Integrated bioinformatics analysis of differentially expressed genes and immune cell infiltration characteristics in Esophageal Squamous cell carcinoma

JCM, 2.11.2021

Esophageal Squamous cell carcinoma

1. Esophageal cancer is the seventh most common cancer worldwide, with an estimated 572,034 new cases and 508,585 deaths occurring in 2018.

Esophageal Squamous cell carcinoma

1. Esophageal cancer is the seventh most common cancer worldwide, with an estimated 572,034 new cases and 508,585 deaths occurring in 2018.
2. Esophageal squamous cell carcinoma (ESCC) accounts for approximately 90% of new incident esophageal cancers each year.

Esophageal Squamous cell carcinoma

1. Esophageal cancer is the seventh most common cancer worldwide, with an estimated 572,034 new cases and 508,585 deaths occurring in 2018.
2. Esophageal squamous cell carcinoma (ESCC) accounts for approximately 90% of new incident esophageal cancers each year.
3. Due to its inconspicuous symptoms and inadequate endoscopic screening, esophageal cancer is often diagnosed at an advanced stage, and the 5-year overall survival (OS) rate ranges from 12 to 20%

Aim of the study

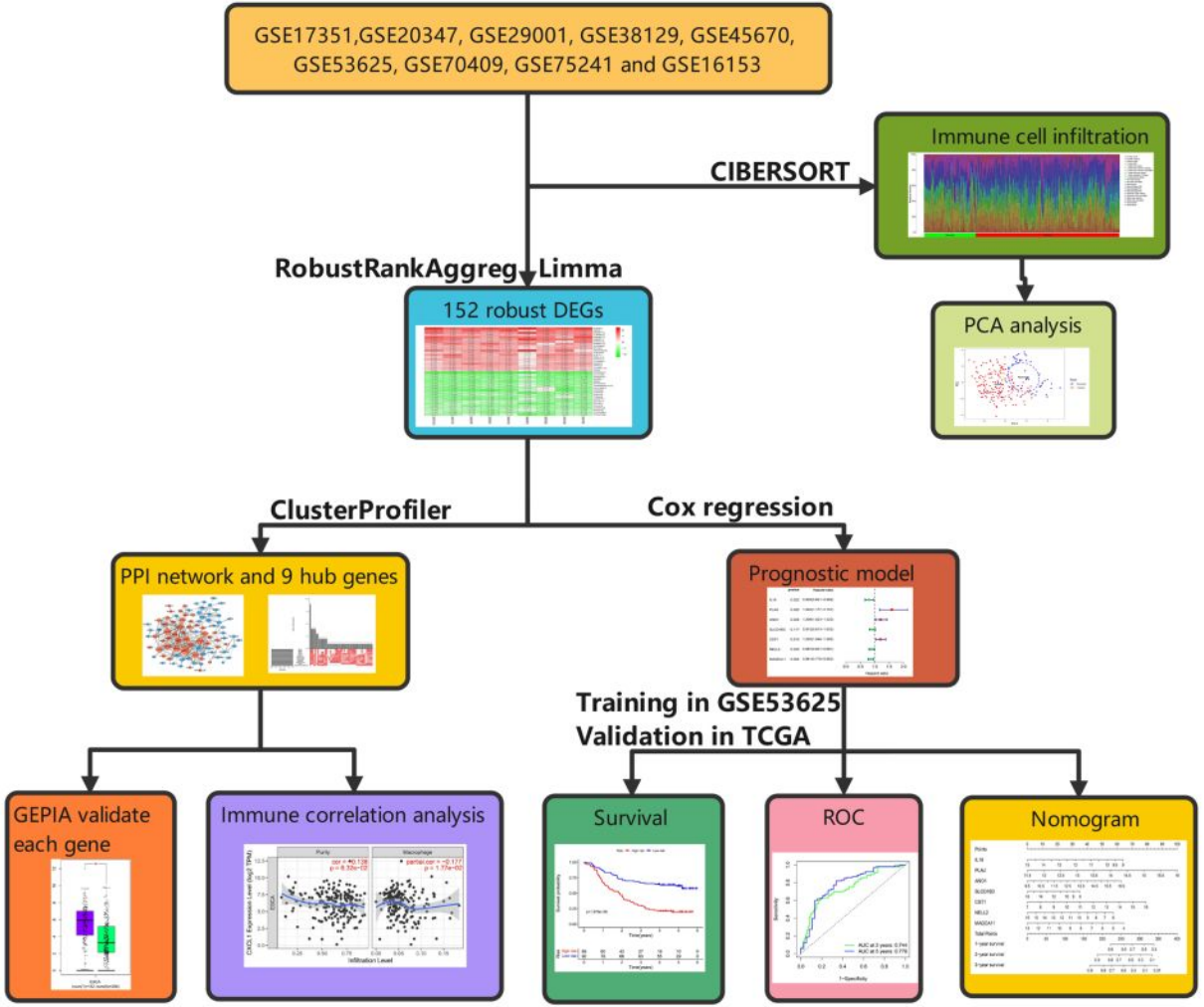
1. This study aimed to identify robust DEGs and characterize the immune cell infiltration distribution in ESCC from as many datasets as possible.

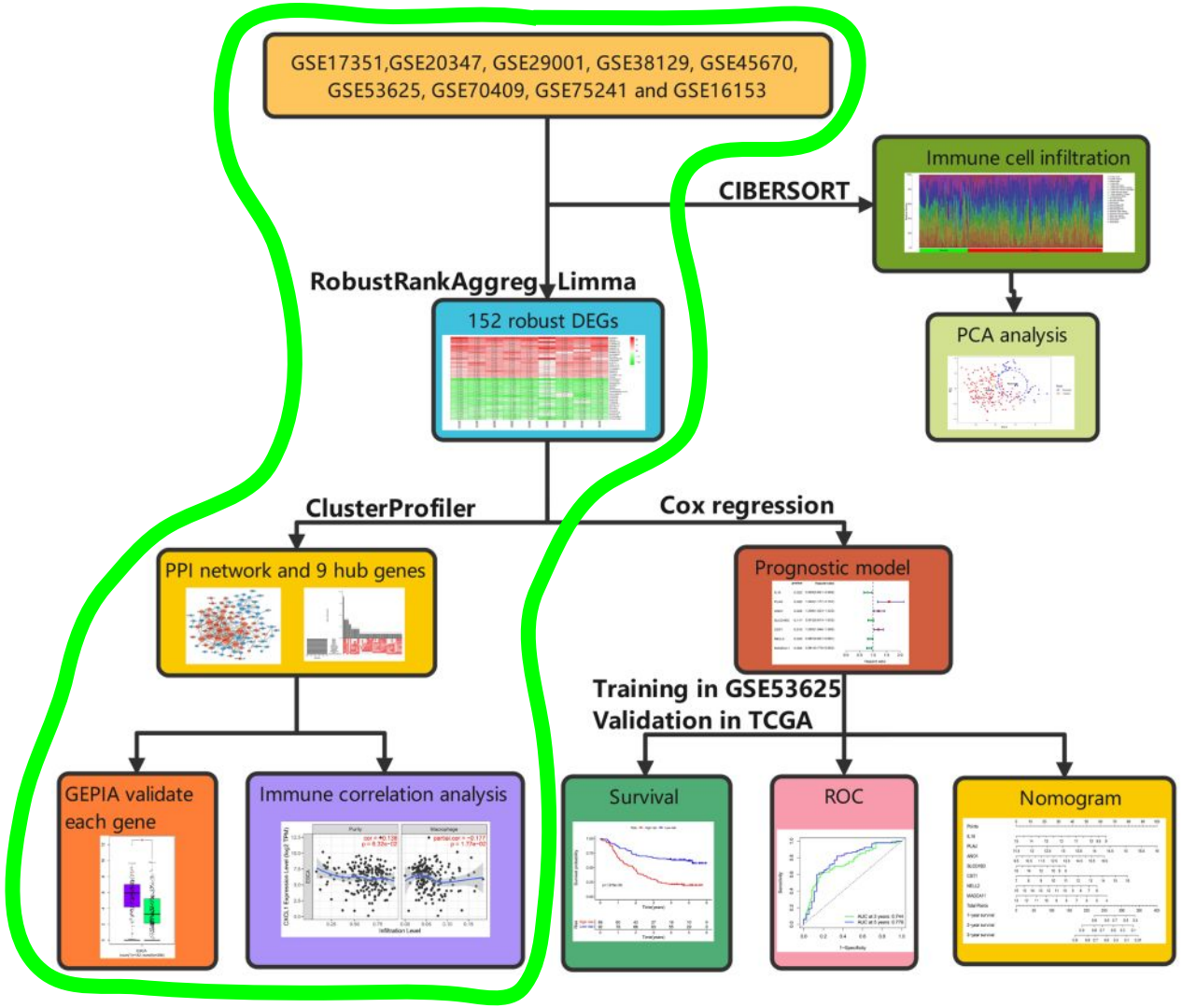
Aim of the study

1. This study aimed to identify robust DEGs and characterize the immune cell infiltration distribution in ESCC from as many datasets as possible.
2. In addition, a prognostic model for ESCC based on the robust DEGs was established.

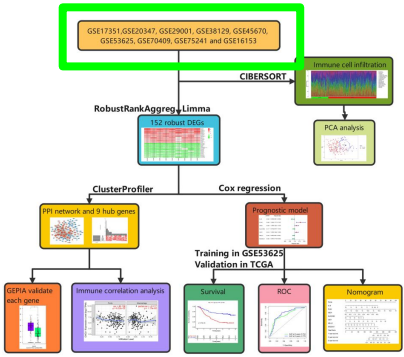
Aim of the study

1. This study aimed to identify robust DEGs and characterize the immune cell infiltration distribution in ESCC from as many datasets as possible.
2. In addition, a prognostic model for ESCC based on the robust DEGs was established.
3. Enrichment analysis and immune infiltration analysis of robust DEGs would improve the understanding of the molecular mechanisms of tumorigenesis and facilitate the development of new therapeutic strategies in ESCC.


GSE17351,GSE20347, GSE29001, GSE38129, GSE45670,
GSE53625, GSE70409, GSE75241 and GSE16153
CIBERSORT
Immune cell infiltration
RobustRankAggreg Limma
152 robust DEGs
PCA analysis
ClusterProfiler
Cox regression
PPI network and 9 hub genes
Prognostic model
Training in GSE53625
Validation in TCGA
GEPIA validate each gene
Immune correlation analysis
Survival
ROC
Nomogram


GSE17351,GSE20347, GSE29001, GSE38129, GSE45670,
GSE53625, GSE70409, GSE75241 and GSE16153
CIBERSORT
Immune cell infiltration
PCA analysis
RobustRankAggreg Limma
152 robust DEGs
ClusterProfiler
Cox regression
PPI network and 9 hub genes
Prognostic model
Training in GSE53625
Validation in TCGA
GEPIA validate each gene
Immune correlation analysis
Survival
ROC
Nomogram

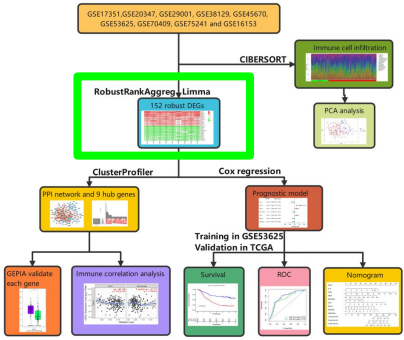
Methods



1. Microarray data collection (Gene Expression Omnibus)

Datasets	Year	Country	Tumor/Normal	Follow-up	Platform	Number of rows
GSE17351	2009	USA	5/5	No	GPL570	54,675
GSE20347	2010	USA	17/17	No	GPL571	22,277
GSE29001	2011	USA	21/24	No	GPL571	22,277
GSE38129	2012	USA	30/30	No	GPL571	22,277
GSE45670	2013	China	28/10	No	GPL570	54,675
GSE53625	2013	China	179/179	Yes	GPL18109	71,584
GSE70409	2013	China	17/17	No	GPL13287	29,187
GSE75241	2015	Brazil	15/15	No	GPL5175	316,919
GSE161533	2020	China	28/28	No	GPL570	54,675

Methods



1. Microarray data collection
2. Differential expression analysis in ESCC (limma + RRA)

   |log2(FC)| > 2
   adjusted p < 0.05

   RobustRankAggreg package

Robust Rank Aggregation

Input

Show Attributes

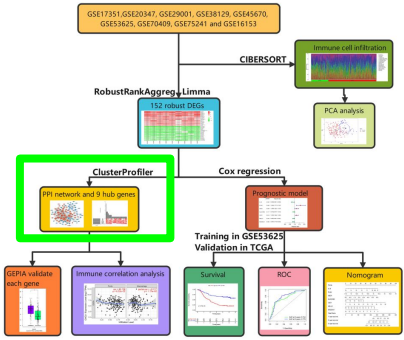
Name	Type	Value
list	list [3]	List of length 3
Set one	character [6]	'A' 'B' 'C' 'D' 'E' 'F'
Set two	character [6]	'A' 'C' 'E' 'B' 'D' 'G'
Set three	character [6]	'B' 'A' 'E' 'C' 'D' 'H'

Output

Filter

	Name	Score
A	A	0.0468750
B	B	0.3750000
C	C	0.3750000
D	D	0.7324219
E	E	0.7324219
F	F	1.0000000
G	G	1.0000000
H	H	1.0000000

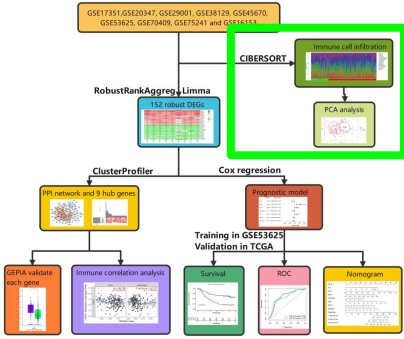
Methods



1. Microarray data collection
2. Differential expression analysis in ESCC
3. Functional and pathway enrichment analysis (GO + KEGG)

   CC + BP + MF
   clusterProfileR package

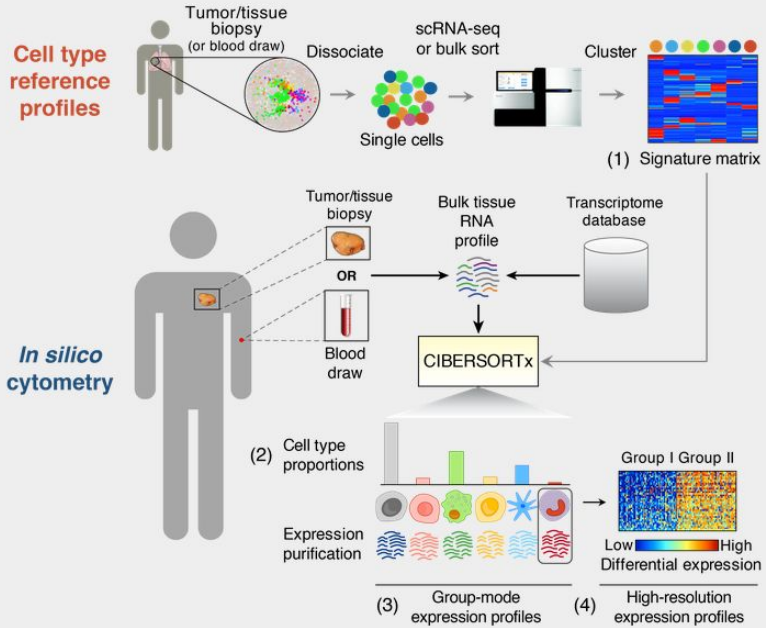
Methods



1. Microarray data collection
2. Differential expression analysis in ESCC
3. Functional and pathway enrichment analyses
4. Analysis of immune cell infiltration with the CIBERSORT algorithm

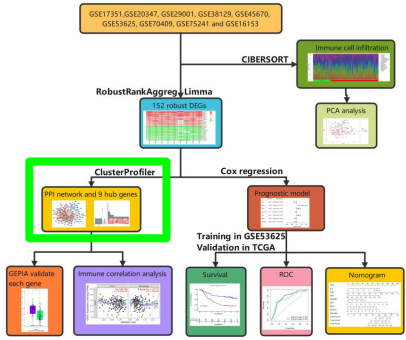
CIBERSORTX

CIBERSORTx is an analytical tool from the Alizadeh Lab and Newman Lab to impute gene expression profiles and provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data.



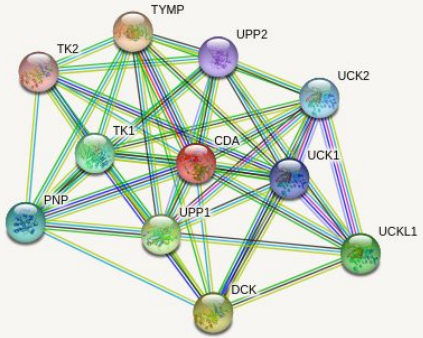
Run CIBERSORTx »

Methods



1. Microarray data collection
2. Differential expression analysis in ESCC
3. Functional and pathway enrichment analyses
4. Analysis of immune cell infiltration with the CIBERSORT algorithm
5. Identification of hub genes

   string-db.org
   Cytoscape + CytoHubba plugin



Viewers › Legend › Settings › Analysis › Exports ⌄ Clusters › More Less

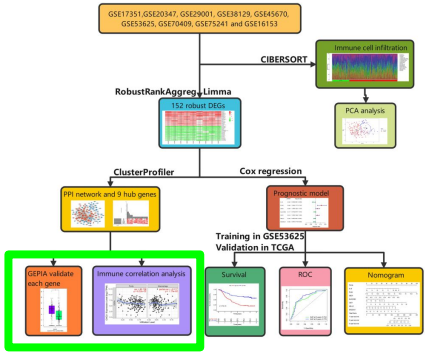
Export your current network: *Open or install latest Cytoscape to send network* Send network to Cytoscape

... as a bitmap image:	download	*file format is 'PNG': portable network graphic*
... as a high-resolution bitmap:	download	*same PNG format, but at higher resolution*
... as a vector graphic:	download	*SVG: scalable vector graphic - can be opened and edited in Illustrator, CorelDraw, Dia, etc*
... as short tabular text output:	download	*TSV: tab separated values - can be opened in Excel and Cytoscape (lists only one-way edges: A-B)*
... as tabular text output:	download	*TSV: tab separated values - can be opened in Excel (lists reciprocal edges: A-B,B-A)*
... as an XML summary:	download	*structured XML interaction data, according to the 'PSI-MI' data standard*
... protein node degrees:	download	*node degree of proteins in your network (given the current score cut-off)*
... network coordinates:	download	*a flat-file format describing the coordinates and colors of nodes in the network*
... protein sequences:	download	*MFA: multi-fasta format - containing the aminoacid sequences in the network*
... protein annotations:	download	*a tab-delimited file describing the names, domains and descriptions of proteins in your network*

Browse interactions in tabular form:

▲node1	node2	node1 accession	node2 accession	node1 annotation	node2 annotation	score
CDA	DCK	ENSP00000364212	ENSP00000286648	*Cytidine deaminase; This enzyme ...*	*Deoxycytidine kinase; Required for ...*	0.985
CDA	PNP	ENSP00000364212	ENSP00000354532	*Cytidine deaminase; This enzyme ...*	*Purine nucleoside phosphorylase; ...*	0.962
CDA	TK1	ENSP00000364212	ENSP00000301634	*Cytidine deaminase; This enzyme ...*	*Thymidine kinase, cytosolic; Thymi...*	0.963
CDA	TK2	ENSP00000364212	ENSP00000299697	*Cytidine deaminase; This enzyme ...*	*Thymidine kinase 2, mitochondrial;...*	0.942
CDA	TYMP	ENSP00000364212	ENSP00000379038	*Cytidine deaminase; This enzyme ...*	*Thymidine phosphorylase; May ha...*	0.996
CDA	UCK1	ENSP00000364212	ENSP00000361285	*Cytidine deaminase; This enzyme ...*	*Uridine-cytidine kinase 1; Phospho...*	0.949
CDA	UCK2	ENSP00000364212	ENSP00000356853	*Cytidine deaminase; This enzyme ...*	*Uridine-cytidine kinase 2; Phospho...*	0.952
CDA	UCKL1	ENSP00000364212	ENSP00000346155	*Cytidine deaminase; This enzyme ...*	*Uridine-cytidine kinase-like 1; May ...*	0.970
CDA	UPP1	ENSP00000364212	ENSP00000330032	*Cytidine deaminase; This enzyme ...*	*Uridine phosphorylase 1; Catalyzes...*	0.973
CDA	UPP2	ENSP00000364212	ENSP00000474090	*Cytidine deaminase; This enzyme ...*	*Uridine phosphorylase 2; Catalyzes...*	0.949

Methods



1. Microarray data collection
2. Differential expression analysis in ESCC
3. Functional and pathway enrichment analyses
4. Analysis of immune cell infiltration with the CIBERSORT algorithm
5. Identification of hub genes
6. Analysis of hub genes

Gene Expression Profiling Interactive Analysis (GEPIA)
GraphPad Prism 8.0
Tumor Immune Estimation Resource (TIMER)

GEPIA

Gene Expression Profiling Interactive Analysis

Single Gene Analysis | Cancer Type Analysis | Multiple Gene Analysis

Enter gene name:

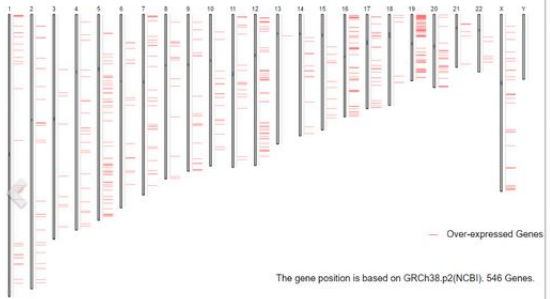
The indicators in search box are "symbol" or "alias (newest symbol)".

e.g. ERBB2/ENSG00000141736/2064

GoPIA!

Profile | Boxplots | Stage Plots | Survival Analysis | Similar

Chromosome distribution



ERBB2 overexpression plot

Differential genes by cancer type

Gene Symbol	Gene ID	Median (Tumor)	Median (Normal)	Log2(Fold Change)	adjp
COL10A1	ENSG00000123500.9	67.461	0.120	5.934	6.52e-208
TFF1	ENSG00000160182.2	133.796	1.270	5.892	2.02e-24
RP11-40C6.2	ENSG00000219928.2	278.049	4.590	5.642	1.98e-85
MMP11	ENSG00000099953.9	167.729	2.560	5.567	9.50e-231
COL11A1	ENSG00000060718.18	37.131	0.500	4.668	1.96e-137
AP000349.2	ENSG00000280178.1	26.150	0.150	4.561	3.61e-121
UBE2C	ENSG00000175063.16	54.579	1.410	4.527	1.12e-227
RPS-940J5.9	ENSG00000269968.1	17.259	0.000	4.191	1.29e-25
AGR3	ENSG00000173467.8	232.904	11.820	4.189	2.97e-24
S100P	ENSG00000163993.6	48.839	1.880	4.113	1.81e-72

BRCA significantly different genes

TIMER: Tumor IMmune Estimation Resource

A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells

The lastest version: TIMER2.0

Home | Gene | Survival | Mutation | SCNA | Diff Exp | Correlation | Estimation

Cancer Type:

ACC (Adrenocortical Carcinoma, 79)

Gene Symbols: (Y-axis)

Example: PDCD1 CTLA4

Gene Symbols: (X-axis)

Example: GZMA PRF1

Correlation Adjusted by:

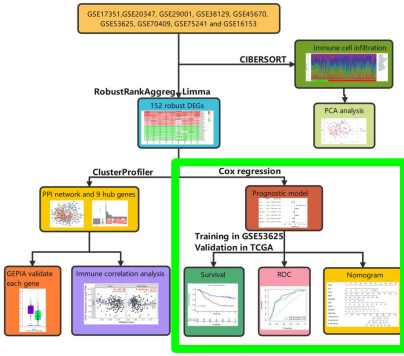
None

Submit

Instruction

Correlation module draws the expression scatterplots between a pair of user-defined genes in a given cancer type, together with the Spearman's rho value and estimated statistical significance. Options for partial correlation conditioned on tumor purity or age are also provided.

Methods



1. Microarray data collection
2. Differential expression analysis in ESCC
3. Functional and pathway enrichment analyses
4. Analysis of immune cell infiltration with the CIBERSORT algorithm
5. Identification of hub genes
6. Analysis of hub genes
7. Construction and validation of the prognostic model.
8. Independence analysis of the prognostic model and construction of the nomogram.

survival package
SurvivalROC package
rms package

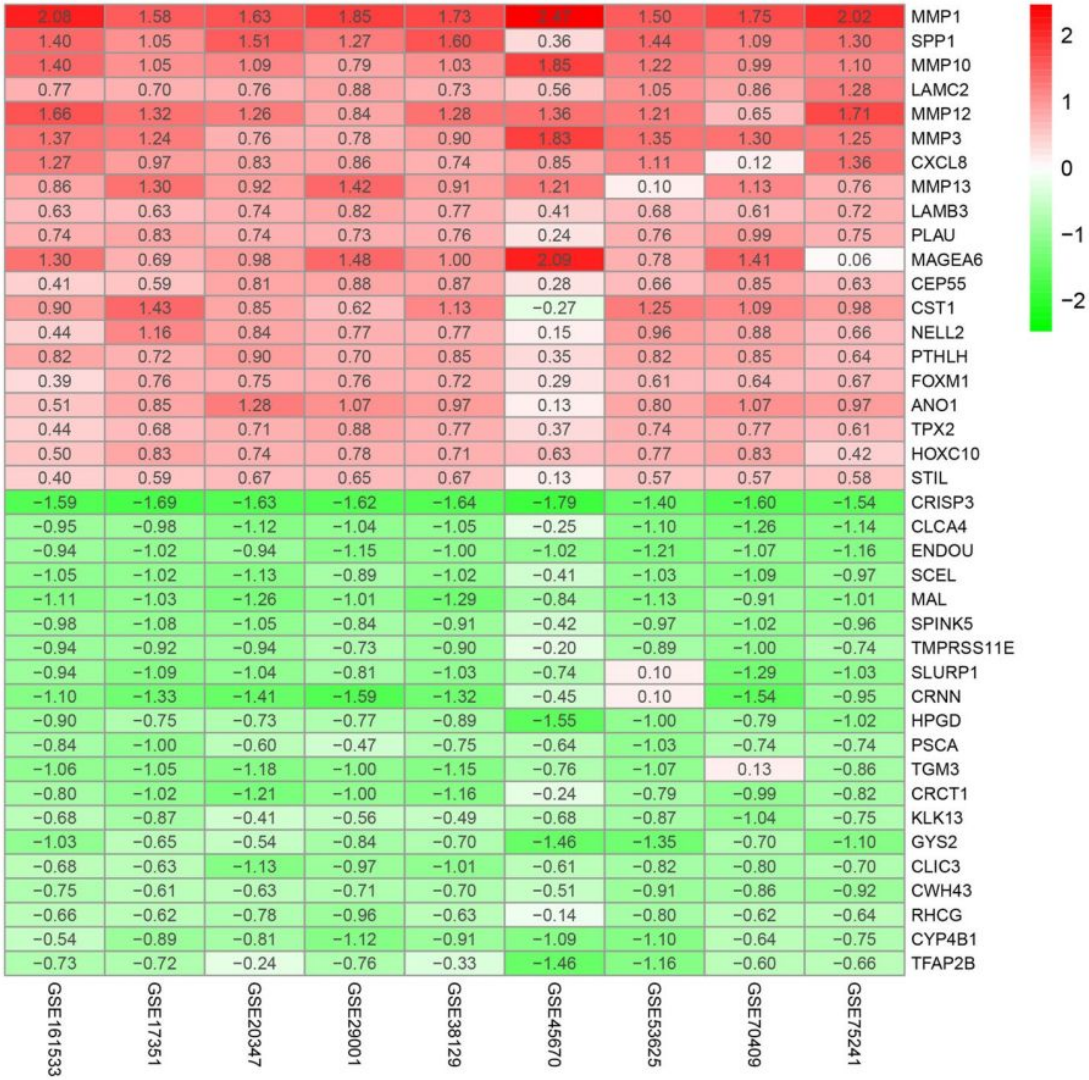
Results

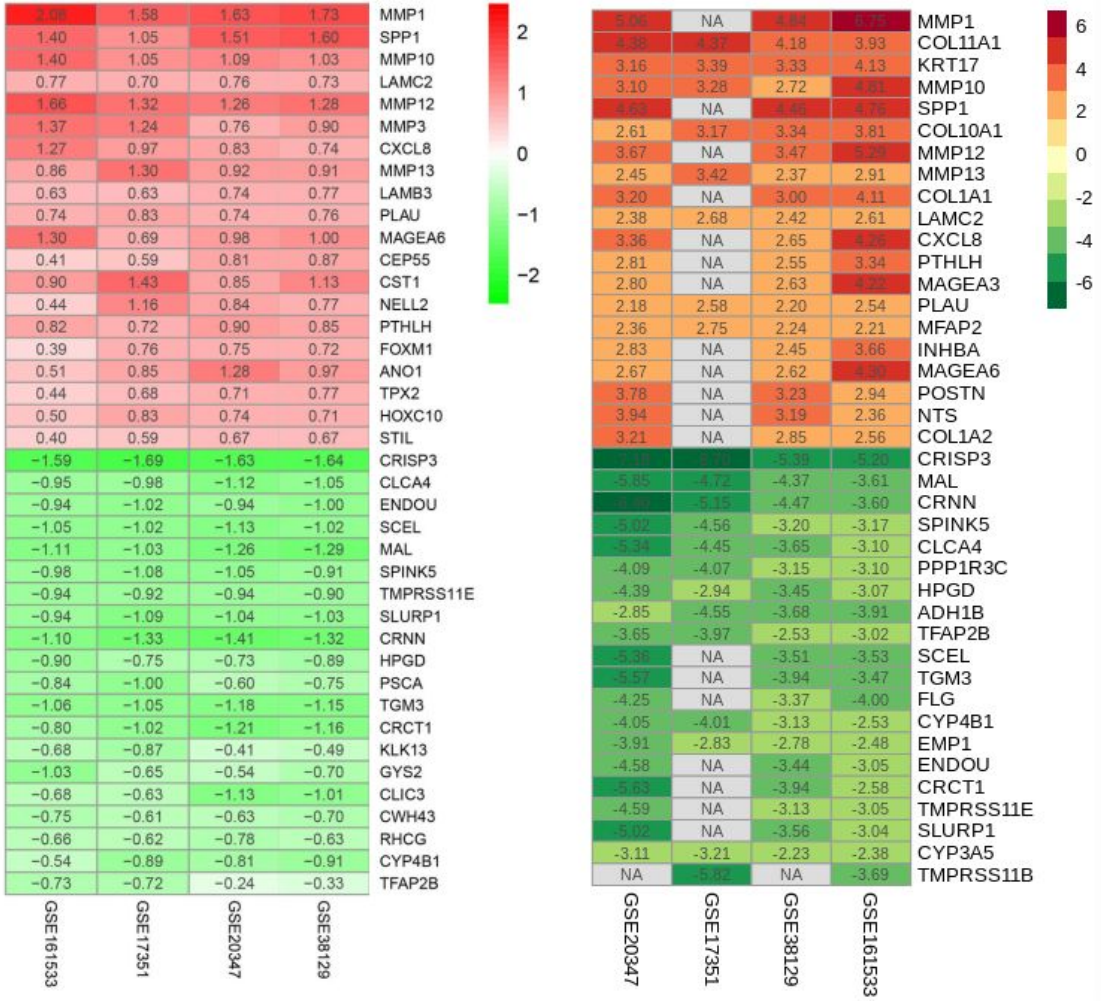
Identification of DEGs and robust DEGs

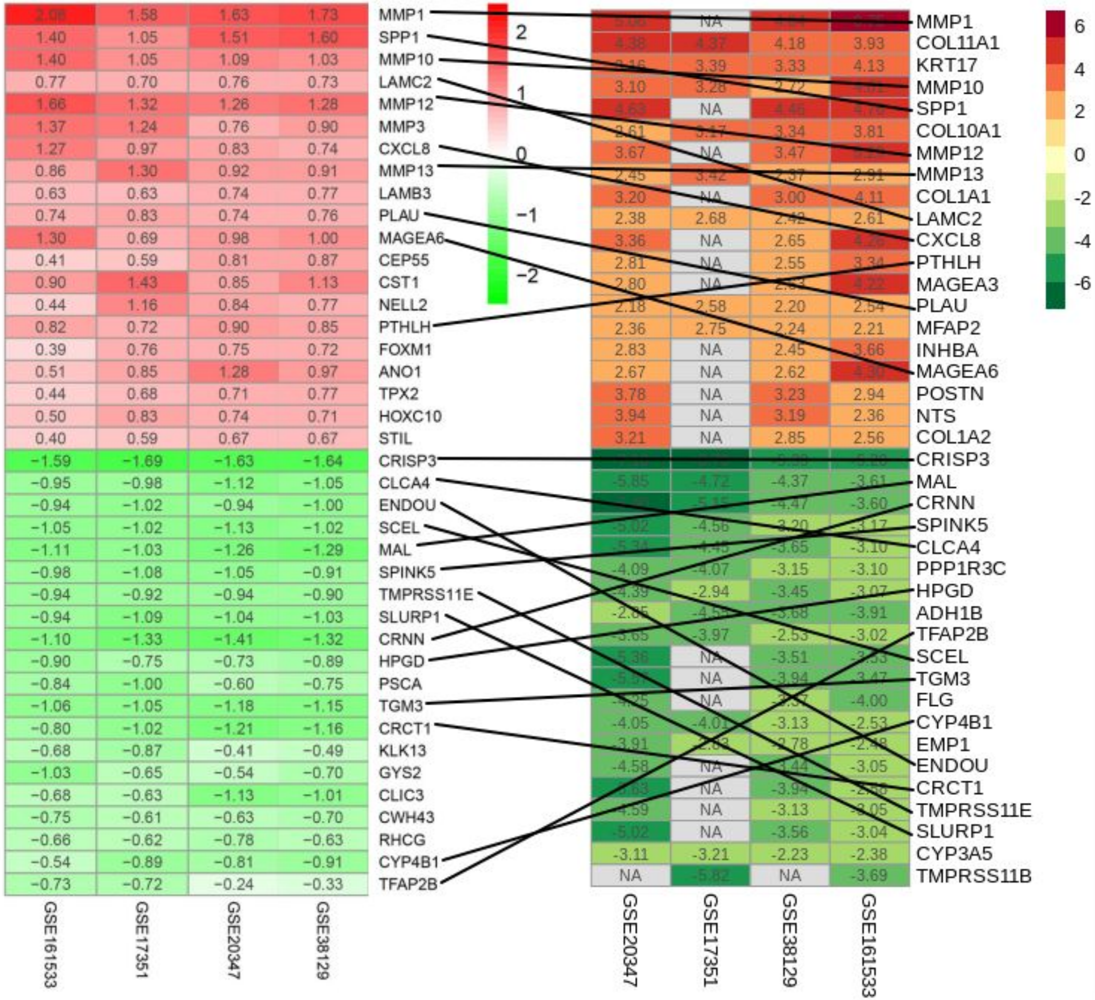
152 robust DEGs were identified; 54 up- and 98 downregulated (FDR < 0.05)

DEGs ranked according to their log2FC values.

Datasets	# of upregulated	# of downregulated
GSE17351	110	116
GSE20347	56	163
GSE29001	168	221
GSE38129	38	70
GSE45670	249	443
GSE53625	204	482
GSE70409	115	272
GSE75241	124	99
GSE161533	57	90


2.08 1.58 1.63 1.85 1.73 2.47 1.50 1.75 2.02 MMP1
1.40 1.05 1.51 1.27 1.60 0.36 1.44 1.09 1.30 SPP1
1.40 1.05 1.09 0.79 1.03 1.85 1.22 0.99 1.10 MMP10
0.77 0.70 0.76 0.88 0.73 0.56 1.05 0.86 1.28 LAMC2
1.66 1.32 1.26 0.84 1.28 1.36 1.21 0.65 1.71 MMP12
1.37 1.24 0.76 0.78 0.90 1.83 1.35 1.30 1.25 MMP3
1.27 0.97 0.83 0.86 0.74 0.85 1.11 0.12 1.36 CXCL8
0.86 1.30 0.92 1.42 0.91 1.21 0.10 1.13 0.76 MMP13
0.63 0.63 0.74 0.82 0.77 0.41 0.68 0.61 0.72 LAMB3
0.74 0.83 0.74 0.73 0.76 0.24 0.76 0.99 0.75 PLAU
1.30 0.69 0.98 1.48 1.00 2.09 0.78 1.41 0.06 MAGEA6
0.41 0.59 0.81 0.88 0.87 0.28 0.66 0.85 0.63 CEP55
0.90 1.43 0.85 0.62 1.13 −0.27 1.25 1.09 0.98 CST1
0.44 1.16 0.84 0.77 0.77 0.15 0.96 0.88 0.66 NELL2
0.82 0.72 0.90 0.70 0.85 0.35 0.82 0.85 0.64 PTHLH
0.39 0.76 0.75 0.76 0.72 0.29 0.61 0.64 0.67 FOXM1
0.51 0.85 1.28 1.07 0.97 0.13 0.80 1.07 0.97 ANO1
0.44 0.68 0.71 0.88 0.77 0.37 0.74 0.77 0.61 TPX2
0.50 0.83 0.74 0.78 0.71 0.63 0.77 0.83 0.42 HOXC10
0.40 0.59 0.67 0.65 0.67 0.13 0.57 0.57 0.58 STIL
−1.59 −1.69 −1.63 −1.62 −1.64 −1.79 −1.40 −1.60 −1.54 CRISP3
−0.95 −0.98 −1.12 −1.04 −1.05 −0.25 −1.10 −1.26 −1.14 CLCA4
−0.94 −1.02 −0.94 −1.15 −1.00 −1.02 −1.21 −1.07 −1.16 ENDOU
−1.05 −1.02 −1.13 −0.89 −1.02 −0.41 −1.03 −1.09 −0.97 SCEL
−1.11 −1.03 −1.26 −1.01 −1.29 −0.84 −1.13 −0.91 −1.01 MAL
−0.98 −1.08 −1.05 −0.84 −0.91 −0.42 −0.97 −1.02 −0.96 SPINK5
−0.94 −0.92 −0.94 −0.73 −0.90 −0.20 −0.89 −1.00 −0.74 TMPRSS11E
−0.94 −1.09 −1.04 −0.81 −1.03 −0.74 0.10 −1.29 −1.03 SLURP1
−1.10 −1.33 −1.41 −1.59 −1.32 −0.45 0.10 −1.54 −0.95 CRNN
−0.90 −0.75 −0.73 −0.77 −0.89 −1.55 −1.00 −0.79 −1.02 HPGD
−0.84 −1.00 −0.60 −0.47 −0.75 −0.64 −1.03 −0.74 −0.74 PSCA
−1.06 −1.05 −1.18 −1.00 −1.15 −0.76 −1.07 0.13 −0.86 TGM3
−0.80 −1.02 −1.21 −1.00 −1.16 −0.24 −0.79 −0.99 −0.82 CRCT1
−0.68 −0.87 −0.41 −0.56 −0.49 −0.68 −0.87 −1.04 −0.75 KLK13
−1.03 −0.65 −0.54 −0.84 −0.70 −1.46 −1.35 −0.70 −1.10 GYS2
−0.68 −0.63 −1.13 −0.97 −1.01 −0.61 −0.82 −0.80 −0.70 CLIC3
−0.75 −0.61 −0.63 −0.71 −0.70 −0.51 −0.91 −0.86 −0.92 CWH43
−0.66 −0.62 −0.78 −0.96 −0.63 −0.14 −0.80 −0.62 −0.64 RHCG
−0.54 −0.89 −0.81 −1.12 −0.91 −1.09 −1.10 −0.64 −0.75 CYP4B1
−0.73 −0.72 −0.24 −0.76 −0.33 −1.46 −1.16 −0.60 −0.66 TFAP2B
GSE161533
GSE17351
GSE20347
GSE29001
GSE38129
GSE45670
GSE53625
GSE70409
GSE75241
2
1
0
−1
−2


2.08 1.58 1.63 1.73 MMP1
1.40 1.05 1.51 1.60 SPP1
1.40 1.05 1.09 1.03 MMP10
0.77 0.70 0.76 0.73 LAMC2
1.66 1.32 1.26 1.28 MMP12
1.37 1.24 0.76 0.90 MMP3
1.27 0.97 0.83 0.74 CXCL8
0.86 1.30 0.92 0.91 MMP13
0.63 0.63 0.74 0.77 LAMB3
0.74 0.83 0.74 0.76 PLAU
1.30 0.69 0.98 1.00 MAGEA6
0.41 0.59 0.81 0.87 CEP55
0.90 1.43 0.85 1.13 CST1
0.44 1.16 0.84 0.77 NELL2
0.82 0.72 0.90 0.85 PTHLH
0.39 0.76 0.75 0.72 FOXM1
0.51 0.85 1.28 0.97 ANO1
0.44 0.68 0.71 0.77 TPX2
0.50 0.83 0.74 0.71 HOXC10
0.40 0.59 0.67 0.67 STIL
−1.59 −1.69 −1.63 −1.64 CRISP3
−0.95 −0.98 −1.12 −1.05 CLCA4
−0.94 −1.02 −0.94 −1.00 ENDOU
−1.05 −1.02 −1.13 −1.02 SCEL
−1.11 −1.03 −1.26 −1.29 MAL
−0.98 −1.08 −1.05 −0.91 SPINK5
−0.94 −0.92 −0.94 −0.90 TMPRSS11E
−0.94 −1.09 −1.04 −1.03 SLURP1
−1.10 −1.33 −1.41 −1.32 CRNN
−0.90 −0.75 −0.73 −0.89 HPGD
−0.84 −1.00 −0.60 −0.75 PSCA
−1.06 −1.05 −1.18 −1.15 TGM3
−0.80 −1.02 −1.21 −1.16 CRCT1
−0.68 −0.87 −0.41 −0.49 KLK13
−1.03 −0.65 −0.54 −0.70 GYS2
−0.68 −0.63 −1.13 −1.01 CLIC3
−0.75 −0.61 −0.63 −0.70 CWH43
−0.66 −0.62 −0.78 −0.63 RHCG
−0.54 −0.89 −0.81 −0.91 CYP4B1
−0.73 −0.72 −0.24 −0.33 TFAP2B
GSE161533
GSE17351
GSE20347
GSE38129
2
1
0
−1
−2
5.06 NA 4.84 6.75 MMP1
4.38 4.37 4.18 3.93 COL11A1
3.16 3.39 3.33 4.13 KRT17
3.10 3.28 2.72 4.81 MMP10
4.63 NA 4.45 4.75 SPP1
2.61 3.17 3.34 3.81 COL10A1
3.67 NA 3.47 5.29 MMP12
2.45 3.42 2.37 2.91 MMP13
3.20 NA 3.00 4.11 COL1A1
2.38 2.68 2.42 2.61 LAMC2
3.36 NA 2.65 4.26 CXCL8
2.81 NA 2.55 3.34 PTHLH
2.80 NA 2.63 4.22 MAGEA3
2.18 2.58 2.20 2.54 PLAU
2.36 2.75 2.24 2.21 MFAP2
2.83 NA 2.45 3.66 INHBA
2.67 NA 2.62 4.30 MAGEA6
3.78 NA 3.23 2.94 POSTN
3.94 NA 3.19 2.36 NTS
3.21 NA 2.85 2.56 COL1A2
-5.10 -5.63 -5.39 -5.20 CRISP3
-5.85 -4.72 -4.37 -3.61 MAL
-5.60 -5.15 -4.47 -3.60 CRNN
-5.02 -4.56 -3.20 -3.17 SPINK5
-5.34 -4.45 -3.65 -3.10 CLCA4
-4.09 -4.07 -3.15 -3.10 PPP1R3C
-4.39 -2.94 -3.45 -3.07 HPGD
-2.85 -4.55 -3.68 -3.91 ADH1B
-3.65 -3.97 -2.53 -3.02 TFAP2B
-5.36 NA -3.51 -3.53 SCEL
-5.57 NA -3.94 -3.47 TGM3
-4.25 NA -3.37 -4.00 FLG
-4.05 -4.01 -3.13 -2.53 CYP4B1
-3.91 -2.83 -2.78 -2.48 EMP1
-4.58 NA -3.44 -3.05 ENDOU
-5.63 NA -3.94 -2.58 CRCT1
-4.59 NA -3.13 -3.05 TMPRSS11E
-5.02 NA -3.56 -3.04 SLURP1
-3.11 -3.21 -2.23 -2.38 CYP3A5
NA -5.82 NA -3.69 TMPRSS11B
GSE20347
GSE17351
GSE38129
GSE161533
6
4
2
0
-2
-4
-6


MMP1
SPP1
MMP10
LAMC2
MMP12
MMP3
CXCL8
MMP13
LAMB3
PLAU
MAGEA6
CEP55
CST1
NELL2
PTHLH
FOXM1
ANO1
TPX2
HOXC10
STIL
CRISP3
CLCA4
ENDOU
SCEL
MAL
SPINK5
TMPRSS11E
SLURP1
CRNN
HPGD
PSCA
TGM3
CRCT1
KLK13
GYS2
CLIC3
CWH43
RHCG
CYP4B1
TFAP2B
GSE161533
GSE17351
GSE20347
GSE38129
2
1
0
−1
−2
MMP1
COL11A1
KRT17
MMP10
SPP1
COL10A1
MMP12
MMP13
COL1A1
LAMC2
CXCL8
PTHLH
MAGEA3
PLAU
MFAP2
INHBA
MAGEA6
POSTN
NTS
COL1A2
CRISP3
MAL
CRNN
SPINK5
CLCA4
PPP1R3C
HPGD
ADH1B
TFAP2B
SCEL
TGM3
FLG
CYP4B1
EMP1
ENDOU
CRCT1
TMPRSS11E
SLURP1
CYP3A5
TMPRSS11B
GSE20347
GSE17351
GSE38129
GSE161533
6
4
2
0
-2
-4
-6

GSEA

Many biological functions enriched with the DEGs were associated with the tumor microenvironment (TME) and growth of cancer cells

GSEA

Many biological functions enriched with the DEGs were associated with the tumor microenvironment (TME) and growth of cancer cells

BP: extracellular matrix organization, extracellular structure organization and leukocyte chemotaxis

GSEA

Many biological functions enriched with the DEGs were associated with the tumor microenvironment (TME) and growth of cancer cells

BP: extracellular matrix organization, extracellular structure organization and leukocyte chemotaxis

CC: collagen-containing extracellular matrix, apical part of cell and endoplasmic reticulum lumen

GSEA

Many biological functions enriched with the DEGs were associated with the tumor microenvironment (TME) and growth of cancer cells

BP: extracellular matrix organization, extracellular structure organization and leukocyte chemotaxis

CC: collagen-containing extracellular matrix, apical part of cell and endoplasmic reticulum lumen

MF: receptor ligand activity, signaling receptor activator activity, extracellular matrix structural, cytokine activity and CXCR chemokine receptor binding

GSEA

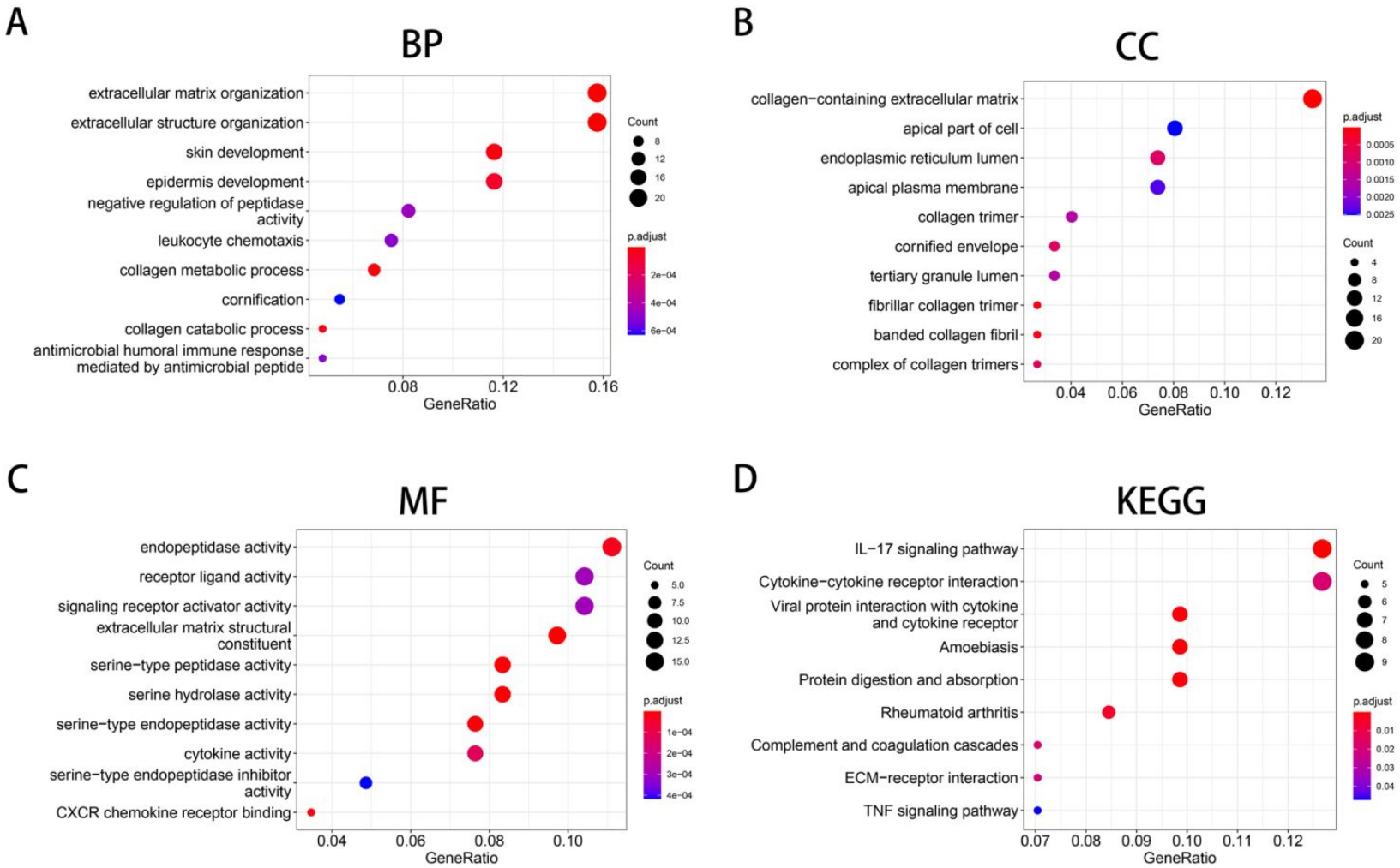
Many biological functions enriched with the DEGs were associated with the tumor microenvironment (TME) and growth of cancer cells

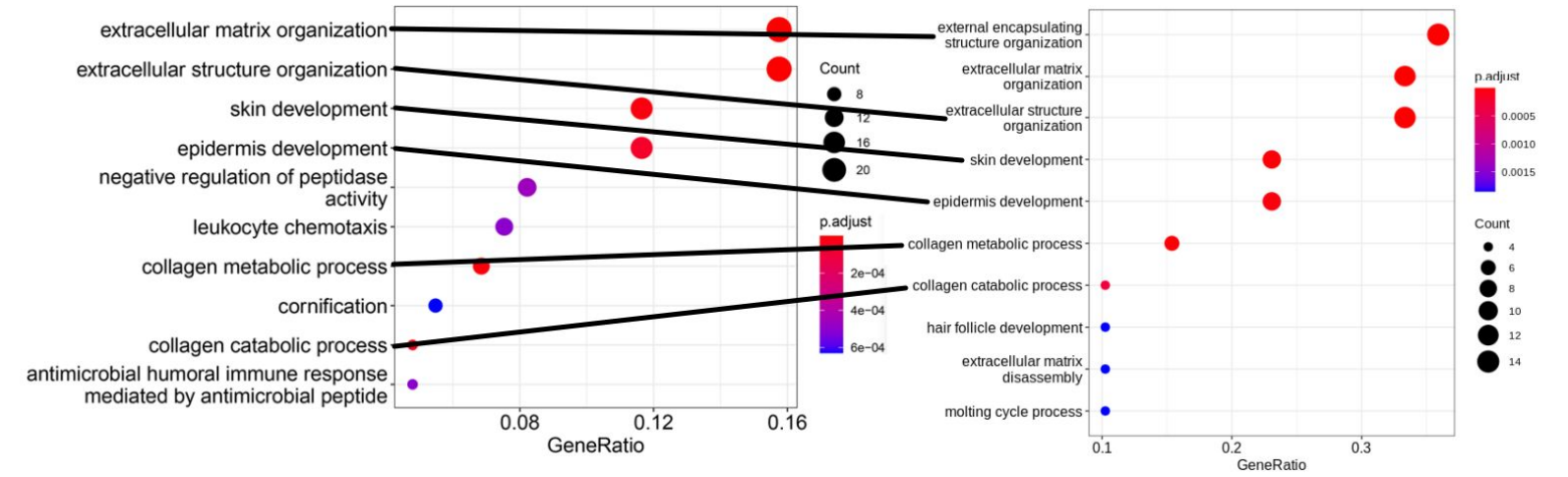
BP: extracellular matrix organization, extracellular structure organization and leukocyte chemotaxis

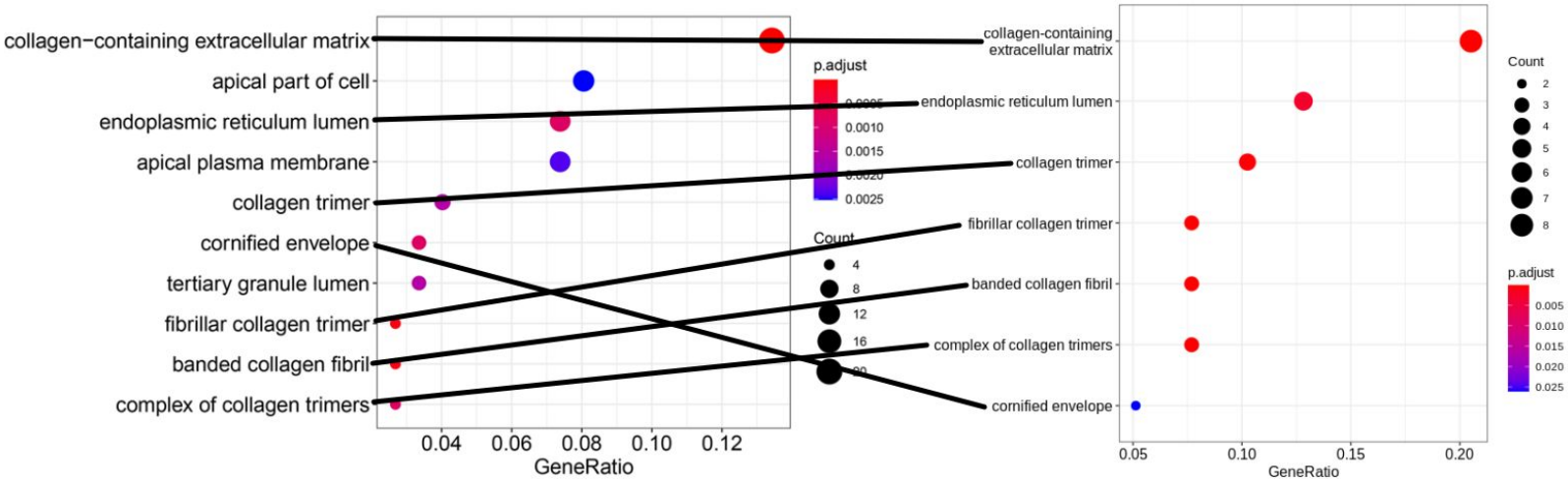
CC: collagen-containing extracellular matrix, apical part of cell and endoplasmic reticulum lumen

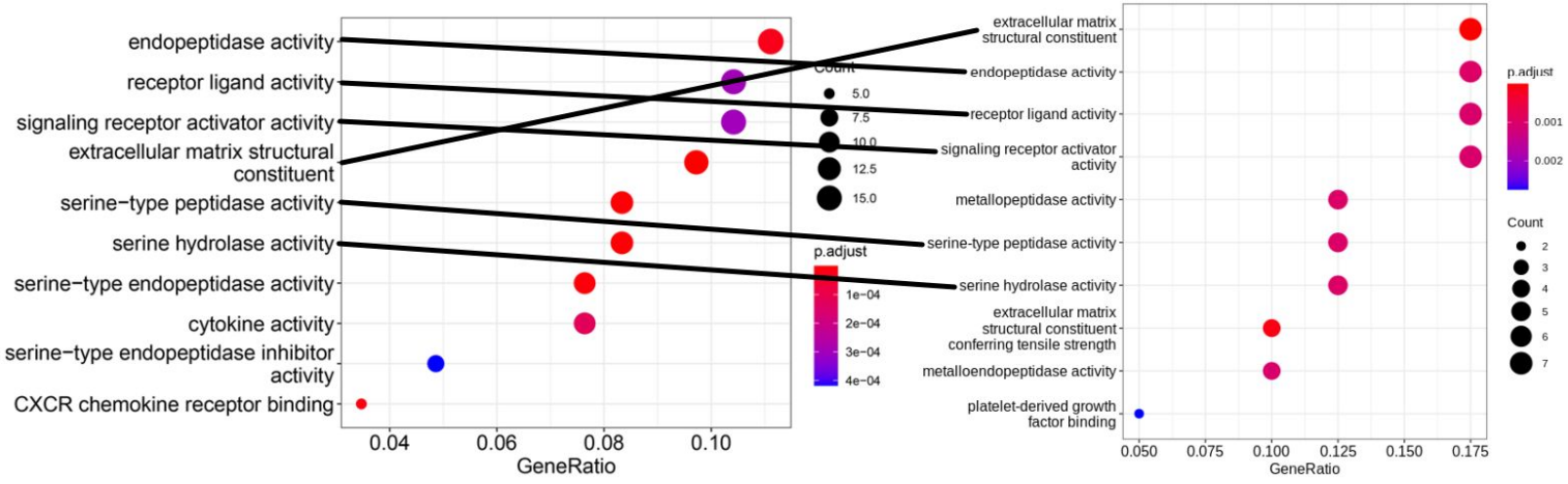
MF: receptor ligand activity, signaling receptor activator activity, extracellular matrix structural, cytokine activity and CXCR chemokine receptor binding

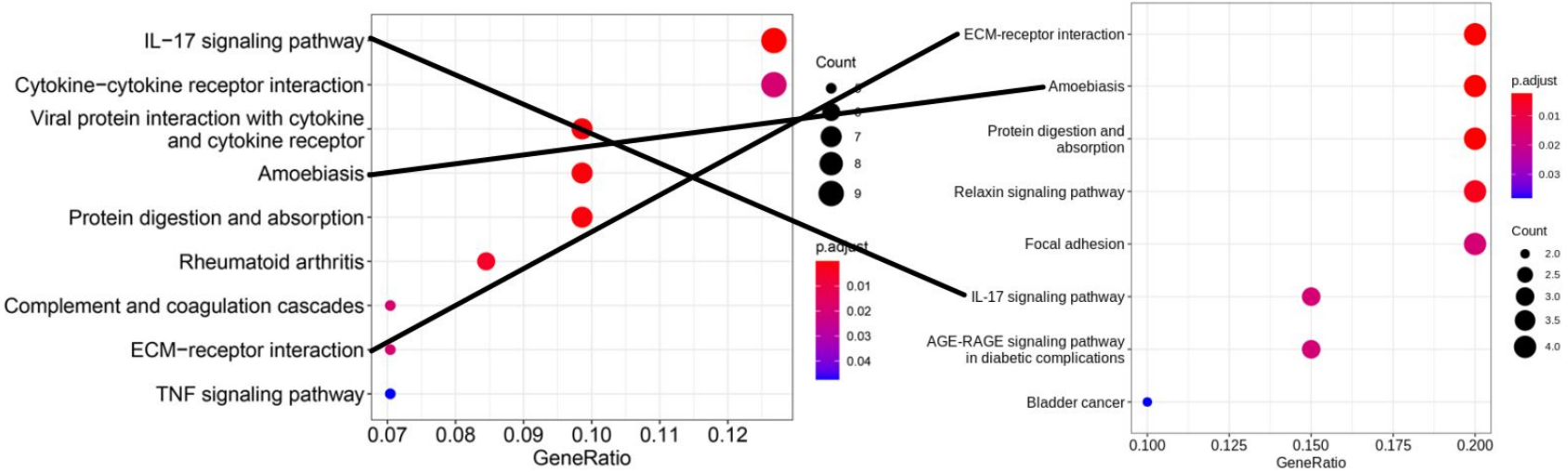
KEGG: IL-17 signaling pathway, cytokine-cytokine receptor interaction, ECM − receptor interaction and TNF signaling pathway


A
BP
extracellular matrix organization
extracellular structure organization
skin development
epidermis development
negative regulation of peptidase activity
leukocyte chemotaxis
collagen metabolic process
cornification
collagen catabolic process
antimicrobial humoral immune response mediated by antimicrobial peptide
0.08
0.12
0.16
GeneRatio
Count
8
12
16
20
p.adjust
2e-04
4e-04
6e-04
B
CC
collagen-containing extracellular matrix
apical part of cell
endoplasmic reticulum lumen
apical plasma membrane
collagen trimer
cornified envelope
tertiary granule lumen
fibrillar collagen trimer
banded collagen fibril
complex of collagen trimers
0.04
0.06
0.08
0.10
0.12
GeneRatio
p.adjust
0.0005
0.0010
0.0015
0.0020
0.0025
Count
4
8
12
16
20
C
MF
endopeptidase activity
receptor ligand activity
signaling receptor activator activity
extracellular matrix structural constituent
serine-type peptidase activity
serine hydrolase activity
serine-type endopeptidase activity
cytokine activity
serine-type endopeptidase inhibitor activity
CXCR chemokine receptor binding
0.04
0.06
0.08
0.10
GeneRatio
Count
5.0
7.5
10.0
12.5
15.0
p.adjust
1e-04
2e-04
3e-04
4e-04
D
KEGG
IL-17 signaling pathway
Cytokine-cytokine receptor interaction
Viral protein interaction with cytokine and cytokine receptor
Amoebiasis
Protein digestion and absorption
Rheumatoid arthritis
Complement and coagulation cascades
ECM-receptor interaction
TNF signaling pathway
0.07
0.08
0.09
0.10
0.11
0.12
GeneRatio
Count
5
6
7
8
9
p.adjust
0.01
0.02
0.03
0.04


extracellular matrix organization
extracellular structure organization
skin development
epidermis development
negative regulation of peptidase activity
leukocyte chemotaxis
collagen metabolic process
cornification
collagen catabolic process
antimicrobial humoral immune response mediated by antimicrobial peptide
0.08
0.12
0.16
GeneRatio
Count
8
12
16
20
p.adjust
2e-04
4e-04
6e-04
external encapsulating structure organization
extracellular matrix organization
extracellular structure organization
skin development
epidermis development
collagen metabolic process
collagen catabolic process
hair follicle development
extracellular matrix disassembly
molting cycle process
0.1
0.2
0.3
GeneRatio
p.adjust
0.0005
0.0010
0.0015
Count
4
6
8
10
12
14


collagen−containing extracellular matrix
apical part of cell
endoplasmic reticulum lumen
apical plasma membrane
collagen trimer
cornified envelope
tertiary granule lumen
fibrillar collagen trimer
banded collagen fibril
complex of collagen trimers
0.04
0.06
0.08
0.10
0.12
GeneRatio
p.adjust
0.0010
0.0015
0.0025
Count
4
8
12
16
collagen-containing extracellular matrix
endoplasmic reticulum lumen
collagen trimer
fibrillar collagen trimer
banded collagen fibril
complex of collagen trimers
cornified envelope
0.05
0.10
0.15
0.20
GeneRatio
Count
2
3
4
5
6
7
8
p.adjust
0.005
0.010
0.015
0.020
0.025


endopeptidase activity
receptor ligand activity
signaling receptor activator activity
extracellular matrix structural constituent
serine−type peptidase activity
serine hydrolase activity
serine−type endopeptidase activity
cytokine activity
serine−type endopeptidase inhibitor activity
CXCR chemokine receptor binding
0.04
0.06
0.08
0.10
GeneRatio
Count
5.0
7.5
10.0
12.5
15.0
p.adjust
1e−04
2e−04
3e−04
4e−04
extracellular matrix structural constituent
endopeptidase activity
receptor ligand activity
signaling receptor activator activity
metallopeptidase activity
serine-type peptidase activity
serine hydrolase activity
extracellular matrix structural constituent conferring tensile strength
metalloendopeptidase activity
platelet-derived growth factor binding
0.050
0.075
0.100
0.125
0.150
0.175
GeneRatio
p.adjust
0.001
0.002
Count
2
3
4
5
6
7


IL−17 signaling pathway
Cytokine−cytokine receptor interaction
Viral protein interaction with cytokine and cytokine receptor
Amoebiasis
Protein digestion and absorption
Rheumatoid arthritis
Complement and coagulation cascades
ECM−receptor interaction
TNF signaling pathway
0.07
0.08
0.09
0.10
0.11
0.12
GeneRatio
Count
7
8
9
p.adjust
0.01
0.02
0.03
0.04
ECM-receptor interaction
Amoebiasis
Protein digestion and absorption
Relaxin signaling pathway
Focal adhesion
IL-17 signaling pathway
AGE-RAGE signaling pathway in diabetic complications
Bladder cancer
0.100
0.125
0.150
0.175
0.200
GeneRatio
p.adjust
0.01
0.02
0.03
Count
2.0
2.5
3.0
3.5
4.0

Immune infiltration

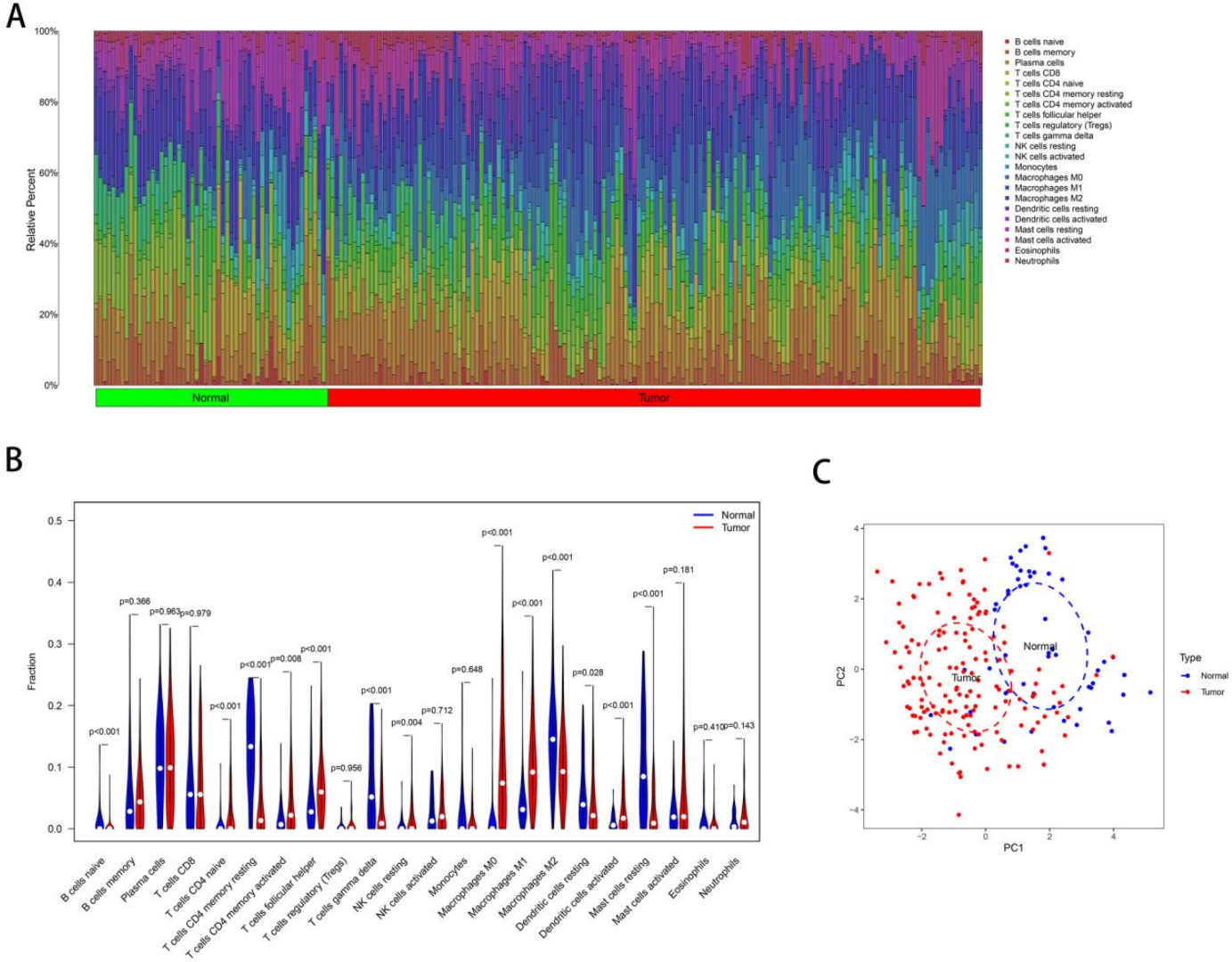
The CIBERSORT algorithm was used to analyse immune cell infiltration in all 665 samples from the 9 GEO normalized expression matrices

Seven types of immune cells were more abundant in ESCC tissues:

1. naïve $CD4^+$ T cells,
2. activated memory $CD4^+$ T cells
3. follicular helper T cells
4. resting natural killer (NK) cells
5. M0 macrophages
6. M1 macrophages
7. activated dendritic cells

Six types of immune cells were more abundant in normal tissues:

1. naïve B cells
2. resting memory $CD4^+$ T cells
3. gamma delta T cells
4. M2 macrophages
5. resting dendritic cells
6. resting mast cells


A
100%
80%
60%
40%
20%
0%
Relative Percent
Normal
Tumor
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
B
Fraction
0.5
0.4
0.3
0.2
0.1
0.0
Normal
Tumor
p<0.001
p=0.366
p=0.963
p=0.979
p<0.001
p<0.001
p=0.008
p<0.001
p=0.956
p<0.001
p=0.004
p=0.712
p=0.648
p<0.001
p<0.001
p<0.001
p=0.028
p<0.001
p<0.001
p=0.181
p=0.410
p=0.143
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
C
PC2
PC1
Normal
Tumor
Type
Normal
Tumor

Protein-protein interaction

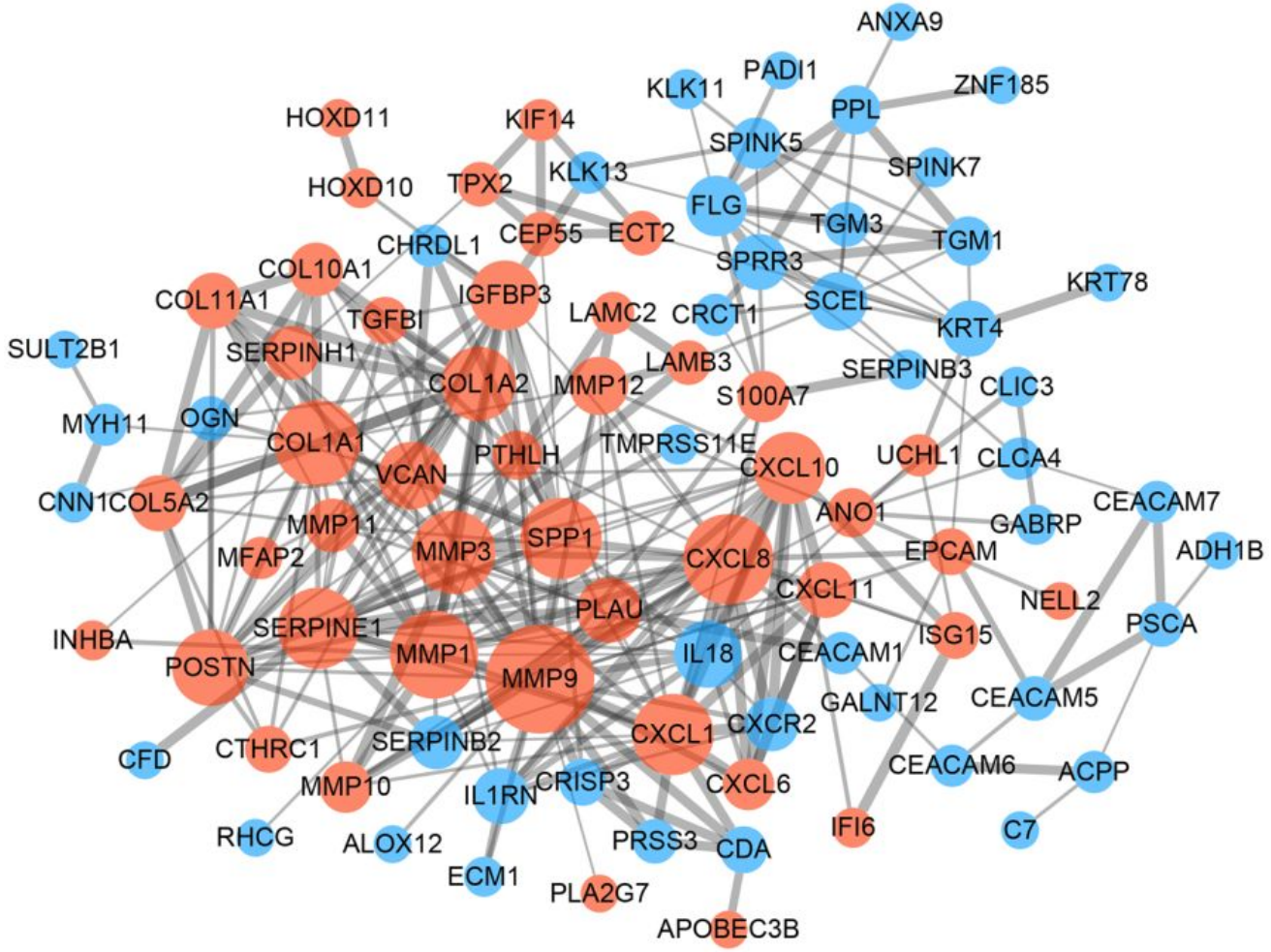
To further study the interaction of the 152 robust DEGs, we constructed a PPI network using the Search Tool for the Retrieval of Interacting Genes (STRING) database with a combined score > 0.4 as the cutoff criterion.

Protein-protein interaction

To further study the interaction of the 152 robust DEGs, we constructed a PPI network using the Search Tool for the Retrieval of Interacting Genes (STRING) database with a combined score > 0.4 as the cutoff criterion.

Nine identified hub genes:

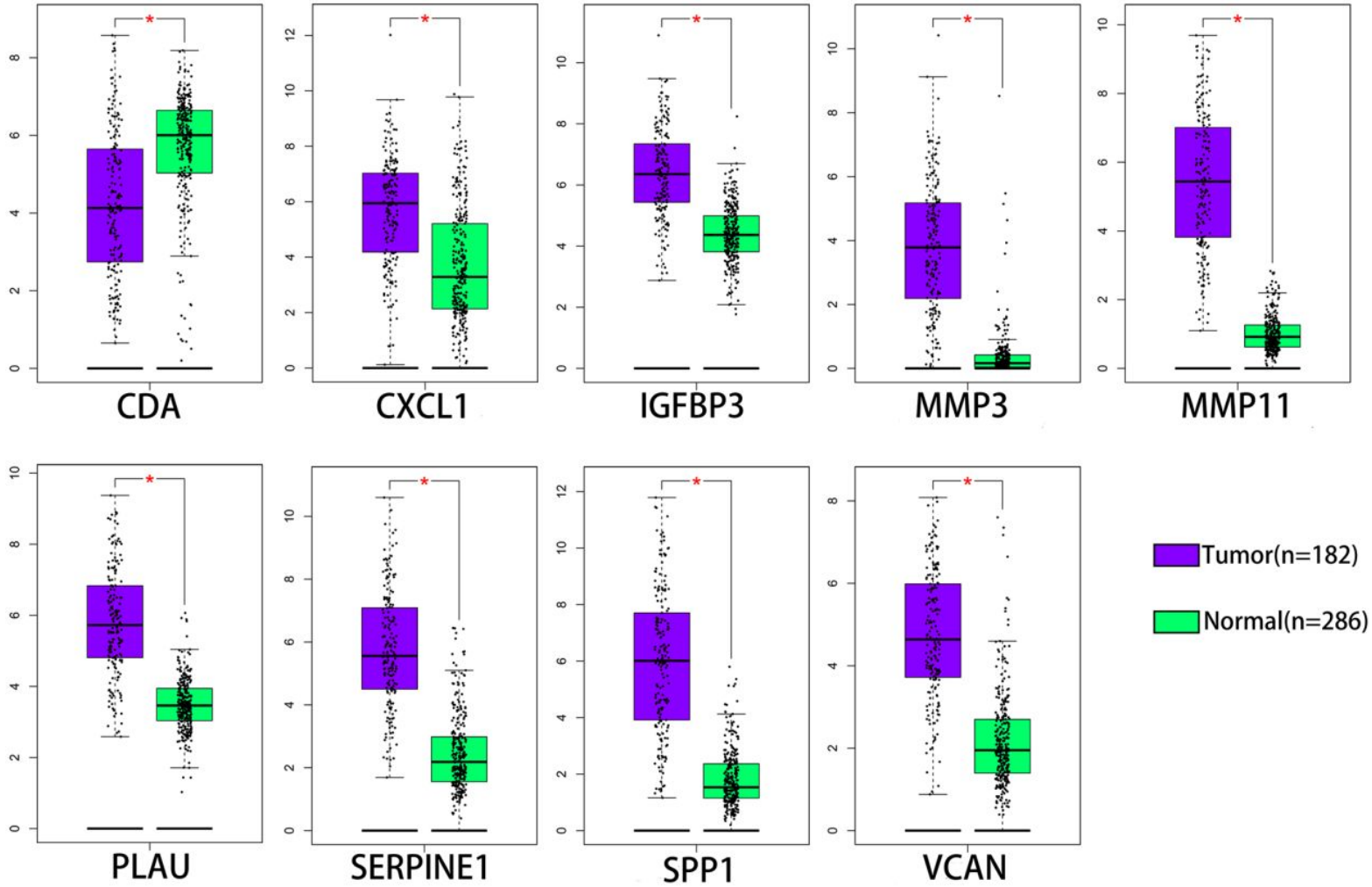
1. cytidine deaminase (CDA)
2. chemokine ligand 1 (CXCL1)
3. insulin-like growth factor binding protein 3 (IGFBP3)
4. matrix metallopeptidase 3 (MMP3)
5. matrix metallopeptidase 11 (MMP11)
6. plasminogen activator urokinase (PLAU, also named uPA)
7. serpin peptidase inhibitor member 1 (SERPINE1)
8. secreted phosphoprotein 1 (SPP1)
9. versican (VCAN)

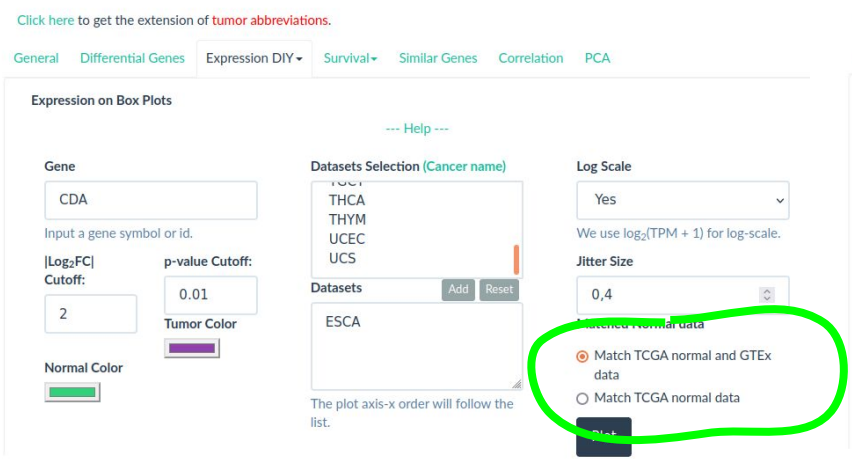

ANXA9
PADI1
KLK11
ZNF185
PPL
HOXD11
KIF14
SPINK5
SPINK7
HOXD10
TPX2
KLK13
FLG
TGM3
TGM1
CHRDL1
CEP55
ECT2
SPRR3
COL10A1
KRT78
COL11A1
IGFBP3
LAMC2
CRCT1
SCEL
KRT4
TGFBI
SULT2B1
SERPINH1
LAMB3
SERPINB3
COL1A2
MMP12
S100A7
CLIC3
MYH11
OGN
COL1A1
PTHLH
TMPRSS11E
UCHL1
CLCA4
VCAN
CXCL10
CNN1COL5A2
ANO1
CEACAM7
MMP11
GABRP
MFAP2
MMP3
SPP1
CXCL8
EPCAM
ADH1B
CXCL11
NELL2
PLAU
INHBA
SERPINE1
ISG15
PSCA
MMP1
IL18
CEACAM1
POSTN
MMP9
GALNT12
CEACAM5
CXCR2
CFD
CTHRC1
SERPINB2
CXCL1
CEACAM6
ACPP
MMP10
IL1RN
CRISP3
CXCL6
IFI6
C7
RHCG
ALOX12
PRSS3
CDA
ECM1
PLA2G7
APOBEC3B

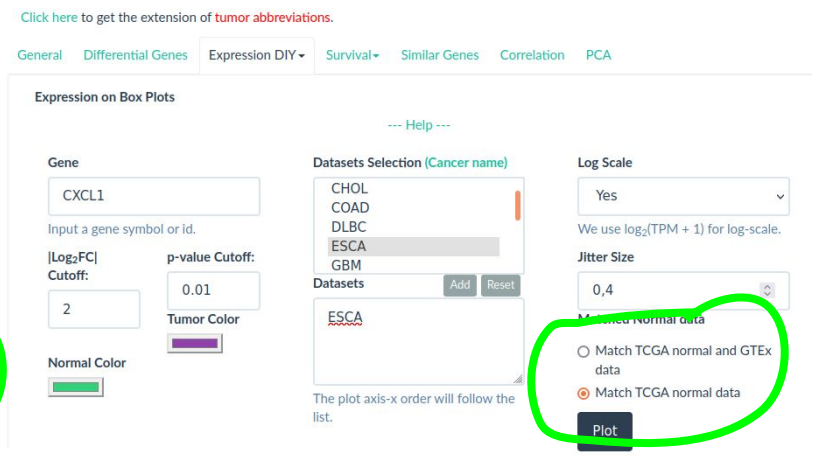
Validation of hub genes

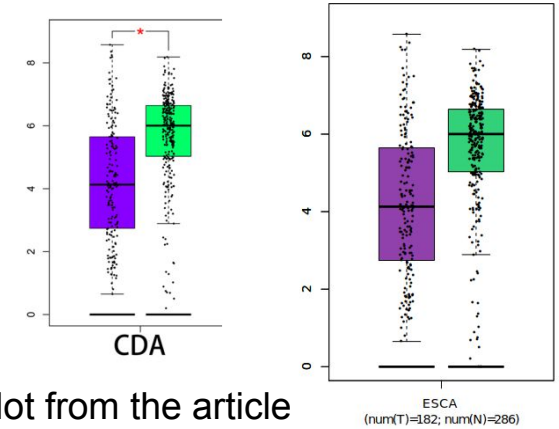
The mRNA expression of the 9 hub genes was validated using the Gene Expression Profiling Interactive Analysis (GEPIA) database.

Consistent with the results of the GEO analysis, the mRNA expression of CXCL1, IFGFBP3, MMP3, MMP11, PLAU, SERPINE1, SPP1 and VCAN was markedly upregulated but the mRNA expression of CDA was markedly downregulated in esophageal carcinoma tissues ($P < 0.01$)

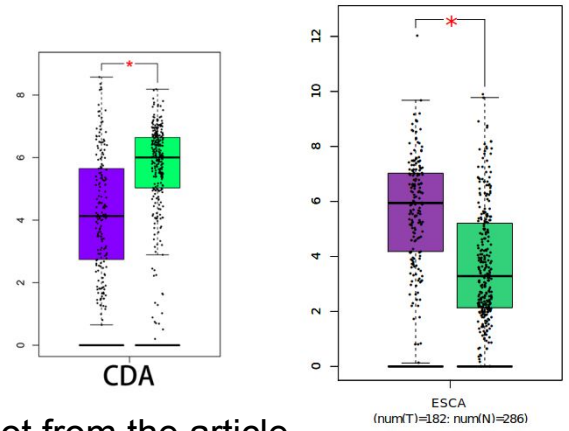

CDA
CXCL1
IGFBP3
MMP3
MMP11
PLAU
SERPINE1
SPP1
VCAN
Tumor(n=182)
Normal(n=286)



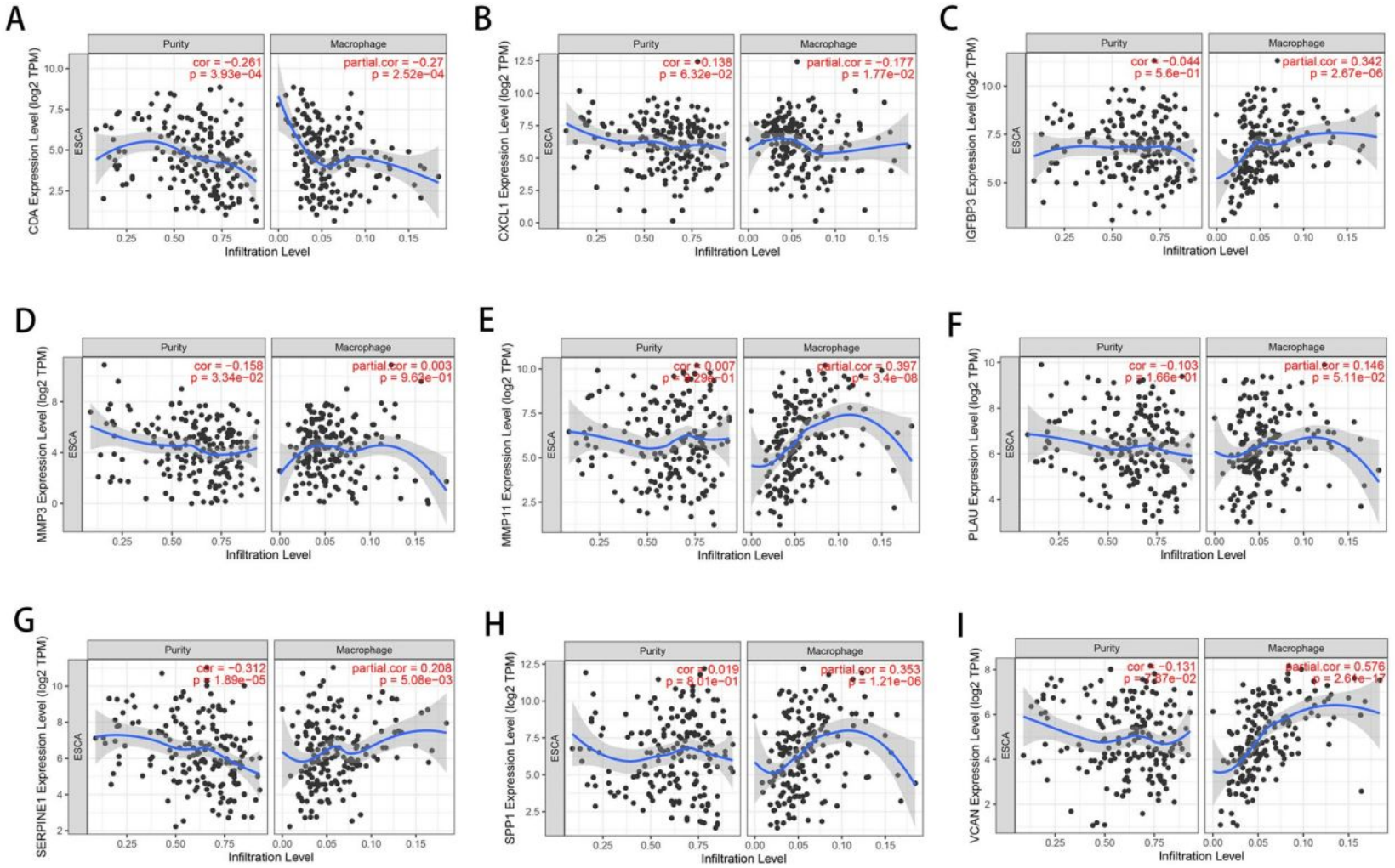


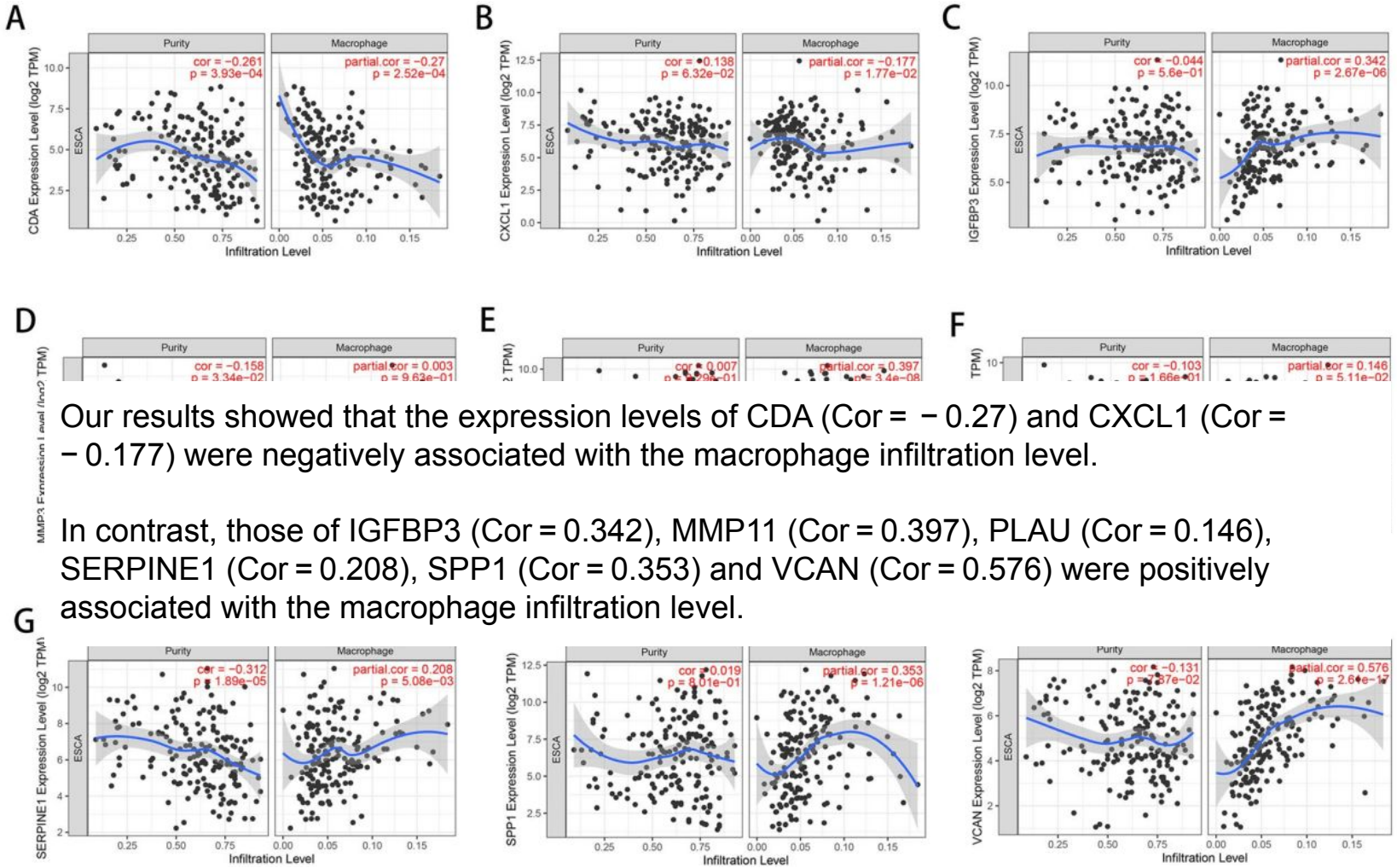


Plot from the article



Plot from the article


A
Purity
Macrophage
cor = −0.261
p = 3.93e−04
partial.cor = −0.27
p = 2.52e−04
CDA Expression Level (log2 TPM)
ESCA
Infiltration Level
B
Purity
Macrophage
cor = 0.138
p = 6.32e−02
partial.cor = −0.177
p = 1.77e−02
CXCL1 Expression Level (log2 TPM)
ESCA
Infiltration Level
C
Purity
Macrophage
cor = −0.044
p = 5.6e−01
partial.cor = 0.342
p = 2.67e−06
IGFBP3 Expression Level (log2 TPM)
ESCA
Infiltration Level
D
Purity
Macrophage
cor = −0.158
p = 3.34e−02
partial.cor = 0.003
p = 9.62e−01
MMP3 Expression Level (log2 TPM)
ESCA
Infiltration Level
E
Purity
Macrophage
cor = 0.007
p = 9.29e−01
partial.cor = 0.397
p = 3.4e−08
MMP11 Expression Level (log2 TPM)
ESCA
Infiltration Level
F
Purity
Macrophage
cor = −0.103
p = 1.66e−01
partial.cor = 0.146
p = 5.11e−02
PLAU Expression Level (log2 TPM)
ESCA
Infiltration Level
G
Purity
Macrophage
cor = −0.312
p = 1.89e−05
partial.cor = 0.208
p = 5.08e−03
SERPINE1 Expression Level (log2 TPM)
ESCA
Infiltration Level
H
Purity
Macrophage
cor = 0.019
p = 8.01e−01
partial.cor = 0.353
p = 1.21e−06
SPP1 Expression Level (log2 TPM)
ESCA
Infiltration Level
I
Purity
Macrophage
cor = −0.131
p = 7.37e−02
partial.cor = 0.576
p = 2.64e−17
VCAN Expression Level (log2 TPM)
ESCA
Infiltration Level



Our results showed that the expression levels of CDA (Cor = − 0.27) and CXCL1 (Cor = − 0.177) were negatively associated with the macrophage infiltration level.

In contrast, those of IGFBP3 (Cor = 0.342), MMP11 (Cor = 0.397), PLAU (Cor = 0.146), SERPINE1 (Cor = 0.208), SPP1 (Cor = 0.353) and VCAN (Cor = 0.576) were positively associated with the macrophage infiltration level.

Prognostic model

To investigate the prognostic significance of the 152 robust DEGs, 17 survival-related genes ($P < 0.05$) were identified by univariate Cox regression analysis in the GSE53625 datase.

After selecting the most suitable combination of candidate genes by multiple stepwise Cox regression, seven genes were used to construct a prognostic model:

1. interleukin 18 (IL18)
2. PLAU
3. anoctamin 1 (ANO1)
4. solute carrier organic anion transporter family member 1B3 (SLCO1B3)
5. cystatin SN (CST1)
6. neural EGFL like 2 (NELL2)
7. melanoma antigen family A11 (MAGEA11)

Prognostic model

To validate the risk model constructed with the 179 patients in GSE53625, we selected 185 patients in TCGA as the validation cohort. The patients in the two cohorts were divided into the low-risk and high-risk groups according to the median risk score. Kaplan–Meier survival analysis demonstrated that in both cohorts, the prognosis of the low-risk group was significantly better than that of the high-risk group.

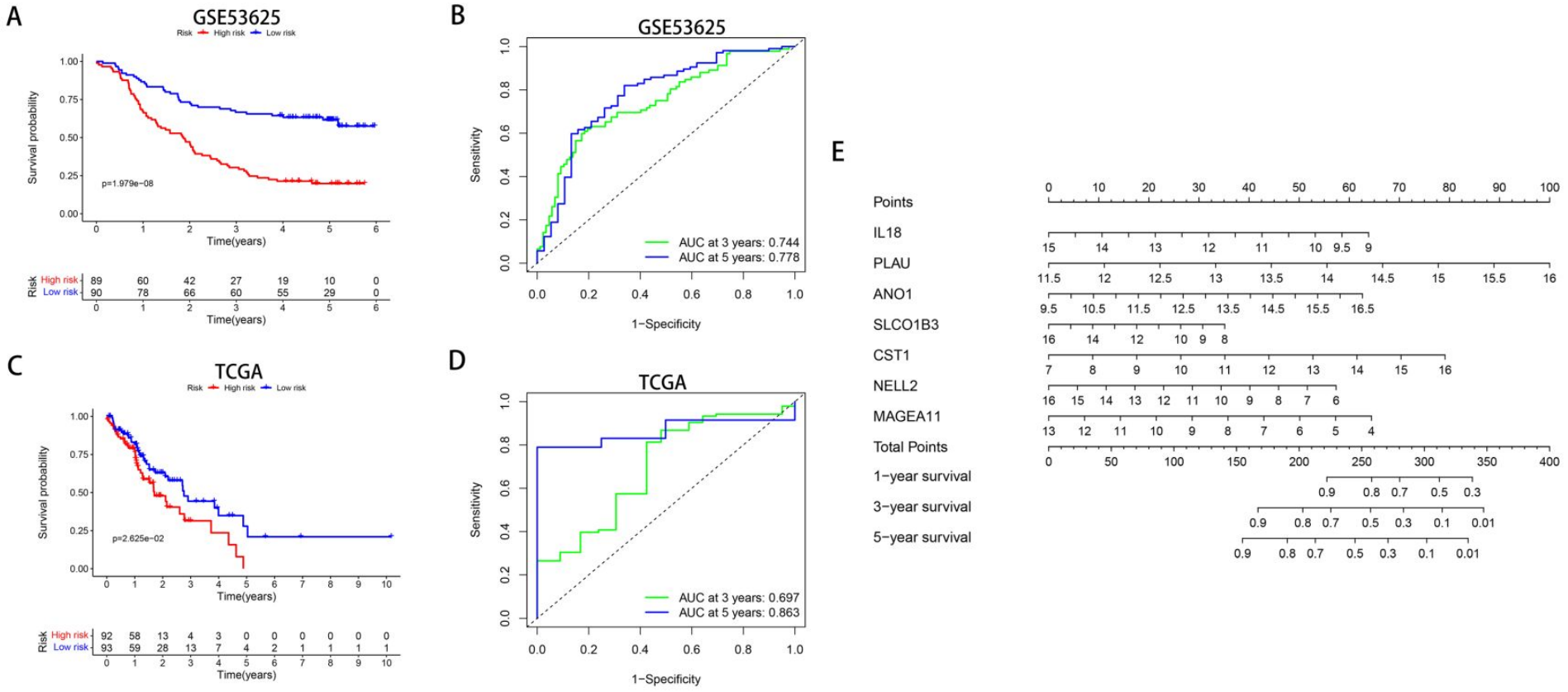

A
GSE53625
Risk High risk Low risk
Survival probability
p=1.979e−08
Time(years)
Risk High risk 89 60 42 27 19 10 0
Low risk 90 78 66 60 55 29 0
Time(years)
B
GSE53625
Sensitivity
1−Specificity
AUC at 3 years: 0.744
AUC at 5 years: 0.778
C
TCGA
Risk High risk Low risk
Survival probability
p=2.625e−02
Time(years)
Risk High risk 92 58 13 4 3 0 0 0 0 0 0
Low risk 93 59 28 13 7 4 2 1 1 1 1
Time(years)
D
TCGA
Sensitivity
1−Specificity
AUC at 3 years: 0.697
AUC at 5 years: 0.863
E
Points
IL18
PLAU
ANO1
SLCO1B3
CST1
NELL2
MAGEA11
Total Points
1−year survival
3−year survival
5−year survival

Figure S1

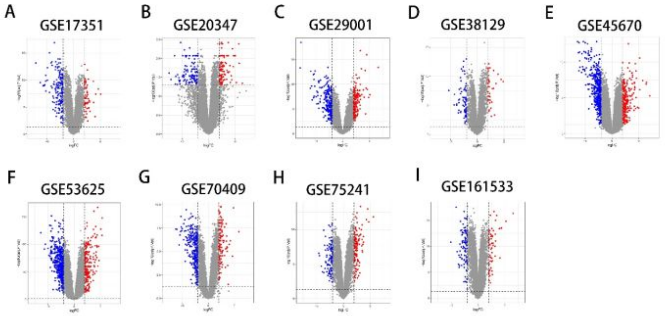
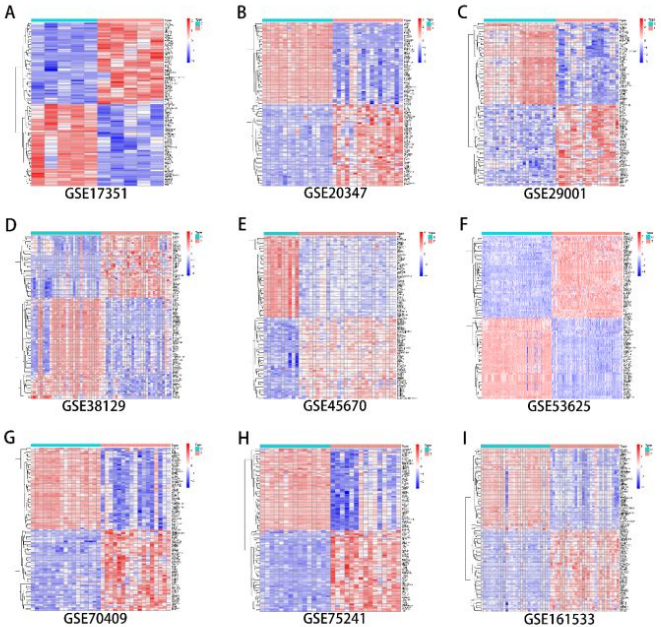

A
GSE17351
B
GSE20347
C
GSE29001
D
GSE38129
E
GSE45670
F
GSE53625
G
GSE70409
H
GSE75241
I
GSE161533

Figure S2


A
GSE17351
B
GSE20347
C
GSE29001
D
GSE38129
E
GSE45670
F
GSE53625
G
GSE70409
H
GSE75241
I
GSE161533

A