Integrating the Molecular Subtypes of Breast Cancer into a Novel Prognostic Model

Haibe-Kains B[1,2,*], Desmedt C[1,*], Rothé F[1], Piccart MJ[1], Bontempi G[2,#], and Sotiriou C[1,#]

[1]Functional Genomics Unit, Institut Jules Bordet, Brussels, Belgium
[2]Machine Learning Group, Université Libre de Bruxelles, Brussels, Belgium
[*]Co-first authors, [#]Co-last authors.

1. Background

THE early gene expression studies in breast cancer have provided a molecular classification of these tumors into at least three clinically relevant subtypes (ER-/HER2-, HER2+ or ER+/HER2). Our group recently introduced a robust method for subtype identification, exhibiting numerous advantages compared to the hierarchical clustering used in these initial publications [7, 1].
From a prognostic point of view, several signatures have been identified in the global population of patients. However, the majority of these only yielded good prognostic performance in the ER+/HER2- subtype [7, 1]. Additionally, we showed that clinical and genomic prognostic factors dramatically depend on the molecular subtype of breast tumors.

In this work, we propose a novel prognostic model which takes into the account the molecular heterogeneity of breast cancer.

2. Methods

WE developed a two-step prognostic model called **GENIUS** (**G**ene **E**xpression prog**N**ostic **I**ndex **U**sing **S**ubtypes) as depicted in Figure 1:

1. Accurate assessment of the probabilities for a tumor to belong to each of the three breast cancer molecular subtypes (*subtype clustering*).
2. Combination of these probabilities with subtype-specific prognostic signatures (*subtype prognostic gene signatures*), which results in the final GENIUS risk predictions.

Subtype Clustering

Model-based clustering (mixture of Gaussians) in a two-dimensional space defined by the ESR1 and ERBB2 module scores representing the ER and HER2 phenotypes respectively [1]. Once fitted to the training set, this clustering model returns a set of probabilities of a tumor belonging to each *subtype*, denoted by ρ_j where $j \in \{1{=}\text{ER-/HER2-}, 2{=}\text{HER2+}, 3{=}\text{ER+/HER2-}\}$ are the subtypes.

Subtype Prognostic Gene Signatures

Stability-based feature ranking method [2] using a weighted version of the concordance index [3] as scoring function:

$$C_{wted}(x_i, y, \rho_j) = \frac{\sum_{k,l \in \Omega} w_{kl} 1\{x_{ki} > x_{li}\}}{\sum_{k,l \in \Omega} w_{kl}}$$

where x_i are the expressions of gene i, y are the survival data and $w_{kl} = \rho_j(x_k)\rho_j(x_l)$ is the weight for the pair of *comparable* patients $\{k, l\}$ with $\rho_j(x)$ being the probability for a patient's tumor x to belong to the subtype j.

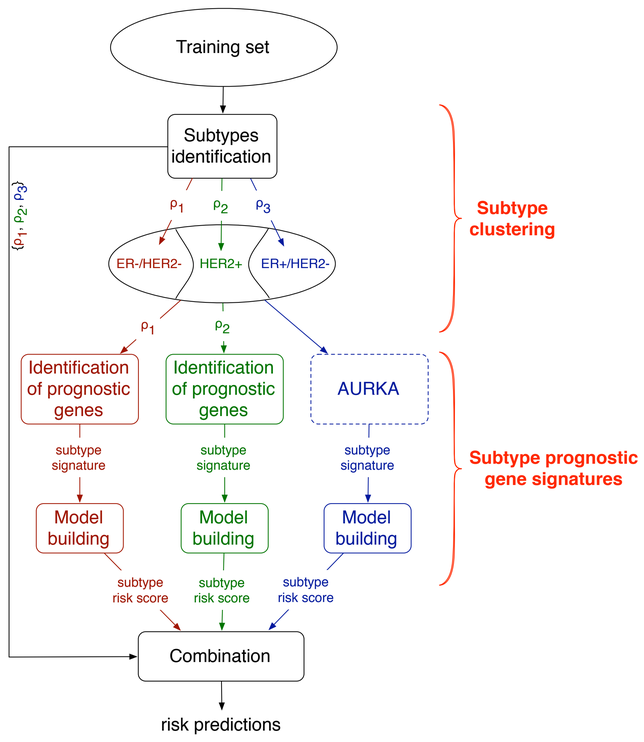

Figure 1: *Design of GENIUS prognostic model.*

The subtype prognostic gene signatures were therefore used to build the subtype risk prediction models such that, for a subtype j, the *subtype risk score*, denoted by r_j, was defined as the weighted combination of all the gene expressions in the corresponding signature:

$$r_j = \frac{\sum_{i \in Q} w_i x_i}{n_Q}$$

where Q is the set of genes in the prognostic signature specific to subtype j, n_Q the number of genes in Q and w_i is either $+1$ or -1 depending on the concordance index of gene i.

The final risk predictions were computed by combining the three subtype risk scores (Figure 1) in a local Model Network [4]:

$$r = \sum_j \rho_j r_j$$

3. Results

We focussed our analysis on untreated node-negative patients in order to build a prognostic model for early stage breast cancer and to avoid any confounding factors due to the treatment effects on survival (untreated). We used VDX [6, 5] as training set since this population contained the largest set of untreated patients with early breast cancer.

Subtype Clustering

Once subtype clustering was fitted on the training set (see Figures 2, we were able to robustly identify three subtypes: ER-/HER2- (99), HER2+ (54) and ER+/HER2- (191) (Figure 3). We validated the robustness of this clustering model in 21 breast cancer microarray datasets (data not shown).

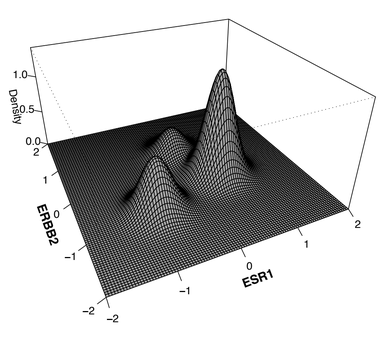

Figure 2: *The subtype clustering model is composed of a mixture of three Gaussians.*

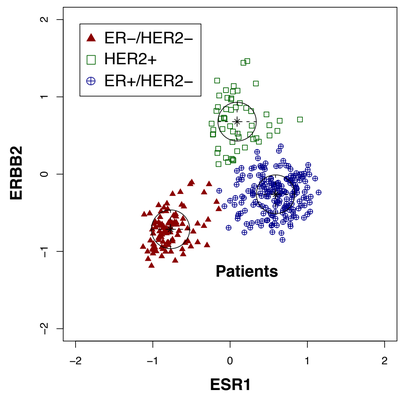

Figure 3: *Identification of breast cancer molecular subtypes in the training set (VDX).*

Subtype Prognostic Gene Signatures

Since we showed in [7, 1] that most of current gene signatures are only prognostic in the ER+/HER2- subtype and that proliferation-related genes are their driving force, we did not generate a new prognostic signature for this subtype but considered instead the proliferation module (AURKA) introduced in [1], as subtype signature.

In contrast, very few prognostic signatures have been reported thus far in the ER-/HER2- and HER2+ subtypes. Therefore, we identified two stable signatures composed of 63 and 22 genes for the ER-/HER2- and HER2+ subtypes respectively.

Using these subtype prognostic gene signatures, we computed the GENIUS risk predictions for a set of 745 untreated node-negative patients retrieved from 5 public datasets.

Performance Assessment and Comparison

As sketched in Figure 5, GENIUS risk score predictions significntly outperformed current gene signatures in the global population of patients (C-index of 0.71, superiority test p-values < 0.05) and was very competitive in each subtype (C-indices of 0.7, 0.66 and 0.66 in the ER+/HER2-, ER-/HER2- and HER2+ subtypes, respectively). GENIUS also outperformed clinical guidelines (superiority test p-values < 0.05) in most cases (Figure 4).

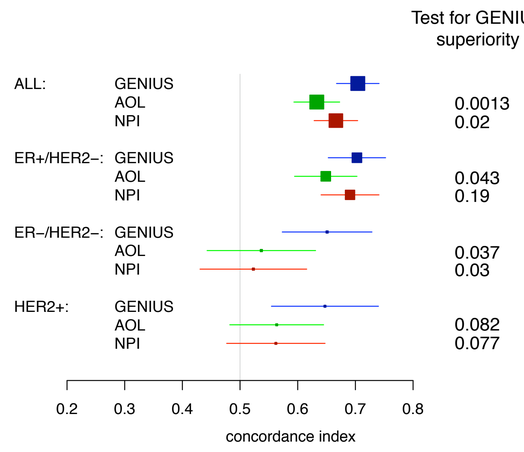

Figure 4: *Performance of risk score predictions for GENIUS and prognostic clinical models wrt the molecular subtype.*

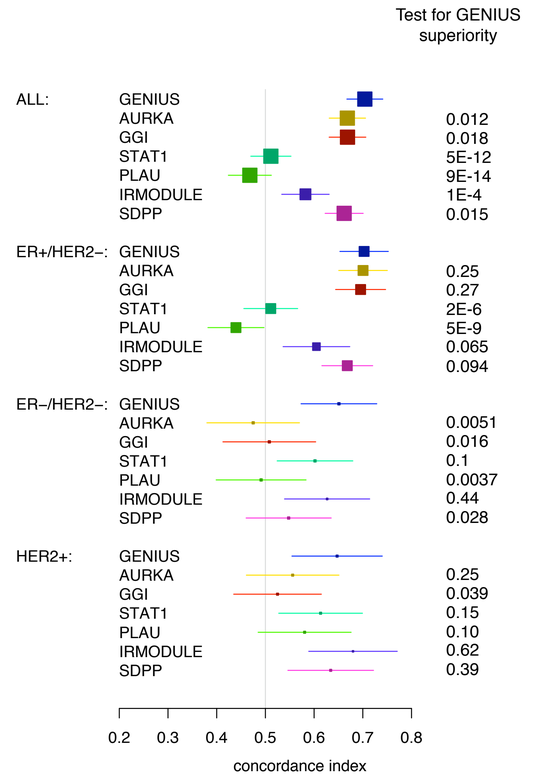

Figure 5: *Performance of risk score predictions for GENIUS and current prognostic gene signatures wrt the molecular subtype.*

The performance of the risk group predictions computed by GENIUS is illustrated in Figure 6. We observed a significant difference between the survival curves of low- and high-risk groups for both the global population (HR: 3.7; 95%CI [2.7,5]; p=1E-16) and all the subtypes: HRs of 3.7 (95%CI [2.5,5.5]; p=1E-10), 2.7 (95%CI [1.3,5.6]; p=7E-3) and 3.9 (95%CI [1.8,8.8]; p=8E-4) in the ER+/HER2-, ER-/HER2- and HER2+ subtypes, respectively. The probability of distant metastasis or relapse free survival of the low-risk group at 5 years was estimated at 91% in the global population, and 92%, 83% and 89% in the ER+/HER2-, ER-/HER2- and HER2+ subtypes, respectively.

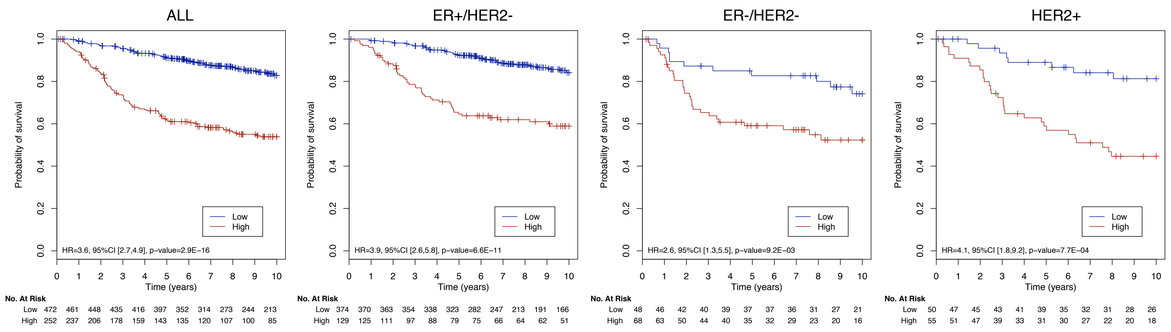

Figure 6: *Survival curves of GENIUS risk group predictions wrt the molecular subtype.*

4. Conclusions

THIS novel prognostic model, which considers the molecular heterogeneity of breast cancer, outperforms the current clinical models and gene signatures. GENIUS was the only signature to be highly prognostic in all the molecular subtypes. Additionally, the modular architecture of the model allows plugging in any other gene expression signatures, potentially enlarging its usability.

References

[1] C. Desmedt, B. Haibe-Kains, P. Wirapati, et al. Biological Processes Associated with Breast Cancer Clinical Outcome Depend on the Molecular Subtypes. *Clin Cancer Res*, 14(16):5158–5165, 2008. doi:10.1158/1078-0432.CCR-07-4756.

[2] B. Haibe-Kains. *Identification and Assessment of Dene Signatures in Human Breast Cancer*. Ph.D. thesis, Université Libre de Bruxelles, 2009.

[3] F. J. Harrell, K. Lee, and D. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 15(4):361–387, 1996. doi:10.1002/(SICI)1097-0258(19960229)15:4¡361::AID-SIM168¿3.0.CO;2-4.

[4] T. A. Johansen and B. A. Foss. Constructing NARMAX models using ARMAX models. *International Journal of Control*, 58:1125–1153, 1993.

[5] A. J. Minn, G. P. Gupta, D. Padua, et al. Lung metastasis genes couple breast tumor size and metastatic spread. *Proceedings of the National Academy of Sciences*, 104(16):6740–6745, 2007. doi:10.1073/pnas.0701138104.

[6] 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.

[7] P. Wirapati, C. Sotiriou, S. Kunkel, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Research*, 10(4):R65, 2008. ISSN 1465-5411. doi:10.1186/bcr2124.