Supplementary Materials and Methods

Process public sequencing data.

Single cell RNA-sequencing data set GSE149614 was downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). we collected transcriptome data from 33 types of tumor patients in the TCGA (https://portal.gdc.cancer.gov/) database for 11,057 samples (10,327 tumor samples and 730 matched paraneoplastic samples), and we also downloaded 9 types of GSE data set about HCC from the GEO database including 2,081 cases in the GEO database. In addition, RNA sequencing (RNA-Seq) and clinicopathological data of liver cancer patients were downloaded from Genomic Data Commons (GDC)database (https://portal.gdc.cancer.gov/), including 374 tumors samples for the next analysis. The data downloaded above is analyzed and processed by R language.

Single-cell RNA-sequencing (scRNA-seq).

In order to clear the cell populations of the HCC, we used the R (version 3.5.2, https://www.r-project.org/) and Seurat15,16 R package (version 3.1, https://satijalab.org/seurat/). Subsequently, we carried out pseudotime trajectories analysis by the Monocle R package (version2.10.1, http://cole-trapnell-lab. github.io/monocle-release/) to understand the fate of tumor cell differentiation.

Correlation between ENG Expression and Prognosis of Tumor Patients.

We downloaded the normal tissue samples and cancer tissue samples of 33 kinds of cancer patients in TCGA database, and the number of normal tissue samples of 33 kinds of tumors in GTEx (https://toil.xenahubs.net/download/gtex\_RSEM\_gene\_tpm.gz) database. We also downloaded the normal tissue samples and cancer tissue samples of 9 kinds of GSE data sets in GEO database. After the data was normalized, we used Wilcoxon test to analyze whether there are differences in the expression of ENG in these tumors and healthy tissues. *P* < 0.05 indicates a statistically significant difference in expression.

Survival Analysis

We first downloaded the survival time information (including overall survival, disease-specific survival, disease-free interval, and progression-free interval) and clinical information of 374 samples of HCC tumors in TCGA cohort. Subsequently, we extracted the clinical information (AFP, invasion levels, virus, BMI, age, sex, stage and grade) of patients in the downloaded and ECs marker ENG. These variables were analyzed using univariate and multivariate Cox regression models. According to the median expression value of ENG, each HCC sample was sorted and divided into high expression group and low expression group. Then, the Kaplan-Meier survival method based on TCGA HCC cohort was used to analyze the relationship between ENG expression level and overall survival (OS), disease-specific survival (DSS), disease-free interval (DFI), or progression-free interval (PFI). A *p* value < 0.05 was regarded as statistically significant.

**Enrichment Analysis**

We first divided the TCGA-LIHC RNA-Seq data into high expression and low expression according to the expression level of ENG. Using a threshold of |lgFC| > 1 and P Val < 0.05, we found that a total of 7,689 differential expressed genes (DEGs) were screened out from differential expression analysis. Subsequently, the DEGs were used for enrichment analysis, including functional Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, and the GO consist of biological process (BP), cellular component (CC), molecular function (MF). In this study, the ‘clusterProfiler’ R package was used to analyze GO enrichment and KEGG pathway. In order to accurately reveal the biological function of DEGs, P-value [ adjusted by false discovery rate (FDR)] <0.01 was set as the cut off criteria.

Gene Set Enrichment Analysis (GSEA)

In order to further clarify the biological pathways that differed significantly in high and low ENG expressing HCC patients, we performed GSEA (version 4.0.3) gene enrichment analysis according to the MSigDB molecular signatures database (version 7). A *P* value less than 0.05 was considered as statistically significant difference.

Correlation analysis between TME and ENG expression levels

ESTIMATE (https ://bioinformatics. mdanderson .org/public-software/estimate/) is a tool used to obtain the immune and stromal scores for each HCC sample. After removing normal samples, according to the ratio of stromal cells and immune cells in each HCC tumor sample, estimate and limma package were used to estimate tumor purity. Then, we used the spearman correlation test to calculate and analyze the correlation between the ENG expression data and the TME score data, and used the ggplot2 (https ://CRAN.R-project.org/package=ggplot2), ggpubr and ggExtra (https ://CRAN.R-project.org/package=ggExtra) packages to plot the correlation to map the correlation distribution.

Correlation analysis of immune cell infiltration and ENG expression

we first use immune cells marker and Gene Set Variation Analysis (GSVA) to quantify the immune cells of HCC patients in TCGA. Subsequently, we calculated 22 immune cell infiltration scores in each HCC tumor sample by using the CIBERSORT algorithm in R software. Then, the correlation between the level of individual immune cell infiltration in 22 immune cells in HCC and the expression of ENG was analyzed by Spearman’s correlation test.