## Abstract

Predicting the response to the treatments is obviously an important issue in clinical oncology and in oncological research. Efficient predictors would help the clinicians in choosing the treatment the most likely to benefit each specific patient.

Furthermore they would allow focusing new researches on the specific classes of non responder patients from which could result higher healing rates, a better understanding of the mechanisms involved in the resistance to the treatments, and a better understanding of the disease itself.

Clinical trials were conducted in which each patient case was allocated her or his gene expression profile measured prior to the treatment (aprox. 20,000 expression levels per profile.)

Given a classifier model the process of predictive modeling is selecting a subset of genes (a genomic signature), fitting the classifier's parameters to the selected signatures and treatment's responses (the phenotypes), then assessing the robustness of the fitted predictor.

In some subsets of cancers the response phenotype could never be predicted with enough accuracy. Faced to this unpredictability the oncologists use to think that these cancers are << heterogeneous >> but they do not assign heterogeneity a quantitative definition. Triple negative breast cancers (TNBC) is such a subset of cancers.

We have addressed the question of a quantitative definition of the heterogeneity in TNBC expression profiles without any gene selection, we have assessed the correlation between the expression profiles' heterogeneities and the response phenotypes, and we have assessed the improvement of the predictions when the heterogeneity was taken into account by the predictive modeling.

We showed that the heterogeneity of TNBC expression profiles was systemic: no proper subset of genes was responsible of the heterogeneity and phenotype correlation. Then, to understand how the profiles' heterogeneities and response phenotypes correlation came out of the individual genes' expressions, we gathered the expression levels of each gene across the profiles of the dataset and we defined the heterogeneity that was local to each single gene. These local heterogeneities too correlated the response phenotypes in TNBC.

Importantly, the information that was carried out by the gene expressions was not the information carried out by the heterogeneity of the gene expressions: the genes whose distributions of expressions were highly different in the two phenotypes were not those whose distributions of local heterogeneities were highly different (and conversely.)

In TNBC no predictive modeling relying on the gene expressions ever brought predictors accurate enough to be used clinical routine. Can one expect more successful outcomes from predictive modelings relying on the heterogeneities of the gene expressions?