

Table 2. Characteristics of patients in the learning and test subsets. No significant difference was found between the two subsets.

	Training set (N= 148)	Validation set (N=147)	All (N=295)	P-value
Overall Survival (10 years)	69.6%	70.9%	70.4%	0.96
Metastasis-free probability (10 years)	66.8%	63.3%	65.2%	0.89
T1 vs. T2	53%-47%	52%-48%	53%-47%	0.77
pN0-pN1a-pN2a/3a	51%-36%-13%	51%-36%-13%	51%-36%-13%	1
Mastectomy vs. Breast Conserving Therapy	45% - 55%	46% - 54%	45%-55%	0.96
ER+ vs. ER-	72% - 28%	81% - 19%	77%-23%	0.08
Grade I-II-III	27%-30%-43%	24%-38%-38%	24%-35%-40%	0.38
Age \geq 40	81%	76%	79%	0.31
Adjuvant chemo-hormonal therapy	38%	37%	37%	0.84
70 genes poor	62%	60%	61%	0.69
Wound signature activated	43%	42%	43%	0.85

Table 3. Multivariate analysis of prognostic gene expression signatures and clinical risk factors using a linear additive Cox proportional hazard model.

	Death		Metastasis		<i>P</i> value
	Hazard Ratio (95% CI)	<i>P</i> value	Hazard Ratio (95% CI)	<i>P</i> value	
Wound response signature*	6.17 (1.11-34.48)	0.034	3.60 (0.71-18.17)	0.11	
70-gene poor prognosis signature	4.46 (1.71-11.63)	0.002	4.53 (2.10-9.77)	<0.0001	
Molecular subtypes					
Basal	0.45 (0.047-4.20)	0.47	0.244 (0.042-1.40)	0.11	
ErbB2	0.74 (0.085-6.43)	0.78	0.532 (0.11-2.69)	0.44	
Luminal A	0.79 (0.085-7.38)	0.83	0.679 (0.13-3.53)	0.64	
Luminal B	0.59 (0.068-5.12)	0.62	0.458 (0.092-2.29)	0.33	
Indeterminate	0.51 (0.061-4.20)	0.52	0.438 (0.094-2.04)	0.28	
Age (per decade)	0.75 (0.51-1.10)	0.13	0.821 (0.57-1.18)	0.27	
Diameter of tumor (per cm)	1.03 (1.00-1.05)	0.081	1.046 (1.02-1.08)	0.001	
Lymph node status (per positive node)	1.10 (0.98-1.24)	0.11	1.148 (1.04-1.27)	0.007	
Tumor grade					
Grade 2 vs. 1	1.93 (0.62-6.08)	0.25	1.262 (0.54-2.91)	0.58	
Grade 3 vs. 1	1.70 (0.51-5.69)	0.38	0.972 (0.39-2.42)	0.95	
Vascular invasion					
1-3 vessels vs. 0 vessels	0.72 (0.26-2.00)	0.52	0.623 (0.25-1.55)	0.30	
> 3 vessels vs. 0 vessels	1.74 (1.01-2.98)	0.040	1.539 (0.93-2.56)	0.09	
Estrogen receptor status (Positive vs. negative)	1.85 (0.83-4.12)	0.12	1.400 (0.65-3.03)	0.38	
Mastectomy (vs. breast conserving therapy)	0.85 (0.51-1.41)	0.52	0.836 (0.52-1.36)	0.46	
No adjuvant chemotherapy	1.86 (0.99-3.50)	0.050	2.795 (1.53-5.11)	0.001	
No adjuvant hormonal therapy	1.25 (0.50-3.16)	0.63	1.713 (0.73-4.03)	0.21	

*Per 1.0 increment in correlation value to the serum-activated fibroblast centroid. The correlation value to the serum-activated fibroblast centroid was modeled as a continuous variable; the hazard ratio per +1.0 correlation value is reported. CI=confidence interval. The hazard ratio per +0.1 correlation value for death and metastasis are 1.20 (95% CI=1.01-1.42) and 1.14 (95% CI=0.97-1.34) respectively. Each molecular subtype was compared to all other subtypes.. Parameters found to be significant ($p < 0.05$) are shown in bold. Note that the 70-gene signature was identified based on metastasis prediction of a subset of these patients, thus its performance in this data set maybe optimistic.

Table 4. Sensitivity and specificity for predicting distant metastasis as first recurrence: comparison of gene expression profiles and clinical criteria.

	Sensitivity	Specificity	False Negative
NIH high risk	96.6%	3.9%	3.4%
St. Gallen high risk	93.2%	7.7%	6.8%
Wound response signature*	59.1%	64.3%	40.1%
70-gene signature**	85.2%	49.3%	14.8%
Wound response criterion+	90.9%	29.0%	9.1%

* Activated vs. Quiescent by hierarchical clustering.

** Good vs. Poor

+ Activated vs. Quiescent; cut off by correlation level -0.15 to the serum-activated fibroblast centroid.

Figure 4. Clinical outcomes of patients with indeterminate expression of the wound response signature (yellow bar in Fig.1) are intermediate between patients with activated and quiescent wound response signatures.

Figure 5. Wound response signature predicts decreased survival independent of tumor size or lymph node status.

Left: In tumors ≤ 20 mm (pT1) (N=155, 48 Activated, 107 Quiescent), the 10 year overall survival (OS) for the Activated vs. Quiescent groups are 62% vs. 85%, respectively (p=0.0009). Middle: in lymph node negative patients (N=151, 48 Activated, 103 Quiescent), 10 year OS for the Activated vs. Quiescent groups are 52% vs. 80% respectively (p<0.00001). Right: In lymph node positive patients, (N=144, 64 Activated vs. 80 Quiescent), 10 year OS for the Activated vs. Quiescent group are 51% vs. 90% respectively (p=0.00002).

Figure 6. Expression of the 5 molecular subtypes in early breast cancer and improved risk stratification by addition of wound response and 70-gene signatures.

(A) Correlation of gene expression pattern in 295 breast cancer samples to the centroids of the 5 molecular subtypes. The strongest positive correlation of at least ≥ 0.10 determines the subtype (1). The individual patient branches are colored according to the subclass as defined by centroid correlation. Note that the basal subtype is most clearly defined, but > 100 samples were not able to be assigned to any subtype.

(B) Tabular summary of patients in each tumor subtype with the activated wound response signature or poor prognosis 70-gene signature. Classification by the unsupervised wound response signature from Fig. 1 was applied for consistency.

(C) Improve risk stratification by integration of signatures. Patients in the ErbB2 (left) or Luminal B (right) subtypes were stratified by whether they have both the wound response and 70-gene signatures. Expression of the activated wound and poor prognosis 70-gene signatures conferred additive risk of death.

Figure 7. Nonlinear multivariate analysis of prognostic gene expression signatures and clinical risk factors in early stage breast cancer.

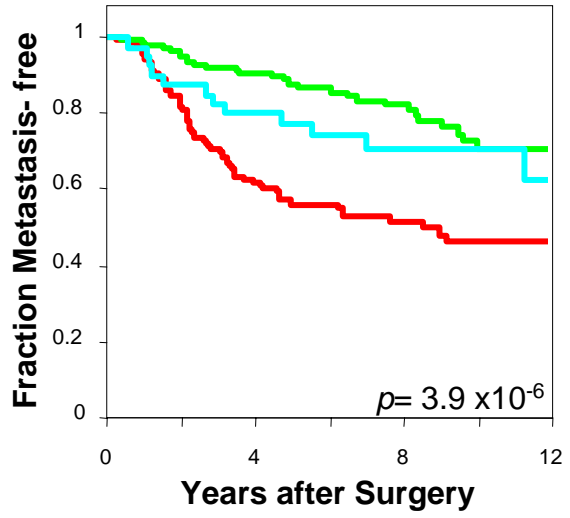
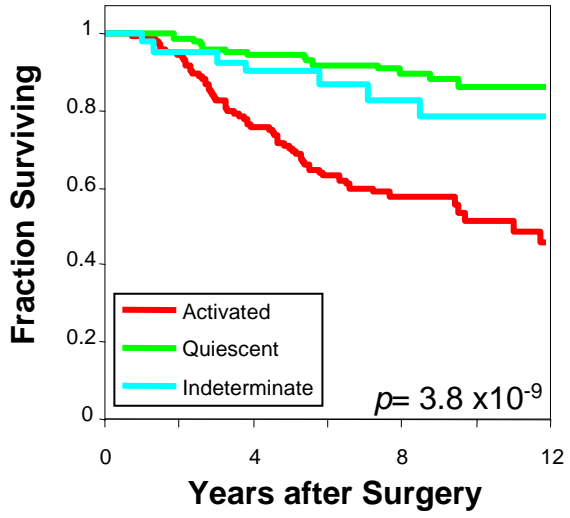
Shown are the additive contributions of the Wound signature (top row) and the 70 gene Good Prognosis signature (bottom row) to the log-relative-risk in Cox proportional hazard models, in the presence of all standard risk factors (Table 1). In the left column, the outcome is time to distant metastasis, while in the right it is patient survival time. The black curve in each case represents the contribution of the signature as a smooth function, using a basis of natural cubic splines with 4 interior knots (3). The green curves are pointwise-standard-error curves about the smooth curves. The blue lines are the result when these continuous scores are fit instead by a pair of constants, obtained by thresholding the scores at the values indicated. Because the thresholds were obtained from the decision tree analysis (Fig 4B); their mapping to the linear part of the smoothed curves indicates the congruence between the two models. The piecewise-constant fit summarizes the contribution of each of these scores, while the curves give a more detailed contribution. We note that the bends on the extreme two ends of the curves are fitted with less confidence (thus much larger confidence intervals). Although some simple tests indicate evidence for these details, a larger dataset would be required to establish them convincingly.

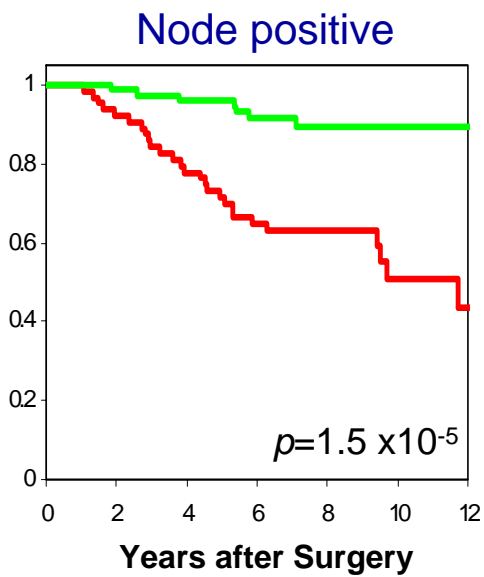
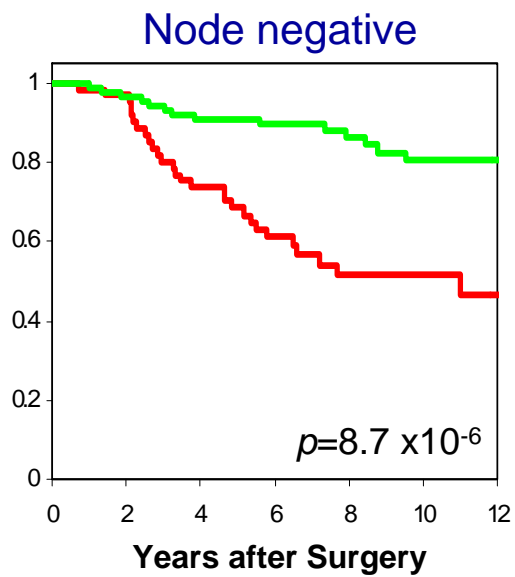
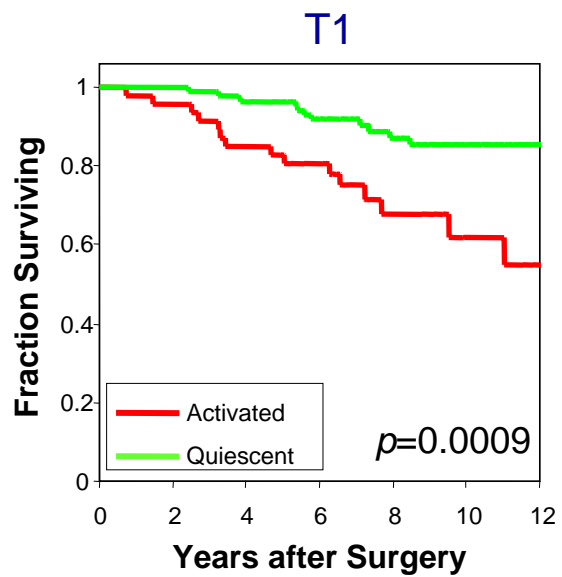
Method: Using the “survival” function in statistical package R, we compared models that augmented the baseline model of clinical risk factors (detailed in Table 1) with (i) linear terms in these correlations, (ii) non-linear terms using splines, and (iii) models using binary variables obtained by thresholding the probabilities. The non-linear spline models, using 4 degrees of freedom per signature correlation term, were significantly better than those models that augmented the baseline with linear terms ($P < 0.02$ for DMFP and OS), which were not significant for either response. These nonlinear functions suggest the use of threshold functions. For the latter we used the threshold 0.4 for the 70 gene signature (2) and for the wound signature we used a threshold of 0.05 (suggested by both the tree analysis above and the spline models). Both the spline models and the threshold models fit equally well, and we chose the threshold model in our subsequent analysis due to their simplicity. Adding the 70 gene signature threshold alone is highly significant ($P < 0.003$ for DMFP, < 0.028 for OS); subsequently adding the thresholded wound signature was also significant ($P < 0.04$ for DMFP, $P < 0.02$ for OS). Thus, despite their correlation, both signatures have significant and separate contributions to make in assessing risk. Because the two-level categorical variables was equivalent to adding the three-level tree described above in terms of the fits of the models, the data are unable to distinguish between an interaction as described by the tree, or an additive fit.

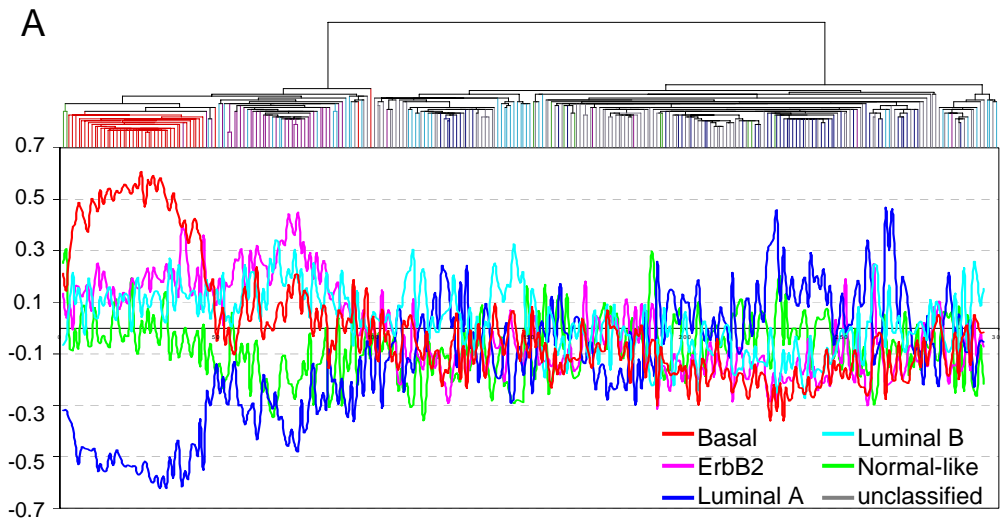
References

1. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L. & Botstein, D. (2003) *Proc Natl Acad Sci U S A* 100, 8418-8423.

2. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. & Friend, S. H. (2002) *Nature* 415, 530-6.
3. Hastie, T. and R. Tibshirani, (1990) *Generalized Additive Models*. Chapman and Hall: London.

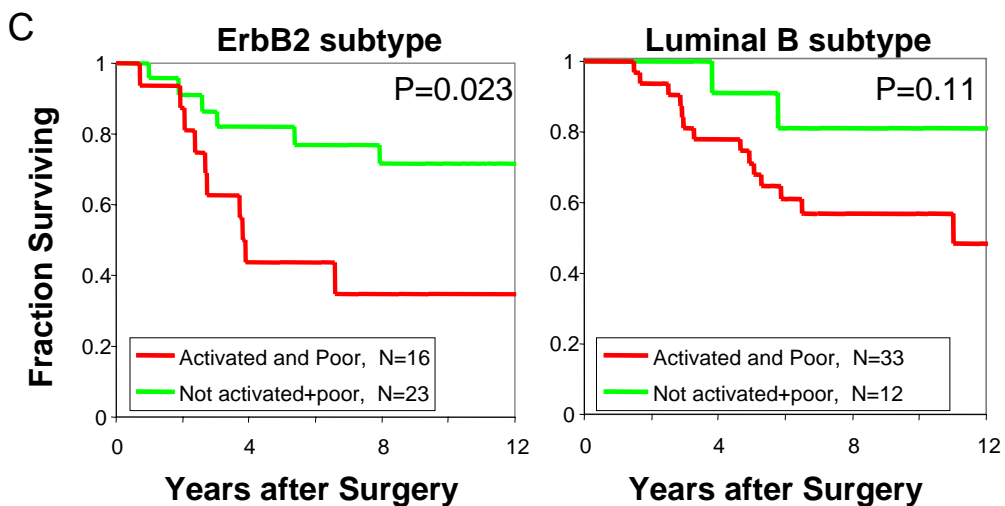




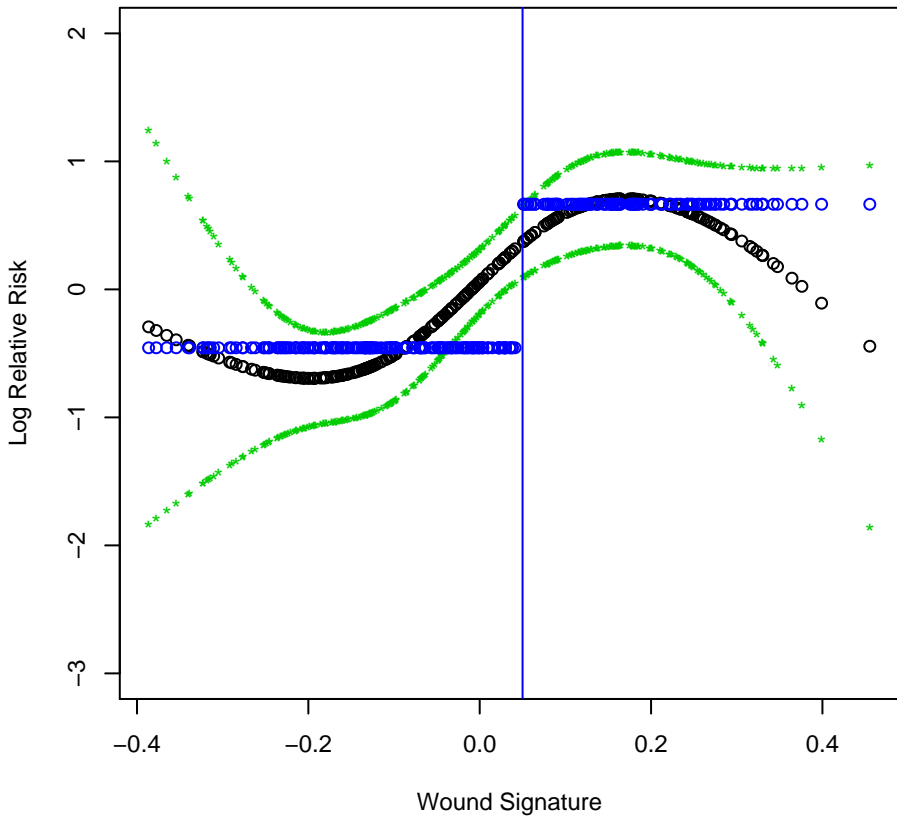


B

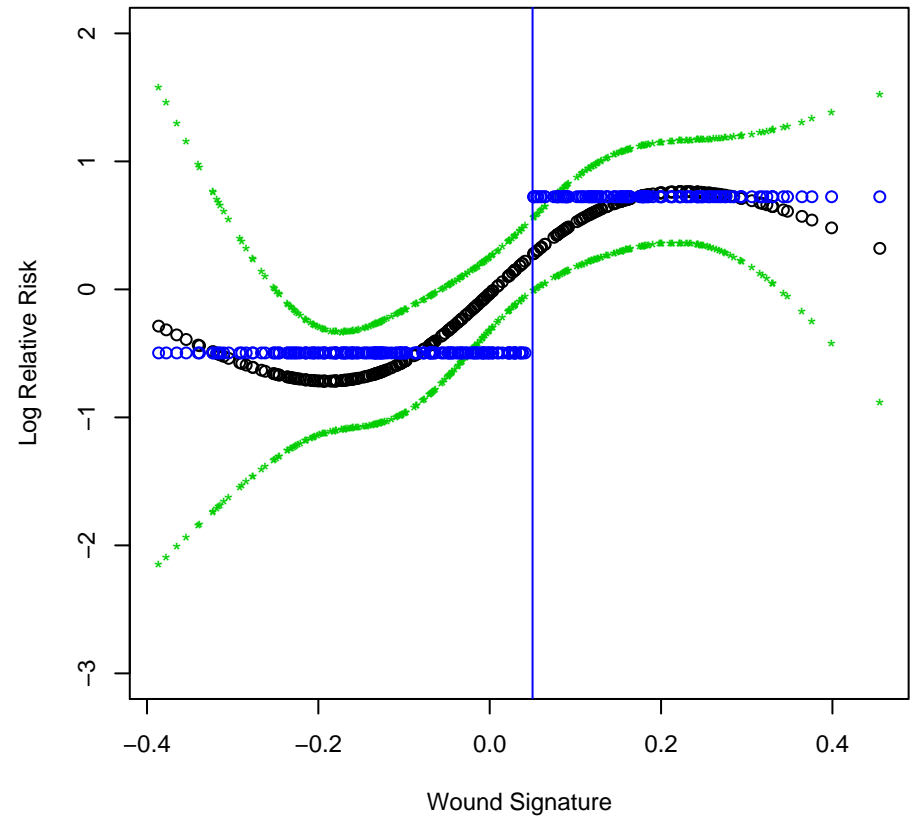
	<i>N</i>	Wound Signature "Activated"	70 Gene Signature "Poor"
Basal	45	42	44
ErbB2	39	18	32
Luminal A	47	10	11
Luminal B	45	33	42
Normal-like	10	0	3
Unclassified	109	23	48



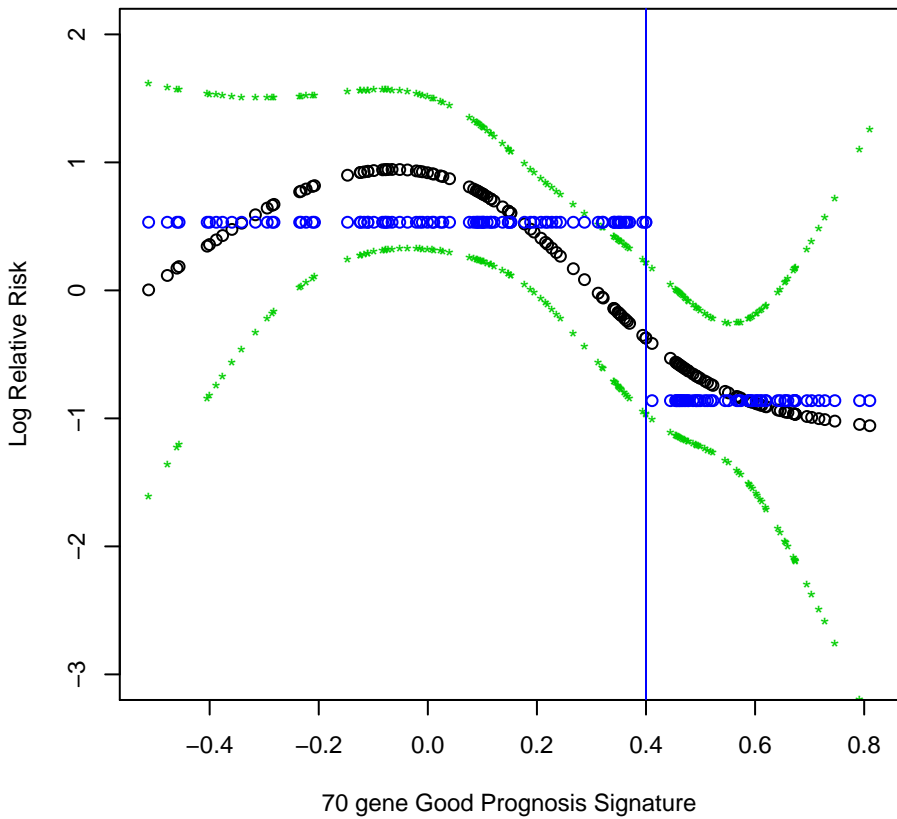
Cox Model: Distant Metastasis Free Probability



Cox Model: Survival



Cox Model: Distant Metastasis Free Probability



Cox Model: Survival

