

Table S1 Primer sequences

No.	Genes	Primer sequences (5'→3')
1	<i>CCL14</i>	F: ACAGCTAAAGTTGGTGGGG R: CCCCTTTTGGTGATGAAGACAA
2	<i>CLEC4G</i>	F: GAGAGTCCGGTTCAGAACAGCTCCT R: TCCCCCAGATTCCAGTGGCTGAAG
3	<i>FCN2</i>	F: CCAGCTCCCATGTCTAAAGG R: TTACAAACCGTAGGGCCAAG
4	<i>IGFBP3</i>	F: AAATGCTAGTGAGTCGC R: TGTCTATGGGTCTTGAA
5	<i>CXCL14</i>	F: TATCATCACCACCAAGAGCG R: CTATTCTTCGTAGACCCTGCG
6	<i>VIPR1</i>	F: TCATCCGAATCCTGCTTCAGA R: AGGCGAACATGATGTAGTGACT

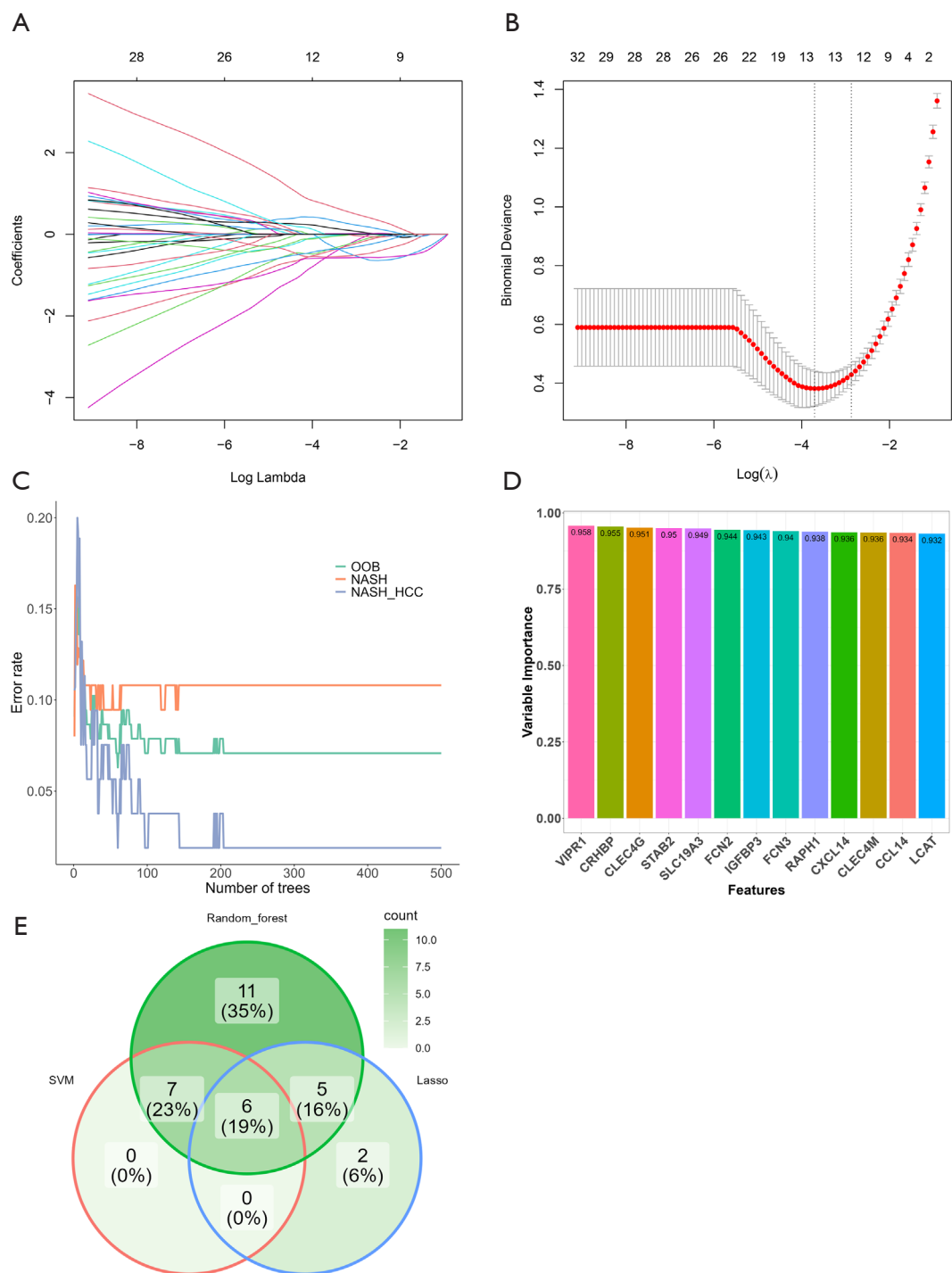


Figure S1 The machine algorithms for signature genes. (A,B) Penalty plot and cross-validation curve of LASSO model. (C) The error rate confidence intervals for random forest model. (D) The relative importance of genes selected by SVM-RFE. (E) Venn diagram of intersecting genes screened by three machine learning models.

Table S2 Feature genes screened by three machine learning models

Model	N	Selected feature genes
LASSO	13	<i>CAP2, CCL14, CDHR2, CLEC4G, CXCL14, FCN2, IGFBP1, IGFBP3, IL13RA2, OBSL1, PLVAP, VIPR1, ZIC2</i>
Random forest	29	<i>STAB2, CRHBP, VIPR1, CLEC4G, SLC19A3, FCN3, IGFBP3, FAM13A, ZIC2, CLEC4M, FCN2, CAP2, OBSL1, CCL14, ACACB, DGAT2, CXCL14, LPA, LCAT, IL13RA2, LOC340888, LPAL2, CYP2C9, CDHR2, NR1I2, RAPH1, DNASE1L3, CTGF, INS-IGF2</i>
SVM-REF	13	<i>VIPR1, CRHBP, CLEC4G, STAB2, SLC19A3, FCN2, FCN3, IGFBP3, CXCL14, LCAT, CLEC4M, RAPH1, CCL14</i>
Intersection	6	<i>CCL14, CLEC4G, FCN2, IGFBP3, CXCL14, VIPR1</i>

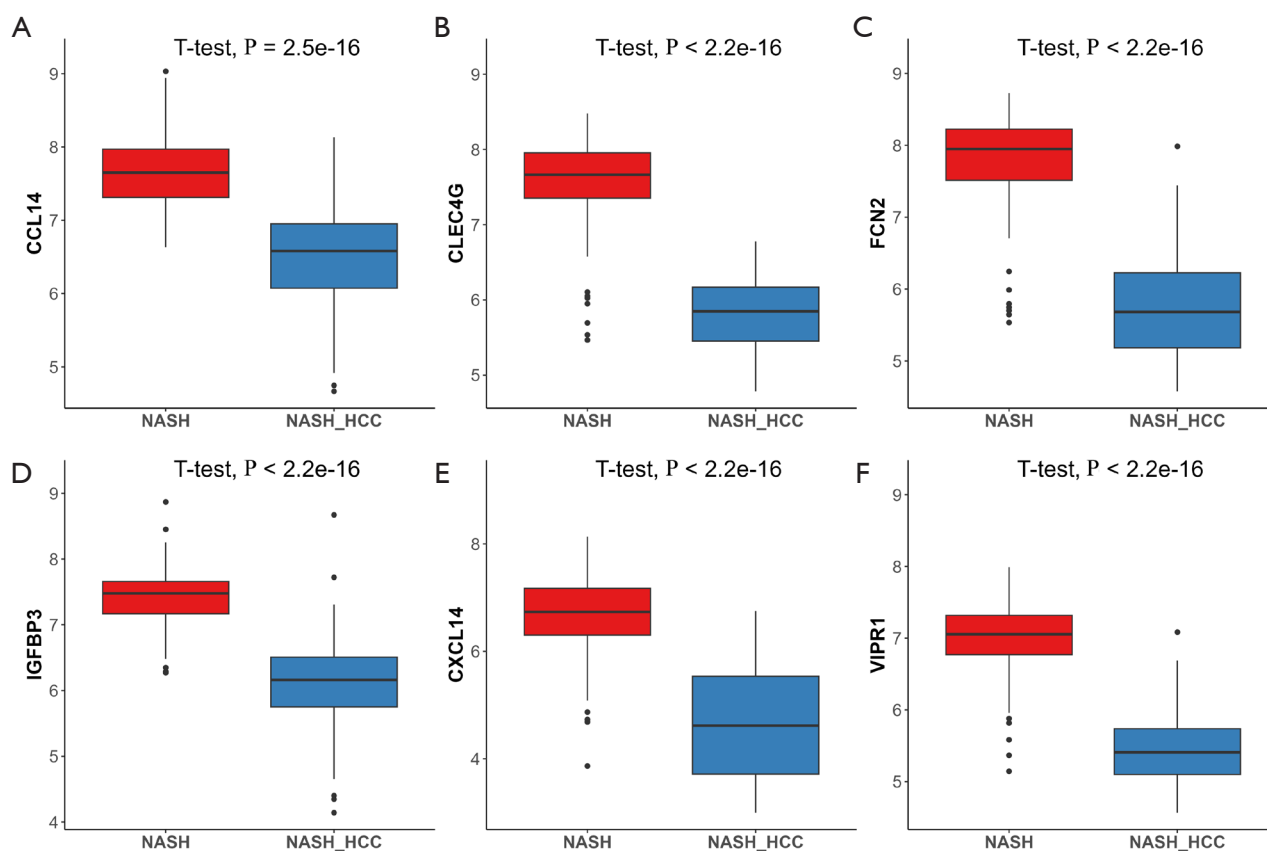


Figure S2 Box plots of the six genes screened in the GSE164760 dataset. (A) *CCL14*. (B) *CLEC4G*. (C) *FCN2*. (D) *IGFBP3*. (E) *CXCL14*. (F) *VIPR1*.

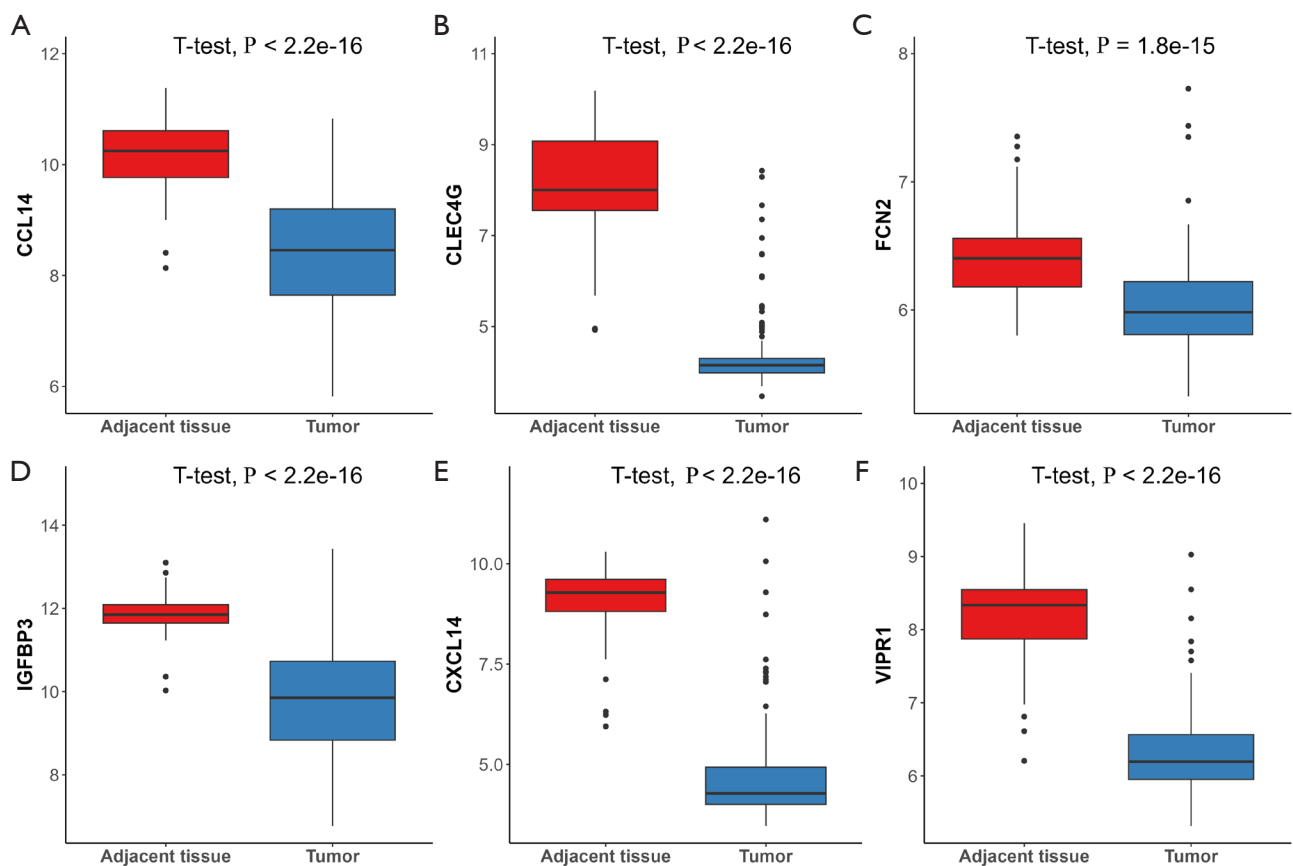


Figure S3 Box plots of the six genes in the GSE102079 dataset (external validation dataset). (A) *CCL14*. (B) *CLEC4G*. (C) *FCN2*. (D) *IGFBP3*. (E) *CXCL14*. (F) *VIPR1*.

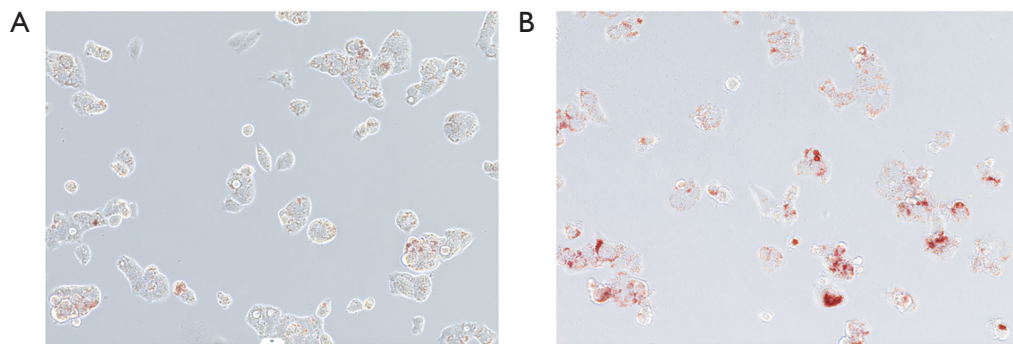


Figure S4 The Oil Red O staining results were observed using a fluorescence microscope, with a magnification of 20 $\times$. (A) Untreated HepG2 cells. (B) HepG2 cells were treated with 0.3 mmol/L sodium palmitate for 24 hours.