

SUPPLEMENTARY MATERIAL

Figure S1: Significant DEGs of the combined microarray datasets of TGF- β -induced EMT in NSCLC cell lines

Figure S2: Protein-protein interaction network of DEGs in EMT of NSCLC cells

Figure S3: Prognostic value of genes associated with EMT in NSCLC – Survival until first progression (FP)

Figure S4: Prognostic value of genes associated with EMT in NSCLC – Post (first) progression survival (PPS)

Table S1: List of the 20 most highly ranked upregulated genes in EMT of NSCLC cells with their z-scores, listed in order of their FDR values

Table S2: List of the 20 most highly ranked downregulated genes in EMT of NSCLC cells with their z-scores, listed in order of their FDR values

Table S3: Genes with their z-scores from the PPI network-based clustered modules in EMT of NSCLC cells, listed in order of their FDR values

Table S4: List of datasets sourced for patient survival analysis

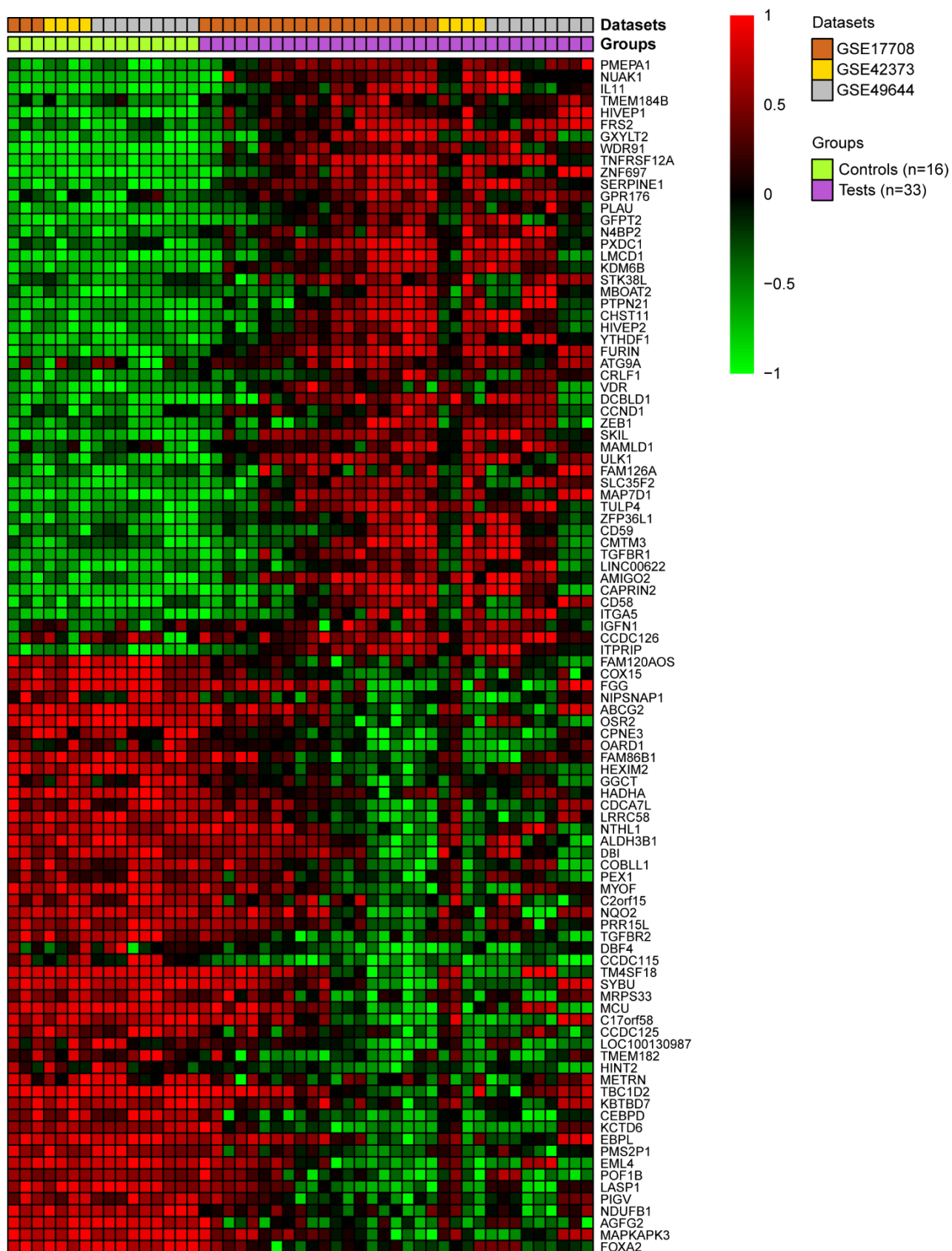


Figure S1. Significant DEGs of the combined microarray datasets of TGF- β -induced EMT in NSCLC cell lines. Heatmap of the top 100 DEGs from the combined microarray datasets listed according to their ranking based on their z-scores.

Table S1. List of the 20 most highly ranked upregulated genes in EMT of NSCLC cells with their z-scores, listed in order of their FDR values.

Gene symbol	Gene name	FDR*	Z-score
<i>PMEPA1</i>	prostate transmembrane protein, androgen induced 1	6.20×10^{-5}	5.888
<i>NUAK1</i>	NUAK family kinase 1	1.08×10^{-4}	5.679
<i>FRS2</i>	fibroblast growth factor receptor substrate 2	1.69×10^{-4}	5.332
<i>GXYLT2</i>	glucoside xylosyltransferase 2	1.69×10^{-4}	5.323
<i>HIVEP1</i>	human immunodeficiency virus type I enhancer binding protein 1	1.69×10^{-4}	5.372
<i>IL11</i>	interleukin 11	1.69×10^{-4}	5.402
<i>WDR91</i>	WD repeat domain 91	1.69×10^{-4}	5.320
<i>ZNF697</i>	zinc finger protein 697	3.69×10^{-4}	5.153
<i>GPR176</i>	G protein-coupled receptor 176	4.89×10^{-4}	5.041
<i>GFPT2</i>	glutamine-fructose-6-phosphate transaminase 2	5.99×10^{-4}	4.961
<i>N4BP2</i>	NEDD4 binding protein 2	5.99×10^{-4}	4.956
<i>PLAU</i>	plasminogen activator, urokinase	5.99×10^{-4}	4.966
<i>PXDC1</i>	PX domain containing 1	7.12×10^{-4}	4.913
<i>LMCD1</i>	LIM and cysteine rich domains 1	1.04×10^{-3}	4.810
<i>KDM6B</i>	lysine demethylase 6B	1.07×10^{-3}	4.795
<i>STK38L</i>	serine/threonine kinase 38 like	1.10×10^{-3}	4.774
<i>PTPN21</i>	protein tyrosine phosphatase, non-receptor type 21	1.23×10^{-3}	4.736
<i>HIVEP2</i>	human immunodeficiency virus type I enhancer binding protein 2	1.60×10^{-3}	4.669
<i>YTHDF1</i>	YTH N6-methyladenosine RNA binding protein 1	1.82×10^{-3}	4.629
<i>ATG9A</i>	autophagy related 9A	2.29×10^{-3}	4.539

* $p < 0.05$ corrected by BH FDR

Table S2. List of the 20 most highly ranked downregulated genes in EMT of NSCLC cells with their z-scores, listed in order of their FDR values.

Gene symbol	Gene name	FDR*	Z-score
<i>FOXA2</i>	forkhead box A2	1.69 x 10 ⁻⁴	-5.316
<i>AGFG2</i>	ArfGAP with FG repeats 2	4.37 x 10 ⁻⁴	-5.076
<i>MAPKAPK3</i>	mitogen-activated protein kinase-activated protein kinase 3	4.37 x 10 ⁻⁴	-5.079
<i>NDUFB1</i>	NADH:ubiquinone oxidoreductase subunit B1	5.54 x 10 ⁻⁴	-5.005
<i>PIGV</i>	phosphatidylinositol glycan anchor biosynthesis class V	7.41 x 10 ⁻⁴	-4.895
<i>LASP1</i>	LIM and SH3 protein 1	7.58 x 10 ⁻⁴	-4.882
<i>EML4</i>	echinoderm microtubule associated protein like 4	1.78 x 10 ⁻³	-4.640
<i>EBPL</i>	emopamil binding protein like	2.00 x 10 ⁻³	-4.596
<i>PMS2P1</i>	PMS1 homolog 2, mismatch repair system component pseudogene 1	2.00 x 10 ⁻³	-4.600
<i>CEBPD</i>	CCAAT enhancer binding protein delta	2.23 x 10 ⁻³	-4.556
<i>KCTD6</i>	potassium channel tetramerization domain containing 6	2.23 x 10 ⁻³	-4.557
<i>KBTBD7</i>	kelch repeat and BTB domain containing 7	2.29 x 10 ⁻³	-4.541
<i>TBC1D2</i>	TBC1 domain family member 2	2.55 x 10 ⁻³	-4.511
<i>METRNL</i>	meteorin, glial cell differentiation regulator	2.59 x 10 ⁻³	-4.498
<i>HINT2</i>	histidine triad nucleotide binding protein 2	2.75 x 10 ⁻³	-4.470
<i>TMEM182</i>	transmembrane protein 182	2.98 x 10 ⁻³	-4.443
<i>LOC100130987</i>	uncharacterized LOC100130987	3.11 x 10 ⁻³	-4.430
<i>CCDC125</i>	coiled-coil domain containing 125	3.22 x 10 ⁻³	-4.418
<i>C17orf58</i>	chromosome 17 open reading frame 58	3.38 x 10 ⁻³	-4.403
<i>MCU</i>	mitochondrial calcium uniporter	3.53 x 10 ⁻³	-4.381

* $p < 0.05$ corrected by BH FDR

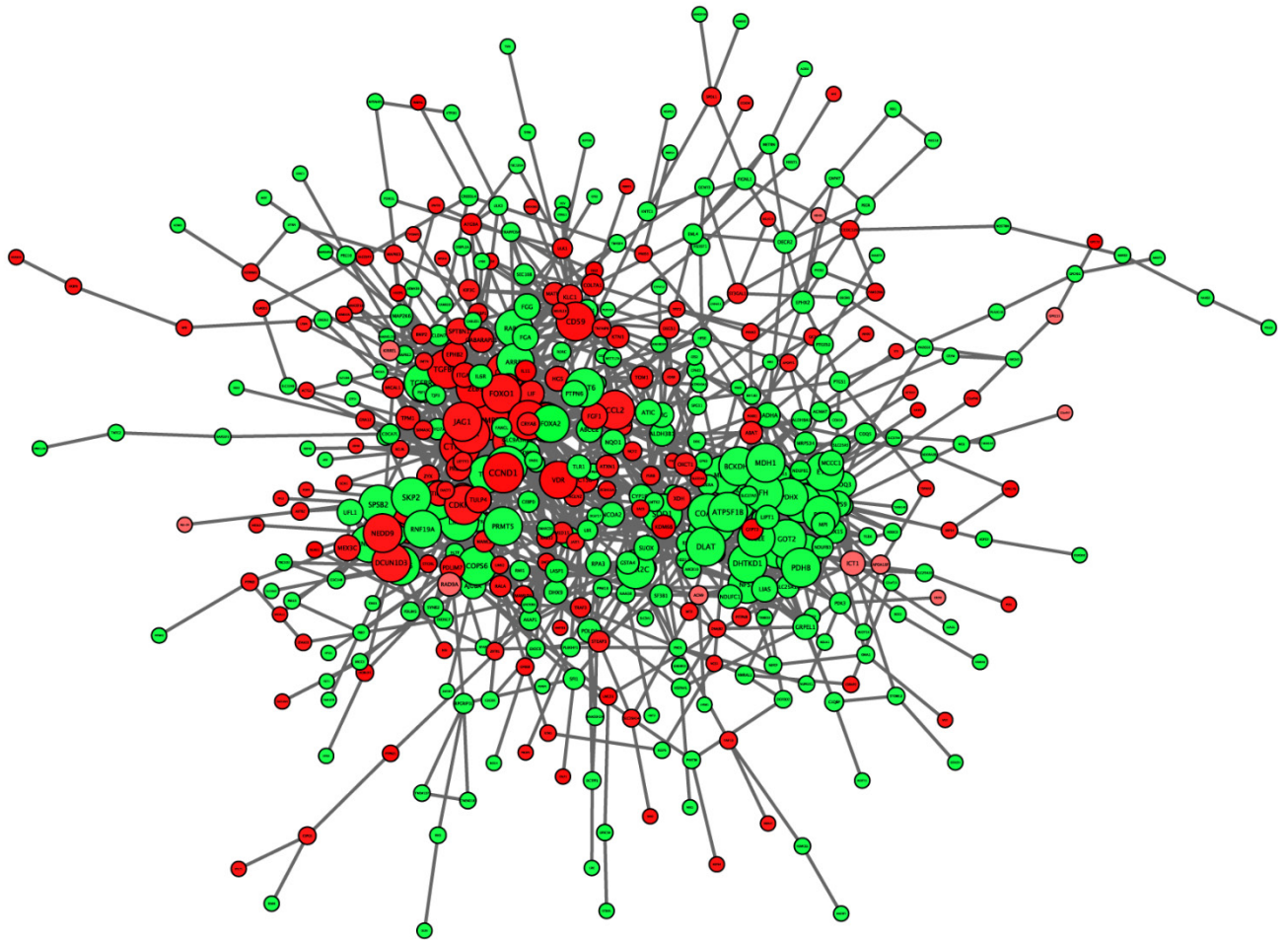


Figure S2. Protein-protein interaction network of DEGs in EMT of NSCLC cells. Network construction of proteins encoded by the total DEGs was carried out. All nodes with a combined interaction score > 0.4 are displayed, whereas isolated nodes are ignored. Larger nodes have a central regulatory role with a centrality degree ≥ 11.0 . Red indicates upregulated and green shows downregulated node genes.

Table S3. Genes with their z-scores from the PPI network-based clustered modules in EMT of NSCLC cells, listed in order of their FDR values.

Gene symbol	Gene name	FDR*	Z-score
Module 1			
<i>KCTD6</i>	potassium channel tetramerization domain containing 6	2.23×10^{-3}	-4.557
<i>KBTBD7</i>	kelch repeat and BTB domain containing 7	2.29×10^{-3}	-4.541
<i>TULP4</i>	tubby like protein 4	4.88×10^{-3}	4.255
<i>KLHL25</i>	kelch like family member 25	6.58×10^{-3}	4.024
<i>FBXO32</i>	F-box protein 32	8.94×10^{-3}	3.911
<i>COPS6</i>	COP9 signalosome subunit 6	9.32×10^{-3}	-3.875
<i>LMO7</i>	LIM domain 7	1.19×10^{-2}	-3.758
<i>SKP2</i>	S-phase kinase associated protein 2	1.53×10^{-2}	-3.658
<i>SPSB2</i>	splA/ryanodine receptor domain and SOCS box containing 2	2.07×10^{-2}	-3.495
<i>MEX3C</i>	mex-3 RNA binding family member C	2.73×10^{-2}	3.347
<i>RNF115</i>	ring finger protein 115	2.96×10^{-2}	-3.306
<i>UFL1</i>	UFM1 specific ligase 1	3.78×10^{-2}	-3.187
<i>RNF19A</i>	ring finger protein 19A, RBR E3 ubiquitin protein ligase	4.77×10^{-2}	-3.075
<i>DCUNID3</i>	defective in cullin neddylation 1 domain containing 3	4.81×10^{-2}	3.071
Module 2			
<i>FOXA2</i>	forkhead box A2	1.69×10^{-4}	-5.316
<i>PLAU</i>	plasminogen activator, urokinase	5.99×10^{-4}	4.966
<i>CCND1</i>	cyclin D1	2.82×10^{-3}	4.460
<i>TGFBRI</i>	transforming growth factor beta receptor 1	5.16×10^{-3}	4.205
<i>FGG</i>	fibrinogen gamma chain	5.26×10^{-3}	-4.162
<i>MMP2</i>	matrix metalloproteinase 2	5.60×10^{-3}	4.108
<i>DHTKD1</i>	dehydrogenase E1 and transketolase domain containing 1	5.60×10^{-3}	-4.117
<i>CTNNB1</i>	catenin beta 1	5.69×10^{-3}	4.098
<i>KTN1</i>	kinectin 1	6.45×10^{-3}	4.040
<i>BMP2</i>	bone morphogenetic protein 2	1.11×10^{-2}	3.790
<i>FGA</i>	fibrinogen alpha chain	1.24×10^{-2}	-3.736
<i>TIMP3</i>	TIMP metalloproteinase inhibitor 3	2.06×10^{-2}	3.500
<i>GSTA4</i>	glutathione S-transferase alpha 4	2.34×10^{-2}	-3.422
<i>CDH2</i>	cadherin 2	2.45×10^{-2}	3.395
<i>SOD1</i>	superoxide dismutase 1	2.45×10^{-2}	-3.392
<i>MATN3</i>	matrilin 3	2.96×10^{-2}	3.308
<i>GOLM1</i>	Golgi membrane protein 1	2.96×10^{-2}	-3.308
<i>PDHX</i>	pyruvate dehydrogenase complex component X	3.20×10^{-2}	-3.269
<i>CDH1</i>	cadherin 1	3.20×10^{-2}	-3.267
<i>JAG1</i>	jagged 1	3.57×10^{-2}	3.214
<i>CDKN1A</i>	cyclin dependent kinase inhibitor 1A	3.78×10^{-2}	3.187
<i>NQO1</i>	NAD(P)H quinone dehydrogenase 1	3.91×10^{-2}	-3.170
<i>GPI</i>	glucose-6-phosphate isomerase	3.99×10^{-2}	-3.160
<i>CCL2</i>	C-C motif chemokine ligand 2	4.01×10^{-2}	3.158
<i>GCLC</i>	glutamate-cysteine ligase catalytic subunit	4.04×10^{-2}	-3.154
<i>MDH1</i>	malate dehydrogenase 1	4.63×10^{-2}	-3.087
Module 3			

<i>TXNRD1</i>	thioredoxin reductase 1	1.63×10^{-2}	-3.617
<i>PDHB</i>	pyruvate dehydrogenase E1 beta subunit	1.72×10^{-2}	-3.589
<i>DLAT</i>	dihydrolipoamide S-acetyltransferase	2.20×10^{-2}	-3.461
<i>FH</i>	fumarate hydratase	2.89×10^{-2}	-3.322
<i>GSR</i>	glutathione-disulfide reductase	3.84×10^{-2}	-3.177
Module 4			
<i>MRPS33</i>	mitochondrial ribosomal protein S33	3.72×10^{-3}	-4.362
<i>MRPL24</i>	mitochondrial ribosomal protein L24	7.27×10^{-3}	-3.980
<i>MRPS34</i>	mitochondrial ribosomal protein S34	1.72×10^{-2}	-3.591
<i>MRPL58</i>	mitochondrial ribosomal protein L58	3.61×10^{-2}	-3.209
<i>MRPS9</i>	mitochondrial ribosomal protein S9	3.92×10^{-2}	-3.168

* $p < 0.05$ corrected by BH FDR

Table S4. List of datasets sourced for patient survival analysis.

OS	FP	PPS
caArray	caArray	caArray
GSE14814	GSE29013	GSE29013
GSE19188	GSE31210	GSE31210
GSE29013	GSE31908	GSE31908
GSE30219	GSE50081	GSE50081
GSE31210	GSE8894	
GSE3141		
GSE31908		
GSE37745		
GSE4573		
GSE50081		
TCGA		

OS: overall survival; FP: survival until first progression;

PPS: post (first) progression survival

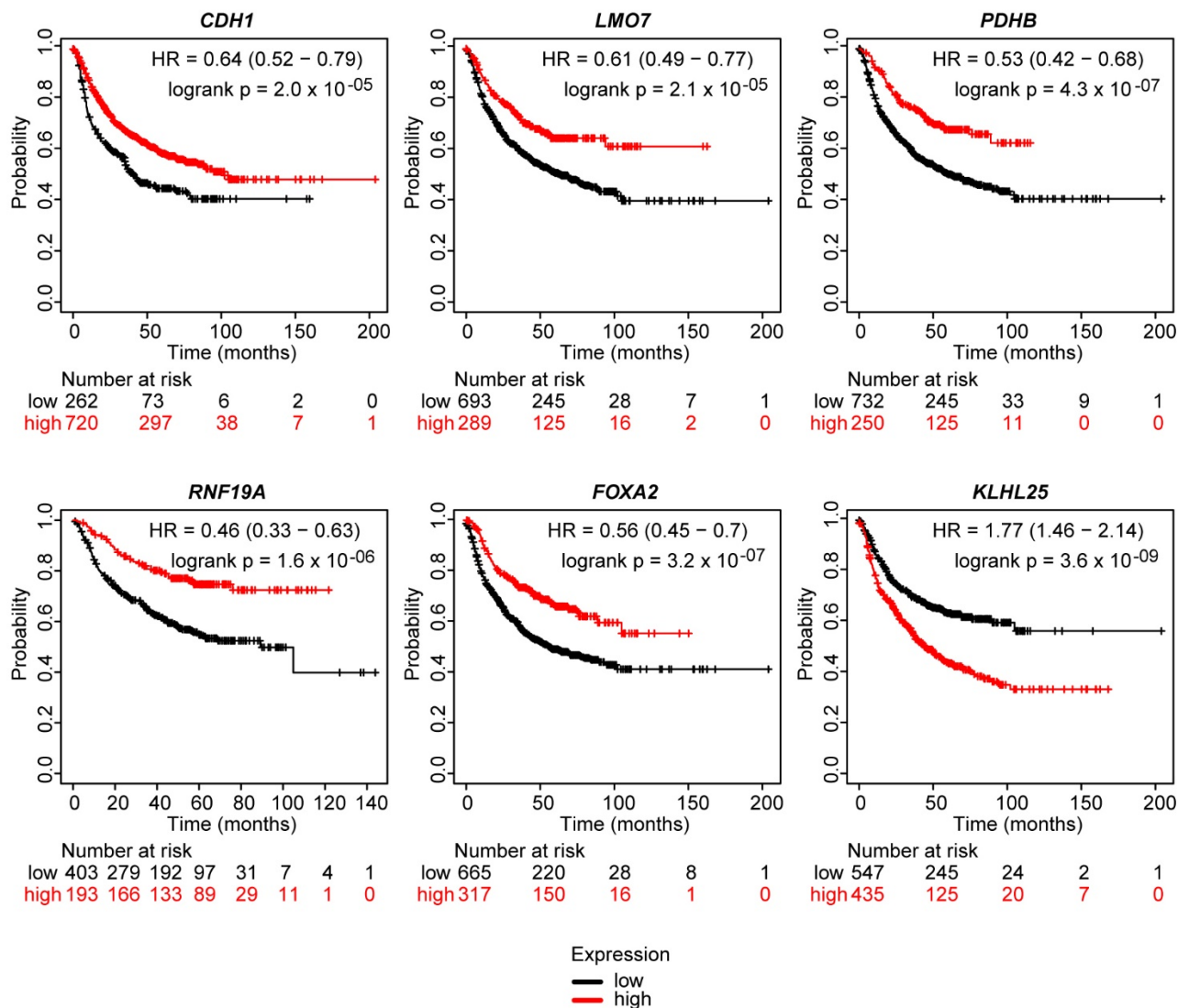


Figure S3. Prognostic value of genes associated with EMT in NSCLC – Survival until first progression (FP). Kaplan–Meier survival curves showing the correlation between marker gene expression and survival until FP in patients with NSCLC (n=596). Significance thresholds for potential gene biomarkers include a log-rank test $p < 0.001$ and a corrected $p < 0.05$ by BH FDR for the patient cut-off selection method. Low expression levels of *CDH1*, *LMO7*, *PDHB*, *RNF19A* and *FOXA2* whereas high expression level of *KLHL25* were associated with reduced survival until FP. HR: hazard ratio.

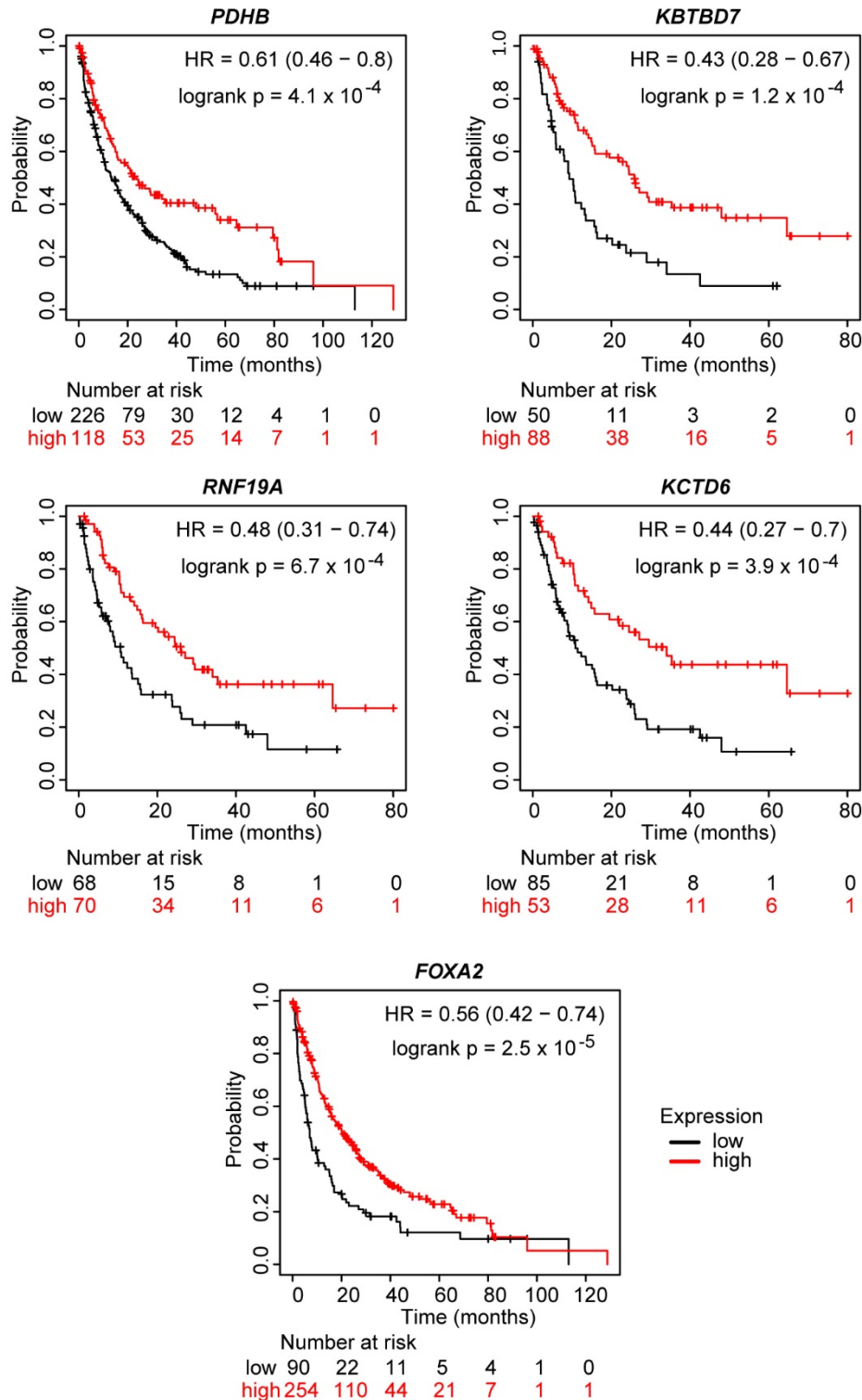


Figure S4. Prognostic value of genes associated with EMT in NSCLC – Post (first) progression survival (PPS). Kaplan–Meier survival curves showing the correlation between marker gene expression and PPS in patients with NSCLC (n=138). Significance thresholds for potential gene biomarkers include a log-rank test $p < 0.001$ and a corrected $p < 0.05$ by BH FDR for the patient cut-off selection method. Low expression levels of *PDHB*, *KBTBD7*, *RNF19A*, *KCTD6* and *FOXA2* were associated with reduced PPS. HR: hazard ratio.