

Figure S1. (A) Kaplan-Meier survival curve of the gene expression levels of each selected MMP in prediction model based on GSE53625 dataset. (B) ROC curves of each single MMP in the prediction model based on the GSE53625 dataset.

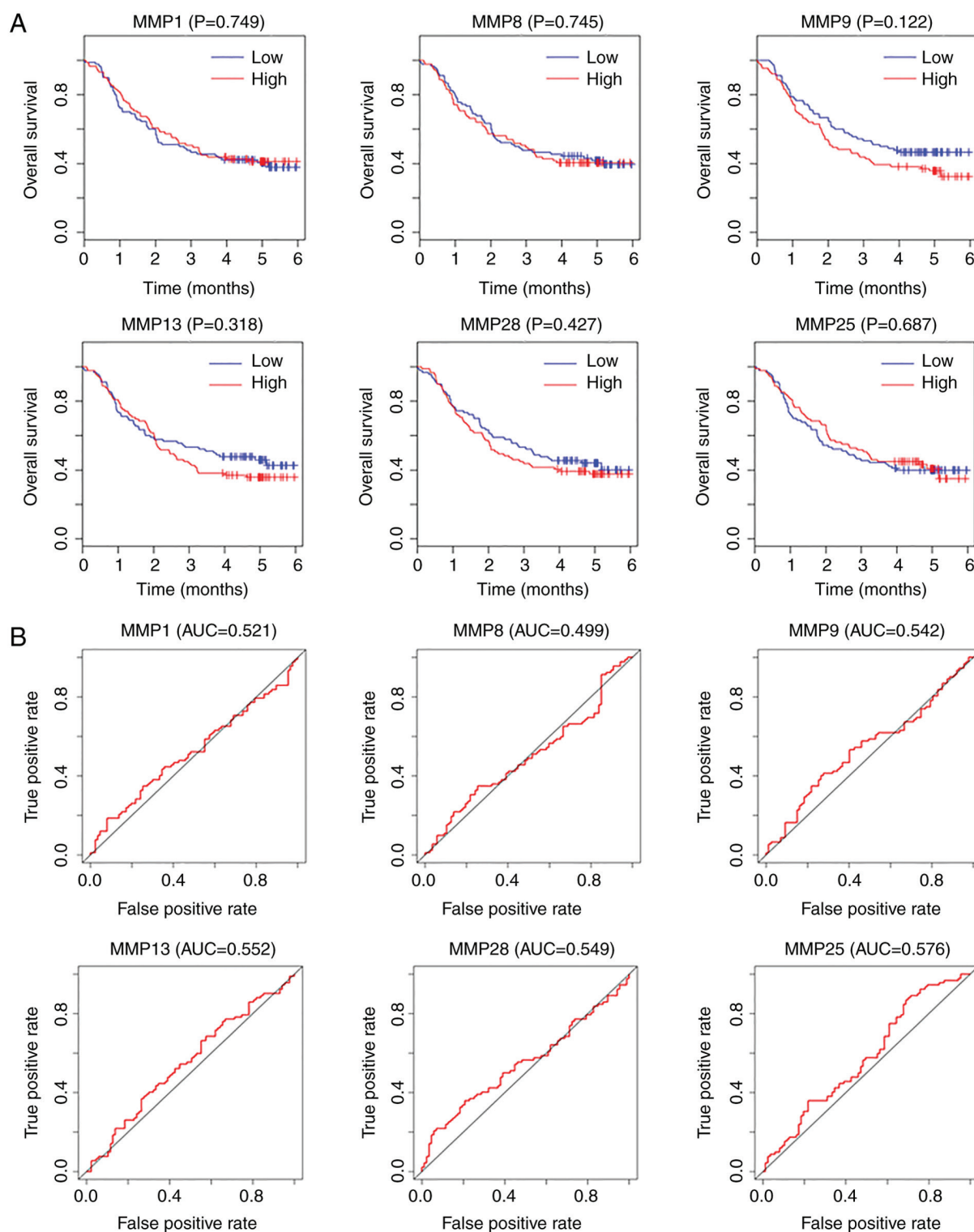


Figure S1. Continued. (C) Association diagram of the MMP family members based on the GSE53625 dataset. Red and green colors represent positive and negative correlation, respectively. (D and E) ROC curves of both the MMP model and the Tumor-Node-Metastasis stage model based on TCGA database. (F and G) Kaplan-Meier curves of the risk scores calculated using the prediction models based on TCGA database. MMP, matrix metalloproteinase; ROC, receiver operating characteristic; AUC, area under the curve; TCGA, The Cancer Genome Atlas.

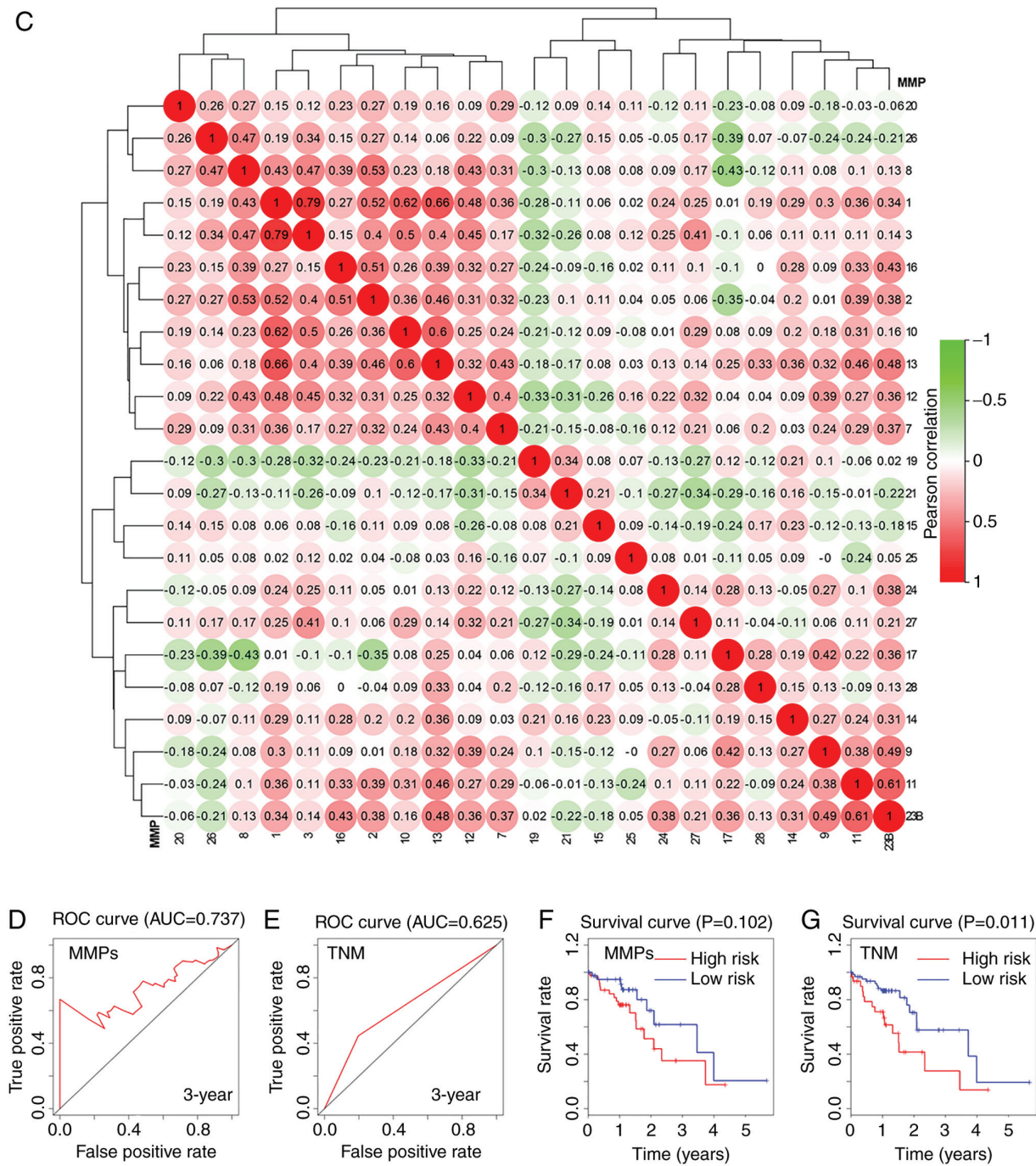


Table SI. Clinical information of the 179 patients with ESCC in the GSE53625 dataset.

Characteristic	No. of cases (%)
Age	
<60	91 (50.8%)
≥60	88 (49.2%)
Sex	
Male	146 (81.6%)
Female	33 (18.4%)
TNM stage	
I+II	87 (48.6%)
III	92 (51.4%)
T stage	
T1 + T2	39 (21.8%)
T3 + T4	140 (78.2%)
Tumor Grade	
G1 + G2	99 (55.3%)
G3	80 (44.7%)
N stage	
N0 + N1	145 (81.0%)
N2 + N3	34 (19.0%)
Tumor Location	
Upper + Middle	117 (65.4%)
Lower	62 (34.6%)
Alcohol Use	
No	73 (40.8%)
Yes	106 (59.2%)
Tobacco Use	
No	65 (36.3%)
Yes	114 (63.7%)
Pneumonia	
No	164 (91.6%)
Yes	15 (8.4%)

ESCC, esophageal squamous cell carcinoma; TNM stage, Tumor-Node-Metastasis stage; T stage, stage of tumor invasion; N stage, stage of regional lymph node invasion.

Table SII. Clinical information of 95 patients with ESCC in TCGA database.

Characteristic	No. of cases (%)
Age	
<60	56 (58.9%)
≥60	39 (41.1%)
Sex	
Male	80 (84.2%)
Female	15 (15.8%)
TNM stage	
I + II	63 (66.3%)
III + IV	31 (32.6%)
Missing	1 (1.1%)
T stage	
T1 + T2	40 (42.1%)
T3 + T4	54 (56.8%)
Missing	1 (1.1%)
N stage	
N0 + N1	84 (88.4%)
N2 + N3	9 (9.5%)
Missing	2 (2.1%)
Tumor grade	
G1 + G2	65 (68.4%)
G3	21 (22.1%)
Missing	9 (9.5%)
Tumor location	
Upper + Middle	50 (52.6%)
Lower	44 (46.3%)
Missing	1 (1.1%)
Alcohol use	
No	25 (26.3%)
Yes	68 (71.6%)
Missing	2 (2.1%)
Tobacco use	
No	44 (46.3%)
Yes	51 (53.7%)
Race	
Asian	45 (47.4%)
White + Other	47 (49.5%)
Missing	3 (3.2%)

ESCC, esophageal squamous cell carcinoma; TCGA, The Cancer Genome Atlas; TNM stage, Tumor-Node-Metastasis stage; T stage, stage of tumor invasion; N stage, stage of regional lymph node invasion.

Table SIII. Results of expression difference analysis of MMP family members in the GSE53625 dataset.

Gene	logFC	Average expression	P-value	Adjusted P-value
MMP1	5.49	11.48	0.000 ^c	0.000
MMP2	0.36	10.27	0.000 ^c	0.000
MMP3	5.13	8.51	0.000 ^c	0.000
MMP7	1.21	11.89	0.000 ^c	0.000
MMP8	1.32	4.36	0.000 ^c	0.000
MMP9	0.98	11.46	0.000 ^c	0.000
MMP10	4.18	9.32	0.000 ^c	0.000
MMP11	3.92	12.08	0.000 ^c	0.000
MMP12	4.15	10.37	0.000 ^c	0.000
MMP13	3.78	9.48	0.000 ^c	0.000
MMP14	1.01	12.04	0.000 ^c	0.000
MMP15	0.24	10.06	0.001 ^b	0.001
MMP16	0.82	7.28	0.000 ^c	0.000
MMP17	0.22	12.13	0.116	0.144
MMP19	0.10	11.64	0.013 ^a	0.019
MMP20	0.83	4.80	0.000 ^c	0.000
MMP21	-0.10	7.96	0.127	0.157
MMP23B	0.25	10.83	0.019 ^a	0.027
MMP24	-0.40	9.12	0.000 ^c	0.000
MMP25	0.55	8.66	0.000 ^c	0.000
MMP26	-0.06	3.69	0.246	0.288
MMP27	-0.95	5.36	0.000 ^c	0.000
MMP28	-0.39	8.56	0.000 ^c	0.000

^aP<0.05, ^bP<0.01, ^cP<0.001. FC, fold change; MMP, matrix metalloproteinase.

Table SIV. Results of expression difference analysis of MMP family members in the GSE23400 dataset.

Gene	logFC	Average expression	P-value	Adjusted P-value
MMP1	4.44	7.76	0.000 ^b	0.000
MMP2	0.88	10.03	0.000 ^b	0.000
MMP3	1.24	6.75	0.000 ^b	0.000
MMP7	0.48	7.30	0.010 ^a	0.018
MMP8	-0.02	5.37	0.410	0.475
MMP9	1.47	8.15	0.000 ^b	0.000
MMP10	2.19	6.82	0.000 ^b	0.000
MMP11	1.24	7.62	0.000 ^b	0.000
MMP12	2.24	8.07	0.000 ^b	0.000
MMP13	1.29	6.52	0.000 ^b	0.000
MMP14	0.59	8.45	0.000 ^b	0.000
MMP15	0.09	7.05	0.041 ^a	0.063
MMP16	-0.03	4.71	0.050	0.076
MMP17	-0.05	6.83	0.187	0.242
MMP19	0.03	7.79	0.367	0.432
MMP20	0.04	5.27	0.162	0.213
MMP24	-0.11	6.23	0.000 ^b	0.000
MMP25	-0.15	6.29	0.000 ^b	0.000
MMP26	-0.04	5.63	0.206	0.263
MMP27	-0.05	5.37	0.016 ^a	0.027
MMP28	-0.10	5.46	0.236	0.296

^aP<0.05, ^bP<0.001. FC, fold change; MMP, matrix metalloproteinase.

Table SV. Results of expression difference analysis of MMP family members in the GSE38129 dataset.

Gene	logFC	Average expression	P-value	Adjusted P-value
MMP1	4.84	6.66	0.000 ^c	0.000
MMP2	1.47	8.83	0.000 ^c	0.002
MMP3	2.35	5.28	0.000 ^c	0.000
MMP7	1.19	6.71	0.000 ^c	0.001
MMP8	-0.03	4.06	0.678	0.766
MMP9	1.64	6.64	0.000 ^c	0.000
MMP10	2.72	5.44	0.000 ^c	0.000
MMP11	1.67	6.44	0.000 ^c	0.000
MMP12	3.47	7.10	0.000 ^c	0.000
MMP13	2.37	5.01	0.000 ^c	0.000
MMP14	0.41	7.04	0.016 ^a	0.042
MMP15	0.42	6.25	0.002 ^b	0.006
MMP16	-0.04	3.89	0.274	0.393
MMP17	-0.04	5.71	0.468	0.587
MMP19	0.07	6.01	0.478	0.595
MMP20	0.21	3.46	0.194	0.303
MMP24	-0.04	5.31	0.373	0.496
MMP25	-0.14	5.36	0.074	0.144
MMP26	-0.03	3.91	0.780	0.849
MMP27	0.01	4.08	0.883	0.923
MMP28	-0.75	5.19	0.003 ^b	0.011

^aP<0.05, ^bP<0.01, ^cP<0.001. FC, fold change; MMP, matrix metalloproteinase.

Table SVI. Results of expression difference analysis of MMP family members in TCGA database.

Gene	logFC	logCPM	P-value	FDR
MMP1	3.41	7.43	0.000 ^c	0.000
MMP2	0.46	8.47	0.344	0.531
MMP3	4.57	4.67	0.000 ^c	0.000
MMP7	1.14	4.88	0.129	0.268
MMP8	2.39	-1.53	0.004 ^b	0.016
MMP9	3.87	5.78	0.000 ^c	0.000
MMP10	5.09	4.99	0.000 ^c	0.000
MMP11	6.09	7.24	0.000 ^c	0.000
MMP12	4.90	5.87	0.000 ^c	0.000
MMP13	7.18	6.03	0.000 ^c	0.000
MMP14	2.01	8.61	0.000 ^c	0.000
MMP15	-1.64	5.54	0.000 ^c	0.000
MMP16	0.22	2.30	0.701	0.834
MMP17	2.47	2.68	0.000 ^c	0.000
MMP19	-1.02	3.44	0.001 ^b	0.007
MMP20	2.86	-2.30	0.008 ^b	0.032
MMP21	-0.98	-3.48	0.013 ^a	0.046
MMP23B	0.18	-0.52	0.696	0.830
MMP24	-0.75	1.21	0.007 ^b	0.026
MMP25	0.91	1.72	0.011 ^a	0.039
MMP26	-0.78	-4.12	0.330	0.516
MMP27	-1.62	-2.97	0.002 ^b	0.010
MMP28	1.12	4.56	0.065	0.160

^aP<0.05, ^bP<0.01, ^cP<0.001. FC, fold change; CPM, counts per million; FDR, false discovery rate; MMP, matrix metalloproteinase.

Table SVII. Overlapping enriched gene sets based on Hallmark Gene Sets in GSE53625 dataset.

Name	Size	NES	NOM p-val	FDR q-val
HALLMARK_APICAL_JUNCTION	190	1.81	0.004 ^b	0.081
HALLMARK_COAGULATION	135	1.78	0.004 ^b	0.070
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	194	1.82	0.012 ^a	0.153
HALLMARK_ESTROGEN_RESPONSE_EARLY	195	1.56	0.017 ^a	0.118
HALLMARK_NOTCH_SIGNALING	31	1.68	0.020 ^a	0.122
HALLMARK_APICAL_SURFACE	43	1.60	0.022 ^a	0.127
HALLMARK_HYPOXIA	193	1.68	0.024 ^a	0.104
HALLMARK_HEME_METABOLISM	190	1.57	0.035 ^a	0.145
HALLMARK_ESTROGEN_RESPONSE_LATE	197	1.53	0.036 ^a	0.123
HALLMARK_UV_RESPONSE_DN	140	1.66	0.039 ^a	0.108

^aP<0.05; ^bP<0.01. NES, normalized enrichment score; NOM p-val, nominal P value; FDR, false discovery rate.

Table SVIII. Overlapping enriched gene sets according to KEGG Curated Gene Sets in the GSE53625 dataset.

Name	Size	NES	NOM p-val	FDR q-val
KEGG_PANCREATIC_CANCER	69	1.81	0.000 ^b	0.319
KEGG_FOCAL_ADHESION	195	1.80	0.000 ^b	0.178
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	51	1.69	0.010 ^a	0.118
KEGG_VEGF_SIGNALING_PATHWAY	70	1.55	0.012 ^a	0.158
KEGG_VIRAL_MYOCARDITIS	65	1.64	0.014 ^a	0.121
KEGG_PATHWAYS_IN_CANCER	319	1.47	0.014 ^a	0.178
KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	96	1.68	0.016 ^a	0.103
KEGG_CELL_ADHESION_MOLECULES_CAMS	128	1.60	0.016 ^a	0.125
KEGG_EPITHELIAL_CELL_SIGNALING_IN_				
HELICOBACTER_PYLORI_INFECTION	66	1.67	0.018 ^a	0.108
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	209	1.58	0.020 ^a	0.140
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	127	1.70	0.020 ^a	0.136
KEGG_LEISHMANIA_INFECTION	67	1.63	0.021 ^a	0.120
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	93	1.62	0.022 ^a	0.120
KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	68	1.70	0.023 ^a	0.153
KEGG_ALLOGRAFT_REJECTION	33	1.55	0.029 ^a	0.152
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	113	1.53	0.032 ^a	0.160
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	73	1.62	0.040 ^a	0.117
KEGG_SMALL_CELL_LUNG_CANCER	84	1.44	0.047 ^a	0.200
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	74	1.55	0.047 ^a	0.149

^aP<0.05, ^bP<0.001. NES, normalized enrichment score; NOM p-val, nominal P value; FDR, false discovery rate.

Table SIX. Overlapping enriched gene sets according to Oncogenic Signatures Gene Sets in the GSE53625 dataset.

Name	Size	NES	NOM p-val	FDR q-val
RB_DN.V1_DN	119	1.87	0.000 ^c	0.097
TBK1.DF_UP	265	1.83	0.000 ^c	0.082
PRC2_SUZ12_UP.V1_DN	173	1.63	0.000 ^c	0.106
GCNP_SHH_UP_LATE.V1_DN	174	1.63	0.004 ^b	0.099
TGFB_UP.V1_UP	179	1.68	0.004 ^b	0.080
ESC_J1_UP_EARLY.V1_UP	159	1.47	0.004 ^b	0.116
NOTCH_DN.V1_UP	171	1.57	0.006 ^b	0.093
KRAS.AMP.LUNG_UP.V1_DN	134	1.49	0.006 ^b	0.106
ATF2_S_UP.V1_DN	176	1.70	0.006 ^b	0.082
ATF2_UP.V1_DN	170	1.71	0.008 ^b	0.095
SNF5_DN.V1_DN	152	1.46	0.008 ^b	0.115
ATF2_S_UP.V1_UP	179	1.54	0.010 ^a	0.090
ESC_V6.5_UP_EARLY.V1_DN	165	1.65	0.010 ^a	0.092
IL15_UP.V1_UP	169	1.54	0.011 ^a	0.094
RB_P130_DN.V1_DN	130	1.75	0.012 ^a	0.148
BMI1_DN_MEL18_DN.V1_UP	136	1.70	0.012 ^a	0.076
IL21_UP.V1_UP	173	1.46	0.012 ^a	0.116
EIF4E_UP	89	1.62	0.012 ^a	0.095
NRL_DN.V1_UP	129	1.53	0.014 ^a	0.088
SRC_UP.V1_DN	154	1.53	0.014 ^a	0.089
E2F1_UP.V1_DN	179	1.73	0.015 ^a	0.113
CRX_DN.V1_UP	125	1.54	0.016 ^a	0.091
BMI1_DN.V1_UP	139	1.72	0.016 ^a	0.096
LTE2_UP.V1_UP	177	1.60	0.016 ^a	0.103
EIF4E_DN	94	1.67	0.016 ^a	0.081
PKCA_DN.V1_DN	152	1.44	0.017 ^a	0.113
CAHOY_OLIGODENDROCTIC	87	1.55	0.018 ^a	0.096
P53_DN.V1_UP	181	1.55	0.018 ^a	0.100
HOXA9_DN.V1_UP	177	1.74	0.019 ^a	0.119
PTEN_DN.V2_UP	131	1.54	0.020 ^a	0.093
CAHOY_ASTROGLIAL	93	1.59	0.021 ^a	0.105
KRAS.LUNG_UP.V1_DN	130	1.53	0.021 ^a	0.090
P53_DN.V1_DN	181	1.53	0.023 ^a	0.089
CRX_NRL_DN.V1_UP	129	1.42	0.024 ^a	0.121
MEL18_DN.V1_UP	133	1.70	0.028 ^a	0.091
ESC_V6.5_UP_LATE.V1_DN	169	1.55	0.028 ^a	0.096
ALK_DN.V1_UP	129	1.46	0.037 ^a	0.115

^aP<0.05, ^bP<0.01, ^cP<0.001; NES, normalized enrichment score; NOM p-val, nominal P value; FDR, false discovery rate.