

SUPPLEMENTARY TABLE 3**HUMAN MICROARRAY DATA PROCESSING AND SURVIVAL STATISTICAL ANALYSIS**

To generate a high confidence human *Ets2*-TAM signature, 407 probe sets (357 genes) with absolute INT values > 1.5 were compared to the 98 lymph-node negative Rosetta cohort ((30), www.rii.com/publications/2002) divided in to 2 groups based on lymphocyte/leukocyte infiltration status. Out of 2856 probe sets that had a significant differential expression profile between the two groups ($p < 0.001$, Student's t-test) 341 probe sets matched the 407 mouse homologues, with 142 mouse probe sets having p-value less than 0.05.

Microarray datasets (GSE2034 and GSE1456) were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/geo). The Rosetta dataset was downloaded from Rosetta Company website (www.rii.com/publications/2002/vantveer.html). The Affymetrix probe IDs denoting the *Ets2*-TAM Gene Signature were used to extract expression data from the GSE2034 and GSE1456 Series Matrix Table. Both of these studies were performed on an Affymetrix HU133 A platform. A total of 239 probes, corresponding to 118 genes, were retrieved from these datasets. For the NKI dataset, which is a custom array, HUGO Gene Symbols of the *Ets2*-TAM signature were utilized to extract out expression values for the corresponding genes present on the NKI microarray chip. This resulted in a dataset that had expression values for 113 genes (128 counting replicates for some genes).

Analysis of gene expression in breast cancer datasets were performed using BRB-Array Tools developed by Dr. Richard Simon and BRB-Array Tools Development team. Expression data were loaded onto BRB-Array Tools using the Data Import Wizard. Normalization and log-transformation (to base 2) was performed on the Wang dataset (GSE2034) using median center array on BRB-Array Tools. The expression values provided for the other two datasets were already normalized and log-transformed. Unsupervised clustering of each dataset was performed by using the samples only clustering option of BRB ArrayTools. K-means clustering was performed by using the uncentered correlation {for NKI and

GSE1456 (Stockholm)} and absolute uncentered correlation (for Wang) metrics. Samples were assigned into two groups based on this clustering, and Kaplan–Meier survival analysis was performed by using the Survival Analysis module of the software package StatsDirect. Significance of survival analysis was performed by using the Log-Rank (Peto) test. Hazard ratio indicates the relative risk for the ‘bad’ group versus the ‘good’ group for tumor relapse/recurrence.

(Note: The significance analysis has been repeated for each dataset using the Cox Regression test and the results are consistent with the Log-Rank p-value.)

HUMAN *Ets2*-TAM SIGNATURE

- Genes highlighted in red are genes that are differentially regulated in lymphocyte/leukocyte infiltration –positive versus –negative breast cancer samples with high statistical significance ($p < 10^{-4}$). Those highlighted in yellow have a p-value between 10^{-4} and 10^{-3} . Genes highlighted in blue have a p-value between 10^{-3} and 0.05
- ‘Up’ and ‘down’ represents the trend of gene expression in lymphocyte/leukocyte infiltration –positive breast cancer samples when compared to the –negative samples

SUPPLEMENTARY TABLE 3. HUMAN *Ets2* -TAM SIGNATURE

Gene genes highlighted in red have $p < 0.0001$
 genes highlighted in yellow have $p < 0.001$
 genes highlighted in blue have $p < 0.05$

Accession#	Gene Symbol	P-value	Trend
U62325	APBB2	$<10^{-7}$	Down
NM_001673	ASNS	$<10^{-7}$	Up
NM_017680	ASPN	$<10^{-7}$	Down
NM_004049	BCL2A1A	$<10^{-7}$	Up
NM_001770	CD19	0.000003	Up
NM_001797	CDH11	0.000001	Down
AF018081	COL18A1	0.000006	Down
Contig2578_RC	COL1A1	0.000005	Down
NM_000089	COL1A2	$<10^{-7}$	Down
NM_000090	COL3A1	$<10^{-7}$	Down
NM_000393	COL5A2	0.000065	Down
NM_004369	COL6A3	$<10^{-7}$	Down
NM_001920	DCN	$<10^{-7}$	Down
NM_006892	DNMT3B	$<10^{-7}$	Down
NM_001387	DPYSL3	0.000003	Down
NM_000136	FANCC	0.000038	Down
NM_000138	FBN1	$<10^{-7}$	Down
NM_002775	HTRA1	$<10^{-7}$	Down
NM_001553	IGFBP7	$<10^{-7}$	Down
NM_014214	IMPA2	0.000037	Up
NM_005544	IRS1	$<10^{-7}$	Down
NM_002290	LAMA4	$<10^{-7}$	Down
NM_002293	LAMC1	0.000008	Down
NM_005576	LOXL1	$<10^{-7}$	Down
NM_003829	MPDZ	0.000006	Down
NM_003953	MPZL1	$<10^{-7}$	Up
NM_002487	NDN	$<10^{-7}$	Down
NM_003894	PER2	$<10^{-7}$	Down
NM_007173	PRSS23	0.000002	Down
AL049670	RPL34P1	0.000012	Up
NM_002964	S100A8	0.000002	Up
NM_002965	S100A9	0.000005	Up
NM_004694	SLC16A6	0.000014	Down
Contig2237_RC	SMOC2	$<10^{-7}$	Down
NM_006941	SOX10	0.000047	Up
NM_014467	SRPX2	$<10^{-7}$	Down
NM_003247	THBS2	0.000082	Down
NM_000362	TIMP3	$<10^{-7}$	Down
NM_006670	TPBG	0.000042	Down
NM_014232	VAMP2	0.000001	Down
NM_004385	VCAN	$<10^{-7}$	Down
NM_005161	APLNR	0.00015	Up
NM_001657	AREG	0.00032	Down
NM_001681	ATP2A2	0.000146	Down
NM_004327	BCR	0.000572	Up
NM_006129	BMP1	0.000106	Up
NM_001250	CD40	0.000648	Down

Gene genes highlighted in red have $p < 0.0001$
 genes highlighted in yellow have $p < 0.001$
 genes highlighted in blue have $p < 0.05$

Accession#	Gene Symbol	P-value	Trend
NM_000093	COL5A1	0.000278	Down
NM_004378	CRABP1	0.000553	Up
U67784	CXCR7	0.000516	Down
NM_001943	DSG2	0.000209	Up
NM_004461	FARSB	0.000448	Up
AL050228	FSTL1	0.000312	Down
NM_001530	HIF1A	0.000208	Up
X58529	IGH-6	0.000291	Down
AJ005273	KIN	0.000191	Up
NM_001206	KLF9	0.000138	Down
NM_000224	KRT18	0.000366	Down
NM_005556	KRT7	0.000108	Up
NM_002345	LUM	0.000512	Down
NM_004552	NDUFS5	0.000281	Up
NM_004289	NFE2L3	0.000246	Up
NM_014057	OGN	0.000878	Down
NM_017456	PSCD1	0.00027	Up
NM_006868	RAB31	0.000448	Down
Contig50913_RC	SMOC1	0.000607	Up
NM_004175	SNRPD3	0.000263	Up
NM_006307	SRPX	0.00025	Down
NM_003312	TST	0.000771	Up
NM_003474	ADAM12	0.003375	Down
NM_006988	ADAMTS1	0.030525	Down
NM_007038	ADAMTS5	0.009199	Down
NM_012098	ANGPTL2	0.004436	Down
NM_001149	ANK3	0.049777	Down
NM_001630	ANXA8	0.029727	Up
AK001565	AP3M1	0.046754	Down
NM_000648	CCR2	0.002891	Down
NM_020404	CD248	0.018763	Down
NM_000616	CD4	0.018636	Down
AF098641	CD44	0.004086	Up
NM_001778	CD48	0.005904	Up
NM_004931	CD8B1	0.016986	Down
Y16521	CDS2	0.004179	Down
NM_013943	CLIC4	0.038808	Up
NM_017649	CNNM2	0.041078	Down
M92642	COL16A1	0.039434	Up
NM_001881	CREM	0.00746	Up
NM_001896	CSNK2A2	0.041543	Up
NM_001901	CTGF	0.00325	Down
NM_004734	DCLK1	0.002595	Down
NM_001937	DPT	0.02876	Down
NM_000121	EPOR	0.003027	Down
NM_000131	F7	0.003918	Down
NM_003902	FUBP1	0.009583	Up
NM_005300	GPR34	0.001835	Down
NM_012259	HEY2	0.028872	Down
NM_001528	HGFAC	0.036221	Up
X69111	ID3	0.008904	Down

Gene genes highlighted in red have $p < 0.0001$
 genes highlighted in yellow have $p < 0.001$
 genes highlighted in blue have $p < 0.05$

Accession#	Gene Symbol	P-value	Trend
NM_000877	IL1R1	0.011431	Down
X52015	IL1RN	0.02708	Up
NM_007035	KERA	0.011078	Down
NM_000421	KRT10	0.013334	Down
NM_002273	KRT8	0.002858	Down
NM_000595	LTA	0.013368	Up
NM_005926	MFAP1A	0.001466	Down
NM_006983	MMP23B	0.001343	Down
NM_000139	MS4A1	0.001404	Down
AF234532	MYO10	0.004972	Up
NM_000434	NEU1	0.007816	Down
NM_006164	NFE2L2	0.027787	Up
X72631	NR1D1	0.0059	Down
NM_012327	PIGN	0.031566	Down
NM_003628	PKP4	0.012	Down
NM_006347	PPIH	0.001564	Up
M30773	PPP3R1	0.007799	Down
Contig50675	PRRX1	0.042689	Down
NM_000963	PTGS2	0.002694	Down
NM_002852	PTX3	0.002254	Up
NM_006505	PVR	0.020424	Down
NM_003702	RGS20	0.034815	Up
AL137448	RIPK4	0.002276	Down
NM_000994	RPL32	0.046897	Up
NM_001001	RPL36A	0.008991	Down
Contig2236_RC	RPL37A	0.002411	Up
NM_001000	RPL39	0.009187	Up
Contig1508_RC	SDC2	0.019969	Down
NM_012431	SEMA3E	0.026089	Down
NM_014331	SLC7A11	0.043474	Down
NM_002354	TACSTD1	0.010319	Up
NM_000361	THBD	0.001286	Down
NM_003313	TSTA3	0.008105	Up
AB002347	UBR2	0.00775	Down
NM_006634	VAMP5	0.002309	Up