

Supplemental Table 1. The EMT signature genes*

Affymetrix Probe	Gene Symbol	Gene Name	E-cadherin	Vimentin	N-cadherin	Fibronectin 1	Bimodal index	Accession	LocusLink	Chromosome	Cytoband
212764_at	ZEB1	Zinc finger E-box binding homeobox 1	-0.78	0.62	0.38	-0.05	1.75	BX647794	6935	10	10p11.22
210875_s_at	ZEB1	Zinc finger E-box binding homeobox 1	-0.68	0.54	0.16	-0.17	2.25	NM_030751	6935	10	10p11.22
225793_at	LIX1L	Lix1 homolog (mouse)-like	-0.67	0.54	0.28	-0.12	1.81	AK128733	128077	1	1q21.1
201426_s_at	VIM	Vimentin	-0.55	1.00	0.42	0.30	1.68	NM_003380	7431	10	10p12.33
202686_s_at	AXL	AXL receptor tyrosine kinase	-0.45	0.60	0.54	0.24	1.84	NM_021913	568	19	19q13.2
201069_at	MMP2	"Matrix metalloproteinase 2 (gelatinase A,	-0.27	0.30	0.56	0.22	1.83	NM_004530	4313	16	16q12.2
225524_at	ANTXR2	Anthrax toxin receptor 2	-0.26	0.40	0.09	0.55	1.74	NM_058172	118429	4	4q21.2
226891_at	C3orf21	Chromosome 3 open reading frame 21	-0.14	-0.14	-0.04	-0.54	2.19	NM_152531	152002	3	3q29
214702_at	FN1	Fibronectin 1	-0.08	0.27	0.09	0.58	1.62	NM_054034	2335	2	2q35
212298_at	NRP1	Neuropilin 1	-0.01	0.15	0.01	0.69	1.54	NM_003873	8829	10	10p11.22
201506_at	TGFB1	"Transforming growth factor, beta-inducer	0.07	0.09	-0.02	0.58	1.89	NM_000358	7045	5	5q31.1
229555_at	GALNT5	UDP-N-acetyl-alpha-D-galactosamine pol	0.15	0.22	0.14	0.55	1.82	NM_014568	11227	2	2q24.1
208510_s_at	PPARG	Peroxisome proliferator-activated recepto	0.15	0.03	-0.13	0.56	1.73	NM_015869	5468	3	3p25.2
211719_x_at	FN1	Fibronectin 1	0.15	0.27	0.11	0.97	1.50	NM_212482	2335	2	2q35
211732_x_at	HNMT	Histamine N-methyltransferase	0.24	-0.03	-0.02	0.57	1.99	NM_00102407	3176	2	2q22.1
204112_s_at	HNMT	Histamine N-methyltransferase	0.33	-0.07	0.02	0.63	1.55	NM_006895	3176	2	2q22.1
224414_s_at	CARD6	"Caspase recruitment domain family, mer	0.40	-0.16	0.02	0.61	1.83	NM_032587	84674	5	5p13.1
209488_s_at	RBPMS	RNA binding protein with multiple splicing	0.41	-0.22	-0.20	0.54	1.94	NM_00100871	11030	8	8q12
218856_at	TNFRSF21	"Tumor necrosis factor receptor superfam	0.48	-0.22	-0.06	0.56	1.51	NM_014452	27242	6	6p12.3
226226_at	TMEM45B	Transmembrane protein 45B	0.53	-0.41	-0.54	0.20	1.84	NM_138788	120224	11	11q24.3
238742_x_at			0.63	-0.67	-0.37	-0.14	2.22	NA	NA		
238778_at	MPF7	"Membrane protein, palmitoylated 7 (MAC	0.65	-0.44	-0.34	0.21	1.62	AL832380	143098	10	10p11.23
219919_s_at	SSH3		0.65	-0.48	-0.26	0.08	1.51	NM_018276	NA	11	11q13.1
234970_at	MTAC2D1		0.66	-0.64	-0.42	0.13	2.03	NA	NA		
207847_s_at	MUC1	"Mucin 1, cell surface associated"	0.66	-0.51	-0.43	0.19	1.63	NM_002456	4582	1	1q22
232164_s_at	EPPK1	Eppiklin 1	0.66	-0.47	-0.23	-0.08	1.95	NM_031308	83481	8	8q24.3
225548_at	SHROOM3	Shroom family member 3	0.67	-0.36	-0.25	0.24	1.84	NM_020859	57619	4	4q21.1
220318_at	EPN3	Epsin 3	0.67	-0.70	-0.48	0.08	2.07	NM_017957	55040	17	17q21.33
205847_at	PRSS22	"Protease, serine, 22"	0.67	-0.50	-0.41	0.16	1.72	NM_022119	64063	16	16p13.3
65517_at	AP1M2	"Adaptor-related protein complex 1, mu 2	0.67	-0.46	-0.29	0.14	3.39	NM_005498	10053	19	19p13.2
229842_at			0.68	-0.62	-0.31	0.21	2.07	AC099676	NA		
204019_s_at	SH3YL1	"SH3 domain containing, Ysc84-like 1 (S.	0.68	-0.56	-0.40	0.13	1.58	NM_015677	26751	2	2p25.3
239853_at	KLC3	Kinesin light chain 3	0.68	-0.33	-0.12	0.01	1.85	NM_177417	147700	19	19q13.32
235988_at			0.68	-0.30	-0.31	0.40	1.74	AB065679	NA		
224762_at	SERINC2	Serine incorporator 2	0.69	-0.45	-0.24	0.06	1.63	NM_178865	347735	1	1p35.2
204503_at	EVPL	Envoplakin	0.69	-0.47	-0.42	0.22	1.78	NM_001988	2125	17	17q25.1
202489_s_at	FXYP3	FXYP domain containing ion transport reg	0.69	-0.70	-0.33	0.05	2.47	NM_021910	5349	19	19q13.11
201428_at	CLDN4	Claudin 4	0.69	-0.43	-0.35	0.40	2.11	NM_001305	1364	7	7q11.23
232609_at	CRB3	Crumbs homolog 3 (Drosophila)	0.69	-0.43	-0.35	0.05	1.68	NM_174881	92359	19	19p13.3
219476_at	LRRC54	"CDNA FLJ25280 fis, clone STM06543"	0.69	-0.41	-0.22	0.32	2.18	AK059009	NA	11	11q13.5
210058_at	MAPK13	Mitogen-activated protein kinase 13	0.69	-0.41	-0.39	0.23	1.54	NM_002754	5603	6	6p21.31
232165_at	EPPK1	Eppiklin 1	0.70	-0.51	-0.27	-0.11	2.13	AL137725	83481	8	8q24.3
203397_s_at	GALNT3	UDP-N-acetyl-alpha-D-galactosamine pol	0.70	-0.45	-0.29	0.24	1.81	NM_004482	2591	2	2q24.3
235144_at		"CDNA FLJ32320 fis, clone PROST2003	0.70	-0.49	-0.40	0.28	1.97	AK056882	NA	9	9q21.32
221610_s_at	STAP2	Signal transducing adaptor family membe	0.70	-0.49	-0.25	0.18	1.57	NM_00101384	55620	19	19p13.3
218261_at	AP1M2	"Adaptor-related protein complex 1, mu 2	0.70	-0.45	-0.16	0.14	2.71	NM_005498	10053	19	19p13.2
200606_at	DSP	Desmoplakin	0.70	-0.56	-0.30	-0.09	1.57	NM_004415	1832	6	6p24.3
219411_at	ELMO3	Engulfment and cell motility 3	0.71	-0.52	-0.41	0.09	1.71	NM_024712	79767	10	16q22.1
235148_at	KRTCAP3	Keratinocyte associated protein 3	0.71	-0.59	-0.42	0.02	2.50	NM_173853	200634	2	2p23.3
224650_at	MAL2	"Mal, T-cell differentiation protein 2"	0.71	-0.50	-0.44	0.21	2.55	NM_052886	114569	8	8q24.12
224097_s_at	F11R		0.72	-0.45	-0.38	0.09	1.57	NM_144504	NA	1	1q23.3
238689_at	GPR110	G protein-coupled receptor 110	0.72	-0.48	-0.41	0.32	1.79	NM_153840	286977	6	6p12.3
228441_s_at			0.72	-0.38	-0.29	0.12	1.64	AC092611	NA		
212070_at	GPR56	G protein-coupled receptor 56	0.72	-0.53	-0.33	0.27	1.80	NM_201525	9289	16	16q13
201650_at	KRT19	Keratin 19	0.73	-0.43	-0.33	0.35	2.58	NM_002276	3880	17	17q21.2
222830_at	GRHL1	Grainyhead-like 1 (Drosophila)	0.73	-0.52	-0.45	0.09	1.85	NM_198182	29841	2	2p25.1
218792_s_at	BSPRY	B-box and SPRY domain containing	0.73	-0.51	-0.37	0.09	1.53	NM_017688	54836	9	9q32
228865_at	C1orf116	Chromosome 1 open reading frame 116	0.73	-0.30	-0.10	0.44	1.58	NM_023938	79098	1	1q32.1
218677_at	S100A14	S100 calcium binding protein A14	0.73	-0.65	-0.37	0.09	1.96	NM_020672	57402	1	1q21.3
210715_s_at	SPINT2	"Serine peptidase inhibitor, Kunitz type, 2	0.73	-0.44	-0.35	0.07	1.91	NM_021102	10653	19	19q13.2
236489_at			0.74	-0.38	-0.28	0.32	1.89	AB065679	NA		
238439_at	ANKRD22	Ankyrin repeat domain 22	0.74	-0.49	-0.48	0.17	1.76	NM_144590	118932	10	10q23.31
216905_s_at	ST14	Suppression of tumorigenicity 14 (colon c	0.74	-0.50	-0.35	0.15	2.09	NM_021978	6788	11	11q24.3
219388_at	GRHL2	Grainyhead-like 2 (Drosophila)	0.74	-0.59	-0.44	0.11	1.86	NM_024915	79977	8	8q22.3
205980_s_at	PRR5	Rho GTPase activating protein 8	0.74	-0.46	-0.32	0.00	2.98	NM_00101752	55615	22	22q13.31
222746_s_at	BSPRY	B-box and SPRY domain containing	0.75	-0.44	-0.39	0.18	1.77	NM_017688	54836	9	9q32
35148_at	TJP3	Tight junction protein 3 (zona occludens	0.75	-0.61	-0.38	0.05	1.74	NM_014428	27134	19	19p13.3
202286_s_at	TACSTD2	Tumor-associated calcium signal transdu	0.75	-0.49	-0.30	0.18	2.15	NM_002353	4070	1	1p32.1
203256_at	CDH3	"Cadherin 3, type 1, P-cadherin (placenta	0.75	-0.42	-0.31	0.22	2.40	NM_001793	1001	16	16q22.1
236058_at	C1orf172	Chromosome 1 open reading frame 172	0.76	-0.64	-0.40	0.18	2.39	NM_152365	126895	1	1p36.11
205709_s_at	CDS1	CDP-diacylglycerol synthase (phosphatid	0.76	-0.50	-0.49	0.16	1.56	NM_001263	1040	4	4q21.23
37117_at	PRR5	Rho GTPase activating protein 8	0.76	-0.48	-0.32	0.01	1.77	NM_00101752	55615	22	22q13.31
203780_at	MPZL2	Myelin protein zero-like 2	0.76	-0.50	-0.36	0.13	1.69	NM_005797	10205	11	11q23.3
223681_s_at	INADL		0.76	-0.57	-0.15	0.09	1.59	AB044807	NA	1	1p31.3
223895_s_at	EPN3	Epsin 3	0.76	-0.65	-0.40	0.13	2.20	NM_017957	55040	17	17q21.33
219121_s_at	RBM35A	RNA binding motif protein 35A	0.76	-0.54	-0.38	0.22	2.12	NM_017697	54845	8	8q22.1
226403_at	TMC4	Transmembrane channel-like 4	0.77	-0.56	-0.38	0.18	1.53	NM_144686	147798	19	19q13.42
226535_at	ITGB6	"Integrin, beta 6"	0.77	-0.45	-0.37	0.36	2.25	AK026736	3694	2	2q24.2
225822_at	TMEM125	Transmembrane protein 125	0.78	-0.55	-0.40	0.18	2.33	NM_144626	128218	1	1p34.2
205977_s_at	EPHA1	EPH receptor A1	0.78	-0.54	-0.44	0.24	2.05	NM_005232	2041	7	7q34
226185_at	CDS1	"CDNA: FLJ23044 fis, clone LNG02454"	0.78	-0.59	-0.37	0.24	2.25	AK026697	NA	4	4q21.23
227803_at	ENPP5	Ectonucleotide pyrophosphatase/phosph	0.79	-0.45	-0.21	0.23	1.88	NM_021572	59084	6	6p12.3
202005_at	ST14	Suppression of tumorigenicity 14 (colon c	0.79	-0.51	-0.36	0.17	2.34	NM_021978	6788	11	11q24.3
229292_at	EPB41L5	Erythrocyte membrane protein band 4.1 li	0.79	-0.55	-0.49	-0.05	1.92	BC032822	57669	2	2q14.2
202454_s_at	ERBB3	V-erb-b2 erythroblastic leukemia viral onc	0.79	-0.53	-0.34	0.15	1.64	NM_001982	2065	12	12q13.2
218186_at	RAB25	"RAB25, member RAS oncogene family"	0.80	-0.50	-0.34	0.21	2.88	NM_020387	57111	1	1q22
202525_at	PRSS8	"Protease, serine, 8"	0.80	-0.58	-0.37	0.19	2.01	NM_002773	5652	16	16p11.2
239148_at			0.80	-0.63	-0.39	-0.02	1.99	AC090907	NA		
213285_at	TMEM30B	Transmembrane protein 30B	0.80	-0.60	-0.40	0.16	2.03	NM_00101797	161291	14	14q23.1
242354_at			0.80	-0.73	-0.46	-0.04	2.17	NA	NA		
202790_at	CLDN7	Claudin 7	0.80	-0.51	-0.35	0.19	2.12	NM_001307	1366	17	17p13.1
225846_at	RBM35A	RNA binding motif protein 35A	0.81	-0.63	-0.52	0.06	2.60	NM_00103491	54845	8	8q22.1
236279_at			0.81	-0.55	-0.27	0.15	2.37	AC010503	NA		
201839_s_at	TACSTD1	Tumor-associated calcium signal transdu	0.82	-0.58	-0.39	-0.02	2.49	NM_002354	4072	2	2q21
226187_at	CDS1	"CDNA: FLJ23044 fis, clone LNG02454"	0.82	-0.57	-0.37	0.21	1.58	AK026697	NA	4	4q21.23
203453_at	SCNN1A	"Sodium channel, nonvoltage-gated 1 alp	0.83	-0.52	-0.28	0.28	1.97	NM_001038	6337	12	12p13.31
201131_s_at	CDH1		1.00	-0.55	-0.22	0.15	2.44	NM_004360	NA	16	16q22.1

*Ninety-six probes representing 76 unique bimodally distributed genes that correlated with E-cadherin (*CDH1*), vimentin (*VIM*), N-cadherin (*CHD2*), and/or fibronectin 1 (*FN1*) were identified

in the NSCLC training set. Individual probes are ranked in the table by their correlation with E-cadherin.