Subtype-dependent prognostic relevance of an interferon-induced pathway metagene in node-negative breast cancer

Maurizio Callari, Valeria Musella, Eleonora Di Buduo, Marialuisa Sensi, Patrizia Miodini, Matteo Dugo, Rosaria Orlandi, Roberto Agresti, Biagio Paolini, Maria Luisa Carcangi, Vera Cappelletti, Maria Grazia Daidone

Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Amadeo, 42, 20133 Milan, Italy
Department of Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian, 1, 20133 Milan, Italy
Department of Pathology, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian, 1, 20133 Milan, Italy

ABSTRACT
The majority of gene expression signatures developed to predict the likelihood to relapse in breast cancer (BC) patients assigns a high risk score to patients with Estrogen Receptor (ER) negative or highly proliferating tumors. We aimed to identify a signature of differentially expressed (DE) metagenes, rather than single DE genes, associated with distant metastases beyond classical risk factors.

We used 105 gene expression profiles from consecutive BCs to identify metagenes whose prognostic role was defined on an independent series of 92 $ESR1^+$/ERBB2/$^-$/node-negative BCs (42 cases developing metastases within 5 years from diagnosis and 50 cases metastasis-free for more than 5 years, comparable for age, tumor size, ER status and surgery). Findings were validated on publicly available datasets of 684 node-negative BCs including all the subtypes.

Only a metagene containing interferon-induced genes (IFN metagene) proved to be predictive of distant metastasis in our series of patients with $ESR1^+$/ERBB2/$^-$/ tumors ($P = 0.029$), and such a finding was validated on 457 $ESR1^+$/ERBB2/$^-$/ BCs from public datasets ($P = 0.0424$). Conversely, the IFN metagene was associated with a low risk of metastasis in 104 ERBB2/$^+$ tumors ($P = 0.0099$) whereas it did not prove to significantly affect prognosis in 123 $ESR1^-$/ERBB2/$^-$/ tumors ($P = 0.2235$). A complex prognostic interaction was revealed in $ESR1^+$/ERBB2/$^-$/ and ERBB2/$^+$ tumors when the association between the IFN metagene and a T-cell metagene was considered.

The study confirms the importance of analyzing prognostic variables separately within BC subtypes, highlights the advantages of using metagenes rather than genes, and finally identifies in node-negative $ESR1^+$/ERBB2/$^-$/ BCs, the unfavorable role of high IFN metagene expression.

© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
Research on the potential for invasion and metastasis has accelerated dramatically over the last decade as new research tools and pre-clinical models have become available, but we are still faced with many unanswered questions both at biological and clinical level.

Studies on metastasis biology have pinpointed the central role of epithelial mesenchymal transition and of a set of transcriptional factors (Snail, Slug, Twist, Zeb1/2) orchestrating the process and the related migratory ability of cells. In parallel, the role of crosstalk between cancer cells and cells of the microenvironment has been acknowledged as a major mechanism leading to acquisition of invasive growth and metastasis as well as plasticity (Egeblad et al., 2010).

From a clinical point of view, as metastatic disease offers still unmet challenges for a curative treatment of breast cancer (BC), it is important to early identify patients at risk of relapse especially within the subset of node-negative women defined with a relatively indolent disease based on pathologic and molecular features. In fact, over-treatment is still an unresolved problem in clinical management as up to 75% of women receive unnecessary adjuvant chemotherapy (Reis-Filho and Pusztai, 2011; van’t Veer and Bernards, 2008).

Gene expression profiling (GEP) was initially acclaimed as the magic bullet for a proper risk prediction. Most of the studies were based on the concept that, even years before eliciting a clinically evident metastasis, primary tumors are able to send specific signals to distant sites in an attempt to prepare a sort of landing dock for circulating tumor cells which will settle and develop a secondary lesion (Wels et al., 2008). Although profiling of the primary tumor may give hints about its future metastatic potential, mistakes in designing the studies, mainly due to analyses carried out without taking into consideration BC subtypes, led to the development of a plethora of signatures still attributing the risk to lack of estrogen receptor (ER) or the presence of HER2 alterations (Reis-Filho et al., 2010; Reis-Filho and Pusztai, 2011). An exception to this was the so-called 76-gene Rotterdam signature (Wang et al., 2005) which predicts risk of distant metastases in node-negative patients independently of age, tumor size, grade and ER status, partially taking into account the BC subtype, as 60 and 16 genes of the signature were specifically derived from women with ER+ or ER− tumors respectively.

Overall, such analyses minimally improved the predictive value over classical predictors, whereas the clinical benefit associated with treatment decision based on gene expression profiling is still under clinical validation (Cardoso et al., 2007; Sparano and Paik, 2008). In addition, the information gained from functional annotation of genes derived from prognostic signatures so far has shown that most of them are proliferation related, associated with ER status or with immune response (Iwamoto and Pusztai, 2010). Thus, globally, gene expression in the context of clinical studies is of limited help in giving hints concerning possible new targets or pathways associated with the metastatic process whereas, in the context of preclinical studies, it suffers from the intrinsic limited availability of models representing the molecular heterogeneity of clinical tumors (Vargo-Gogola and Rosen, 2007).

With the aim to identify additional features associated with the likelihood of metastatic spread beyond the classical risk factors, we performed gene expression analysis on a homogeneous clinical setting of primary node-negative BC with a study design: i) accounting for a comparable pattern of classical risk factors (age, tumor size, ER) between women who early developed metastasis and those free of disease for more than 5 years to avoid biases linked to correlations between driver molecular signatures predictive of metastasis and routine clinico-pathological features; ii) investigating the expression of metagenes instead of single genes; iii) initially focused on ESR1+/ERBB2− tumors, since it is becoming increasingly clear that BC represents an heterogeneous group of diseases and that prognosis is associated with distinct pathways according to molecular subtypes (Iwamoto et al., 2011) and iv) successively extended to public independent datasets also including ESR1−/ERBB2− and ERBB2+ subtypes.

Based on these premises, we identified and validated the pivotal role for a metagene containing IFN-induced genes associated with a different metastatic risk as a function of the specific BC subtype and of its interaction with a T-cell metagene.

2. Materials and methods

2.1. Cases series

2.1.1. Discovery datasets

Our case series consisted of a total of 228 GEPs from primary tumors, carefully sampled by experienced Pathologists, immediately snap-frozen in liquid nitrogen and stored at −80 °C in the Institutional Tissue Bank, and included:

a) one metagene identification dataset, consisting of tissues collected at the time of diagnosis from 97 consecutive previously untreated patients with operable BC undergoing radical or conservative surgery at the Fondazione IRCCS Istituto Nazionale Tumori di Milano (INTM) during the year 2007 and profiled for an independent study (manuscript in preparation) including 8 technical replicates (105 total profiles);

b) one independent metagene test dataset consisting of 123 patients with operable BCs previously untreated and undergoing surgery at INTM between 1990 and 1998. All these patients underwent radical or conservative surgery plus radiotherapy and axillary lymph node dissection, were defined as node-negative by pathologic analysis of axillary lymph nodes, and were not submitted to any type of adjuvant systemic treatment until relapse. This case series included 59 patients who developed distant metastases within 5 years of the surgical treatment and 64 patients free of any distant metastasis for more than 5 years and selected to have similar age, tumor size and ER status. A written informed consent signed by the patients...
authorized the use of material leftover from the diagnosis for research purposes. The study was approved by the Independent Ethics Committee and INTM Review Board.

The cellularity of frozen specimens was evaluated by Pathologists, and only tissues with at least 60% of tumor cells were submitted to molecular analyses (about 12% of the initially collected samples did not meet such criteria). Clinico-pathological information for the discovery datasets is summarized in Table S1.

2.1.2. Validation dataset
We combined three publicly available datasets (GSE2034, GSE7390, GSE11121) downloaded from the Gene Expression Omnibus website (http://www.ncbi.nlm.nih.gov/geo/) yielding a total of 684 gene expression profiles of primary node-negative BC from patients receiving no adjuvant or neo-adjuvant systemic treatments. All data were obtained using the Affymetrix U133A chips. Mas5-processed data were used and an empirical Bayes method (Johnson et al., 2007) to reduce inter-dataset variation was applied (Supplementary Figure 1). Probesets with IQR < 0.6 were filtered out.

A flow-chart describing both the discovery and validation datasets and the type of analytical steps applied to each dataset is given in Figure 1.

2.2. RNA isolation and microarray hybridization
Tissue was pulverized using a Mikrodismembrator (Braun Biotech International, Germany). Total RNA was extracted with the Trizol reagent (Invitrogen, Carlsbad, CA) according to manufacturer instructions and an additional DNase digestion was performed using the RNeasy kit (Qiagen, Valencia, CA). After each extraction, a small fraction of RNA was used for quality and yield assessment. RNA total concentration and purity were determined by UV spectrophotometry. Total RNA electrophoretic profile was analyzed by the Agilent RNA 6000 NanoLabChip kit on the Agilent 2100 Bioanalyzer (Agilent Technologies, Falo Alto, CA) using the software provided by the manufacturer for determination of RNA integrity number.

In the metagene test dataset 8/137 (6%) of samples, and 6/137 (4%) were rated as degraded and yielding insufficient RNA, respectively.

RNA samples were processed for microarray hybridization by INTM Functional Genomics Core Facility. Briefly, 800 ng of total RNA was reverse transcribed, labeled with biotin and amplified overnight using the Illumina RNA TotalPrep Amplification kit (Ambion, Austin, TX) according to the manufacturer’s protocol. One μg of the biotinylated cRNA sample was mixed with the Hyb E1 hybridization buffer containing 37.5% (w/w) formamide and then hybridized at 58 °C overnight on the HumanRef-6_v3 (Illumina, Inc, San Diego, CA) in the case of the metagene test dataset and on the Human-HT 12 v3 (Illumina, Inc, San Diego, CA) in the case of the metagene identification dataset. Array chips were washed with the manufacturer’s E1BC solution, stained with 1 ug/ml Cy3-streptavidine (Amer sham Biosciences, Buckinghamshire, UK) and eventually scanned with an Illumina BeadArray Reader.

2.3. Immunohistochemical determinations
Formalin-fixed, paraffin-embedded whole tumor sections (4 μm thick) were dewaxed in xylene and rehydratated, and

---

**Figure 1** – Workflow of the analyses.
tissue sections stained with haematoxylin and eosin were evaluated by an expert Pathologist with respect to the extent of lymphocyte infiltration (LI) with a 0 to 3 score (none, mild, moderate or marked).

Tumor sections were also processed for autoclave antigen retrieval at 95 °C in 10 mM Na-citrate buffer for 6 min. A rabbit polyclonal anti MX1 antibody (Sigma Life Science, St Louis, Mo) was used at a 1:200 dilution and incubated for 1 h at room temperature. Negative controls were run on each slide by omitting the primary antibody. The Ultravision detection system HPR Polymer (Thermo Scientific, Astmoor, UK) was used for developing. Slides were counterstained with haematoxylin and examined under a light microscope by two independent investigators.

2.4. Data analysis and statistics

2.4.1. Array processing
Microarray raw data where generated using the Illumina BeadStudio 3.8 software and processed using the lumi package (Du et al., 2008) of Bioconductor. After quality control, the Robust Spline Normalization was applied and probes with a detection of \( p < 0.01 \) in at least one sample were selected. Array data of the metagene test dataset were deposited at the Gene Expression Omnibus data repository (GSE37181).

2.4.2. Metagenes identification
After applying the normalization and filtering procedures described above, the most variable probes (IQR>0.5;11334 probes) were selected and subjected to hierarchical clustering using 1-Pearson correlation as distance metrics and an average linkage method. A correlation threshold of \( r > 0.7 \) was used and only clusters containing at least 7 probes were considered.

Each probe cluster identified in the metagene identification dataset described above was re-clustered in our metagene test dataset. Probes with correlation coefficients \( \leq 0.7 \) were removed and only clusters with at least 3 probes left after re-clustering were considered. Probe signals were median-scaled and, for each probes cluster, a metagene was computed as the mean expression of the probes in the cluster. A dendrogram of all the genes also showing the applied cut-off is reported in Supplementary Figure 2.

To define clusters on Affymetrix data, probe sets targeting the genes belonging to each cluster of interest were re-clustered on the validation dataset and only those with a correlation \( >0.7 \) were retained and used to compute the metagene values as described.

2.4.3. Breast cancer subtypes identification
Patients were subdivided into three subtypes based on ESR1 and ERBB2 expression levels in a way similar to that described in (Bianchini et al., 2010a): ESR1−/ERBB2− (roughly corresponding to basal-like subtype), ESR1+/−/ERBB2+ (roughly corresponding to HER2+ enriched subtype), ESR1+/ERBB2− (roughly corresponding to luminal A and B subtypes). The ILMN_15142 and ILMN_28003 probes for Illumina data and the “205225_at” and “216836_s_at” probe sets for Affymetrix data were considered as reporters for, respectively, ESR1 and ERBB2 gene expression. The threshold values to define gene expression positivity were selected according to the strong bimodal distribution observed.

2.4.4. Class comparison
The class comparison analyses were performed using a linear modeling approach and empirical Bayes methods as implemented in the limma Bioconductor package (R version 2.10) (Smyth, 2004). A two-sided \( p < 0.05 \) was considered significant if not otherwise specified.

2.5. Survival analysis
Univariable and multivariable Cox regressions, as implemented in the survival Bioconductor package, were used to correlate metagenes with outcome. Metagenes values were dichotomized defining as positive samples those with expression values higher than 67th percentile. Such a cut-off roughly corresponded to the deviation from normality observed in the metagene distributions. Kaplan–Meier survival curves were also plotted using the same cut-offs. Distant Metastasis Free Survival (DMFS) was the main endpoint and survival differences were evaluated using the log-rank test.

3. Results

3.1. Identification of prognostic metagenes in the metagene test dataset
All analyses were carried out considering the expression of metagenes rather than single genes. The metagene approach is very effective in simplifying biological interpretation of data, reducing multiple testing-associated statistical issues and allowing an easier cross-platform analysis.

Clusters of highly correlated genes were first identified by hierarchical clustering as described in the Material and Methods section. Of the 84 clusters identified in the metagene identification dataset of 105 GEP from consecutive BC (see Supplementary Table 1 for clinicopathological information), 79 were validated on the independent metagene test dataset of 123 BC. A complete list of these clusters and of genes belonging to each one is reported in Supplementary file 1. For each cluster, a metagene was computed as the mean expression of all probes after median centering of the signals.

For all derived metagenes, their association with metastasis was tested in the metagene test dataset consisting of 123 node-negative BC patients who developed distant metastases within 5 years of surgery (59 patients) or were free of distant metastases (64 patients) for at least 5 years. In order to control the effect of classical prognostic factors on clinical outcome, patients were selected to obtain similar distributions for age \( t \)-test, \( P = 0.8200 \) and tumor size \( t \)-test, \( P = 0.2600 \) between the subsets “relapsed in distant sites” and “disease-free”. In keeping with the criticism associated with identification of prognostic signatures on patients not subdivided by BC subtype, we ran all subsequent analyses focusing on the subset of 92 patients with ESR1+/−/ERBB2− tumors. To identify such a subset in a possibly homogeneous way, ER and HER2 statuses were defined based on the strong bimodal expression of the
corresponding probes (ILMN_15142, ILMN_28003) overcoming in this way problems associated with technical variations in ER and HER2 determinations over the years. Again, age (P = 0.2110, t-test) and size (P = 0.6400, t-test) were not significantly different between the two subgroups of patients developing (N = 42) or not (N = 50) distant metastases. Data are reported in Table S1.

By class comparison analysis, only the IFN metagene (metagene #42) was found to be differentially expressed at statistically significant level (P = 0.0290, modified t-test, N = 92) between tumors from patients relapsed or disease-free, with a higher expression level in the former group (Supplementary Figure 3). This metagene (Table 1) includes several genes, collectively known as interferon-stimulated genes (ISGs), whose products are classically linked to the host innate immune response to viral infections. Among those are IFIT (IFN-induced protein with tetratricopeptide repeats) genes encoding IFTT1, IFTT2 and IFFT3, recently shown to form complexes able to bind and inhibit specific viral nucleic acids (Pichlmair et al., 2011); OAS (2'-5'-oligoadenylate synthetase) genes encoding OAS1, OAS2 and OAS3, proteins which cause viral RNA degradation through their ability to synthesize 2',5'-oligoadenylates; the MX1 (myxovirus resistance 1) gene encoding a protein that physically inhibits the assembly of viral particles; genes encoding members of the IFN-regulated ubiquitin-like modifier response (ISGylation; HERC5 and ISG15) which bind numerous protein substrates, modulate pleiotropic cellular response and negatively affect viral protein synthesis (Sadler and Williams, 2008).

To exclude that such findings could be affected by percentage of tumor cells in the samples we verified that average tumor cell content was the similar in the two groups (71% for samples from disease-free patients and 75% for samples from patients who developed distant metastasis) and that the IFN metagene was not significantly correlated with tumor cell percentage (ρ0 = 0.1206; P = 0.2033).

To investigate whether expression of the IFN metagene was associated with LI, we determined the LI score (ranging from 0 to 3) on histological sections of all samples. There was no correlation between LI score and IFN metagene expression (P = 0.4933, modified F-test) and ESR1+/ERBB2− tumors with higher LI (scores 2 and 3) were frequently metastasis free (45.6% disease free vs 30% with metastasis, P = 0.1360, χ² test).

To establish the identity of cells expressing ISGs, the subset of ESR1+/ERBB2− tumors with the highest (3rd tertile) expression of the IFN metagene was stained for MX-1 protein expression. A representative immunostained section is reported in Figure 2A. MX1 was almost always expressed by a higher percentage of tumoral cells than stromal cells, and only in 3 out of 22 cases the stromal expression prevailed (Figure 2B).

### 3.2. Validation of the prognostic relevance of the IFN metagene

To validate the prognostic role of the IFN metagene in ESR1+/ERBB2− tumors and to investigate its role in the other BC subtypes we used a combined dataset obtained by pooling 3 publicly available untreated node-negative BC datasets. This allowed us to separately test the IFN metagene on the clinical outcome of 457 ESR1+/ERBB2−, 104 ERBB2+ and 123 ESR1−/ERBB2− tumors, and to overcome the size limits of our discovery dataset concerning non-ESR1+/ERBB2− tumors. Clinico-pathological information on the three datasets is reported in Supplementary Table 2.

Univariable Cox regression analysis for the IFN metagene (Supplementary file 2) was performed separately for each BC subtype (Figure 3A). A significant increase in metastasis risk

<table>
<thead>
<tr>
<th>Table 1 — Annotation of the IFN cluster.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Illumina probe ID</strong></td>
</tr>
<tr>
<td>ILMN_1797001</td>
</tr>
<tr>
<td>ILMN_179749</td>
</tr>
<tr>
<td>ILMN_1654639</td>
</tr>
<tr>
<td>ILMN_1835092</td>
</tr>
<tr>
<td>ILMN_2058782</td>
</tr>
<tr>
<td>ILMN_1760062</td>
</tr>
<tr>
<td>ILMN_1723912</td>
</tr>
<tr>
<td>ILMN_234779</td>
</tr>
<tr>
<td>ILMN_1781373</td>
</tr>
<tr>
<td>ILMN_1699331</td>
</tr>
<tr>
<td>ILMN_1739428</td>
</tr>
<tr>
<td>ILMN_2239754</td>
</tr>
<tr>
<td>ILMN_2054019</td>
</tr>
<tr>
<td>ILMN_1662358</td>
</tr>
<tr>
<td>ILMN_2410826</td>
</tr>
<tr>
<td>ILMN_1672606</td>
</tr>
<tr>
<td>ILMN_1736729</td>
</tr>
<tr>
<td>ILMN_2184262</td>
</tr>
<tr>
<td>ILMN_1681721</td>
</tr>
<tr>
<td>ILMN_1857871</td>
</tr>
<tr>
<td>ILMN_1814305</td>
</tr>
</tbody>
</table>
3.3. Relation of the IFN metagene with the immune axis

In addition to eliciting anti-viral response through induction of ISGs genes, interferons are also key regulators of adaptive immunity and have emerged as central coordinators of tumor-immune-system interactions. Given the frequent involvement of immune-response related genes in published BC signatures with prognostic value (Bianchini et al., 2010b; Gu-Trantien et al., 2013; Nagalla et al., 2013; Rody et al., 2009; Teschendorff et al., 2007) we analyzed, in the above defined BC subtypes, the prognostic impact of a T-cell metagene (metagene#55, Supplementary file 2) compared with our IFN metagene.

The T-cell metagene was composed of 22 genes (ARHGAP15, CCL5, CD2, CD247, CD27, CD3D, CD48, CD53, CORO1A, EVI1B, GZMA, GZMK, HCLS1, IL10RA, IRF8, ITK, LFXN, LTβ, PLAC8, RAC2, SELL, SH2D1A), 13 of which were involved in different aspects of innate and T-cell mediated adaptive immunity such as the CD3/T cell receptor complex (CD3D, CD247) of T-lymphocytes, serine protease components necessary for lysis of target cells (GZMA, GZMK) by cytotoxic CD8+ T lymphocytes and natural killer (NK) cells, and surface antigens providing co-stimulatory signals (CD2, CD48).

T-cell metagene expression was associated with a reduced risk of distant metastases (Figure 4A) both in ERBB2+ (HR = 0.44, 95% CI 0.22–0.88, P = 0.0214) as well as in ESR1+/ERBB2– tumors (HR = 0.40, CI 0.23–0.71, P = 0.0016) and similar results were obtained by varying the cut-off (Supplementary Figure 4C and D). As also confirmed by Kaplan–Meier analysis (Figure 4B), high T-cell metagene expression was not able to significantly affect DMFS in women with ESR1–/ERBB2– tumors (P = 0.2821), but it was associated with a longer DMFS in patients with ERBB2+ (P = 0.0179) and with ESR1+/ERBB2– tumors (P = 0.0011). T-cell and IFN metagenes have been shown to impact similarly on prognosis in patients with ESR1–/ERBB2– and ERBB2+ tumors, but seem to play opposite roles in ESR1+/ERBB2– tumors.

To directly compare the roles of these two metagenes, we performed multivariable analysis for T-cell and IFN metagenes in the three subtypes of tumors (Table 2). In women with ESR1+/ERBB2– tumors the two metagenes did maintain an independent but opposite prognostic relevance on outcome; the T-cell metagene was found to be associated with a 2.94-fold reduction in metastasis risk (HR = 0.36, 95% CI 1.24–2.74, P = 0.0027). On the contrary, a significant protective role was observed for the IFN metagene (HR = 0.36, 95% CI 0.15–0.87, P = 0.0236) and a protective trend for the T-cell metagene (HR = 0.51, 95% CI 0.25–1.05, P = 0.0678) in ERBB2+ tumors. Such results are in agreement with the different pattern of correlation between these two metagenes according to BC subtype, with a correlation stronger in ESR1–/ERBB2– and ERBB2+ tumors (r = 0.50, r = 0.52) and weaker in ESR1+/ERBB2– tumors (r = 0.35), where high expression of the IFN metagene, but low expression of T-cell genes was observed, as reported in Figure 4C.

The combined effect of T-cell and IFN metagenes on DMFS was also studied. In women with ESR1+/ERBB2– tumors, higher IFN metagene expression levels were associated with
increased risk ($HR = 1.32, 95\% CI 1.12–1.57, P = 0.0013$) only when the tumors were classified as T-cell low whereas in tumors with high expression of the T-cell metagene, the level of IFN metagene expression did not modify the risk ($HR = 1.16, 95\% CI 0.64–2.10, P = 0.6154$). Therefore in Kaplan–Meier curves (Figure 4D) patients with tumors expressing high T-cell levels were not segregated according to IFN metagene expression. As expected, in patients with basal tumors the 10-year DMFS was not affected by the expression of such metagenes. Conversely, in women with ERBB2+ and ESR1+/ERBB2− tumors a high expression of T-cell metagene was associated with a longer DMFS whereas when the T-cell metagene scored low, the outcome was affected by the IFN metagene expression, but with an opposite pattern. In fact, high levels of IFN genes are associated with a risk of relapse low in women with ERBB2+ tumors but high in those with ESR1+/ERBB2− tumors.

Since the prognostic role of the IFN metagene has been initially identified in a dataset ‘balanced’ for classical risk factors, we attempted to confirm whether the IFN metagene was an independent prognostic factor. A multivariable analysis was therefore carried out on patients with ESR1+/ERBB2− tumors using two of the three public datasets (one did not contain information on size) and including the following variables: size, age, proliferation (derived from the genomic grade index), T-cell metagene and IFN metagene. Data are reported in Table 3.

Also in the presence of other prognostic variables, known to be predictive of distant metastasis, an increased expression of IFN metagene was associated with a higher risk ($HR = 2.22, 95\% CI: 1.14–4.30, P = 0.0185$).

Finally, acknowledging the subtype-dependent association with outcome for the IFN metagene, we performed an exploration analysis to evaluate whether it was also associated with the site of metastasis. Only a trend for a higher association with bone metastases was found in the metagene test dataset focusing on ESR1+/ERBB2− tumors, but it was not confirmed in the GSE2034 public data (data not shown), which was the only one containing information on the site of metastasis.
3.4. Biological interpretation of the role of the IFN metagene

To understand the biological basis for the different prognostic role of the IFN metagene in distinct BC subtypes, we selected from the validation dataset tumors with high (>4th quartile) versus low (<1st quartile) expression of the IFN metagene and ran a class comparison analysis separately for each subtype and for ESR1+/ERBB2-- tumors with low T-cell metagene expression.

Probesets both up- and down-regulated showing a fold change $>1.5$ and false discovery rates $<1\%$ were selected as differentially expressed (Table S3) and canonical pathways enriched in the differentially expressed genes compared with all tested genes were identified using Ingenuity Pathway Analysis. By comparing the enrichments in all subtypes, almost all enriched pathways included genes involved in adaptive and innate immunity (Figure 4), except for ESR1+/ERBB2-- tumors with low T-cell metagene expression. Interestingly, in this latter subgroup of tumors, characterized by a relation between high IFN gene expression and worse DMFS, genes differentially expressed as a function of IFN metagene showed an enrichment in a proliferation-related pathway 'Mitotic Roles of Polo-Like Kinase', which was in fact also present among the enriched terms of ESR1+/ERBB2-- tumors, and a DNA damage pathway 'Cell Cycle: G2/M DNA Damage Checkpoint Regulation' (Figure 5).

Finally, the genes belonging to the IFN metagene are known to be activated by IFN-α/β and by their Type I and II receptors. Therefore correlations between the IFN metagene and a series of IFN related genes were investigated to get further insight into the biological role of the IFN metagene. Globally, a low ($r$=0.21--0.28), but statistically significant correlation ($P<0.0215$) was observed between the IFN metagene expression and the IFNγ gene in all subtypes. Other significant positive correlations were observed for IFNα/β receptor II in ESR1+/ERBB2-- and in all tumors, but also for ERBB2+ tumors when using a different probe. In ESR1+/ERBB2-- tumors a correlation was also observed with the IFNγ receptor I (Supplementary Figure S5).

4. Discussion

The present study was planned to identify genes linked to the risk of developing metastases in patients with node-negative BC in order to avoid inter-subtype confounding effects and to control the effect of classical prognostic factors. Under such a perspective, we report the novel finding that the risk for distant metastases in ESR1+/ERBB2-- tumors, characterized by a less aggressive disease, is linked to the expression of an IFN metagene.

Our IFN metagene includes ISGs which are classically upregulated in response to type I interferons, IFNα and IFNβ and to lesser extent to Type II IFNγ, in response to viral infection (Pichlmair et al., 2011; Sadler and Williams, 2008). The same genes can also be expressed as a result of IFN-mediated immune responses. The association with prognosis emerged through the consideration of the metagene instead of the single genes, and the levels of expression of neither IFN Type I family members (which included twelve different subtypes for IFNα/β) nor IFNγ or their receptors, were singly found, to confer increased risk of developing metastases in ESR1+/ERBB2-- tumors (data not shown). Immunohistochemistry indicated that MX1, one of these ISGs, was mainly expressed by neoplastic cells compared to cells of the lymphocytic infiltrate.

The relevance for IFN-regulated genes among primary human BC was already pointed out in gene expression data published starting in 1999 by Perou and colleagues (Hu et al., 2006; Perou et al., 1999, 2000). A subsequent study, by considering the van’t Veer dataset, reported overexpression of a group of ISG by 40% of breast cancer samples (Einar et al., 2005). Furthermore, higher expression of ISGs such as IFI27, IFIT1, IFI35, MX1, GP2 and ISG15 proved to be predictive of adverse clinical outcome following radiation and chemotherapy (Weichselbaum et al., 2008). More recently, high expression of ISG has been described in cancer cells of the NCI-60 panel and in a different set of BC cell lines confirming that this pathway can be activated in neoplastic cells independently of the microenvironment. In such cell lines higher expression of ISG predicted greater sensitivity to TNFα-related apoptosis inducing ligand (TRAIL)-mediated cytotoxicity. Although the mechanism leading to activation of an IFN-related pathway remains to be investigated, all together such evidence indicate that differences in expression levels of IFN-induced genes in cancer cells correlate with distinct biological properties. The higher sensitivity to TRAIL of neoplastic cells with significantly higher expression of ISGs can have clinical implications since it suggests a benefit by recombinant human TRAIL (Chen et al., 2012) for patients with high-risk BCs.

It is important to emphasize that the metastasis-promoting effect of the IFN metagene is limited to ESR1+/ERBB2-- tumors. In fact in the validation dataset, which provided us with a sufficient number of cases to separately analyze each BC subtype (showing as expected different DMFS, Supplementary Figure S6), expression of IFN genes was protective in ERBB2+ tumors whereas the prognosis of the more aggressive basal tumors was not affected by IFN metagene levels. This suggests that ISGs may have an opposite role on tumor progression according to the specific biological background of each distinct breast cancer subtype, bestowing metastatic potential on ESR1+/ERBB2-- tumors while being protective for ERBB2+ tumors and irrelevant for ESR1--/ERBB2-- tumors.

ISG are not only part of antiviral pathways, but are also involved in tumor-immune cells interactions. Several studies have indicated a prognostic value of immune-function genes. The paradoxical role of the immune system during cancer development (de Visser et al., 2006) and metastasis (DeNardo et al., 2008) has been extensively documented for several tumor types. Our results should indicate that in BC the opposite effects of immune genes on prognosis may be related to the specific biological background of each distinct BC subtype. The involvement of genes related to immune response among those predicting metastasis in ER+ tumors has already been reported (Bianchini et al., 2010b; Yu et al., 2007). However, to better understand the prognostic role played by immunity genes it is important to identify the involved genes, and
Figure 4 – (A) Forest plots showing the HRs of the DMFS univariable Cox regression analyses for the T-cell metagene (33% of patients with tumors with the highest T-cell metagene expression were defined as T-cell+, the remaining patients were considered with T-cell- tumors) in the
In fact, Rody et al. (2009) observed that a T-cell gene module was prognostic in patients with ER− tumors and ER+/HER2+ tumors, which should indicate that the protective effect of T-cells may be confined only to highly proliferating tumors. In that dataset, B-cell gene modules did not associate with the risk of metastasis, whereas a B-cell metagene proved to play an important role in highly proliferating tumors from distinct datasets analyzed by Schmidt et al. (Schmidt et al., 2008). Results recently reported by Nagalla et al. (2013) were similar. An immune module with prognostic relevance confined to ERtumors was also reported by Teschendorff et al. (2007) who analyzed three different datasets and emphasized that not all ER− tumors have a poor prognosis, and that down-regulation of a seven-gene module in such tumors is associated with a two-fold risk of metastasis. An IFN gene (STAT1 module) associated with favorable clinical outcome in basal tumors (ESR1−/HERBB2−) was instead reported by Abraham et al. (2010) using an approach based on gene sets rather than on single genes, and recently a novel mechanism centered on STAT1 up-regulation and recruitment of myeloid-derived suppressor cells has been shown to drive BC progression (Hix et al., 2013).

Considering all these findings, we also planned to investigate the prognostic role of a T-cell metagene and its relationship with the IFN metagene in the different BC subtypes. In our analysis of public datasets we confirmed previous results (Alexe et al., 2007) on the protective role of the presence of a T-cell metagene in HER2+ tumors, and extend such evidence to ESR1+/ERBB2− tumors, regardless of the expression level of the IFN metagene. A closer look at the gene expression pattern of tumors with low compared to those with high expression of IFN metagenes indicated that in all BC subtypes, with the exception of ESR1+/ERBB2− tumors with low T-cell expression, the majority of differentially expressed genes belongs to canonical pathways related to innate and adaptive immunity. In ESR1+/ERBB2− tumors, an increased expression level of the IFN metagene was instead associated with an enrichment in genes belonging to proliferation-related pathways with a mechanism that still needs to be clarified.

Results from our study thus add to the discordant data already reported in the literature, underlining the fact that the role of immune gene modules in BC is complex. Discordant results may derive from heterogeneity in treatments and analytical approaches, but also in the different methods used to identify BC subtypes. Breast cancer is a very heterogeneous disease characterized by clinical and biological complexity. The importance of the biological complexity has been increasingly recognized thanks to the contribution of molecular studies which were able to refine the rough classification based on immunohistochemical detection of a few biomarkers. Altogether our data confirm that it is of pivotal importance to analyze the contribution of prognostic variables in homogeneous patient subgroups taking into account tumor molecular entities since it is likely that early drivers of metastatic processes may be different in the different subtypes. In this context, it is interesting to underline the dual role played by the association between low T-cell metagene/high IFN metagene, which is characterized by a strong activation of immune response pathways that overall appears to be protective on ERBB2+ tumors, and conversely linked to genes related to cell proliferation and to a worse prognosis in ESR1+/ERBB2− tumors.

Finally, our study shows that activation of an IFN metagene represents a novel feature associated with an unfavorable outcome in ESR1+/ERBB2− tumors. Biological complexity captured by classification based on intrinsic genes is not sufficient to fully justify the clinical heterogeneity of BC and needs therefore to be complemented with information specifically addressing host/environment-related aspects as the latter can completely modify the prognosis predicted by the BC subtype itself and, as shown in our study, can play a different role accordingly.

| Table 2 – IFN and T-cell metagenes multivariable Cox analysis. |
|-----------------|-----------------|-----------------|
| HR              | P               | 95% CI          |
| ESR1−/ERBB2− T-cell (high vs low) | 0.81 | 0.5024 | 0.43−1.51 |
| IFN (high vs low) | 0.75 | 0.3711 | 0.40−1.41 |
| ERBB2+ T-cell (high vs low) | 0.51 | 0.0678 | 0.25−1.05 |
| IFN (high vs low) | 0.36 | 0.0236 | 0.15−0.87 |
| ESR1+/ERBB2− T-cell (high vs low) | 0.34 | 0.0002 | 0.19−0.61 |
| IFN (high vs low) | 1.84 | 0.0027 | 1.24−2.74 |

Bold values indicate significant P-values.

| Table 3 – Multivariable Cox analysis in ESR1−/ERBB2− samples. |
|-----------------|-----------------|-----------------|
| HR              | P               | 95% CI          |
| IFN (high vs low) | 2.22 | 0.0185 | 1.14−4.30 |
| T-cell (high vs low) | 0.18 | 0.0006 | 0.07−0.48 |
| GSI (high vs low) | 3.73 | 0.0001 | 1.90−7.33 |
| Size (<2 vs ≤2) | 2.01 | 0.0270 | 1.08−3.73 |
| Age (>50 vs ≤50) | 0.98 | 0.9664 | 0.53−1.81 |

Bold values indicate significant P-values.
Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

We thank L. Ventura and L. Gioiosa for skillful assistance in IHC determinations, and personnel and Committees of the INT Tissue Bank for providing breast cancer specimens.

This work was supported by grants from the Associazione Italiana per la Ricerca sul Cancro (AIRC, the Italian Association for Cancer Research, grant number 10611 to M.G.D. and Call for Proposals on Innovative tools for cancer risk assessment and early diagnosis), by the Italian Ministry of Health (Special Project on Female Cancers), and also by funds obtained through an Italian law that allows taxpayers to allocate the “5/C2100” share of their payments to support a research institution of their choice.

Appendix A.
Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2014.04.010.

REFERENCES
