Abstract

Breast cancer during pregnancy is rare and is associated with relatively poor prognosis. No information is available on its biological features at the genomic level. Using a dataset of 54 pregnant and 113 non-pregnant breast cancer patients, we evaluated the pattern of hot spot somatic mutations and did transcriptomic profiling using Sequenom and Affymetrix respectively. We performed gene set enrichment analysis to evaluate the pathways associated with diagnosis during pregnancy. We also evaluated the expression of selected cancer-related genes in pregnant and non-pregnant patients and correlated the results with changes occurring in the normal breast using a pregnant murine model. We finally investigated aberrations associated with disease-free survival (DFS). No significant differences in mutations were observed. Of the total number of patients, 18.6% of pregnant and 23% of non-pregnant patients had a PIK3CA mutation. Around 30% of tumors were basal, with no differences in the distribution of breast cancer molecular subtypes between pregnant and non-pregnant patients. Two pathways were enriched in tumors diagnosed during pregnancy: the G protein-coupled receptor pathway and the serotonin receptor pathway (FDR <0.0001). Tumors diagnosed during pregnancy had higher expression of PD1 (PDCD1; P=0.015), PDL1 (CD274; P=0.014), and gene sets related to SRC (P=0.004), IGF1 (P=0.032), and β-catenin (P=0.019). Their expression increased almost linearly throughout gestation when evaluated on the normal breast using a pregnant mouse model underscoring the potential effect of the
Introduction

Breast cancer arising in young women exhibits distinct biological features and has been shown to pursue an aggressive clinical behavior (Cancello et al. 2010, Azim et al. 2012a, Collins et al. 2012, Narod 2012). Among the unique situations that young women with breast cancer may face is pregnancy. The fact that breast cancer is a hormonally driven tumor, and given the considerable structural changes that occur in the breast during pregnancy, raises concerns on the potential impact of pregnancy on breast cancer biology.

It is estimated that one out of 3000 pregnancies is complicated with breast cancer, and this incidence appears to be on the rise (Peccatori et al. 2013). While in recent years, there have been considerable efforts in defining safe and effective treatment strategies in managing breast cancer arising during pregnancy (Cardonick et al. 2010, Amant et al. 2012, Loibl et al. 2012), very little progress have been made regarding our understanding of the biology of these tumors.

Recently, several groups including ours have published on the prognosis of patients diagnosed with breast cancer during pregnancy (Azim et al. 2012b, Amant et al. 2013, Litton et al. 2013). Results of these studies have witnessed major controversy with studies showing worse outcomes, while others showing no difference in prognosis or even improved survival. These results further call for the importance of elucidating the effects of diagnosis during pregnancy on breast cancer biology.

In the current study, we investigated for the first time the effect of diagnosis during pregnancy on the biology of breast cancer arising in young women at the genomic level; particularly, the pattern of hot spot somatic mutations and gene expression profiles.

Materials and methods

We identified 65 patients who were diagnosed with breast cancer during pregnancy in the period from 1996 to 2010 at the European Institute of Oncology (IEO) in Milan, Italy. For each case, we identified two breast cancer patients who were not matched according to age (±2 years), tumor size, nodal status, year of surgery (±2 years), and whether neoadjuvant chemotherapy was administered. Details regarding patient characteristics, pathological features, and treatments were published previously (Azim et al. 2012b). In brief, we did not find any obvious differences in the pathological features or the expression of estrogen receptor (ER), progesterone receptor (PgR), HER2, and Ki-67. However, at a median follow-up of 4 years, pregnant patients had significantly worse disease-free survival (DFS; Azim et al. 2012b).

All patients had available formalin-fixed paraffin-embedded (FFPE) tissue from the primary breast surgery. For the sake of this study, all samples were pathologically reviewed at IEO and three 1 mm tumor cores (>90% tumor cellularity) were sent to the Breast Cancer Translational Research Laboratory at Institut Jules Bordet for mutational and transcriptomic profiling. DNA and RNA extraction were performed using the ZYMO Research Pinpoint Slide DNA Isolation System and stored at −80°C until use. Quality and quantity were evaluated using the NanoDrop.

All patients provided consent to donate biological samples for research purposes per IEO institutional policies. This study was approved by the Ethics Committee of Institut Jules Bordet in 2010 (number: 1782).

Mutational profiling

The samples were genotyped using the Sequenom MassARRAY Assay Design 3.1 with the default parameters. Multiplex PCR was done in 5 μl volume containing 5–10 ng DNA. Samples were considered to be of sufficient quality when genotyping could be performed for more than 75% of the mutations.

We queried the COSMIC database to identify a broad range of genes for hotspot somatic mutation profiling (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/). In total, we examined 84 hotspot somatic mutations on 19 cancer-related genes. Details regarding the genes and mutations that were evaluated are provided in
Supplementary Table 1, see section on supplementary data given at the end of this article. We ultimately covered 94% (n=29) of the reported PIK3CA mutations.

Transcriptomic profiling

All samples were hybridized using Affymetrix Human Genome U219 array plates following standard Affymetrix protocols. This platform was used based on pilot studies performed in-house showing high reproducibility between transcriptomic information obtained from paired frozen and FFPE tumor samples. Detailed information on the used protocol is provided in Supplementary Materials and Methods, see section on supplementary data given at the end of this article.

We used two methods to determine breast cancer molecular subtypes: PAM50 (Parker et al. 2009) and 3-Gene (Haibe-Kains et al. 2012), as defined previously (Supplementary Materials and Methods). To compare the transcriptomic profiles of patients diagnosed with breast cancer during pregnancy and of those not diagnosed during pregnancy, we performed gene set enrichment analysis (GSEA) using the Gene Ontology classes as a reference gene set using the Bioconductor Package Piano (Varemo et al. 2013). The genes were sorted based on their differential expression between pregnant and non-pregnant patients using a conditional logistic regression model (http://hosho.ees.hokudai.ac.jp/~kubo/Rdoc/library/Epi/html/clogistic.html). Results were cross-checked with those obtained using the hypergeometric probabilities delivered by the DAVID web tool (Huang da et al. 2009).

We then evaluated the expression of selected cancer-related genes such as BRCA1, BRCA2, PD1 (PDCD1), and PDL1 (CD274) in addition to previously published gene sets representing several important oncological biological processes in pregnant and non-pregnant patients. These gene sets included the following: RAS, SRC, MYC, Wnt/β-catenin (Bild et al. 2006), PI3K (Loi et al. 2010), insulin-like growth factor 1 (IGF1; Creighton et al. 2008), AKTmTOR (Majumder et al. 2004), MAPK (Creighton et al. 2006), PTEN (Saal et al. 2008); proliferation gene sets; GGI (Sotiriou et al. 2006) and GENE70 (van ’t Veer et al. 2002), mammary stem cell and luminal progenitor gene sets (Lim et al. 2009); stroma gene sets; DCN (Farmer et al. 2009) and PLAU (Desmedt et al. 2008), and immune gene sets; and STAT1 (Desmedt et al. 2008) and IRM (Teschendorff et al. 2007).

Gene sets (or signature score) were computed by taking the sum of the products of the gene coefficient in the module (si) and the corresponding gene expression value (ei) according to the following formula:

$$\text{Sigscore} = \sum_{i=1}^{n} e_i \times s_i.$$ 

Supplementary Table 2, see section on supplementary data given at the end of this article summarizes the size of the different gene sets and the number of genes that were available in the Affymetrix platform that we used in the current analysis. The raw gene expression data, together with patients’ characteristics, are publically available on GEO http://www.ncbi.nlm.nih.gov/geo/, with accession number GSE53031.

In-silico analysis of publically available gene expression data derived from the normal breast using a pregnant mouse model

In order to evaluate the potential role of pregnancy in the normal breast tissue, we investigated the expression of genes or gene sets that were found to be differentially expressed between the pregnant and non-pregnant groups on a publically available murine data set derived from the normal breast of pregnant mice (GEO ID: GSE8191; Anderson et al. 2007). In the murine model, the breast tissue samples were obtained on days 1, 3, 7, 12, and 17 of pregnancy and also on day 19 (post-partum). Four replicates were available for each time point. Similar to the procedure that was applied in the human data set, the data were re-normalized using the robust multiarray analysis normalization, and the expression values of each set of probes corresponding to one Ensembl identifier were averaged (Bolstad et al. 2003).

Survival analysis

We performed updated DFS comparison between pregnant and non-pregnant breast cancer patients using a multivariate Cox regression model adjusted for age, tumor size, nodal status, histological grade, ER, HER2, and type of systemic therapy.

We performed an exploratory analysis to evaluate genes/gene sets or somatic mutations that were associated with DFS. We used the median expression value in the whole population (i.e. pregnant+non-pregnant) to define the cutoff of high and low gene or gene set expression. Analyses were performed using the SAS Software, version 9.2 (SAS Institute, Cary, NC, USA) and R Software, version 2.12.2 (available at www.r-project.org). All tests were two sided.
Results

Study population

The original data set included 195 breast cancer patients: 65 pregnant and 130 non-pregnant patients. For the current genomic analysis, we opted to exclude patients who received neoadjuvant therapy to avoid potential impact on the obtained results (n=21). Seven more patients were excluded due to poor RNA quality resulting in a total of 167 patients (85% of original data set): 54 pregnant and 113 non-pregnant patients who were included in the current analysis (Supplementary Fig. 1, see section on supplementary data given at the end of this article). Table 1 summarizes patient characteristics in comparison to the original data set. No major differences were observed.

Mutational analysis

No significant differences in the pattern of hot spot somatic mutations evaluated in this study were observed between pregnant and non-pregnant patients (Table 2). Considering all eligible patients, 36 (21.6%) and 9 (5.4%) had a mutation in PIK3CA and TP53 respectively.

Molecular subtyping

We found a high correlation between PAM50 and the 3-Gene classifiers in defining breast cancer molecular subtypes (r=0.857, P<0.0001). There was also a high correlation with immunohistochemistry (IHC)-defined subtypes based on the expression of ER, HER2, and a Ki-67 cutoff of 14% to distinguish luminal A and luminal B (both r>0.7, P<0.0001; Supplementary Table 3, see section on supplementary data given at the end of this article). We did not find any significant differences in breast cancer subtypes defined by either of the above classifiers between pregnant and non-pregnant breast cancer patients (Fig. 1).

Gene set enrichment analysis

After adjusting for multiple testing, two relevant pathways were significantly enriched in tumors diagnosed in pregnant vs non-pregnant breast cancer patients; the G protein-coupled receptor (GPCR) signaling pathway and the serotonin receptor signaling pathway (both P<0.0001; Fig. 2). The results obtained by the GSEA approach were also confirmed using the hypergeometric probabilities delivered by the DAVID web tool on the non-sorted list of significantly differentially expressed genes.

Table 1 Differences in characteristics between the whole study population and the subpopulation (~85%) included in the genomic analysis

<table>
<thead>
<tr>
<th></th>
<th>Breast cancer patients diagnosed during pregnancy, n (%)</th>
<th>Breast cancer patients not diagnosed during pregnancy, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Whole study</td>
<td>Genomic analysis</td>
</tr>
<tr>
<td>Number</td>
<td>65</td>
<td>54</td>
</tr>
<tr>
<td>Tumor size ≥2 cm</td>
<td>37 (56.9%)</td>
<td>28 (51.9%)</td>
</tr>
<tr>
<td>Node positive</td>
<td>35 (53.8%)</td>
<td>28 (51.9%)</td>
</tr>
<tr>
<td>Histological grade III</td>
<td>36 (55.3%)</td>
<td>32 (59.2%)</td>
</tr>
<tr>
<td>ER-positive</td>
<td>43 (66.1%)</td>
<td>36 (66.7%)</td>
</tr>
<tr>
<td>PgR-positive</td>
<td>42 (64.6%)</td>
<td>35 (64.8%)</td>
</tr>
<tr>
<td>HER2-positive</td>
<td>11 (16.9%)</td>
<td>9 (16.7%)</td>
</tr>
<tr>
<td>Ki-67 ≥30%</td>
<td>35 (53.9%)</td>
<td>31 (57.4%)</td>
</tr>
<tr>
<td>No. of DFS events</td>
<td>27 (41.5%)</td>
<td>19 (35.2%)</td>
</tr>
<tr>
<td>No. of OS events</td>
<td>13 (20%)</td>
<td>10 (18.5%)</td>
</tr>
</tbody>
</table>

ER, estrogen receptor; PgR, progesterone receptor; DFS, disease-free survival; OS, overall survival.

Table 2 Differences in the prevalence of hot spot somatic mutations in tumors diagnosed in pregnant and non-pregnant breast cancer patients

<table>
<thead>
<tr>
<th></th>
<th>Patients diagnosed during pregnancy, n (%)</th>
<th>Patients not diagnosed during pregnancy, n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number PIK3CA</td>
<td>54 (100%)</td>
<td>113 (100%)</td>
<td>0.51</td>
</tr>
<tr>
<td>Exon 9 mutations</td>
<td>10 (18.6%)</td>
<td>26 (23%)</td>
<td></td>
</tr>
<tr>
<td>Exon 20 mutations</td>
<td>3 (5.6%)</td>
<td>9 (8%)</td>
<td></td>
</tr>
<tr>
<td>Other PIK3CA mutations</td>
<td>4 (7.4%)</td>
<td>12 (10.6%)</td>
<td></td>
</tr>
<tr>
<td>p53</td>
<td>2 (3.7%)</td>
<td>7 (6.2%)</td>
<td>0.51</td>
</tr>
<tr>
<td>MAP3K1</td>
<td>1 (1.9%)</td>
<td>2 (1.8%)</td>
<td>0.97</td>
</tr>
<tr>
<td>STK1</td>
<td>1 (1.9%)</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>RET</td>
<td>0</td>
<td>1 (0.9%)</td>
<td>0.49</td>
</tr>
</tbody>
</table>

DOI: 10.1530/ERC-14-0111
Differential expression of selected cancer-related genes and published gene sets on tumors diagnosed in pregnant vs non-pregnant breast cancer patients

In pregnant breast cancer patients, we found higher expression of PD1 (P = 0.015), PDL1 (P = 0.014), BRCA1 (P = 0.018), and gene sets related to SRC (P = 0.004), IGF1 (P = 0.032), and β-catenin (P = 0.019) (Fig. 3). Most of them showed a trend of higher expression in patients diagnosed later in the course of pregnancy, although this was not statistically significant. On the contrary, we found lower expression of AKTmTOR gene set (P = 0.05) in pregnant breast cancer patients, with patients diagnosed during the third trimester of pregnancy having the lowest expression (Supplementary Fig. 3, see section on supplementary data given at the end of this article).

No significant difference in expression of the other genes and gene sets was observed between patients diagnosed with breast cancer during pregnancy and non-pregnant patients considered in this study (Supplementary Fig. 4, see section on supplementary data given at the end of this article).

The expression of the differentially expressed genes and gene sets on the normal breast during the course of gestation in a pregnant mouse model

While evaluating the expression of the genes and gene sets that were highly expressed in tumors diagnosed during pregnancy in the normal breast using a pregnant mouse model, we found almost a linear increase in expression from D1 to D17 of pregnancy in all the genes and gene sets evaluated with the exception of BRCA1 (Fig. 4). The opposite trend was observed with AKTmTOR module where the expression was lower in patients diagnosed during pregnancy (P = 0.05) with a linear decrease in expression in the normal breast from D1 to D17 in the pregnant mouse model (Supplementary Fig. 3).

Survival analysis

At a median follow-up of 65 months, patients diagnosed during pregnancy continued to have a poorer DFS compared with the non-pregnant group (HR 2.5 (95% CI, 1.3–4.8), P = 0.005), even after adjusting for classical prognostic factors (Supplementary Fig. 5, see section on supplementary data given at the end of this article).
No genes were found to be associated with DFS in the univariate model using false discovery rate (FDR) <0.05. Among the genes and gene sets that were highly expressed in the pregnant group, only IGF1 was significantly associated with DFS, both in the pregnant (P=0.01) and the non-pregnant groups (P=0.04) (Supplementary Fig. 6, see section on supplementary data given at the end of this article). However, this was not significant in a multivariate model adjusted for classical prognostic parameters (HR=1.0, P=0.4). No association was observed between PIK3CA mutation and DFS as well (Supplementary Fig. 7).

Discussion

In this study, we evaluated the impact of diagnosis during pregnancy on breast cancer biology. We found no obvious differences in classical pathological features, pattern of hot spot somatic mutations, or breast cancer molecular subtypes. However, tumors diagnosed during pregnancy were associated with different gene expression patterns and activated signaling pathways, namely the GPCR pathway and the serotonin receptor pathway.

GPCRs represent the largest family of cell surface molecules involved in signal transmission, which include ~2–4% of genes in the genome. They regulate many cell functions including proliferation, angiogenesis, and cell survival (Dorsam & Gutkind 2007). CXCR4 and GPR30 are GPCRs that were shown to be of higher relevance in breast cancer; the former was found in preclinical models to increase the metastatic potential of breast cancer cells (Muller et al. 2001), while the latter was linked to the process of endocrine resistance (Thomas et al. 2005). The serotonin receptor pathway belongs to the GPCR family (Kvachnina et al. 2005). Serotonin is known to play a role in mammary gland development (Matsuda et al. 2004). It is synthesized within the breast epithelial cells during lactation and acts as a physiological regulator of involution following pregnancy, in part by favoring growth arrest and cell death (Marshall et al. 2010). The application of serotonin to primary human epithelial cells was shown to suppress cell growth (Pai et al. 2009). Conversely, serotonin was found to stimulate tumor growth when applied to invasive breast cancer cell lines, by promoting proliferation and epithelial–mesenchymal transition (Pai et al. 2009). Moreover, serotonin was found to induce the transcription of parathyroid hormone-related protein and the metastasis-associated transcription factor Runx2/Cbfal via the serotonin receptor, 5-HT2 (Hernandez et al. 2012). Hence, it is plausible that the activation of the serotonin receptor pathway in tumors arising during pregnancy may play a role in controlling the aggressive clinical behavior of these tumors.

We found that tumors diagnosed during pregnancy have higher IGF1 gene set expression. IGF1 is known to increase breast cancer development and breast cancer recurrence (Kim et al. 2007, Smith et al. 2011). In the normal breast, the IGF1 pathway is known to be critical for mammary gland development and progesterone...
regulation and is required for estrogen to cause ductal development (Macias & Hinck 2012). As the breast undergoes part of its development during pregnancy, we hypothesize that the increase in IGF1 expression that we observed in the normal breast over the course of gestation could be to support mammary gland development. However, a previous study has demonstrated that ‘previous’ pregnancy down-regulates IGF1 expression in the normal breast, suggesting that this could explain the long-term protective effect of pregnancy on breast cancer (D’Cruz et al. 2002). This is not in direct conflict with our results as pregnancy is known to exert a bi-directional effect on breast cancer development, promoting aggressive tumor growth on the short term with a long-term protective effect (Albrektsen et al. 2005). We speculate that pregnancy modifies the breast microenvironment, which consequentially impacts breast cancer risk and biology with IGF1 emerging as one of the main pathways that appear to be implicated in the crosstalk between pregnancy and breast cancer.

The high expression of PD1 is also intriguing. PD1 is a key immune-checkpoint receptor expressed by activated T-cells and mediates immunosuppression by binding to its ligand, PDL1. During pregnancy, PD1–PDL1 interaction is activated to allow feto-maternal tolerance (D’Addio et al. 2011). On the other hand, in cancer, inhibition of the interaction between PD1 and its ligand PDL1 enhances T-cell responses in vitro and mediates antitumor activity (Dong et al. 2002, Iwai et al. 2002). In a recent study, an anti-PD1 was associated with an impressive clinical activity in patients with refractory solid tumors, mainly those with positive expression of PDL1 (Topalian et al. 2012). Our findings point out that it is likely that patients diagnosed during pregnancy express PDL1, which appears to be mediated by the normally higher PDL1 expression during pregnancy. It is plausible that the high expression of PD1/PDL1 on the tumor mediates immune suppression and hence results in aggressive tumor growth. However, this remains speculative requiring further confirmation.

In line with our findings, it was previously shown that BRCA1 is highly expressed in the breast during the first weeks of pregnancy possibly to counteract the state of proliferation and differentiation that occur in the breast to form mature alveoli (Marquis et al. 1995). It was later

---

**Figure 3**

High expression of different genes and gene sets in pregnant compared with non-pregnant breast cancer patients. Differential expression according to period of gestation at breast cancer diagnosis is shown (i.e. first, second, and third trimesters). One patient had missing information on period of gestation at breast cancer diagnosis. The Y-axis represents the relevant expression value where each gene or gene set was treated as a continuous variable. (A) SRC gene set (pregnant vs non-pregnant, \( P = 0.004 \); first trimester vs third trimester, \( P = 0.28 \)); (B) IGF1 gene set (pregnant vs non-pregnant, \( P = 0.032 \); first trimester vs third trimester, \( P = 0.61 \)); (C) \( \beta \)-catenin gene set (pregnant vs non-pregnant, \( P = 0.019 \); first trimester vs third trimester, \( P = 0.45 \)); (D) PD1 (pregnant vs non-pregnant, \( P = 0.015 \); first trimester vs third trimester, \( P = 0.1 \)); (E) PDL1 (pregnant vs non-pregnant, \( P = 0.014 \); first trimester vs third trimester, \( P = 0.05 \)); and (F) BRCA1 (pregnant vs non-pregnant, \( P = 0.018 \); first trimester vs third trimester, \( P = 0.74 \)).
In conclusion, this is a unique study interrogating the effect of pregnancy on breast cancer biology. Given the rarity of this disease, we believe that it could serve as a valuable resource for future research in this field. While we did not observe differences in selected somatic mutations, the altered breast microenvironment during pregnancy appeared to influence the tumor phenotype, as shown in our transcriptomic analysis. This could potentially impact the behavior of these tumors.

**Figure 4**
Differential expression of genes and gene sets on the normal breast in a pregnant mouse model throughout the period of gestation. D1 represents conception, D17 end of pregnancy, and D19 post-partum. The Y-axis represents the relevant expression value where each gene or gene set was treated as a continuous variable. (A) SRC gene set (days 1 vs 17, \( P=0.035 \)); (B) IGF1 gene set (days 1 vs 17, \( P=0.004 \)); (C) BRCA1 (days 1 vs 17, \( P=0.11 \)); (D) \( \beta \)-catenin gene set (days 1 vs 17, \( P=0.008 \)); and (E) PD1 (days 1 vs 17, \( P=0.001 \)).
Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/ERC-14-0111.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
H A Azim Jr was supported by an ESMO translational research grant to conduct this Study. This research was supported by the fund from Les Amis De L’Institut Bordet and the ‘Fond de la Recherche Scientifique (FNRS)’ (ref: FC 55854).

Author contribution statement
H A Azim Jr, F A Peccatori, N Rotmensz, M Piccart, and C Sotiriou were responsible for study concept and design. P Dell’Orto and G Pruneri, and G Viale were responsible for data collection. N Rotmensz, and P Dell’Orto were responsible for data analysis. H A Azim Jr, S Brohé, A Azim Jr, S Brohé, F A Peccatori, and C Sotiriou were responsible for gene expression profiling analysis. D Lambrechts was responsible for mutational profiling. S Brohé and N Rotmensz were responsible for the statistical analysis. All authors have contributed to data interpretation. H A Azim Jr, S Brohé, F A Peccatori, and C Sotiriou were responsible for manuscript writing. All authors approved the final version of the manuscript.

Acknowledgements
This work was presented as an oral presentation in the 9th European Breast Cancer Conference (EBCC), March 2014, Glasgow, UK.

References


Received in final form 30 April 2014
Accepted 6 May 2014
Made available online as an Accepted Preprint 13 May 2014

Research

H A Azim Jr et al.

Biology of breast cancer during pregnancy

21:4 554

Published by Bioscientifica Ltd.