Supplementary Table 3. Top 100 differentially modified genes between early HCC and controls in the training set.

Gene	Coefficient [a]	p-value	AUC (Early HCC vs. Controls) [b]	TCGA $\log_2$(FC) [c]	Differential Gene (TCGA) [c]
SOX9	0.722	5.3E-44	81.30%	0.142	No
PDX1	0.730	5.5E-35	76.50%	-0.211	No
ESRRG	0.390	1.4E-33	79.50%	-0.583	Yes
EPB41L4A	0.579	7.3E-30	75.00%	-0.249	Yes
TNFRSF11B	0.528	3.6E-28	75.20%	-0.342	Yes
GABRB3	0.329	7.7E-28	76.30%	-0.877	Yes
CHST9	0.398	2.1E-27	75.50%	-0.299	Yes
TMCC3	0.916	4.7E-26	72.60%	0.097	No
A2M	0.761	4.8E-26	72.40%	-0.106	Yes
ADAMTS9	0.580	1.2E-25	74.10%	0.195	Yes
ENPP2	0.876	3.3E-25	67.40%	0.088	No
EVC	0.703	8.4E-25	76.20%	-0.176	No
FAT1	0.335	2.2E-23	74.10%	0.153	Yes
APOA5	0.703	6.7E-22	70.90%	-0.225	Yes
ADH1B	0.448	4.4E-21	72.30%	-0.244	Yes
SERPINA10	0.479	4.5E-21	72.00%	-0.119	Yes
NPBWR1	0.651	2.0E-20	73.10%	-0.877	Yes
APOC1	0.729	2.9E-20	69.80%	-0.045	Yes
CYB5A	0.984	3.5E-20	67.90%	-0.081	Yes
ADH4	0.445	4.2E-20	72.30%	-0.386	Yes
ARSJ	0.397	6.3E-20	69.00%	-0.209	Yes
STAT1	0.809	1.0E-19	69.30%	0.052	No
FRMD6	0.714	1.9E-19	70.20%	-0.125	No
EXT1	0.604	5.6E-19	67.90%	0.038	No
CYP2C18	0.587	6.7E-19	68.40%	-0.215	Yes
ABCA1	0.920	1.0E-18	70.30%	-0.079	Yes
APOE	0.560	2.8E-18	68.80%	0.016	No
CYP26C1	0.673	1.0E-17	68.70%	0.227	No
TNFSF15	0.611	1.2E-17	67.20%	0.416	Yes
RAPH1	0.430	1.4E-17	69.10%	-0.139	No
PDE7B	0.736	1.8E-17	69.90%	-0.330	Yes
RGL1	0.784	4.2E-17	67.50%	-0.078	No
ADAMTS16	0.298	4.6E-17	70.90%	0.723	Yes
MEIS2	0.398	6.5E-17	71.40%	0.089	Yes
PROX1	0.239	6.9E-17	69.70%	-0.058	No
FAM13C	0.361	8.9E-17	69.40%	0.260	Yes
TMCO5A	0.412	1.0E-16	71.80%	0.539	No
PLXNA4	-0.690	1.2E-16	68.00%	-0.534	Yes
APOB	0.286	1.7E-16	70.50%	-0.065	Yes
SMIM12	-0.863	2.8E-16	67.50%	-0.027	No
POR	0.635	4.0E-16	67.50%	-0.029	No
OSMR	0.485	4.5E-16	67.30%	-0.064	No
OTX1	0.575	4.9E-16	69.10%	1.839	Yes
CECR2	-0.829	6.0E-16	66.70%	0.008	No

GPRC5C	0.657	6.4E-16	67.60%	0.003	No
SOWAHC	0.577	8.3E-16	66.90%	-0.090	Yes
FMO3	0.296	1.4E-15	69.60%	-0.167	Yes
TBX3	0.375	1.5E-15	70.00%	0.001	No
RAPGEF5	0.787	1.7E-15	67.50%	0.002	No
LBP	0.597	1.8E-15	70.20%	-0.087	Yes
TAT	0.652	1.9E-15	65.20%	-0.277	Yes
NUAK1	0.373	2.0E-15	68.70%	0.200	Yes
CYP27A1	0.428	2.3E-15	67.80%	-0.088	Yes
CLVS2	0.317	2.3E-15	67.70%	0.860	No
DNAH5	0.320	2.8E-15	71.30%	-0.099	No
ERRFI1	0.871	4.0E-15	66.30%	-0.137	Yes
MFAP3L	-0.675	4.5E-15	67.50%	-0.368	Yes
ROBO1	0.280	5.2E-15	66.10%	0.406	Yes
TEX14	-1.071	6.0E-15	66.30%	0.095	No
FGD6	1.174	6.0E-15	63.10%	0.126	Yes
TSHZ2	0.318	7.0E-15	68.80%	-0.007	No
SLC16A11	0.635	1.0E-14	65.70%	0.096	No
OCA2	0.333	1.4E-14	72.00%	0.338	No
FAM20A	0.453	1.4E-14	68.50%	0.027	No
SLC10A1	0.574	1.5E-14	67.10%	-0.360	Yes
MAT1A	0.617	1.7E-14	66.50%	-0.174	Yes
ZNF532	0.670	1.8E-14	64.10%	0.203	Yes
NNMT	0.723	2.0E-14	68.20%	-0.341	Yes
RELN	0.299	2.2E-14	69.20%	-0.314	Yes
OAF	0.669	2.6E-14	67.10%	-0.070	Yes
YAP1	0.282	2.7E-14	67.70%	0.041	No
SOX13	0.508	2.7E-14	67.80%	0.159	Yes
VCAN	0.432	3.3E-14	68.20%	0.135	No
FBXL7	0.234	3.6E-14	65.20%	0.161	Yes
HOMER2	-0.767	3.7E-14	64.60%	-0.077	Yes
SDC2	0.300	3.8E-14	69.10%	0.041	Yes
SALL1	0.344	4.1E-14	68.10%	-0.080	Yes
ZKSCAN1	0.787	4.1E-14	66.10%	0.084	Yes
KLF15	0.594	4.7E-14	66.30%	0.023	No
SPATA18	0.378	4.7E-14	66.70%	-0.386	Yes
CNN3	0.441	4.9E-14	66.50%	0.054	Yes
PQLC1	0.651	5.6E-14	67.10%	-0.032	No
GOT2	0.675	6.3E-14	67.60%	-0.082	Yes
PAH	0.208	6.9E-14	68.30%	-0.096	Yes
NEDD4L	0.893	7.4E-14	67.20%	0.157	Yes
TOX3	0.190	7.6E-14	68.40%	0.032	No
SDC1	0.605	8.3E-14	66.10%	-0.004	No
HIC1	0.464	8.4E-14	65.70%	-0.041	No
TF	0.445	1.0E-13	68.50%	-0.089	Yes
MT1M	0.532	1.2E-13	65.70%	-0.889	Yes
GABRG3	0.245	1.4E-13	70.30%	0.304	No

RBMS3	0.269	1.4E-13	65.50%	-0.273	Yes
ALDOB	0.440	1.4E-13	68.10%	-0.192	Yes
B3GNT9	0.494	1.5E-13	66.30%	-0.022	No
EVL	-1.135	1.5E-13	65.40%	-0.021	No
FREM2	0.288	1.5E-13	68.50%	-1.240	Yes
NUP214	-1.368	1.8E-13	66.10%	0.080	Yes
MYO1B	0.372	1.9E-13	67.40%	-0.097	Yes
OPRD1	-0.596	1.9E-13	65.30%	0.038	No
ISL2	0.535	2.0E-13	67.00%	2.011	Yes

Notes: [a] Coefficient was based on the logistic regression adjusted for age and gender in the training set. [b] AUC was calculated for each gene in the training set. [c] The $\log_2$ (FC) for gene expression was calculated between tumors and normal tissues from TCGA. A 5% FDR (false discovery rate) was considered significant for the TCGA data. AUC: area under curve; Controls: patients with benign liver lesions and healthy individuals; FC: fold change; HCC: hepatocellular carcinoma; TCGA: The Cancer Genome Atlas.