

Supplementary Information

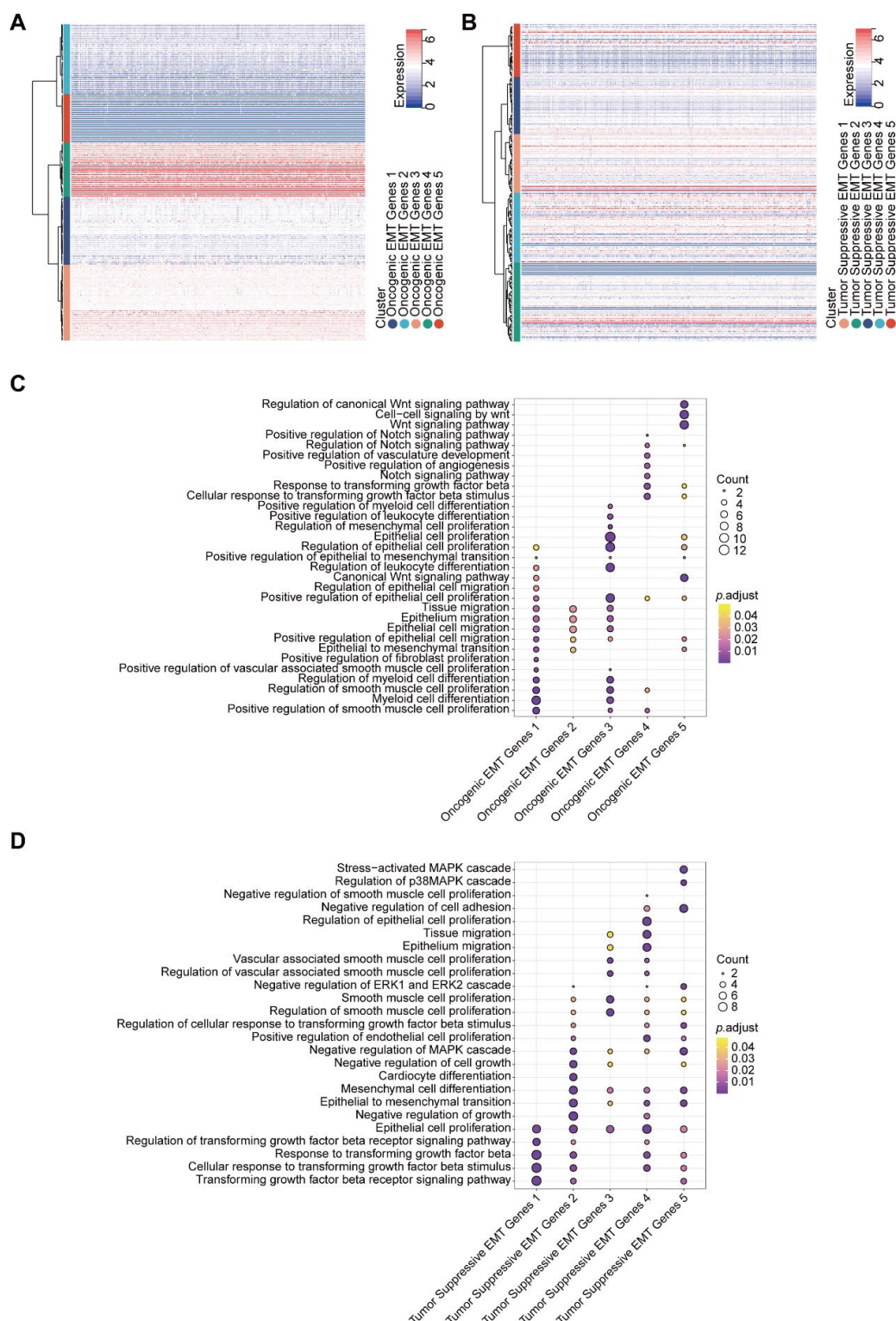


Figure S1. Complete hierarchical clustering of the epithelial-mesenchymal transition (EMT) genes from the dbEMT website. **A–B.** Heatmaps for the complete hierarchical clustering results of the oncogenic EMT genes (**A**) and the tumor suppressive EMT genes (**B**) obtained from dbEMT. **C–D.** Bubble plots of the enriched Gene Ontology (GO) pathways for the five oncogenic EMT gene sets (**C**) and five tumor-suppressive EMT gene sets (**D**).

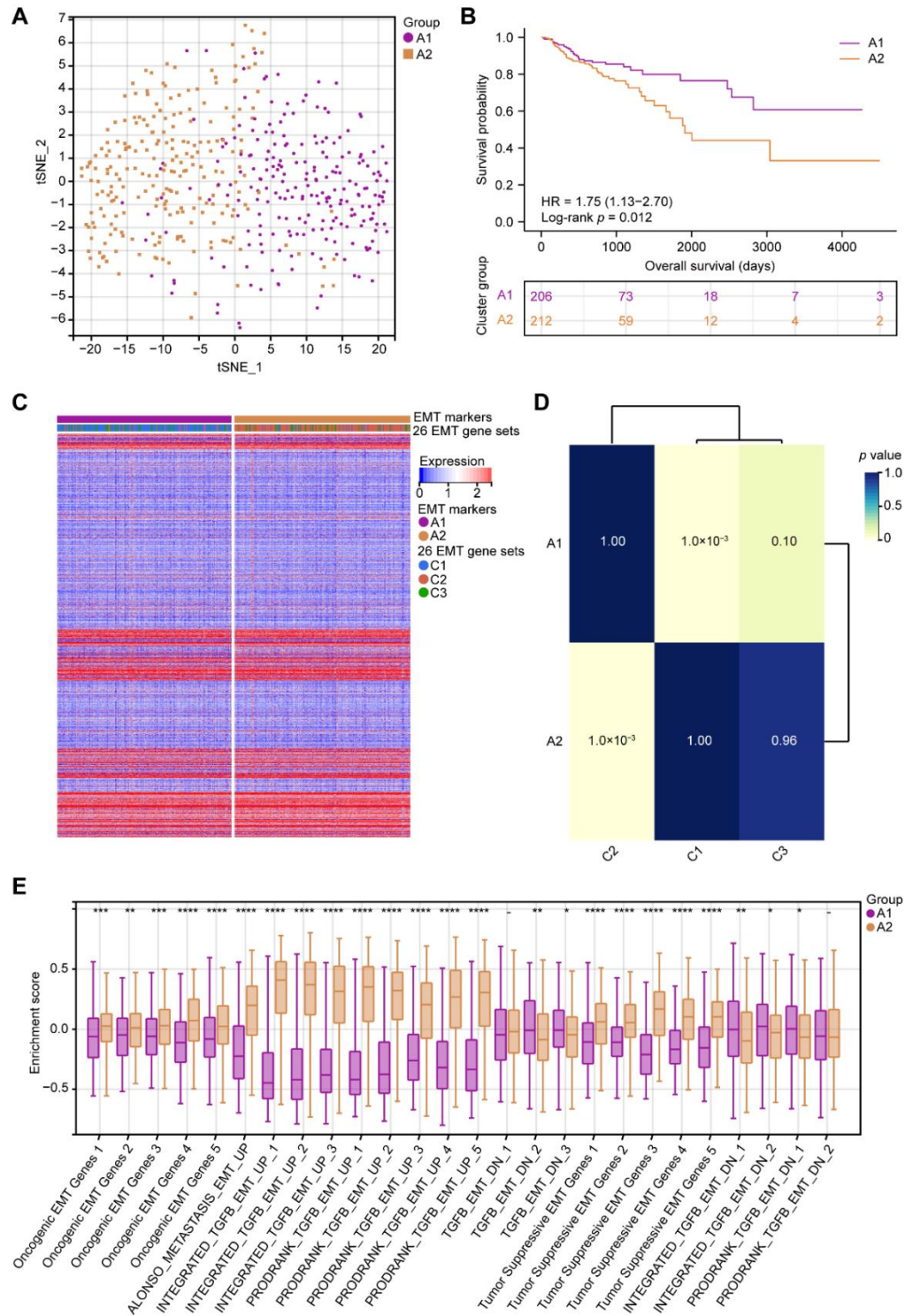


Figure S2. Validation of the three EMT clusters by an independent classification method. **A.** Principal component analysis (PCA) showing the differences between clusters A1 and A2. **B.** Kaplan–Meier overall survival (OS) curves for The Cancer Genome Atlas program (TCGA) colon adenocarcinoma (COAD) patients categorized as clusters A1 and A2. **C.** Heatmap showing the gene expression data in A1, A2, C1, C2, and C3 clusters. **D.** Submap showing the correlation significance (adjusted nominal p by Fisher’s test) between the different clusters. **E.** Box plot of GSVA enrichment scores for 26 EMT gene sets in clusters A1 and A2. Wilcoxon test; -, no significance, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

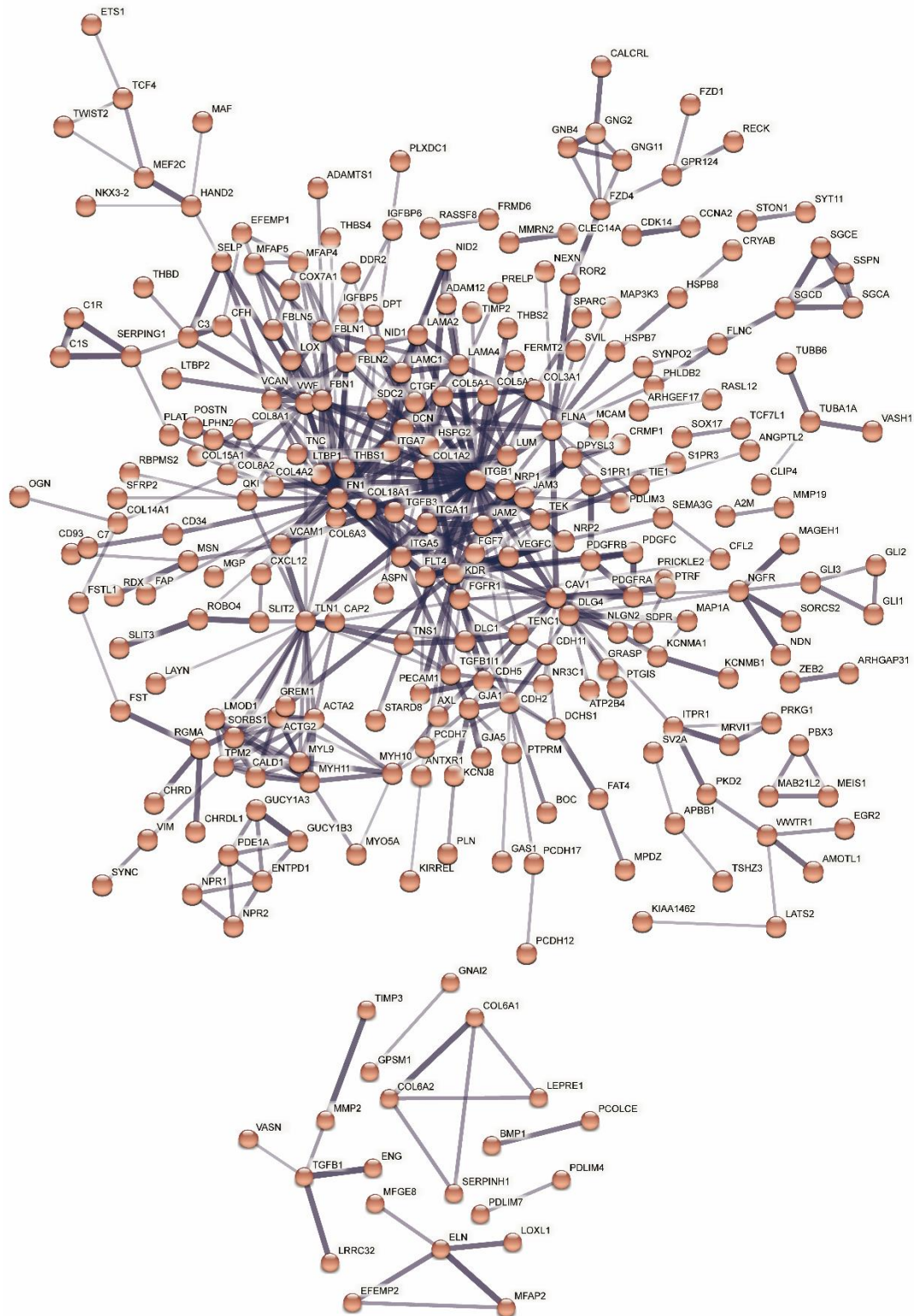


Figure S3. Protein-protein interaction (PPI) network showing the interactions between the genes from the lightcyan and darkolivegreen modules.

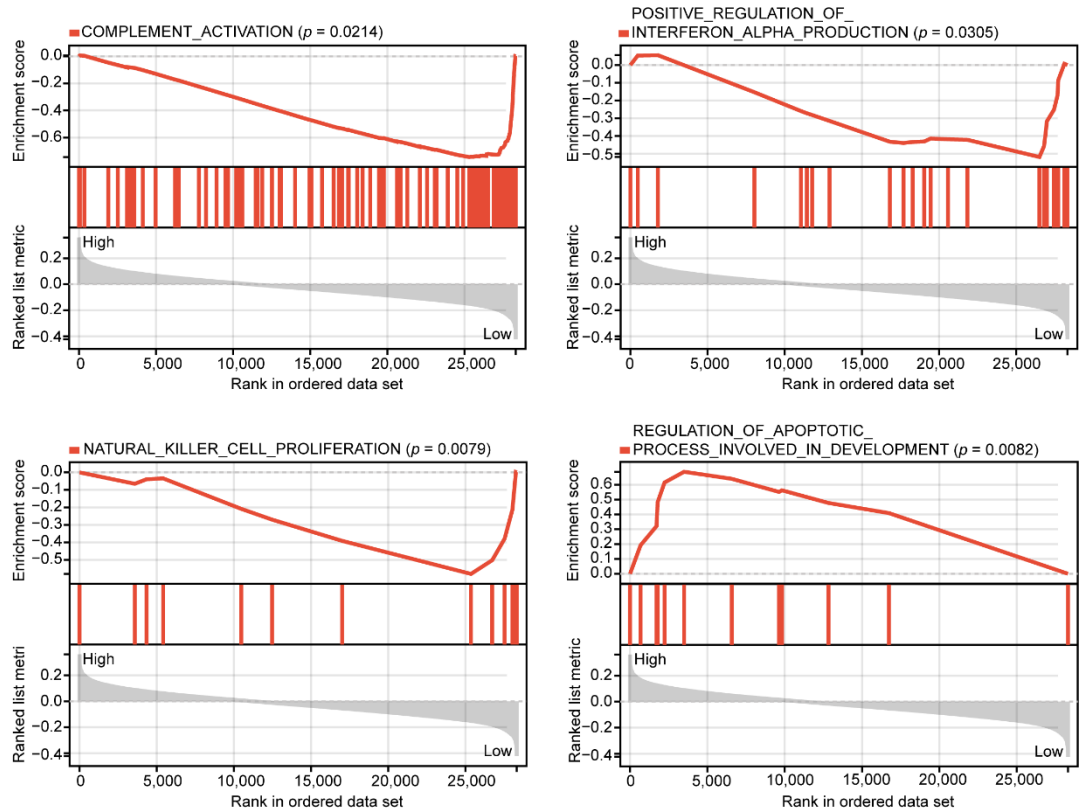


Figure S4. GSEA analysis of different GO biological process gene sets associated with the risk score.

Table S1. The ten EMT gene sets acquired from dbEMT database.

Gene set	Gene									
Oncogenic EMT Genes 1	<i>RAB22A</i>	<i>HOXA9</i>	<i>JUNB</i>	<i>RAF1</i>	<i>EPCAM</i>	<i>HRAS</i>	<i>WWTR1</i>	<i>CDK14</i>	<i>SOX4</i>	<i>SETDB1</i>
	<i>MYB</i>	<i>PIM2</i>	<i>JUN</i>	<i>KRAS</i>	<i>CEACAM6</i>	<i>MYD88</i>	<i>JAK2</i>	<i>FASN</i>	<i>TBX3</i>	<i>MIR221</i>
	<i>PDGFRB</i>	<i>ETS1</i>	<i>S100A4</i>	<i>GATA6</i>	<i>MIR10B</i>	<i>FOXQ1</i>	<i>PIK3CA</i>	<i>FZD2</i>	<i>TAZ</i>	<i>FOXM1</i>
	<i>AKT2</i>	<i>LCN2</i>	<i>PLAC8</i>	<i>FHL2</i>	<i>MYCN</i>	<i>CUL4A</i>	<i>NEAT1</i>			
Oncogenic EMT Genes 2	<i>MIR191</i>	<i>SNAI1</i>	<i>MIR93</i>	<i>MET</i>	<i>ETV4</i>	<i>BRF2</i>	<i>FGFR1</i>	<i>ZEB1-AS1</i>	<i>MALAT1</i>	<i>MIR485</i>
	<i>ELK1</i>	<i>TBX2</i>	<i>TXN</i>	<i>BRD4</i>	<i>PRKCI</i>	<i>SQSTM1</i>	<i>EIF5A2</i>	<i>ZNF217</i>	<i>ACTN4</i>	<i>MIR663A</i>
	<i>ERG</i>	<i>TYMS</i>	<i>CD24</i>	<i>IRS2</i>	<i>AKTIP</i>	<i>ITGA3</i>	<i>NANOG</i>	<i>PPM1D</i>	<i>PTP4A2</i>	<i>CAMK1D</i>
	<i>TRIM28</i>	<i>MIR520G</i>	<i>ROS1</i>	<i>CIP2A</i>	<i>UHRF1</i>	<i>ROCK1</i>	<i>TFCP2</i>			
Oncogenic EMT Genes 3	<i>CTNNB1</i>	<i>SREBF1</i>	<i>LEF1</i>	<i>GNA13</i>	<i>MDM2</i>	<i>YY1</i>	<i>PAX2</i>	<i>AXL</i>	<i>ID1</i>	<i>PRKCA</i>
	<i>BRAF</i>	<i>YWHAZ</i>	<i>NCOA3</i>	<i>PDGFB</i>	<i>EEF1D</i>	<i>FOSL1</i>	<i>HOTAIR</i>	<i>MITF</i>	<i>BCL2L1</i>	<i>SIX1</i>
	<i>ID2</i>	<i>ELAVL1</i>	<i>UCA1</i>	<i>YBX1</i>	<i>MIR135B</i>	<i>MUC1</i>	<i>PAK5</i>	<i>MYC</i>	<i>PTP4A3</i>	<i>KDM5B</i>
	<i>MIR92B</i>	<i>ERBB2</i>	<i>HOXD9</i>	<i>MIR373</i>	<i>RHOC</i>	<i>MIR96</i>	<i>LIN28A</i>	<i>CRK</i>	<i>MIR506</i>	<i>S100A8</i>
Oncogenic EMT Genes 4	<i>CREB1</i>	<i>GOLPH3</i>	<i>SKI</i>	<i>GLI2</i>	<i>AKT1</i>	<i>EIF3I</i>	<i>MTDH</i>	<i>ARHGEF2</i>	<i>GSK3A</i>	<i>PAK1</i>
	<i>MIR301A</i>	<i>ALK</i>	<i>HSPB1</i>	<i>CXCR4</i>	<i>YWHAG</i>	<i>MCL1</i>	<i>IGF1R</i>	<i>MDM4</i>	<i>MAP3K7</i>	<i>TWIST1</i>
	<i>USP22</i>	<i>EPS8</i>	<i>GLO1</i>	<i>AR</i>	<i>BIRC2</i>	<i>NANOGP8</i>	<i>SATB1</i>	<i>LETMD1</i>	<i>EIF4E</i>	<i>KIT</i>
	<i>WNT1</i>	<i>NOTCH4</i>	<i>MACC1</i>	<i>CRKL</i>	<i>UBE3C</i>	<i>BMI1</i>				
Oncogenic EMT Genes 5	<i>FGFR2</i>	<i>TLE1</i>	<i>AURKA</i>	<i>ETV1</i>	<i>RAC1</i>	<i>BCL2</i>	<i>MUC4</i>	<i>HSPA5</i>	<i>GAB2</i>	<i>ELK3</i>
	<i>CBLB</i>	<i>BIRC5</i>	<i>MIR454</i>	<i>ALDH1A1</i>	<i>KLF8</i>	<i>FRAT1</i>	<i>BMP7</i>	<i>CCND1</i>	<i>GMNN</i>	<i>HSPA4</i>
	<i>EGFR</i>	<i>SOX2</i>	<i>MIR128-1</i>	<i>EML4</i>	<i>MTOR</i>	<i>MLLT3</i>	<i>CRYAB</i>	<i>SRC</i>	<i>PLAGL2</i>	
Tumor Suppressive EMT Genes 1	<i>CREBBP</i>	<i>GPC3</i>	<i>NUMB</i>	<i>TP53BP2</i>	<i>AFAP1L2</i>	<i>LOX</i>	<i>HDAC3</i>	<i>VDR</i>	<i>NR1I2</i>	<i>CEACAM1</i>
	<i>CTNNBIP1</i>	<i>IGFBP3</i>	<i>IGFBP7</i>	<i>VHL</i>	<i>YPEL3</i>	<i>CTNND1</i>	<i>EPHB3</i>	<i>LIMA1</i>	<i>RUNX2</i>	<i>HPGD</i>
	<i>ABCG2</i>	<i>HIPK2</i>	<i>RNH1</i>	<i>NEDD4L</i>	<i>SIRT3</i>	<i>DKK3</i>	<i>SIAH1</i>	<i>GSN</i>	<i>KDM5A</i>	<i>INPP4B</i>
	<i>RB1</i>	<i>LEFTY1</i>	<i>PCGF2</i>	<i>SMAD2</i>	<i>ANXA1</i>	<i>DAB2</i>	<i>ARHGEF12</i>	<i>PHLDA2</i>	<i>ZYX</i>	
Tumor Suppressive EMT Genes 2	<i>ESR2</i>	<i>NOTCH2</i>	<i>SCUBE2</i>	<i>ISG15</i>	<i>BBC3</i>	<i>SMAD4</i>	<i>PRKAA1</i>	<i>TXNIP</i>	<i>STK11</i>	<i>AHR</i>
	<i>CLDN1</i>	<i>TCF4</i>	<i>CDH11</i>	<i>FBP1</i>	<i>BMP4</i>	<i>DKK1</i>	<i>AXIN2</i>	<i>ESRRB</i>	<i>TGFB3</i>	<i>EGLN3</i>
	<i>TGFB2</i>	<i>TSC1</i>	<i>HNF4A</i>	<i>STAT5A</i>	<i>SLC9A3R1</i>	<i>S100A2</i>	<i>CLU</i>	<i>BATF2</i>	<i>PEBP1</i>	<i>KDM6A</i>
	<i>NR4A1</i>	<i>GSK3B</i>	<i>EPAS1</i>	<i>GKN2</i>	<i>HRG</i>	<i>ROR2</i>	<i>CD44</i>	<i>PDCD4</i>	<i>IL17A</i>	<i>ITGA5</i>
Tumor Suppressive EMT Genes 3	<i>MTUS1</i>	<i>CXCR2</i>	<i>NR2C2</i>	<i>NUAK1</i>	<i>RACK1</i>	<i>STAT1</i>	<i>TNFAIP8L2</i>	<i>AGTR1</i>	<i>PKD1</i>	<i>NDRG2</i>
	<i>ESR1</i>	<i>ATM</i>	<i>FBXW7</i>	<i>GJB2</i>	<i>FOXC1</i>	<i>FOXO4</i>	<i>KDM3A</i>	<i>WWOX</i>	<i>CD82</i>	<i>ITGB3</i>
	<i>DAPK1</i>	<i>KL</i>	<i>RHOB</i>	<i>DNMT1</i>	<i>PCDH9</i>	<i>SUFU</i>	<i>WNT11</i>	<i>IGF1</i>	<i>TDGF1</i>	<i>PTPN6</i>
	<i>LATS1</i>	<i>FHL1</i>	<i>CSF2</i>	<i>FAS</i>	<i>DICER1</i>	<i>BRD7</i>	<i>DDR2</i>	<i>FBLN1</i>	<i>VSNL1</i>	
Tumor Suppressive EMT Genes 4	<i>UIMC1</i>	<i>EPB41L3</i>	<i>HOXB13</i>	<i>TWIST2</i>	<i>ING4</i>	<i>PARP1</i>	<i>SPOP</i>	<i>DNMT3B</i>	<i>BRMS1</i>	<i>RASAL2</i>
	<i>NF1</i>	<i>CXCL12</i>	<i>CCNDBP1</i>	<i>SCRIB</i>	<i>PROX1</i>	<i>HIF1A</i>	<i>ERRFI1</i>	<i>KLK6</i>	<i>ITGB1</i>	<i>FOXO3</i>
	<i>ESRP1</i>	<i>FLNA</i>	<i>FOXA2</i>	<i>LRIG1</i>	<i>KRT19</i>	<i>SOCS3</i>	<i>AXIN1</i>	<i>PRKAA2</i>	<i>ZFP36</i>	<i>MARVELD1</i>
	<i>TNFSF12</i>	<i>TSC2</i>	<i>THBD</i>	<i>TP53</i>	<i>CDH5</i>	<i>RNF8</i>	<i>SOD2</i>	<i>BRCA1</i>	<i>EGR1</i>	<i>PTEN</i>
Tumor Suppressive EMT Genes 5	<i>IL17RD</i>	<i>VEGFA</i>	<i>UHRF2</i>	<i>ANGPTL4</i>	<i>PAWR</i>	<i>BMP2</i>	<i>RNF111</i>	<i>DAB2IP</i>	<i>NFKB1</i>	<i>PIN1</i>
	<i>MAP3K4</i>	<i>TIMP3</i>	<i>SFRP2</i>	<i>GLS2</i>	<i>TP53BP1</i>	<i>TP53INP1</i>	<i>TRIM62</i>	<i>IFT88</i>	<i>TUSC7</i>	<i>CDKN2A</i>
	<i>MEG3</i>	<i>EPHB2</i>	<i>CDH13</i>	<i>ERF</i>	<i>AJAP1</i>	<i>EED</i>	<i>NOTCH3</i>	<i>NDRG1</i>	<i>IRF8</i>	<i>CXCL14</i>
	<i>CFTR</i>	<i>HIC1</i>								

Table S2. The sixteen EMT gene sets acquired from GSEA database.

Gene set	Gene									
ALONSO_	<i>ANLN</i>	<i>APLP2</i>	<i>CD63</i>	<i>CDH2</i>	<i>CLIC4</i>	<i>C7SB</i>	<i>CX3CR1</i>	<i>DSG2</i>	<i>EDNRB</i>	<i>EMP1</i>
METASTASIS_EMT_UP	<i>ENC1</i>	<i>FGG</i>	<i>FZD1</i>	<i>HMMR</i>	<i>ITGAV</i>	<i>L1CAM</i>	<i>LUM</i>	<i>MFAP1</i>	<i>MMP2</i>	<i>PFN1</i>
	<i>PRKCA</i>	<i>RAB1A</i>	<i>RAN</i>	<i>RRAGA</i>	<i>SDCBP</i>	<i>SELENOP</i>	<i>SERPINA3</i>	<i>SMARCA1</i>	<i>SPA17</i>	<i>SPARC</i>
	<i>TUBA1A</i>	<i>TUBA3C</i>	<i>TUBA4A</i>	<i>TUBB3</i>	<i>VCAN</i>					
INTEGRATED_	<i>ACKR3</i>	<i>ADAM19</i>	<i>ADAMTS6</i>	<i>ALOX5AP</i>	<i>ANGPTL4</i>	<i>AP1S2</i>	<i>ARHGEF40</i>	<i>BMPR2</i>	<i>CALD1</i>	<i>CCN2</i>
TGFB_EMT_UP_1	<i>CDH11</i>	<i>CDH2</i>	<i>CDK14</i>	<i>CHRNA9</i>	<i>COL1A1</i>	<i>COL3A1</i>	<i>COL4A1</i>	<i>COL4A2</i>	<i>COL5A1</i>	<i>COL5A2</i>
	<i>COL6A3</i>	<i>COL7A1</i>	<i>CRLF1</i>	<i>DACT1</i>	<i>DHRS2</i>	<i>DIXDC1</i>	<i>DLC1</i>	<i>DOCK4</i>	<i>DSE</i>	<i>EPHB2</i>
	<i>FBN1</i>	<i>FERMT2</i>	<i>FHOD3</i>	<i>FN1</i>	<i>FOXD1</i>	<i>FSTL3</i>	<i>GADD45B</i>	<i>GAL</i>	<i>GALNT10</i>	
INTEGRATED_	<i>GASK1B</i>	<i>GFPT2</i>	<i>GLIPR1</i>	<i>GREM1</i>	<i>HMOX1</i>	<i>HS3ST3A1</i>	<i>HS3ST3B1</i>	<i>HSF2BP</i>	<i>HTRA1</i>	<i>IGFBP5</i>
TGFB_EMT_UP_2	<i>IGFBP7</i>	<i>IL11</i>	<i>INHBA</i>	<i>ITGA5</i>	<i>ITGB3</i>	<i>JUN</i>	<i>JUNB</i>	<i>KCNMA1</i>	<i>LMCD1</i>	<i>LOX</i>
	<i>LUM</i>	<i>MAF</i>	<i>MFAP2</i>	<i>MMP1</i>	<i>MMP10</i>	<i>MMP2</i>	<i>MN1</i>	<i>MRC2</i>	<i>MYL9</i>	<i>NCF2</i>
	<i>NEDD9</i>	<i>NKX3-1</i>	<i>NREP</i>	<i>NT5E</i>	<i>NUAK1</i>	<i>PDGFC</i>	<i>PDLIM7</i>	<i>PID1</i>	<i>PLEK2</i>	
INTEGRATED_	<i>PMEPA1</i>	<i>PODXL</i>	<i>POSTN</i>	<i>PTHLH</i>	<i>PTPRK</i>	<i>RGS4</i>	<i>SACS</i>	<i>SCG2</i>	<i>SCG5</i>	<i>SERPINE1</i>
TGFB_EMT_UP_3	<i>SERPINE2</i>	<i>SKIL</i>	<i>SLC22A4</i>	<i>SLC26A2</i>	<i>SLN</i>	<i>SMAD7</i>	<i>SNAI2</i>	<i>SPARC</i>	<i>SPHK1</i>	<i>SPOCK1</i>
	<i>SRPX</i>	<i>STC1</i>	<i>TAGLN</i>	<i>TCF4</i>	<i>TGFB1</i>	<i>TGFB11</i>	<i>TGFB1</i>	<i>TGM2</i>	<i>THBS1</i>	<i>TIMP2</i>
	<i>TNFAIP6</i>	<i>TNS1</i>	<i>TP53I3</i>	<i>TPM1</i>	<i>TUFT1</i>	<i>VCAN</i>	<i>WNT5A</i>	<i>WNT5B</i>	<i>XYLT1</i>	<i>ZNF365</i>
PRODRANK_	<i>ABCA1</i>	<i>ACKR3</i>	<i>ACTN1</i>	<i>ADAM12</i>	<i>ADAM19</i>	<i>ADAMTS6</i>	<i>AKT3</i>	<i>ALOX5AP</i>	<i>AMIGO2</i>	<i>ANGPTL4</i>
TGFB_EMT_UP_1	<i>ANKLE2</i>	<i>AP1S2</i>	<i>APBB2</i>	<i>ARFGAP1</i>	<i>ARHGEF40</i>	<i>ARNTL</i>	<i>BHLHE40</i>	<i>BMP1</i>	<i>BMP2</i>	<i>BMPR2</i>
	<i>BPGM</i>	<i>C3orf52</i>	<i>CALD1</i>	<i>CCN2</i>	<i>CD59</i>	<i>CDH11</i>	<i>CDH2</i>	<i>CDK14</i>	<i>CHRNA9</i>	<i>CHST11</i>
	<i>COL1A1</i>	<i>COL4A1</i>	<i>COL4A2</i>	<i>COL5A1</i>	<i>COL7A1</i>	<i>CRLF1</i>	<i>CYTH1</i>			
PRODRANK_	<i>SNAI2</i>	<i>SPARC</i>	<i>SPDL1</i>	<i>SPHK1</i>	<i>SPOCK1</i>	<i>SRPX</i>	<i>SRRD</i>	<i>STC1</i>	<i>TAGLN</i>	<i>TAGLN2</i>
TGFB_EMT_UP_2	<i>TBX3</i>	<i>TCF4</i>	<i>TFPI2</i>	<i>TGFB1</i>	<i>TGFB11</i>	<i>TGFB1</i>	<i>TGM2</i>	<i>THBS1</i>	<i>TIMP2</i>	<i>TMCC1</i>
	<i>TNFAIP6</i>	<i>TNS1</i>	<i>TP53I3</i>	<i>TPM1</i>	<i>TPM4</i>	<i>TPST1</i>	<i>TPST2</i>	<i>TUBA1A</i>	<i>TUBA4A</i>	<i>TUFT1</i>
	<i>VCAN</i>	<i>VEGFC</i>	<i>VIM</i>	<i>WNT5A</i>	<i>WNT5B</i>	<i>XYLT1</i>	<i>ZNF365</i>			
PRODRANK_	<i>NT5E</i>	<i>NUAK1</i>	<i>PALLD</i>	<i>PDGFA</i>	<i>PDGFC</i>	<i>PDLIM7</i>	<i>PEA15</i>	<i>PID1</i>	<i>PIK3CD</i>	<i>PLAUR</i>
TGFB_EMT_UP_3	<i>PLEK2</i>	<i>PMEPA1</i>	<i>PODXL</i>	<i>POSTN</i>	<i>PRR5L</i>	<i>PSMD2</i>	<i>PTHLH</i>	<i>PTPN21</i>	<i>PTPRK</i>	<i>PXDC1</i>
	<i>RALA</i>	<i>RFTN1</i>	<i>RGS4</i>	<i>RUNX2</i>	<i>SACS</i>	<i>SCG5</i>	<i>SEMA3C</i>	<i>SERPINE1</i>	<i>SERPINE2</i>	<i>SIK1</i>
	<i>SKIL</i>	<i>SLC22A4</i>	<i>SLC26A2</i>	<i>SLCO2A1</i>	<i>SLN</i>	<i>SMAD7</i>	<i>SMURF2</i>			
PRODRANK_	<i>INHBA</i>	<i>INPP4B</i>	<i>ITGA5</i>	<i>ITGB3</i>	<i>JAG1</i>	<i>JARID2</i>	<i>JUN</i>	<i>JUNB</i>	<i>KCNJ15</i>	<i>KCNMA1</i>
TGFB_EMT_UP_4	<i>KDELR3</i>	<i>KLF7</i>	<i>LAMC2</i>	<i>LARP6</i>	<i>LBH</i>	<i>LMCD1</i>	<i>LOX</i>	<i>MAF</i>	<i>MAGED2</i>	<i>MAP1LC3B</i>
	<i>MATN3</i>	<i>MBOAT2</i>	<i>MFAP2</i>	<i>MICAL2</i>	<i>MMP1</i>	<i>MMP10</i>	<i>MMP2</i>	<i>MMP9</i>	<i>MN1</i>	<i>MRC2</i>
	<i>MYL9</i>	<i>MYO10</i>	<i>NCF2</i>	<i>NEDD9</i>	<i>NKX3-1</i>	<i>NREP</i>	<i>NRIP3</i>			
PRODRANK_	<i>DAAM1</i>	<i>DACT1</i>	<i>DHRS2</i>	<i>DIXDC1</i>	<i>DLC1</i>	<i>DOCK4</i>	<i>DSE</i>	<i>DUSP10</i>	<i>ELK3</i>	<i>EML1</i>
TGFB_EMT_UP_5	<i>EPHB2</i>	<i>ETS2</i>	<i>FAM114A1</i>	<i>FBN1</i>	<i>FERMT2</i>	<i>FHOD3</i>	<i>FN1</i>	<i>FOXD1</i>	<i>FSTL3</i>	<i>GADD45B</i>
	<i>GAL</i>	<i>GALNT10</i>	<i>GASK1B</i>	<i>GFPT2</i>	<i>GLIPR1</i>	<i>GNG11</i>	<i>GRB10</i>	<i>GREM1</i>	<i>HMGA2</i>	<i>HMOX1</i>
	<i>HRH1</i>	<i>HS3ST3A1</i>	<i>HSF2BP</i>	<i>HTRA1</i>	<i>IGFBP5</i>	<i>IGFBP7</i>	<i>IL11</i>			
TGFB_EMT_DN_1	<i>ABLIM1</i>	<i>ADORA2B</i>	<i>AGR2</i>	<i>ALDH1A3</i>	<i>ALDH3A2</i>	<i>ALDH5A1</i>	<i>ANK3</i>	<i>AQP3</i>	<i>AREG</i>	<i>ARHGAP29</i>
	<i>ATP8B1</i>	<i>BIRC3</i>	<i>C1orf115</i>	<i>C1orf116</i>	<i>CAVIN2</i>	<i>CD9</i>	<i>CDH1</i>	<i>CEACAM6</i>	<i>CEBPD</i>	<i>CFB</i>

