

Original Article

A new risk model for CSTA, FAM83A, and MYCT1 predicts poor prognosis and is related to immune infiltration in lung squamous cell carcinoma

Yu Ding^{1*}, Ting-Ting Bian^{2*}, Qian-Yun Li^{3*}, Jin-Rong He¹, Qiang Guo², Chuang-Yan Wu², Shan-Shan Chen¹

¹Key Laboratory for Molecular Diagnosis of Hubei Province, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430014, Hubei, China; ²Department of Thoracic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, Hubei, China; ³The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu 310030, Zhejiang, China.
*Equal contributors.

Received June 6, 2022; Accepted October 27, 2022; Epub November 15, 2022; Published November 30, 2022

Abstract: Objectives: To create a prognostic model based on differentially expressed genes (DEGs) in early lung squamous cell carcinoma (LUSC) and characterize the relationship between risk scores and tumor immune infiltration. Methods: We identified DEGs in normal and tumor tissues that overlapped between LUSC-related data sets from the Gene Expression Omnibus and the Cancer Genome Atlas and evaluated their roles in the diagnosis and prognosis of LUSC by Kaplan-Meier survival analysis, receiver operating characteristic (ROC) analysis, meta-analysis and nomogram analysis. We then constructed a risk model based on Cox regression analysis and the Akaike information criterion and identified the relationship between LUSC risk scores and immune infiltration. Results: Sixty-two overlapping DEGs were involved with keratinocyte differentiation, epidermal cell differentiation, neutrophil migration, granulocyte chemotaxis, granulocyte migration, leukocyte aggregation, and positive regulation of nuclear factor- κ B (NF- κ B) activity. Overexpression of family with sequence similarity 83 member A (*FAM83A*) and MYC target 1 (*MYCT1*), kallikrein related peptidase 8 (*KLK8*), and downregulation of ADP ribosylation factor like GTPase 14 (*ARL14*), caspase recruitment domain family member 14 (*CARD14*), cystatin A (*CSTA*), dickkopf WNT signaling pathway inhibitor 4 (*DKK4*), desmoglein 3 (*DSG3*), and keratin 6B (*KRT6B*) were associated with a poor prognosis in LUSC and had significant value for LUSC diagnosis. The expression of *CSTA*, *FAM83A*, and *MYCT1* and high-risk scores were independent risk factors for a poor prognosis in LUSC. A risk nomogram revealed that risk scores could predict the prognosis of LUSC. The risk score was associated with neutrophils, naive B cells, helper follicular T cells, and activated dendritic cells. Conclusions: The expression levels of *CSTA*, *FAM83A*, and *MYCT1* are related to the diagnosis and prognosis of LUSC and may have potential as therapeutic targets in LUSC. A risk model and nomogram based on *CSTA*, *FAM83A*, and *MYCT1* can predict the prognosis of LUSC.

Keywords: Differentially expressed genes, receiver operating characteristic, overall survival, lung squamous cell carcinoma, risk model

Introduction

Lung squamous cell carcinoma (LUSC) is a common subtype of non-small cell lung carcinoma, a highly prevalent disease that causes substantial morbidity and mortality [1-5]. In recent years, the prognosis of cancer has improved significantly with advances in treatment methods. Patients with early-stage LUSC can achieve long-term survival with surgical treatment; however, most LUSC is diagnosed at advanced stages with the tumor located in the

hilar of the lungs and thus cannot be treated surgically. Furthermore, chemotherapy is ineffective in patients with advanced LUSC, resulting in a poor prognosis [5, 6]. Targeted therapy has achieved good results in patients with lung adenocarcinoma (LAC). Therefore, it is important to develop novel molecular targets in patients with LUSC for early diagnosis to improve their prognosis and quality of life.

Elevated levels of abnormal gene expression, microRNAs (miRNAs), long noncoding RNAs

A risk model for lung squamous cell carcinoma

(lncRNAs), and other factors are involved in the occurrence and development of LUSC [3, 7-11]. For example, lncRNA nicotinamide nucleotide transhydrogenase antisense RNA 1 (NNT-AS1) and forkhead box protein M1 (FOXM1) are frequently up-regulated, while miR-22 is frequently down-regulated in LUSC tissues and cells. Furthermore, NNT-AS1 deletion was found to inhibit LUSC cell migration and invasion, causing apoptosis and inhibiting carcinogenesis by controlling the miR-22/FOXM1 signaling axis [9]. In another study, LUSC cells showed reduced expression of lncRNA STAR Related Lipid Transfer Domain Containing 13 antisense RN (STARD13-AS), and STARD13-AS overexpression could delay the growth and invasion of LUSC cells by controlling the miR-1248/complement C3 (C3A) signaling axis [10]. In contrast, lncRNA RP11-116G8.5 was overexpressed in LUSC cells, and its inhibition could inhibit LUSC cell proliferation, migration, and invasion while speeding up apoptosis. RP11-116G8.5 regulates the expression of PHD finger protein 12 (PHF12) and forkhead box P4 (FOXP4) by acting as a sponge for miR-3150b-3p and miR-6870-5p. However, overexpression of *PHF12* and *FOXP4* in LUSC cells was shown to reverse the inhibitory effect of RP11-116G8.5 knockdown in cancer cells [11].

Risk models and nomograms are used to assess cancer prognosis [2, 12-14]. Here, we screened differentially expressed genes (DEGs) in normal lung tissues and LUSC that were present in Mascaux et al. data in the Gene Expression Omnibus (GEO) [15] and the Cancer Genome Atlas (TCGA) databases. We then evaluated the clinical values of DEGs that are critical in the progression of LUSC using the Kaplan-Meier (K-M) survival analysis, the receiver operating characteristic (ROC) analysis, and the risk model construction. These investigations provided novel diagnostic target molecules and prognostic biomarkers for better management of patients with LUSC. We also constructed a risk model to predict the prognosis of LUSC.

Materials and methods

LUSC gene expression and clinical data

We retrieved and downloaded the Series Matrix File (s) in the GSE33479 dataset from the GEO database, which was generated using the GPL6480: Agilent-014850 Whole Human

Genome Microarray 4 × 44K G4112F platform and includes gene expression data from 27 normal tissues, 14 tissues with squamous metaplasia, 13 carcinoma *in situ* tissues and 14 LUSC tissues. In addition, we obtained Fragments per Kilobase Million (FPKM)-type gene expression data and clinical data from the TCGA database. After removing entries with missing values or incomplete clinical information, our study sample included 49 normal tissues, 502 LUSC tissues, and clinical data from 490 patients.

Overlapping DEGs of the GSE33479 data set

We use the limma package to identify genes that were differentially expressed during progression from normal lung tissues to squamous metaplasia tissues, carcinomas *in situ*, or LUSC, with a fold change of 1 and an adjusted *P* value < 0.05 as the screening criteria. The DEGs that overlapped the three groups of cancer tissues were visualized using a Venn diagram and a heat map.

The biological functions, signaling pathways, and protein-protein interaction network of the DEGs

Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) are commonly used to analyze biological functions and signaling mechanisms involving multiple genes [2, 16]. The GO type includes three types of information: biological processes, cellular components, and molecular functions. We used GO and KEGG analyses to identify the enriched biological processes and signaling mechanisms of overlapping DEGs, with an adjusted *P* value < 0.05 as the screening criterion. Furthermore, we constructed a protein-protein interaction (PPI) network between the overlapping DEGs and DEGs in the STRING database and visualized it using Cytoscape software (version 3.8.2). The critical DEGs in the PPI network were visualized using the CytoHubba plug-in.

Identification of DEGs in LUSC tissues

We obtained expression data of the identified DEGs in 49 normal lungs and 502 LUSC tissues from TCGA. The expression levels of the DEGs in unpaired LUSC tissues were investigated using the Wilcoxon rank sum test, with *P* < 0.05 as the screening criterion. In addition to the

A risk model for lung squamous cell carcinoma

Table 1. PCR primers used in the study

Gene	Forward primer	Reversed primer
CSTA	5'-AATGATACCTGGAGGCTTATCT-3'	5'-TTTATTATCACCTGCTCGTACC-3'
FAM83A	5'-CCCATCTCAGTCACTGGCATT-3'	5'-CCGCCAACATCTCCTTGTTTC-3'
MYCT1	5'-GCCAGAAAACCTTTGGGAGGA-3'	5'-ATCCAGTTCTGTTGAGGCCG-3'

Note: PCR, Polymerase Chain Reaction; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; MYCT1, MYC Target 1.

unpaired samples, 49 pairs of matched normal lung and LUSC tissues were matched. The expression levels of DEGs in unpaired LUSC tissues were compared with those of paired LUSC tissues, with $P < 0.05$ as the screening criterion.

K-M survival analysis

We paired and merged the DEG expression data with the survival data of patients with LUSC in TCGA. We then performed a K-M survival analysis to investigate the impact of high or low DEG expression on the prognosis of patients with LUSC [2, 16], using $P < 0.05$ as the filter criterion.

Determination of the diagnostic value of LUSC prognostic genes

ROC analysis is typically used to determine the diagnostic value of gene expression levels in cancer, with an area under the curve (AUC) between 0.5 and 1.0 as the evaluation standard. The higher the AUC, the greater the diagnostic value. We used ROC analysis to investigate the importance of expression of ADP ribosylation factor like GTPase 14 (*ARL14*), caspase recruitment domain family member 14 (*CARD14*), cystatin A (*CSTA*), dickkopf WNT signaling pathway inhibitor 4 (*DKK4*), desmoglein 3 (*DSG3*), family with sequence similarity 83 member A (*FAM83A*), kallikrein related peptidase 8 (*KLK8*), keratin 6B (*KRT6B*), and MYC target 1 (*MYCT1*) in LUSC.

Establishment of a risk model

We used univariate Cox regression to investigate the association between the expression levels of *ARL14*, *CARD14*, *CSTA*, *DKK4*, *DSG3*, *FAM83A*, *KLK8*, *KRT6B*, and *MYCT1* and the prognosis of LUSC, with $P < 0.05$ as the elimination criterion. Furthermore, we employed multivariate Cox regression and Akaike information criterion (AIC) to identify the relationship between the expression levels of *CSTA*,

FAM83A, and *MYCT1* and the prognosis of LUSC. We also incorporated risk scoring with LUSC tissue samples to construct a risk model for patients with LUSC.

Validation of risk gene expression in LUSC tissues

We collected tumor tissues and normal tissues from seven patients with LUSC diagnosed with pathology that were surgically treated at our hospital from July to August 2022. Patients gave their signed informed consent, and the ethics committee approved of this study at Wuhan Central Hospital (WHZXKYL2022-192). We extracted and quantified total RNA from tissue samples and performed reverse transcription according to the instructions of the reverse transcription RNA kit. We then performed reverse transcription quantitative polymerase chain reaction (RT-qPCR) and calculated the relative expression levels of *CSTA*, *FAM83A*, and *MYCT1* in the tissue samples. **Table 1** shows the primers for *CSTA*, *FAM83A*, and *MYCT1*.

Construction of a prognostic factor-related nomogram

We constructed a nomogram of *ARL14*, *CARD14*, *CSTA*, *DKK4*, *DSG3*, *FAM83A*, *KLK8*, *KRT6B*, and *MYCT1* expression based on the expression levels in cancer tissues, diagnosis, and prognostic values of *ARL14*, *CARD14*, *CSTA*, *DKK4*, *DSG3*, *FAM83A*, *KLK8*, *KRT6B*, and *MYCT1*.

The values of risk factors and construction of a risk score-related nomogram

We identified the relationship between high and low expression of *CSTA*, *FAM83A*, and *MYCT1*, categorized according to the median expression level of each gene and the clinicopathological characteristics of patients with LUSC using the meta-analysis and the K-M survival analysis functions in the online lung can-

A risk model for lung squamous cell carcinoma

cer Explorer and the UALCAN database. We also analyzed univariate Cox regression to explore the relationship between risk score and LUSC prognosis. Finally, we constructed a prognostic nomogram related to the clinical stage, T stage, age, and risk score of patients with LUSC.

The relationship between the risk score and LUSC immune infiltration

We used the CIBERSORT algorithm to calculate the levels of immune cell infiltration in the LUSC samples in TCGA. We then divided the samples into high-risk and low-risk groups based on the scores from our risk model and used the limma package to analyze the relationship between high- and low-risk scores and LUSC-infiltrating immune cells, with $P < 0.05$ as the screening criterion.

Statistical analysis

Gene expression levels in LUSC were investigated using the Wilcoxon rank sum test and limma package. We use ROC analysis to assess the diagnostic value of gene expression levels, with AUC between 0.5 and 1.0 as the evaluation standard. The higher the AUC, the more significant the diagnostic value. We used univariate and multivariate Cox regression analyses to investigate prognostic risk factors in patients with LUSC. We analyzed the relationship between risk scores and immune cell infiltration of LUSC by correlation analysis, with $P < 0.05$ as the threshold for statistical significance.

Results

Overlapping DEGs in normal tissues and squamous metaplasia tissues, carcinomas in situ, and LUSC tissues

There were 150 significant DEGs in tissues with squamous metaplasia compared to normal tissues (**Table 2**). Among them, 132 were overexpressed in metaplasia tissues, and 18 had reduced expression in metaplasia tissues compared to normal tissues. There were 1427 significant DEGs in carcinoma *in situ* tissues compared to normal tissues (**Table S1**), among which 996 were overexpressed in carcinoma tissues, and 431 had reduced expression in carcinoma tissues. There were 3137 significant DEGs in the LUSC tissues compared with the

normal tissues (**Table S2**), of which 1758 were overexpressed in the LUSC tissues, and 1379 had reduced expression in the LUSC tissues compared with the normal tissues (**Table S2**). The conversion of 70 DEGs from genes that overlapped among the three groups revealed 62 unique overlapping DEGs (**Figure S1** and **Table 3**).

The roles and signaling mechanisms of DEG enrichment and ppi network construction

The functions of the overlapping DEGs from the GEO database included epidermis development, cornification, skin development, keratinization, keratinocyte differentiation, epidermal cell differentiation, glycoside metabolism, neutrophil chemotaxis, and migration, granulocyte chemotaxis and migration, leukocyte aggregation and migration involved in an inflammatory response, positive regulation of nuclear factor- κ B (NF- κ B) transcription factor activity, secondary metabolic processes, and protein nitrosylation, among others (**Figure S2A-S2C**; **Table S3**). The overlapping DEGs were associated with signaling mechanisms for folate biosynthesis, galactose metabolism, fruit and mannose metabolism, pentose and glucuronate interconversions, and glycerolipid metabolism (**Figure S2D**). **Figure S3A** depicts the PPI network between the overlapping DEGs from the GEO data and the DEGs in the STRING database. The key DEGs in the PPI network included *SPRR1B*, *KRT16*, *IVL*, *CSTA*, and *S100A8* (**Figure S3B**).

Identification of crucial DEGs in LUSC

ADAM metallopeptidase domain 22 (*ADAM22*), ADAM like decysin 1 (*ADAMDEC1*), aldo-keto reductase family 1 member B (*AKR1B1*), aldo-keto reductase family 1 member B10 (*AKR1B10*), aldo-keto reductase family 1 member C1 (*AKR1C1*), *ARL14*, calmodulin like 5 (*CALML5*), *CARD14*, *CSTA*, diaphanous related formin 3 (*DIAPH3*), dickkopf WNT signaling pathway inhibitor 1 (*DKK1*), dickkopf WNT signaling pathway inhibitor 4 (*DKK4*), desmocollin 2 (*DSC2*), desmocollin 3 (*DSG3*), family with sequence similarity 25 member A (*FAM25A*), *FAM83A*, FXYD domain containing ion transport regulator 1 (*FXYD1*), grainyhead like transcription factor 3 (*GRHL3*), interleukin 1 receptor antagonist (*IL1RN*), involucrin (*IVL*), potassium voltage-gated channel subfamily B member 1 (*KCNB1*), *KLK8*, keratin 16 (*KRT16*), keratin 16 pseudogene 2 (*KRT16P2*), keratin 16 pseu-

A risk model for lung squamous cell carcinoma

Table 2. Differentially expressed genes in squamous metaplasia tissues

id	logFC	id	logFC	id	logFC
A_24_P175519	1.305668979	A_32_P31744	-1.273349042	A_23_P207213	2.375811323
A_23_P73097	1.676109788	A_32_P63113	1.874350357	A_23_P106806	1.479353625
A_23_P60248	1.453737069	A_23_P216052	1.723853713	A_24_P412088	1.253757953
A_23_P48350	1.257334441	A_23_P4335	1.046156459	A_23_P69537	2.34741673
A_24_P220947	1.093857676	A_24_P673063	1.362193272	A_23_P58266	1.193511551
A_23_P170233	1.959207116	A_23_P41114	2.406915532	A_32_P315178	1.7071763
A_23_P370635	1.00577998	A_32_P149158	-1.016032997	A_23_P81190	85.19582085
A_23_P135257	1.791428294	A_24_P212086	1.444105599	A_23_P155711	1.144977224
A_23_P369343	3.707520566	A_23_P353524	1.957193091	A_32_P198978	1.926114993
A_32_P112452	3.777406989	A_23_P166269	1.001215093	A_23_P24129	2.312543966
A_24_P355006	-1.318793118	A_23_P360329	1.044416316	A_24_P916782	-1.232122428
A_32_P71032	2.691356836	A_23_P4494	1.05280376	A_32_P200238	1.152174573
A_23_P23048	1.672129448	A_23_P111766	5.614856196	A_32_P62963	1.252592288
A_24_P589301	2.552030506	A_23_P115478	1.780520598	A_23_P17134	2.900480239
A_24_P348118	1.254147483	A_24_P7642	1.324696291	A_32_P52153	1.067986958
A_24_P55092	1.104058615	A_23_P310274	3.939282425	A_32_P176790	-1.013091836
A_23_P78248	1.054512256	A_23_P92562	1.276041648	A_24_P306896	1.592324703
A_32_P174121	1.832379447	A_23_P163338	1.562685678	A_32_P158272	1.014369349
A_24_P152845	1.798337561	A_23_P3038	1.215441259	A_32_P141948	1.08559652
A_23_P108062	1.60850786	A_32_P67266	1.828537263	A_23_P92161	1.812216887
A_32_P161855	1.259332336	A_32_P190303	-1.104816576	A_32_P119830	-1.122679204
A_24_P152968	1.199453308	A_23_P153120	1.134293665	A_23_P11644	1.07581238
A_23_P113793	1.583887983	A_32_P168973	1.468658268	A_24_P945059	-1.066996871
A_23_P356494	1.813166276	A_24_P282266	1.558407964	A_23_P74001	1.689310735
A_24_P129341	1.596110416	A_32_P186364	-1.213002845	A_24_P913146	1.853254949
A_23_P208126	1.22910464	A_24_P252155	1.287260405	A_23_P16523	1.088036451
A_23_P258190	1.349123989	A_24_P226755	-1.005791877	A_32_P37867	1.160838789
A_24_P859859	2.760650021	A_23_P136724	1.153515082	A_23_P108216	-1.008609446
A_23_P500010	2.002885752	A_32_P94444	3.755643047	A_23_P8801	2.06619658
A_23_P163336	1.070258728	A_24_P104689	1.442011501	A_32_P34138	3.507513336
A_32_P204676	1.259019153	A_23_P213050	1.224115167	A_23_P76743	1.027252127
A_23_P59877	1.194315142	A_23_P94275	1.571666444	A_23_P128574	-1.064032554
A_24_P238250	1.243422754	A_24_P68908	1.617632117	A_23_P204947	1.072236955
A_23_P18751	2.804584864	A_24_P43810	1.624736366	A_23_P214935	-1.10396774
A_23_P209995	1.430272841	A_32_P51855	1.258839681	A_32_P471485	1.824972945
A_23_P30126	1.982165789	A_24_P918065	1.106835785	A_23_P45751	1.589627155
A_23_P38537	1.133831182	A_23_P151975	1.200385538	A_23_P351148	1.869926295
A_23_P140928	1.06922676	A_23_P124095	2.292832348	A_23_P256425	1.126512375
A_23_P93641	1.593320408	A_32_P71710	2.505112783	A_23_P74723	1.14070976
A_23_P432978	1.190852292	A_24_P153035	1.478314805	A_24_P31627	-1.289108403
A_23_P52067	1.542073304	A_24_P48495	1.050669737	A_23_P55198	-1.113960254
A_32_P141338	1.289107314	A_24_P360674	1.040875417	A_23_P320070	1.03797418
A_23_P434809	6.448052031	A_23_P430718	1.743908998	A_32_P128174	-1.559302387
A_23_P500000	2.56451442	A_24_P408736	1.007930946	A_23_P358917	2.00037122
A_23_P52410	1.236176037	A_24_P71781	-1.37130085	A_32_P157208	-1.172676541
A_24_P416645	1.825724421	A_23_P66739	2.506683538	A_24_P313895	1.150806867
A_23_P324754	1.247052108	A_23_P76249	1.00272057	A_23_P250385	1.699721426
A_23_P217498	2.365430447	A_23_P119015	1.173560213	A_23_P155660	1.012706161
A_23_P159406	1.742366668	A_24_P245379	1.855535138	A_24_P399490	1.055003049
A_23_P201706	1.768636607	A_23_P123234	1.132425178	A_32_P150891	1.040569253

Note: DEGs, Differentially Expressed Genes; FC, Fold Change.

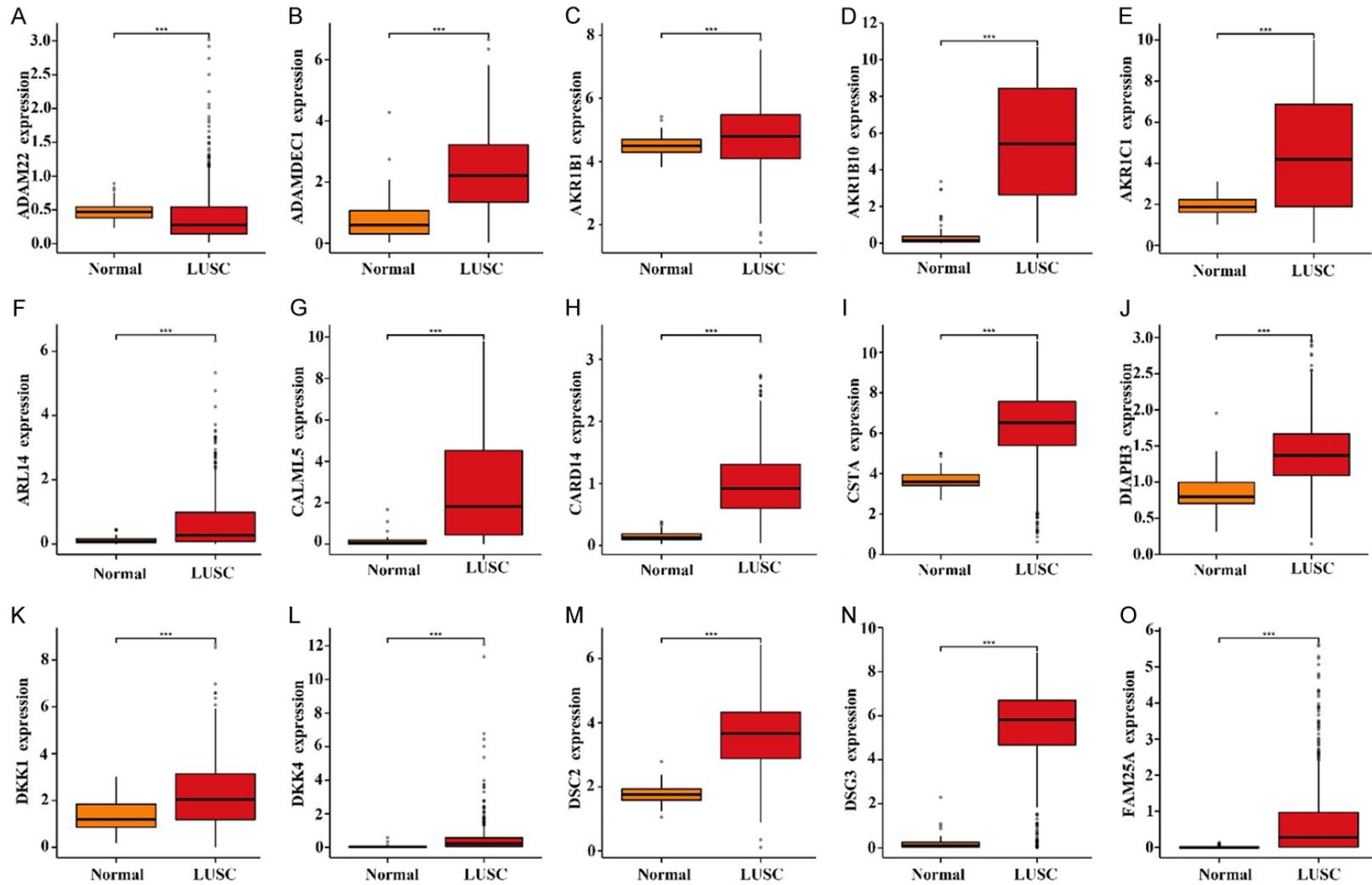
A risk model for lung squamous cell carcinoma

Table 3. Differentially expressed genes associated with lung squamous cell carcinoma progression after gene name conversion

id	Gene	Id	Gene
A_23_P106806	PRSS27	A_23_P73097	RGS20
A_23_P108216	FXYD1	A_23_P74001	S100A12
A_23_P113793	ZBED2	A_23_P76249	KRT6B
A_23_P11644	SPRR2D	A_23_P92161	ARL14
A_23_P124095	CALML5	A_23_P93641	AKR1B10
A_23_P128574	ENOX1	A_23_P94275	DKK4
A_23_P136724	LOC344887	A_24_P129341	AKR1B10
A_23_P151975	RHCG	A_24_P152968	AKR1C1
A_23_P153120	DSG3	A_24_P220947	AKR1C1
A_23_P155660	PPP2R2C	A_24_P226755	TOX
A_23_P155711	NEIL3	A_24_P31627	KCNB1
A_23_P159406	SPRR1B	A_24_P355006	ADAM22
A_23_P170233	CSTA	A_24_P412088	MCM10
A_23_P18751	TMPRSS11E	A_24_P43810	FAM83A
A_23_P201706	S100A2	A_24_P48495	LYPD3
A_23_P209995	IL1RN	A_24_P673063	FABP5
A_23_P216052	FAM83A	A_24_P68908	LOC344887
A_23_P23048	S100A9	A_24_P7642	FABP5
A_23_P24129	DKK1	A_24_P945059	MYCT1
A_23_P256425	ADAMDEC1	A_32_P112452	LOC100652944
A_23_P258190	AKR1B1	A_32_P119830	PEG3-AS1
A_23_P310274	PRSS2	A_32_P149158	PLCL1
A_23_P320070	CARD14	A_32_P150891	DIAPH3
A_23_P353524	IVL	A_32_P157208	LOC572558
A_23_P369343	KLK8	A_32_P168973	KRT16P3
A_23_P38537	KRT16	A_32_P190303	LONRF2
A_23_P434809	S100A8	A_32_P200238	UCA1
A_23_P4494	DSC2	A_32_P204676	FABP5
A_23_P52067	GRHL3	A_32_P34138	FAM25A
A_23_P55198	CNTD1	A_32_P62963	KRT16P2
A_23_P59877	FABP5	A_32_P94444	PRSS2

Note: PRSS27, Serine Protease 27; FXYD1, FXYD Domain Containing Ion Transport Regulator 1; ZBED2, Zinc Finger BED-Type Containing 2; SPRR2D, Small Proline Rich Protein 2D; CALML5, Calmodulin Like 5; ENOX1, Ecto-NOX Disulfide-Thiol Exchanger 1; RHCG, Rh Family C Glycoprotein; DSG3, Desmoglein 3; PPP2R2C, Protein Phosphatase 2 Regulatory Subunit Bgamma; NEIL3, Nei Like DNA Glycosylase 3; SPRR1B, Small Proline Rich Protein 1B; CSTA, Cystatin A; TMPRSS11E, Transmembrane Serine Protease 11E; S100A2, S100 Calcium Binding Protein A2; IL1RN, Interleukin 1 Receptor Antagonist; FAM83A, Family With Sequence Similarity 83 Member A; S100A9, S100 Calcium Binding Protein A9; DKK1, Dickkopf WNT Signaling Pathway Inhibitor 1; ADAMDEC1, ADAM Like Decysin 1; AKR1B1, Aldo-Keto Reductase Family 1 Member B; PRSS2, Serine Protease 2; CARD14, Caspase Recruitment Domain Family Member 14; IVL, Involucrin; KLK8, Kallikrein Related Peptidase 8; KRT16, Keratin 16; S100A8, S100 Calcium Binding Protein A8; DSC2, Desmocollin 2; GRHL3, Grainyhead Like Transcription Factor 3; CNTD1, Cyclin N-Terminal Domain Containing 1; FABP5, Fatty Acid Binding Protein 5; RGS20, Regulator Of G Protein Signaling 20; S100A12, S100 Calcium Binding Protein A12; KRT6B, Keratin 6B; ARL14, ADP Ribosylation Factor Like Gtpase 14; AKR1B10, Aldo-Keto Reductase Family 1 Member B10; DKK4, Dickkopf WNT Signaling Pathway Inhibitor 4; AKR1C1, Aldo-Keto Reductase Family 1 Member C1; TOX, Thymocyte Selection Associated High Mobility Group Box; KCNB1, Potassium Voltage-Gated Channel Subfamily B Member 1; ADAM22, ADAM Metalloproteinase Domain 22; MCM10, Minichromosome Maintenance 10 Replication Initiation Factor; FAM83A, Family With Sequence Similarity 83 Member A; LYPD3, LY6/PLAUR Domain Containing 3; MYCT1, MYC Target 1; PLCL1, Phospholipase C Like 1; DIAPH3, Diaphanous Related Formin 3; KRT16P3, Keratin 16 Pseudogene 3; LONRF2, LON Peptidase N-Terminal Domain And Ring Finger 2; UCA1, Urothelial Cancer Associated 1; FAM25A, Family With Sequence Similarity 25 Member A; KRT16P2, Keratin 16 Pseudogene 2.

A risk model for lung squamous cell carcinoma



A risk model for lung squamous cell carcinoma

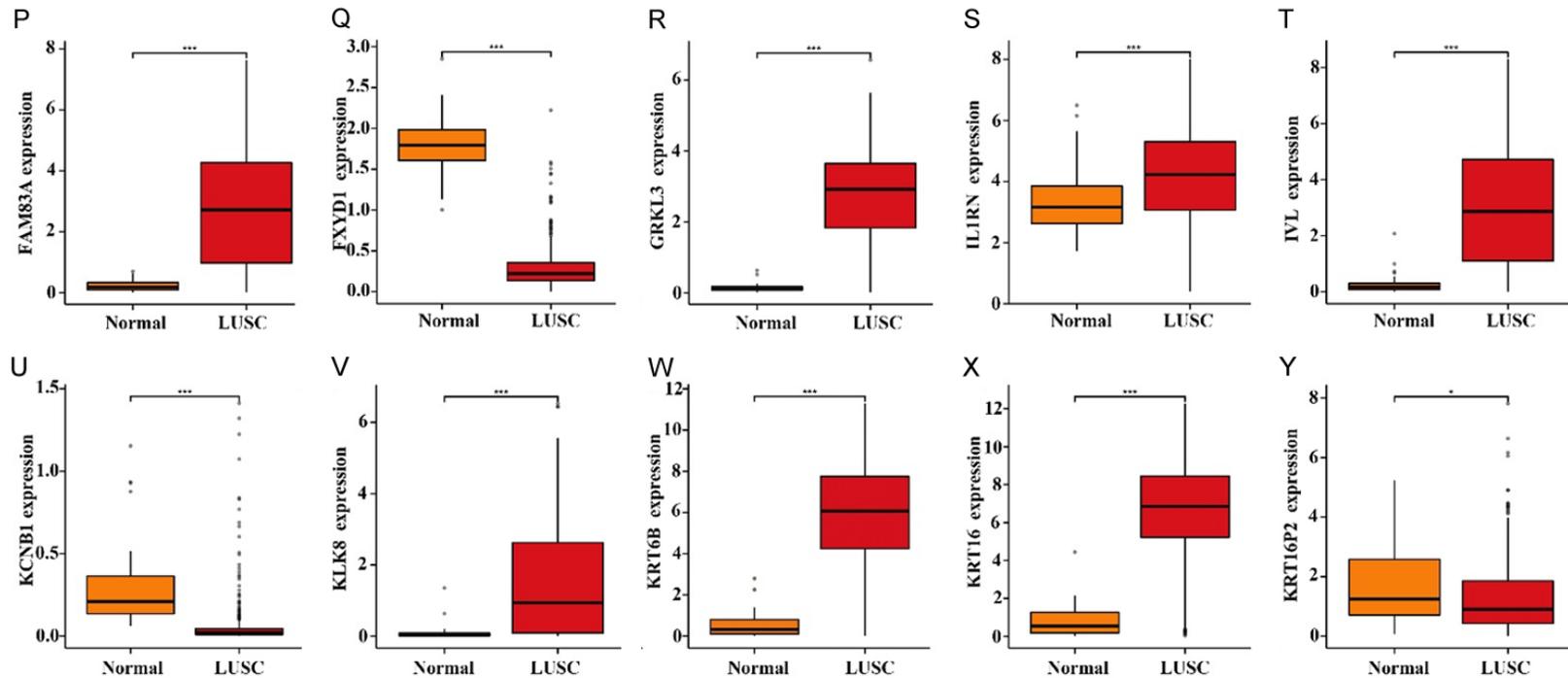
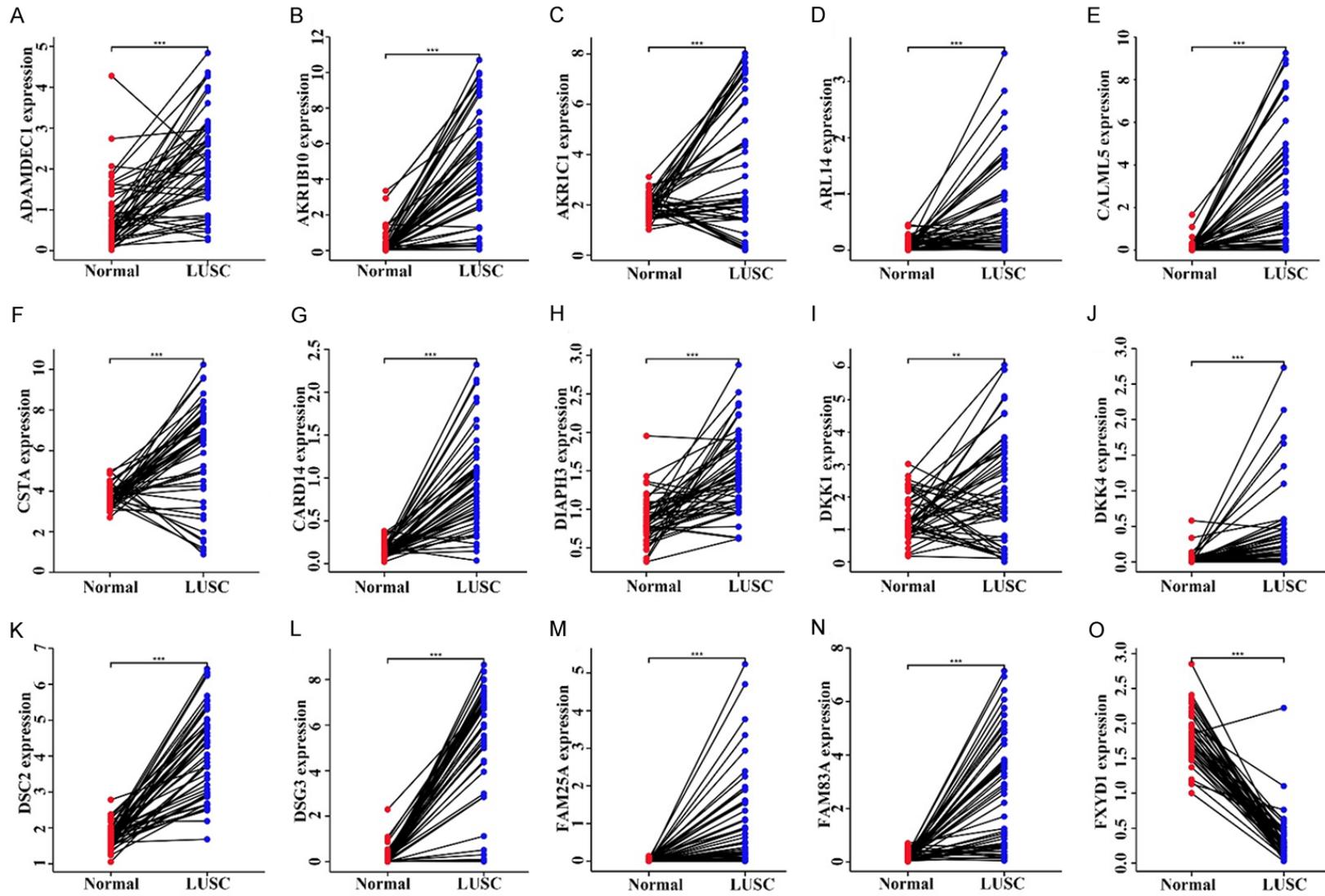


Figure 1. The differentially expressed genes in unpaired lung squamous cell carcinoma tissues of the Cancer Genome Atlas database. A. ADAM22; B. ADAMDEC1; C. AKR1B1; D. AKR1B10; E. AKR1C1; F. ARL14; G. CALML5; H. CARD14; I. CSTA; J. DIAPH3; K. DKK1; L. DKK4; M. DSC2; N. DSG3; O. FAM25A; P. FAM83A; Q. FXYD1; R. GRHL3; S. IL1RN; T. IVL; U. KCNB1; V. KLK8; W. KRT6B; X. KRT16; Y. KRT16P2. Note: LUSC, Lung Squamous Cell Carcinoma; ADAM22, ADAM Metalloproteinase Domain 22; ADAMDEC1, ADAM Like Decysin 1; AKR1B1, Aldo-Keto Reductase Family 1 Member B; AKR1B10, Aldo-Keto Reductase Family 1 Member B10; AKR1C1, Aldo-Keto Reductase Family 1 Member C1; ARL14, ADP Ribosylation Factor Like Gtpase 14; CALML5, Calmodulin Like 5; CARD14, Caspase Recruitment Domain Family Member 14; CSTA, Cystatin A; DIAPH3, Diaphanous Related Formin 3; DKK1, Dickkopf WNT Signaling Pathway Inhibitor 1; DKK4, Dickkopf WNT Signaling Pathway Inhibitor 4; DSC2, Desmocollin 2; DSC3, Desmocollin 3; FAM25A, Family With Sequence Similarity 25 Member A; FAM83A, Family With Sequence Similarity 83 Member A; FXYD1, FXYD Domain Containing Ion Transport Regulator 1; GRHL3, Grainyhead Like Transcription Factor 3; IL1RN, Interleukin 1 Receptor Antagonist; IVL, Involucrin; KCNB1, Potassium Voltage-Gated Channel Subfamily B Member 1; KLK8, Kallikrein Related Peptidase 8; KRT6B, Keratin 6B; KRT16, Keratin 16; KRT16P2, Keratin 16 Pseudogene 2; DSG3, Desmoglein 3.

A risk model for lung squamous cell carcinoma



A risk model for lung squamous cell carcinoma

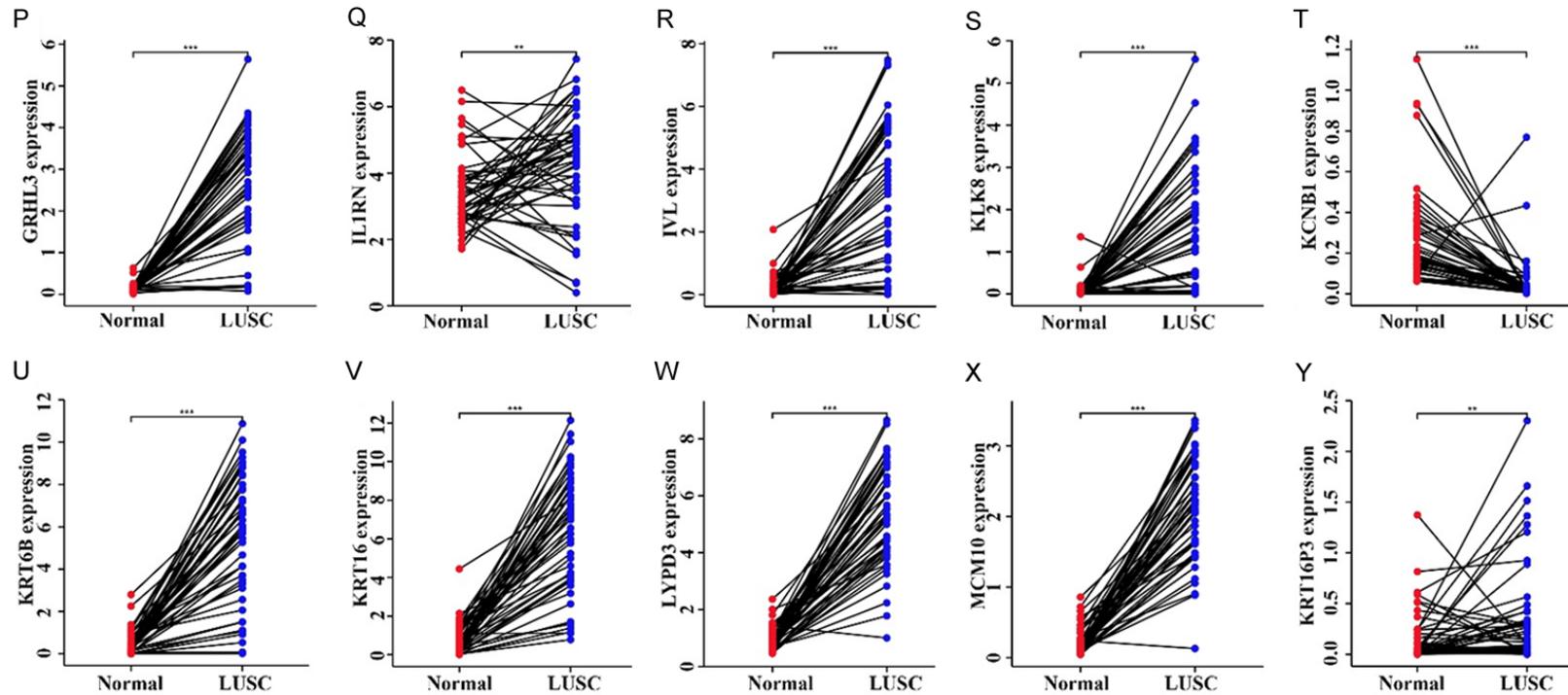


Figure 2. Differentially expressed genes in paired lung squamous cell carcinoma tissues from the Cancer Genome Atlas database. A. ADAMDEC1; B. AKR1B10; C. AKR1C1; D. ARL14; E. CALML5; F. CSTA; G. CARD14; H. DIAPH3; I. DKK1; J. DKK4; K. DSC2; L. DSG3; M. FAM25A; N. FAM83A; O. FXYD1; P. GRHL3; Q. IL1RN; R. IVL; S. KLK8; T. KCNB1; U. KRT6B; V. KRT16; W. LYPD3; X. MCM10; Y. KRT16P3. Note: LUSC, Lung Squamous Cell Carcinoma; ADAMDEC1, ADAM Like Decysin 1; AKR1B10, Aldo-Keto Reductase Family 1 Member B10; AKR1C1, Aldo-Keto Reductase Family 1 Member C1; ARL14, ADP Ribosylation Factor Like Gtpase 14; CALML5, Calmodulin Like 5; CARD14, Caspase Recruitment Domain Family Member 14; CSTA, Cystatin A; DIAPH3, Diaphanous Related Formin 3; DKK1, Dickkopf WNT Signaling Pathway Inhibitor 1; DKK4, Dickkopf WNT Signaling Pathway Inhibitor 4; DSC2, Desmocollin 2; DSC3, Desmocollin 3; FAM25A, Family With Sequence Similarity 25 Member A; FAM83A, Family With Sequence Similarity 83 Member A; FXYD1, FXYD Domain Containing Ion Transport Regulator 1; GRHL3, Grainyhead Like Transcription Factor 3; IL1RN, Interleukin 1 Receptor Antagonist; IVL, Involucrin; KCNB1, Potassium Voltage-Gated Channel Subfamily B Member 1; KLK8, Kallikrein Related Peptidase 8; KRT6B, Keratin 6B; KRT16, Keratin 16; LYPD3, LY6/PLAUR Domain Containing 3; MCM10, Minichromosome Maintenance 10 Replication Initiation Factor; KRT16P3, Keratin 16 Pseudogene 3; DSG3, Desmoglein 3.

A risk model for lung squamous cell carcinoma

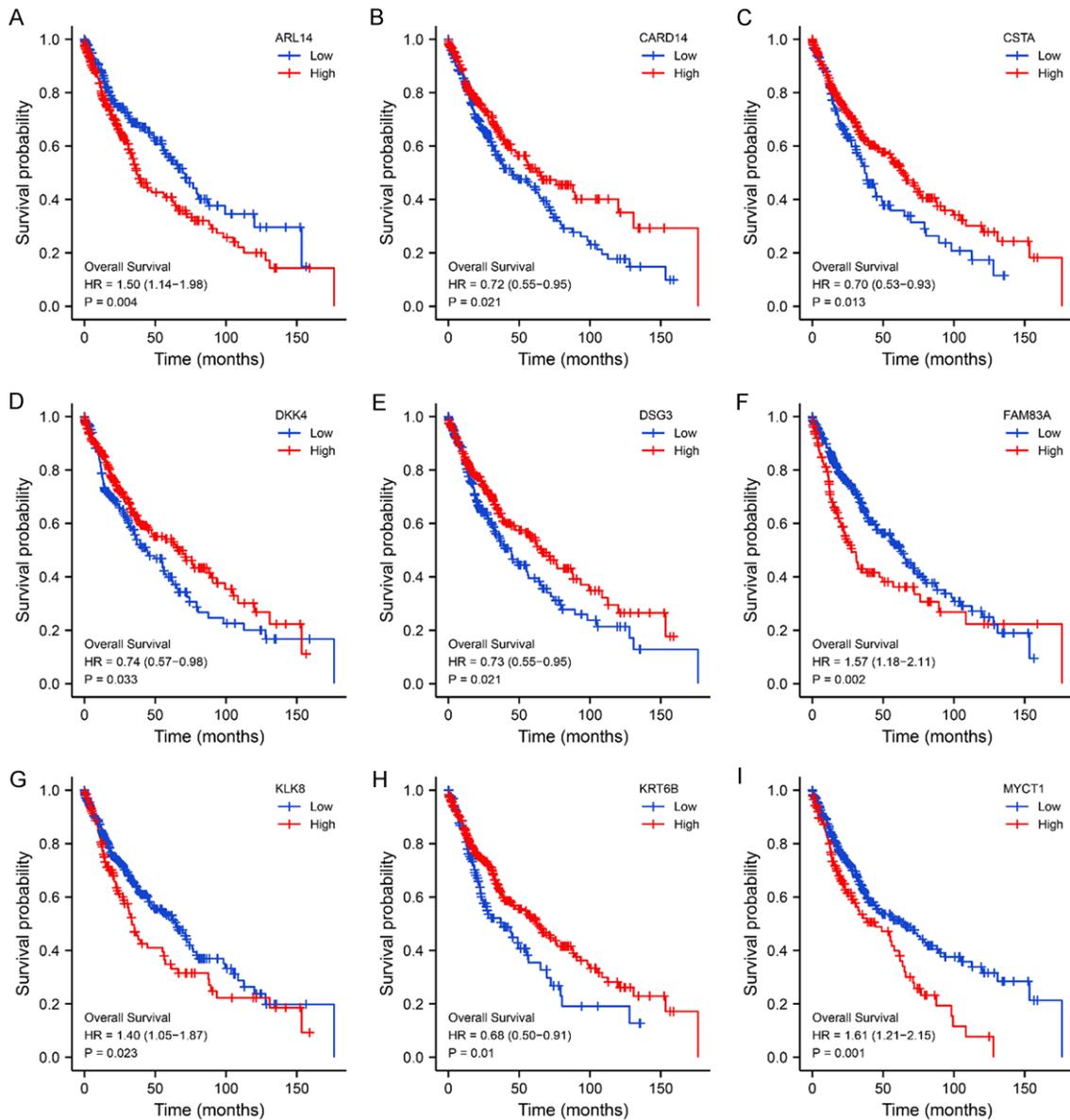


Figure 3. Kaplan-Meier survival analysis of patients with lung squamous cell carcinoma stratified by expression levels of candidate prognostic genes. A. ARL14; B. CARD14; C. CSTA; D. DKK4; E. DSG3; F. FAM83A; G. KLK8; H. KRT6B; I. MYCT1. Note: ARL14, ADP Ribosylation Factor Like Gtpase 14; CARD14, Caspase Recruitment Domain Family Member 14; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; KLK8, Kallikrein Related Peptidase 8; KRT6B, Keratin 6B; DKK4, Dickkopf WNT Signaling Pathway Inhibitor 4; DSG3, Desmoglein 3; MYCT1, MYC Target 1.

dogene 3 (*KRT16P3*), *KRT6B*, LON peptidase N-terminal domain and ring finger 2 (*LONRF2*), LY6/PLAUR domain containing 3 (*LYPD3*), mini-chromosome maintenance 10 replication initiation factor (*MCM10*), *MYCT1*, nei like DNA glycosylase 3 (*NEIL3*), phospholipase C like 1 (*PLCL1*), protein phosphatase 2 regulatory subunit Bgamma (*PPP2R2C*), serine protease 2 (*PRSS2*), serine protease 27 (*PRSS27*), regula-

tor of G protein signaling 20 (*RGS20*), Rh family C glycoprotein (*RHCG*), S100 calcium binding protein A12 (*S100A12*), S100 calcium binding protein A2 (*S100A2*), S100 calcium binding protein A9 (*S100A9*), small proline rich protein 1B (*SPRR1B*), small proline rich protein 2D (*SPRR2D*), thymocyte selection associated high mobility group box (*TOX*), urothelial cancer associated 1 (*UCA1*), and zinc finger BED-type

A risk model for lung squamous cell carcinoma

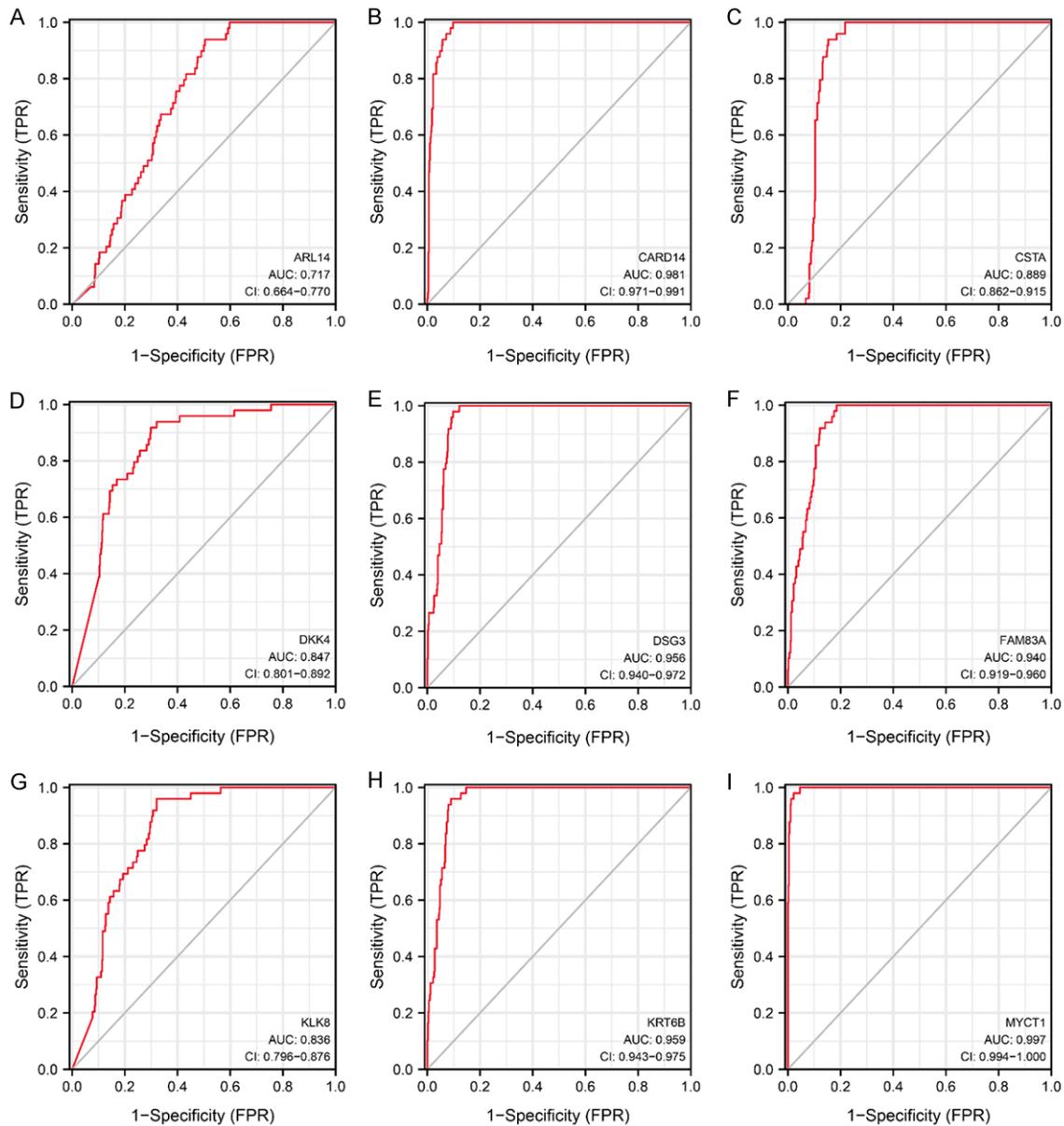


Figure 4. Receiver operating characteristic analysis of the diagnostic values of differentially expressed genes in LUSC. A. ARL14; B. CARD14; C. CSTA; D. DKK4; E. DSG3; F. FAM83A; G. KLK8; H. KRT6B; I. MYCT1. Note: LUSC, Lung Squamous Cell Carcinoma; ARL14, ADP Ribosylation Factor Like Gtpase 14; CARD14, Caspase Recruitment Domain Family Member 14; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; KLK8, Kallikrein Related Peptidase 8; KRT6B, Keratin 6B; DKK4, Dickkopf WNT Signaling Pathway Inhibitor 4; DSG3, Desmoglein 3; MYCT1, MYC Target 1.

containing 2 (*ZBED2*) showed significant changes in expression between unpaired LUSC tissues and normal lung tissues (**Figures 1 and S4**). *ADAMDEC1*, *AKR1B1*, *AKR1B10*, *AKR1C1*, *ARL14*, *CALML5*, *CARD14*, *CSTA*, *DIAPH3*, *DKK1*, *DKK4*, *DSC2*, *DSG3*, *FAM25A*, *FAM83A*, *FXD1*, *GRHL3*, *IL1RN*, *IVL*, *KCNB1*, *KLK8*, *KRT16*, *KRT16P3*, *KRT6B*, *LONRF2*, *LYPD3*, *MCM10*, *MYCT1*, *NEIL3*, *PLCL1*, *PPP2R2C*, *PRSS2*, *PRSS27*, *RGS20*, *RHCG*,

S100A12, *S100A2*, *S100A9*, *SPRR1B*, *SPRR2D*, *TMPRSS11E*, *TOX*, *UCA1*, and *ZBED2* had significantly different expression between the 49 LUSC tissues and the 49 paired normal lung tissues (**Figures 2 and S5**).

Construction of a prognostic nomogram for patients with LUSC

FAM83A, *KLK8*, and *MYCT1* overexpression levels were each correlated with the short over-

A risk model for lung squamous cell carcinoma

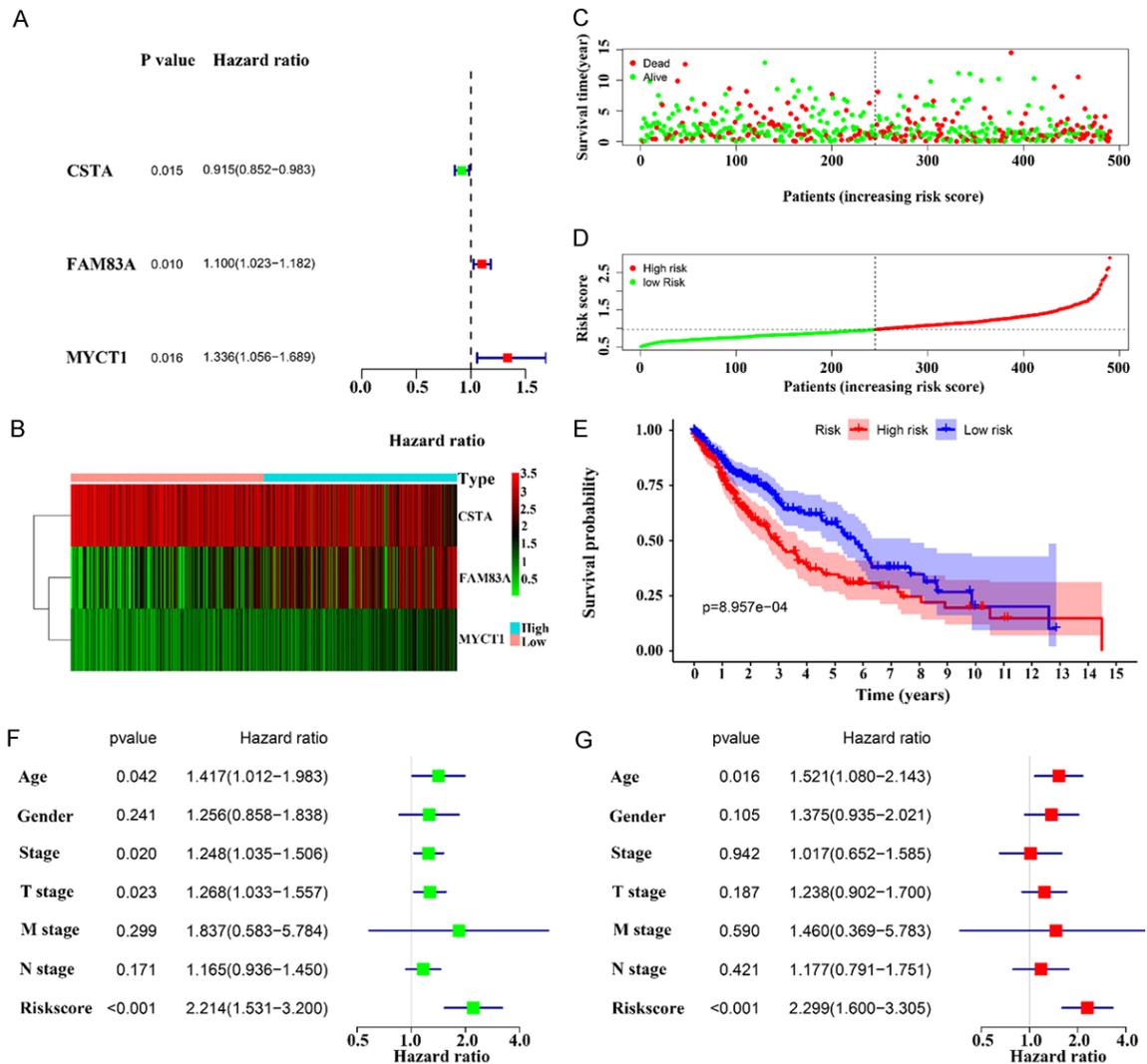


Figure 5. A prognostic-related risk model for LUSC. A. Prognostic risk factors identified by univariate Cox regression analysis. B. Multivariate Cox regression and AIC screening confirmed the relationship between risk factors and patient prognosis. C, D. The relationship between overall survival of patients with lung squamous cell carcinoma and risk factors. E. Patients with high-risk scores in the risk model have a poor prognosis. F, G. Cox regression analysis identified prognostic factors in lung squamous cell carcinoma. Note: LUSC, Lung Squamous Cell Carcinoma; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; MYCT1, MYC Target 1.

Table 4. Univariate Cox regression analysis revealed prognosis-related genes

Gene	HR	P
CSTA	0.915229739	0.015172199
FAM83A	1.099601051	0.010161734
MYCT1	1.335569478	0.015580685

Note: HR, Hazard Ratio; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; MYCT1, MYC Target 1.

all survival (OS) of patients with LUSC. In contrast, lower expression levels of *ARL14*,

CARD14, *CSTA*, *DKK4*, *DSG3*, and *KRT6B* were each associated with a short OS in LUSC patients according to the K-M survival plots (Figure 3). ROC analysis showed that the expression levels of *FAM83A*, *MYCT1*, *ARL14*, *CARD14*, *CSTA*, *DKK4*, *DSG3*, *KLK8*, and *KRT6B* could be used to diagnose LUSC (Figure 4). The AUCs of *ARL14*, *CARD14*, *CSTA*, *DKK4*, *DSG3*, *FAM83A*, *KLK8*, *KRT6B* and *MYCT1* expression in LUSC were 0.717, 0.981, 0.889, 0.847, 0.956, 0.94, 0.836, 0.959, and 0.997, respectively. A nomogram was constructed for DEGs related to LUSC prognosis and diagnosis to

Table 5. Expression of MYCT1, FAM83A, and CSTA was associated with a poor prognosis in patients with lung squamous cell carcinoma

Gene	HR (95% CI)	P	Method	Ref
MYCT1	1.01 (0.89-1.15)	0.84	Meta-analysis	LCE database
FAM83A	1.14 (1.04-1.25)	**	Meta-analysis	LCE database
CSTA	0.90 (0.82-0.99)	*	Meta-analysis	LCE database
MYCT1	NA	*	K-M survival	UALCAN database
FAM83A	NA	***	K-M survival	UALCAN database
CSTA	NA	0.64	K-M survival	UALCAN database

Note: HR, Hazard Ratio; CI, Confidence Interval; LCE, Lung Cancer Explorer; K-M, Kaplan-Meier; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; MYCT1, MYC Target 1; *, P < 0.05; **, P < 0.01; ***, P < 0.001.

assess the prognosis of patients with LUSC (Figure S6).

Constructing a prognostic risk model of LUSC

We used univariate Cox regression to examine the relationship between the expression levels of *ARL14*, *CARD14*, *CSTA*, *DKK4*, *DSG3*, *FAM83A*, *KLK8*, *KRT6B*, and *MYCT1* and the prognosis of patients with LUSC. We found that *CSTA*, *FAM83A*, and *MYCT1* each had a significant prognostic value (Figure 5A). Furthermore, multivariate Cox regression and AIC screening revealed that the expression levels of *CSTA*, *FAM83A*, and *MYCT1* independently influenced the prognosis of LUSC (Figure 5B and Table 4). Using the expression levels of these three genes, we constructed a risk score as follows: risk score = (*CSTA* × -0.099643117) + (*FAM83A* × 0.117245801) + (*MYCT1* × 0.246912866). Figure 5C-E shows the relationship between risk scores and OS in patients with LUSC, confirming that patients with high-risk scores had poor prognoses.

Prognostic risk model nomogram

PCR confirmed that the expression of *CSTA* was reduced in 71.43% (5/7) of patients with LUSC from our hospital (Figure S7A), while the expression of *FAM83A* and *MYCT1* increased in 100% and 71.43% (5/7) of the patients, respectively (Figure S7B and S7C). Meta-analysis and K-M survival analysis using Lung Cancer Explorer and UALCAN databases showed that high expression of *MYCT1* and *FAM83A* and low expression of *CSTA* were associated with poor prognosis in patients with LUSC (Table 5). *CSTA* expression in TCGA LUSC tissues was related to

progression-free interval end-point events in patients with LUSC after the samples were grouped according to the median expression level of *CSTA* (Table 6). On the contrary, expression levels of *FAM83A* and *MYCT1*, similarly grouped by their median values, were associated with primary therapy outcomes of patients with LUSC (Tables 7 and 8). Furthermore, there were significant differences in the expression levels of *CSTA*, *FAM83A*, and *MYCT1* in the risk model score grouping (Figure S8), indicating that the risk model based on the expression of *CSTA*, *FAM83A*, and *MYCT1* could predict disease progression and prognosis in patients with LUSC.

The univariate Cox regression analysis revealed that age, clinical stage, T stage, and risk score all impacted the prognosis of LUSC (Figure 5F). Multivariate Cox regression further revealed that age and risk score were independent prognostic factors in LUSC (Figure 5G). We used age, clinical stage, T stage, and risk score to create a prognostic model nomogram (Figure 6), which showed that the clinical stage is the most important prognostic factor, followed by the risk score and T stage.

The risk score is related to the infiltrating of immune cells in LUSC

Immune cell infiltration is an essential factor in cancer progression. Therefore, we calculated the levels of immune cell infiltration in TCGA LUSC tissues using the CIBERSORT algorithm. We divided the LUSC tissues into high- and low-risk groups using the risk scores and the median value of the risk score among all samples. We observed that the risk score was significantly correlated with the levels of LUSC immune infiltration consisting of naïve B cells, helper follicular T cells, neutrophils, and activated dendritic cells (Figure 7).

Discussion

Patients with LUSC experience high morbidity and mortality [3] and have a worse prognosis than patients with LAC. Therefore, it is impera-

A risk model for lung squamous cell carcinoma

Table 6. Clinicopathological characteristics of patients with high and low CSTA expression in LUSC

Characteristic	Low expression of CSTA	High expression of CSTA	P
T stage			0.120
T1	60 (12%)	54 (10.8%)	
T2	135 (26.9%)	159 (31.7%)	
T3	43 (8.6%)	28 (5.6%)	
T4	13 (2.6%)	10 (2%)	
N stage			0.679
N0	161 (32.5%)	159 (32.1%)	
N1	65 (13.1%)	66 (13.3%)	
N2	20 (4%)	20 (4%)	
N3	1 (0.2%)	4 (0.8%)	
M stage			0.723
M0	208 (49.6%)	204 (48.7%)	
M1	3 (0.7%)	4 (1%)	
Pathologic stage			0.625
Stage I	119 (23.9%)	126 (25.3%)	
Stage II	78 (15.7%)	84 (16.9%)	
Stage III	47 (9.4%)	37 (7.4%)	
Stage IV	3 (0.6%)	4 (0.8%)	
Primary therapy outcome			0.494
PD	19 (5.3%)	12 (3.3%)	
SD	7 (1.9%)	10 (2.8%)	
PR	2 (0.6%)	3 (0.8%)	
CR	151 (41.8%)	157 (43.5%)	
Gender			0.155
Female	73 (14.5%)	58 (11.6%)	
Male	178 (35.5%)	193 (38.4%)	
Race			0.890
Asian	4 (1%)	5 (1.3%)	
Black or African American	14 (3.6%)	16 (4.1%)	
White	176 (45.2%)	174 (44.7%)	
Age			0.482
≤ 65	91 (18.5%)	100 (20.3%)	
> 65	155 (31.4%)	147 (29.8%)	
Smoker			0.824
No	10 (2%)	8 (1.6%)	
Yes	236 (48.2%)	236 (48.2%)	
OS event			0.241
Alive	136 (27.1%)	150 (29.9%)	
Dead	115 (22.9%)	101 (20.1%)	
DSS event			0.276
Alive	177 (39.3%)	184 (40.9%)	
Dead	50 (11.1%)	39 (8.7%)	
PFI event			0.040
Alive	166 (33.1%)	188 (37.5%)	
Dead	85 (16.9%)	63 (12.5%)	

Note: CSTA, Cystatin A; CR, Complete Response; PR, Partial Response; SD, Stable Disease; PD, Progressive Disease; OS, Overall Survival; DSS, Disease-Specific Survival; PFI, Progression-Free Interval; LUSC, Lung Squamous Cell Carcinoma.

tive to find new treatment targets and methods to improve the prognosis of patients with LUSC. Many studies have reported the role of changes in gene expression during LUSC development, and modulation of gene expression is expected to improve the prognosis of LUSC patients [9, 17, 18]. For example, fat mass and obesity-associated protein (*FTO*) influences the prognosis of patients with LUSC and are the main factors causing abnormal m^A modification in LUSC. *FTO* knockdown can effectively promote apoptosis and inhibit the proliferation of L78 and NCI-H520 cells, while overexpression of *FTO* encourages the malignant phenotype of CHLH-1 cells. Furthermore, *FTO* can enhance myeloid zinc finger 1 (*MZF1*) expression by reducing the levels of m^A and the stability of *MZF1* mRNA transcripts, thus exerting oncogenic functionalities [17]. As another example, lncRNA NNT-AS1 deletion inhibits LUSC cell migration and invasion and induces apoptosis. Overexpression of miR-22 impedes LUSC progression by targeting *FOXM* expression. NNT-AS1 directly regulates *FOXM1* expression by binding to miR-22 in LUSC cells, thus affecting the growth and migration of LUSC cells [9].

The TCGA and GEO databases contain expression data for many cancer genes, miRNAs, lncRNAs, and other RNAs, and are used in many cancer studies [2, 19, 20]. For example, Takeda et al. used transcriptome data from the TCGA and GEO databases to show that high expression of insulin like growth factor 2 receptor (*IGF2R*), a tumor suppressor gene, is related to a poor prognosis in cervical cancer. Further research using various cellular models showed that interference with *IGF2R* expression in

A risk model for lung squamous cell carcinoma

Table 7. Clinicopathological characteristics of patients with high and low FAM83A expression in LUSC

Characteristic	Low expression of FAM83A	High expression of FAM83A	P
T stage			0.262
T1	60 (12%)	54 (10.8%)	
T2	150 (29.9%)	144 (28.7%)	
T3	28 (5.6%)	43 (8.6%)	
T4	13 (2.6%)	10 (2%)	
N stage			0.762
N0	161 (32.5%)	159 (32.1%)	
N1	70 (14.1%)	61 (12.3%)	
N2	18 (3.6%)	22 (4.4%)	
N3	2 (0.4%)	3 (0.6%)	
M stage			0.720
M0	210 (50.1%)	202 (48.2%)	
M1	3 (0.7%)	4 (1%)	
Pathologic stage			0.736
Stage I	120 (24.1%)	125 (25.1%)	
Stage II	87 (17.5%)	75 (15.1%)	
Stage III	40 (8%)	44 (8.8%)	
Stage IV	3 (0.6%)	4 (0.8%)	
Primary therapy outcome			0.034
PD	12 (3.3%)	19 (5.3%)	
SD	5 (1.4%)	12 (3.3%)	
PR	1 (0.3%)	4 (1.1%)	
CR	168 (46.5%)	140 (38.8%)	
Gender			1.000
Female	65 (12.9%)	66 (13.1%)	
Male	186 (37.1%)	185 (36.9%)	
Race			0.757
Asian	4 (1%)	5 (1.3%)	
Black or African American	13 (3.3%)	17 (4.4%)	
White	177 (45.5%)	173 (44.5%)	
Age			0.059
≤ 65	106 (21.5%)	85 (17.2%)	
> 65	140 (28.4%)	162 (32.9%)	
Smoker			0.810
No	8 (1.6%)	10 (2%)	
Yes	237 (48.4%)	235 (48%)	
OS event			0.528
Alive	147 (29.3%)	139 (27.7%)	
Dead	104 (20.7%)	112 (22.3%)	
DSS event			1.000
Alive	180 (40%)	181 (40.2%)	
Dead	45 (10%)	44 (9.8%)	
PFI event			0.769
Alive	179 (35.7%)	175 (34.9%)	
Dead	72 (14.3%)	76 (15.1%)	

Note: FAM83A, Family With Sequence Similarity 83 Member A; CR, Complete Response; PR, Partial Response; SD, Stable Disease; PD, Progressive Disease; OS, Overall Survival; DSS, Disease-Specific Survival; PFI, Progression-Free Interval; LUSC, Lung Squamous Cell Carcinoma.

cervical cancer cells could induce apoptosis, reduce viability, and increase susceptibility to the anticancer drug cisplatin. *IGF2R* can also exert carcinogenic effects by transporting M6P-labeled cargo [20]. The transition from normal tissue to carcinoma *in situ*, or LUSC, takes time. Therefore, the GEO database can discover critical DEGs in LUSC progression, providing novel targets for LUSC diagnosis and treatment. In our study, DEGs that were overlapped with squamous metaplasia, carcinoma *in situ*, and LUSC were involved the development of the epidermis, cornification, epidermal cell differentiation, glycoside metabolic process, neutrophil chemotaxis, and migration, granulocyte chemotaxis and migration, leukocyte aggregation, migration involved in the inflammatory response, positive regulation of Nf-κB transcription factor activity, secondary metabolic processes, protein nitrosylation, and other roles. The appearance and development of LUSC may be induced by chronic stimulation and injury to columnar epithelial cells of the bronchial mucosa, loss of cilia, and squamous metaplasia basal cells. The overlapping DEGs are enriched with biological functions related to these possible causes of LUSC.

We found that the expression of *FAM83A*, *MYCT1*, *ARL14*, *CARD14*, *CSTA*, *DKK4*, *DSG3*, *KLK8*, and *KRT6B* were associated with LUSC prognosis and had diagnostic potential. Various studies have confirmed associations between *ARL14*, *CARD14*, *DKK4*, *DSG3*, *KLK8*, *FAM83A*, *MYCT1*, and *CSTA* expression levels and cancer progression [21-34]. For example, *ARL14* expression levels are associated with a

A risk model for lung squamous cell carcinoma

Table 8. Clinicopathological characteristics of patients with high and low MYCT1 expression in LUSC

Characteristic	Low expression of MYCT1	High expression of MYCT1	P
T stage			0.958
T1	55 (11%)	59 (11.8%)	
T2	147 (29.3%)	147 (29.3%)	
T3	37 (7.4%)	34 (6.8%)	
T4	12 (2.4%)	11 (2.2%)	
N stage			0.468
N0	157 (31.7%)	163 (32.9%)	
N1	69 (13.9%)	62 (12.5%)	
N2	22 (4.4%)	18 (3.6%)	
N3	1 (0.2%)	4 (0.8%)	
M stage			1
M0	205 (48.9%)	207 (49.4%)	
M1	3 (0.7%)	4 (1%)	
Pathologic stage			0.481
Stage I	116 (23.3%)	129 (25.9%)	
Stage II	89 (17.9%)	73 (14.7%)	
Stage III	42 (8.4%)	42 (8.4%)	
Stage IV	3 (0.6%)	4 (0.8%)	
Primary therapy outcome			0.036
PD	17 (4.7%)	14 (3.9%)	
SD	5 (1.4%)	12 (3.3%)	
PR	5 (1.4%)	0 (0%)	
CR	166 (46%)	142 (39.3%)	
Gender			0.067
Female	56 (11.2%)	75 (14.9%)	
Male	195 (38.8%)	176 (35.1%)	
Race			0.395
Asian	3 (0.8%)	6 (1.5%)	
Black or African American	13 (3.3%)	17 (4.4%)	
White	180 (46.3%)	170 (43.7%)	
Age			0.508
≤ 65	99 (20.1%)	92 (18.7%)	
> 65	146 (29.6%)	156 (31.6%)	
Smoker			0.223
No	6 (1.2%)	12 (2.4%)	
Yes	240 (49%)	232 (47.3%)	
OS event			0.321
Alive	149 (29.7%)	137 (27.3%)	
Dead	102 (20.3%)	114 (22.7%)	
DSS event			0.605
Alive	188 (41.8%)	173 (38.4%)	
Dead	43 (9.6%)	46 (10.2%)	
PFI event			0.922
Alive	178 (35.5%)	176 (35.1%)	
Dead	73 (14.5%)	75 (14.9%)	

Note: MYCT1, MYC Target 1; CR, Complete Response; PR, Partial Response; SD, Stable Disease; PD, Progressive Disease; OS, Overall Survival; DSS, Disease-Specific Survival; PFI, Progression-Free Interval; LUSC, Lung Squamous Cell Carcinoma.

prognosis in patients with LAC. *ARL14* silencing inhibits LAC cell proliferation, cell cycle progression, migration, and invasion ability. It also reduces radiation damage to cancer cells but does not affect normal lung cell proliferation. Interference with *ARL14* expression can effectively block the extracellular signal-regulated kinase (ERK)/p38 signaling pathway [21]. The expression of *CARD14* was higher in breast cancer samples than in normal breast tissues, and its inhibition can delay cell proliferation and migration, leading to cell cycle arrest in the G/S phase and promoting apoptosis [23]. *DKK4* is related to cancer progression and negatively regulates the Wnt/ β -catenin signaling pathway. The expression of *DKK4* in A549/DTX cells increased compared to that in A549 cells. *DKK4* overexpression increases the resistance of A549 cells to docetaxel, whereas interference with *DKK4* expression can inhibit growth and reduce colony formation and invasion properties of A549/DTX cells. Furthermore, because it is associated with caspase-3 activation and *BCL-2* down-regulation, *DKK4* suppression enhances docetaxel's apoptosis-promoting ability [26]. These findings suggest that the prognostic and diagnostic genes identified in our study play an important role in LUSC and may be able to predict the prognosis of patients with LUSC.

The expression of *FAM83A* in cervical cancer tissues is significantly increased compared to that in normal cervical tissues. This expression of *FAM83A* is related to the differentiation, stage of TNM, lymph node metastasis, and prognosis of cervical cancer. Interference with *FAM83A* expression can inhibit the proliferation, colony forma-

A risk model for lung squamous cell carcinoma

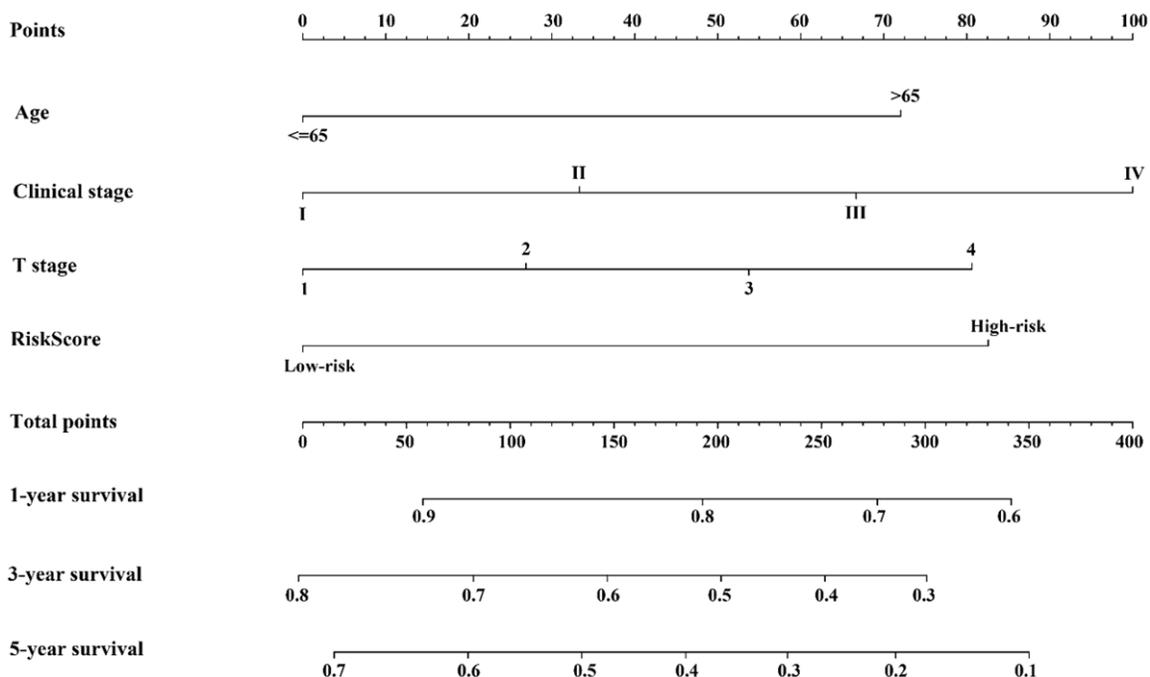


Figure 6. A prognosis-related risk model nomogram in LUSC. Note: LUSC, Lung Squamous Cell Carcinoma.

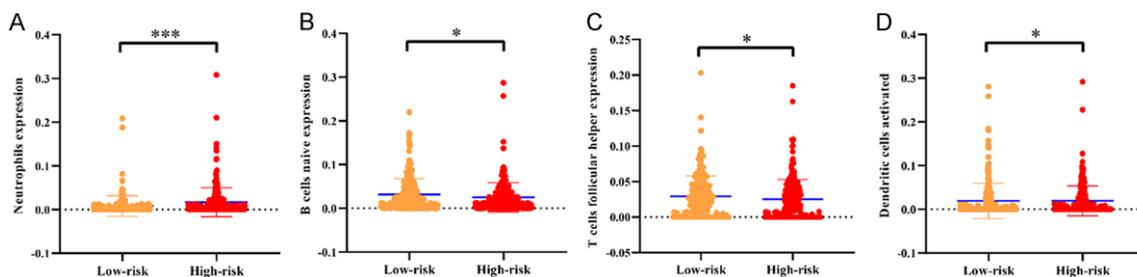


Figure 7. Immune cell infiltration in lung squamous cell carcinoma of patients with high- and low-risk scores. A. Neutrophils; B. Naïve B cells; C. T cells helper follicular; D. Activated dendritic cells.

tion, and invasion of cervical cancer cells. In lung cancer, overexpression of *FAM83A* promotes the epithelial-mesenchymal transition (EMT) and Wnt signaling pathways [30] and is associated with poor patient survival. Interference with *FAM83A* expression can inhibit the proliferation, migration, and invasion of H1355 and A549 lung cancer cells and promotes the inactivation of the epidermal growth factor receptor/mitogen-activated protein kinase (MAPK)/choline kinase alpha signaling pathway [29]. *MYCT1* inhibits the adhesion and migration of laryngeal cancer cells by regulating the expression of the *COL6* target [31]. Down-regulation of *CSTA* is associated with high tumor grade, lymph node metastasis, and

short OS in patients with oral squamous cell carcinoma (OSCC). *CTSA* overexpression can also inhibit OSCC cell migration and invasion *in vitro* [33]. *CSTA* is down-regulated in lung cancer cells compared to normal lung epithelial cells, and its high expression is correlated with low tumor grade. Stable *CSTA* transfection reduces cathepsin B activity; inhibits colony formation, migration, and invasion; and enhances gemcitabine-induced apoptosis. *CSTA* overexpression also reduces *ERK*, *p38*, and *AKT* activities and inhibits the *ERK*/*MAPK* pathway to block EMT [34]. Two of the genes that contribute to our risk score, *FAM83A*, and *CSTA*, have been reported in previous studies of lung cancer and LAC. However, *MYCT1* has

A risk model for lung squamous cell carcinoma

not been reported in the literature in lung cancer. Our results using TCGA and GEO data, as well as patient samples collected in our hospital, together with previous findings in the literature, indicate that *FAM83A*, *CSTA*, and *MYCT1* are of significant biological importance in LUSC.

The immune microenvironment is closely related to cancer progression, and immunotherapy is an effective approach that promises to improve cancer outcomes. The immune microenvironment can influence the efficacy of immunotherapy [35]. We analyzed the relationship between our risk score and the LUSC immune microenvironment and found that the risk score was significantly correlated with the levels of LUSC immune infiltration consisting of naïve B cells, helper follicular T cells, neutrophils, and activated dendritic cells. However, the relationship between our risk model and the immune microenvironment must be further confirmed using basic research.

Our study based on PCR results and data from the TCGA and GEO databases had a large sample size and it is therefore expected to be highly reliable and provides novel molecular targets for diagnosing LUSC and determining LUSC prognosis. A nomogram is also required to determine the prognoses of patients, and more research is needed to confirm our findings. The roles of immune cell infiltration in LUSC progression and metastasis are worth exploring. Therefore, we explore the functions and signaling mechanisms of *CSTA*, *FAM83A*, and *MYCT1* in the infiltration of certain immune cells based on the relationship between risk scores and immune cells identified in LUSC tissues. More LUSC tissue samples and clinical data will need to be collected to verify the clinical values of *ARL14*, *CARD14*, *DKK4*, *DSG3*, *KLK8*, *FAM83A*, *MYCT1*, and *CSTA* in LUSC. Additionally, a LUSC cell model must be constructed, and the impact of *CSTA*, *FAM83A*, and *MYCT1* on the growth and migration of LUSC cells must be validated through future proliferation, apoptosis, and migration experiments. However, our results show that *CSTA*, *FAM83A*, and *MYCT1* are abnormally expressed in LUSC tissues and significantly related to LUSC diagnosis and prognosis, suggesting that they may have potential as molecular targets for the treatment of LUSC. Furthermore, our risk model and the nomogram

based on *CSTA*, *FAM83A*, and *MYCT1* expression can potentially evaluate the prognosis of patients with LUSC.

Acknowledgements

This work was supported by the Wuhan Municipal Health Commission Foundation (No. wx21Q38).

Disclosure of conflict of interest

None.

Address correspondence to: Shan-Shan Chen, Key Laboratory for Molecular Diagnosis of Hubei Province, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, No. 26 Shengli Street, Jiang'an District, Wuhan 430014, Hubei, China. E-mail: chenshanshan@zxhospital.com; Chuang-Yan Wu, Department of Thoracic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, No. 1277 Jiefang Avenue, Jianghan District, Wuhan 430022, Hubei, China. E-mail: caffeein@hust.edu.cn

References

- [1] Duruisseaux M and Esteller M. Lung cancer epigenetics: from knowledge to applications. *Semin Cancer Biol* 2018; 51: 116-128.
- [2] Zhang YQ, Yuan Y, Zhang J, Lin CY, Guo JL, Liu HS and Guo Q. Evaluation of the roles and regulatory mechanisms of PD-1 target molecules in NSCLC progression. *Ann Transl Med* 2021; 9: 1168.
- [3] Ke D, Guo Q, Fan TY and Xiao X. Analysis of the role and regulation mechanism of hsa-mir-147b in lung squamous cell carcinoma based on the cancer genome atlas database. *Cancer Biother Radiopharm* 2021; 36: 280-291.
- [4] Woodman C, Vundu G, George A and Wilson CM. Applications and strategies in nanodiagnosis and nanotherapy in lung cancer. *Semin Cancer Biol* 2021; 69: 349-364.
- [5] Liu Y, Jia W, Li J, Zhu H and Yu J. Identification of survival-associated alternative splicing signatures in lung squamous cell carcinoma. *Front Oncol* 2020; 10: 587343.
- [6] Wu J, Xu C, Guan X, Ni D, Yang X, Yang Z and Wang M. Comprehensive analysis of tumor microenvironment and identification of an immune signature to predict the prognosis and immunotherapeutic response in lung squamous cell carcinoma. *Ann Transl Med* 2021; 9: 569.

A risk model for lung squamous cell carcinoma

- [7] Liu HS, Guo Q, Yang H, Zeng M, Xu LQ, Zhang QX, Liu H, Guo JL and Zhang J. SPDL1 overexpression is associated with the 18F-FDG PET/CT metabolic parameters, prognosis, and progression of esophageal cancer. *Front Genet* 2022; 13: 798020.
- [8] Haque S, Raina R, Afroze N, Hussain A, Alsulmani A, Singh V, Mishra BN, Kaul S and Kharwar RN. Microbial dysbiosis and epigenetics modulation in cancer development—a chemopreventive approach. *Semin Cancer Biol* 2021; 86: 666-681.
- [9] Ma J, Qi G and Li L. LncRNA NNT-AS1 promotes lung squamous cell carcinoma progression by regulating the miR-22/FOXM1 axis. *Cell Mol Biol Lett* 2020; 25: 34.
- [10] Li G and Guo X. LncRNA STARD13-AS blocks lung squamous carcinoma cells growth and movement by targeting miR-1248/C3A. *Pulm Pharmacol Ther* 2020; 64: 101949.
- [11] Li H, Zhao Q and Tang Z. LncRNA RP11-116G8.5 promotes the progression of lung squamous cell carcinoma through sponging miR-3150b-3p/miR-6870-5p to upregulate PHF12/FOXP4. *Pathol Res Pract* 2021; 226: 153566.
- [12] Shi H, Zhong F, Yi X, Shi Z, Ou F, Xu Z and Zuo Y. Application of an autophagy-related gene prognostic risk model based on TCGA database in cervical cancer. *Front Genet* 2020; 11: 616998.
- [13] Zhang Y, Tang M, Guo Q, Xu H, Yang Z and Li D. The value of erlotinib related target molecules in kidney renal cell carcinoma via bioinformatics analysis. *Gene* 2022; 816: 146173.
- [14] Sun L, Zhang Z, Yao Y, Li WY and Gu J. Analysis of expression differences of immune genes in non-small cell lung cancer based on TCGA and ImmPort data sets and the application of a prognostic model. *Ann Transl Med* 2020; 8: 550.
- [15] Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, Buttard B, Rothe F, Willard-Gallo K, Haller A, Ninane V, Burny A, Sculier JP, Spira A and Galon J. Immune evasion before tumour invasion in early lung squamous carcinogenesis. *Nature* 2019; 571: 570-575.
- [16] Wu Y, Wang J, Ge L and Hu Q. Significance of a PTEN mutational status-associated gene signature in the progression and prognosis of endometrial carcinoma. *Oxid Med Cell Longev* 2022; 2022: 5130648.
- [17] Liu J, Ren D, Du Z, Wang H, Zhang H and Jin Y. MA demethylase FTO facilitates tumor progression in lung squamous cell carcinoma by regulating MZF1 expression. *Biochem Biophys Res Commun* 2018; 502: 456-464.
- [18] Chen TJ, Zheng Q, Gao F, Yang T, Ren H, Li Y and Chen MW. MicroRNA-665 facilitates cell proliferation and represses apoptosis through modulating Wnt5a/ β -Catenin and Caspase-3 signaling pathways by targeting TRIM8 in LUSC. *Cancer Cell Int* 2021; 21: 215.
- [19] Wu Z, Liu Z, Jiang X, Mi Z, Meng M, Wang H, Zhao J, Zheng B and Yuan Z. Depleting PTOV1 sensitizes non-small cell lung cancer cells to chemotherapy through attenuating cancer stem cell traits. *J Exp Clin Cancer Res* 2019; 38: 341.
- [20] Takeda T, Komatsu M, Chiwaki F, Komatsuzaki R, Nakamura K, Tsuji K, Kobayashi Y, Tominaga E, Ono M, Banno K, Aoki D and Sasaki H. Up-regulation of IGF2R evades lysosomal dysfunction-induced apoptosis of cervical cancer cells via transport of cathepsins. *Cell Death Dis* 2019; 10: 876.
- [21] Guo F, Yuan D, Zhang J, Zhang H, Wang C, Zhu L, Zhang J, Pan Y and Shao C. Silencing of ARL14 gene induces lung adenocarcinoma cells to a dormant state. *Front Cell Dev Biol* 2019; 7: 238.
- [22] Lim JY, Kim SW, Kim B and Park SJ. Knockdown of CARD14 inhibits cell proliferation and migration in breast cancer cells. *Anticancer Res* 2020; 40: 1953-1962.
- [23] He S, Shen J, Hu N, Xu X and Li J. DKK4 enhances resistance to chemotherapeutics 5-Fu and YN968D1 in colorectal cancer cells. *Oncol Lett* 2017; 13: 587-592.
- [24] Yang X, Liu Y, Li W, Li A and Sun Q. DKK4-knockdown enhances chemosensitivity of A549/DTX cells to docetaxel. *Acta Biochim Biophys Sin (Shanghai)* 2017; 49: 899-906.
- [25] Cai X, Yao Z, Li L and Huang J. Role of DKK4 in tumorigenesis and tumor progression. *Int J Biol Sci* 2018; 14: 616-621.
- [26] Chao TB, Li CF, Lin CY, Tian YF, Chang IW, Sheu MJ, Lee YE, Chan TC and He HL. Prognostic significance of DSG3 in rectal adenocarcinoma treated with preoperative chemoradiotherapy. *Future Oncol* 2016; 12: 1457-67.
- [27] Chen YJ, Lee LY, Chao YK, Chang JT, Lu YC, Li HF, Chiu CC, Li YC, Li YL, Chiou JF and Cheng AJ. DSG3 facilitates cancer cell growth and invasion through the DSG3-plakoglobin-TCF/LEF-Myc/cyclin D1/MMP signaling pathway. *PLoS One* 2013; 8: e64088.
- [28] Zhang M, Luo C, Lin D, Cui K, Chen Z and Liu J. Human tissue kallikrein 1 is downregulated in elderly human prostates and possesses potential in vitro antioxidative and antifibrotic effects in rodent prostates. *Oxid Med Cell Longev* 2021; 2021: 8877540.
- [29] Liu PJ, Chen YH, Tsai KW, Yeah HY, Yeh CY, Tu YT and Yang CY. Involvement of MicroRNA-1-FAM83A axis dysfunction in the growth and motility of lung cancer cells. *Int J Mol Sci* 2020; 21: 8833.

A risk model for lung squamous cell carcinoma

- [30] Lan C, Liu CC, Nie XC, Lei L, Xiao ZX, Li MX, Tang XN, Jia MY and Xu HT. FAM83A promotes the proliferative and invasive abilities of cervical cancer cells via epithelial-mesenchymal transition and the wnt signaling pathway. *J Cancer* 2021; 12: 6320-6329.
- [31] Wang PP, Ding SY, Sun YY, Li YH and Fu WN. MYCT1 inhibits the adhesion and migration of laryngeal cancer cells potentially through repressing collagen VI. *Front Oncol* 2020; 10: 564733.
- [32] Kabir AU, Subramanian M, Lee DH, Wang X, Krchma K, Wu J, Naismith T, Halabi CM, Kim JY, Pulous FE, Petrich BG, Kim S, Park HC, Hanson PI, Pan H, Wickline SA, Fremont DH, Park C and Choi K. Dual role of endothelial Myct in tumor angiogenesis and tumor immunity. *Sci Transl Med* 2021; 13: eabb6731.
- [33] Wang Y, Wang L, Li X, Qu X, Han N, Ruan M and Zhang C. Decreased CSTA expression promotes lymphatic metastasis and predicts poor survival in oral squamous cell carcinoma. *Arch Oral Biol* 2021; 126: 105116.
- [34] Ma Y, Chen Y, Li Y, Grün K, Berndt A, Zhou Z and Petersen I. Cystatin A suppresses tumor cell growth through inhibiting epithelial to mesenchymal transition in human lung cancer. *Oncotarget* 2018; 9: 14084-14098.
- [35] Yu L, Ding Y, Wan T, Deng T, Huang H and Liu J. Significance of CD47 and its association with tumor immune microenvironment heterogeneity in ovarian cancer. *Front Immunol* 2021; 12: 768115.

A risk model for lung squamous cell carcinoma

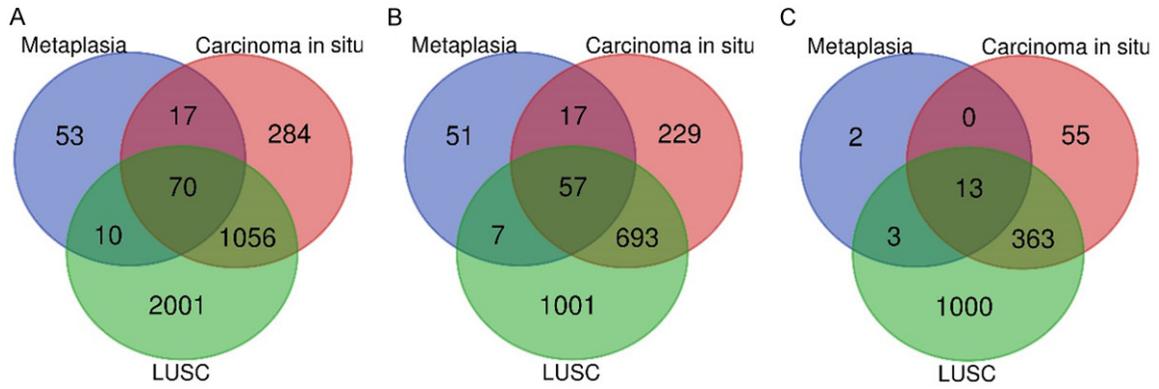


Figure S1. The overlapping genes in the progression of LUSC using a VEEN diagram. A. All overlapping DEGs; B. Overlapping overexpressed genes; C. Overlapping down-regulated genes. Note: LUSC, Lung Squamous Cell Carcinoma.

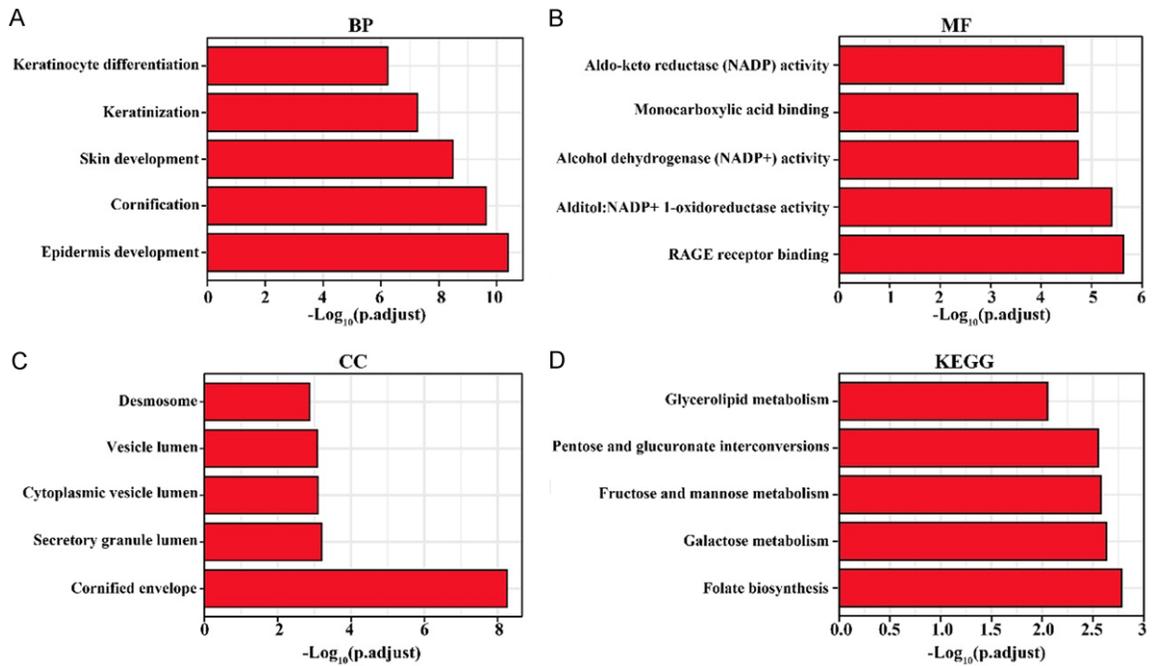


Figure S2. The roles and signaling mechanisms of DEG enrichment. A. BP; B. MF; C. CC; D. KEGG. Note: DEGs, Differentially Expressed Genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, Biological Processes; CC, Cellular Components; MF, Molecular Functions.

A risk model for lung squamous cell carcinoma

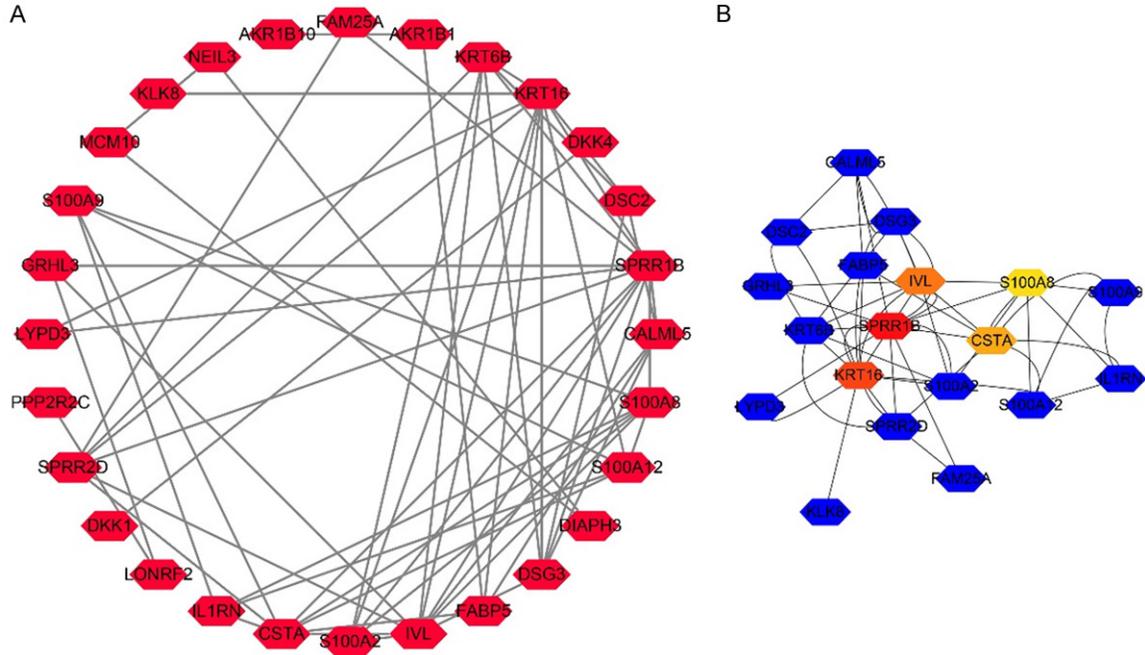
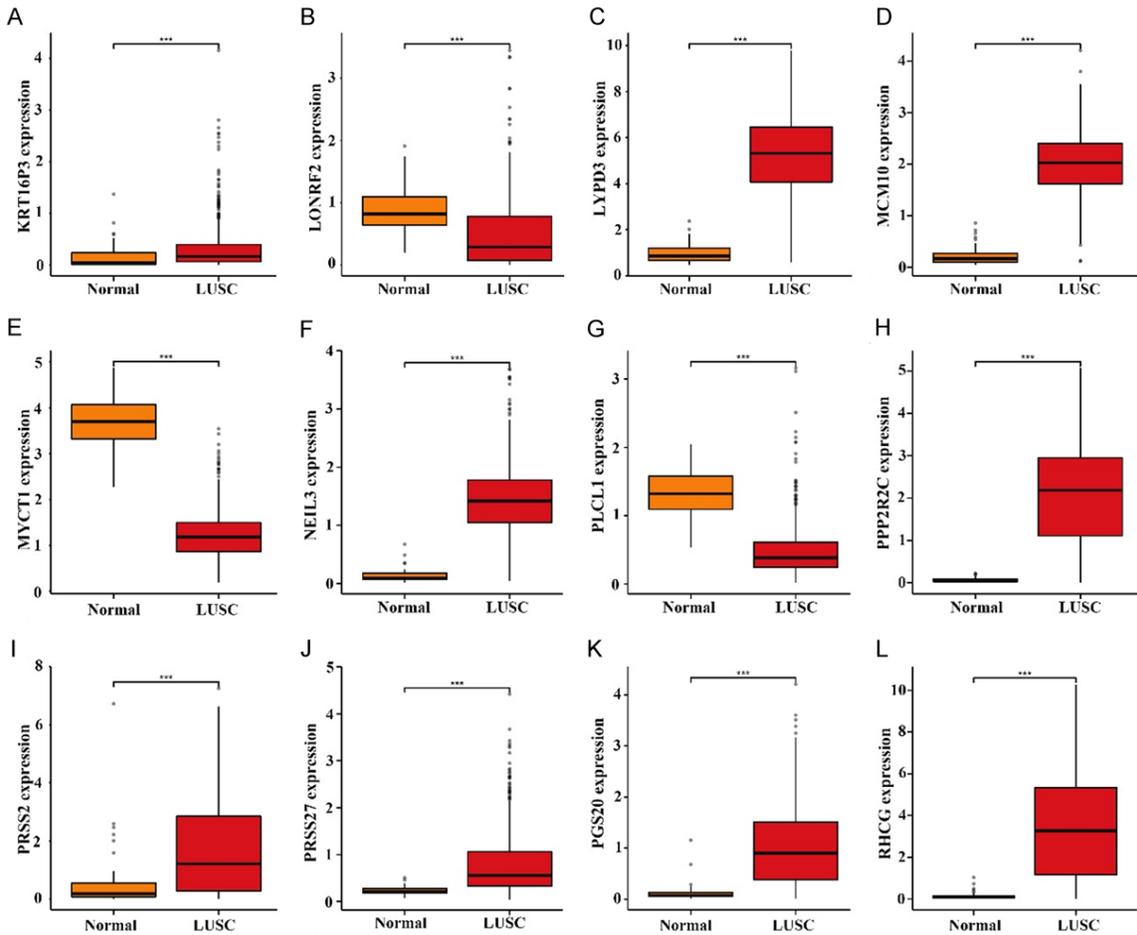


Figure S3. The PPI network of the DEGs. A. PPI network; B. Hub genes of the PPI network. Note: PPI, Protein-Protein Interaction, DEGs, Differentially Expressed Genes.



A risk model for lung squamous cell carcinoma

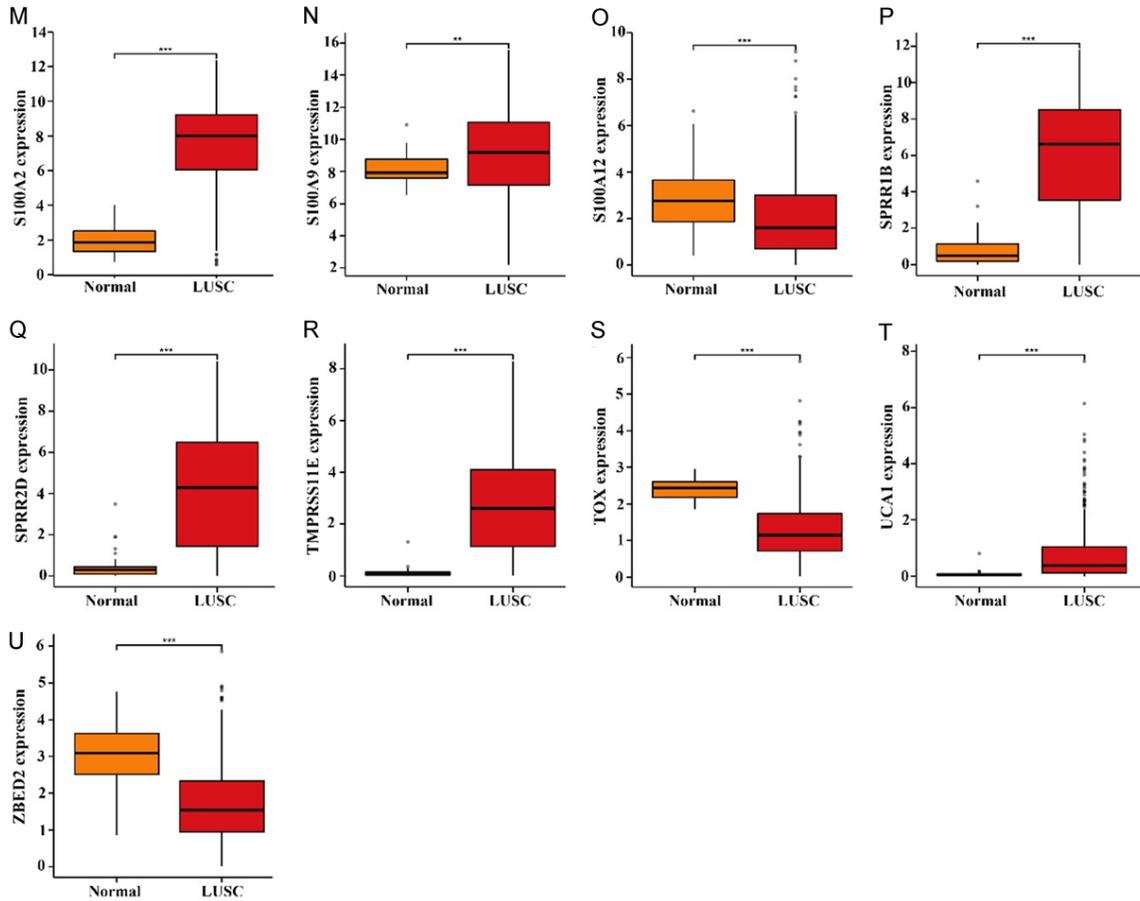


Figure S4. The DEGs in unpaired LUSC tissues from the TCGA database. A. KRT16P3; B. LONRF2; C. LYPD3; D. MCM10; E. MYCT1; F. NEIL3; G. PLCL1; H. PPP2R2C; I. PRSS2; J. PRSS27; K. RGS20; L. RHCG; M. S100A2; N. S100A9; O. S100A12; P. SPRR1B; Q. SPRR2D; R. TMPRSS11E; S. TOX; T. UCA1; U. ZBED2. Note: Degs, Differentially Expressed Genes; LUSC, Lung Squamous Cell Carcinoma; TCGA, The Cancer Genome Atlas; KRT16P3, Keratin 16 Pseudogene 3; LONRF2, LON Peptidase N-Terminal Domain and Ring Finger 2; LYPD3, LY6/PLAUR Domain Containing 3; MCM10, Minichromosome Maintenance 10 Replication Initiation Factor; MYCT1, MYC Target 1; NEIL3, Nei Like DNA Glycosylase 3; PLCL1, Phospholipase C Like 1; PPP2R2C, Protein Phosphatase 2 Regulatory Subunit Bgamma; PRSS2, Serine Protease 2; PRSS27, Serine Protease 27; RGS20, Regulator Of G Protein Signaling 20; RHCG, Rh Family C Glycoprotein; S100A2, S100 Calcium Binding Protein A2; S100A9, S100 Calcium Binding Protein A9; S100A12, S100 Calcium Binding Protein A12; SPRR1B, Small Proline Rich Protein 1B; SPRR2D, Small Proline Rich Protein 2D; TMPRSS11E, Transmembrane Serine Protease 11E; TOX, Thymocyte Selection Associated High Mobility Group Box; UCA1, Urothelial Cancer Associated 1; ZBED2, Zinc Finger BED-Type Containing 2.

A risk model for lung squamous cell carcinoma

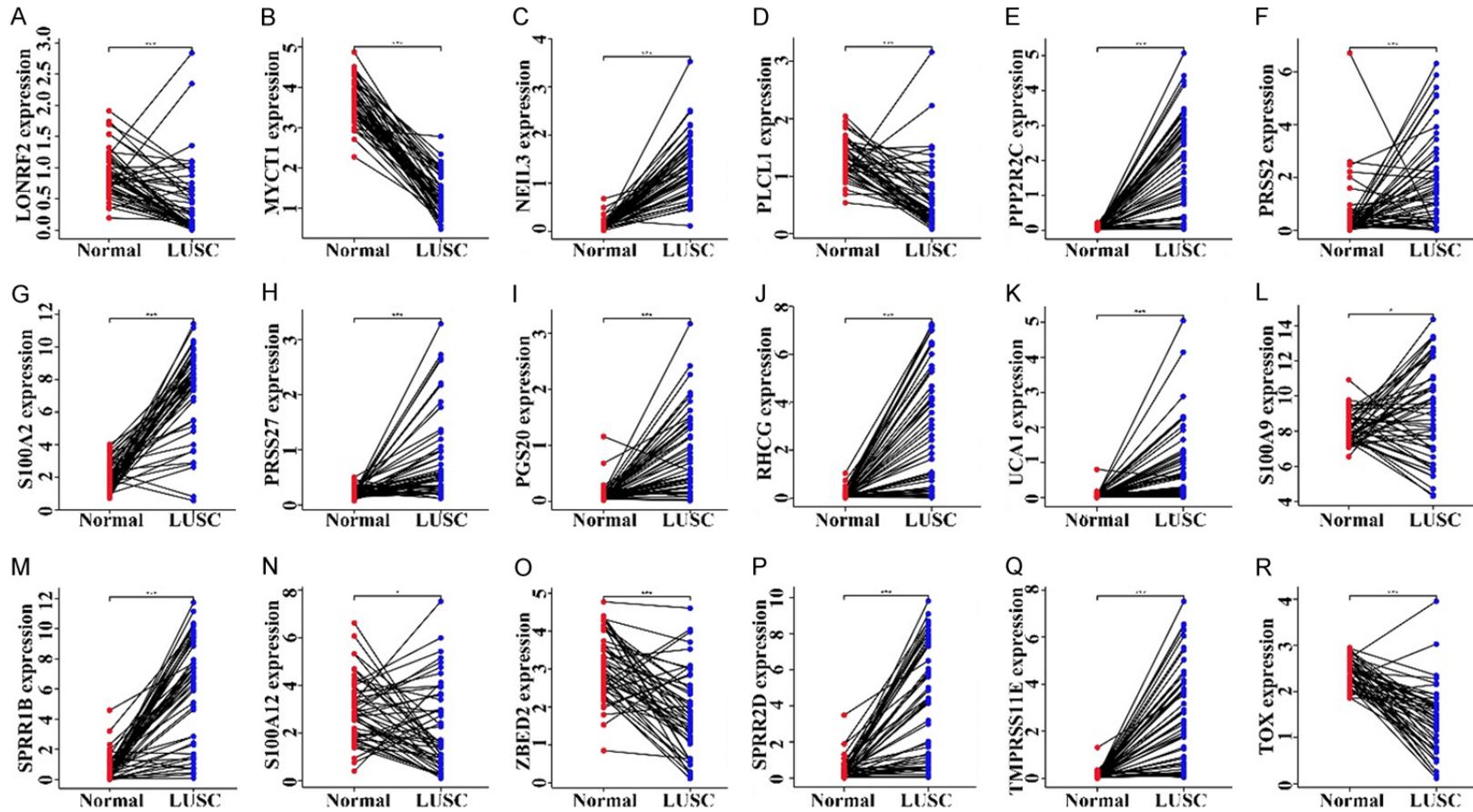


Figure S5. DEGs in paired LUSC tissues of the TCGA database. A. LONRF2; B. MYCT1; C. NEIL3; D. PLCL1; E. PPP2R2C; F. PRSS2; G. S100A2; H. PRSS27; I. RGS20; J. RHCG; K. UCA1; L. S100A9; M. SPRR1B; N. S100A12; O. ZBED2; P. SPRR2D; Q. TMPRSS11E; R. TOX. Note: Degs, Differentially Expressed Genes; LUSC, Lung Squamous Cell Carcinoma; TCGA, The Cancer Genome Atlas; LONRF2, LON Peptidase N-Terminal Domain and Ring Finger 2; MYCT1, MYC Target 1; NEIL3, Nei Like DNA Glycosylase 3; PLCL1, Phospholipase C Like 1; PPP2R2C, Protein Phosphatase 2 Regulatory Subunit Bgamma; PRSS2, Serine Protease 2; PRSS27, Serine Protease 27; RGS20, Regulator Of G Protein Signaling 20; RHCG, Rh Family C Glycoprotein; S100A2, S100 Calcium Binding Protein A2; S100A9, S100 Calcium Binding Protein A9; S100A12, S100 Calcium Binding Protein A12; SPRR1B, Small Proline Rich Protein 1B; SPRR2D, Small Proline Rich Protein 2D; TMPRSS11E, Transmembrane Serine Protease 11E; TOX, Thymocyte Selection Associated High Mobility Group Box; UCA1, Urothelial Cancer Associated 1; ZBED2, Zinc Finger BED-Type Containing 2.

A risk model for lung squamous cell carcinoma

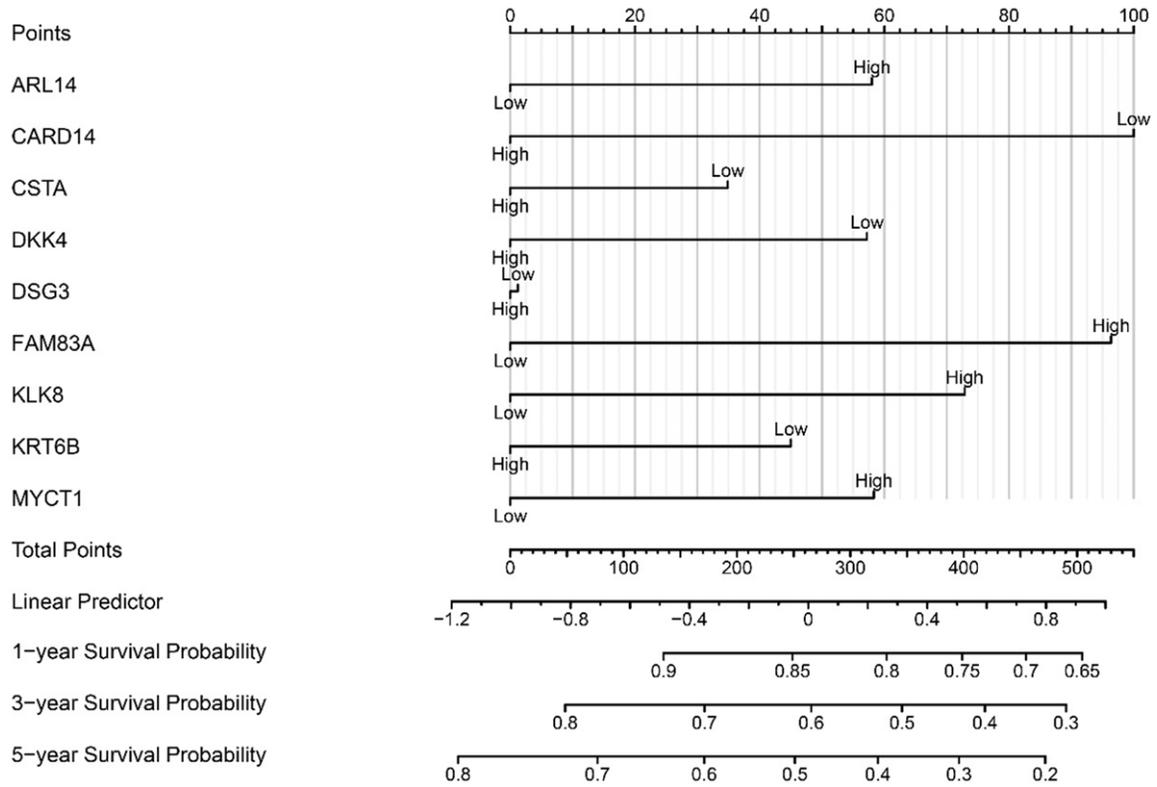


Figure S6. Construction of a nomogram for the prognosis and diagnosis-related DEGs in LUSC. Note: Degs, Differentially Expressed Genes; LUSC, Lung Squamous Cell Carcinoma; ARL14, ADP Ribosylation Factor Like Gtpase 14; CARD14, Caspase Recruitment Domain Family Member 14; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; KLK8, Kallikrein Related Peptidase 8; KRT6B, Keratin 6B; DKK4, Dickkopf WNT Signaling Pathway Inhibitor 4; DSG3, Desmoglein 3; MYCT1, MYC Target 1.

A risk model for lung squamous cell carcinoma

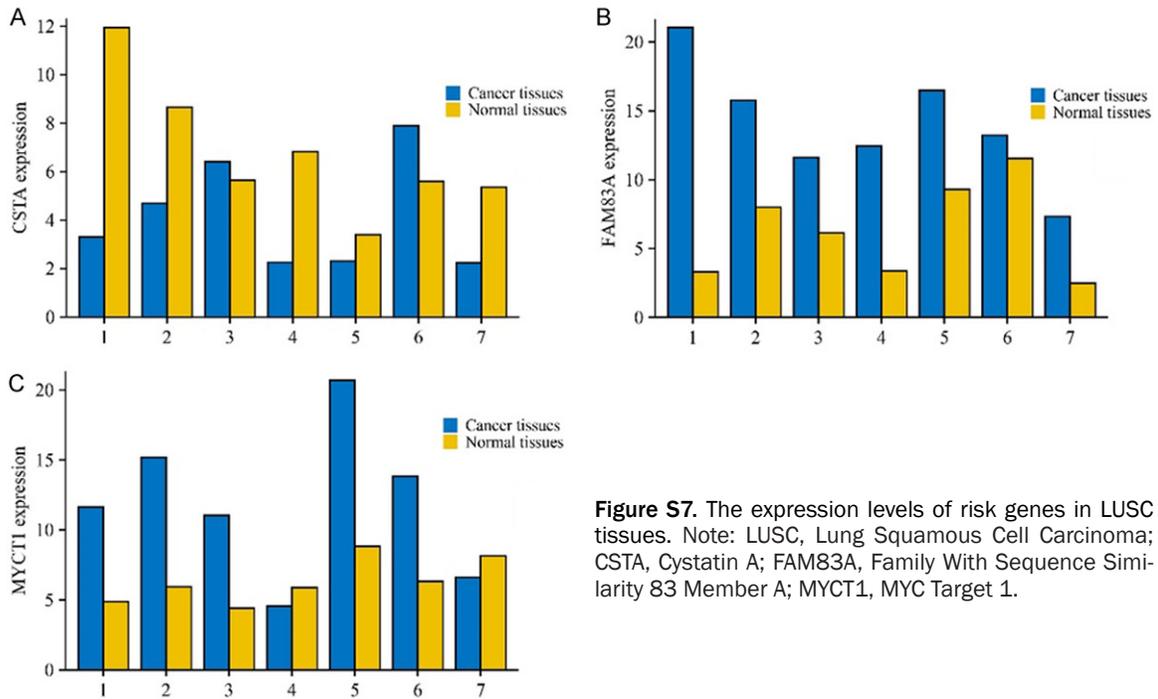


Figure S7. The expression levels of risk genes in LUSC tissues. Note: LUSC, Lung Squamous Cell Carcinoma; CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; MYCT1, MYC Target 1.

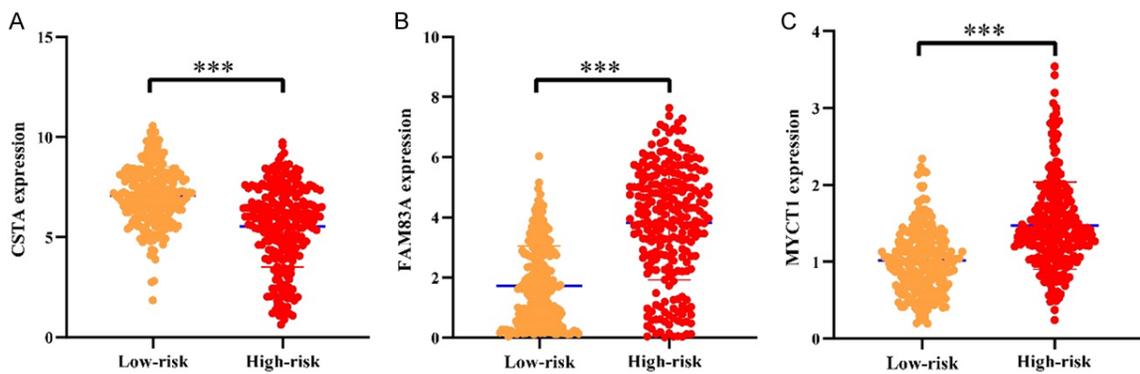


Figure S8. The expression levels of the risk model factor in cancer tissues from high- and low-risk patients. A. CSTA; B. FAM83A; C. MYCT1. Note: CSTA, Cystatin A; FAM83A, Family With Sequence Similarity 83 Member A; MYCT1, MYC Target 1.