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 (Gene Expression prognOstic Index Using Subtypes) 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 Prognostic Gene Signatures

Stability-based feature ranking method [2] using a weighted version of the concordance index [3] as scoring function: $$c_{\text{AUC}}(x,y,p_j) = \sum \frac{w_y w_x}{w_y + w_x} \left( 1 - \frac{1}{n} \sum_{i=1}^{n} I \left( x_i > x \right) \right)$$

where $$x$$ are the expressions of gene $$i$$, $$y$$ are the survival data and $$w_{x,y} = p_j / (p_j + p_i)$$ is the weight for the pair of comparable patients $$(x,y)$$ with $$p_i$$ being the probability for a patient’s tumor $$x$$ to belong to the subtype $$j$$. The subtype prognostic gene signatures were therefore used to build the subtype 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:

$$r_j = \sum \frac{w_x w_y}{w_x + w_y} \left( 1 - \frac{1}{n} \sum_{i=1}^{n} I \left( x_i > x \right) \right)$$

where $$\chi$$ is the set of genes in the prognostic signature specific to subtype $$j$$, $$n_j$$, the number of genes in $$\chi$$ and $$w_j$$ is either (1 - 0.9j) 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]:

$$r = \sum p_j r_j$$

3. Results

We focused 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 VXD [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).

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 superior to the current clinical models and gene signatures. GENIUS was the only signature to be

References


Figure 1: Design of GENIUS prognostic model.

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 (VXD).

Figure 4: Performance of risk score predictions for GENIUS and prognostic clinical models w.r.t. the molecular subtype.

Figure 5: Performance of risk score predictions for GENIUS and current prognostic gene signatures w.r.t. the molecular subtype.

Figure 6: Survival curves of GENIUS risk group predictions w.r.t. the molecular subtype.