Supplementary Material

1. Supplementary Methods

1.1 Machine learning algorithms developed a prognostic model [1]

10 machine learning algorithms (stepwise Cox, random survival forest [RSF], elastic network [Enet], supervised principal components [SuperPC], partial least squares regression for Cox [plsRcox], CoxBoost, survival support vector machine [survival-SVM], Lasso, Ridge and generalized boosted regression modeling [GBM]) were integrated into 100 kinds of machine learning combination. The stepwise Cox model was preformed via survival package. Stepwise Cox regression is a stepwise variable selection process based on the Akaike Information Criterion (AIC), identifying variables that significantly influence survival time to construct an optimal Cox proportional hazards model. RSF was preformed via randomForestSRC package. RSF included two parameters, ntree and mtry. The ntree was the number of trees in the forest and mtry represented the number of variables randomly selected for splitting at each node. A grid search was conducted for ntree and mtry using 10-fold cross-validation. All pairs of (ntree, mtry) were obtained, and the pair with the highest C-index value was determined as the optimization parameter. Enet, Lasso and Ridge were preformed via glmnet package. All the three algorithms contribute to addressing the problem of overfitting. SuperPC was implemented via superpc package. SuperPC is an improvement of principal component analysis (PCA), focusing on dimension reduction of variables that are closely associated with survival time. The plsRcox model was implemented via plsRcox package. The plsRcox extracts the most relevant features and constructs a generalized linear model using partial least squares regression. CoxBoost was preformed via CoxBoost package. CoxBoost is a Cox regression model based on boosting algorithms, fitting the Cox proportional hazards model through componentwise likelihood-based boosting. Survival-SVM was preformed via survivalsvm package. Survival-SVM is an extension of SVM for survival data analysis, predicting survival time and risk levels by maximizing the decision boundary of the data. The GBM model was implemented via the gbm package. GBM is an ensemble learning method that iteratively combines multiple weak learners, typically decision trees, to enhance the accuracy and robustness of model.

1.2 Cell culture

The culture protocol for BLCA cells is described below. Four BLCA cell lines (T24, UMUC3, RT4, and BIU87) and normal cells (SVHUC) were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). BIU87 cells were treated with 10% PMI-1640 medium, T24 cells with 10%McCoy's 5A medium, RT4 cells with 10% DMEM medium, and UM-UC-3 cells with 10% MEM medium. SV-HUC-1 was cultured using 10% Ham's F-12K medium in a constant temperature incubator at 37° C with 5% CO2 volume fraction using culture bottles. All were cultured at 37°C in a humidified atmosphere containing 5% CO2.

1.3 ceRNA related database

TargetScan: https://www.targetscan.org/vert_80/ [2]

ENCORI: https://rnasysu.com/encori/ [3]

miRWalk: http://mirwalk.umm.uni-heidelberg.de/ [4]

miRDB: https://mirdb.org/ [5]

2. Supplementary Tables

Supplementary Table1: clinical data of TCGA-BLCA

Supplementary Table2: clinical data of GSE13507

Supplementary Table3: clinical data of GSE31684

Supplementary Table4: clinical data of GSE32894

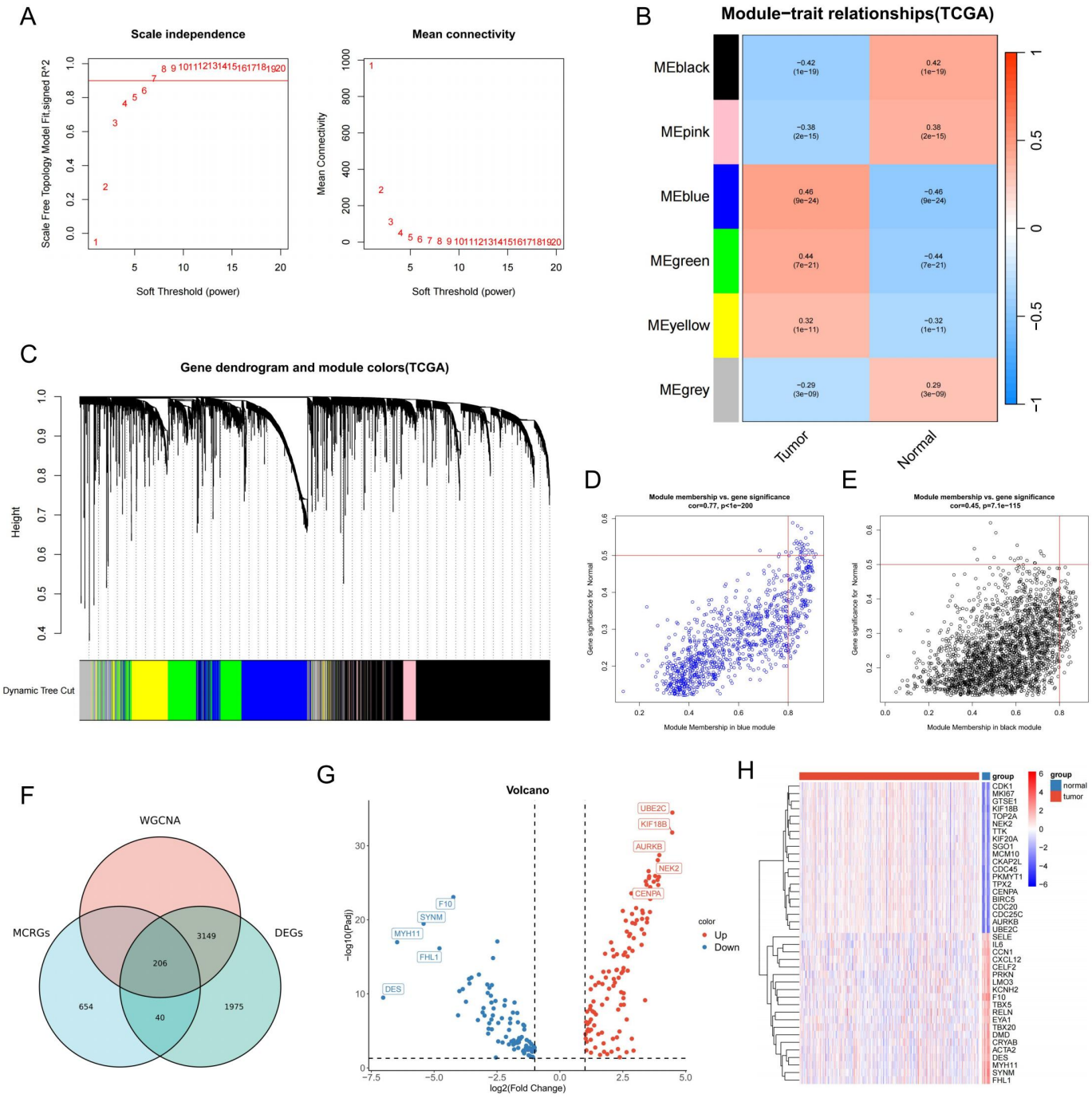
Supplementary Table5: clinical data of GSE135337

Patient ID	Sex	Age	Histology type	Tumor grade	Tumor stage	Basal/luminal	Tumor size(cm)	Surgically resection
BC1	Male	67	UCC	Low	pTa	Luminal	<3cm	TURBT
BC2	Male	67	UCC	High	pT1	Luminal	<3cm	RC
BC3	Male	38	UCC	High	pT1	Luminal	>3cm	TURBT
BC4	Male	80	UCC	High	pT2N0M0	Basal	<3cm	RC
BC5	Male	81	UCC	High	pT3N0M0	Luminal	<3cm	RC
BC6	Male	58	UCC	Low	pTa	Luminal	<3cm	TURBT
BC7	Male	66	UCC	Low	pTa	Luminal	<3cm	TURBT

All tumors are newly diagnosed and treatment-naive.

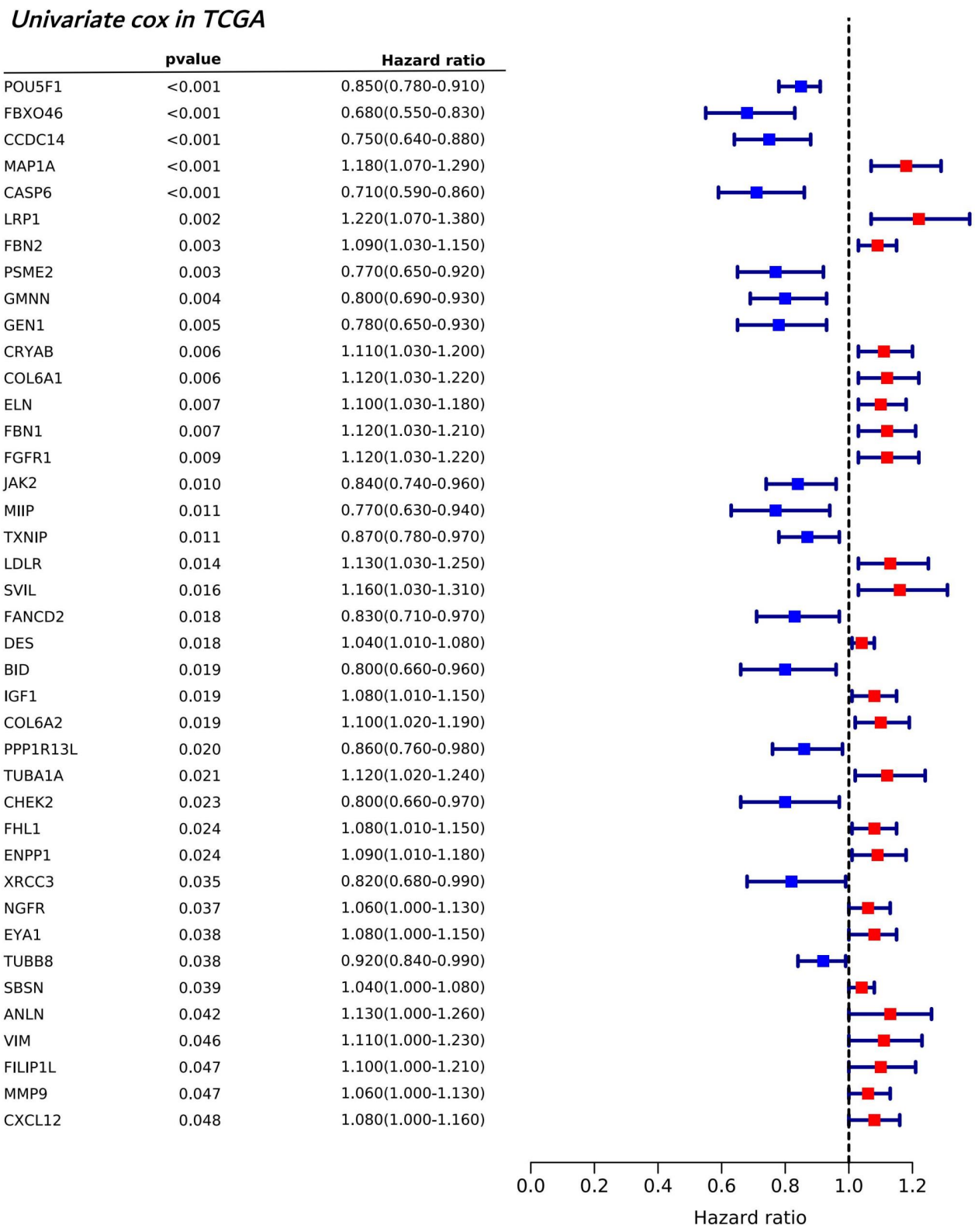
UCC: urothelial cell carcinoma; TURBT: transurethral resection of bladder tumor; RC: radical cystectomy

3. Supplementary Figures



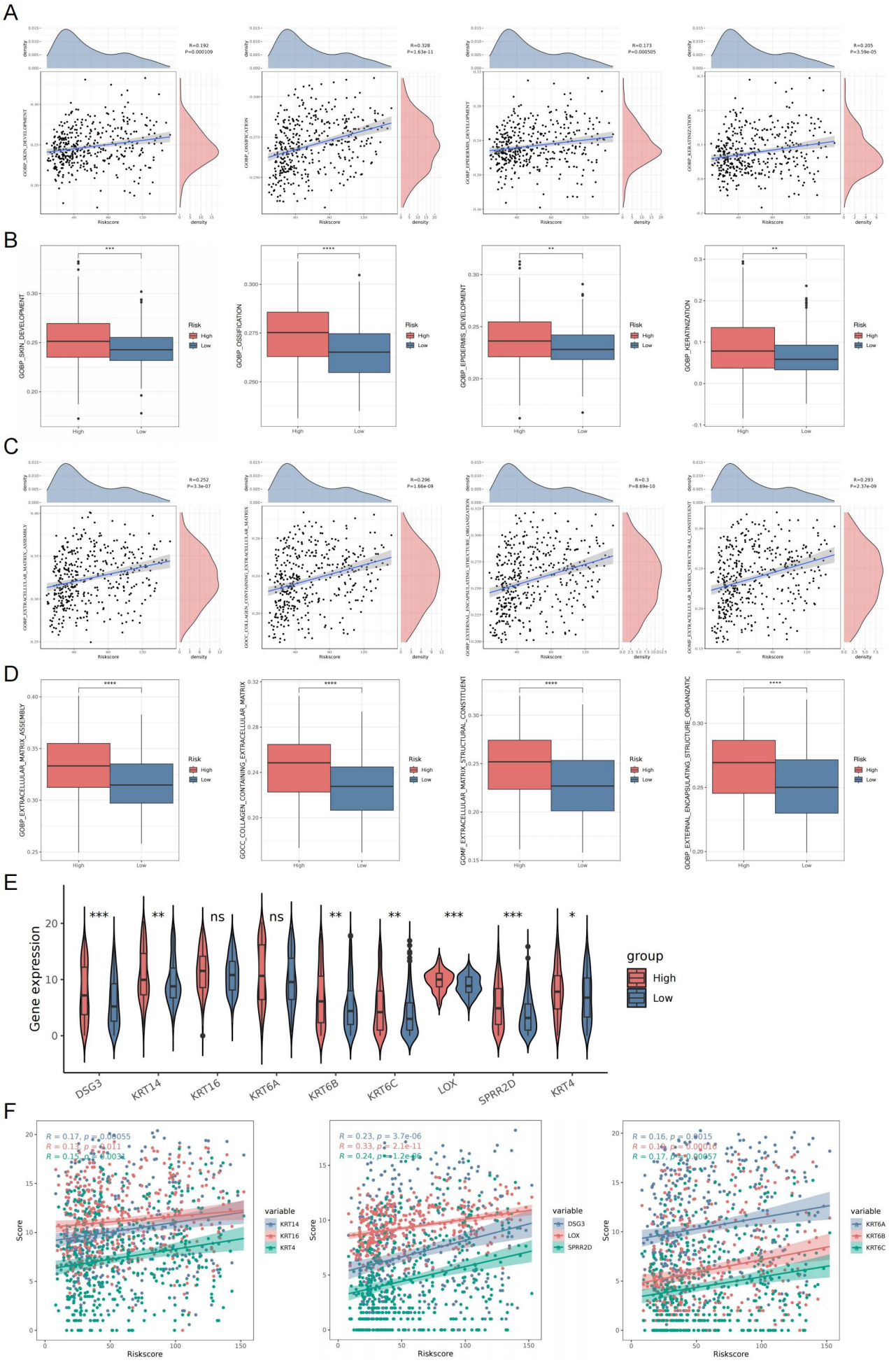
Supplementary Fig S1

WGCNA (A) Identification of soft threshold (B) Heatmap of gene modules correlating with disease phenotypes (C) Gene clustering tree (D-E) Significance of gene modules (F) Identification of key dysregulated MCRGs (G) Volcano plot of key dysfunctional MCRGs (H) Heatmap of key dysfunctional MCRGs



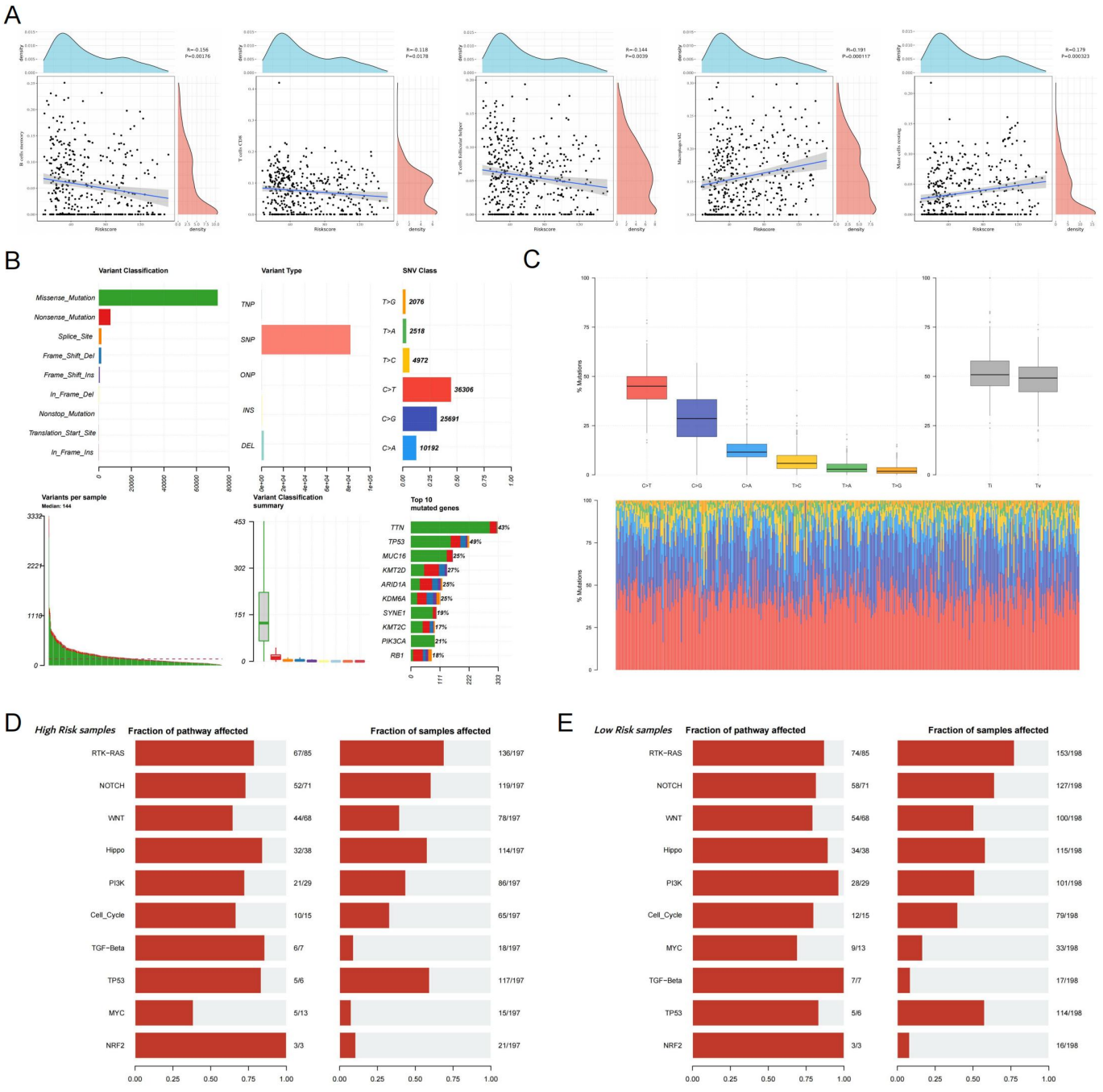
Supplementary Fig S2

Univariate cox regression analysis based on key dysregulated MCRGs.



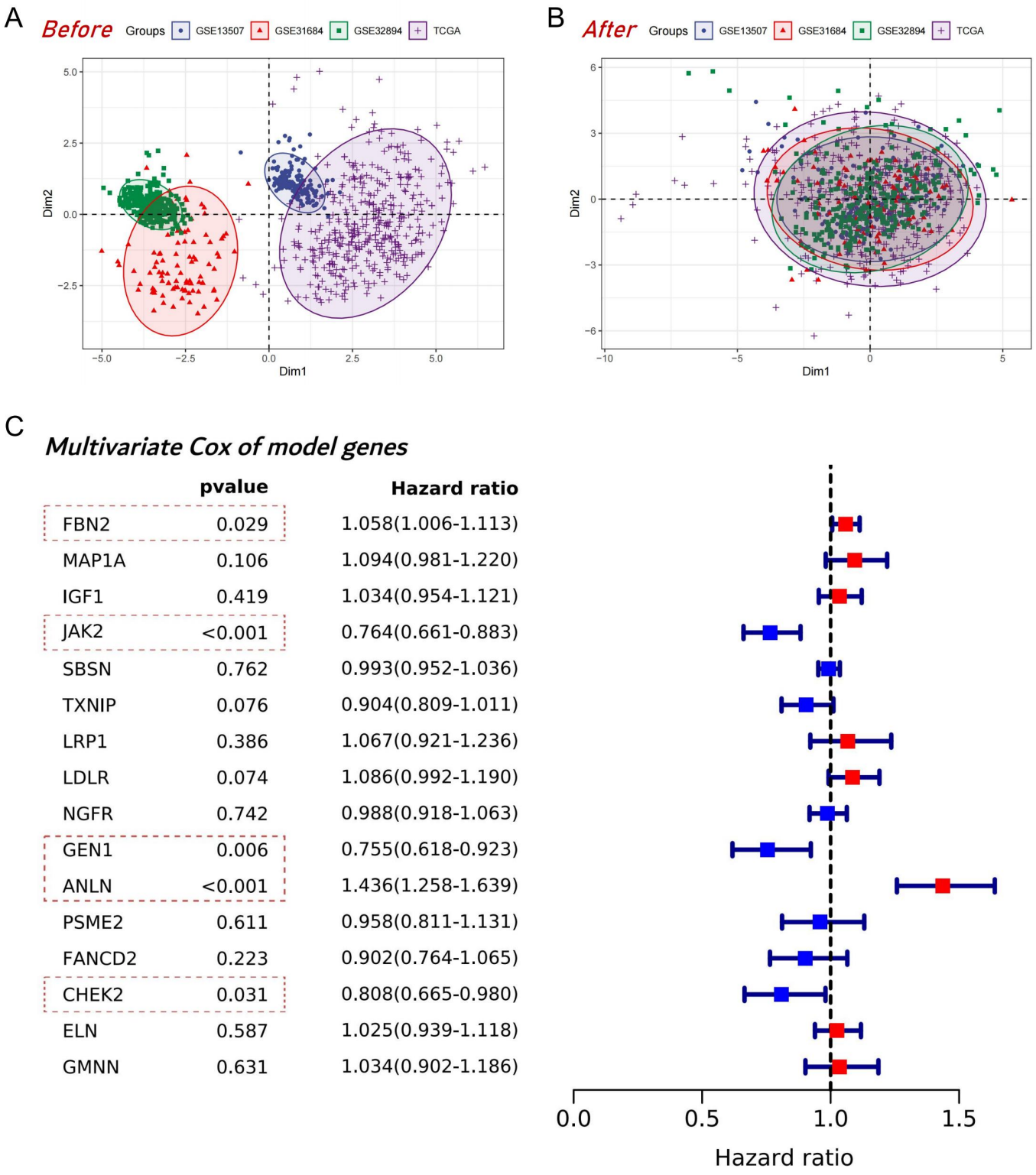
Supplementary Fig S3

Correlation of risk scores with squamous carcinoma characterization and extracellular mesenchymal activity (A) Correlation of squamous cancerization-related features with Riskscore (B) Differences of squamous cancerization-related features between high- and low-risk groups (C) Correlation of extracellular matrix activation-related characteristics with Riskscore (D) Differences of extracellular matrix activation-related characteristics between high- and low-risk groups (E) Differential expression of squamous carcinogenicity markers in high and low risk groups (F) Correlation of squamous carcinogenicity markers with Riskscore



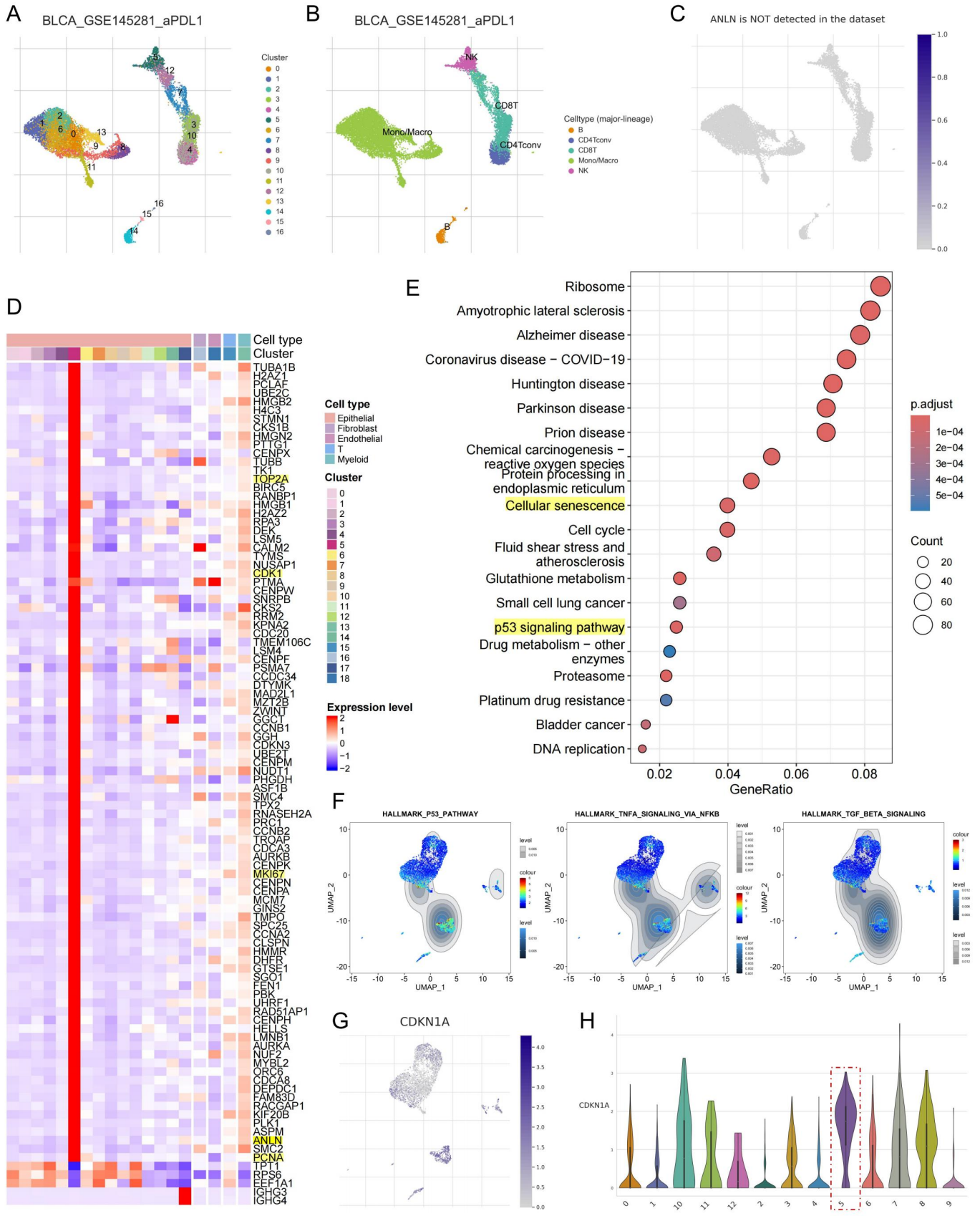
Supplementary Fig S4

(A) Correlations of several immune cells and riskscore (B) Overall mutation statistics (C) Base mutation direction statistics (D-E) Pathway mutation rate in high-risk group and low-risk group



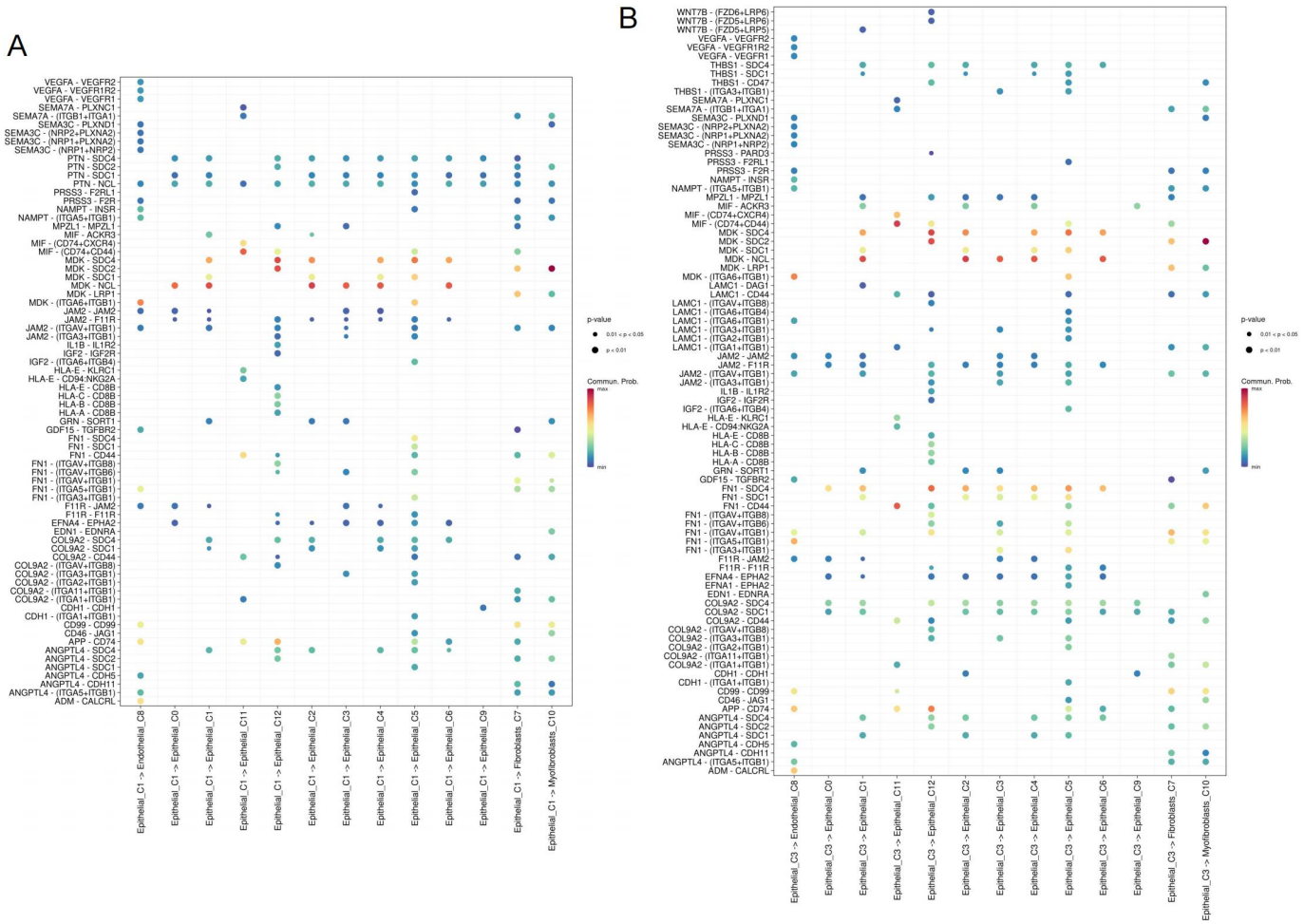
Supplementary Fig S5

De-batching and Multivariate Cox regression of model genes. (A) Distribution of datasets before de-batching (B) Distribution of data sets after de-batching (C) Forest plot of multivariate cox regression



Supplementary Fig S6

Single cell analysis (A) Clusters of GSE145281 (B) Cell type of GSE145281 (C) ANLN expression in GSE145281 (D) Heatmap of 98 DEGs (E) GO enrichment of DEGs (F) Function Score Landscape (G) Expression of CDKN1A (H) Violin plot of CDKN1A expression

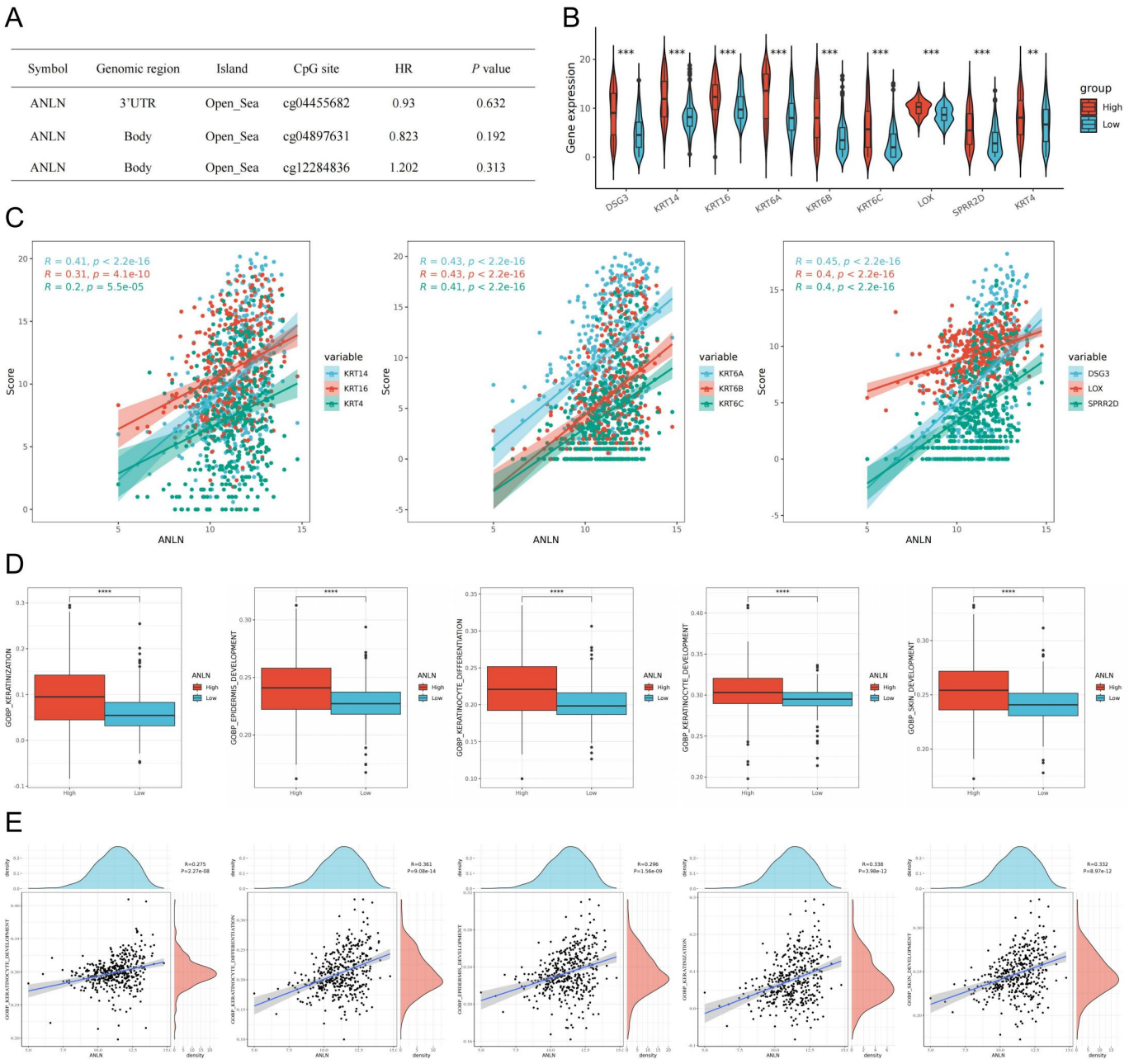


Supplementary Fig S7

Intercellular communication analysis (A) Intercellular communication between C1 and other cell clusters (B) Intercellular communication between C3 and other cell clusters

A

Symbol	Genomic region	Island	CpG site	HR	*P* value
ANLN	3'UTR	Open_Sea	cg04455682	0.93	0.632
ANLN	Body	Open_Sea	cg04897631	0.823	0.192
ANLN	Body	Open_Sea	cg12284836	1.202	0.313



Supplementary Fig S8

Methylation and squamous carcinoma characterization of ANLN (A) Methylation sites of ANLN (B) Differences in squamous carcinogenicity marker expression between high and low ANLN expression groups (C) Correlation of squamous carcinogenicity markers with ANLN (D) Differences in squamous carcinoma-related scores between high and low ANLN expression groups (E) Correlation of squamous carcinomatosis-related scores with ANLN

Reference:

1. Zhan Y, Weng M, Guo Y, Lv D, Zhao F, Yan Z, Jiang J, Xiao Y, Yao L. Identification and validation of the nicotine metabolism-related signature of bladder cancer by bioinformatics and machine learning. Front Immunol. 2024 Dec 17;15:1465638.

2. McGeary SE, Lin KS, Shi CY, Pham TM, Bisaria N, Kelley GM, Bartel DP. The biochemical basis of microRNA targeting efficacy. Science. 2019 Dec 20;366(6472):eaav1741. doi: 10.1126/science.aav1741. Epub 2019 Dec 5.

3. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2014 Jan;42(Database issue):D92-7.

4. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. PLoS One. 2018 Oct 18;13(10):e0206239. doi: 10.1371/journal.pone.0206239.

5. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020 Jan 8;48(D1):D127-D131.