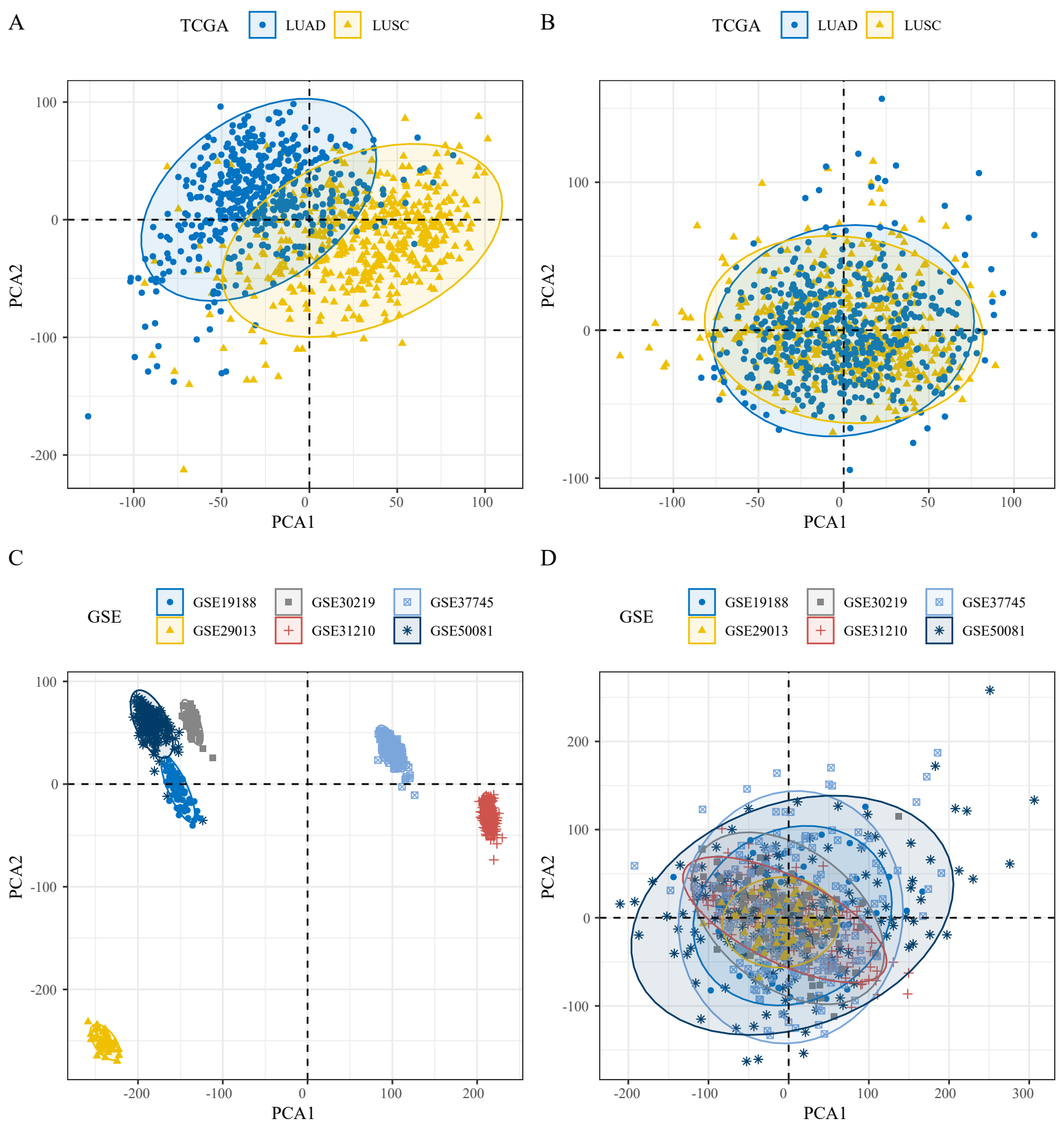
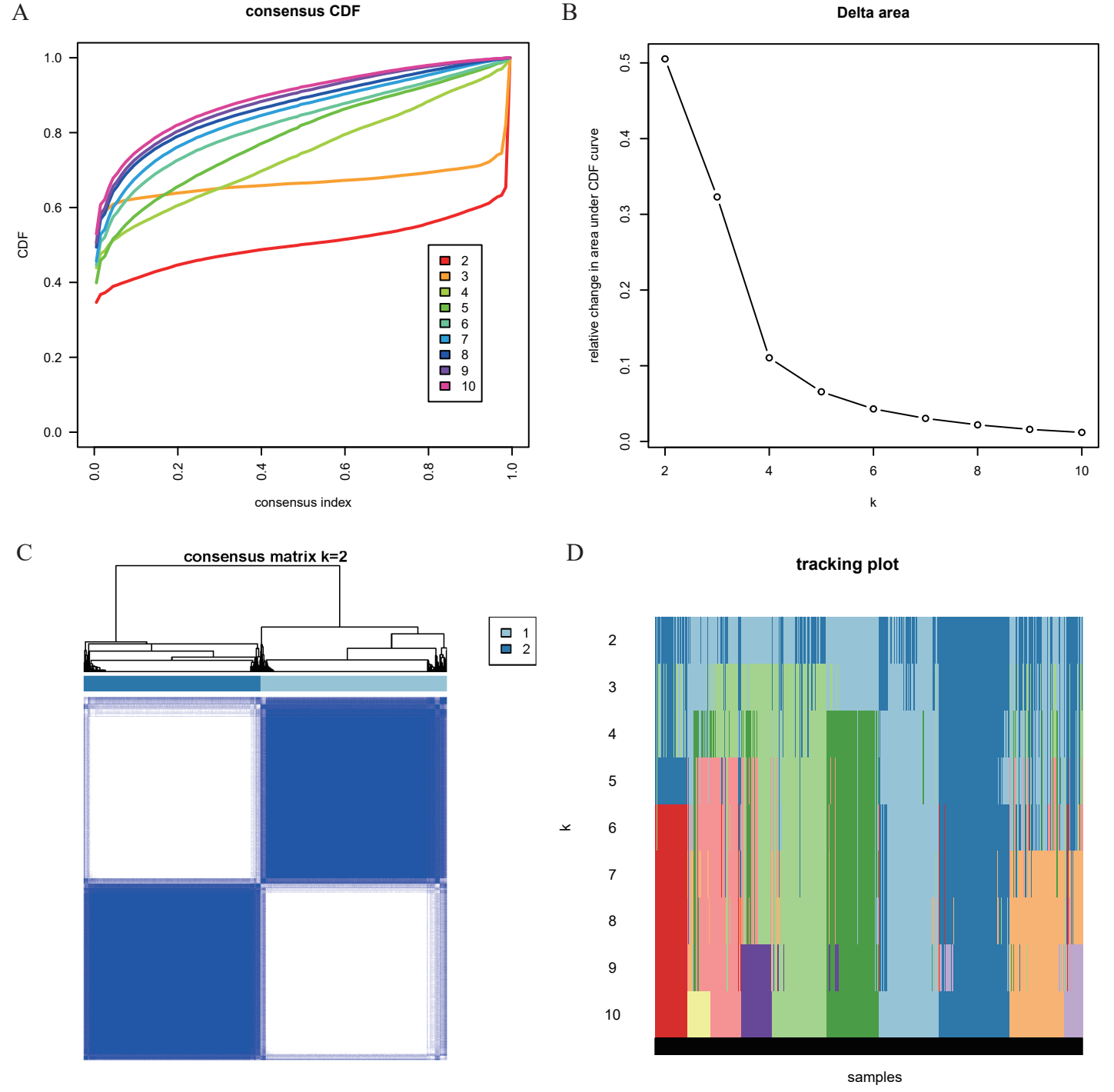
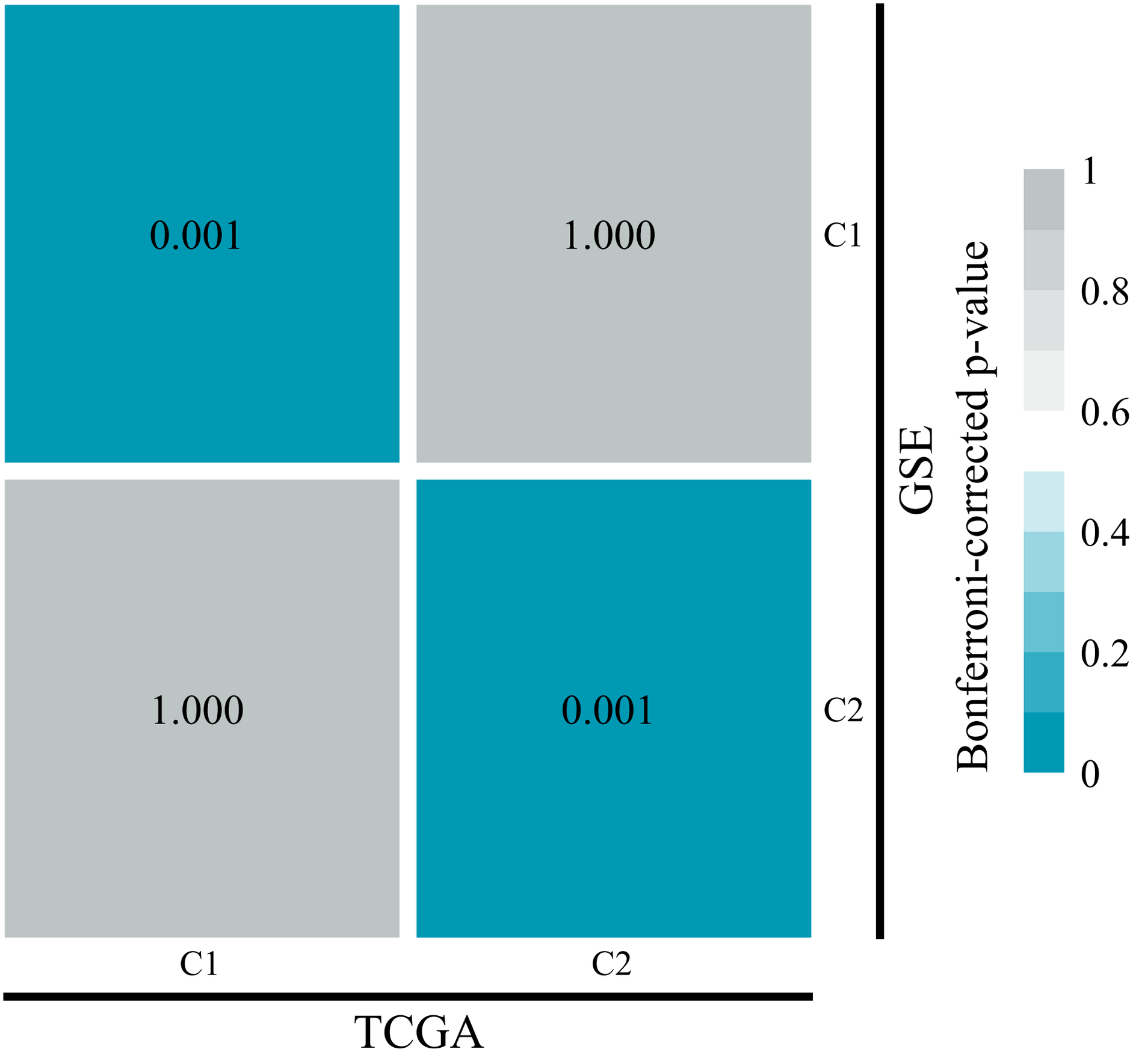
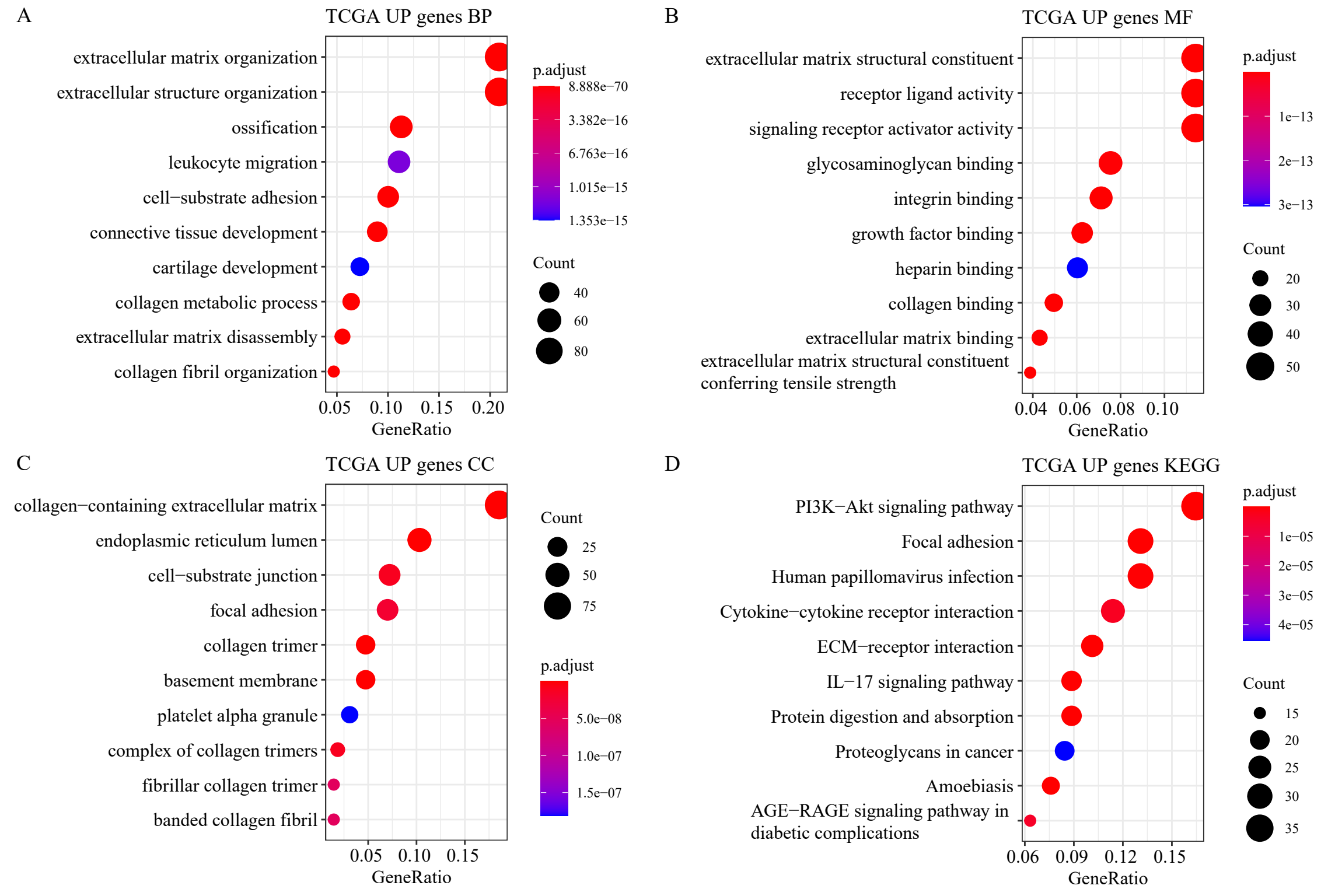
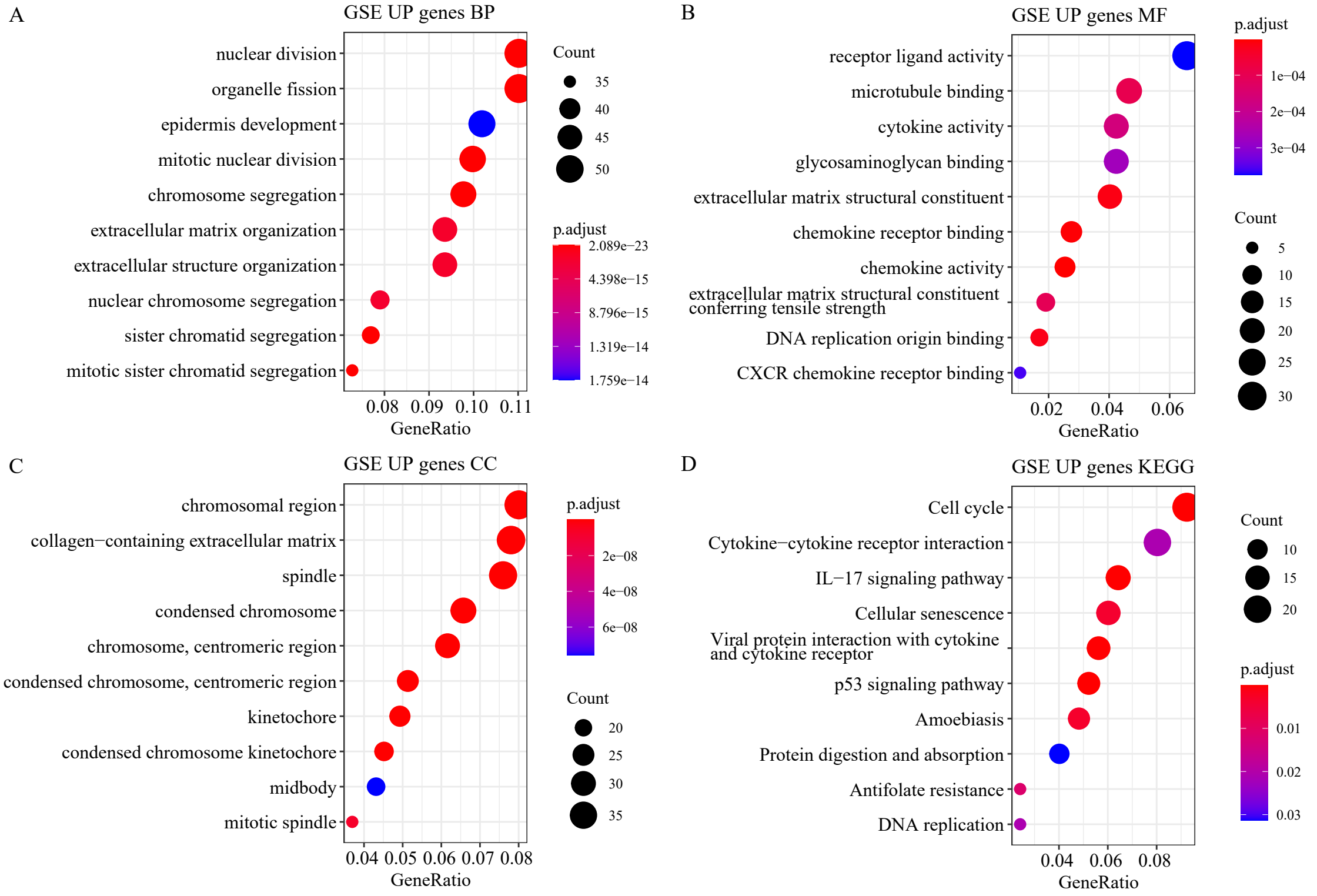
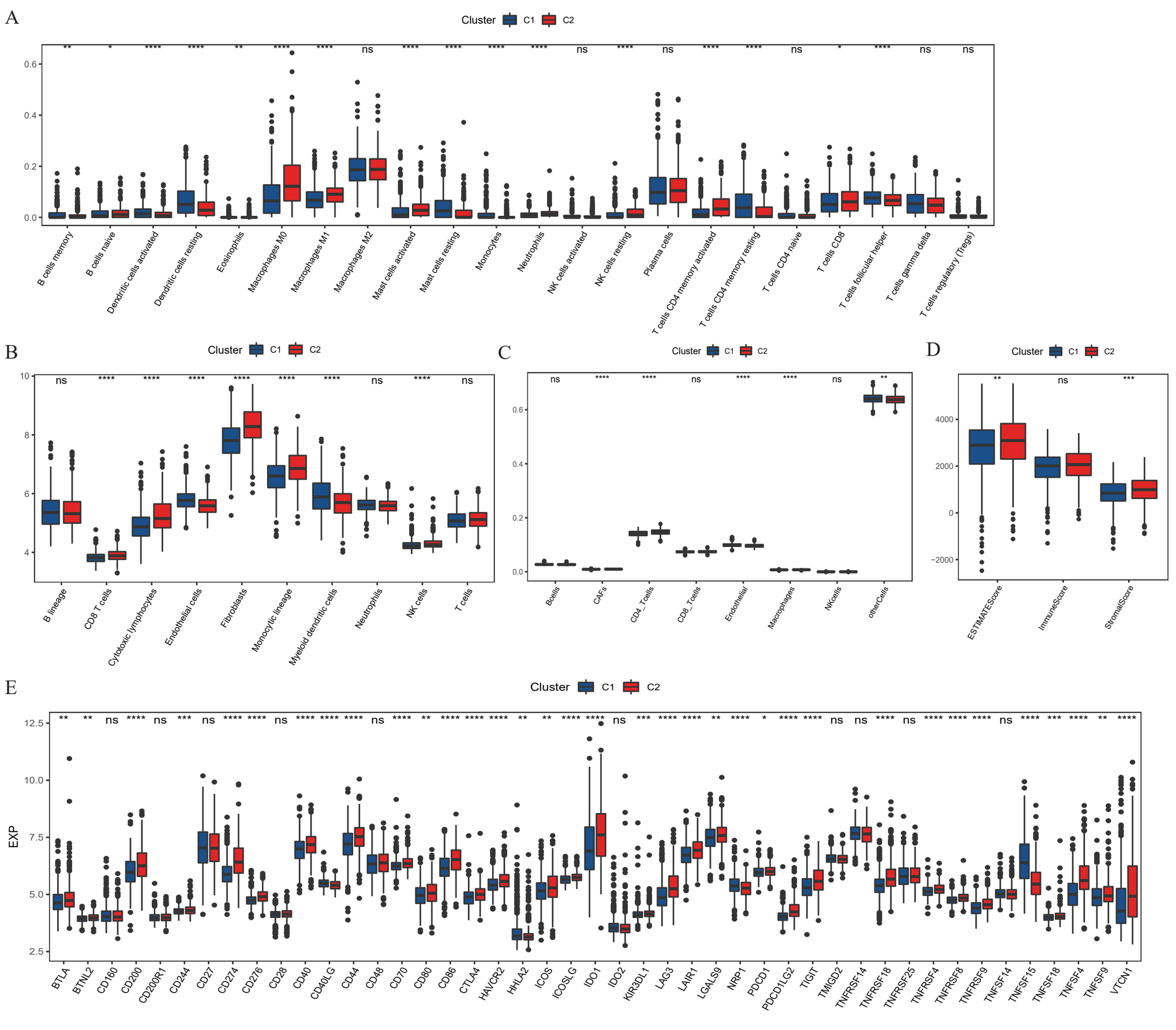

A
TCGA
LUAD
LUSC
PCA2
PCA1
B
TCGA
LUAD
LUSC
PCA2
PCA1
C
GSE
GSE19188
GSE29013
GSE30219
GSE31210
GSE37745
GSE50081
PCA2
PCA1
D
GSE
GSE19188
GSE29013
GSE30219
GSE31210
GSE37745
GSE50081
PCA2
PCA1

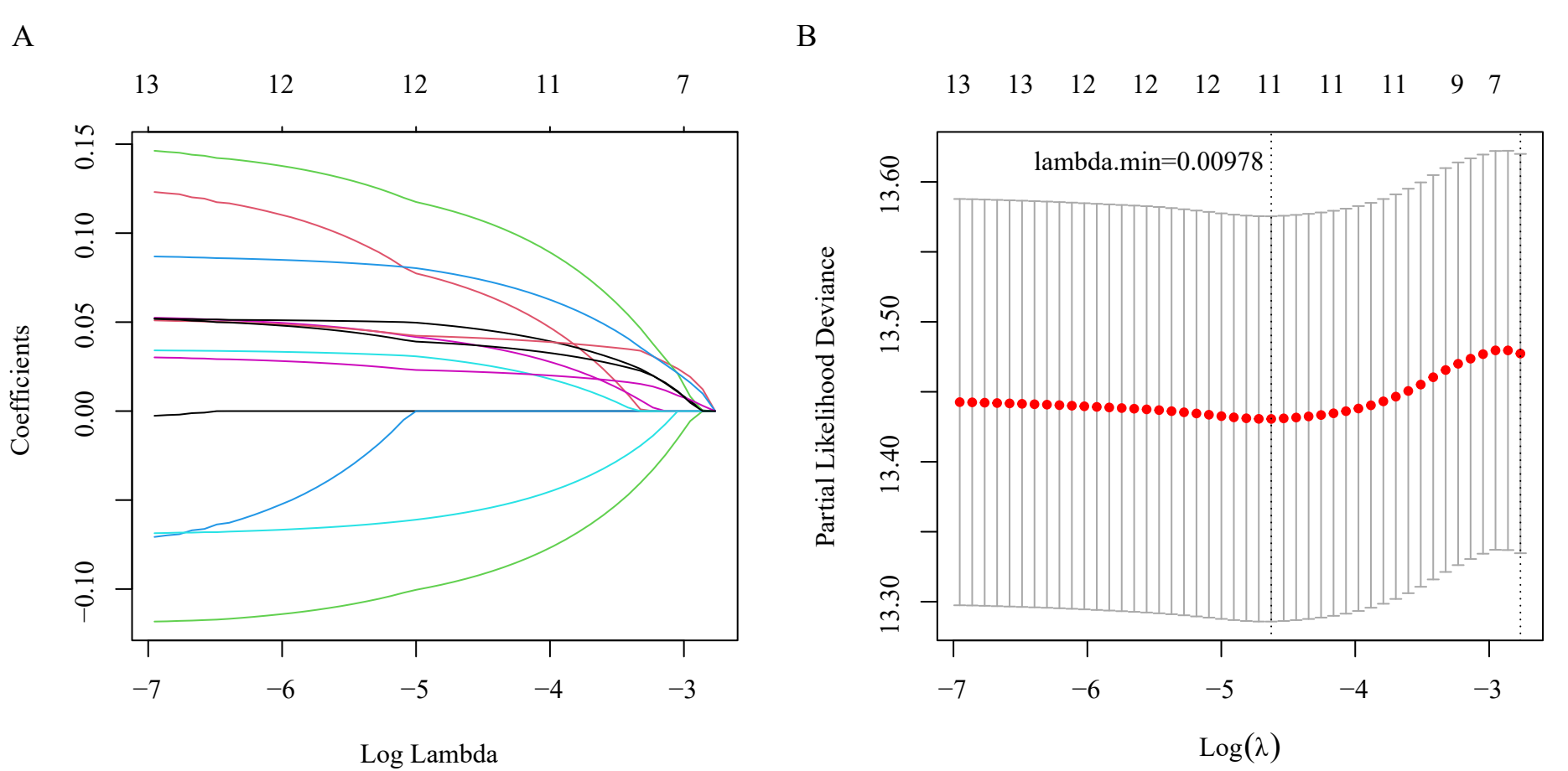

A
consensus CDF
CDF
consensus index
2
3
4
5
6
7
8
9
10
B
Delta area
relative change in area under CDF curve
k
C
consensus matrix k=2
1
2
D
tracking plot
k
samples


0.001
1.000
1.000
0.001
C1
C2
C1
C2
TCGA
GSE
Bonferroni-corrected p-value
1
0.8
0.6
0.4
0.2
0


A
TCGA UP genes BP
extracellular matrix organization
extracellular structure organization
ossification
leukocyte migration
cell−substrate adhesion
connective tissue development
cartilage development
collagen metabolic process
extracellular matrix disassembly
collagen fibril organization
0.05
0.10
0.15
0.20
GeneRatio
p.adjust
8.888e−70
3.382e−16
6.763e−16
1.015e−15
1.353e−15
Count
40
60
80
B
TCGA UP genes MF
extracellular matrix structural constituent
receptor ligand activity
signaling receptor activator activity
glycosaminoglycan binding
integrin binding
growth factor binding
heparin binding
collagen binding
extracellular matrix binding
extracellular matrix structural constituent conferring tensile strength
0.04
0.06
0.08
0.10
GeneRatio
p.adjust
1e−13
2e−13
3e−13
Count
20
30
40
50
C
TCGA UP genes CC
collagen−containing extracellular matrix
endoplasmic reticulum lumen
cell−substrate junction
focal adhesion
collagen trimer
basement membrane
platelet alpha granule
complex of collagen trimers
fibrillar collagen trimer
banded collagen fibril
0.05
0.10
0.15
GeneRatio
Count
25
50
75
p.adjust
5.0e−08
1.0e−07
1.5e−07
D
TCGA UP genes KEGG
PI3K−Akt signaling pathway
Focal adhesion
Human papillomavirus infection
Cytokine−cytokine receptor interaction
ECM−receptor interaction
IL−17 signaling pathway
Protein digestion and absorption
Proteoglycans in cancer
Amoebiasis
AGE−RAGE signaling pathway in diabetic complications
0.06
0.09
0.12
0.15
GeneRatio
p.adjust
1e−05
2e−05
3e−05
4e−05
Count
15
20
25
30
35


A
GSE UP genes BP
nuclear division
organelle fission
epidermis development
mitotic nuclear division
chromosome segregation
extracellular matrix organization
extracellular structure organization
nuclear chromosome segregation
sister chromatid segregation
mitotic sister chromatid segregation
0.08 0.09 0.10 0.11
GeneRatio
Count
35
40
45
50
p.adjust
2.089e−23
4.398e−15
8.796e−15
1.319e−14
1.759e−14
B
GSE UP genes MF
receptor ligand activity
microtubule binding
cytokine activity
glycosaminoglycan binding
extracellular matrix structural constituent
chemokine receptor binding
chemokine activity
extracellular matrix structural constituent conferring tensile strength
DNA replication origin binding
CXCR chemokine receptor binding
0.02 0.04 0.06
GeneRatio
p.adjust
1e−04
2e−04
3e−04
Count
5
10
15
20
25
30
C
GSE UP genes CC
chromosomal region
collagen−containing extracellular matrix
spindle
condensed chromosome
chromosome, centromeric region
condensed chromosome, centromeric region
kinetochore
condensed chromosome kinetochore
midbody
mitotic spindle
0.04 0.05 0.06 0.07 0.08
GeneRatio
p.adjust
2e−08
4e−08
6e−08
Count
20
25
30
35
D
GSE UP genes KEGG
Cell cycle
Cytokine−cytokine receptor interaction
IL−17 signaling pathway
Cellular senescence
Viral protein interaction with cytokine and cytokine receptor
p53 signaling pathway
Amoebiasis
Protein digestion and absorption
Antifolate resistance
DNA replication
0.04 0.06 0.08
GeneRatio
Count
10
15
20
p.adjust
0.01
0.02
0.03


A
Cluster C1 C2
B cells memory
B cells naive
Dendritic cells activated
Dendritic cells resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast cells activated
Mast cells resting
Monocytes
Neutrophils
NK cells activated
NK cells resting
Plasma cells
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
T cells follicular helper
T cells gamma delta
T cells regulatory (Tregs)
B
Cluster C1 C2
B lineage
CD8 T cells
Cytotoxic lymphocytes
Endothelial cells
Fibroblasts
Monocytic lineage
Myeloid dendritic cells
Neutrophils
NK cells
T cells
C
Cluster C1 C2
Bcells
CAFs
CD4_Tcells
CD8_Tcells
Endothelial
Macrophages
NKcells
otherCells
D
Cluster C1 C2
ESTIMATEScore
ImmuneScore
StromalScore
E
Cluster C1 C2
EXP
BTLA
BTNL2
CD160
CD200
CD200R1
CD244
CD27
CD274
CD276
CD28
CD40
CD40LG
CD44
CD48
CD70
CD80
CD86
CTLA4
HAVCR2
HHLA2
ICOS
ICOSLG
IDO1
IDO2
KIR3DL1
LAG3
LAIR1
LGALS9
NRP1
PDCD1
PDCD1LG2
TIGIT
TMIGD2
TNFRSF14
TNFRSF18
TNFRSF25
TNFRSF4
TNFRSF8
TNFRSF9
TNFSF14
TNFSF15
TNFSF18
TNFSF4
TNFSF9
VTCN1


A
13
12
12
11
7
0.15
0.10
0.05
0.00
−0.10
Coefficients
−7
−6
−5
−4
−3
Log Lambda
B
13
13
12
12
12
11
11
11
9
7
lambda.min=0.00978
13.60
13.50
13.40
13.30
Partial Likelihood Deviance
−7
−6
−5
−4
−3
Log(λ)

Supplementary figure legends

Supplementary Figure S1. PCA for eliminating batch effects. (A-B) PCA plot before (A) and after (B) removing batch effects in TCGA datasets. (C-D) PCA plot before (C) and after (D) removing batch effects in GSE cohorts.

Supplementary Figure S2. Unsupervised consensus clustering for identifying molecular subtypes. (A) CDF curve of different cluster number k = 2 to 10. (B) Relative change in area under CDF curve when k = 2 to 10. (C) Consensus matrix when k = 2. (D) The classification of samples when k = 2 to 10.

Supplementary Figure S3. SubMap analysis on C1 and C2 molecular subtypes between TCGA dataset and GSE cohorts.

Supplementary Figure S4. Functional analysis of up-regulated genes in TCGA-NSCLC dataset. (A-C) The top 10 terms of GO annotation including biological process (BP), molecular function (MF) and cellular component (CC). (D) The top 10 enriched KEGG pathways. Dot size represents the gene counts.

Supplementary Figure S5. Functional analysis of up-regulated genes in GSE cohorts. (A-C) The top 10 terms of GO annotation including biological process (BP), molecular function (MF) and cellular component (CC). (D) The top 10 enriched KEGG pathways. Dot size represents the gene counts.

Supplementary Figure S6. Characterization of tumor microenvironment between two subtypes in GSE cohorts. (A) Enrichment of 22 immune cells evaluated by CIBERSORT. (B) Enrichment of 10 immune cells analyzed by MCP-Counter. (C) EPIC measurement for describing 8 immune cells. (D) ESTIMATE method for calculating

stromal score and immune score. (E) Expression of 47 immune checkpoints presented as log2(TPM). Student *t* test was performed. ns, no significance. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. TPM, transcript per million.

Supplementary Figure S7. LASSO Cox regression analysis for optimizing the prognostic model. (A) Correlation coefficients of each variable (gene) changing with the increasing lambda value. (B) Confidence interval of the changing lambda value. The first dotted line represents the optimal lambda = 0.00978.