Comparison of Prognostic Gene Expression Signatures for Breast Cancer

B. Haibe-Kains[1,2], C. Desmedt[2], C. Sotiriou[2], and G. Bontempi[1]

[1]Functional Genomic Unit, Institut Jules Bordet, Brussels, Belgium
[2]Machine Learning Group, Université Libre de Bruxelles, Brussels, Belgium

1. Background

DURING the last years, several groups have identified prognostic gene expression signatures [11, 12, 9, 1, 3] that consistently outperform traditional clinical parameters (e.g. tumor size, age, and histological grade) and guidelines (e.g. Adjuvant! Online, aka AOL [6]). Previous publications reported that gene signatures exhibited similar classification and performance for survival prediction [4, 10]. Although these studies yielded promising conclusions, some issues remained open:

- The dataset which was considered for these studies was the one employed also for the identification of some gene sets and as such it could not be considered as an independent test set.
- Since some of these signatures were developed on another platform, the initial algorithms could not be applied due to different or missing probe sequences and difference in gene expressions measurement.

Therefore, these gene signatures were never compared on an independent population of untreated breast cancer patients, where classification was computed using the original algorithms and microarray platforms.

2. Materials and Methods

WE compared three gene expression signatures, the GENE70 [11], the GENE76 [12] and the GGI [9] signatures, in terms of predicting time to distant metastasis (TDM), distant metastasis free survival (DMFS), and overall survival (OS) for the individual patient. To this end, we used the previously published TRANSBIG independent validation series of node-negative untreated primary breast cancer patients [1, 3].

We used Cramer's V statistic [2] to quantify the strength of the association between two gene signature classifications, i.e. the low- and high-risk groups of patients. We estimated hazard ratios between the low- and high-risk groups using the Cox's proportional hazards model. We used the concordance index [5] to quantify the predictive ability of a survival model. Standard errors, confidence intervals and p-values for the concordance index were computed making an assumption of asymptotic normality [7].

The statistical performance comparison between the different gene signatures in terms of hazard ratios and concordance indices were performed using a Student t test for two dependent.

3. Results

THE classification of the tumor samples according to the prognostic signatures is illustrated in Figure 1. We observed agreement in prediction in 137/198 patients (69%) when considering the three signatures. When comparing the signatures two by two, agreement in prediction was higher with the GENE70 and the GGI signatures, i.e. 89% (Cramer's V of 0.75), compared to 71% (Cramer's V of 0.33) and 78% (Cramer's V of 0.49) for the GENE76 and the GENE70 signatures, and the GENE76 and the GGI signatures, respectively.

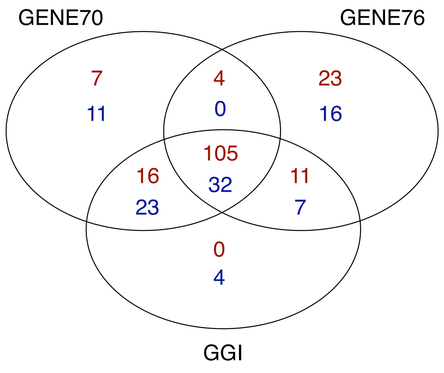

Figure 1: *Venn diagram illustrating the classification of the tumor sample according to the prognostic signatures. The high-risk patients are in red and the low-riks patients are in blue.*

The three signatures had similar capabilities of predicting TDM, DMFS, and OS in adding significant prognostic information to that provided by the classical parameters. Figure 2 shows the concordance indices and the hazard ratios for the three gene signatures and AOL that is based on the classical parameters. The three gene signatures exhibited similar performances for survival prediction. Indeed the statistical performance comparison allowed us to show that none of the gene signatures were significantly better than another, although GENE76 tended to be worst. On the contrary, GENE70 and GGI were significantly better than AOL and the same trend was observed for GENE76.

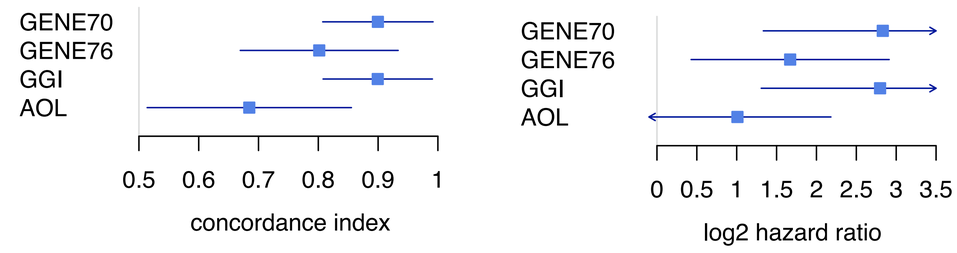

Figure 2: *Forest plots (and 95% CI) for the three gene signatures and the Adjuvant! Online classification showing the concordance indices at left, and the log2 hazard ratios at right.*

Figure 3 shows the Kaplan-Meier survival curves for DMFS for the concordant and discordant cases of all the gene signature pairwise comparisons. We can see that most of the discordant cases were patients with intermediate survival, i.e. who developed a distant metastasis after 5 years. Moreover, we observed again the high concordance between the three gene signature classifications, especially for GENE70 an GGI.

The analyses were performed for TDM and OS as well with similar results leading to the same observations (data not shown).

4. Conclusions

DESPITE the difference in development of these signatures and the small overlap in gene identity, they showed similar prognostic performance, confirming that these prognostic signatures are of wide clinical relevance. Moreover, the nature of the GGI, i.e. a signed average of expressions of proliferation genes, suggested that proliferation might be the common driving force of several prognostic signatures.

Until now, the generation of the prognostic signatures has been done on global sets of breast cancer patients. However, since it is clear that breast cancer is a molecular heterogeneous disease, with subgroups defined primarily by the estrogen (ER) and HER2 receptors, prognosis could be refined to these molecularly homogeneous subgroups of patients. We showed for example in our meta-analysis that proliferation is the strongest parameter predicting clinical outcome in the ER+/HER2- subgroup of patients only, whereas immune response and tumor invasion appear to be the main biological processes associated with prognosis in the ER-/HER2- and HER2+ subgroups respectively [8]. A new gene classifier taking into account these different subgroups of breast cancer will be the subject of our future research.

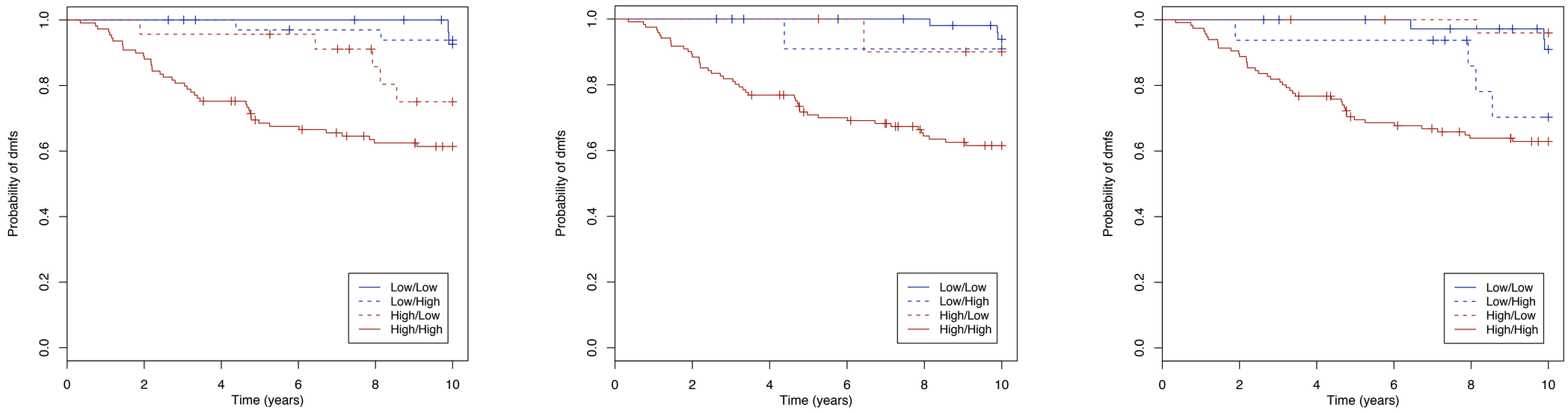

Figure 3: *Kaplan-Meier survival curves for distant metastasis free survival for: GENE70 vs GENE76 at left, GENE70 vs GGI at the middle, and GENE76 vs GGI at right.*

References

[1] M. Buyse, S. Loi, L. van't Veer, et al. Validation and clinical utility of a 70-gene prognostic signature for patients with node-negative breast cancer. *Journal of National Cancer Institute*, 98:1183–1192, 2006.

[2] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, 1999.

[3] C. Desmedt, F. Piette, S. Loi, et al. Strong time-dependency of the 76-gene prognostic signature for node-negative breast cancer patients in the transbig multi-centre independent validation series. *Clinical Cancer Research*, 13(11):3207–3214, 2007.

[4] C. Fan, D. S. Oh, L. Wessels, et al. Concordance among gene-expression–based predictors for breast cancer. *New England Journal of Medicine*, 355:560–569, 2006.

[5] F. E. Harrell Jr, K. L. Lee, and D. B. Mark. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15:361–387, 1996.

[6] I. A. Olivotto, C. D. Bajdik, P. M. Ravdin, et al. Population-based validation of the prognostic model adjuvant! for early breast cancer. *Journal of Clinical Oncology*, 23(12):2716–2725, 2005.

[7] M. J. Pencina and R. B. D'Agostinno. Overall C as a measure of discrimination in survival analysis: model specic population value and condence interval estimation. *Statistics in Medicine*, 23:2109–2123, 2004.

[8] C. Sotiriou, C. Desmedt, B. Haibe-Kains, et al. Biological mechanisms that trigger breast cancer (bc) tumor progresion are molecular subtype dependent. In *American Society of Clinical Oncology*, 10581. 2007.

[9] C. Sotiriou, P. Wirapati, S. M. Loi, et al. Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *Journal of National Cancer Institute*, 98:262–272, 2006.

[10] M. Thomassen, Q. Tan, F. Eiriksdottir, et al. Comparison of gene sets for expression profiling: Prediction of metastasis from low-malignant breast cancer. *Clinical Cancer Research*, 13(18):5355–5360, 2007.

[11] L. J. van't Veer, H. Dai, M. J. van de Vijver, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–536, 2002.

[12] Y. Wang, J. G. Klijn, Y. Zhang, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, 365:671–679, 2005.