Supplementary Information

Manuscript Title: Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm

Authors: Gennadi V. Glinsky, Takuya Higashiyama, Anna B. Glinskii

Affiliations: Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, CA 92121

Table of Contents

1. Protocol of discovery and validation of the breast cancer survival predictor algorithm. We hypothesized that clinically relevant genetic signatures could be found by searching for clusters of co-regulated genes that display highly concordant transcript abundance behavior across multiple experimental models and clinical settings which are modeling or representing malignant phenotypes of interest (1-4). Thus, according to this model the primary criterion in transcript selection process should be the concordance of changes in expression rather than a magnitude of changes (e.g., fold change). One of the predictions of this model is that transcripts of interest would be expected to have a tightly controlled "rank order" of expression within a cluster of co-regulated genes reflecting a balance of up- and down-regulated mRNAs as a desired regulatory end-point in a cell. A degree of resemblance of the transcript abundance rank order within a gene cluster between a test sample and reference standard is measured by a Pearson correlation coefficient and designated as a phenotype association index (PAI). To identify genes with consistently concordant expression patterns across multiple data sets and various experimental conditions, we compared the expression profile of 70 genes in clinical samples (test samples) to the expression profiles of transcripts differentially regulated in multiple established human breast cancer cell lines (reference standard).

The transcripts comprising each breast cancer survival predictor signature were selected based on Pearson correlation coefficients ($r > 0.95$) reflecting a degree of similarity of expression profiles in clinical tumor samples (metastatic versus non-metastatic tumors) and experimental samples using the following protocol.

Step 1. Expression profiles of transcripts comprising 70-gene signature were independently quantified for each experimental conditions (see below) and clinical

samples (training set of 78 breast cancer samples) using the quantitative RT-PCR protocol and Affymetrix microarray processing and statistical analysis software package (for U95Av2 microarray data sets) as described in Materials and Methods.

Step 2. Sub-sets of transcripts exhibiting concordant expression changes in clinical and experimental samples were identified. From a set of 70 transcripts, sub-sets of transcripts were identified with concordant changes of transcript abundance behavior in metastatic (34 samples) versus non-metastatic (44 samples) clinical breast tumor samples and experimental conditions independently defined for each signature by comparing gene expression in a given breast cancer cell line to the normal human breast epithelial cell lines. Thus, from a set of 70 transcripts seven concordant sub-sets of transcripts were identified corresponding to each binary comparison of clinical (34 poor prognosis samples versus 44 good prognosis samples) and experimental samples (MCF7; MDA-MB-435; MDA-MB-468; MDA-MB-231; MDA-MB-435Lung2; MDA-MB-435Br1; MDA-MB-435BL3 human breast carcinoma cell lines versus primary cultures of normal human breast epithelial cells). For each breast cancer cell line concordant sets of genes were identified exhibiting both positive and negative correlation between fold expression changes in cancer cell lines versus control cell line and poor prognosis group (34 samples) versus good prognosis group (44 samples).

Step 3. Next, minimum segregation sets were selected from corresponding concordance sets and individual phenotype association indices were calculated. Selection of small gene clusters (minimum segregation sets) was performed from sub-sets of genes exhibiting concordant changes of transcript abundance behavior in poor prognosis versus good prognosis clinical tumor samples and experimental conditions defined for each

signature. Expression profiles were presented as Log10 average fold changes for each transcript and processed for visualization and Pearson correlation analysis using Microsoft Excel software. Cut-off criterion was set to exceed an absolute value of a Pearson correlation coefficient 0.95.

Step 4. Identified small gene clusters exhibiting highly concordant pattern of expression (Pearson correlation coefficient, $r > 0.95$; Table S4) in clinical and experimental samples were evaluated for their ability to discriminate clinical samples with distinct outcome after the therapy. To assess a potential prognostic relevance of individual gene clusters, we calculated a Pearson correlation coefficient for each of 78 tumor samples (training data set) by comparing the expression profiles of individual samples to the reference expression profiles of relevant experimental samples defined for each signature and an "average" expression profile of poor prognosis versus good prognosis clinical samples. Based on expected correlation of expression profiles of identified gene clusters with clinical behavior of breast cancer, we named the corresponding correlation coefficients calculated for individual samples the phenotype association indices (PAIs). We evaluated the prognostic power of identified clusters of co-regulated transcripts based on their ability to segregate the patients with metastatic and non-metastatic breast tumors into distinct sub-groups and selected four best performing clusters for follow-up validation experiments (Figure 1; Tables 1 and S1).

Step 5. We used the Kaplan-Meier survival analysis to assess the prognostic power of each best performing cluster in predicting the probability that patients would remain disease-free after therapy (Figures 1-7; Tables 1, 2, and S1). We selected the prognosis discrimination cut-off value for each signature based on highest level of

statistical significance in patient's stratification into poor and good prognosis groups as determined by the log-rank test (lowest P value and highest hazard ratio; Table S1 & Figures 1, 2, & S1). Clinical samples having the Pearson correlation coefficient at or higher the cut-off value were identified as having the poor prognosis signature. Clinical samples with the Pearson correlation coefficient lower the cut-off value were identified as having the good prognosis signature.

Step 6. We developed a breast cancer survival predictor algorithm taking into account calls from all four individual signatures. We accepted the prognosis discrimination cut-off value for all four signatures based on highest level of statistical significance in patient's stratification into poor and good prognosis groups as determined by the Kaplan-Meier survival analysis (lowest P value and highest hazard ratio defined by the log-rank test; Table S1 & Figures 2 and S1). Clinical samples having at least two poor survival signatures were stratified into the poor prognosis sub-group. Clinical samples exhibiting at least three good survival signatures were stratified into the good prognosis sub-groups (Table S1).

Step 7. We validated the prognostic power of breast cancer survival predictor signatures alone and in combination with the established markers of outcome using an independent set of clinical samples obtained from 295 breast cancer patients (Figures 2-7; Table S1).

References

1. Glinsky, G.V., Krones-Herzig, A., Glinskii, A.B., Gebauer, G. Microarray analysis of xenograft-derived cancer cell lines representing multiple

experimental models of human prostate cancer. Molecular Carcinogenesis, 37: 209-221, 2003.

2. Glinsky, G.V., Krones-Herzig, A., Glinskii, A.B. Malignancy-associated regions of transcriptional activation: gene expression profiling identifies common chromosomal regions of a recurrent transcriptional activation in human prostate, breast, ovarian, and colon cancers. Neoplasia, 5: 21-228.

3. Glinsky, G.V., Ivanova, Y.A., Glinskii, A.B. Common malignancy-associated regions of transcriptional activation (MARTA) in human prostate, breast, ovarian, and colon cancers are targets for DNA amplification. Cancer Letters, 201: 67-77, 2003.

4. Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M., Gerald, W.L. Gene expression profiling predicts clinical outcome of prostate cancer. Journal of Clinical Investigation, in press, 2004.

2. Quantitative reverse-transcription polymerase chain reaction (Q-RT-PCR) analysis protocol. Quantitative RT-PCR analysis of transcripts abundance levels for genes of the breast cancer survival predictor cluster was performed using an ABI7900 instrument (Applied Biosystems, Foster City, CA). Primer design, assay validation, and Q-PCR analysis were performed as previously described (27-29) and according to the vendor's recommended protocols (http://appliedbiosystems.com/support/tutorials/). For quantification, a reference curve was generated for each gene by amplifying serial dilution of cDNA and expression values were normalized using GAPDH and mRNA from normal human breast epithelial cell line (Clonetics/BioWhittaker, San Diego, CA) as controls. Each 15 µl reaction contained 5 µl primers (0.3 µM); 2.5 µl cDNA; 7.5 µl CyBr Green Myx (Apllied Biosystems, Foster City, CA) and was started at 95^0C for 10 min and carried out for 45 cycles (94^0C, 10 sec; 60^0C, 20 sec; 72^0C, 30 sec). Primer sequences are shown in the Supplementary Table S3.

Table S1. Stratification of 295 breast cancer patients at the time of diagnosis into poor and good prognosis groups using different therapy outcome predictor signatures

Outcome signature (cut off value)	Poor prognosis, 5-(10)-year survival	Good prognosis, 5-(10)-year survival	Correct predictions, poor outcome	Correct predictions, good outcome	Hazard ratio	95% Confidence interval	P value
70-gene (0.45)	75% (56%)	97% (92%)	70 of 79 (89%)	106 of 216 (49%)	6.327	2.498 to 6.077	<0.0001
70-gene (0.00)	64% (46%)	91% (80%)	42 of 79 (53%)	174 of 216 (81%)	3.867	3.405 to 9.809	<0.0001
13-gene (0.12)	73% (56%)	98% (93%)	71 of 79 (90%)	106 of 216 (49%)	7.005	2.560 to 6.237	<0.0001
13-gene (0.04)	73% (54%)	97% (92%)	69 of 79 (87%)	115 of 216 (53%)	6.519	2.728 to 6.610	<0.0001
13-gene (0.00)	73% (54%)	96% (90%)	67 of 79 (85%)	118 of 216 (55%)	5.698	2.663 to 6.450	<0.0001
14-gene (0.37)	77% (62%)	96% (91%)	72 of 79 (91%)	79 of 216 (37%)	5.220	1.912 to 4.874	<0.0001
14-gene (0.28)	76% (59%)	95% (89%)	69 of 79 (87%)	95 of 216 (44%)	4.701	2.038 to 5.016	<0.0001
14-gene (0.00)	75% (55%)	92% (85%)	58 of 79 (73%)	130 of 216 (60%)	3.637	2.217 to 5.419	<0.0001
14-gene (-0.55)	65% (45%)	91% (81%)	45 of 79 (57%)	176 of 216 (81%)	4.171	3.632 to 10.21	<0.0001
6-gene (-0.12)	78% (62%)	96% (88%)	70 of 79 (89%)	85 of 216 (39%)	4.543	1.901 to 4.756	<0.0001
6-gene (0.00)	78% (60%)	92% (86%)	64 of 79 (81%)	101 of 216 (47%)	3.314	1.757 to 4.282	<0.0001
4-gene (0.20)	73% (53%)	93% (85%)	60 of 79 (76%)	136 of 216 (63%)	4.389	2.723 to 6.735	<0.0001
4-gene (0.00)	75% (58%)	93% (84%)	60 of 79 (76%)	119 of 216 (55%)	3.519	2.050 to 4.983	<0.0001
Algorithm	77% (60%)	98% (94%)	75 of 79 (95%)	84 of 216 (39%)	10.05	2.355 to 5.956	<0.0001

Legend: 295 breast cancer patients were classified according to whether they had a good-prognosis signature or poor-prognosis signature defined by individual therapy outcome predictor signatures. Kaplan-Meier analysis was performed to evaluate the probability that patients would survive according to whether they had a poor-prognosis or a good-prognosis signature and determine the proportion of patients who would survive at least 5 or 10 years after therapy in poor-prognosis and good-prognosis sub-groups. Hazard ratios, 95% confidence intervals, and P values were calculated with use of the log-rank test. The number of correct predictions in poor-prognosis and good-prognosis groups is shown as a fraction of patients with the observed clinical outcome after therapy (79 patients died and 216 patients remained alive). The classification performance of different signatures were evaluated using one common threshold level (0.00) and optimized threshold levels adjusted for each gene cluster to achieve the most statistically significant (highest hazard ratio and lowest P value) discrimination in survival probability between patients assigned to poor and good prognosis groups.

Table S2. Estimated therapeutic benefits of using gene expression survival predictor signatures for classification of breast cancer patients

Classification category	5-year survival	10-year survival	Number (%) of patients	Good outcome (current)	Good outcome (projected)	Estimated increase in 10-year survival, %
LN-negative	82%	69%	151/295 (51%)			
LN-positive	85%	72%	144/295 (49%)			
LN- Good signature	92%	82%	95/151 (63%)	95	95	0.00
LN- Poor signature	64%	46%	56/151 (37%)	0	17 (56 x 0.3)	23%
LN+ Good signature	98%	98%	43/144 (30%)	43	43	0.00
LN+ Intermediate	86%	73%	67/144 (47%)	0	20 (67 x 0.3)	10%
LN+ Poor signature	68%	43%	34/144 (24%)	0	10 (34 x 0.3)	13%
Overall				**138/295 (47%)**	**185/295 (63%)**	**5%**
ER+ tumors	90%	77%	226/295 (77%)			
ER- tumors	62%	47%	69/295 (23%)			
ER+ LN-						
Good signature	97%	86%	69/109 (63%)	69	69	0.00
Poor signature	76%	57%	40/109 (37%)	0	15 (40 x 0.37)	17%
ER- LN-						
Good signature	74%	74%	16/42 (38%)	16	16	0.00
Poor signature	50%	34%	26/42 (62%	0	11 (25 x 0.41)	44%
ER+ LN+						
Good signature	98%	98%	43/117 (37%)	43	43	0.00
Poor signature	86%	68%	74/117 (63%)	0	27 (74 x 0.37)	16%
ER- LN+						
Good signature	82%	82%	11/27 (41%)	11	11	0.00
Poor signature	47%	24%	16/27 (59%)	0	7 (16 x 0.41)	100%
Overall				**139/295 (47%)**	**199/295 (67%)**	**6%**

Legend: The estimate of potential therapeutic benefits is made in the cohort of 295 breast cancer patients (21) and based on the assumption that the use of additional cycle(s) of adjuvant systemic therapy would be prescribed to patients classified into poor prognosis sub-groups. In the cohort of 295 breast cancer patients, ten of 151 (6.6%) patients who had lymph node-negative disease and 120 of the 144 (83.3%) patients who had lymph node-positive disease had received adjuvant systemic therapy (21). We accepted the actual 5- and 10-year survival in the corresponding classification categories as the expected therapy outcome for a given sub-group. We assumed that each additional cycle of adjuvant systemic therapy would result in the same therapy outcome as was actually

documented in the most relevant sub-groups of the 295 patients. Therapy outcome for patients classified into poor prognosis sub-groups and treated with additional cycle(s) of adjuvant systemic therapy is expected to be in 37% of patients in good therapy outcome category for ER+LN+ and ER+LN- poor signature sub-groups and in 41% of patients in good therapy outcome category for ER-LN+ and ER-LN- poor signature sub-groups.

Table S3

Gene	Primer sequence	Primer sequence	Product
TRAG3	GCCCATTGTCCAACAACCA	TTGAAGCGGCGGTCTTTTA	123
KIAA1750	TACCTCAAATAGCCCCAGGAGT	ACCCGGAGATCCAAAACAGA	109
AI813331	GATCAAGAATTTCGGCAGCAC	GGACCATAAGGCAATTGAGCA	131
AI554061	CAGGATTTTCAGCAGAGCGTC	TTGCCTAAAGGTGGCATCCAG	122
AA555029RC	CCAAATCTGCCTGAAGCAAA	GCGGTCCTATGCAGTTCAAA	107
DIAPH3 Diaphanous	AACTACATGAATGCTGGCTCCC	AGCGTTGT TTTCTGATCTGCTG	104
FLT1	CAGAGGCATGGAGTTCCTGTCT	GGCAAGGCCAAAATCACAA	112
MMP9	ACGACGTCTTCCAGTACCGAGA	TAGGTCACGTAGCCCACTTGGT	112
DC13	GCTGTGATCCATCCTCATCTCC	CTCAACTCCCGATCAACATCAT	161
EXT1	CAGAGGTGGATTTCGCCTTCAC	TCCCCACCGCTCCTTAGAGTTA	104
PK428	TTAAGAGTTACGGCTGGTCCC	GCTCAAGGCGCTTAATTCTTCT	112
HEC	TGCCAGTGAGCTTGAGTCCTT	TTCAGTCGTGGTTTGCACAAC	136
ECT2	AGCTGAGCATTCCCTTTCCATA	GAAGGAACTGGAGTGGAGCTTT	126
GMPS	CAGAGAGTCAAAGCCTGCACA	CACCCTGCACACCTACAGTTTT	115
AI377418	CATGTCTCGCTAATAACCCCAGC	CCCTTCCCCTTTTGGCAAG	126
UCH37	CCAGGATGTCCATTTAGGCGA	CCAGGATGTCCATTTAGGCGA	139
AI283268	GTGATCTCGGCTCATTGCAAC	ACATGGTGAAACCCCGTCTCT	142
KIAA1067	TCTCTTAAAACACATTCCCGCC	AAAATGGCCAGTGAGCCTG	104
GNAZ	GCGGCTACGACCTGAAACTCTA	TGAGTGAGGTGTTGATGAACCAG	111
SERF1A	GCCCGCCAGAAAAACATGA	GCTGCCTTCTGCTTTTCTTGC	125
OXCT	AGTGTCCAGTGCGAAAACCA	ATGCGGTTGACACATTGCTT	120
ORC6L	GCACTGCTTTCAGCATGCAA	TCGACCTGCTGTCCAATCTTC	131
L2DTL	AAAACTGGTTGTTGGCCATGG	TGGCAATGTCTTTCCGCTCT	110
PRC1	ACGCTTTCTCGGCGTTTGT	ACAGCTCAACCCATTGGAACA	127
AF052162	TCACCTGAAGTCCCTCTGACATC	AAACCCGGAGCCTTTGCTT	126
COL4A2	CGGAGTTTGTGGATCGGATA	CATTCGATGAATGGTGTGGC	125
KIAA0175	GGAAGATGTGACCGCAAGTGA	TCCACCCCATTTGATTCTGTC	141
RAB6B	TTCCCATGACACTCCTTGCTTG	CCCATCCCACCCTACTCCTAAA	128
AI992158	ACTGCAGAAACCCAGACTGCTG	AGTTGCAGATTCCTCGACAAGG	138
DCK	AGCCACTCCAGAGACATGCTTA	CGAAGTTGGTTTTCAGTGTCCT	140
CENPA	TCCGAAAGCTTCAGAAGAGCA	TTGCCAATTGAAGTCCACACC	110
SM-20	AGCAGCATGGACGACCTGATA	ACATAACCCGTTCCATTGCC	113
MCM6	TGCTCCCAAAGCCTCCTTAAG	CCAGTTAACAAGCTCGCTCCTC	133
AKAP2	TGATGCAGACCCTCATGGA	TTCGATTAACCCGTGTGGC	138
AI741117	AGGAGGAGCCCGAATACCAC	CAGGTTGAGAGGTCCCCTCA	126
RFC4	GCAGCAACTCAGCTCGTCAAT	GCTAGGCATTTGTCAACTTCGG	113
DKFZP564D0462	ATCGGTCTCAGTGCATGCTGT	GCAGATTGGCTTTGCCAGAA	123
SLC2A3	CCAACTTCCTAGTCGGATTGCT	ATCCTCAAAAGTCCTGCCACG	146
MP1	CGAGAAAAAGGCGGTGCTTAT	TCGACAGCCTTCCCAAAAGA	122
AI224578	GGCAAGCTTTCCACGTCCT	AACCGCTGCTACCAAAGAGTG	101
AW024884	GTGCAGGTAGGCGAAGAAAGC	GCCCTTACCCCTCAAGACCA	101

FLJ11190	AGTGATGCCACCAGCACAGG	GGACCATTGCTCAGTGTCTTCAG	123
AI817737	CAGGATGGCAGGAATAGCACA	ATCTGGTCTGGCATTCACAGC	109
IGFBP5	CCAATTGTGACCGCAAAGGA	TGGCAGCTTCATCCCGTACTT	107
CCNE2	CTATTTGGCTATGCTGGAGGAA	TGCTCTTCGGTGGTGTCATAAT	104
ESM1	GGAAAATGCCTGAAATTCCCC	ACCCGGCAGCATTCTCTTT	142
AA834945	TGGATGAAACAGCTGAGCAGA	CCCGAATGTTGGAAGCTGA	114
NMU	GTCAGTTGTGCATCCGTTGCT	TTCCATTCCGTGGCCTGAA	140
LOC57110	CCGAGCGATAAGTACCGTTGA	TCCTCCTCCCAAATTCCTTCA	143
AI583960	CCTGAGTTGAGCCGGAAATTT	ATTGTAGCGCCTCAGGTGAAG	126
PECI	AGAAGTCTGGACCAGGCTGAA	CCCTGAAGGACATTGCATTCTT	135
AP2B1	CAGCCAGGAAACCCCAATTAC	TCTTGACTTGCAC CGATCACA	150
CFFM4	TGTTTTGGCATTACTGGATCCC	TGCTGTCAGCAAGGAGGAAGAG	136
TGFB3	GGCTTTGGACACCAATTACTGC	CCCTTAGGTTCATGGACCCACT	114
AA528243	CAGAAGTGCCTGAGATGGTTCA	CCCACGGTTCAACTGCTAAATT	150
AI918032	ATTCCTCTGCTCACCTCCCAA	CAGCTGCTGGTTACCGATTTG	127
HSA250839	TGGCAAAGAACAGATCCAGG	GAATATGATGAATTCCTCCCGG	101
GSTM3	GAAGGCTTTCATGTGCCGTT	TCAGCATACAGGCTTGTTGCC	124
BBC3	ACGACCTCAACGCACAGTACG	TCCCATGATGAGATTGTACAGG	95
CEGP1	TGGTGATGCGGAAAACCTCTT	TAGCGCTGTTCCCTTCATTGG	138
AI694320	AGTGGCTGTGAACTCCCTGTCA	AGGATTTAATTGGACCAGCCCC	101
WISP1	CGGATCTCCAATGTTAACGCC	AAGTTCATGGATGCCTCTGGC	146
ALDH4	TTATCCCATCGGCCATGTG	AAAGGATCCCAAGAGCTGCTC	123
KIAA1442	TGACCCTCTCCTACAAGTCCAA	TTCTGTAGCCTCTGGAATCCG	109
AA404325	TTGGTGCCTCAGCTTTCCCTT	TACTCCACCCCCACCAAACAGT	122
FGF18	CGGTAGTCAAGTCCGGATCAA	AACACACACTCCTTGCTGGTGC	108
GAPD	CCCTCAACGACCACTTTGTCA	TTCCTCTTGTGCTCTTGCTGG	144

Table S4

4-gene signature	NBEC	MDA-MB-435BL3	r = 0.952
Gene	Log10FoldC	Log10FoldChange	Rosetta Array ID
HEC	1	1.093580498	NM_006101
MCM6	1	0.867324462	NM_005915
GSTM3	1	0.177651907	NM_000849
FGF18	1	-0.300315323	NM_003862

6-gene signature	NBEC	MDA-MB-468	r = 0.981
Gene	Log10FoldC	Log10FoldChange	Rosetta Array ID
FLT-1	1	2.485746116	NM_002019
BBC3	1	1.653815502	U82987
TGFB3	1	0.903686276	NM_003239
CFFM4	1	0.887972293	AF201951
GSTM3	1	0.65708388	NM_000849
FGF18	1	0.490464867	NM_003862

13-gene signature	NBEC	MCF7	r = - 0.992
Gene	Log10FoldC	Log10FoldChange	Rosetta Array ID
CEGP1	1	3.110682867	NM_020974
FGF18	1	1.601769472	NM_003862
GSTM3	1	1.538610306	NM_000849
TGFB3	1	0.849094266	NM_003239
CFFM4	1	0.819324227	AF201951
AI918032	1	0.470298048	Contig55377_RC
AP2B1	1	0.391249134	NM_001282
CCNE2	1	0.152420909	NM_004702
KIAA0175	1	0.074875381	NM_014791
EXT1	1	-0.253876939	NM_000127
AI813331	1	-0.433512401	Contig46218_RC
PK428	1	-0.566454237	NM_003607
AI554061	1	-0.569885453	Contig38288_RC

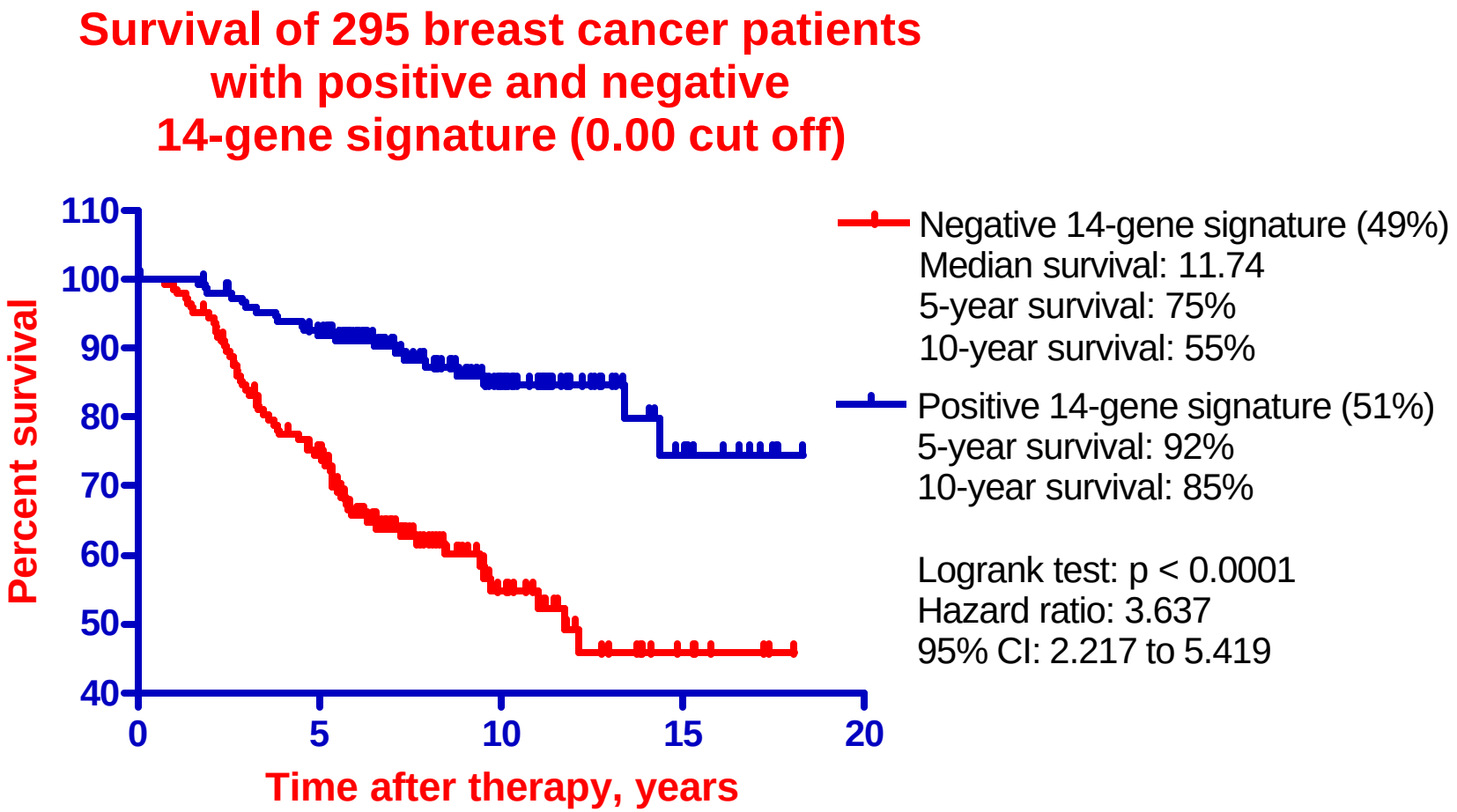
14-gene signature	NBEC	MDA-MB-435Br1	r = - 0.992
Gene	Log10FoldC	Log10FoldChange	Rosetta Array ID
CFFM4	1	1.392573779	AF201951
TGFB3	1	1.360251708	NM_003239
BBC3	1	0.48989205	U82987
AP2B1	1	0.429447816	NM_001282
ALDH4	1	0.321802308	NM_003748
FLJ11190	1	-0.226325951	NM_018354
DC13	1	-0.235413188	NM_020188
GMPS	1	-0.251668057	NM_003875
AKAP2	1	-0.271262874	Contig57258_RC
DCK	1	-0.300637148	NM_000788
ECT2	1	-0.466667713	Contig25991
AI554061	1	-0.491217835	Contig38288_RC
OXCT	1	-0.504651147	NM_000436
EXT1	1	-0.581483128	NM_000127

NBEC, normal human breast epithelial cells
r, Pearson correlation coefficient to the average expression value in the poor prognosis sub-group of the 78 breast cancer patients

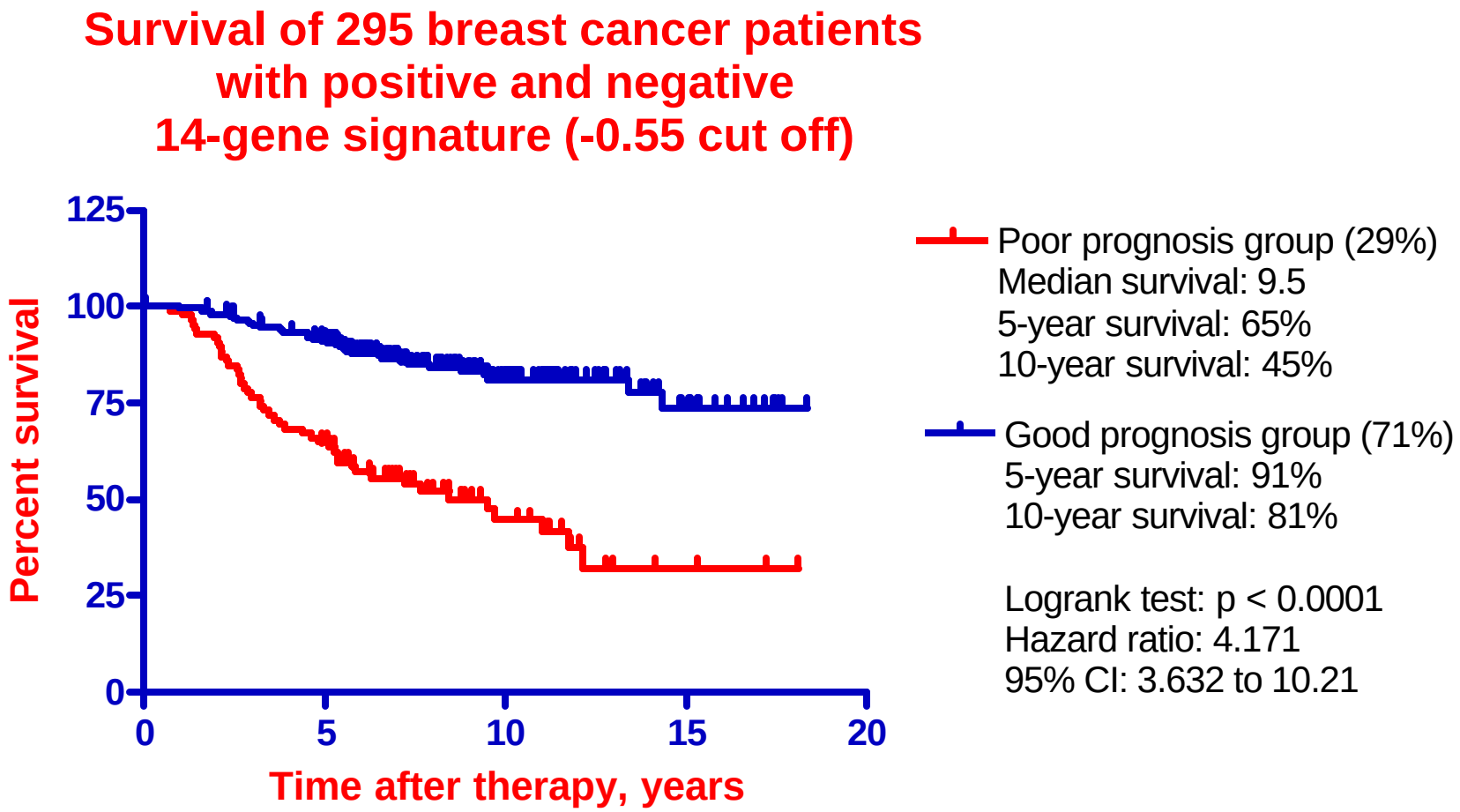
Figure S1. Kaplan-Meier analysis of the survival probability among 295 breast cancer patients comprising a signature validation group according to whether they had a good-prognosis or poor-prognosis defined by the 14-gene survival predictor signature. 295 patients were stratified into sub-groups using the values of the 14-gene expression profile at the different cut-off levels (panels A-E) or into sub-groups using a 10% increment from bottom to top values of the 14-gene expression profile (panel F). Statistical significance of the differences in the survival probability between sub-groups was assessed using Chi square and log rank tests. In the panel F, differences in the survival probability between sub-groups was at the

$p < 0.0001$ levels.

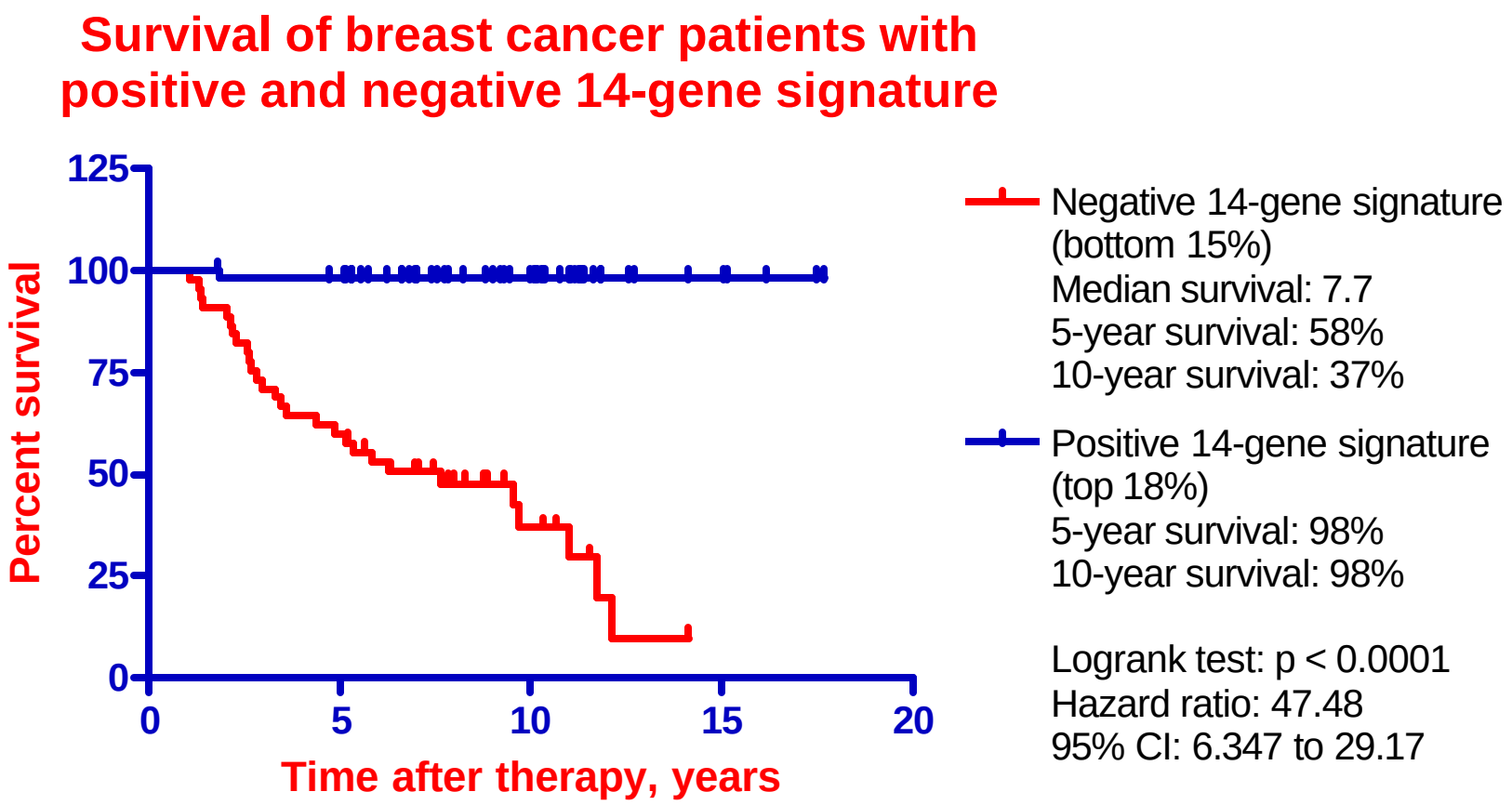
A


Survival of 295 breast cancer patients
with positive and negative
14-gene signature (0.00 cut off)
Percent survival
110
100
90
80
70
60
50
40
0
5
10
15
20
Time after therapy, years
Negative 14-gene signature (49%)
Median survival: 11.74
5-year survival: 75%
10-year survival: 55%
Positive 14-gene signature (51%)
5-year survival: 92%
10-year survival: 85%
Logrank test: p < 0.0001
Hazard ratio: 3.637
95% CI: 2.217 to 5.419

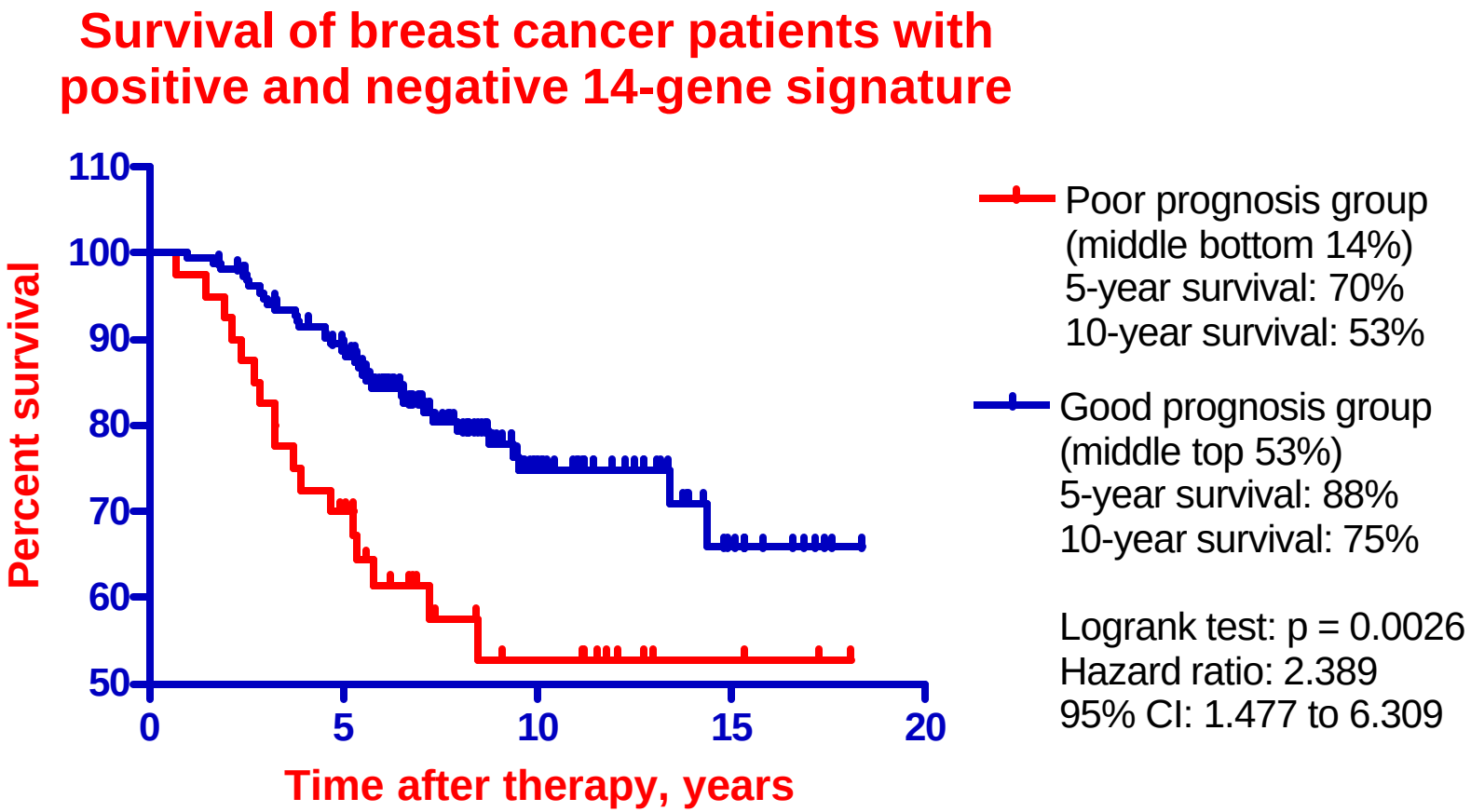
B


Survival of 295 breast cancer patients
with positive and negative
14-gene signature (-0.55 cut off)
Percent survival
125
100
75
50
25
0
0
5
10
15
20
Time after therapy, years
Poor prognosis group (29%)
Median survival: 9.5
5-year survival: 65%
10-year survival: 45%
Good prognosis group (71%)
5-year survival: 91%
10-year survival: 81%
Logrank test: p < 0.0001
Hazard ratio: 4.171
95% CI: 3.632 to 10.21

C

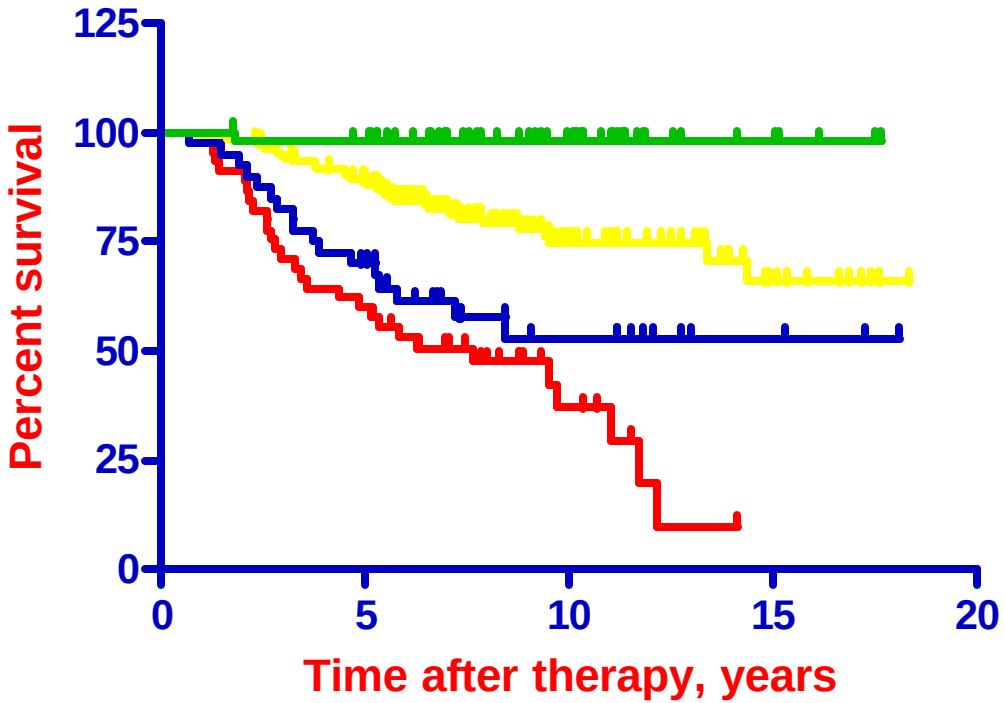

Survival of breast cancer patients with positive and negative 14-gene signature
Percent survival
125
100
75
50
25
0
0
5
10
15
20
Time after therapy, years
Negative 14-gene signature (bottom 15%)
Median survival: 7.7
5-year survival: 58%
10-year survival: 37%
Positive 14-gene signature (top 18%)
5-year survival: 98%
10-year survival: 98%
Logrank test: p < 0.0001
Hazard ratio: 47.48
95% CI: 6.347 to 29.17

D


Survival of breast cancer patients with positive and negative 14-gene signature
Percent survival
110
100
90
80
70
60
50
0
5
10
15
20
Time after therapy, years
Poor prognosis group
(middle bottom 14%)
5-year survival: 70%
10-year survival: 53%
Good prognosis group
(middle top 53%)
5-year survival: 88%
10-year survival: 75%
Logrank test: p = 0.0026
Hazard ratio: 2.389
95% CI: 1.477 to 6.309

E

Survival of breast cancer patients classified based on relative values of the 14-gene signature



Worst prognosis (15%)
Median survival: 7.7
5-year survival: 58%
10-year survival: 37%

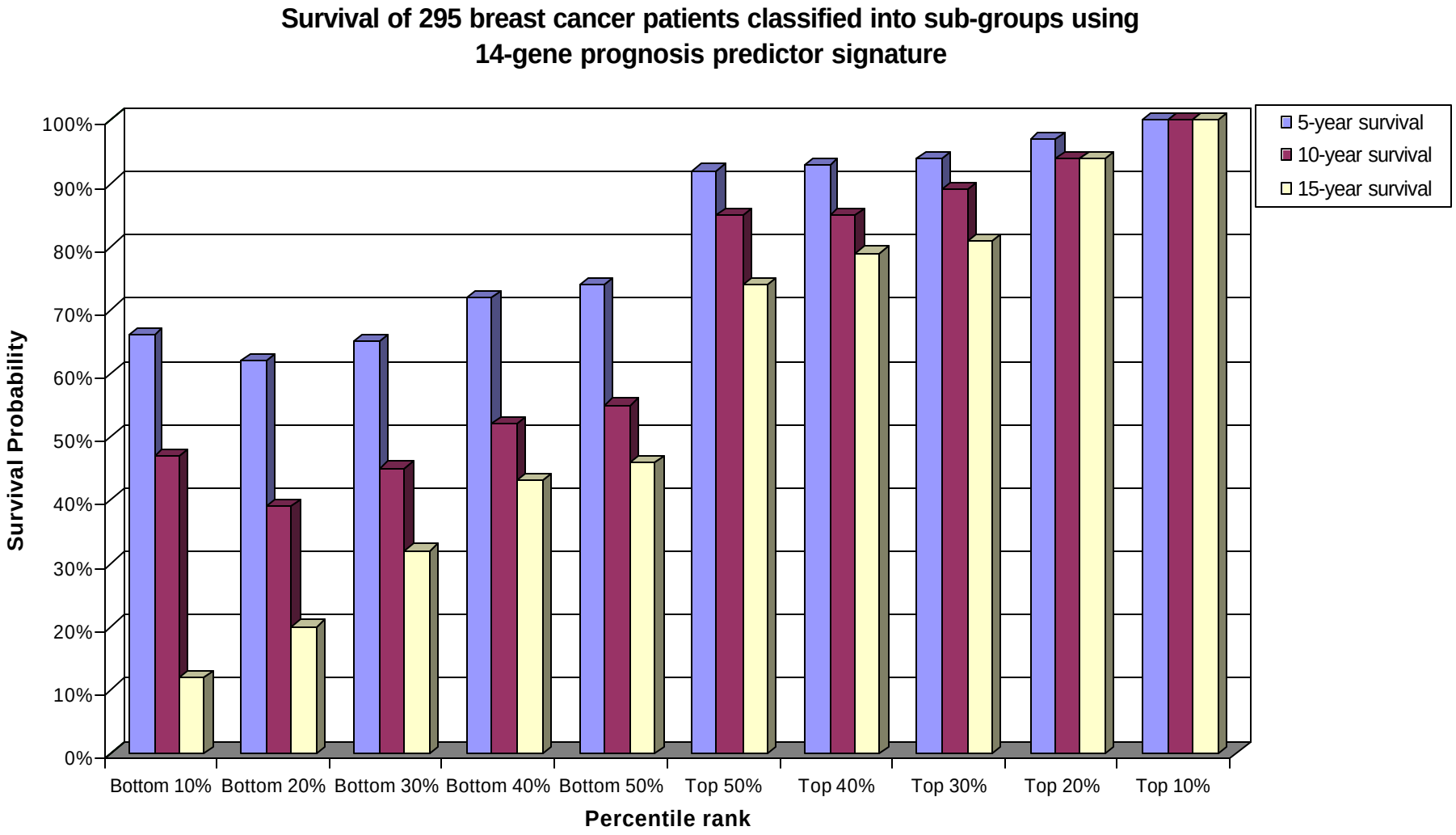
Poor prognosis (14%)
5-year survival: 70%
10-year survival: 53%

Good prognosis (53%)
5-year survival: 88%
10-year survival: 75%

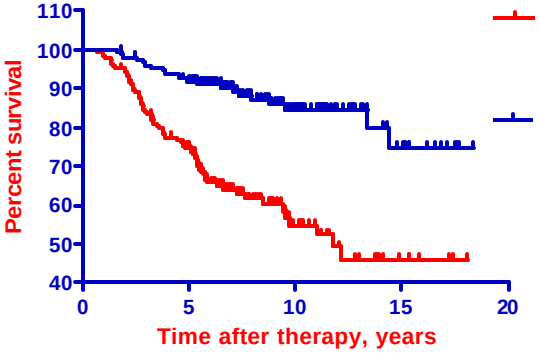
Best prognosis (18%)
5-year survival: 98%
10-year survival: 98%

Logrank test: $p < 0.0001$
Logrank test for trend: $p < 0.0001$

F


Survival of 295 breast cancer patients classified into sub-groups using 14-gene prognosis predictor signature
Survival Probability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
5-year survival
10-year survival
15-year survival
Bottom 10%
Bottom 20%
Bottom 30%
Bottom 40%
Bottom 50%
Top 50%
Top 40%
Top 30%
Top 20%
Top 10%
Percentile rank

Survival of 295 breast cancer patients with positive and negative 14-gene signature (0.00 cut off)

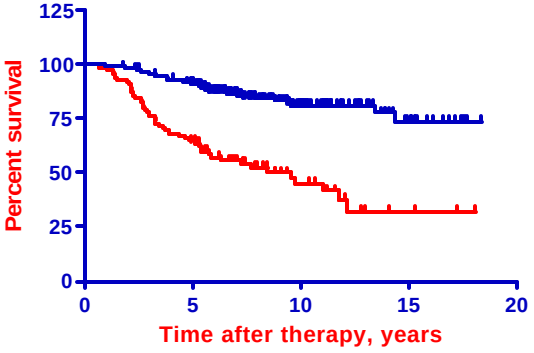


Negative 14-gene signature (49%)
Median survival: 11.74
5-year survival: 75%
10-year survival: 55%

Positive 14-gene signature (51%)
5-year survival: 92%
10-year survival: 85%

Logrank test: p < 0.0001
Hazard ratio: 3.637
95% CI: 2.217 to 5.419

Survival of 295 breast cancer patients with positive and negative 14-gene signature (-0.55 cut off)

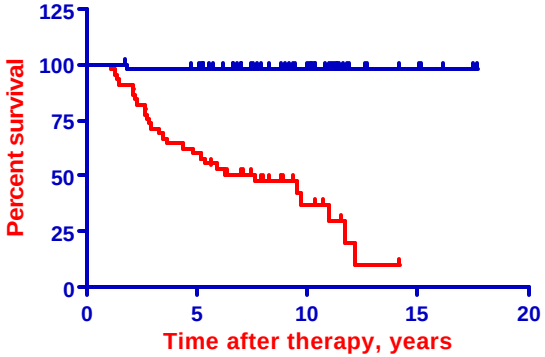


Poor prognosis group (29%)
Median survival: 9.5
5-year survival: 65%
10-year survival: 45%

Good prognosis group (71%)
5-year survival: 91%
10-year survival: 81%

Logrank test: p < 0.0001
Hazard ratio: 4.171
95% CI: 3.632 to 10.21

Survival of breast cancer patients with positive and negative 14-gene signature

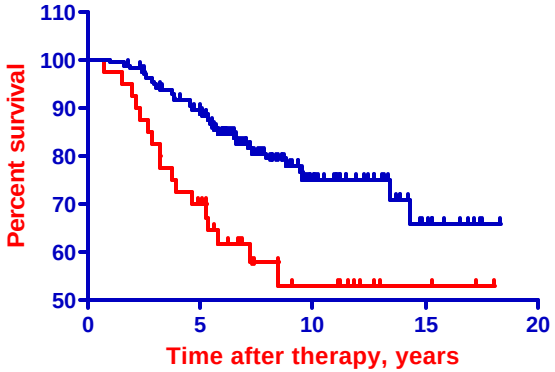


Negative 14-gene signature (bottom 15%)
Median survival: 7.7
5-year survival: 58%
10-year survival: 37%

Positive 14-gene signature (top 18%)
5-year survival: 98%
10-year survival: 98%

Logrank test: p < 0.0001
Hazard ratio: 47.48
95% CI: 6.347 to 29.17

Survival of breast cancer patients with positive and negative 14-gene signature

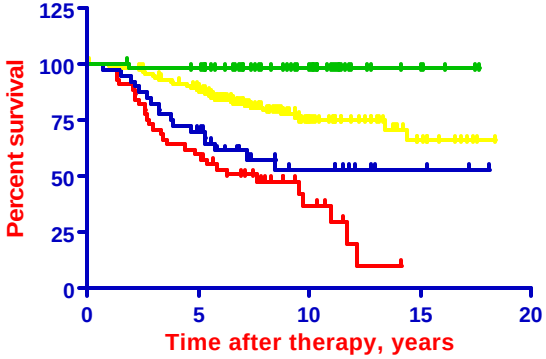


Poor prognosis group (middle bottom 14%)
5-year survival: 70%
10-year survival: 53%

Good prognosis group (middle top 53%)
5-year survival: 88%
10-year survival: 75%

Logrank test: p = 0.0026
Hazard ratio: 2.389
95% CI: 1.477 to 6.309

Survival of breast cancer patients classified based on relative values of the 14-gene signature



Worst prognosis (15%)
Median survival: 7.7
5-year survival: 58%
10-year survival: 37%

Poor prognosis (14%)
5-year survival: 70%
10-year survival: 53%

Good prognosis (53%)
5-year survival: 88%
10-year survival: 75%

Best prognosis (18%)
5-year survival: 98%
10-year survival: 98%

Logrank test: p < 0.0001
Logrank test for trend: p < 0.0001

Survival of 295 breast cancer patients classified into sub-groups using 14-gene prognosis predictor signature