Published in final edited form as:

Clin Cancer Res. 2013 September 01; 19(17): 4685–4696. doi:10.1158/1078-0432.CCR-12-2549.

Uncovering the molecular secrets of Inflammatory Breast Cancer biology: An integrated analysis of three distinct Affymetrix gene expression data sets

Steven J. Van Laere^{#1,2}, Naoto T. Ueno^{#3}, Pascal Finetti^{#4}, Peter Vermeulen¹, Anthony Lucci³, Fredika M. Robertson³, Melike Marsan^{1,2}, Takayuki Iwamoto³, Savitri Krishnamurthy³, Peter van Dam¹, Wendy A. Woodward³, Patrice Viens⁴, Massimo Cristofanilli⁵, Daniel Birnbaum⁴, Luc Dirix¹, James M. Reuben³, and François Bertucci^{#4}

¹Translational Cancer Research Unit Antwerp, Oncology Center, General Hospital Sint-Augustinus, Wilrijk, Belgium

²Department Oncology, KU Leuven, Leuven, Belgium

³Morgan Welch Inflammatory Breast Cancer Program and Clinic, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

⁴Département d'Oncologie Moléculaire, Centre de Recherche en Cancérologie de Marseille, UMR891 Inserm, Institut Paoli-Calmettes (IPC), Marseille, France

⁵Department of Medical Oncology, G. Morris Dorrance Jr. Endowed Chair in Medical Oncology, Fox Chase Cancer Center, Philadelphia, PA, USA

[#] These authors contributed equally to this work.

Abstract

Background—Inflammatory breast cancer (IBC) is a poorly characterized form of breast cancer. So far, the results of expression profiling in IBC are inconclusive due to various reasons including limited sample size. Here we present the integration of 3 Affymetrix gene expression data sets collected through the World IBC Consortium, which allows us to interrogate the molecular profile of IBC using the largest series of IBC samples ever reported.

Materials and methods—Affymetrix profiles (HGU133-series) from 137 IBC patients and 252 non-IBC patients (nIBC) were analyzed using unsupervised and supervised techniques. **All samples were classified according to the molecular subtypes using the PAM50-algorithm.** Regression models were used to delineate IBC-specific and molecular subtype-independent changes in gene expression, pathway activation and transcription factor activation.

Co-corresponding authors: **Pr Steven J. Van Laere**, Translational Cancer Research Unit, Department Medical Oncology, Sint-Augustinus, Oosterveldlaan 24, 2610 Wilrijk, Belgium., Phone : +32 3 443 36 37 - Fax : +32 3 443 30 36, Steven.VanLaere@GZA.be. **Pr François Bertucci**, Institut Paoli-Calmettes, Department of Medical Oncology, 232, bd Ste-Marguerite, 13273 Marseille Cedex 09, France, Phone: 33 4 91 22 35 37 – Fax: 33 4 91 22 36 70, bertuccif@marseille.fnclcc.fr. **Pr. Naoto T. Ueno**, Morgan Welch Inflammatory Breast Cancer Program and Clinic, Section Chief, Section of the Translational Breast Cancer Research, Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe, Unit 1354, Houston, Texas 77030, Phone: 713-792-8754 - Fax: 713-794-4385, nueno@mdanderson.org.

Conflict of interest: None declared

Results—Four robust IBC-sample clusters were identified, associated with the different molecular subtypes ($P<0.001$), all of which were identified in IBC with a similar prevalence as in nIBC, **except for the Luminal A subtype (19% vs. 42%; $P<0.001$) and the ErbB2+ subtype (22% vs. 9%; $P<0.001$)**. Supervised analysis identified and validated an IBC-specific, molecular subtype-independent **79-gene signature**, which held independent prognostic value in a series of **871 nIBCs**. Functional analysis revealed attenuated TGF β signaling in IBC.

Conclusion—We show that IBC is transcriptionally heterogeneous and that all molecular subtypes described in nIBC are detectable in IBC, albeit with a different frequency. The molecular profile of IBC, bearing molecular traits of aggressive breast tumor biology, shows attenuation of TGF β signaling, potentially explaining the metastatic potential of IBC tumor cells in an unexpected manner.

Keywords

World IBC Consortium; IBC; molecular heterogeneity; TGF β signaling; EMT

Introduction

Inflammatory breast cancer (IBC) is an aggressive form of locally advanced breast cancer. At the time of diagnosis, virtually all patients have lymph node metastases, and one third of the patients have metastases in distant organs. Consequently, patient prognosis is poor (1). The rate of pathological complete response (pCR) after primary anthracycline-based chemotherapy (CT) ranges from 15 to 30%, and the 5-year survival remains around 40%. Currently, IBC remains a poorly characterized disease lacking specific molecular targets for therapy, although ErbB2 is often amplified.

The low frequency of occurrence (approximately 5% of all breast tumors) combined with the small size of diagnostic samples, in addition to the fact that patients with IBC are treated with neo-adjuvant chemotherapy, are the factors that have hampered past molecular studies, specifically those that were directed at deciphering the molecular biology of this disease by genome-wide approaches (2). The major stumbling block has been the poor statistical power when multiple genes (typically a multitude of 10,000) are tested on a small number of samples. High genes/samples ratios, even after removing noisy and non-informative data, result in dramatic downscaling of the significance level (i.e. multiple testing correction) leading to loss of statistical power (3).

Despite the issues outlined above, several studies aimed at defining the molecular signature of IBC using genome-wide gene expression profiling. Most of these studies have demonstrated that a molecular signature of IBC is definable (4–10). Nevertheless, when comparing the crude gene lists, only a limited number of genes were commonly identified across all studies (7, 9). Also, when translating the gene lists into biological processes or signal transduction pathways, there was significant ambiguity. Potential reasons for this observation have been discussed previously and include: A. distinct case definitions of IBC used across different studies, B. variability of the nIBC control group, C. differences in the characteristics of the IBC and nIBC groups between different studies, notably with respect to

the hormone receptor status, and D. inter-study technological differences, i.e. use of different platforms and subsequently different input gene lists (2).

In 2008, the World IBC Consortium (www.ibcconsortium.org) was founded with the goal of fostering collaborations between international research groups who focus on this rare but aggressive form of breast cancer. The research described in the current paper is the first project spearheaded by the World IBC Consortium and aims at redefining the molecular profile of IBC by taking into account the points of criticism raised with respect to the earlier studies described above. Here, we used a uniform case definition of IBC that has been put forward by an international expert panel (11). Using gene expression profiles from Affymetrix (HG133-series) derived from 3 different sites involved in IBC research, we were able to obtain an unprecedented number of IBC samples (N=137), which allows us to resolve both the sample size-related and the platform-related issues identified in the previous studies.

Materials and Methods

Patients and samples

Tumor samples were obtained from patients with breast adenocarcinoma treated in our three institutions: the Institut Paoli-Calmettes (IPC, Marseille, France: 71 IBC and 139 nIBC), the MD Anderson Cancer Center (MDA, Houston, TX, USA: 25 IBC and 58 nIBC), and the General Hospital Sint-Augustinus (TCRU, Antwerp, Belgium: 41 IBC and 55 nIBC). Each patient gave written informed consent and this study was approved by the institutional review boards of all 3 participating centers. The present data set includes 137 samples from patients with IBC and 252 samples from patients with nIBC. Patients with IBC were selected by strictly adhering to the consensus diagnostic criteria published by Dawood and colleagues (11). IBC cases included locally advanced (AJCC Stage III) and metastatic cases (AJCC Stage IV). nIBC cases included both early stage disease (AJCC Stages I and II) and advanced stage disease (locally advanced: AJCC Stage III, and metastatic: AJCC Stage IV). ER and ErbB2 expression were defined using probe sets 205225_at (*ESR1*) and 216836_s_at (*ERBB2*) as described previously (12). **As a control, we verified the correlation between mRNA status and immunohistochemistry (IHC)-based protein status as discrete variables. The positive or negative mRNA status was defined using a 2-component Gaussian mixture distribution model as described previously (13). Data are provided in the supplementary data file (Supplementary Figure 1 and Supplementary Table 1). Tumor grade was determined using the genomic grade index (GGI) (14). All patients were treated using a multidisciplinary approach according to standard guidelines. Clinicopathological data of our series are shown in Table 1 and are consistent with literature.**

Data processing and normalization

RNA extraction from the 389 samples and hybridization onto Affymetrix GeneChips (HG133-series) was performed as described before (6, 8, 9, 15). Expression data were normalized by GCRMA (16). For each of the 3 data sets separately, probe sets with expression values above $\log_2(100)$ in at least 1% of the arrays were filtered in. Next, the list

of common informative probe sets (N=9,926) was identified. This list was used to merge the distinct data sets. Therefore, we performed regression normalization, using the Limma-package in BioConductor, to remove technical, lab-specific, variation in gene expression between the distinct data sets. A principal component analysis (PCA) was done on the merged data set prior and after the regression normalization to verify the accuracy of the regression normalization in removing the lab-specific variation in gene expression.

Molecular subtypes

We classified each sample in the merged data set according to the molecular subtypes (Luminal A, Luminal B, Basal-like, ErbB2+ and Normal-like) using the **PAM50**-methodology described by **Parker** and colleagues (17). In addition, samples predicted to belong to the Claudin-low subtype by the nine-cell line Claudin-low predictor (18) were considered Claudin-low. To further investigate the effect of the regression normalization on the biological variation in gene expression, the molecular subtype classification was compared with the expression of an ER activation signature (19).

Unsupervised analysis

To investigate common themes in gene expression within each tumor phenotype (IBC or nIBC) separately as well as for both tumor phenotypes combined, we adopted the following strategy. We randomly selected 100 samples from the 3 groups of interest (Group 1: IBC, Group 2: nIBC, and Group 3: combined IBC/nIBC). On this selection, unsupervised hierarchical cluster analysis (UHCA) using Manhattan distance as distance measure and Ward linkage as the dendrogram drawing method was performed only for those genes having a standard deviation greater than 2. UHCA was followed by a cluster robustness analysis using the silhouette algorithm. The silhouette algorithm divides the dendrogram into an increasing number of clusters. We chose to analyze a maximum number of 20 clusters. For each increase in the number of clusters, the algorithm generates a score, the magnitude of which is proportional to the robustness of the identified clusters. This analysis was repeated 10 times, resulting in a series of 10 silhouette scores for each number of clusters. The number of clusters in each group of interest was then determined by sequentially comparing the silhouette scores (silhouette scores for 2 clusters vs. silhouette scores for 3 clusters, silhouette scores for 3 clusters vs. silhouette scores for 4 clusters, and so on) and equals the highest number of clusters for which the silhouette score significantly drops when increasing the number of clusters with one unit.

Supervised analyses

Supervised analysis, identifying differentially expressed genes between IBC and nIBC, was performed to evaluate the influence of alternative stage-matching approaches on the identifiable biological differences between IBC and nIBC. Therefore, global differences in gene expression between IBC and 3 alternatively composed nIBC control groups (Stage I-IV, Stage I-II and Stage III-IV) were investigated using the global test (20). In addition, for each comparison, we sought to identify lists of differentially expressed genes using Significance Analysis of Microarrays (SAM) (21). The resulting gene lists were compared.

To further explore functional differences between IBC and nIBC, we repeated supervised analyses at the pathway level. We applied 18 pathway GES originally described by Gatz and colleagues using the procedures outlined in the original publication (19) and one additional GES for VEGF-activation (22). In addition, for each of the previously defined comparisons, we explored differences in patterns of transcription factor activation. For that purpose, we designed an algorithm to quantify the extent of transcription factor activation in samples based on the gene expression data. Details are provided in the supplementary data file and Supplementary Table 2.

Molecular subtype-independent supervised analysis

Due to the association of the tumor phenotype (IBC/nIBC) with the molecular subtypes, the list of probe sets resulting from the IBC/nIBC comparison probably contains many genes related to molecular subtype, and therefore we hypothesize that this gene list is not IBC-specific. In order to dissect IBC-specific variations in gene expression from molecular subtype-specific variations in gene expression, supervised analysis (IBC vs. nIBC) was repeated by applying linear regression modeling (Limma-package). Prior to analysis, we divided the data set into a training set of 250 samples (84 IBC samples and 166 nIBC samples) and an independent validation set of 139 samples (53 IBC samples and 86 nIBC samples). **The distributions of the molecular subtypes, the tumor stage and the institution from which the samples were derived (IPC, MDA, TCRU) were compared between both data sets. The training data set was analyzed using linear regression models incorporating the molecular subtype classification and the IBC/nIBC classification to identify probe sets with IBC-specific expression components. Using the global test (20), lists of probe sets were analyzed for global expression differences between IBC and nIBC in the validation set. Classifier models based on the identified lists of probe sets with an IBC-specific expression component were constructed using the nearest shrunken centroid algorithm implemented in the PAMR-package. For this purpose, 10-fold cross-validation on the training set was performed to select an appropriate δ -value minimizing the cross-validated training error rate. The resulting models were applied on the validation set to estimate accuracy, sensitivity and specificity. In addition, we tested the performance of the classifier models to discriminate between sample from patients with and without IBC in groups stratified by the molecular subtypes. Finally, the strategy of dissecting IBC-specific effects from molecular subtype-specific effects using linear regression models was repeated on the entire data set for the pathway and transcription factor activation data.**

Analysis of publicly available nIBC data

Given the higher metastatic risk of IBC compared to nIBC, we tested the hypothesis that our 79-gene IBC/nIBC signature, if biologically relevant, is associated with metastatic relapse in nIBC. Therefore, we collected expression data from 6 publicly available series of adjuvant chemotherapy- and hormone therapy-naïve lymph node-negative early breast cancer patients. These series have been used for similar purposes in earlier studies and follow-up data in terms of metastatic relapse are available. An overview of included data sets is provided in Supplementary Table 3. Redundancy related to patient samples included in multiple data sets was eliminated, yielding 871

samples available for analysis. Before analysis, we mapped hybridization probes for the differentially expressed genes across the two oligonucleotide-based platforms used across the series. When multiple probes were mapped to the same GeneID (EntrezGene identification number), the one with the highest variance in a particular dataset was selected to represent the GeneID. Analysis of each data set was done separately to guarantee a larger number of genes common with our gene list. After having identified the common genes, we used Distance Weighted Discrimination (DWD) (23) to normalize each public dataset in order to be comparable to our present pooled Affymetrix data set. The samples were classified as “IBC-like” or “nIBC-like” using a Support Vector Machine (SVM) function based on the list of **79** differentially expressed genes. The initial outcome of interest for this SVM classifier was the separation of IBC from nIBC. **The SVM model was established in our original training set (84 IBC and 166 nIBC samples) with a linear kernel as the parameter.**

Statistical analyses

In the results section, only nominal P-values are reported; however in case of multiple comparisons, P-values were corrected and were considered significant only if the false discovery rate (FDR) was smaller than 0.05 (9). To compare the distribution according to categorical variables, Chi-square testing or the Fisher Exact test were used when appropriate. To compare continuous data between two or more groups, respectively the Mann-Whitney U-test and the Kruskal-Wallis test were used. To compare continuous data with one reference value, the Wilcoxon signed rank test was used. To compare distributions of continuous variables, Spearman correlation analysis was performed. **The distant metastasis-free survival (DMFS) for nIBC patients included in the public data set was calculated from the date of diagnosis until the date of first metastatic relapse.** The follow-up was calculated from the date of diagnosis to the date of last news for event-free patients. Survival was calculated using the Kaplan-Meier method and compared between groups with the log-rank test. Univariate and multivariate analyses were done using generalized linear models or Cox’s proportional hazards in case of survival analysis.

Results

Unsupervised analysis and heterogeneity of IBC

First, we assessed the effect of the normalization strategy on the technical and biological variation in gene expression. Results of this analysis are presented in Supplementary Figure 2. PCA demonstrates that normalization was effective in removing technical, inter-laboratory variation in gene expression. After normalization, ER activation scores were significantly different between samples classified according to the various molecular subtypes ($P < 0.001$), suggesting that the normalization procedure did not remove biological differences.

Second, using UHCA followed by silhouette score analysis, we identified 4 robust clusters in our series of 137 IBC samples (Figure 1A). The details of the silhouette analysis are provided in supplementary data file (Supplementary Figures 3–4). A significant association was observed between these sample clusters and the molecular subtypes ($P < 0.001$). In

agreement with this observation, ER-, PR-, and ERBB2-pathway activation scores were also significantly different between the different clusters ($P<0.001$). When a similar analysis was performed on the nIBC series ($N=252$), 6 sample clusters, again related to the molecular subtypes ($P<0.001$), were identified (Figure 1B). Arguably, the increase in number of clusters observed in nIBC can be explained by the increase in sample size (137 vs. 252). Therefore, we analyzed 10 sets of 137 randomly selected nIBC samples. The median number of clusters in these series of samples equaled 6 (range: 4–10), which was significantly different from the number of sample clusters observed in the series of IBC samples ($P=0.002$), suggesting that IBC is indeed less heterogeneous than nIBC.

When the above outlined analysis was repeated on the combined data set of IBC and nIBC samples, seven robust sample clusters were identified (Figure 1C). Downstream of the first bifurcation, Luminal A, **Luminal B** and Normal-like samples were segregated from Basal-like, ErbB2+ and Claudin-Low samples ($P<0.001$). In addition, a significant segregation of IBC and nIBC samples was observed ($P<0.001$). Linear regression analysis using the cluster label as the dependent variable and both the tumor phenotype and the different molecular subtypes as independent variables demonstrated that molecular subtypes, and not the tumor phenotype, are independent predictors of the cluster label (tumor phenotype: $P=0.104$; molecular subtypes: $P<0.050$). These data suggest that differences in gene expression between IBC and nIBC are dominated by the molecular subtype-related differences in gene expression.

IBC/nIBC molecular comparison and influence of the molecular subtypes

We sought to determine differences in gene expression between IBC and nIBC in a gene-by-gene supervised manner. Table 2 provides an overview of the results of this comparison. When comparing IBC to nIBC, we observed significant differences in gene expression, regardless of the type of stage-matching performed for the nIBC control group. On initial review, differences between IBC and advanced stage (III-IV) nIBC seemed smaller than between IBC and all nIBC samples, as suggested by the lower Q-value. However, when differences between IBC samples and non-stage-matched nIBC samples were analyzed in ten equally large series or randomly selected samples, no significant differences for Q-values were recorded ($P=0.345$), suggesting that the determinant responsible for this difference is sample size and is not due to the type of stage-matching (Table 2 and Figure 2A). Similar analyses were performed for the numbers of differentially expressed genes at $FDR<0.050$ and $FDR<0.100$ (Table 2 and Figures 2B-C). Lists of differentially expressed genes ($FDR<0.100$) identified in the comparisons of IBC to alternatively composed nIBC control groups (i.e. non-stage matched, advanced stage (III-IV) only, or low stage (I-II) only) were evaluated to assess the effect of the type of stage-matching on the gene list content; results are presented by a Venn diagram (Figure 2D). Of the genes identified in the smallest gene list (IBC vs. Stage III-IV nIBC), 98% and 91% of the genes were present in the gene lists comparing respectively IBC to all nIBC samples and IBC to the stage I-II nIBC samples only. When comparing IBC to either all nIBC samples or to the stage I-II nIBC samples only, an average concordance of approximately 87% was observed between both gene lists (data not shown). These results suggest that stage matching had limited influence, allowing us to retain all nIBC control samples for further analysis.

We then compared IBC and nIBC for several expression profiles, including molecular subtypes and pathway and transcription factor activation signatures. Results are shown in Supplementary Table 4. Regarding the molecular subtypes, similar frequencies for all subtypes were observed except for the Luminal A subtype (**42% in nIBC vs. 19% in IBC; $P<0.001$**) and the ErbB2+ subtype (**9% in nIBC vs. 22% in IBC; $P<0.001$**). Out of 19 tested pathways, 12 (63%) were differentially activated between IBC and nIBC: 8 pathways were more activated in IBC (CTNB, HER2, MYC, RAS, INF α , INF γ , TNF α and VEGF), whereas 4 pathways were attenuated in IBC (ER, PR, P53, and TGF β). When analyzing the transcription factor activation profiles, we identified 78 out of 234 (33%) differentially activated transcription factors, from which 38 (49%) were more activated in IBC and 40 (51%) were more activated in nIBC. Among the transcription factors hyperactivated in IBC we identified RELA, corroborating the results from previous studies (24, 25).

Molecular subtype-independent comparison of IBC and nIBC

Because of the strong correlation between the IBC/nIBC phenotype and the molecular subtypes, we hypothesized that the results presented above could be related to the molecular subtypes, and hence were not actually IBC-specific. Thus, we reiterated the IBC/nIBC comparative analysis (gene expression, pathway activation, and transcription factor activation) using linear regression models to **identify probe sets that are differentially expressed between IBC and nIBC in molecular subtype independent manner**. In order to be able to evaluate the extent by which the identified probe sets truly reflect IBC biology, we divided our series into a training and validation set. **No differences between both sets were observed with respect to the distribution of the molecular subtypes ($P=0.955$), the tumor stage ($P=0.954$) and the institution from which the samples were derived ($P=0.761$).**

In the training set, we identified 491 probe sets, representing 443 unique and IBC-specific genes, of which 231 (47%) and 260 (53%) were respectively up- or downregulated in IBC (Supplementary Table 5). For comparison, 2743 genes were differentially expressed ($FDR<0.050$) between IBC and nIBC in the training set when molecular subtype-dependent gene expression differences were not considered. Within the list of 491 probe sets, 79 probe sets (16%) were uniquely IBC-specific, whereas the remaining probe sets ($N=412$) showed additional molecular subtype-specific gene expression variation. Both the complete list of probe sets with an IBC-specific expression component ($N=491$) and the list of probe sets that are uniquely IBC-specific ($N=79$) were used for further validation. Both lists are differentially expressed between IBC and nIBC ($P<0.001$) in the independent validation set with a slightly greater expression difference for the list of uniquely IBC-specific probe sets (Observed $Q=51.677$ vs. Observed $Q=50.893$). Further details are provided in the supplementary data file (Supplementary Figure 5).

Using the nearest shrunken centroid method, classifier models were constructed on the training set and applied onto the validation data set. Cross-validated probabilities for the training set and posterior probabilities for the validation set are provided for both models in Figure 3. For the full list of probe sets with an IBC-specific expression component we observed an overall test accuracy of 68% with a sensitivity of 66% and a specificity of 70%

(OR=4.434, $P<0.001$). Using the list of probe sets with a uniquely IBC-specific expression pattern we observed an overall test accuracy of 71% with a sensitivity of 65% and a specificity of 74% (OR=5.136, $P<0.001$). We next evaluated the performance of both models to discriminate between samples from patients with and without IBC stratified per molecular subtype. For the sake of statistical power, we combined the tumor samples with Luminal A and Luminal B phenotypes (Luminal group) and the tumor samples with Basal-like and Claudin-low phenotypes (Triple Negative Breast Cancer group). The model based on the list of uniquely IBC-specific probe sets ($N=79$) performed well in discriminating between IBC and nIBC samples of the luminal subtype ($N=70$, OR=3.909, $P=0.019$), the triple negative breast cancer subtype ($N=40$, OR=4.839, $P=0.027$) and the ErbB2+ subtype ($N=22$, OR=7.066, $P=0.081$). In addition, in the Normal-like subtype we obtained 70% classification accuracy ($N=7$, OR=4.450, $P=0.486$). Evaluating the model based on the full list of probe sets ($N=491$) in the above described subgroups demonstrated poor performance, probably due to the presence of a molecular subtype-specific expression component for the majority of the involved probe sets.

To identify relevant pathways, gene functions and regulatory networks associated with IBC biology, we decided to analyze the full list of probe sets with an IBC-specific expression component. We reason that the presence of an IBC-specific expression component for each of the involved probe sets lends credit to the use of this list of probe sets as a starting point to decipher the molecular biology of IBC. Detailed results of the Ingenuity Pathway Analysis are provided in Supplementary Tables 6 and 7 and Supplementary Figures 6 and 7. We observed that inflammation- and immune-related processes, including NF κ B-signaling, characterize the IBC tumor phenotype. We identified a regulatory gene network composed of genes repressed in IBC and centered on TGF β .

Next, we focused on IBC-specific pathway and transcription factor activation patterns. When molecular subtypes were incorporated into the model, activation of the INF α -pathway and inhibition of the P53-, TGF β - and PR-pathways were associated with IBC (FDR <0.050). Several other pathways known to be associated with heterogeneity in breast cancer, including ER, HER2 and EGFR, ranked among the lesser significant pathways (FDR=0.153, 0.124 and 0.343 respectively), suggesting that the influence of the molecular subtypes on these results was removed. Upon reassessment of the transcription factor activation profiles, we identified 18 IBC-specific transcription factors, 10 of which had attenuated activity (MZFI, GC, FOXO3/STAT5, BCL3, PITX2, MSX2, TCF3/TCF4, MYB, PEG10 and ID1) and 8 of which are activated (EGR, LXR, STAT3, HNF4A, THRA, SHH, FOXM1 and MKX2-2) in IBC (*data not shown*). Inhibition of SMAD in IBC was also observed with a significant nominal P-value although the FDR exceeded our thresholds for significance (FDR=0.084, $P=0.009$, Rank=23rd/234). As expected, ER ranked among the lesser significant transcription factors (FDR=0.246, $P=0.055$, Rank=52nd/234).

Prognostic value of the IBC/nIBC signature in nIBC

Given the poor prognosis of IBC, we hypothesized that our 79-gene list, if biologically relevant with respect to the IBC/nIBC distinction, might be associated with the occurrence of metastatic relapse in nIBC in general. We thus tested its prognostic value in a series of

871 clinically annotated and chemotherapy- and hormone therapy-naïve node-negative early breast cancers. The SVM model classified 83.6% of the samples with a “nIBC-like” profile (N=728) and 16.4% with an “IBC-like” profile (N=143). We compared the DMFS of these two nIBC classes. With a median follow-up of 110 months after diagnosis (range: 1 to 286), the “nIBC-like” class had a 5-year DMFS of 75%, which was significantly longer than the 5-year DMFS of 62% observed for samples classified as “IBC-like” (P=0.002). Kaplan-Meier plots are shown in Supplementary Figure 8. In univariate analysis (Supplementary Table 8), the risk of distant metastatic recurrence in the “IBC-like” class was significantly elevated as compared to the “nIBC-like” class (HR=1.530, 95% C.I.=1.160–2.020, P=0.002). The other significant variables (P<0.050) included the histological grade, the molecular subtypes and the pathological tumor size. Patient age at diagnosis was not significant. Interestingly, our SVM model remained significant (HR=1.480, 95% C.I.=1.040–2.100, P=0.029) in multivariate analysis incorporating all variables significant in univariate analysis (pT status, histological grade, and molecular subtypes), suggesting that the 79-gene signature contains molecular subtype-independent prognostic information that might explain the worse prognosis of IBC patients.

Discussion

This whole transcriptome study of IBC, the largest one reported so far, was made possible thanks to collaboration through the World IBC Consortium. The present study addressed several criticisms raised with respect to the earlier studies: a uniform IBC definition put forward by an international expert panel (11) was used, pre-treatment samples were profiled using the same Affymetrix platform and a high number of IBC samples was included in addition to a high number of nIBC control samples encompassing different disease stages.

We have shown that IBC is transcriptionally heterogeneous, as demonstrated by the identification of 4 robust sample clusters. The identification of all existing molecular subtypes in IBC corroborates this finding. Compared to nIBC, the transcriptional heterogeneity was less extensive and the distribution pattern of the molecular subtypes between two tumor groups differed, particularly with respect to the Luminal A and ErbB2+ samples. Overall, **75%** of the IBC samples belong to the classically more aggressive subtypes, Basal-like, ErbB2+, Claudin-low or Luminal B subtypes, whereas these subtypes account for **54%** of the nIBC samples. The Luminal A subtype represents **19%** of the IBC samples, whereas in nIBC this subgroup represents **42%**. These data corroborate the results reported in previous publications where the presence of all molecular subtypes in IBC, albeit with a different frequency, was reported (9, 26, 27). The major difference resides in the fact that all subtypes, except for the Luminal A and the ErbB2+ tumor samples, exhibit a similar prevalence in IBC as in nIBC, whereas in earlier studies IBC was reported to be enriched for the Basal-like and ErbB2+ molecular subtypes (9, 27).

One of the aims of this study was to redefine the molecular profile of IBC, taking into account the points of criticism raised in earlier studies. The issues related to sample size, the definition of IBC cases and the inter-study differences in technical platform are, to the best of our efforts, resolved by the study design but the objections with respect to the composition of the control group remain. Our results clearly demonstrate that the differential

distribution of the molecular subtypes between both tumor groups needs to be considered. As shown by unsupervised hierarchical clustering, IBC and nIBC are significantly segregated downstream of the first bifurcation, suggesting that there are major differences in gene expression between the IBC and nIBC tumor phenotypes. However, multivariate regression analysis revealed that the presence of the different molecular subtypes is the main driver of the clustering pattern. Therefore, differences in gene expression between IBC and nIBC are dominated by the differential distribution pattern of the molecular subtypes, which is corroborated by the fact that only **18%** of the genes initially reported to be differentially expressed between IBC and nIBC remain differentially expressed after performing the linear regression analysis to account for the influence of the molecular subtypes. **Of note, the number of genes with a uniquely IBC-specific gene expression profile is even smaller and represents only 3% of the global expression differences.** The differential pattern of the molecular subtypes, at least in part, explains the higher proportion of samples with a molecular “poor-prognosis” profile in IBC than in nIBC when classified according to previously published prognostic gene signatures, as well as the differential activation of biological pathways (data not shown). On the other hand, the influence of the composition of the control group according to tumor stage appears to be limited.

We thus decided to compare IBC with nIBC, incorporating all stages (Stage I-IV) of breast cancer and to use regression modeling to discriminate between molecular subtype-specific effects and tumor phenotype-specific effects. Alternatively, we could have compared IBC to a non-stage-matched but molecular subtype-matched nIBC control group. However, this would lead to smaller sample sizes and, hence, considerable loss of statistical power, a disadvantage not encountered when using the regression models on the entire data set.

The identified IBC-specific molecular changes both confirm results obtained in previous studies and reveal novel findings. *MARCKS*, a gene that ranked amongst the top 10 differentially expressed genes between IBC and nIBC, is involved in regulating cell motility via its function as a regulator of the actin cytoskeleton (28). Interestingly, Van der Auwera et al reported that *MARCKS* was one of the targets of miR-30b, a PI3K-targeting microRNA specifically downregulated in IBC (29). Other examples of genes overexpressed in IBC and involved in the actin-based cell migration include *RAC1*, *RHOF*, and *FNBPI*. The regulation of cell motility and invasion was one of the most consistent molecular aspects associated with IBC across different studies (4–6, 9). Also, the role of NFκB in IBC biology reported in earlier studies (24, 25) is corroborated by the findings in the present study.

By far the most important biological observation made within the scope of this analysis relates to TGFβ signaling. Given the aggressiveness of IBC and the role of TGFβ in tumor biology (30), the observation of IBC-specific TGFβ-repression is unexpected. Careful examination of the literature provided several interesting aspects with respect to TGFβ-signaling in breast cancer. First, Giampieri *et al.* showed that TGFβ-signaling switches the activity of breast cancer cells from cohesive to single cell motility. In addition, cells restricted to collective invasion were capable of lymphatic invasion but not blood-borne metastasis (31, 32). Thus, the reduced TGFβ-signaling as a signature of IBC implies that IBC tumor cells migrate collectively and invade primarily into the lymph vessels. This is consistent with the presentation of IBC, with the presence of numerous tightly aggregated

nests of tumor cells, defined as tumor emboli that undergo dermal lymphatic invasion as the pathological hallmark of IBC. Based on the central role of TGF β in inducing the process of epithelial-to-mesenchymal transition (EMT) (30), our observations suggest that EMT is not the primary means of tumor cell migration/invasion in IBC. This agrees with the overexpression of E-Cadherin in IBC (33–35), which is a protein involved in retention of an epithelial phenotype that is lost during EMT, during which there is a gain of expression of genes involved in mesenchymal cell differentiation. Second, recent studies illustrate the importance of active TGF β -signaling in preventing the proliferation of ER+ tumor cells, and suggest that the loss of this inhibitory interaction may be important in early breast cancer progression (36). Therefore, we hypothesize that the reduced TGF β -signaling in IBC might lead to increased proliferation of ER+ tumor cells, which might be one of the reasons for the low frequency of Luminal A samples in IBC. Interestingly, we recently described that CDKN1B, a TGF β -target responsible for inhibition of cell proliferation, is repressed in IBC (36). In the current series, we observed a molecular subtype independent repression of CDKN1B in IBC (*data not shown*).

Our results also identified several transcription factors with an IBC-specific activation profile. In line with the observations made on repression of TGF β signaling, we observed reduced IBC-specific activity of the SMAD transcription factor, which acts downstream of TGF β -signaling. In fact, correlation analysis using independent gene sets suggests that both observations are linked (*data not shown*). Further corroborating our hypothesis that EMT is not the main mode of invasion observed in IBC is the finding of IBC-specific attenuation of the TCF3/TCF4 transcription factor activity, which is involved in the induction and maintenance of EMT through the repression of E-Cadherin (37).

In summary, the results obtained within the scope of this study provide novel insights into the biology of IBC and open an entirely new avenue of research topics. Without doubt, the most remarkable identified feature of IBC is the attenuated TGF β -signaling, and by extension the potential effects thereof on IBC tumor cell migration and invasion, which warrants further investigation. Our data also show that the differential distribution pattern of the molecular subtypes between IBC and nIBC should be considered when doing IBC research (e.g. in the present study, up to **82%** of the differentially expressed genes between IBC and nIBC are not IBC-specific), not only in translational studies but also in basic research. For example, the SUM149 cell line, although derived from a patient with IBC, remains a Basal-like cell line and therefore observations made using this model system might not be IBC-specific. **In fact, application of the 79 gene classifier onto a collection of 7 IBC cell lines obtained through the World IBC Consortium revealed that out of multiple replicates for each cell line, only the SUM149, KPL4 and FC-IBC-01 cell lines were reliably classified as IBC (*data not shown*).**

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The Institut National du Cancer (INCa) Translational Research Grant 2007 (FB). The National Institute of Health R01CA138239-01 and R01 CA138239 ARRA supplement (WAW, JMR); The State of Texas Grant for Rare and Aggressive Cancers (NU, JMR, AL, WAW); The American Airlines Komen Foundation Promise Grant KGO81287 (WAW, JMR, AL), Komen Foundation Grant KG101478 (WAW).

References

1. Dirix LY, van Dam P, Prové A, Vermeulen PB. Inflammatory breast cancer: current understanding. *Curr Opin Oncol* 2006;18:563–71. [PubMed: 16988576]
2. Bertucci F, Finetti P, Birnbaum D, Viens P. Gene expression profiling of inflammatory breast cancer. *Cancer* 2010;116:2783–93. [PubMed: 20503410]
3. Clarke R, Renshaw HW, Wang A, Xuan J, Liu MC, Gehan EA, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer* 2008;8:37–49. [PubMed: 18097463]
4. Van Laere S, Van der Auwera I, Van den Eynden GG, Fox SB, Bianchi F, Harris AL, et al. Distinct molecular signature of inflammatory breast cancer by cDNA microarray analysis. *Breast Cancer Res Treat* 2005;93:237–46. [PubMed: 16172796]
5. Bertucci F, Finetti P, Rougemont J, Charafe-Jauffret E, Nasser V, Lloriod B, et al. Gene expression profiling for molecular characterization of inflammatory breast cancer and prediction of response to chemotherapy. *Cancer Res* 2004;64:8558–65. [PubMed: 15574762]
6. Iwamoto T, Bianchini G, Qi Y, Cristofanilli M, Lucci A, Woodward WA, et al. Different gene expressions are associated with the different molecular subtypes of inflammatory breast cancer. *Breast Cancer Res Treat* 2011;125:785–95. [PubMed: 21153052]
7. Boersma BJ, Reimers M, Yi M, Ludwig JA, Luke BT, Stephens RM, et al. A stromal gene signature associated with inflammatory breast cancer. *Int J Cancer* 2008;122:1324–32. [PubMed: 17999412]
8. Bekhouche I, Finetti P, Adelaïde J, Ferrari A, Tarpin C, Charafe-Jauffret E, et al. High-resolution comparative genomic hybridization of inflammatory breast cancer and identification of candidate genes. *PLoS ONE* 2011;6:e16950. [PubMed: 21339811]
9. Van Laere S, Van der Auwera I, Van den Eynden G, Van Hummelen P, van Dam P, Van Marck E, et al. Distinct molecular phenotype of inflammatory breast cancer compared to non-inflammatory breast cancer using Affymetrix-based genome-wide gene-expression analysis. *Br J Cancer* 2007;97:1165–74. [PubMed: 17848951]
10. Bièche I, Lerebours F, Tozlu S, Espie M, Marty M, Lidereau R. Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. *Clin Cancer Res* 2004;10:6789–95. [PubMed: 15501955]
11. Dawood S, Merajver SD, Viens P, Vermeulen PB, Swain SM, Buchholz TA, et al. International expert panel on inflammatory breast cancer: consensus statement for standardized diagnosis and treatment. *Ann Oncol* 2011;22:515–23. [PubMed: 20603440]
12. Gong Y, Yan K, Lin F, Anderson K, Sotiriou C, Andre F, et al. Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study. *Lancet Oncol* 2007;8:203–11. [PubMed: 17329190]
13. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 2011;121:2750–67. [PubMed: 21633166]
14. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262–72. [PubMed: 16478745]
15. Van der Auwera I, Yu W, Suo L, Van Neste L, van Dam P, Van Marck EA, et al. Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS One* 2010;5:e12616. [PubMed: 20830311]

16. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–64. [PubMed: 12925520]
17. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160–7. [PubMed: 19204204]
18. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JJ, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 2010;12:R68. [PubMed: 20813035]
19. Gatz ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD, et al. A pathway-based classification of human breast cancer. *Proc Natl Acad Sci U S A* 2010;107:6994–9. [PubMed: 20335537]
20. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004;20:93–9. [PubMed: 14693814]
21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21. [PubMed: 11309499]
22. Trinh XB, Tjalma WA, Dirix LY, Vermeulen PB, Peeters DJ, Bachvarov D, et al. Microarray-based oncogenic pathway profiling in advanced serous papillary ovarian carcinoma. *PLoS One* 2011;6:e22469. [PubMed: 21799864]
23. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, et al. Adjustment of systematic microarray data biases. *Bioinformatics* 2004;20:105–14. [PubMed: 14693816]
24. Van Laere SJ, Van der Auwera I, Van den Eynden GG, Elst HJ, Weyler J, Harris AL, et al. Nuclear factor-kappaB signature of inflammatory breast cancer by cDNA microarray validated by quantitative real-time reverse transcription-PCR, immunohistochemistry, and nuclear factor-kappaB DNA-binding. *Clin Cancer Res* 2006;12:3249–56. [PubMed: 16740744]
25. Lerebours F, Vacher S, Andrieu C, Espie M, Marty M, Lidereau R, et al. NF-kappa B genes have a major role in inflammatory breast cancer. *BMC Cancer* 2008;8:41. [PubMed: 18248671]
26. Bertucci F, Finetti P, Rougemont J, Charafe-Jauffret E, Cervera N, Tarpin C, et al. Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. *Cancer Res* 2005;65:2170–8. [PubMed: 15781628]
27. Van Laere SJ, Van den Eynden GG, Van der Auwera I, Vandenberghe M, van Dam P, Van Marck EA, et al. Identification of cell-of-origin breast tumor subtypes in inflammatory breast cancer by gene expression profiling. *Breast Cancer Res Treat* 2006;95:243–55. [PubMed: 16261404]
28. Chen X, Rotenberg SA. PhosphoMARCKS drives motility of mouse melanoma cells. *Cell Signal* 2010;22:1097–103. [PubMed: 20211725]
29. Van der Auwera I, Limame R, van Dam P, Vermeulen PB, Dirix LY, Van Laere SJ. Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype. *Br J Cancer* 2010;103:532–41. [PubMed: 20664596]
30. Gotzmann J, Mikula M, Eger A, Schulte-Hermann R, Foisner R, Beug H, et al. Molecular aspects of epithelial cell plasticity: implications for local tumor invasion and metastasis. *Mutat Res* 2004;566:9–20. [PubMed: 14706509]
31. Giampieri S, Manning C, Hooper S, Jones L, Hill CS, Sahai E. Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility. *Nat Cell Biol* 2009;11:1287–96. [PubMed: 19838175]
32. Giampieri S, Pinner S, Sahai E. Intravital imaging illuminates transforming growth factor beta signaling switches during metastasis. *Cancer Res* 2010;70:3435–39. [PubMed: 20424121]
33. Colpaert CG, Vermeulen PB, Benoy I, Soubry A, Van Roy F, van Beest P, et al. Inflammatory breast cancer shows angiogenesis with high endothelial proliferation rate and strong E-cadherin expression. *Br J Cancer* 2003;88:718–25. [PubMed: 12618881]
34. Kleer CG, van Golen KL, Braun T, Merajver SD. Persistent E-cadherin expression in inflammatory breast cancer. *Mod Pathol* 2001;14:458–64. [PubMed: 11353057]
35. Charafe-Jauffret E, Tarpin C, Bardou V-J, Bertucci F, Ginestier C, Braud A-C, et al. Immunophenotypic analysis of inflammatory breast cancers: identification of an “inflammatory signature.” *J Pathol* 2004;202:265–73. [PubMed: 14991891]

36. Ren Y, Wu L, Frost AR, Grizzle W, Cao X, Wan M. Dual effects of TGF-beta on ERalpha-mediated estrogenic transcriptional activity in breast cancer. *Mol Cancer* 2009;8:111. [PubMed: 19943940]
37. Hugo HJ, Kokkinos MI, Blick T, Ackland ML, Thompson EW, Newgreen DF. Defining the E-cadherin repressor interactome in epithelial-mesenchymal transition: the PMC42 model as a case study. *Cells Tissues Organs* 2011;193:23–40. [PubMed: 21051859]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Statement of Translational Relevance

The clinical relevance of the research described in this paper is manifold. First, due to the extent of our series, which includes by far the largest series of IBC samples ever reported, we were able to delineate the molecular profile of IBC with enhanced accuracy, taking into account various points of criticism raised with respect to earlier studies. Our report is the first to analyze the expression profiles of IBC samples in light of a uniform case definition for IBC put forward by an international expert panel. Also, the present study is the first to deal with the fact that the differential distribution pattern of the molecular subtypes between IBC and nIBC affects its outcome. Further, our data **suggest** that differentially expressed genes between IBC and nIBC in a molecular subtype independent manner hold the fingerprints of aggressive breast tumor behavior. Therefore, our expression series not only suits IBC research, but **might also be** instrumental for scientists focused on breast cancer invasion and metastasis in general. Finally, this study gives voice to the World IBC Consortium, which is founded to foster collaborations between researchers active within the IBC-community. We believe that our effort is an example of how collaborations with a strong international backbone can prosper research well into the 21st century.

Figure 1: Unsupervised hierarchical cluster analysis

Clustering was done with Manhattan distance as similarity measure and Ward linkage as the dendrogram drawing method for probe sets having a standard deviation greater than 2 in our series of IBC (A: 298 probe sets) and nIBC (B: 251 probe sets) samples only, as well as in the combined data set (C: 270 probe sets). Median-centered gene expression data are represented in a matrix format with rows indicating genes and columns indicating samples. Overexpressed genes (compared with the median) are color-coded red and repressed genes are color-coded green. Color saturation indicates the level of overexpression or repression. The clusters that were considered stable using the silhouette algorithm are indicated in the dendrogram using alternating grey and blue colors. The classification of the samples according to the molecular subtypes (Basal-like, Claudin-low, ErbB2+, Luminal A, Luminal B, and Normal-like): is represented underneath the heatmaps in matrix format. Tumor subtype classification (IBC *vs.* nIBC) is also indicated underneath the heatmap representing the combined data set. Black vertical ticks indicate class membership. For convenience, the shading pattern of the dendrograms is repeated on the class-membership matrix.

Figure 2: Supervised analysis between IBC and nIBC according to nIBC stage-matching
 Comparison of parameters of differential expression (Q-values (A), number of significant genes at FDR<0.050 (B), and number of significant genes at FDR<0.100 (C)) between IBC and nIBC based on a comparison with stage-matching (Stage III-IV) and for the comparison with no stage-matching (Stage I-IV). The blue vertical line indicates the value for the corresponding parameter for the stage-matched comparison. The black line indicates the distribution of the values of the same parameters when analyzing differential gene expression in the non-stage-matched analysis using 10 random series of equal sample sizes as was the case for the stage-matched comparison. The Venn diagram (D) shows the concordance in the list of differentially expressed genes between IBC and nIBC samples for stage matched nIBC and non-stage matched nIBC (Stage I-II, and Stage III-IV, Stage I-IV). Numbers indicate the number of genes in each section.

Figure 3: Construction and assessment of the IBC/nIBC classification models

Using PAM, classification models to discriminate between IBC and nIBC were constructed. The first model was based on the full list of 491 probe sets with an IBC-specific expression component. The second model was based on the list of 79 uniquely IBC-specific probe sets. Both models were constructed using the training set and tested on the validation set. The leave-one-out cross-validated probabilities for all samples in the training set based on the full list of probe sets with an IBC-specific expression component (A) and the list of uniquely IBC-specific probe sets (C) are provided in a dot plot format. Each sample is represented by two coloured dots, with red corresponding to the cross-validated probability that the related sample belongs to the IBC phenotype and green corresponding to the cross-validated probability that the related sample belongs to the nIBC phenotype. The real classification of the samples according to tumour phenotype is provided along the X-axis, with samples to the left of the vertical dashed line being IBC samples and vice versa. One can clearly observe that the majority of the IBC and nIBC samples are correctly classified using both models, when using a cross-validated probability of 0.500 as threshold for classification. Both models were subsequently evaluated in an independent test set. The posterior probabilities to be classified as IBC for all samples in the validation set based on the full list of probe sets with an IBC-specific expression component (B) and the list of uniquely IBC-

specific probe sets (D) are provided in a dot plot format. A red dot indicates an IBC sample; a green dot indicated a nIBC sample. The posterior probabilities are plotted in ascending order. Most nIBC samples have low posterior probabilities to be classified as IBC and vice versa. Again, when using a posterior probability of 0.500 as threshold for classification we obtain a classification accuracy of 68% and 71% for the model based on the full list of probe sets and the model based on the list of uniquely IBC-specific probe sets.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

Histo-clinical data of IBC and nIBC samples

Characteristic	IBC (n=137)	nIBC (n=252)	<i>p-value</i>	odds ratio
Median age, years (range)	50 (24–82)	55 (24–89)	0.002	NA
AJCC stage			0.000	NA
1	0 (0%)	65 (26%)		
2	0 (0%)	97 (39%)		
3	106 (76%)	66 (27%)		
4	31 (24%)	20 (8%)		
Histological type			0.042	NA
ductal	122 (91%)	203 (81%)		
lobular	7 (5%)	23 (9%)		
other	6 (4%)	26 (10%)		
pN			NR	NR
negative	NR	58 (43%)		
positive	NR	77 (57%)		
pT			NR	NR
pT1	NR	46 (34%)		
pT2	NR	65 (49%)		
pT3	NR	23 (17%)		
ER (mRNA level)			0.000	2.36 (1.50–3.74)
negative	66 (48%)	71 (28%)		
positive	71 (52%)	181 (72%)		
ERBB2 (mRNA level)			0.001	2.38 (1.39–4.00)
negative	95 (69%)	212 (84%)		
positive	42 (31%)	40 (16%)		
Genomic Grade Index			0.003	1.97 (1.24–3.18)
high	98 (72%)	141 (56%)		
low	39 (28%)	111 (44%)		
Systemic therapy (CT and/or HT and/or T)			NR	NR
yes	134 (98%)	219 (86%)		
no	0 (0%)	33 (13%)		
not available	3 (2%)	20 (1%)		
Median follow-up, months (range) *	56 [2–165]	80 [6–182]	NR	NR
DMFS-5y **	53% [43–65]	82% [77–88]	0.000	NA

NR, not relevant; CT, chemotherapy; HT, hormone therapy; T, trastuzumab

* calculated only for M0 patients without any metastatic relapse during follow-up, from the date of diagnosis to the date of last news

** numbers in brackets represent CI95%

Table 2:

Results of supervised analyses between IBC and nIBC according to the disease stage of nIBC

<i>COMPARISON</i>	<i>N IBC</i>	<i>N NIBC</i>	<i>N GENES</i>	<i>OBS Q</i>	<i>EXP Q</i>	<i>SD Q</i>	<i>P-VALUE</i>	<i>FDR<0,05</i>	<i>FDR<0,10</i>
IBC vs, Stage I-II nIBC	137	162	9,926	48.589	10	2.493	<0,001	3,363	4,524
IBC vs, Stage III-IV nIBC	137	86	9,926	30.176	10	2.393	<0,001	1,134	2,164
IBC vs, Stage I-IV nIBC	137	252	9,926	53.574	10	2.385	<0,001	3,746	5,009
IBC vs, Stage I-IV nIBC(Perm)	137	86	9,926	33.672	10	2.440	<0,001	1,930	3,037
<i>Range</i>				(28,625 – 39,679)		(2,395 – 2,483)		(1.082 – 2.630)	(2.132 – 3.731)