Integrative Meta-Analysis of Gene-Expression Profiles in Breast Cancer: Toward a Unified Understanding of Breast Cancer Sub-typing and Prognosis Signatures

Christos Sotiriou[1,3], Pratyaksha Wirapati[2], Susanne Kunkel[2], Pierre Farmer[2], Sylvain Pradervand[2], Benjamin Haibe-Kains[1,3], Christine Desmedt[1], Thierry Sengstag[2], Frédéric Schütz[2], Darlene R. Goldstein[2], Mauro Delorenzi[2], Martine Piccart[1].
[1]Jules Bordet Institute, [2]Swiss Institute of Bioinformatics, [3]Université Libre de Bruxelles

Abstract

Backgrounds: Breast cancer sub-typing and prognosis have been extensively studied by gene expression profiling, resulting in disparate signatures with little overlap in their constituent genes. The biological roles of individual genes in a signature, the equivalence of several signatures and their relation to conventional prognostic factors are still unclear.

Methods: Here we undertook a comprehensive meta-analysis of publicly available gene-expression and clinical data from 18 studies totaling 2833 breast tumor samples. The concept of co-expression modules (comprehensive lists of genes with highly correlated expression) was used extensively to reveal the common thread connecting molecular sub-typing and several prognostic signatures, as well as conventional clinico-pathological prognostic factors.

Results: Breast tumors were consistently grouped into three main subtypes corresponding roughly to ER-/ERBB2- (basal), ERBB2+ and ER+ (luminal) tumors.

ERBB2+ tumors showed an intermediate estrogen receptor module score which is not obvious from the traditional ER and ERBB2 marker status combination.

Both, ER-/ERBB2- and ERBB2+ subtypes were characterized by high proliferation, whereas the ER+ subtype appeared to be more heterogeneous.

Using our meta-analytical approach we were able to identify 524 genes which were significantly associated with survival. Of the 524 prognostic genes, 65% were strongly co-expressed with proliferation, 14% with ER, 0.6% with ERBB2, 2.7% with tumor invasion, 1.5% with immune response and 16% with none of our co-expression modules.

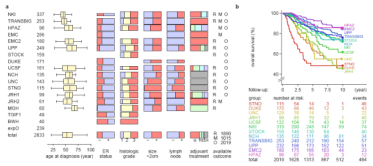
All previously reported prognostics signatures examined in this meta-analysis (N=9) , despite the disparity in their gene lists, carried similar information with regard to prognostication, with proliferation genes being the common driving force.

They were all very useful for determining the risk of recurrence in the ER+ subgroup and much less informative for ER- and ERBB2+ disease. Combining the signatures did not improve their performances.

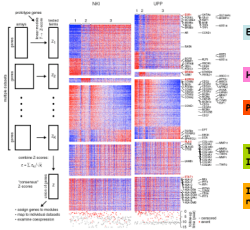
Finally, in multivariate analysis nodal status and tumor size still retained independent prognostic information.

Conclusions: This meta-analysis unifies various results of previous gene-expression studies in breast cancer. It reveals connections between traditional prognostic factors, expression-based sub-typing and prognostic signatures, highlighting the important role of proliferation in breast cancer prognosis.

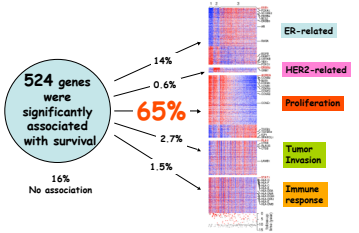
Patient demographics
Meta-Analysis, 18 studies, N≈3000 pts



Modules Representation



Relationship Between Prognostic Power of Individual Genes and Modular Associations



Dissecting Gene Espression Signatures

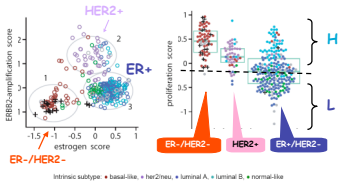
9 Prognostic Signatures

⇩

Two Partial Prognostic Signatures

Proliferation Genes | Non-Proliferation Genes

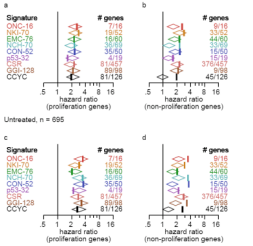
Molecular Modules Divide Breast Cancers in 4 reproducible Subtypes
(ER-/HER2-, HER2+, ER+ low, ER+ high proliferation)



a) Joint distribution between the estrogen and ERBB2-amplification scores in example datasets. Clusters are identified by Gaussian mixture models with three components. The ellipses correspond to the 95% cumulative probability around the cluster centers. The clusters are designated as tumor type: ER-/ERBB2-, HER2+ and ER-/HER2-. HER2+ tumors show intermediate estrogen scores. b) Dot histograms showing dependence of proliferation score on the subtypes. The median and quartiles for each group are shown by the box plot. ER-/ERBB2- and HER2+ tumors show high proliferation scores while ER+/HER2- tumors show a wide range of proliferation scores. The distributions of the intrinsic subtypes (colored dots), BRCA1 mutations and p53 mutations are shown in datasets where they are available.

Proliferation=driving force

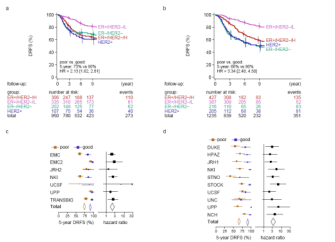
Proliferation Genes | Non-Proliferation Genes



Signature comparison. The prognostic performance of the signatures are compared by the forest plots of hazard ratio and plotted as vertical color bars for comparison. Most signatures show similar performance. Prognostic performance for DRFS of the signatures using partial signatures containing only proliferation genes in the untreated (a) and treated (c) populations. The performance of most signatures is not degraded, and even improved for p53-32. Prognostic performance for DRFS of the signatures using partial signatures containing non-proliferation genes, in the untreated (b) and treated (d) populations.

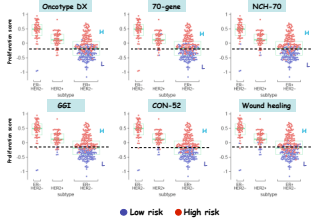
Clinical Relevance?

Untreated | Treated



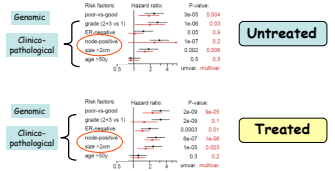
Survival analysis of groups based on module scores. Kaplan–Meier analysis for DRFS of systemically untreated (figure a) and treated (figure b) patient groups. ER- subgroup is split into ER+/HER2-/L and ER+/HER2-/H (low and high proliferation, respectively). Vertical bars on the curves are 95% confidence intervals for the Kaplan–Meier survival estimates. Forest plot representation of the 5-year survival estimates and hazard ratios for DRFS of individual datasets in the systemically untreated (figure c) and treated (figure d) populations. The length of horizontal bars and the width of the diamonds of the "Total" correspond to 95% confidence intervals. Missing bars are unavailable data.

Connexion Between Prognostic Signatures and Molecular Classification



Patient classifications made by example signatures applied to representative datasets, showing that the different signatures are essentially detecting as low risk the low-proliferation subset of ER+/ERBB2- tumors.

Clinico-Pathological Information is still needed!



Key Messages
Prognosis

- **All signatures show similar performance**
- **Proliferation is the common denominator (better quantification…!)**
- **Informative only in ER+ tumors!**
- **Still need stage information.**

Acknowledgments
FNRS, MEDIC Foundation,
Breast Cancer Research Foundation
(BCRF, Evelyne Lauder)