Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer

Juan M. Cejalvo1, 2, Eduardo Martínez de Dueñas3, Patricia Galván4, Susana García-Recio1, Octavio Burgués Gasión5, Laia Paré1, Silvia Antolín6, Rosella Martinello1, Isabel Blancas7, Barbara Adamo1, Ángel Guerrero-Zotano8, Montserrat Muñoz1, Paolo Nuñíforo4, Maria Vidal1, Ramón M. Pérez9, Jose I. Chacón López-Muniz10, Rosalia Caballero11, Vicente Peg12 Eva Carrasco11, Federico Rojo13, Charles M. Perou14, Javier Cortés4, 15, Vincenzo Adamo16, Joan Albanell17, Roger R. Gomis2, 18, Ana Lluch19 and Aleix Prat1, 4.

1Translational Genomics and Targeted Therapeutics in Solid Tumors, August Pi i Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain.
2Oncology Program, Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain.
3Hospital Provincial de Castellón, Castellón, Spain.
4Translational Genomics Group, Vall d’Hebron Institute of Oncology (VHIO), Barcelona, Spain.
5Department of Pathology. Hospital Clínico Universitario de Valencia, Spain
6Complejo Hospitalario Universitario A Coruña, A Coruña, Spain
7Hospital Clínico San Cecilio. Complejo Hospitalario de Granada, Granada, Spain.
8Instituto Valenciano de Oncología, Valencia, Spain.
9Hospital Universitario Quirón de Madrid, Madrid, Spain
10Hospital Virgen de la Salud, Toledo, Spain.
11GEICAM, Spanish Breast Cancer Group, Madrid, Spain.
12Pathology Department, Hospital Vall d’Hebron
13Fundación Jiménez Díaz, Madrid, Spain
14University of North Carolina, Chapel Hill, NC, United States
15Ramón y Cajal University Hospital, Madrid
16University of Messina, Messina, Italy
17Hospital del Mar, Barcelona, Spain
18ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain.
Running Title: Primary versus metastatic breast cancer

Key words: primary, metastasis, breast cancer, intrinsic subtype, gene expression
Financial support: This work was supported by funds from Instituto de Salud Carlos III - PI13/01718 (A.P.), by a Career Catalyst Grant from the Susan Komen Foundation (A.P.), by Banco Bilbao Vizcaya Argentaria (BBVA) Foundation (A.P.) and by the Breast Cancer Research Foundation. This work was also supported by funds from FEDER (RETICC): RD12/0036/0076 (JA), RD12/0036/0051 (JA), RD12/0036/0070 (AL) and RD12/0036/0076 (MM). JMC holds a fellowship from “PhD4MD”, a Collaborative Research Training Programme for Medical Doctors (IDIBAPS, August Pi i Sunyer Institute for Biomedical Research and IRB Barcelona, Institute for Research in Biomedicine) partially funded by Instituto de Salud Carlos III, (ISCIII, project: II14/00019). JMC and RRG research support is provided by the Spanish Ministry of Science and Innovation grant SAF2013-46196 (FEDER Funds) and the Generalitat de Catalunya AGAUR 2014-SGR grant 535. The funders did not have any role in the study design, conduct, or decision to submit the manuscript for publication. CERCA Programme / Generalitat de Catalunya

#, corresponding author:

Aleix Prat, MD PhD
Department of Medical Oncology
Hospital Clinic de Barcelona
Casanova 170, 08036
Barcelona, Spain
(+34) 93 227 54 00
Email: alprat@clinic.cat

Competing interests. Advisory role of A.P. and M.M. for Nanostring Technologies. The other authors declare that they have no competing interests.

Word Count: 3342
Number of Figures: 5
Number of Tables: 2
Abstract

Biological changes that occur during metastatic progression of breast cancer are still incompletely characterized. In this study, we compared intrinsic molecular subtypes and gene expression in 123 paired primary and metastatic tissues from breast cancer patients. Intrinsic subtype was identified using a PAM50 classifier and Chi-square tests determined the differences in variable distribution. The rate of subtype conversion was 0% in basal-like tumors, 23.1% in HER2-enriched (HER2-E) tumors, 30.0% in luminal B tumors and 55.3% in luminal A tumors. In 40.2% of cases, luminal A tumors converted to luminal B tumors, whereas in 14.3% of cases luminal A and B tumors converted to HER2-E tumors. We identified 47 genes that were expressed differentially in metastatic versus primary disease. Metastatic tumors were enriched for proliferation-related and migration-related genes and diminished for luminal-related genes. Expression of proliferation-related genes were better at predicting overall survival in metastatic disease (OSmet) when analyzed in metastatic tissue rather than primary tissue. In contrast, a basal-like gene expression signature was better at predicting OSmet in primary disease compared to metastatic tissue. We observed correlations between time to tumor relapse and the magnitude of changes of proliferation, luminal B or HER2-E signatures in metastatic versus primary disease. Although the intrinsic subtype was largely maintained during metastatic progression, luminal/HER2-negative tumors acquired a luminal B or HER2-E profile during metastatic progression, likely reflecting tumor evolution or acquisition of estrogen independence. Overall, our analysis revealed the value of stratifying gene expression by both cancer subtype and tissue type, providing clinicians more refined tools to evaluate prognosis and treatment.
Introduction

Despite new systemic treatment advances, metastatic breast cancer is still an incurable disease and a major cause of cancer death (1). Median overall survivals of patients with triple-negative, hormone receptor (HR)-positive/HER2-negative and HER2-positive diseases are $\sim 12$, $\sim 20$ and $\sim 56$ months, respectively (2-5). Thus, there is a need to better understand the biology behind the progression of tumor cells towards metastasis. To date, evidence suggests that both intrinsic properties of breast cancer cells and host organ microenvironment participate actively to this matter (6).

In general, detectable distant breast cancer metastases occur years, or even decades, after primary tumor diagnosis. These secondary lesions are believed to originate from disseminated tumor cells that underwent a period of dormancy (7). Tumor dormancy is the result of equal rates of cell proliferation and cell death (8). However, the molecular factors that promote the formation of detectable metastasis from disseminated tumor cells are largely unknown. To try to approach it, studies have started to identify the molecular differences between primary tumor and their matched metastatic lesions (9). At the DNA level, although significant differences have been observed, the vast majority (80-85%) of molecular alterations are similar in both settings. For example, the discordance of HER2 gene amplification by FISH in primary versus metastatic tissue is 3-10% (10). Similarly, at the protein level, estrogen and progesterone receptors by immunohistochemistry (IHC) are discordant in 13-28% of the cases (11). Overall, these results suggest that minor but important molecular changes occur during metastatic progression such as ESR1 mutations (12).

In terms of global gene expression, 4 main molecular subtypes (Luminal A, Luminal B, HER2-enriched [HER2-E] and Basal-like), and a normal breast-like group, have been identified and intensively studied for the last 15 years in early breast cancer
(13-16). Known as the ‘intrinsic subtypes of breast cancer’, these groups of tumors have revealed critical differences in incidence (17, 18), survival (19-21), and response to treatment (22, 23). Importantly, the information provided by the intrinsic subtypes complements and expands the information provided by classical clinical parameters (e.g. age, node status, tumor size, histological grade) and pathological markers (estrogen receptor [ER], progesterone receptor [PR] and human epidermal growth factor receptor 2 [HER2]) (24, 25), all of which are routinely used today in the clinic to stratify patients for prognostic predictions and to select treatments.

Recently, we evaluated the prognostic value of the intrinsic subtypes in 821 samples from a phase 3 clinical trial, where postmenopausal patients with metastatic HR+ breast cancer were treated with first-line letrozole +/- lapatinib (26). The vast majority of samples (~80%) were from the primary tumor years before the patient relapsed. Interestingly, we observed that intrinsic subtype provided the largest amount of prognostic information in this setting beyond HER2 status, treatment, visceral disease and other clinical-pathological variables. Overall, these results suggested that intrinsic subtype does not change substantially when recurrence occurs.

**Methods and Material**

*Study population*

This retrospective study included non-consecutive female patients over the age of 18 with a histological diagnosis of metastatic breast cancer detected at the time of diagnosis, at first relapse or after successive disease progressions. Tissues were collected from 5 independent sources: GEICAM/2009-03 ConvertHER trial (11), Hospital Clínico Universitario de Valencia, Vall d’Hebrón Institute of Oncology, University-AO Papardo and Hospital Clinic of Barcelona. To be included, samples were
required to have a formalin-fixed paraffin-embedded (FFPE) tissue sample from primary and metastatic tumor. Biopsies were performed by core biopsy or surgical process, according to the routine clinical practice of the hospitals. For each sample, receptor status (ER, PR and HER2) were analysed at the local laboratory.

*Gene expression analysis*

All primary and metastatic tissues were analysed using the same methodology. A section of FFPE breast tissue was first examined with a hematoxylin and eosin staining to confirm the diagnosis and determine the tumour surface area. For RNA purification, three 10μm FFPE slides were cut for each tumour, and macrodissection was performed, when needed, to avoid normal breast contamination. A minimum of ∼100 ng of total RNA was used to measure the expression of 105 breast cancer-related genes and 5 house-keeping genes (ACTB, MRPL19, PSMC4, RPLP0 and SF3A1) using the *nCounter* platform (Nanostring Technologies, Seattle, WA, US) (27). Data was log base 2 transformed and normalized using the house-keeping genes. Raw data have been deposited in the Gene Expression Omnibus under the accession number GSE92977.

*Gene list*

The list of 105 breast cancer-related genes includes genes from the following 3 signatures: PAM50 intrinsic subtype predictor (n=50) (28), Claudin-low subtype predictor (n=43)(29), VEGF/Hypoxia signature (n=13) (30). In addition, we included 8 individual genes that have been found to play an important role in breast cancer (i.e. CD24 (29), CRYAB (31), ERBB4 (32), PIK3CA (13), PTEN (13), RAD17 (33), RAD50 (33), and RB1 (13)). The complete list of genes can be found in ([Table S1](#)).
**Intrinsic subtype**

All tumors were assigned to an intrinsic molecular subtype of breast cancer (Luminal A, Luminal B, HER2-E, and Basal like) and the Normal-like group using the previously reported PAM50 subtype predictor (28).

**Claudin-low intrinsic subtype**

We applied the previously reported 9-Cell line claudin-low predictor (29). A sample was identified as Claudin-low independently of the PAM50 subtype call, as previously reported.

**Gene signatures**

The expression of 10 independent signatures was evaluated as a continuous variable. The PAM50 predictor calculates, for each sample, the correlation coefficient to each of the 5 PAM50 centroids (Luminal A, Luminal B, Basal-like, HER2-enriched and Normal-like). Each centroid was considered a single signature. In addition, the PAM50 predictor outputs a Risk Of Recurrence (ROR) score at 10-years. The ROR score based on subtype (ROR-S) and subtype and proliferation (ROR-P) were developed in a microarray-based cohort of node-negative, untreated early breast cancer (34). In addition, we evaluated the following 3 signatures: proliferation score, which is the mean expression of 11 proliferation-related genes (21), VEGF/Hypoxia signature (30), which is the mean expression of 13 hypoxia-related genes, and Claudin-low signature (29) (as a continuous variable).

**Statistical analysis**

Chi-square tests were performed to determine the differences in the distribution
of variables. To identify genes whose expression was significantly different between paired primary and metastatic samples, we used a paired two-class Significance of Microarrays (SAM) with a False Discovery Rate (FDR) <5% (35). Biologic analysis of gene lists was performed with DAVID annotation tool (http://david.abcc.ncifcrf.gov/) (36). Time to tumor recurrence (TTR) was defined as the period of time from surgery to the date of the first distant relapse. Overall survival from metastatic disease (OSmet) was defined as the period of time of metastatic disease to death or last follow-up. Estimates of survival were from the Kaplan-Meier curves and tests of differences by the log-rank test. Univariate Cox-models were used to test the independent prognostic significance of each variable. All statistical computations were carried out in R v2.15.1 (http://cran.r-project.org). All statistical tests were two sided, and the statistical significance level was set to less than 0.05.

Results

Clinical-pathologic characteristics

A total of 123 patients were included (Table S2). The median age at breast cancer diagnosis was 52.5 years (range, 28-90). In primary disease, the immunohistochemical analyses showed 73.17% (n=90) of patients had HR-positive (HR+), 15.45% (n=19) HER2-positive (HER2+) and 9.76% (n=12) triple-negative disease. In metastatic disease, 69.92% (n=86) of patients had HR+, 19.51% (n=24) HER2+ and 9.76% (n=12) triple-negative disease. No significant differences (p>0.502) were observed in the distribution of the three IHC groups in primary vs. metastatic disease. Fourteen patients (11.38%) presented with de novo metastatic disease. Median follow-up and OSmet were 76.5 and 84 months, respectively (Fig. S1).
Type of metastatic tissues

The organs of origin of the metastatic biopsies analysed in this study were skin (n=35; 28.4%), lymph nodes (n=24; 19.5%), liver (n=20; 16.3%), bone (n=16; 13%), lung (n=7; 5.7%), ovarian and peritoneum (n=7; n=5.7%), pleural (n=6; 4.9%) and others (n=8; 6.5%), including brain, pericardial fluid and colon metastases (Fig. S2).

Subtype distribution

The distribution of the PAM50 intrinsic subtype classification in primary tumor vs. metastatic disease was 39% vs. 26% for Luminal A (p=0.029), 26% vs. 35.8% for Luminal B (p=0.097), 11.4% vs. 22% for HER2-E (p=0.026) and 9.8% vs. 12.2% for Basal-like tumors (p=0.540) (Fig. 1). Individually, subtype concordance was high for Basal-like (100%), HER2-E (76.9%) and Luminal B (70.0%) tumors (Table 1). Regarding Luminal A primary tumors, 44.7% remained Luminal A in the metastasis, switching to Luminal B in 40.4% and HER2-E in 14.9% of the cases. Overall, primary luminal tumors (A and B combined) changed to a HER2-E in 14.28%, despite 81% of them being clinically HER2 negative. Cohen’s kappa coefficient was 0.38 (95% CI 0.27 to 0.5, p<0.001). These results were not affected when the Claudin-low classification was investigated since no Claudin-low tumor was identified in this series. Finally, we observed that liver and lung metastases showed the highest and lowest subtype conversion rate (75% and 14%), respectively. However, these results by site of metastasis need further validation due to the small sample sizes (Table S3).

Expression changes of individual signatures

We evaluated the expression changes of each individual signature between primary and their metastatic samples. Luminal A and normal-like signatures were found
significantly less expressed in metastatic tumors than in primary tumor (Fig. 2). In contrast, Luminal B, HER2-E and proliferation signatures were found more expressed in metastatic tumors than in primary tumors. Finally, the expression of Basal-like, VEGF/Hypoxia and Claudin-low signatures was similar between primary and metastatic disease (Fig. 2).

Expression changes of individual genes

Among 105 breast cancer-related genes, 16 and 31 genes were found up- and down-regulated in metastatic tissues compared to primary tissues (FDR<5%) (Table 2). The up-regulated gene list was enriched for genes involved in survival and migration (e.g. FGFR4), cell cycle (e.g. CDC6 and CCNB1) and DNA repair (e.g. TYMS). The down-regulated gene list was enriched for genes involved in response to hormone stimulus (e.g. BCL2 and PGR) (Fig. 2), differentiation (e.g. GATA3) and chromatin regulation (e.g. CXXC5).

A similar analysis was performed within each of the subtypes identified in primary disease. Concordant with the subtype changes, 25, 8, 7 and 0 genes were found differentially expressed in Luminal A, Luminal B, HER2-E and Basal-like primary disease, respectively, compared to metastatic disease (Table S4).

Association with overall survival

We evaluated the ability of the 10 signatures to predict OSmet in primary (Fig. 3A) vs. metastatic (Fig. 3B) disease. Interestingly, no signature consistently predicted OSmet in both primary and metastatic disease. In primary disease, Basal-like signature was found associated with worse outcome (Hazard ratio = 1.50, p=0.007), while the VEGF/Hypoxia signature was associated with a better outcome (Hazard ratio = 0.65,
p=0.016). In metastatic disease, proliferation was found associated with worse outcome (Hazard ratio 1.40, p=0.047).

These results suggested that OSmet might be better predicted by measuring either the primary tumor or the metastatic tumor depending on the biological process (e.g. proliferation) being evaluated. To further explore this, we evaluated the ability of each individual gene to predict OSmet in primary vs. metastatic disease. Among 105 genes, 14 and 10 genes were found associated with OSmet in primary and metastatic disease, respectively. Interestingly, only 1 gene (GATA3) consistently predicted favorable outcome in both settings (Fig. 4). In primary disease, high expression of 13 of the 14 genes was found associated with better outcome. These 13 genes (e.g. PGR, ESR1 and FOXA1) were mostly tracking luminal-related biological processes. On the contrary, high expression of 8 of the 10 genes in metastatic disease was found associated with worse outcome. These 8 genes (e.g. MYC, CCNE1 and CCNB1) were mostly tracking cell cycle/proliferation-related biological processes.

Finally, we explored the ability of each gene signature to predict OSmet in patients with tumors with no subtype conversion (n=59) versus patients with tumors without subtype conversion (n=49). The results revealed that in patients with no subtype conversion, the associations of signatures with OSmet were very similar when the primary or the metastatic tumors were evaluated. In patients with subtype conversion, the associations of signatures with OSmet were generally different when the primary or the metastatic tumors were evaluated. Among them, the HER2-E signature was found significantly associated with poor outcome (Hazard ratio = 1.86, p=0.046) when evaluated in metastatic tumors but not when evaluated in primary disease (Fig. S3).

Magnitude of gene expression changes vs. time to tumor recurrence
To evaluate if the gene expression changes observed in metastatic tissues are a reflection of tumor evolution over time, we plotted the magnitude of change of the expression of each signature versus time to tumor recurrence (TTR) (Fig. 5). The results revealed a positive correlation between TTR and HER2-E (corr=0.324, p<0.001), Luminal B (corr=0.27, p=0.004), Proliferation score (corr=0.291, p=0.002) and ROR-P (corr=0.295, p=0.001). In contrast, Normal-like and Luminal A signatures showed a negative correlation with TTR (corr=-0.285, p=0.002; corr=-0.219, p=0.019, respectively).

Gene by gene analysis revealed a positive correlation between TTR and the magnitude of change of genes implicated in cell proliferation (CEP55: corr=0.244, p=0.024), mitogenesis and differentiation biological process (FGFR4: corr=0.211, p=0.044). In contrast, a negative correlation was observed with genes that participate in cell-to-cell adhesion (CLDN4: corr=-0.207, p=0.027; F11R: corr=-0.237, p=0.01), regulation of DNA damage repair (RAD17: corr: -0.226, p= 0.017), tumor suppression (GRHL2: corr=-0.186, p=0.05), mammary gland development (PGR: corr=-0.203, p=0.045), and that attenuate cell migration (ESRP1: corr=-0.252, p=0.006).

**Discussion**

Here, we explored RNA-based expression differences between paired primary and metastatic breast tumors and made the following observations: 1) intrinsic molecular subtype is largely maintained during metastatic recurrence, except for luminal A disease which converted to Luminal B and HER2-enriched in 55% of the cases; 2) metastatic tissues show higher expression of proliferative and lower expression of luminal-related genes compared to primary tumors, except for Basal-like disease which seems to be very stable from a RNA-based perspective; 3) different biological
processes can predict overall survival from recurrence when evaluated in primary vs. metastatic disease; 4) an intriguing relationship seems to exist between the time taken to develop detectable metastases and the aggressiveness of the tumor, indicating that a tumor might evolve towards a more aggressive phenotype as time evolves.

Previous studies have evaluated the rates of change of the three classical pathological biomarkers (i.e. ER, PR and HER2) between primary and metastatic tumors (37, 38). Overall, the rates of ER, PR and HER2 conversion were 13%, 28% and 3-10%, respectively (11). Among the 3 genes, we also observed PGR as the top downregulated gene in metastatic compared to primary tissues. Nonetheless, the three classical biomarkers are largely maintained in the metastatic setting which is concordant with our findings using the Basal-like, HER2-enriched and Luminal A/B intrinsic subtype classification. At the same time, prior gene expression-based studies with smaller number of patients are concordant with our findings (39-41). However, Lee and colleagues evaluated the PAM50 intrinsic subtypes in 17 paired samples of primary and brain metastasis, and subtype conversion was observed in 47.1% of the cases, which is higher than the 30.9% conversion rate observed in our study. However, similar to our study, a large proportion of Luminal A primary tumors (1 out of 6) changed to non-Luminal A disease, and all Basal-like primary tumors (n=6) remaining Basal-like at recurrence (42).

Other studies have evaluated changes in somatic mutations and gene copy-number aberrations (CNA) between primary and metastasis. For example, Meric-Bernstam and colleagues (43) performed targeted DNA sequencing of 3,320 exons of 182 cancer-related genes plus 37 introns from 14 genes in 74 tumors. In 33 matched primary and recurrent tumors, 97 of 112 (86.6%) somatic mutations were concordant. Of identified CNAs, 136 of 159 (85.5%) were concordant. There was an increased
frequency of CDK4/MDM2 amplifications in recurrences, as well as gains and losses of other actionable alterations. The authors concluded that analysis of recurrent tumors before treatment may provide additional insights, as both gains and losses of targets are observed. In another study, Ding and colleagues (41) described the genomic analyses of four DNA samples from an African-American patient with Basal-like breast cancer: peripheral blood, the primary tumor, a brain metastasis, and a xenograft derived from the primary tumor. Of the 50 validated point mutations and small indels, 48 were detectable in all three tumors. Overall, while additional somatic mutations, copy number alterations, and structural variations occurred during the clinical course of the disease, most of the original mutations and structural variants present in the primary tumor were propagated.

Similar to prior studies looking at DNA alterations, we did not identify large absolute expression changes at the RNA level between primary and metastatic disease. Nonetheless, 47 genes were found differentially expressed, mostly within luminal A/B disease. Among them, FGFR4 was detected as the top upregulated gene in metastatic disease. Interestingly, this gene is found in the PAM50 gene list and its overexpression is characteristic of the HER2-E intrinsic subtype. Fibroblast growth factor receptors are involved in development, differentiation, cell survival, migration, angiogenesis and carcinogenesis (44). Dimerization of the receptor leads to intracellular phosphorylation of receptor kinase domains and intracellular signal transduction, including RAS/RAF/MEK and PI3K/AKT pathways (45). These evidences suggest that FGFR4 could drive the HER2-E phenotype in metastatic lesions with a HER2-negative/HER2-E profile. Indeed, we observed that the 8 patients whose tumors changed from Luminal A/B in primary disease to HER2-E in metastatic disease showed an increase in FGFR4 expression but not ERBB2 expression (Fig. S4). Of note, HER2-E subtype has been
associated with estrogen-independent growth and poor outcome in patients with HR+/HER2-negative breast cancer treated with anti-estrogens (46, 47). Further mechanistic studies are needed to elucidate the role of FGFR4 in metastatic disease.

Currently, large phase III clinical trials, especially within HR+/HER2-negative disease, are not taking into account this biological heterogeneity such as proliferation which is not well captured by HR and HER2 statuses. For example, patients with a Luminal A profile following endocrine therapy might be treated with second-line endocrine therapy while those that change to a HER2-negative/HER2-E or Luminal B profile might be treated with chemotherapy or other novel combinatory strategies such as endocrine therapy and CDK4/6 inhibition. Overall, this result suggests that, although there is some stability of the intrinsic subtype, ∼40% of the tumors will change subtype, highlighting the need to biopsy metastatic disease in order to better understand the clinical and biological evolution of a tumor.

Another interesting observation was the significant correlation between the magnitude of gene expression changes of various signatures between primary and metastasis disease and the time from diagnosis to tumor recurrence. Specifically, we observed that the longer the time to recurrence, the more aggressive the tumors become based on proliferation and expression of luminal genes. This suggests that there is an intrinsic evolution of tumor cells towards a more aggressive phenotype as time elapses. However, the correlation coefficients were weak and thus the magnitude of gene expression changes might also be explained by other variables such as the treatments received in (neo)adjuvant setting.

This study has several limitations worth noting. First, this is a retrospective study using tumor samples from different hospitals and a selection bias is plausible. Second, patients received different adjuvant and/or metastatic systemic treatments and
thus we could not evaluate treatment effects on tumor biology or survival. However, subtype conversion of the 14 patients with \textit{de novo} metastatic disease was found to be 57.1\%, suggesting that subtype conversion is independent of treatment effects. More studies are needed to address this particular question. Third, metastatic tumor biopsies were not always collected at the time of the diagnosis of recurrent disease. Fourth, we did not analyse DNA mutations such as ESR1 whose incidence is known to increase during tumor progression (12). Further studies will be able to evaluate if the gene expression changes observed during progression of luminal breast cancer are related to the appearance of ESR1 mutations.

To conclude, most biological changes occurring during metastatic progression of breast cancer are largely unknown today. Here, we compared intrinsic molecular subtype and expression of individual genes in paired primary and metastatic tissues. Our results suggest that although intrinsic subtype is largely maintained during metastatic progression, luminal/HER2-negative tumors acquire a Luminal B or HER2-E profile during metastatic progression, likely reflecting tumor evolution and/or acquisition of estrogen-independency. Moreover, our study highlights the importance of molecular characterization of metastatic disease.
References


Table 1. Subtype concordance between primary and metastatic disease.

<table>
<thead>
<tr>
<th>Primary disease</th>
<th>Metastatic disease</th>
<th>Genes differentially expressed (FDR&lt;5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basal-like</td>
<td>HER2-E</td>
</tr>
<tr>
<td>Basal-like</td>
<td>12 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>HER2-E</td>
<td>2 (15.38%)</td>
<td>10</td>
</tr>
<tr>
<td>Luminal A</td>
<td>0 (14.9%)</td>
<td>7</td>
</tr>
<tr>
<td>Luminal B</td>
<td>0 (13.33%)</td>
<td>4</td>
</tr>
</tbody>
</table>

*, FDR, false discovery rate.
Table 2: List of up- and down-regulated genes differentially expressed between metastatic vs. primary disease across all samples (FDR<5%).

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene Symbol</th>
<th>Score (d)</th>
<th>Fold Change</th>
<th>FDR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibroblast growth factor receptor 4</td>
<td>FGFR4</td>
<td>3.38</td>
<td>1.74</td>
<td>0</td>
</tr>
<tr>
<td>Cell division cycle 6</td>
<td>CDC6</td>
<td>2.14</td>
<td>1.29</td>
<td>1.90</td>
</tr>
<tr>
<td>Maternal embryonic leucine zipper kinase</td>
<td>MELK</td>
<td>1.99</td>
<td>1.27</td>
<td>1.90</td>
</tr>
<tr>
<td>Pituitary Tumor-Transforming</td>
<td>PTTG1</td>
<td>1.80</td>
<td>1.21</td>
<td>1.90</td>
</tr>
<tr>
<td>Cell division cycle 20</td>
<td>CDC20</td>
<td>1.79</td>
<td>1.23</td>
<td>1.90</td>
</tr>
<tr>
<td>Thymidylates Synthetase</td>
<td>TYMS</td>
<td>1.75</td>
<td>1.22</td>
<td>1.90</td>
</tr>
<tr>
<td>Centrosomal Protein</td>
<td>CEP55</td>
<td>1.73</td>
<td>1.24</td>
<td>1.90</td>
</tr>
<tr>
<td>Cyclin B1</td>
<td>CCNB1</td>
<td>1.71</td>
<td>1.19</td>
<td>1.90</td>
</tr>
<tr>
<td>Phosphoglycerate Dehydrogenase</td>
<td>PHGDH</td>
<td>1.69</td>
<td>1.23</td>
<td>1.90</td>
</tr>
<tr>
<td>Keratin 8, Type II</td>
<td>KRT8</td>
<td>1.68</td>
<td>1.17</td>
<td>1.90</td>
</tr>
<tr>
<td>Ribonucleotide Reductase M2</td>
<td>RRM2</td>
<td>1.67</td>
<td>1.22</td>
<td>1.90</td>
</tr>
<tr>
<td>Epithelial Cell Adhesion Molecule</td>
<td>EPCAM</td>
<td>1.50</td>
<td>1.15</td>
<td>3.60</td>
</tr>
<tr>
<td>BCL2-Associated Athanogene</td>
<td>BAG1</td>
<td>1.42</td>
<td>1.14</td>
<td>3.60</td>
</tr>
<tr>
<td>Ubiquitin-Conjugating Enzyme E2T</td>
<td>UBE2T</td>
<td>1.38</td>
<td>1.17</td>
<td>4.82</td>
</tr>
<tr>
<td>V-Myb Myeloblastosis Viral Oncogene Homolog-Like 2</td>
<td>MYBL2</td>
<td>1.36</td>
<td>1.21</td>
<td>4.82</td>
</tr>
<tr>
<td>Serine Peptidase Inhibitor, Kunitz Type, 2</td>
<td>SPINT2</td>
<td>1.33</td>
<td>1.13</td>
<td>4.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Symbol</th>
<th>Score (d)</th>
<th>Fold Change</th>
<th>FDR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmembrane Protein 158</td>
<td>TMEM158</td>
<td>-6.29</td>
<td>-1.34</td>
<td>0</td>
</tr>
<tr>
<td>Keratin 17, Type I</td>
<td>KRT17</td>
<td>-5.55</td>
<td>-1.22</td>
<td>0</td>
</tr>
<tr>
<td>Matrix Metallopeptidase 11</td>
<td>MMP11</td>
<td>-5.39</td>
<td>-1.32</td>
<td>0</td>
</tr>
<tr>
<td>Secreted Frizzled-Related Protein 1</td>
<td>SFRP1</td>
<td>-4.26</td>
<td>-1.39</td>
<td>0</td>
</tr>
<tr>
<td>Zinc Finger E-Box Binding Homeobox 1</td>
<td>ZEB1</td>
<td>-4.19</td>
<td>-1.57</td>
<td>0</td>
</tr>
<tr>
<td>Progesterone Receptor</td>
<td>PGR</td>
<td>-3.72</td>
<td>-1.40</td>
<td>0</td>
</tr>
<tr>
<td>Crystallin, Alpha B</td>
<td>CRYAB</td>
<td>-2.90</td>
<td>-1.59</td>
<td>0</td>
</tr>
<tr>
<td>Melanoma Inhibitor Activity</td>
<td>MIA</td>
<td>-2.82</td>
<td>-1.55</td>
<td>0</td>
</tr>
<tr>
<td>V-Myc Myelocytomatosis Viral</td>
<td>MYC</td>
<td>-2.69</td>
<td>-1.67</td>
<td>0</td>
</tr>
<tr>
<td>Oncogene Homolog</td>
<td>Gene</td>
<td>Ratio</td>
<td>Log2 Fold Change</td>
<td>p.Value</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>------</td>
<td>-------</td>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>N-Acetyltransferase 1</td>
<td>NAT1</td>
<td>-2.31</td>
<td>-1.62</td>
<td>0</td>
</tr>
<tr>
<td>Microtubule-Associated Protein Tau</td>
<td>MAPT</td>
<td>-2.25</td>
<td>-1.64</td>
<td>0</td>
</tr>
<tr>
<td>Estrogen Receptor</td>
<td>ESR1</td>
<td>-2.07</td>
<td>-1.65</td>
<td>0</td>
</tr>
<tr>
<td>Retinoid-Inducible Nuclear Factor</td>
<td>CXXC5</td>
<td>-2.07</td>
<td>-1.80</td>
<td>0</td>
</tr>
<tr>
<td>Fibrin 1</td>
<td>FBN1</td>
<td>-1.99</td>
<td>-1.72</td>
<td>0</td>
</tr>
<tr>
<td>AXL Receptor Tyrosine Kinase</td>
<td>AXL</td>
<td>-1.88</td>
<td>-1.78</td>
<td>0</td>
</tr>
<tr>
<td>Forkhead Box A1</td>
<td>FOXA1</td>
<td>-1.87</td>
<td>-1.72</td>
<td>0</td>
</tr>
<tr>
<td>Solute Carrier Family 16</td>
<td>SLC16A3</td>
<td>-1.82</td>
<td>-1.78</td>
<td>0</td>
</tr>
<tr>
<td>Lipoma HMGIC Fusion Partner</td>
<td>LHFP</td>
<td>-1.74</td>
<td>-1.78</td>
<td>0</td>
</tr>
<tr>
<td>Solute Carrier Family 39 Member 6</td>
<td>SLC39A6</td>
<td>-1.73</td>
<td>-1.75</td>
<td>0</td>
</tr>
<tr>
<td>Caveolin 1</td>
<td>CAV1</td>
<td>-1.59</td>
<td>-1.78</td>
<td>0</td>
</tr>
<tr>
<td>GATA Binding Protein 3</td>
<td>GATA3</td>
<td>-1.54</td>
<td>-1.79</td>
<td>0</td>
</tr>
<tr>
<td>Erb-B2 Receptor Tyrosine Kinase 2</td>
<td>ERBB2</td>
<td>-1.52</td>
<td>-1.82</td>
<td>0</td>
</tr>
<tr>
<td>B-Cell CLL/ Lymphoma 2</td>
<td>BCL2</td>
<td>-1.41</td>
<td>-1.80</td>
<td>1.35</td>
</tr>
<tr>
<td>DNA Damage Inducible Transcript 4</td>
<td>DDIT4</td>
<td>-1.17</td>
<td>-1.85</td>
<td>3.60</td>
</tr>
<tr>
<td>Claudin 7</td>
<td>CLDN7</td>
<td>-1.10</td>
<td>-1.87</td>
<td>3.60</td>
</tr>
<tr>
<td>Cadherin 3</td>
<td>CDH3</td>
<td>-1.10</td>
<td>-1.82</td>
<td>3.60</td>
</tr>
<tr>
<td>Adrenomedullin</td>
<td>ADM</td>
<td>-1.07</td>
<td>-1.83</td>
<td>4.82</td>
</tr>
<tr>
<td>Phosphatase and Tensin Homolog</td>
<td>PTEN</td>
<td>-1.06</td>
<td>-1.91</td>
<td>4.82</td>
</tr>
<tr>
<td>Forkhead Box C1</td>
<td>FOXC1</td>
<td>-1.03</td>
<td>-1.82</td>
<td>4.82</td>
</tr>
<tr>
<td>MET Proto-Oncogene, Receptor Tyrosine Kinase</td>
<td>MET</td>
<td>-1.02</td>
<td>-1.83</td>
<td>4.82</td>
</tr>
<tr>
<td>Vimentin</td>
<td>VIM</td>
<td>-0.99</td>
<td>-1.88</td>
<td>4.82</td>
</tr>
</tbody>
</table>
Figure 1
Distribution of intrinsic subtype in primary vs. metastatic disease

PAM50 subtypes primary

- LumA: 39%
- LumB: 13.8%
- HER2-E: 26%
- Basal-like: 9.8%
- Normal: 11.4%

PAM50 subtypes metastasis

- LumA: 4%
- LumB: 35.8%
- HER2-E: 22%
- Basal-like: 12.2%
- Normal: 26%
Figure 2
Gene and signature expression changes between primary and metastasis. P-value has been obtained after performing a paired t-test.

LumA PAM50 Signature
P-value = 0.0001446

Her2 PAM50 Signature
P-value = 1.095e-05

Proliferation score PAM50 Signature
P-value = 1.261e-06

ESR1 Gene
P-value = 0.1753

ERBB2 Gene
P-value = 0.3081

PGR Gene
P-value = 0.001071
Figure 3
Association of 10 signatures with OSmet when evaluated in primary (A) and metastatic (B) disease. Each signature has been evaluated as a continuous variable and standardised to have a mean of 0 and a standard deviation of 1. The size of the square is inversely proportional to the standard error; horizontal bars represent the 95% CIs of hazard ratios. Statistically significant variables are shown in blue. Each gene signature has been evaluated in a univariate analysis.
Figure 4
Venn diagram of genes that predict overall survival from the data of recurrence when analyzed in primary vs. metastatic disease. Green: genes associated with good prognosis. Red: genes associated with poor prognosis.