

Preclinical study

Identification of cell-of-origin breast tumor subtypes in inflammatory breast cancer by gene expression profiling

Steven J. Van Laere¹, Gert G. Van den Eynden¹, Ilse Van der Auwera¹, Melanie Vandenberghe¹, Peter van Dam¹, Eric A. Van Marck¹, Kenneth L. van Golen², Peter B. Vermeulen^{1,3}, and Luc Y. Dirix¹

¹Translational Cancer Research Group, Lab Pathology University of Antwerp and Oncology Center, General Hospital Sint-Augustinus, Wilrijk, Belgium; ²Department of Internal Medicine, Division of Hematology and Oncology, The University of Michigan Comprehensive Cancer Center, Ann Arbor, MI, USA; ³Department of Pathology, AZ Sint-Augustinus, Oosterveldlaan 24, B2610, Wilrijk, Belgium

Key words: cell-of-origin subtypes, gene-expression profiling, inflammatory breast cancer, microarray

Abstract

Inflammatory breast cancer (IBC) is an aggressive form of locally advanced breast cancer with high metastatic potential. Most patients have lymph node involvement at the time of diagnosis and 1/3 of the patients have distant metastases. In a previous study, we demonstrated that IBC is a distinct form of breast cancer in comparison with non-IBC. The aim of this study was to investigate the presence of the different molecular subtypes in our data set of 16 IBC and 18 non-IBC specimen. Therefore, we selected an ‘intrinsic gene set’ of 144 genes, present on our cDNA chips and common to the ‘intrinsic gene set’ described by Sorlie et al. [PNAS, 2003]. This set of genes was tested for performance in the Norway/Stanford data set by unsupervised hierarchical clustering. Expression centroids were then calculated for the core members of each of the five subclasses in the Norway/Stanford data set and used to classify our own specimens by calculating Spearman correlations between each sample and each centroid. We identified the same cell-of-origin subtypes in IBC as those already described in non-IBC. The classification was in good agreement with immunohistochemical data for estrogen receptor protein expression and cytokeratin 5/6 protein expression. Confirmation was done by an alternative unsupervised hierarchical clustering method. The robustness of this classification was assessed by an unsupervised hierarchical clustering with an alternative gene set of 141 genes related to the cell-of-origin subtypes, selected using a discriminating score and iterative random permutation testing. The contribution of the different cell-of-origin subtypes to the IBC phenotype was investigated by principal component analysis. Generally, the combined ErbB2-overexpressing and basal-like cluster was more expressed in IBC compared to non-IBC, whereas the combined luminal A, luminal B and normal-like cluster was more pronounced in non-IBC compared to IBC. The presence of the same molecular cell-of-origin subtypes in IBC as in non-IBC does not exclude the specific molecular nature of IBC, since gene lists that characterize IBC and non-IBC are entirely different from gene lists that define the different cell-of-origin subtypes, as evidenced by principal component analysis.

Introduction

Breast cancer is a major cause of morbidity and mortality [1], characterized by heterogeneity in its natural history and in the responsiveness to treatments. To tackle this heterogeneity of breast cancer, a cell-of-origin classifier was described and identified five cell-of-origin subtypes related to mammary epithelial biology: luminal A, luminal B, basal-like, ErbB2-overexpressing, and normal-like [2,3,4]. These molecular subtypes have been associated with different histoclinical features including clinical outcome. Specific genome alterations and

distinct gene expression changes in response to chemotherapy have been associated to some subtypes, further suggesting that these molecular subtypes represent distinct diseases [5,6]. Recently, some of the cell-of-origin subtypes have been associated with cancer stem cell biology [7].

Inflammatory breast cancer (IBC) is the most lethal form of breast cancer, characterized by a very aggressive behavior and high metastatic potential. At time of diagnosis most women have axillary lymph node involvement and 1/3 of the patients have metastases in distant organs. The disease presents at a younger age

than non-IBC, with a rapidly progressing inflammatory tumor associated redness, warmth, pain, induration and edema. A palpable tumor mass is not necessarily present [8]. Tumor emboli in dermal lymphatics may be apparent on skin biopsy, but, in the absence of clinical findings do not indicate IBC. Hence, IBC is primarily a clinical diagnosis, classified T4d in the TNM classification of the American Joint Committee on Cancer [9]. About two-thirds of the IBC tumors are of high histological grade, a ratio far higher than in non-IBC. Given the high histological grade, IBC is more frequently estrogen receptor (ER)-negative than non-IBC and up to 60% lack hormone receptor expression, whereas other forms of locally advanced breast cancer (LABC) are more likely to be ER-positive [8]. In addition, the prevalence of the TP53 mutation as well as the prevalence of the amplification of the ErbB2 oncogene were found to be much higher in IBC compared to non-IBC [10]. Despite advances in the multidisciplinary treatment, the prognosis of IBC is less favorable than of non-IBC, with a 3-year survival of about 40% [8]. Because of its relative scarcity and difficulty of obtaining diagnostic samples of sufficient size, little is known about the molecular basis of IBC. The study of the molecular mechanisms of IBC can, however, lead to new therapeutic targets, and, to insight in mechanisms of motility and metastasis of breast cancer cells.

Recently, both *in vivo* and *in vitro* experiments have identified genes that contribute to the aggressive phenotype of IBC. Overexpression of RhoC GTPase is observed in over 90% of IBC in contrast to 36% of non-IBC [11,12]. Overexpression of RhoC in human mammary epithelial (HME) cell lines nearly recapitulated the IBC phenotype with regards to invasion, motility [13] and production of angiogenic growth factors [14]. Activation of the mitogen activated protein kinase (MAPK) pathway was responsible for these features, whereas anchorage independent growth was induced through the Phosphatidylinositol 3-phosphate Kinase (PI3K) pathway [15]. Addition of a farnesyl transferase inhibitor led to the reversion of the RhoC-induced inflammatory breast cancer phenotype [16]. Moreover, expression of WISP3, a tumor growth and angiogenesis inhibitory protein, is lost in over 90% of the IBC tumors [17]. It has been shown that WISP3 attenuates RhoC protein expression in HME cell lines and in the SUM149 IBC cell line, through the modulation of IGF signaling, inducing opposite changes in invasive, motile and angiogenic characteristics. Loss of WISP3 again recapitulates the IBC phenotype [17,18,19]. In addition to the modulation of angiogenic parameters by RhoC and WISP3, increased angiogenesis in IBC was evident by both an increased number of microvessels and a higher fraction of proliferating endothelial cells [20]. Increased expression of several angiogenic and lymphangiogenic growth factors and growth factor receptors in IBC compared to non-IBC was demonstrated by quantitative real-time RT-PCR [21]. These findings suggest that IBC constitutes a subset of breast carcinoma with a specific biology.

Despite the recent advances in characterizing the molecular basis of IBC, much remains to be elucidated. Therefore genome-wide expression profiling has been performed to investigate additional mechanisms contributing to the biology of IBC [22,23]. We have identified a molecular signature, based on the expression of 756 genes, using 16 IBC and 18 non-IBC specimens. Application of this molecular signature by unsupervised hierarchical clustering methods resulted in a perfect separation of IBC and non-IBC specimens. In the list of differentially expressed genes a high number of NF-kappaB target genes as well as upstream activators of the NF-kappaB pathway (IL1 and Toll Like Receptor pathways) were found. Using a class prediction method based on the expression of 50 genes, we were able to assign 14/16 (87%) IBC and non-IBC samples to the correct class. In addition, we found an increased expression of markers representing the breast basal-like cell-of-origin subtype in IBC. Conversely, markers representing the luminal A cell-of-origin subtype were more pronounced in non-IBC. In agreement with the findings of specific molecular alterations in IBC tumor samples, these data suggested that IBC and non-IBC have to be regarded as two separate entities with important diagnostic and therapeutic consequences [23].

The objective of this study was to determine to which extent the cell-of-origin subtypes, as described by Sorlie et al. [4], determine the specific phenotype of IBC. These cell-of-origin subtypes have been studied once in IBC [24]. In this study the cell-of-origin subtype profile did not discriminate IBC from non-IBC. The present study indicates that the IBC phenotype is mainly determined by a set of genes independent of the gene sets that define the cell-of-origin subtypes, although the latter gene sets contribute.

Materials and methods

Patients and samples

Breast cancer samples were obtained from 34 patients with breast adenocarcinoma treated in the General Hospital Sint-Augustinus, Wilrijk, Belgium. Each patient gave written informed consent. Samples were stored in liquid nitrogen within 15 min after excision. Thirty-four samples included 16 pretreatment samples of patients with IBC, diagnosed by strictly respecting the criteria mentioned in the TNM classification of the American Joint Committee on Cancer as T4d [9]. The presence of tumor emboli in dermal lymphatics was, as an isolated pathological finding, not sufficient for the diagnosis of IBC. Of the 18 non-IBC samples, 10 represented LABC (7 T3, 3 T4a), 3 samples represented T2-tumors, and five represented T1-tumors. Thirteen patients with non-IBC had pathological axillary lymph node involvement.

RNA Isolation, processing and cDNA microarray hybridization

RNA was isolated and processed as described before [23], using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA). High quality RNA was reverse transcribed, amplified and Cy5 labeled using the Amino Allyl MessageAmp aRNA kit (Ambion Inc., Austin, TX, USA). Universal Human Reference RNA (Stratagene, La Jolla, CA, USA) was processed similarly and Cy3 labeled for competitive hybridization. cDNA chips were obtained from the Sanger Center and hybridized during 16 h at 47 °C in a volume of 40 µl. Information regarding the clone set and the microarray production can be obtained from the world wide web: www.sanger.ac.uk/Projects/Microarrays. After hybridization, slides were washed and scanned immediately using ScanArray software. Data were generated using QuantArray software. The methodology has been described in Van Laere et al. [23].

Data analysis

Biologically relevant subtypes of non-IBC have been defined using an 'intrinsic' set of ~500 genes [2,3,4]. These genes were selected from a group of ~8000 genes, based on a lower variation in expression between paired samples from the same patients than between samples from different patients. To test the generality of these subtypes in our series of samples, we analyzed our samples with genes common to the intrinsic gene set and ours. Of the 500-gene set used by Sorlie et al. [4], 165 genes were present on our chip represented by ~240 clones. To prevent batch related clustering, genes with a batch related gene expression difference were filtered out. This resulted in a gene set of 144 genes represented by ~200 clones.

Prior to analyzing our series of samples using the reduced intrinsic set of genes, we first investigated whether these 144 genes were still able to identify five cell-of-origin subtypes, and their associated gene clusters in breast cancer. Therefore, we submitted the Norway/Stanford series of 122 breast tissue samples to hierarchical clustering based on the expression of the 144 gene set, represented by ~150 clones in the original Norway/Stanford data set. Normalized and log₂ transformed expression data for 150 clones were extracted, median-centered on genes and analyzed using unsupervised hierarchical clustering with Pearson correlation as similarity metric. Clustering was visualized using TreeView (freeware, <http://rana.lbl.gov/EisenSoftware.htm>).

For each cell-of-origin subtype, we subsequently computed the typical expression profile for the 144 genes, hereafter designated centroid. The samples from the luminal A, the ErbB2-overexpressing, and the basal centroids were easily defined based on the expression pattern for the ER, the ErbB2, and the basal gene clusters, respectively, and by selecting tumors with the highest correlation with each other within the subgroup.

The samples for the normal, and the luminal B centroids were defined based on their clustering as a separate node in the dendrogram, representing tumor samples with a high correlation with each other within the subgroup. The centroid expression for each of the five tumor subgroups was calculated as the average expression for each of the 144 genes in the corresponding samples. Centroid expression for genes represented by multiple clones was calculated by averaging the centroid expression of the representing clones. The classification power of the centroids was tested by classifying the remaining samples ($n = 56$) from the Norway/Stanford data set by calculating Spearman correlation coefficients between each remaining sample and each centroid. The samples were classified according to the highest correlation coefficient between a sample and a centroid.

In order to identify cell-of-origin groups in our series of breast tumor samples, normalized gene expression data for 200 clones, representing 144 genes were extracted from our data set. Gene expression data for multiple clones representing a single gene were averaged. Next, data were log₂ transformed and median centered on genes. In order to classify our breast tumor samples, Spearman correlations were calculated between each sample in our series and each centroid based on the Norway/Stanford data set. Again, a breast tumor sample was classified according to the highest correlation coefficient between its molecular profile and any of the calculated centroids. Unsupervised hierarchical clustering, with data median-centered on genes, average linkage clustering, and the Pearson correlation coefficient as similarity metric was applied to investigate the reliability of the subtype classification.

To investigate the robustness of this classification, supervised analysis was applied to identify genes that discriminate between the cell-of-origin subtypes. Each subtype was compared with the others by using a discriminating score (DS) combined with iterative random permutation testing. The DS was calculated for each gene [25] as $DS = (M1 - M2)/(S1 + S2)$, where M1 and S1, respectively, represent mean and SD of expression levels of the gene in subgroup 1, and M2 and S2 in subgroup 2. Confidence levels were estimated by 200 iterative random permutations of samples as previously described [26] with a significance threshold of 0.05. The final list of discriminator genes for the analysis included at most the 15 top-ranked genes identified as discriminator in at least one comparison of cell-of-origin groups. Selection of the list of discriminator genes was performed using GeneCluster 2.0 (Broad Institute, Cambridge, MA, USA). Once identified, the classification power of the discriminator signature was illustrated by performing unsupervised hierarchical clustering using data median-centered on genes, Pearson correlation as similarity metric and average-linkage clustering using GeneSpring software (Silicon Genetics, Redwood City, CA, USA).

To identify to which extent the different cell-of-origin subtypes are responsible for the variability seen in the

entire data set, we executed a principal component analysis (PCA) for the five cell-of-origin subtypes on a gene list with 5914 non-batch-related genes, using GeneSpring software. In a second approach we investigated to what extent the different cell-of-origin subtypes characterize IBC. Therefore, PCA was executed for the same five subgroups on a gene lists with 953 genes with a statistically significant ($p < 0.05$), 1.5-fold difference in expression level between IBC and non-IBC. PCA is a decomposition technique that produces a set of expression patterns (eigenvectors) known as the principal components. Each principal component is associated with a value (eigenvalue) representing the amount of variability explained by that expression pattern. The expression pattern with the largest value is the first principal component, the expression pattern with the second largest value is the second principal component, and so on. Once the principal components were identified we executed unsupervised hierarchical clustering with genes strongly correlated ($r > 0.95$) with any of the principal components to investigate common biological themes defined by the principal components.

Immunohistochemistry (IHC)

ER-alpha protein expression was detected by the antibody clone ID5 (M7047, DakoCytomation, Glostrup, Denmark). Tumors were regarded as ER-alpha-positive when at least 10% of the tumor cells had nuclear protein expression. Staining was done on biopsy specimens used for diagnosis in the surgical pathology lab of the General Hospital Sint-Augustinus. Scoring was done by a surgical pathologist (P.B.V.).

Cytokeratin 5/6 (CK5/6) protein expression was detected with the monoclonal antibody, clone D5/16 B4 (M7237, DakoCytomation, Glostrup, Denmark). Formalin fixed, paraffin embedded tissue sections from 16 IBC and 18 non-IBC were rehydrated through sequential changes of alcohol and distilled water. Antigen retrieval was performed for 30 min in citrate buffer (pH 6) at 95 °C. Sections were incubated with the primary antibody for 1 h at room temperature using a dilution of 1400 µg/l. The Dako Envision system on the Dako Cytomation autostainer was used for visualization of antibody binding. Tumor samples were regarded CK5/6+ when tumor cells showed membranous or cytoplasmic staining.

Statistical analysis

Statistical analyses were done using SPSS software (SPSS Inc., version 12.0). The fraction of the combined luminal A, luminal B and normal-like cluster in IBC and non-IBC was compared with the fraction of the combined basal-like and ErbB2-overexpressing cluster in IBC and non-IBC by using χ^2 test. Correlations between immunohistochemical data for ER en CK5/6 and sample groups were studied by Fisher's exact test or χ^2 test when appropriate.

Results

Selection of a common intrinsic gene set and performance in the Norway/Stanford tumors

One hundred and sixty-five genes, represented by ~240 clones, were common to our cDNA chip and the 'intrinsic gene set' used by Sorlie et al. [4] to identify five molecular subtypes in breast cancer. After exclusion of 21 genes with a batch related gene expression profile, 144 genes represented by ~200 clones, were left over. We investigated whether these 144 genes were still able to discriminate between the same five cell-of-origin subtypes in the Norway/Stanford series. Gene expression data for 122 breast tissue samples from the Norway/Stanford data set were retrieved from the world wide web (http://genome-www.stanford.edu/breast_cancer/robustness/) and submitted to hierarchical clustering using data median-centered on genes, average linkage clustering and Pearson correlation coefficient as similarity metric. Results are shown in Figure 1. Gene clustering revealed groups of coordinately expressed genes, some of which represented known expression signatures corresponding to cell types: the 'ER cluster', including ESR1, MUC1 and SIAH2, playing a prominent role in the classification of samples, the 'ErbB2 cluster' including ERBB2 and GRB7, and the 'basal cluster' including KRT5, TRIM29, S100A1, ANXA8 and FZD7. Samples clustered in two main branches (left and right) and five major subgroups. The left branch was composed of the normal-like, the basal-like, ErbB2-overexpressing and the luminal B subgroups, whereas the luminal A samples clustered in the right branch. The luminal A, basal-like and ErbB2-overexpressing clusters were characterized by clear gene clusters (ER, basal and ErbB2, respectively) as described above. Upon visual inspection of the color-coded expression profile (Figure 1), the normal-like subgroup displayed a strong expression of the 'basal cluster' and low expression of the 'ER cluster'. The cluster containing luminal B samples was the only cluster that was poorly identified. Out of 17 samples originally identified as belonging to the luminal B cluster, only 9 samples clustered together using gene expression data for our 144 'intrinsic genes'. Hence, no gene cluster was evident, although the luminal B subgroup did show lower expression of the 'ER cluster', compared to the luminal A subgroup. Despite the limited number of genes, these subgroups were very similar to those previously described and 84% of the Norway/Stanford samples clustered together in the same manner as described in the original manuscript [4]. These data indicate that our gene set, composed of 144 genes common to our cDNA chip and the intrinsic gene set, was able to identify the different cell-of-origin subtypes.

Expression centroids were then calculated for the core members of each of the five subclasses in the Norway/Stanford data set and used to classify our

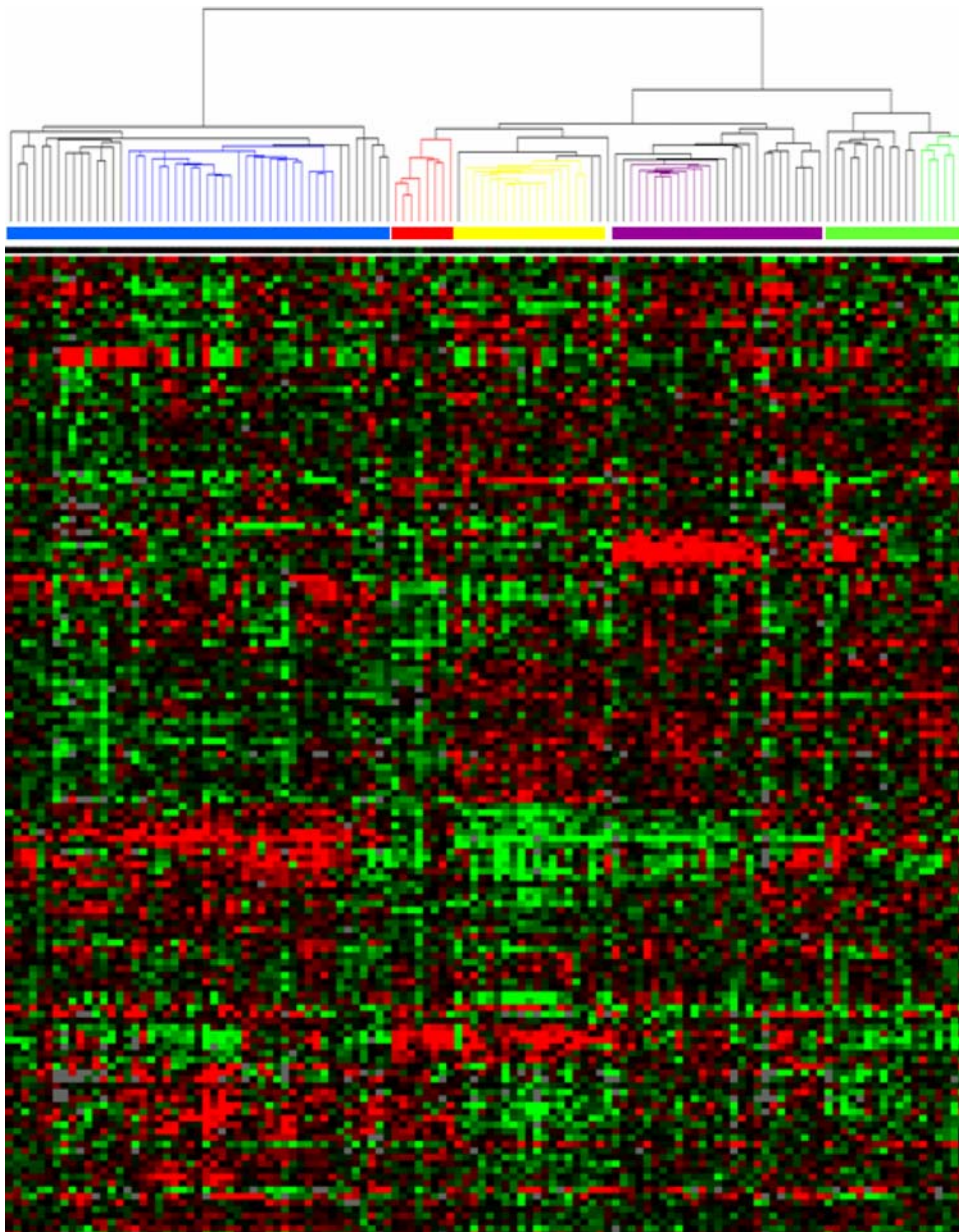


Figure 1. Hierarchical clustering of the Norway/Stanford data set using ~ 150 clones representing 144 genes common to the intrinsic gene set and our cDNA chip. Data are presented in matrix format with rows corresponding to genes and columns corresponding to samples. The expression level of each single gene is relative to its median across all samples, and is color coded red or green indicating genes above and below the median expression level, respectively. The magnitude of deviation from the median is represented by the color saturation. Grey represents genes with missing data. Based on the expression pattern of 144 genes, five cell-of-origin subgroups were clearly identified, three of which were represented by a corresponding gene cluster. The clusters are indicated by the colored bars beneath the dendrogram: luminal A (blue), normal-like (red), basal-like (yellow), ErbB2-overexpressing (purple), and Luminal B (green). Branches of the core samples used for calculation of the centroid expressions are similarly color coded. The corresponding gene clusters for luminal A, basal-like and ErbB2-overexpressing samples are indicated on the right: luminal A (blue), basal-like (yellow) and ErbB2-overexpressing (purple).

specimen. The core members for the luminal A, the ErbB2-overexpressing and the basal centroids were easily defined based on the expression pattern of the ER, ErbB2 and the basal gene clusters, respectively, and by selecting tumors with the highest correlation with each other within the subgroup: 27 samples for luminal A (correlation > 0.43), 11 samples for ErbB2-overexpressing (correlation > 0.47) and 16 samples for basal-like (correlation > 0.53). For the normal-like and luminal B subgroups, respectively 8 (correlation > 0.29)

and 5 (correlation > 0.32) samples were chosen. Altogether, 78% of the samples selected in this manner were the same as those selected by Sorlie et al. [4] for calculating their centroid expressions. The basal centroid proved to be the most homogeneous one. Centroid expression for each of the five cell-of-origin subtypes was calculated by averaging the expression for each of the 144 genes in the corresponding samples.

The classification power of the centroids was then assessed by classifying the remaining samples in the

Norway/Stanford data set. In total, 70% (39/56) of the remaining samples were assigned to the same cell-of-origin subtype as described by the authors in the original manuscript [4]. All of the basal-like (3/3), 77% of the ErbB2-overexpressing samples (10/13), 66% of the luminal A and the normal-like samples (18/27 and 2/3, respectively), and 60% of the luminal B samples (6/10) were correctly assigned to a cell-of-origin subtype with the centroid classification method.

Identification of cell-of-origin subtypes in inflammatory breast cancer

After calculating the centroid expression profile for each of the cell-of-origin subtypes, we looked for common themes in our series of 34 breast cancer samples (16 IBC and 18 non-IBC). We calculated Spearman correlations between each sample and each centroid. A breast tumor sample was classified according to the highest correlation of its molecular profile and any of the calculated centroids. Twelve samples (5 IBC and 7 non-IBC) were closer to the luminal A centroid, 5 (1 IBC and 4 non-IBC) to the luminal B centroid, 9 (6 IBC and 3 non-IBC) to the basal-like centroid, 2 (2 IBC and 0 non-IBC) to the ErbB2-overexpressing centroid, and 6 (2 IBC and 4 non-IBC) to the normal-like centroid. There was a good agreement between these assignments and protein expression data for ER and CK5/6, a representative marker for the basal gene cluster. Two out of 11 samples in the combined basal-like and ErbB2-overexpressing subgroup were ER-positive, whereas 12/23 tumors belonging to the combined luminal A, luminal B and normal-like cluster were ER-positive (Fisher's Exact test, $p = 0.063$). Six out of 9 samples in the basal cluster were CK5/6+ whereas only 6 out of 25 samples in the remaining clusters were CK5/6-positive (Fisher's Exact test, $p = 0.040$). A higher amount of basal-like and ErbB2-overexpressing tumors and a lower amount of luminal A, luminal B or normal-like tumors in IBC as compared to non-IBC was found. In IBC, 8/16 samples (50%) belonged to the combined basal-like and ErbB2-overexpressing cluster compared to only 3/18 non-IBC samples (16.7%). However, 15/18 non-IBC samples (83.3%) belonged to the combined luminal A, luminal B and normal-like cluster compared to 8/16 IBC samples (50%) (Pearson χ^2 , $p = 0.038$). This was in contrast with the results of Bertucci et al. [24]: 19/36 IBC samples (52.8%) belonged to the combined luminal A, luminal B and normal-like cluster compared to 18/32 non-IBC samples (56.3%). 17/36 IBC samples (47.2%) belonged to the combined ErbB2-overexpressing and basal-like tumors compared to 14/32 non-IBC samples (43.7%) (Pearson χ^2 , $p = 0.774$).

We applied hierarchical clustering to the expression levels of 144 genes in our series of 16 IBC and 18 non-IBC samples. Results are shown in Figure 2. We identified two major groups and four subgroups related

to the cell-of-origin subtype: normal-like, luminal A, luminal B and basal-like. No ErbB2-overexpressing subgroup was identified, probably due to the low number of samples correlated with the ErbB2-overexpressing centroid. The two major groups identified were in close agreement with the immunohistochemistry status for ER: in the left group 10 out of 16 samples (62.5%) were ER-positive whereas in the right group 4 out of 18 samples (22.2%) were ER-positive (Pearson χ^2 , $p = 0.01$). Four samples were not assigned to any subgroup, 2 of which were weakly correlated to a centroid. There was a strong association between the grouping of samples and the centroid they were closer to. In the left group, 4 out of 6 normal-like samples and 10 out of 12 luminal A samples clustered together. In the right group, 4 out of 5 luminal B samples and 5 out of 9 basal-like tumors clustered together. The luminal B subgroup was characterized by the presence of 2 samples, closely correlated with both the basal-like and the luminal B centroid, but designated basal-like due to a stronger correlation coefficient between these samples and the basal-like centroid. The basal-like subgroup was entirely composed of IBC specimens. The five cell-of-origin subtypes concerned all samples, the IBC samples were intermingled with the non-IBC samples across most subgroups. This indicates that the different cell-of-origin subtypes are present in both IBC and non-IBC.

Robustness of the taxonomy

To test for the robustness and reliability of the cell-of-origin classification, we performed an unsupervised analysis using an alternative gene set. Based upon our previous cell-of-origin classification, genes were selected by comparing each cell-of-origin subtype with the others, using a discriminating score combined with permutations tests. Due to the low number of samples correlated with the ErbB2-overexpressing centroid, permutation testing for any comparison involving the ErbB2-overexpressing subgroup was unstable and the ErbB2-overexpressing subgroup was left out of the analysis. We developed a molecular signature discriminating between the four centroid-based subtypes by selecting at most 15 top-ranked discriminating genes in each comparison. In this way, we identified 141 genes as discriminatory between the four subtypes.

This list of discriminator genes was then submitted to hierarchical clustering. Results are shown in Figure 3. We identified two major groups and four major subgroups, in close agreement with the clustering reported in Figure 2. The left branch was mainly composed of ER-positive samples (10/16), whereas in the right branch ER-negative samples (10/13) were predominant (Pearson χ^2 , $p = 0.03$). Three samples were not assigned to a specific subgroup, two of which showed low correlation with any of the four cell-of-origin centroids. The left branch was characterized by the presence of both the luminal A and the normal-like subgroup. Eleven out of

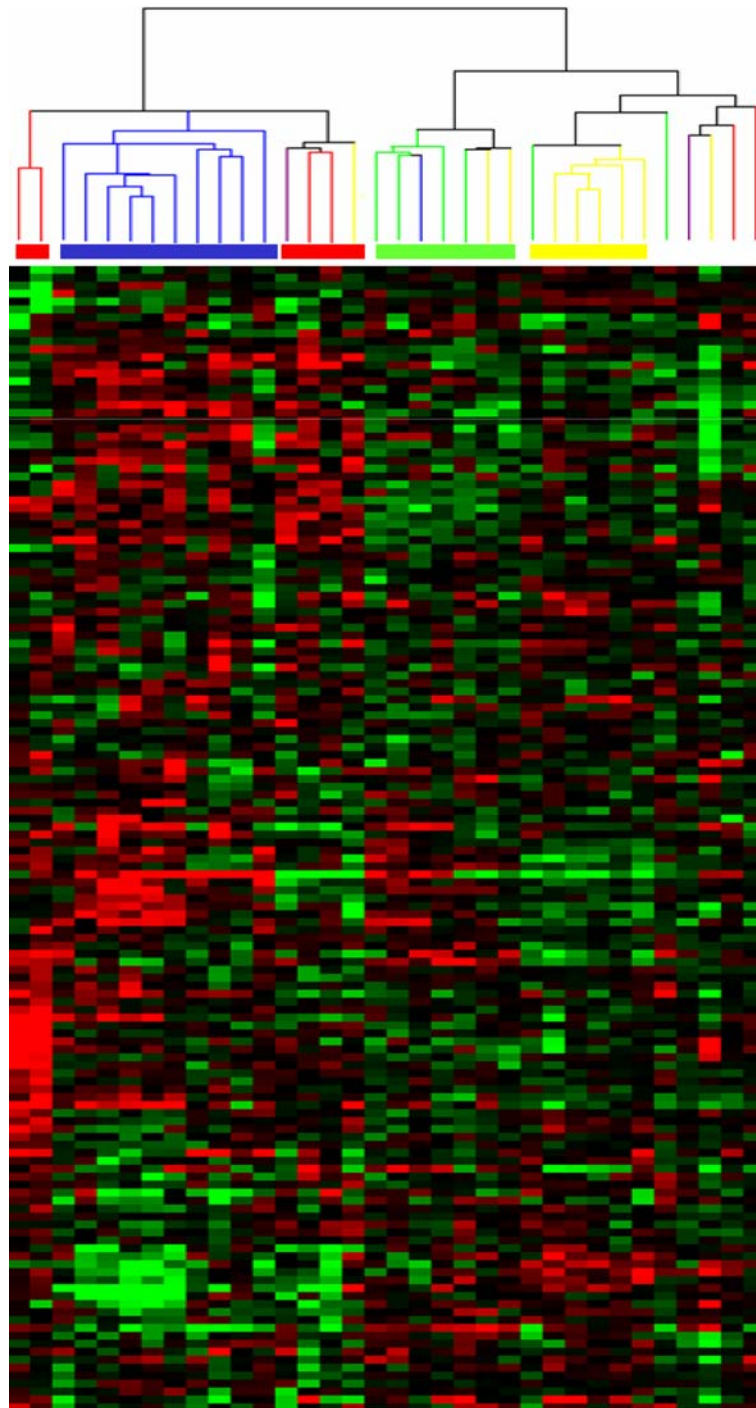


Figure 2. Hierarchical clustering revealing relationships among samples in the data set of 16 IBC and 18 non-IBC specimen using ~200 clones representing 144 genes common to the intrinsic gene set and our cDNA chip. Data are represented in a matrix format with rows representing genes and columns representing samples. The expression level of each gene is relative to its median across all samples, and is color coded red or green indicating genes above and below the median expression level, respectively. The magnitude of deviation from the median is represented by the color saturation. Grey represents genes with missing data. Based on the expression pattern of 144 genes, four out of five cell-of-origin subtypes were identified. No ErbB2-overexpressing cluster was evident, possibly due to a low number of ErbB2-overexpressing samples. The clusters are indicated by the colored branches beneath the dendrogram: luminal A (blue), normal-like (red), luminal B (green), and basal-like (yellow). Within each cluster, branches of the samples are color coded according to the expression centroid they were closest to.

12 luminal A samples and 4 out of 6 normal-like samples clustered together on terminal branches. The right branch was characterized by the presence of both the basal-like cluster and the luminal B cluster. Six out of 9 basal-like samples and 4 out of 5 luminal B samples clustered together on terminal branches. The luminal B

subgroup was again characterized by the presence of the same 2 samples, closely correlated with both the basal-like and the luminal B centroid, but designated basal-like due to a stronger correlation coefficient between these samples and the basal-like centroid. All of the samples present in the basal-like subgroup were IBC.

Overall, 25 out of 32 (78%) samples clustered on terminal branches, correctly identifying the four different cell-of-origin subtypes present in our series of breast tumors, independently of the IBC or non-IBC phenotype.

Contribution of the cell-of-origin subtypes to the IBC phenotype

To evaluate the contribution of each cell-of-origin subtype to the amount of variability seen in the entire

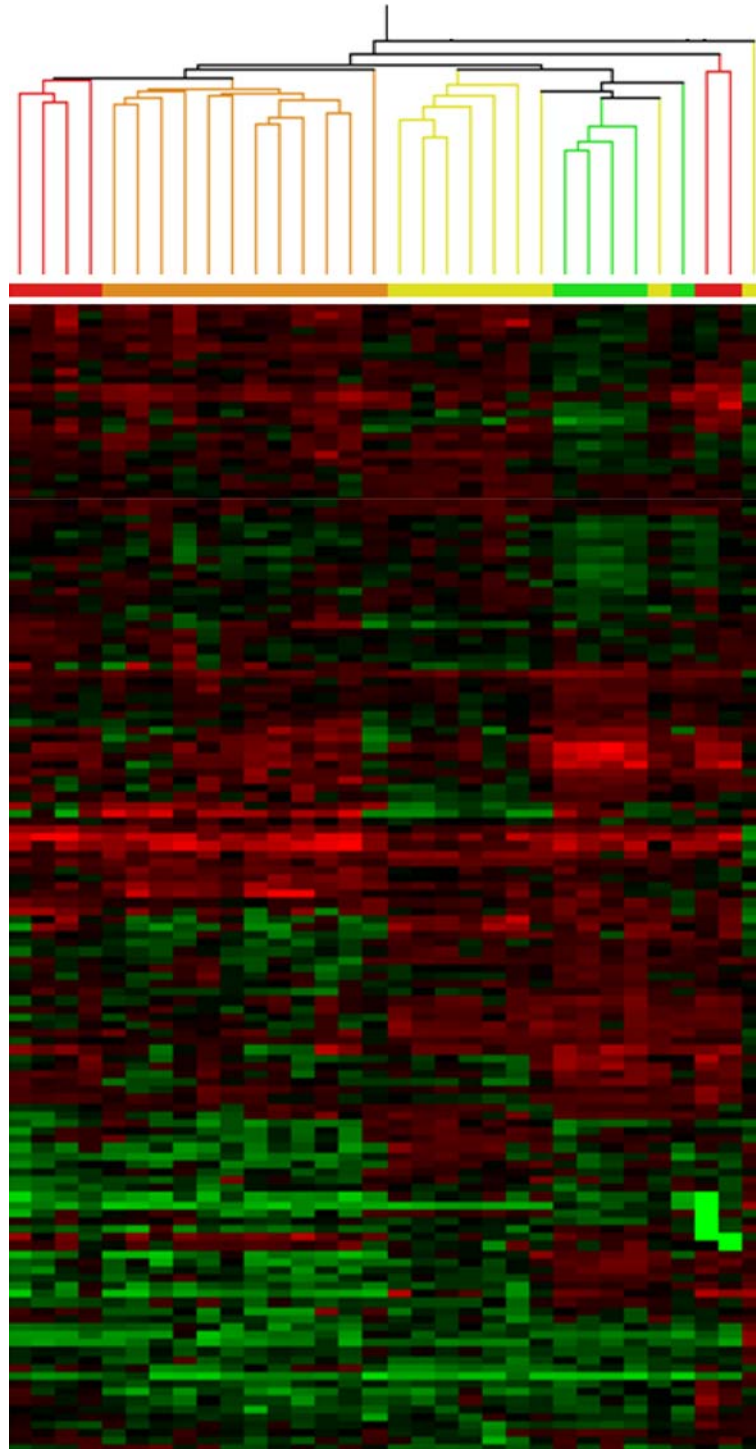


Figure 3. Hierarchical clustering revealing relationships among samples in the data set of 16 IBC and 18 non-IBC specimen using 141 genes selected by comparing four subgroups with each other by using a discriminating score and iterative random permutation testing. Data are represented in a matrix format with rows representing genes and columns representing samples. Expression level of each gene is relative to its median across all samples, and is color coded red or green indicating genes above and below the median expression level, respectively. The magnitude of deviation from the median is represented by the color saturation. Grey represents genes with missing data. Based on the expression pattern of 141 genes, four cell-of-origin subtypes were identified by using unsupervised hierarchical clustering using an alternative gene set, hereby demonstrating the robustness of the taxonomy. The clusters are indicated by the colored branches beneath the dendrogram: luminal A (orange), normal-like (red), luminal B (green), and basal-like (yellow).

data set, we executed a principal component analysis for the five molecular subtypes on a gene list with 5914 non-batch-related genes. Since there are never more principal components than there are conditions in the data, we identified five principal components (PC): PC1, PC2, PC3, PC4, and PC5 accounting for, respectively, 65.2, 17.5, 8.9, 5.5, and 3.0% of the entire variability seen in the data set. The expression patterns of each principal component are pictured in Figure 4. Four out of five principal components have an expression pattern that is clearly linked to the cell-of-origin subtypes: PC2 shows a peak expression for the normal-like cell-of-origin subtype, PC3 is high in ErbB2-overexpressing tumors and low in luminal B tumors, PC4 is high in luminal A tumors and slightly elevated in basal-like tumors, and PC5 is most pronounced in basal-like tumors. PC1 was elevated in all cell-of-origin subtypes but was not specifically associated with a particular cell-of-origin subtype. We then executed an unsupervised hierarchical clustering on our set of 16 IBC and 18 non-IBC specimens, using 719 genes with gene expression patterns strongly correlated

($r > 0.95$) to PC1. This identified two main branches, left and right. The left branch was entirely composed of IBC specimens (13/13), whereas the right branch was mainly composed of non-IBC specimens (18/21) (Pearson χ^2 , $p < 0.0001$). In contrast, unsupervised hierarchical clustering of our 16 IBC and 18 non-IBC specimens, using the entire gene set of 5914 not batch related genes did not clearly separate IBC and non-IBC specimen. This indicates that most of the variability (65.2%) seen in the entire data set of not batch related genes is attributable to the distinction between IBC and non-IBC. Overall, gene expression patterns related to the different cell-of-origin subtypes explain 34.8% of the variability seen within this data set. Hierarchical clustering using genes with gene expression patterns strongly correlated ($r > 0.95$) to PC2 revealed a cluster with normal-like samples (normal-like samples in: normal-like cluster: 4/4 versus non-normal-like cluster: 2/30; Pearson χ^2 , $p < 0.0001$). This agrees with the observation that the expression pattern of PC2 is specifically pronounced in the normal-like cell-of-origin subtype. Similarly, a luminal B cluster (luminal B

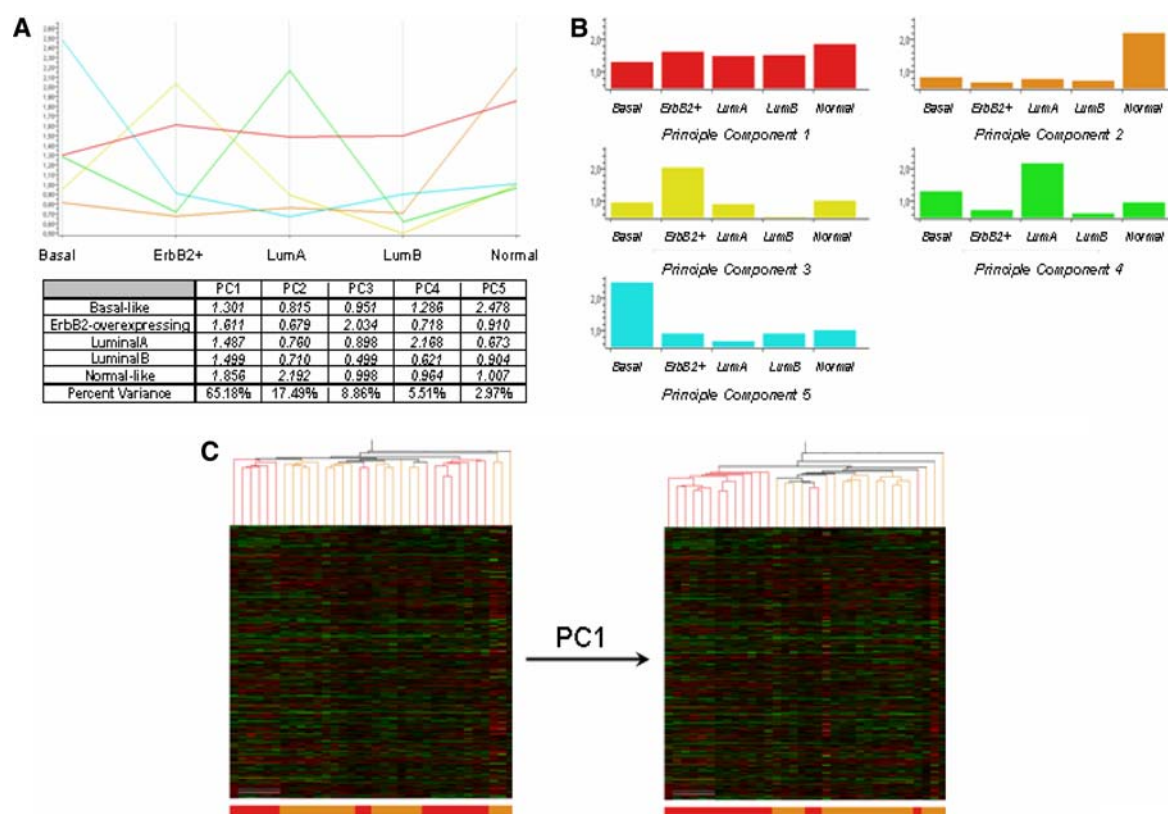


Figure 4. Principal component analysis for each of the cell-of-origin subtypes using a data set with gene expression data from 5914 genes in 16 IBC and 18 non-IBC specimens. Principal component analysis identified five principal components. The gene expression patterns and the associated gene expression data of the principal components (PC) across the different cell-of-origin subtypes are represented in (a): PC1 (red), PC2 (orange), PC3 (yellow), PC4 (green), and PC5 (blue). The principal components account for respectively 65.18, 17.49, 8.86, 5.51, and 2.97% of the total variance observed in the data set. Four out of five principal components have an expression pattern that is linked to the cell-of-origin subtypes: PC2 shows a peak expression for the normal-like cell-of-origin subtype, PC3 is high in ErbB2-overexpressing tumors and low in luminal B tumors, PC4 is high in luminal A tumors and slightly elevated in basal-like tumors, PC5 is most pronounced in basal-like tumors. PC1 was elevated in all cell-of-origin subtypes but was not specifically associated to any cell-of-origin subtype. Gene expression data for each principal component in each cell-of-origin subtype are pictured in (b). Unsupervised hierarchical clustering using a gene set with genes that are strongly correlated ($r > 0.95$) to PC1 was executed to investigate the relationships among samples present in this data set. This revealed that most of the variation seen in this data set is attributable to the distinction between IBC and non-IBC (c).

samples in: luminal B cluster: 4/6 versus non-luminal B cluster: 1/28; Fisher's Exact, $p = 0.002$), a combined basal-like and luminal A cluster (basal-like and luminal A samples in: combined basal-like and luminal A cluster: 15/16 versus combined luminal B, normal-like and ErbB2-overexpressing cluster: 6/18; Pearson χ^2 , $p < 0.0001$), and a basal-like cluster (basal-like samples in: basal-like cluster: 6/9 versus non-basal-like cluster: 3/25; Fisher's Exact, $p = 0.004$) were observed when using a gene list with genes closely correlated with, respectively, PC3, PC4, and PC5.

To evaluate the contribution of each of the cell-of-origin subtypes to the IBC phenotype, a PCA for the five molecular subtypes using a gene list with 953 genes with a statistically significant ($p < 0.05$) 1.5-fold, difference in expression in IBC and non-IBC was performed. Again five principal components were identified: PC1, PC2, PC3, PC4, and PC5 explaining, respectively, 70.54, 11.64, 10.96, 3.98, and 2.88% of the variability seen in the data set. Figure 5 shows the expression pattern associated with each of the principal components. PC3 and PC5 clearly have a peak expression in the normal-like and basal-like subgroup, respectively. PC2 has a peak expression in the ErbB2-overexpressing subgroup and a low expression in the luminal B subgroup. PC4 is characterized by a peak expression in both the luminal B and the ErbB2-overexpressing subgroups and a reduced expression in the luminal A subgroup. However, the principal component, accounting for more than 70% of the variability seen in this data set does not show a pronounced expression in any of the cell-of-origin subtypes. Therefore, the different cell-of-origin subtypes together account for less than 30% of the variability seen between IBC and non-IBC.

Discussion

In this study we have analyzed the presence of different cell-of-origin subtypes in IBC. The different cell-of-origin subtypes have been repeatedly observed in non-IBC [2,3,4] but only once in IBC [24]. Expression centroids were calculated for each of the five cell-of-origin subtypes, by using 144 genes common to our cDNA chip and the intrinsic gene set defined by Sorlie et al. [4]. Using these expression centroids, our samples were classified by calculating correlation coefficients between each sample and each centroid. Classification was validated by unsupervised hierarchical clustering and robustness was shown by classifying our samples using an alternative gene set. Our results showed that different cell-of-origin subtypes are present in IBC but the combined ErbB2-overexpressing and basal-like subtype was significantly more prevalent in IBC in comparison to non-IBC. Taking into account the poor clinical outcome associated with these cell-of-origin subtypes, this observation agrees with the fact that IBC is characterized by a poor 3-year survival rate of only 40%, compared to 85% in non-IBC [8]. Univariate survival analysis comparing the cell-of-origin subtypes with respect to overall survival showed a highly significant difference, with the basal-like and ErbB2-overexpressing clusters associated with the shortest survival times. Similar results were obtained with respect to relapse-free survival [3]. These observations were repeated in other patient cohorts by performing univariate Kaplan–Meier analysis with time to development of distant metastasis as a variable. The probability of remaining disease-free was significantly different between the subtypes, with metastasis-free survival times being significantly shorter in basal-like and ErbB2-overexpressing subtypes [4].

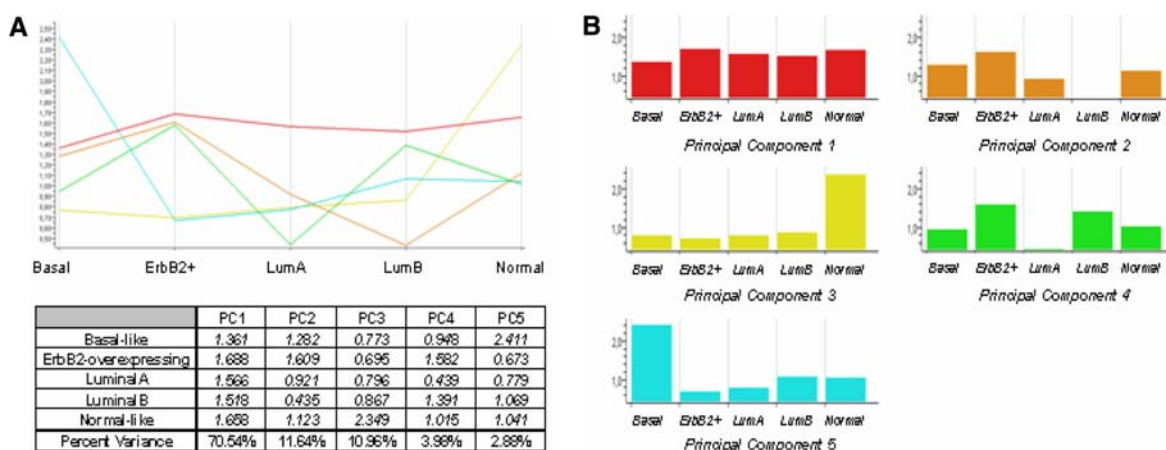


Figure 5. Principal component analysis for each of the cell-of-origin subtypes using a data set with gene expression data from 969 genes with a significant ($p < 0.05$) different expression of 1.5 between IBC and non-IBC specimens. Principal component analysis identified five principal components (PC). The gene expression patterns and the associated gene expression data of the principal components across the different cell-of-origin subtypes are represented in (a): PC1 (red), PC2 (orange), PC3 (yellow), PC4 (green), and PC5 (blue). The principal components account for respectively 70.54, 11.64, 10.96, 3.98, and 2.88% of the total variance observed in the data set. Four out of five principal components have an expression pattern that is clearly linked to the cell-of-origin subtypes: PC2 shows a peak expression for the ErbB2-overexpressing cell-of-origin subtype and low expression in the luminal B subtype, PC3 is high in the normal-like tumors, PC4 is elevated in ErbB2-overexpressing and luminal B tumors and reduced in luminal A tumors, PC5 is most pronounced in basal-like tumors. PC1 was elevated in all cell-of-origin subtypes but was not specifically associated to any cell-of-origin subtype. Gene expression data for each principal component in each cell-of-origin subtype is pictured in (b).

Kaplan–Meier survival analysis of 829 cases immunohistochemically stained for CK5/6 and cytokeratin 17 (CK17), two markers of the basal-like cell-of-origin subtype, showed that positivity for either of these basal markers correlated with shorter disease-specific survival than for negative cases. Positivity for either of these basal markers was also associated with a significantly poorer outcome in the lymph node positive group, whereas in the lymph node negative group a trend towards statistical significance was observed [27]. In the same study, HER1 protein expression was observed to correlate well with the expression of CK5/6 and CK17 and was associated with poor survival. It has been described that *EFGR* transcripts were observed more frequently in IBC than in non-IBC [8,28]. Another molecular marker implicated in breast cancer is *TP53*, mutated in approximately 30% of breast tumors. Mutations are more frequently observed in advanced-stage breast cancer and/or aggressive breast cancer. Likewise, *TP53* is mutated more often in the IBC compared to the non-IBC [8]. The fraction of tumors with *TP53* mutations was significantly different in the cell-of-origin subtypes, with higher mutation frequencies in the ErbB2-overexpressing and basal-like tumor groups [3]. Altogether, the basal-like subtype of breast cancer may represent a distinct clinical entity and is generally characterized by expression of a set of markers also expressed in IBC. Moreover, expression of *EGFR* and *KIT*, another marker associated with the basal-like cell-of-origin subtype [27], correlates with mesenchymal and myoepithelial differentiation of tumor cells [29]. This suggests the intriguing possibility that the tumors belonging to the basal-like subgroup are tumors that acquire the basal-like characteristics through a process of transdifferentiation. Plasticity of human breast carcinoma has been investigated and it has been shown that breast cancer cell lines can acquire a myofibroblastic phenotype through epithelial-to-mesenchymal transition (EMT), a process in which tumor cells lose their epithelial phenotype and acquire a more motile and invasive phenotype [30,31,32]. This can explain the poor prognosis associated with the breast basal-like subgroup. We have preliminary data, indicating that EMT is a process more active in IBC in comparison to non-IBC. It has also been shown that the transcription factor *NF-kappaB* is important for induction and maintenance of EMT [33,34]. This agrees with our previous findings indicating that *NF-kappaB* plays a major role in the biology of IBC [23].

Although the presence of different molecular subtypes in IBC and non-IBC is unquestionable, the individual classification of samples based on the expression of our gene set of 144 genes should be taken with care. When classifying the remaining samples in the Norway/Stanford data set, not considered for calculation of the expression centroids, 30% of these samples were misclassified using the centroid classification method. Particularly luminal B and normal-like samples were often closer to the wrong cell-of-origin centroid. This can be

explained by the fact that in our reduced gene set of 144 genes, few markers for the luminal B and normal-like cell-of-origin subtype are present, since no specific gene cluster for these subgroups was observed. Alternatively, one can also reason that the classification power of the expression centroids is more pronounced in more homogeneous subgroups, e.g. basal-like and ErbB2-overexpressing subtypes, than in less homogeneous subgroups, e.g. luminal A, luminal B and normal-like. While classifying the 16 IBC and 18 non-IBC samples, several samples were often significantly correlated with more than one expression centroid. This suggests a strong relationship between the cell-of-origin subtypes. Basal-like and ErbB2-overexpressing samples are usually correlated with each other and anti-correlated with luminal A samples. Conversely, luminal A samples are usually anti-correlated with basal-like and ErbB2-overexpressing samples. Luminal A and luminal B specimens are often correlated with each other and occasionally luminal B samples are correlated with basal-like samples, as indicated by the hierarchical clustering pattern (Figures 2 and 3). This can be explained by the hypothesis that basal-like and luminal B tumors arise from transformation of the most primitive ER-negative stem cell/early progenitor cell, but luminal B specimens become ER-positive through additional mutations, hence their correlation with the ER-positive luminal A subtype. Nevertheless, ER expression remains heterogeneous in luminal B samples, as evidenced by the weaker expression of the ER gene cluster [7].

Despite the presence of different cell-of-origin subtypes in IBC, as was reported by Bertucci et al. [24], we identified a higher fraction of ErbB2-overexpressing and basal-like samples and a lower fraction of luminal A, luminal B and normal-like samples in IBC in comparison with non-IBC. This is in contrast to the results published by Bertucci et al. [24], where no difference was observed between IBC and non-IBC for the combined ErbB2-overexpressing and basal-like and the combined luminal A, luminal B and normal-like clusters. This discrepancy can be, in part, explained by platform differences but the main reason is probably lack of rigidity in diagnosing IBC in the study of Bertucci et al. [24]. Diagnosis in their study is based on clinical ‘and/or’ pathological criteria, while strict adherence to the TNM classification of the American Joint Committee on cancer [9] was adopted in our study. The diagnosis based on clinical ‘and/or’ pathological criteria resulted in the absence of non-IBC tumors with dermal lymphatic emboli, whereas it is well known that dermal lymphatic invasion can be associated with all breast carcinoma subtypes. Even more, IBC can reliably be diagnosed in the absence of dermal lymphatic emboli [8]. Therefore, it might be that some non-IBC samples have been misclassified as IBC samples due to the presence of dermal lymphatic emboli. This hypothesis is even further strengthened by the observation of a high number of ER-positive IBC samples, whereas IBC is reported to be more frequently ER-negative than non-IBC. Up to 60% of IBC tumors lack hormone receptor expression,

whereas other forms of LABC are more likely to be ER-positive [8]. In the study of Bertucci et al. [24] overall, 56% (20/36) of the IBC samples showed ER-positivity, whereas only 31% (5/16) of the IBC specimen in our study were positive for ER. There was a statistical significant difference when comparing the fraction of ER-positive IBC samples in the combined luminal A, luminal B and normal-like cluster between both studies. We reported 4/8 (50%) ER-positive IBC samples in the combined luminal A, luminal B and normal-like cluster, whereas all IBC samples in the corresponding cluster reported by Bertucci et al. [24] were ER-positive (Fisher's Exact test, $p = 0.004$). This suggests that some of the ER-positive 'IBC' samples in the study by Bertucci et al. [24] might be non-IBC with emboli in dermal lymph vessels.

The identification of the cell-of-origin subtypes in the IBC gene expression pattern does not exclude the specific molecular nature of IBC. Based on the expression of 756 differentially expressed genes, we were able to separate IBC from non-IBC specimen. This clustering was not based on the expression of ER- and ErbB2-related molecules, since these genes were filtered out prior to performing the clustering analysis. A class prediction algorithm based on 50 differentially expressed genes was able to correctly classify 14/16 (86%) IBC and non-IBC specimens [23]. Bertucci et al. [24] described a 109-gene signature that discriminates IBC from non-IBC with an accuracy rate of 84% [22]. We have shown by principal component analysis, that the different cell-of-origin subtypes explain approximately 30% of the variation between IBC and non-IBC. This agrees with the observation that IBC specimens more often express markers related to the basal-like and ErbB2-overexpressing tumors, whereas non-IBC specimens more often express markers related to the combined luminal A, luminal B and normal-like cluster. If no difference would have been found between IBC and non-IBC for the cell-of-origin subtypes, markers representative for the different cell-of-origin subtypes would be absent in the gene list with differentially expressed genes in IBC and non-IBC. Hence no principal components with a cell-of-origin subtype specific expression pattern would have emerged. On the other hand, 70% of gene expression differences of IBC and non-IBC cannot be explained by cell-of-origin subtype specific expression patterns. This was corroborated by principal component analysis on a gene list not enriched in differentially expressed genes between IBC and non-IBC. In this data set, 65% of the total variance seen, was explained by other differences between IBC and non-IBC than those related to cell-of-origin subtype, as evidenced by the hierarchical clustering analysis with genes strongly correlated to the first principal component.

In conclusion, the present data support previous results indicating that IBC and non-IBC have to be regarded as separate breast cancer types [22,23], with a phenotype largely defined by a distinct gene expression profile independent of the cell-of-origin subtypes, although the latter contribute: a significantly higher frac-

tion of IBC than non-IBC belongs to the basal-like and ErbB2-overexpressing subtypes.

Acknowledgements

We would like to thank Professor Charles M. Perou of the Departments of Genetics and Pathology at the Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill for his contribution to the data analysis and for critically reading the manuscript. Gert Van den Eynden is a research assistant of the Fund for Scientific Research Flanders, Belgium. Steven Van Laere is supported by a Research Grant of the University Hospital Antwerp (University of Antwerp, Wilrijk, Belgium). Supported by grant G.0100.04 of the Fund for Scientific Research Flanders, Belgium.

References

1. Key TJ, Verkasalo PK, Banks E: Epidemiology of breast cancer. *Lancet Oncol* 2: 133–140, 2001
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: Molecular portraits of human breast tumors. *Nature* 406: 747–752, 2000
3. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98: 10869–10874, 2001
4. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D: Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 100: 8418–8423, 2003
5. Troester MA, Hoadley KA, Sorlie T, Herbert BS, Borresen-Dale AL, Lonning PE, Shay JW, Kaufmann WK, Perou CM: Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res* 64: 4218–4226, 2004
6. Wang ZC, Lin M, Wei LJ, Li C, Miron A, Lodeiro G, Harris L, Ramaswamy S, Tanenbaum DM, Meyerson M, Iglehart JD, Richardson A: Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. *Cancer Res* 64: 64–71, 2004
7. Dontu G, El-Ashry D, Wicha MS: Breast cancer, stem/progenitor cells and the estrogen receptor. *Trends Endocrinol Metab* 15: 193–197, 2004
8. Lerebours F, Bieche I, Lidereau R: Update on inflammatory breast cancer. *Breast Cancer Res* 7: 52–55, 2005
9. American Joint Committee on Cancer AJCC Cancer Staging Manual Sixth Edition. Springer Verlag, New York, Berlin, Heidelberg, 2002, pp. 221–240
10. Turpin E, Bieche I, Bertheau P, Plassa LF, Lerebours F, de Roquancourt A, Olivi M, Espie M, Marty M, Lidereau R, Vidaud M, de The H: Increased incidence of ERBB2 overexpression and TP53 mutation in inflammatory breast cancer. *Oncogene* 21: 7593–7597, 2002
11. Van Golen KL, Davies S, Wu ZF, Wang Y, Bucana CD, Root H, Chandrasekharappa S, Strawderman M, Ethier SP, Merajver SD: A novel putative low-affinity insulin-like growth factor binding protein, LIBC (lost in inflammatory breast cancer), and RhoC GTPase correlate with the inflammatory breast cancer phenotype. *Clin Cancer Res* 5: 2511–2519, 1999

12. Van den Eynden GG, Van der Auwera I, Van Laere S, Colpaert CG, van Dam P, Merajver S, Kleer CG, Harris AL, Van Marck EA, Dirix LY, Vermeulen PB: Validation of a tissue microarray to study differential protein expression in inflammatory and non-inflammatory breast cancer. *Breast Cancer Res Treat* 85: 13–22, 2004
13. Van Golen KL, Wu ZF, Qiao XT, Bao LW, Merajver SD: RhoC GTPase, a novel transforming oncogene for human mammary epithelial cells that partially recapitulates the inflammatory breast cancer phenotype. *Cancer Res* 60: 5832–5838, 2000
14. Van Golen KL, Wu ZF, Qiao XT, Bao L, Merajver SD: RhoC GTPase overexpression modulates induction of angiogenic factors in breast cells. *Neoplasia* 2: 418–425, 2000
15. Van Golen KL, Bao LW, Pan Q, Miller FR, Wu ZF, Merajver SD: Mitogen activated protein kinase pathway is involved in RhoC GTPase induced motility, invasion and angiogenesis in inflammatory breast cancer. *Clin Exp Metast* 19: 301–311, 2002
16. Van Golen KL, Bao L, DiVito MM, Wu Z, Prendergast GC, Merajver SD: Reversion of RhoC GTPase-induced inflammatory breast cancer phenotype by treatment with a farnesyl transferase inhibitor. *Mol Cancer Ther* 1: 575–583, 2002
17. Kleer CG, Zhang Y, Pan Q, van Golen KL, Wu ZF, Livant D, Merajver SD: WISP3 is a novel tumor suppressor gene of inflammatory breast cancer. *Oncogene* 21: 3172–3180, 2002
18. Kleer CG, Zhang Y, Pan Q, Gallagher G, Wu M, Wu ZF, Merajver SD: WISP3 and RhoC guanosine triphosphatase cooperate in the development of inflammatory breast cancer. *Breast Cancer Res* 6: R110–115, 2004
19. Kleer CG, Zhang Y, Pan Q, Merajver SD: WISP3 (CCN6) is a secreted tumor-suppressor protein that modulates IGF signaling in inflammatory breast cancer. *Neoplasia* 6: 179–185, 2004
20. Colpaert CG, Vermeulen PB, Benoy I, Soubry A, van Roy F, van Beest P, Goovaerts G, Dirix LY, van Dam P, Fox SB, Harris AL, van Marck EA: Inflammatory breast cancer shows angiogenesis with high endothelial proliferation rate and strong E-cadherin expression. *Br J Cancer* 88: 718–725, 2003
21. Van der Auwera I, Van Laere SJ, Van den Eynden GC, Benoy I, van Dam P, Colpaert CG, Fox SB, Turley H, Harris AL, Van Marck EA, Vermeulen PB, Dirix LY: Increased angiogenesis and lymphangiogenesis in inflammatory versus noninflammatory breast cancer by real-time reverse transcriptase-PCR gene expression quantification. *Clin Cancer Res* 10: 7965–7971, 2004
22. Bertucci F, Finetti P, Rougemont J, Charafe-Jauffret E, Nasser V, Lloriod B, Camerlo J, Tagett R, Tarpin C, Houvenaeghel G, Nguyen C, Maraninchi D, Jacquemier J, Houlgatte R, Birnbaum D, Viens P: Gene expression profiling for molecular characterization of inflammatory breast cancer and prediction of response to chemotherapy. *Cancer Res* 64: 8558–8565, 2004
23. Van Laere SJ, Van der Auwera I, Van den Eynden GG, Fox SB, Bianchi F, Harris AL, van Dam P, Van Marck EA, Vermeulen PB, Dirix LY: Distinct molecular signature of inflammatory breast cancer by cDNA microarray analysis. *Breast Cancer Res Treat*, accepted for publication, 2005
24. Bertucci F, Finetti P, Rougemont J, Charafe-Jauffret E, Cervera N, Tarpin C, Nguyen C, Xerri L, Houlgatte R, Jacquemier J, Viens P, Birnbaum D: Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. *Cancer Res* 65: 2170–2178, 2005
25. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531–537, 1999
26. Magrangeas F, Nasser V, Avet-Loiseau H, Lloriod B, Decaux O, Granjeaud S, Bertucci F, Birnbaum D, Nguyen C, Harousseau JL, Bataille R, Houlgatte R, Minvielle S: Gene expression profiling of multiple myeloma reveals molecular portraits in relation to the pathogenesis of the disease. *Blood* 101: 4998–5006, 2003
27. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, Akslen LA, Ragaz J, Gown AM, Gilks CB, van de Rijn M, Perou CM: Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 10: 5367–5374, 2004
28. Guerin M, Gabillot M, Mathieu MC, Travagli JP, Spielmann M, Andrieu N, Riou G: Structure and expression of c-erbB-2 and EGF receptor genes in inflammatory and non-inflammatory breast cancer: prognostic significance. *Int J Cancer* 43: 201–208, 1989
29. Tsuda H, Morita D, Kimura M, Shinto E, Ohtsuka Y, Matsubara O, Inazawa J, Tamaki K, Mochizuki H, Tamai S, Hiraide H: Correlation of KIT and EGFR overexpression with invasive ductal breast carcinoma of the solid-tubular subtype, nuclear grade 3, and mesenchymal or myoepithelial differentiation. *Cancer Sci* 96: 48–53, 2005
30. Vincent-Salomon A, Thiery JP: Host microenvironment in breast cancer development: epithelial-mesenchymal transition in breast cancer development. *Breast Cancer Res* 5: 101–106, 2003
31. Deugnier MA, Faraldo MM, Janji B, Rousselle P, Thiery JP, Glukhova MA: EGF controls the *in vivo* developmental potential of a mammary epithelial cell line possessing progenitor properties. *J Cell Biol* 159: 453–463, 2002
32. Petersen OW, Lind Nielsen H, Gudjonsson T, Villadsen R, Ronnov-Jessen L, Bissell MJ: The plasticity of human breast carcinoma cells is more than epithelial to mesenchymal conversion. *Breast Cancer Res* 3: 213–217, 2001
33. Huber MA, Beug H, Wirth T: Epithelial-mesenchymal transition: NF-kappaB takes center stage. *Cell Cycle* 3: 1477–1480, 2004
34. Huber MA, Azoitei N, Baumann B, Grunert S, Sommer A, Pehamberger H, Kraut N, Beug H, Wirth T: NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J Clin Invest* 114: 569–581, 2004

Address for offprints and correspondence: Dr. Peter B Vermeulen, Department Pathology-AZ Sint-Augustinus, Oosterveldlaan 24, B2610 Wilrijk, Belgium; *Tel.:* + 323-443-52-39; *Fax:* + 323-443-30-36; *E-mail:* Peter.Vermeulen@GVAgroup.be