Supplementary

Table S1 The clinical information of HCC samples with RNA-seq data in TCGA and HCCDB18 datasets

Clinical Features	TCGA-LIHC	HCCDB18
OS		
0	235	168
1	130	35
T Stage		
T1	180	33
T2	91	96
T3	78	59
T4	13	15
TX	3	
N Stage		
N0	248	
N1	4	
NX	113	
M Stage		
M0	263	
M1	3	
MX	99	
Stage		
I	170	
II	84	
III	83	
IV	4	
X	24	
Grade		
G1	55	
G2	175	
G3	118	
G4	12	
GX	5	
Gender		
Male	246	153
Female	119	50
Age		
≤60	173	43
>60	192	160

HCC, hepatocellular carcinoma; TCGA, The Cancer Genome Atlas; OS, overall survival.

https://dx.doi.org/10.21037/jgo-22-1101

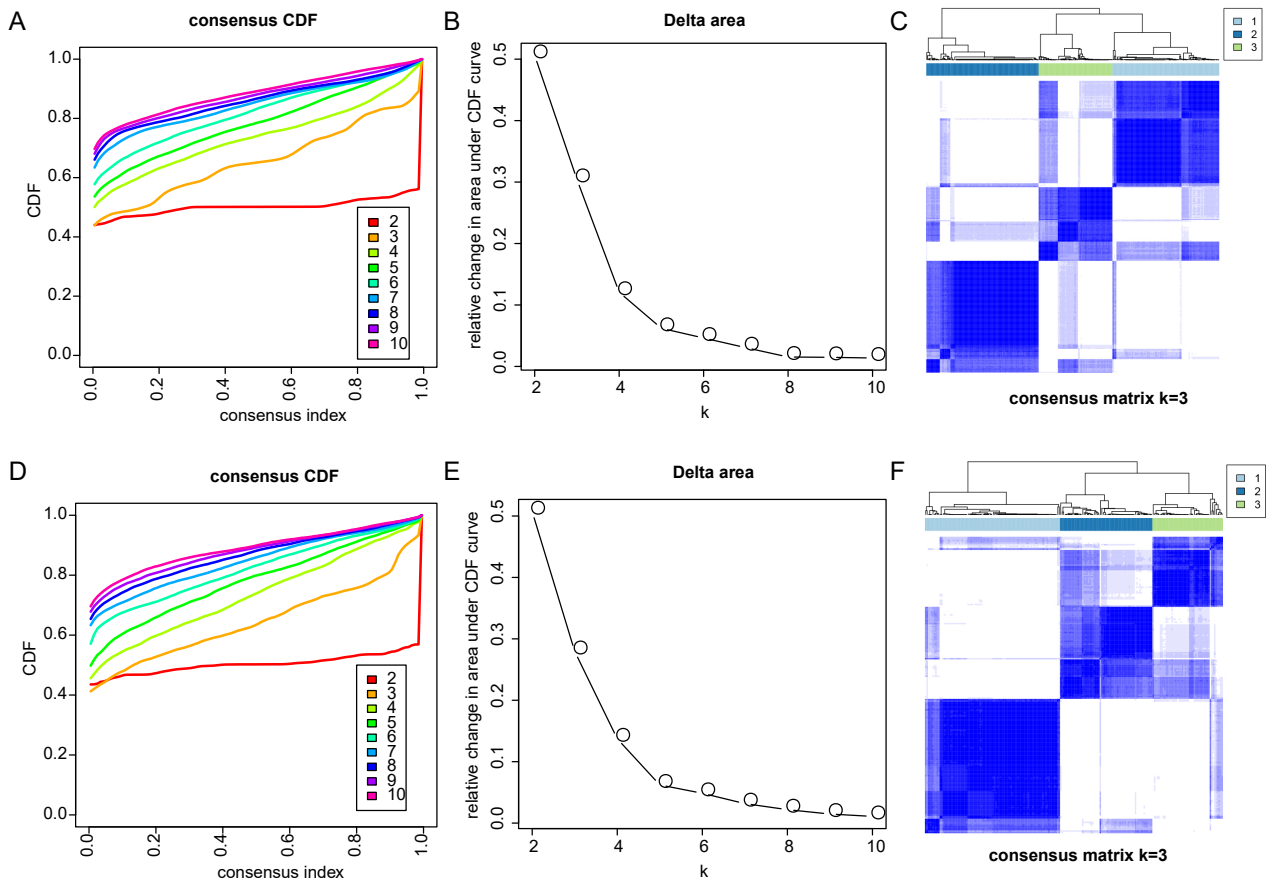

Figure S1 Consensus clustering of HCC samples in TCGA (A-C) and HCCDB18 (D-F) datasets. HCC, hepatocellular carcinoma; CDF, cumulative distribution function.

https://dx.doi.org/10.21037/jgo-22-1101

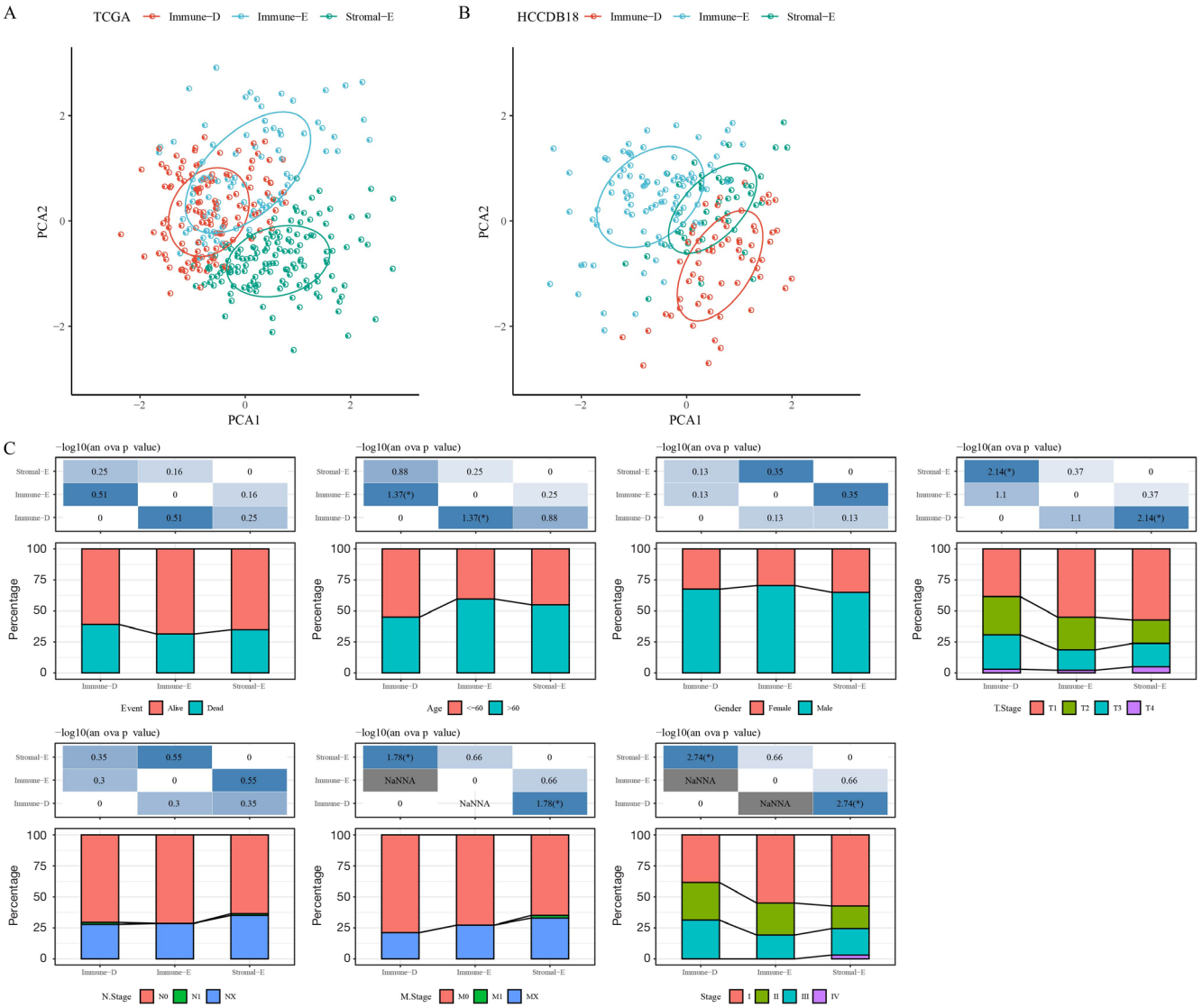

Figure S2 (A,B) PCA plots of the samples in TCGA and HCCDB18 datasets grouped by 3 clusters. (C) The distribution of different clinical characteristics in three clusters in TCGA dataset. ANOVA was conducted. *, P<0.05. NaNNA, no statistical test was performed as the unbalanced distribution of samples. PCA, principal component analysis; TCGA, The Cancer Genome Atlas; ANOVA, analysis of variance.

https://dx.doi.org/10.21037/jgo-22-1101

Figure S3 The methylation beta values of cg sites of EMT and DNA repair-related genes in three clusters. ns, not significant. *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001. EMT, epithelial-mesenchymal transition.

https://dx.doi.org/10.21037/jgo-22-1101

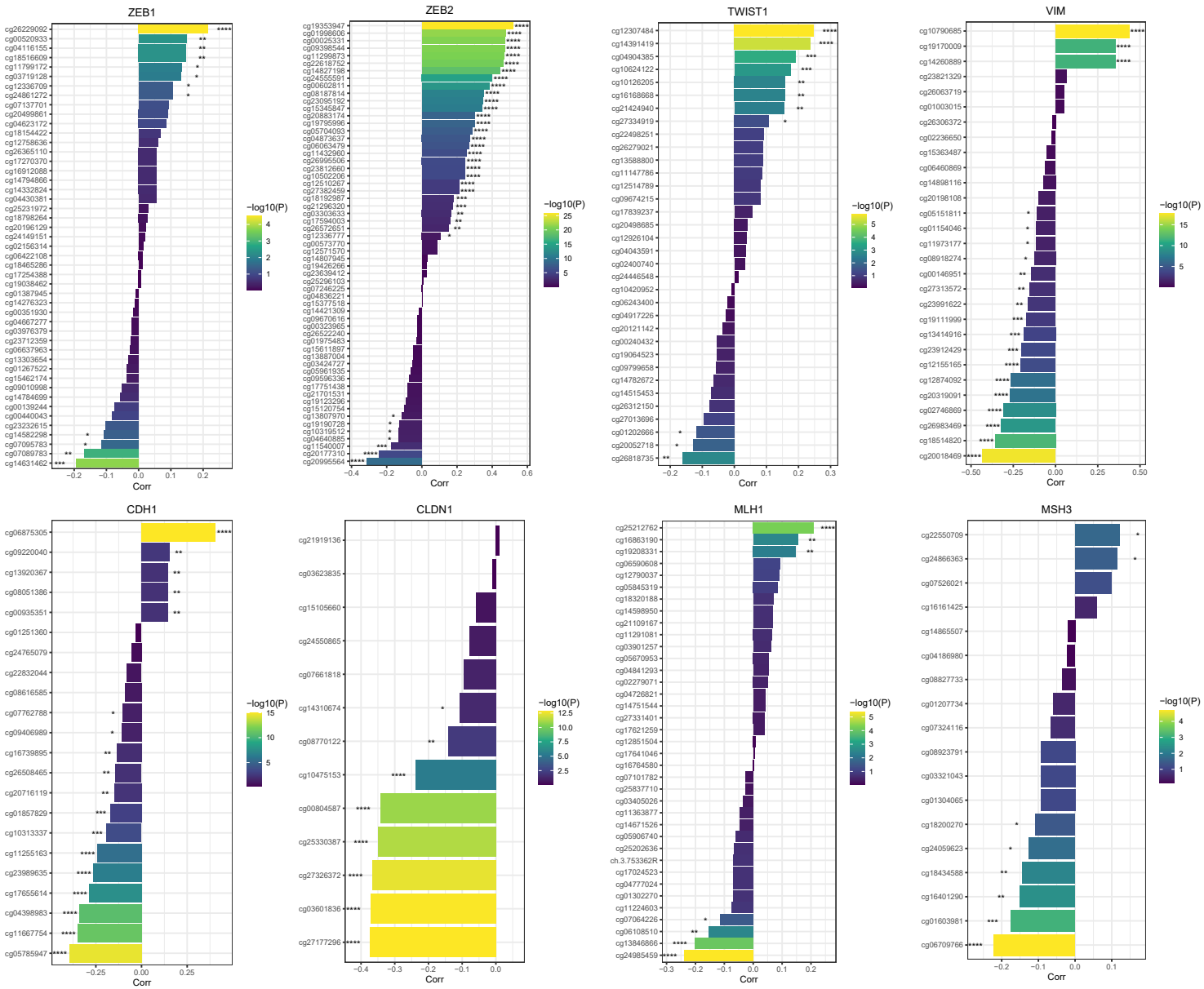

Figure S4 Pearson correlation analysis of gene expression with the methylation beta values of cg sites in EMT and DNA repair-related genes. EMT, epithelial-mesenchymal transition. *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001.

https://dx.doi.org/10.21037/jgo-22-1101

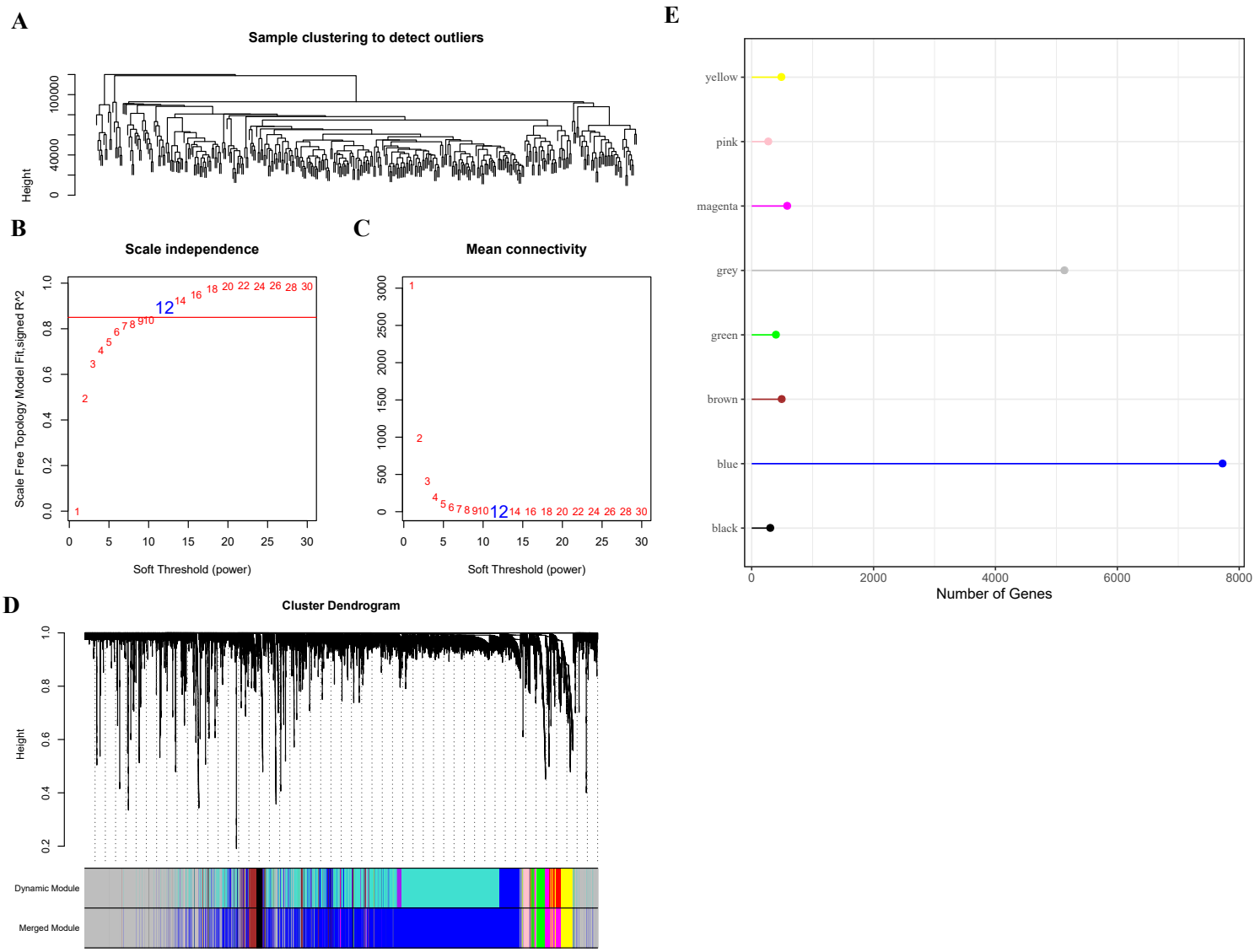

Figure S5 WGCNA for the samples in TCGA dataset. (A) Clustering of samples to screen co-expression modules. (B,C) Determining the soft threshold for constructing a scale-free network. Red line indicates the power of soft threshold = 12 and the fit of scale-free topology model =0.85. (D) Combination of adjacent modules based on epigengenes. (E) Number of genes in 8 gene modules. WGCNA, weighted correlation network analysis; TCGA, The Cancer Genome Atlas.

https://dx.doi.org/10.21037/jgo-22-1101

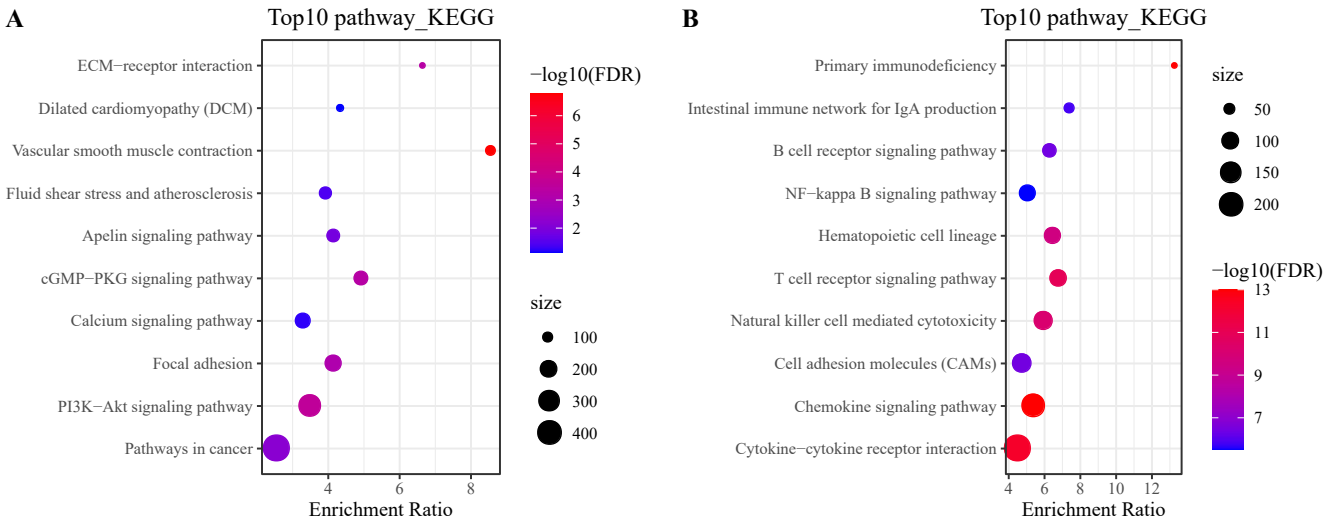

Figure S6 The top 10 enriched KEGG pathways in the pink and yellow modules. KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate.

Table S2 Univariate Cox regression of 15 pathways

Pathways	P value	HR	Low 95% CI	High 95% CI
Natural killer cell mediated cytotoxicity	0.057829	0.05748	0.003005	1.09946
Antigen processing and presentation	0.152237	0.214945	0.026199	1.763487
T cell receptor signaling	0.123136	0.110703	0.006746	1.816703
B cell receptor signaling	0.189311	0.146961	0.008389	2.574611
Fc gamma R-mediated phagocytosis	0.261965	4.876934	0.306055	77.71299
ECM-receptor interaction	0.818133	1.219522	0.224672	6.619584
Focal adhesion	0.771591	0.686531	0.054181	8.699041
Tight junction	0.108424	0.033959	0.000546	2.111249
p53 signaling	0.024364	43.97972	1.631817	1185.314
Mismatch repair	0.009223	10.69646	1.796656	63.68177
Homologous recombination	0.000454	26.7835	4.26408	168.2323
PI3K-Akt signaling	0.479291	0.266955	0.006877	10.36299
Wnt signaling	0.455062	3.219096	0.149815	69.16904
TGF-beta signaling	0.630128	0.530694	0.040264	6.994662
Cell cycle	1.56E-06	137.4686	18.43778	1024.941

CI, confidence interval.

https://dx.doi.org/10.21037/jgo-22-1101

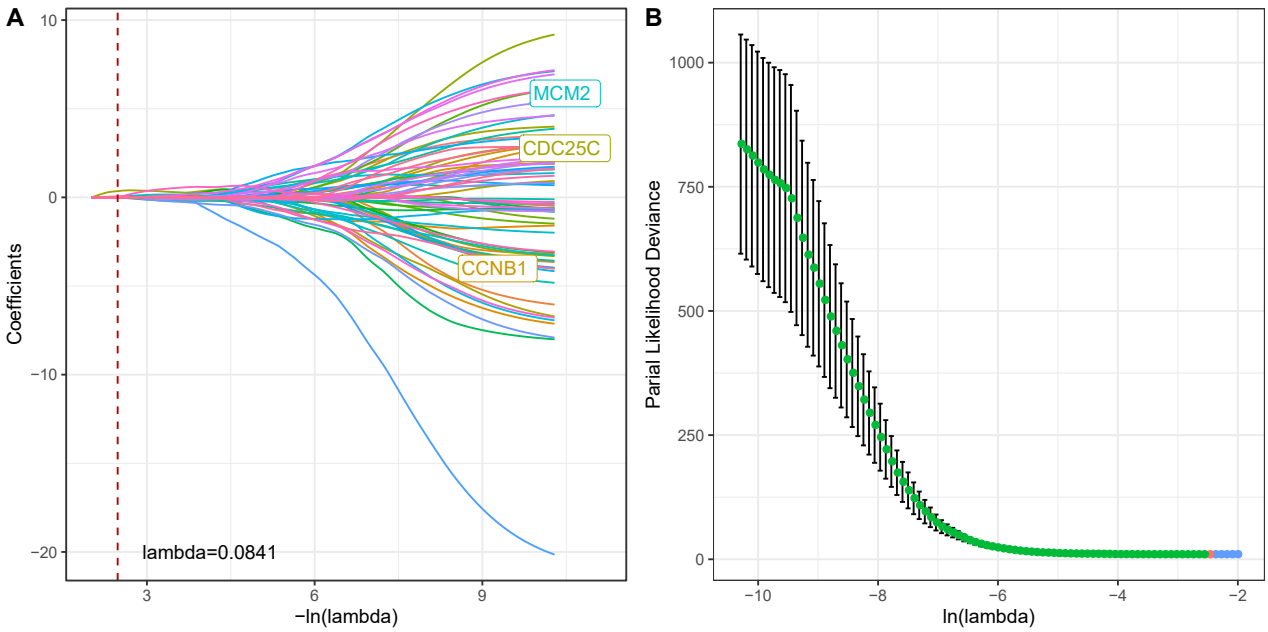

Figure S7 LASSO regression analysis for 70 genes. Dotted red line in (A) and red dot in (B) indicates lambda = 0.0841. KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator.

https://dx.doi.org/10.21037/jgo-22-1101

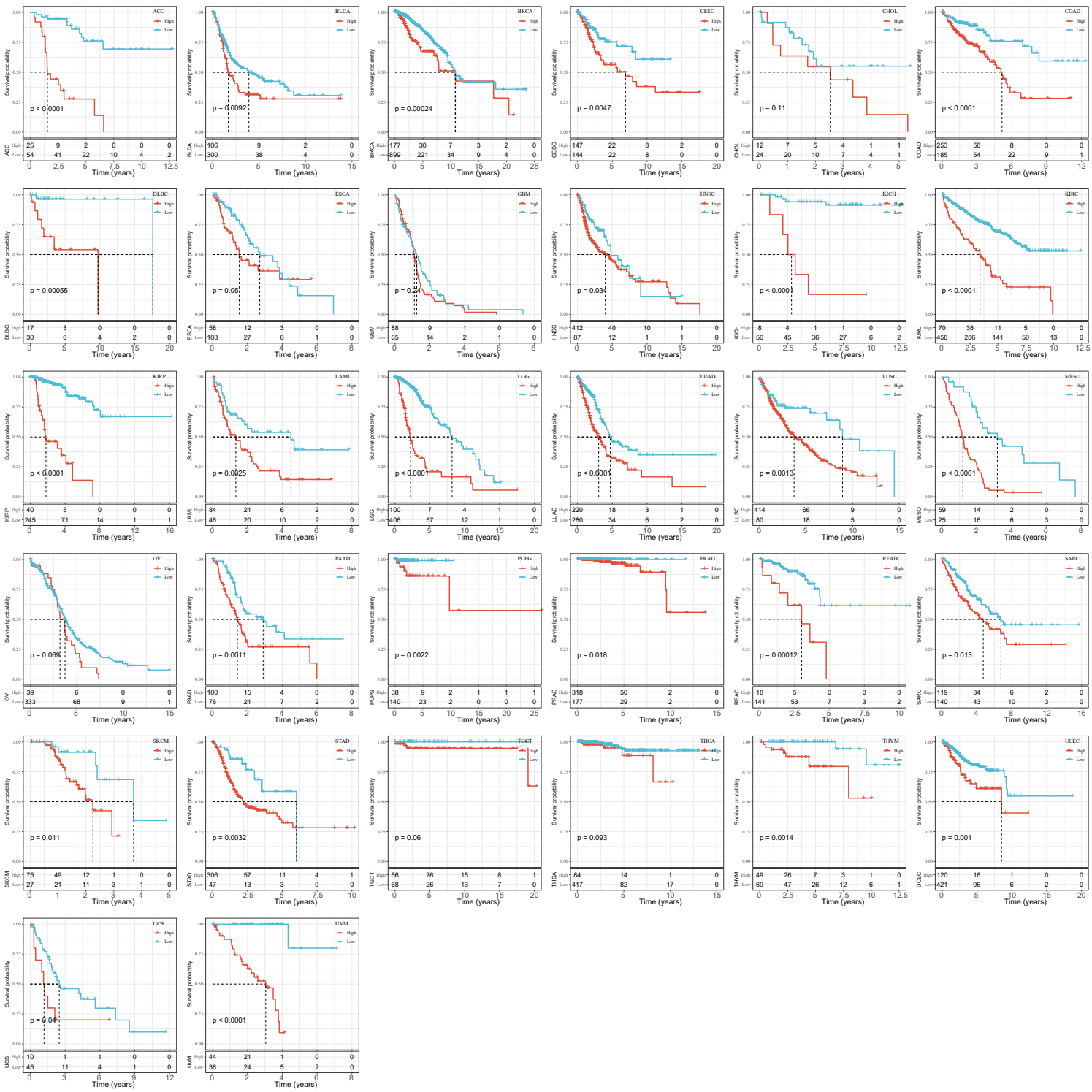

Figure S8 The performance of the IMScore model in 32 cancer types. Log-rank test was conducted.

https://dx.doi.org/10.21037/jgo-22-1101