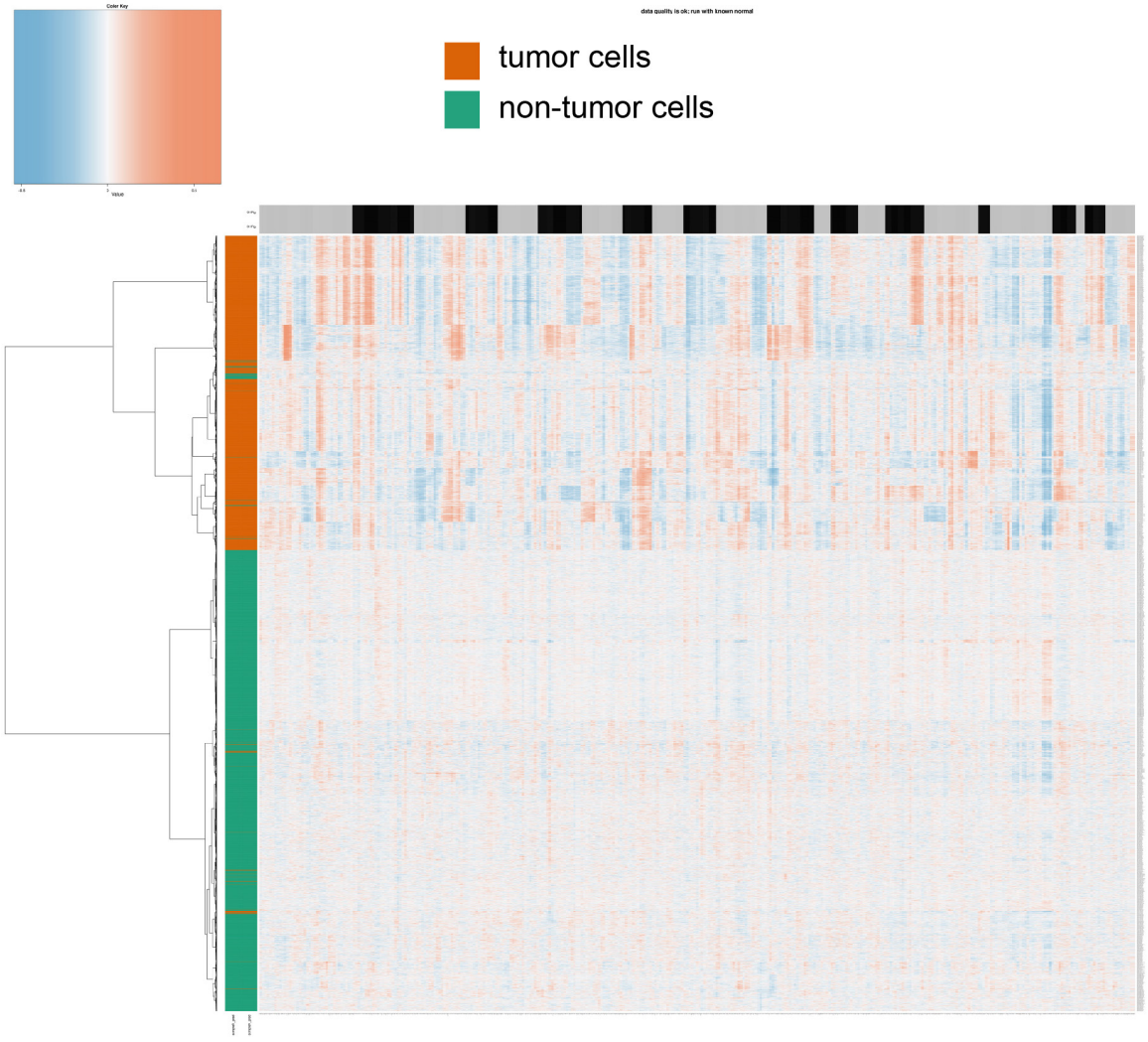


Figure S6. The estimated CNVs of 5746 cells analyzed by CopyKAT.

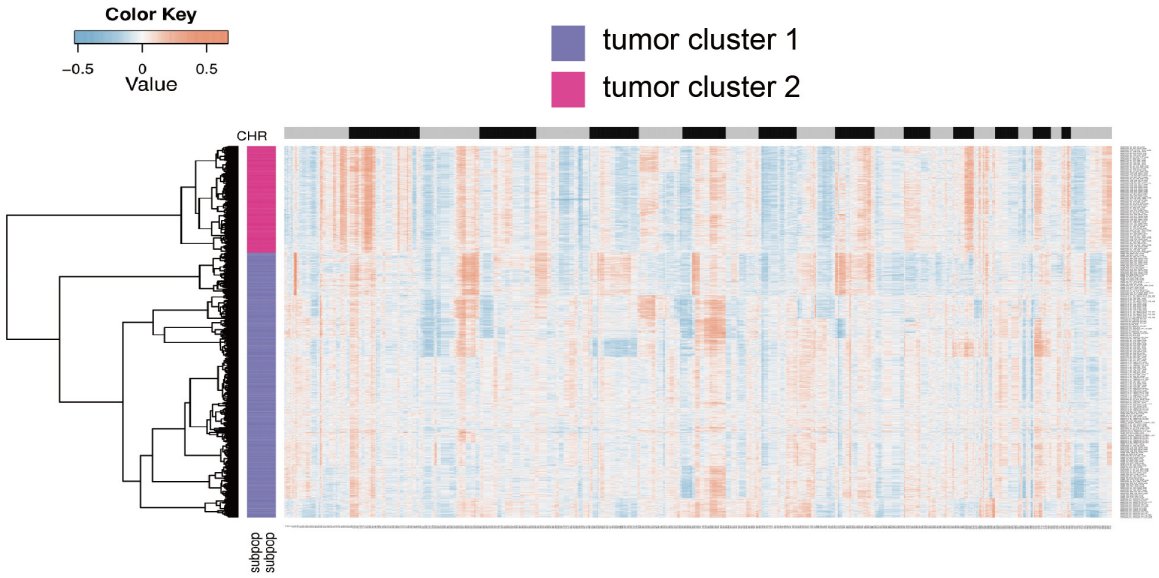


Figure S7. The estimated CNVs of tumor cells analyzed by CopyKAT.