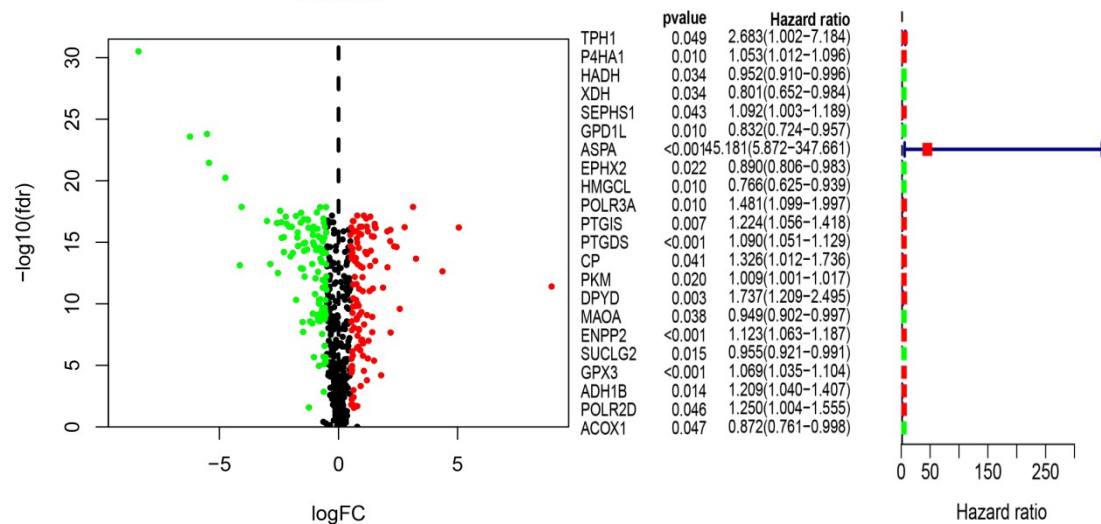(A)

(B)   (C)

Figure 1: Differentially expressed metabolic genes related to survival (A) Heat map showing hierarchical cluster analysis of DEG in TCGA-COAD. (B) Volcano plot of gene expression data. Green dots are down-regulated RNAs. Red dots are up-regulated RNA; black dots are RNA that is not differentially expressed. P Value <0.05 and |log2-fold change|> 0.5 are considered to be statistically significant. (C) Univariate Cox regression analysis forest plot shows the differential genes related to survival in TCGA-COAD. DEG, differentially expressed genes; TCGA-COAD, colon adenocarcinoma gene expression profile.
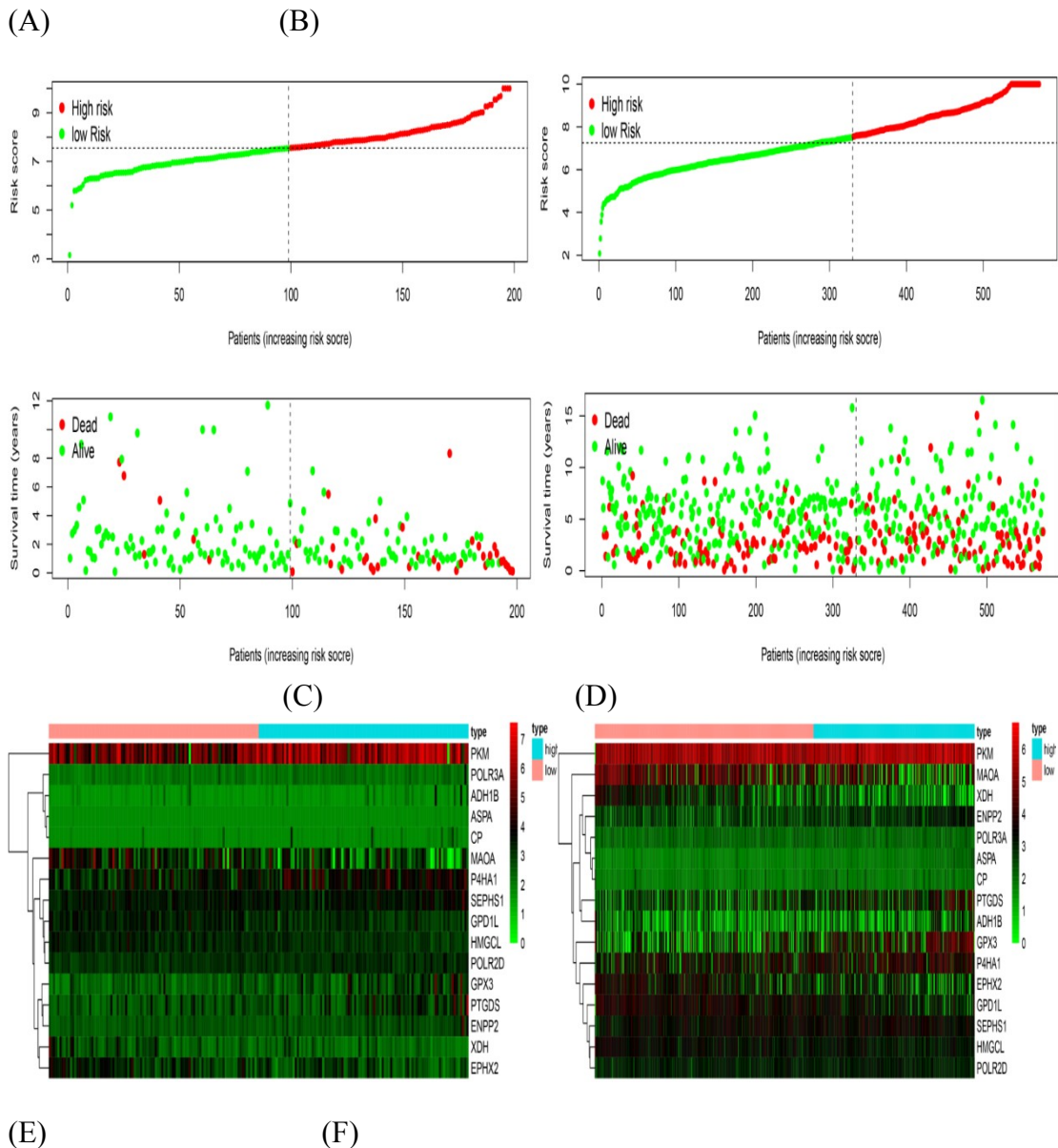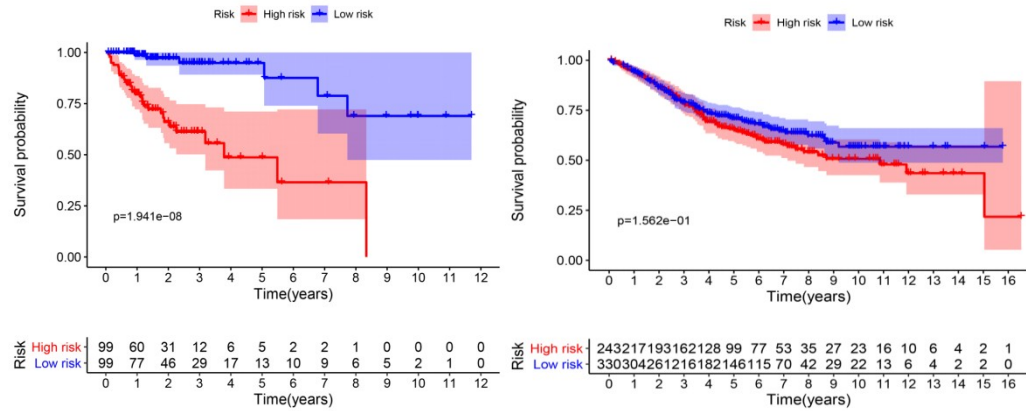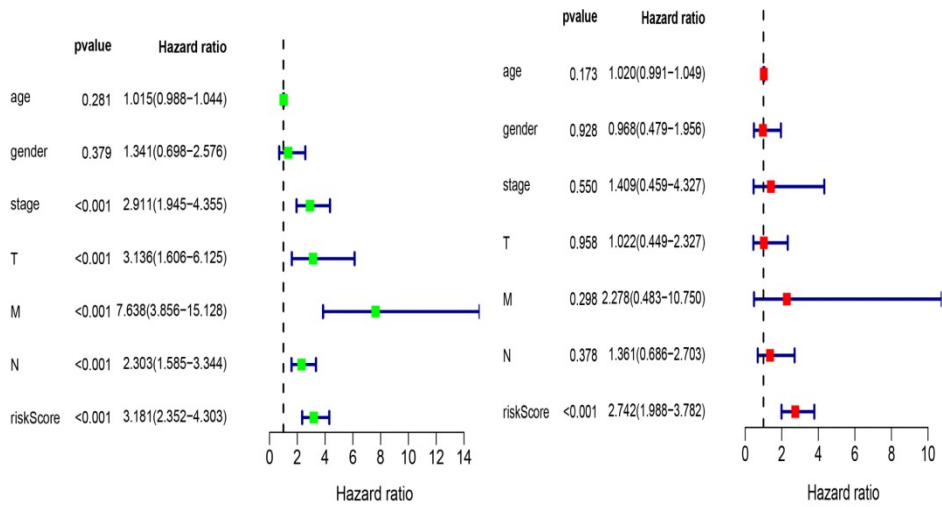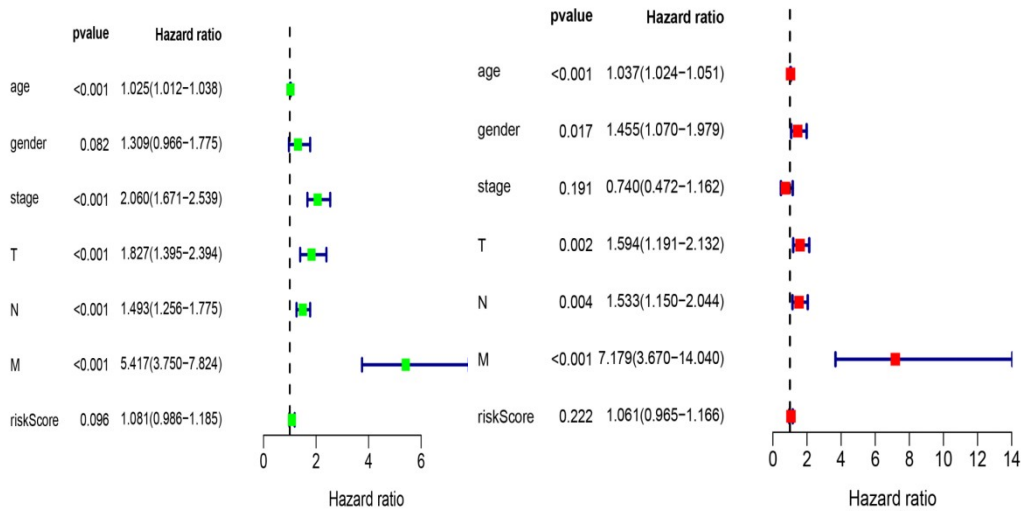
Figure 2: Risk score for predicting survival. (A, B) Kaplan–Meier analysis of the TCGA-COAD and GSE40976 prognostic models has a longer overall survival (OS) time for the high-risk group than the low-risk group. **(C, D)** Distribution of risk scores of sixteen genes in TCGA-COAD and GSE40976 (each point represents a sample). **(E, F)** Cluster enrichment
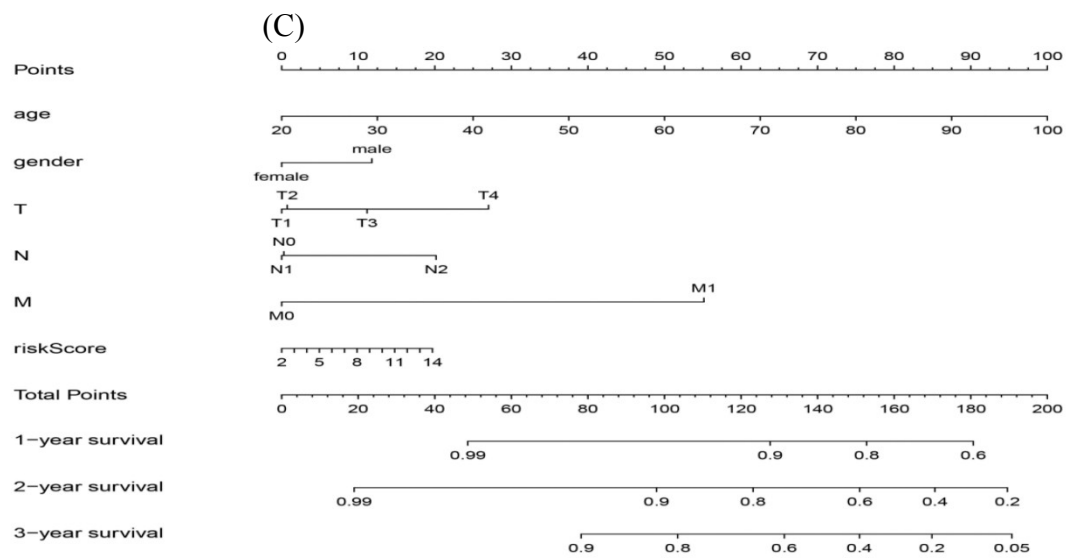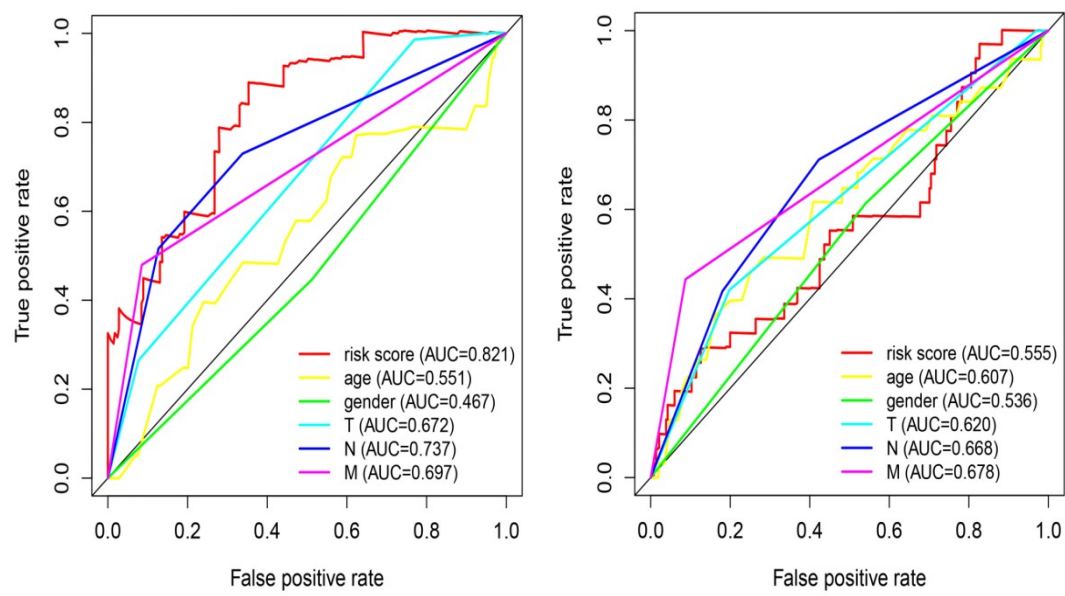
analysis of sixteen genes in TCGA-COAD and GSE40976. GEO, Gene Expression Omnibus.

(A)

(B)

True positive rate

risk score (AUC=0.821)
age (AUC=0.551)
gender (AUC=0.467)
T (AUC=0.672)
N (AUC=0.737)
M (AUC=0.697)

False positive rate

True positive rate

risk score (AUC=0.555)
age (AUC=0.607)
gender (AUC=0.536)
T (AUC=0.620)
N (AUC=0.668)
M (AUC=0.678)

False positive rate

(C)

Points   0   10   20   30   40   50   60   70   80   90   100

age   20   30   40   50   60   70   80   90   100

gender   male
              female

T   T2   T4
              T1   T3

N   N0
              N1   N2

M   M1
              M0

riskScore   2   5   8   11   14

Total Points 0   20   40   60   80   100   120   140   160   180   200

1-year survival   0.99   0.9   0.8   0.6

2-year survival   0.99   0.9   0.8   0.6   0.4   0.2

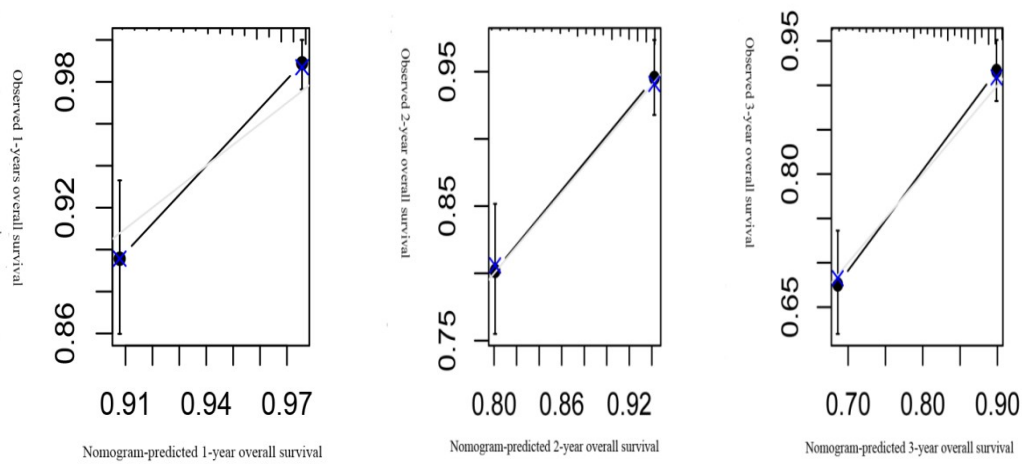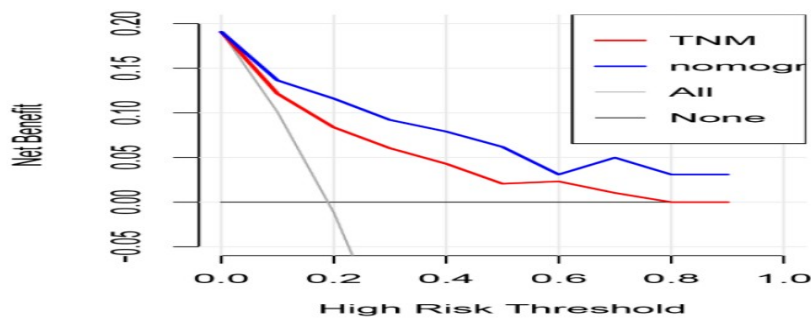3-year survival   0.9   0.8   0.6   0.4   0.2   0.05

(D)

(E)

(F)

Figure 3: Establishment and verification of the nomogram prediction model. (A,B) univariate and multivariate analysis forest plots in TCGA-COAD and GSE40976. (C) Time-dependent ROC analysis in TCGA-COAD and GSE40976 showed the most AUC curve area. (D) Establish a nomogram on GEO. (E) Calibration curve for nomogram1-year, 2-year and 3-year overall suvival. (F)The clinical decision curve shows that the clinical benefit rate of the nomogram is higher than that of the TNM staging system. Black horizontal line: all patients died; black oblique line: no patient died; blue solid line: nomogram prediction model ; Red solid line: TNM staging system model. AUC, Area Under the ROC curve; ROC, Receiver Operating Characteristic.
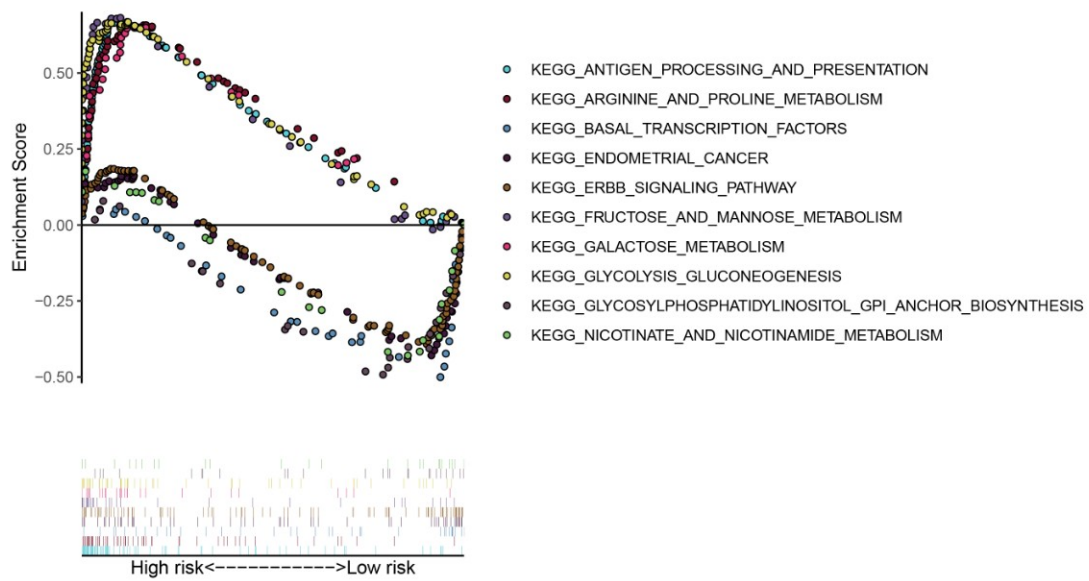
Figure 4: 10 representative KEGG enrichment pathways analyzed by GSEA. Each group contains 5 KEGG pathways. Table 2 lists the relevant parameters of the channel. GESA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes.