Subtype-Specific Metagene-Based Prediction of Outcome after Neoadjuvant and Adjuvant Treatment in Breast Cancer

Maurizio Callari¹, Vera Cappelletti¹, Francesca D’Aiuto¹, Valeria Musella¹, Antonio Lembo², Fabien Petel³, Thomas Karp³, Takayuki Iwamoto⁵, Paolo Provero⁵, Maria Grazia Daidone¹, Luca Gianni⁷, and Giampaolo Bianchini⁷

Abstract

Purpose: In spite of improvements of average benefit from adjuvant/neoadjuvant treatments, there are still individual patients with early breast cancer at high risk of relapse. We explored the association with outcome of robust gene cluster-based metagenes linked to proliferation, ER-related genes, and immune response to identify those high-risk patients.

Experimental Design: A total of 3,847 publicly available gene-expression profiles were analyzed (untreated, N = 826; tamoxifen-treated, N = 685; chemotherapy-treated, N = 1,150). Genes poorly performing in formalin-fixed samples were removed. Outcomes of interest were pathologic-complete response (pCR) and distant metastasis-free survival (DMFS). In ER-HER2+, the proliferation and ER-related metagenes were combined to define three risk groups. In HER2+ and ER-HER2+ risk groups were defined by tertiles of an immune-related metagene.

Results: The high-proliferation/low-ER group of ER-HER2+ breast cancer had significantly higher pCR rate [OR, 5.01 (1.76–17.99), P = 0.005], but poorer outcome [HR = 3.73 (1.63–8.51), P = 0.0018] than the low-proliferation/high-ER. A similar association with outcome applied to patients with residual disease (RD) after neoadjuvant chemotherapy (P = 0.01). In ER-HER2+ and HER2+ breast cancer, immune metagene in the high tertile was linked to higher pCR [33.7% vs. 11.6% in high and low tertile, respectively; OR, 3.87 (1.79–8.93), P = 0.0009]. In ER-HER2+, after adjuvant/neoadjuvant chemotherapy, 5-year DMFS was 85.4% for high-tertile immune metagene, and 43.9% for low tertile. The outcome association was similar in patients with RD (P = 0.0055). In HER2+ breast cancer treated with chemotherapy the association with risk of relapse was not significant.

Conclusions: We developed metagene-based predictors able to define low and high risk of relapse after adjuvant/neoadjuvant therapy. High-risk patients so defined should be preferably considered for trials with investigational agents. Clin Cancer Res; 22(2): 337–45. © 2015 AACR.

Introduction

Early breast cancer is a molecularly, biologically, and clinically heterogeneous disease (1–3). The prognosis in patients with early breast cancer has significantly improved over the last two decades by introducing new adjuvant treatments with an "add-on" strategy (4, 5). For instance, the addition of anthracyclines to early polychemotherapy regimens improved the chance of cure, which was further increased by the later addition of taxanes to the anthracycline-containing regimens (5). However, the drawback of this strategy is that at each sequential step of treatment improvement the portion of overtreated patients increases due to the progressive decrease of the residual risk, whereas for the same reason similar relative benefits translate into progressively smaller absolute benefits (6). Therefore, new drug development in early breast cancer has become challenging. Indeed, to demonstrate an additional benefit over an overall relative good outcome, very large clinical trials are needed to provide statistically significant results, and a small average incremental benefit might appear nonclinically meaningful or the related treatments may be deemed cost-ineffective. In this context, there is an urgent need for the identification of prognostic and predictive biomarkers able to distinguish patients who will do very well with standard treatments and could be excluded from trials with investigational drugs, from patients who will have a significant residual risk despite standard treatment. These biomarkers would improve and optimize the design of clinical trials and increase the chance of a successful development of new drugs in the early setting (7).

A large number of gene expression profiles (GEP) have been generated during the last decades to discover, develop, and
validate prognostic and predictive gene signatures. Some of these signatures are commercially available to define the residual risk in ER+HER2− tumors after receiving adjuvant endocrine treatment (Mammaprint, Oncotype DX, Breast Cancer Index, PAM50, EndoPredict; refs. 8–12). None of these signatures was specifically aimed or assembled to define the residual risk after standard chemotherapy or chemoendocrine therapy.

A general bottleneck in the development of new signatures is the difficulty to reach the level of evidence required for their clinical implementation by running expensive and long prospective clinical trials (13, 14). To overcome this limitation, it was suggested that the retrospective use of samples collected within prospective clinical trials could more efficiently provide this level of evidence (15). Formalin-fixed, paraffin-embedded (FFPE) samples are routinely archived within clinical trials. Although obtaining reliable GEPs from FFPE samples using commercially available chip (i.e., Affymetrix and Illumina) has been considered challenging for a long time, we and others have recently shown that it is feasible (16–18), in particular, by improving and optimizing the processing approach (19). However, predictors developed using frozen samples would underperform on FFPE-derived GEPs due to the unpredictable lower performances of some probesets.

In this study, building on the knowledge acquired during the last decade on the most relevant prognostic and predictive factors in breast cancers, we aimed to develop metagene-based risk predictors (MBRPs) suitable for application on FFPE-derived GEPs with the objective to predict the risk of relapse in patients receiving neoadjuvant/adjuvant chemotherapy followed by endocrine treatment as appropriate. In particular, we were interested in defining patients at high recurrence risk despite standard treatment, which could be ideal candidates for trials with investigational drugs in early breast cancer. Coherently with our previous works (20–22), these MBRPs were developed and tested separately for the three main breast cancer subtypes (ER+HER2−, HER2+, and ER−HER2).

**Materials and Methods**

**Data sets collection and processing**

We collected a total of 25 publicly available breast cancer data sets (N = 3847) from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress website (http://www.ebi.ac.uk/arrayexpress/). These data sets contained expression profiles generated from the Affymetrix GeneChip Human Genome U133A or U133 Plus 2.0 chips. Data sets were grouped according to the clinical information available and the type of treatment received to generate four distinct groups: (i) the GENERIC collection (N = 1186), including samples without meaningful clinical information available; (ii) the PROGNOSTIC collection (N = 826), including node negative patients with early breast cancers not treated with any systemic therapy until relapse; (iii) the TAM (tamoxifen-treated) collection (N = 685), including breast cancer profiles from patients receiving 5 years of adjuvant tamoxifen; and (iv) the CHEMO collection (N = 1150), including patients treated with either neoadjuvant or adjuvant chemotherapy followed by endocrine treatment as appropriate (Table 1, Supplementary Section S1). Only two patients with HER2+ tumor were treated with adjuvant trastuzumab (CHEMO collection, Petel study, E-MTAB-365). We also used an additional series of 44 samples for which GEPs were derived from FFPE tissue (18).

---

**Table 1.** Summary of public data sets and number of samples included in the study

<table>
<thead>
<tr>
<th>Collection</th>
<th>Data set ID</th>
<th>Data set name</th>
<th>Number of samples</th>
<th>Suitable samples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic</strong></td>
<td>GSE2109</td>
<td>expO</td>
<td>353</td>
<td>301</td>
</tr>
<tr>
<td></td>
<td>GSE3744</td>
<td>Richardson</td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>GSE5460</td>
<td>Lu</td>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>GSE0780</td>
<td>Chen</td>
<td>185</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>GSE12276</td>
<td>Massague</td>
<td>204</td>
<td>204</td>
</tr>
<tr>
<td></td>
<td>GSE12763</td>
<td>Hoeflich</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>GSE13787</td>
<td>Marty</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>GSE6391</td>
<td>Desmedt</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>GSE6195</td>
<td>Li</td>
<td>115</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>GSE20711</td>
<td>Dedeuwaerder</td>
<td>90</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>GSE23653</td>
<td>Sabatier</td>
<td>266</td>
<td>266</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1,495</td>
<td>1,186</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prognostic</strong></td>
<td>GSE2034</td>
<td>Wang</td>
<td>286</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>GSE2990</td>
<td>Sotiriou</td>
<td>189</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>GSE5327</td>
<td>Minn</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>GSE7390</td>
<td>Transig</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td>GSE121</td>
<td>Mainz</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>931</td>
<td>826</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TAM</strong></td>
<td>GSE9195</td>
<td>Loi</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>GSE6532</td>
<td>Loi2</td>
<td>277</td>
<td>277</td>
</tr>
<tr>
<td></td>
<td>GSE12093</td>
<td>Zhang</td>
<td>136</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>GSE17075</td>
<td>Symmans</td>
<td>289</td>
<td>195</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>788</td>
<td>685</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CHEMO</strong></td>
<td>GSE25055</td>
<td>Hatzis_disc</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td></td>
<td>GSE25065</td>
<td>Hatzis_val</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td>E-MTAB-365</td>
<td>Petel</td>
<td>537</td>
<td>243</td>
</tr>
<tr>
<td></td>
<td>GSE6446</td>
<td>Desmedt</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>GSE4998</td>
<td>Horak</td>
<td>279</td>
<td>279</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1,444</td>
<td>1,150</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Raw signals were processed using RMA normalization and an alternative Chip Description File (CDF) as previously described (19). We considered only the probes common between U133A and U133 Plus 2.0 chips as detailed in Supplementary Section S2.

Definition of molecular subtypes

Three molecular subtypes were defined according to ER and HER2 status: ER+/HER2−, HER2+, and ER−/HER2+. Within the HER2+ subtype, we also assessed outcome predictors in ER+ and ER− groups. A key step in our strategy was to develop outcome predictors specifically in each subtype. To define consistently ER and HER2 status in all samples avoiding the heterogeneous assessment across data sets, we developed two metagene-based predictors to define ER and HER2 status as detailed in the Supplementary Section S3. The metagene-based assessment was highly concordant with standard pathologic assessment where this information was available (Supplementary Figs. S2–S5). Seven genes (CCDC170, ESR1, ETV1, ABAT, SLC39A6, GATA3, and SCUBE2) were included in the ER status and 10 (ERBB2, PAP3, STARD3, GBR7, PNM1, PSMD3, GSDMB, RPL19, FGFR4, and CAP1) in the HER2 status predictor, respectively.

Strategy for development and refinement of MBRPs

We first used the GENERIC collection of samples and a data-split approach to identify robust clusters of genes with a reciprocal correlation higher than 0.4. Their composition was subsequently refined by removing genes showing a correlation below the threshold in our series of FFPE-derived GEPs (ref. 18; Supplementary Section S4). In this step, we used the correlation as a simple metric for the assessment and removal at the very beginning of probesets that have suboptimal hybridizations performance on Affymetrix, when GEPs are derived starting from fragmented FFPE-derived mRNA. After a Gene Ontology evaluation of the genes within each cluster, three clusters were selected based on their known relevant prognostic and/or predictive biologic functions in specific molecular subtypes (immune system, proliferation, and ER-related genes; refs. 1, 23, 24). These clusters were used to develop subtype-specific MBRPs in the PROGNOSTIC and TAM collection. Instead of the commonly used approach of simply calculating the average expression of all the cluster genes (unrefined metagene; refs. 20, 21, 25, 26), we introduced a refinement step to select with a cross-validation approach the optimal number of genes to maximize the prognostic/predictive performance (refined metagenes or MBRPs; Supplementary Section S5). Metagene scores were calculated as the average of the expression of the selected genes without fitting any weight.

Statistical analyses

Survival analysis. Univariable and multivariable Cox regression were used to correlate metagenes or clinicopathologic variables with outcome (survival R package). Concordance indices (c indices), as computed by the coplot function, were used to evaluate the refinement procedure efficacy. Results were also plotted using the Kaplan–Meier method (rms R package) and differences tested by log-rank test. Genomic predictors were categorized by tertiles if not otherwise specified. Distant metastasis-free survival (DMFS) was the main outcome endpoint. For consistency, all data sets were right censored at 5 years because longer term outcome was not available for the CHEMO collection and because late relapse seemed to be associated with different molecular features (27).

Logistic regression. Logistic regression analysis was performed to associate metagene scores by tertiles (high, intermediate, and low) with pathologic complete response (pCR). OR, 95% confidential interval (CI), and P values were derived from the fitted model and plotted as forest plot.

Results

Overall workflow and patient population

The general workflow of the analysis used to derive breast cancer MBRPs is illustrated in Fig. 1. A total of 3,847 GEPs of invasive breast cancers were used in our study (Table 1). These samples were grouped into four collections (GENERIC, PROGNOSTIC, TAM, and CHEMO) as detailed in the Materials and Methods section. Clinicopathologic features for the PROGNOSTIC, TAM, and CHEMO series are summarized in Supplementary Tables S1–S3. All samples were stratified in ER−/HER2–, HER2+, and ER−/HER2+ subtypes.

Definition of MBRPs

Three robust gene clusters representative of biologic functions with known associations with clinical outcomes (proliferation, immune-related, and ER-related clusters), were identified in the GENERIC collection. Genes poorly performing in FFPE-derived samples were removed (i.e., genes lacking the expected correlation; Supplementary Section S4). These FFPE-adapted clusters (thereafter called unrefined metagenes) were used to generate subtype-specific MBRPs (thereafter also called refined metagenes).

In ER−/HER2– and HER2+ tumors, biomarkers associated with immune functions have been reported as prognostic and predictive (21, 24–26, 28–31). Therefore, starting from the unrefined FFPE-adapted immune cluster we developed two subtype-specific refined immune metagenes (Supplementary Section S5). In untreated ER−/HER2+ cases (n = 179, PROGNOSTIC collection), 25 genes were selected for the refined immune metagene (cross-validated P value = 0.01, Table 2). In untreated HER2+ tumors (n = 122, PROGNOSTIC collection), a 10-gene refined metagene was developed (cross-validated P value = 9.5e-5, Table 2). Interestingly, six of these selected genes were in common (CXCL13, PRR1, IRF1, IKZF1, GZMB, and HLA-E). Notably, some of them are associated with cytotoxic T cells. Using these genes, a consensus T cell–related metagene (CTM) was defined. This consensus metagene showed comparable performances to each subtype-specific refined immune metagene (Supplementary Fig. S8A and S8B), and it was also similarly prognostic in ER−/HER2+ (P = 0.024) and ER−/HER2+ (P = 0.0001) subtypes (Supplementary Fig. S8C and S8D). Therefore, it was used for validation in the CHEMO collection.

In untreated ER−/HER2+ tumors (n = 508, PROGNOSTIC collection), starting from the FFPE-adapted proliferation cluster, we defined a refined proliferation metagene including 10 genes, whose low expression was associated with favorable prognosis (cross-validated P = 1.91e-13, Table 2 and Supplementary Section S5). We applied this proliferation metagene to ER−/HER2+ tamoxifen-treated patients (n = 588, TAM collection) to validate its prognostic performance in an independent patient cohort (P = 2.2e-6, Supplementary Fig. S9). Indeed, proliferation was not
associated with any predictive value of endocrine treatment benefit (32) resulting in a similar prognostic value in untreated and endocrine-treated patients (33). In addition, we used the proliferation metagene in the TAM cohort to identify those patients at low risk of relapse even without tamoxifen (low proliferation tertile). This group had a very low risk of relapse (95.8% 5-year DMFS, 98.7% 5-year DMFS in node negative). To develop a context-specific prognostic score to define patients at low and high residual risk of relapse despite tamoxifen administration, we considered only patients with intermediate/high proliferation (n = 395), excluding patients with low proliferation who are less informative because they tend to do well anyway. Starting with the FFPE-adapted, ER-related cluster, we defined a refined ER-related metagene including 10 genes, whose high expression was associated with an excellent prognosis (cross-validated \( P = 1.10 \times 10^{-5} \), Table 2 and Supplementary Section S5).

**Figure 1.** Workflow of the analysis. Schematic representation of data sets used, and the analyses performed to derive MBRPs (Metagene-Based Risk Predictors).

**Table 2.** Refined prognostic metagenes and internally cross-validated performances

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Data collection</th>
<th>No. of points</th>
<th>Metagene</th>
<th>Number of genes</th>
<th>Genes</th>
<th>5-years DMFS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER+HER2-</td>
<td>PROGNOSTIC</td>
<td>179</td>
<td>Immune</td>
<td>25</td>
<td>CXCL13, PLEX, IFNG, SLAMF7, IL2RB, PRF1, IRF1, PTPN2, IKZF1, APOBEC3G, IL2RA, ITGAL, CXCL9, GZMA, GZMB, HLA-E, CCR5, CD8A, SIRPG, CST7, GNY, CECR1, PNOC, LCP1, HLA-DMB</td>
<td>77.79 75.27 57.92 0.0089</td>
</tr>
<tr>
<td>HER2+</td>
<td>PROGNOSTIC</td>
<td>122</td>
<td>Immune</td>
<td>10</td>
<td>HLA-E, GIMAP5, IRF1, CXCL13, SEL1L3, GZMB, IKZF1, PRF1, FGL2, BIN2</td>
<td>86.56 81.34 51.26 9.55E-05</td>
</tr>
<tr>
<td>ER+HER2-</td>
<td>PROGNOSTIC</td>
<td>508</td>
<td>Proliferation</td>
<td>10</td>
<td>NCAPG, BUB1B, PRCI, CCNB2, RAD51AP1, ORC6, FANCI, UBE2C, AURKA, KIF20A</td>
<td>91.47 88.34 63.61 5.25E-14</td>
</tr>
<tr>
<td>ER+HER2 (intermediate/high proliferation)*</td>
<td>TAM</td>
<td>394</td>
<td>ER-related</td>
<td>10</td>
<td>ABT, CA12, MCC2, SCUBE2, LRIG1, PAMG3A, CCDC176, MYB, CACNA1D, GATA3</td>
<td>93.99 85.11 75.24 9.12E-06</td>
</tr>
</tbody>
</table>

**NOTE:** Gene selected in both ER+HER2- and HER2+ cases are in bold.

*Only intermediate and high tertiles by proliferation metagene were included in this analysis.
We also explored the association with outcome of the ER-related metagene in ER/^HER2/^- patients, in which it was not significant (P = 0.97; Supplementary Fig. S10).

Overall, we have defined in the ER/^HER2/^- group, a pure prognostic factor (refined proliferation metagene) and a context-specific predictive factor (refined ER-related metagene). By combining these two metagenes through a median splitting, we defined a low-risk (low proliferation and high ER-related metagenes), an intermediate-risk (high proliferation and high ER-related or low proliferation and low ER-related metagenes), and a high-risk (high proliferation and low ER-related metagenes) group (Supplementary Fig. S11).

**Prognostic and predictive value of MBRP in ER/^HER2/^- and HER2/^HER2/+ chemotherapy-treated patients**

We aimed to evaluate without any further refinement the six-gene CTM in the context of ER/^HER2/^- and HER2/^HER2/+ patients treated with neoadjuvant or adjuvant CT.

The association with pCR was assessed in patients treated with neoadjuvant chemotherapy (anthracycline or anthracycline-taxane-based regimen; n = 260, CHEMO collection). The tumors with high, intermediate, and low expression of this CTM were associated with different pCR rates [33.7%, 35.2%, and 11.6%, respectively; high versus low OR = 3.87 (1.79–8.95), P = 0.0009; intermediate versus low OR = 4.13 (1.93–9.52), P = 0.0004; Fig. 2A]. The association between the CTM expression and pCR was similar in ER/^HER2/^- and HER2/^HER2/+ subtypes, but it reached the statistical significance only in ER/^HER2/-, probably because of the larger sample size. The same association pattern was found by stratifying the analysis according with different data sets (Supplementary Fig. S12).

The context-specific prognostic value of the CTM was assessed in ER/^HER2/^- and HER2/^HER2/+ patients treated with either adjuvant or neoadjuvant CT and having available outcome information (n = 205, CHEMO collection). The CTM was prognostic in the overall group (P = 0.001; Fig. 2B). However, when analyzed per subtype, it resulted statistically significant in ER/^HER2/^- (P = 0.0001) but not in HER2/^HER2/+ (P = 0.5079) group (Fig. 2C and D and Supplementary Table S8). The association was also not significant when ER/^HER2/^- and ER/^HER2/+ subtypes were considered separately (Supplementary Fig. S13). In the ER/^HER2/+ subtype, the 5-year DMFS in the high and low tertile were 85.4% and 43.9%, respectively. The association was similar when separately investigated in the two different data sets (Supplementary Figs. S14 and S15 and Supplementary Table S8). Notably, in the ER/^HER2/+ subtype the CTM was significantly prognostic also in patients with residual disease after neoadjuvant CT (P = 0.0055; Supplementary Fig. S16).

The predictive and prognostic performances of our CTM were compared with two T-cell–related immune signatures [LK (25) and Th (29)]. These immune markers showed a similar pattern of association with pCR and long-term outcome but weaker and sometime not statistically significant (Supplementary Figs. S17 and S18). In a multivariable analysis including the three immune signatures, only our refined metagene remain significant (Supplementary Table S9).

Finally, we correlated our CTM with the above-mentioned T-cell–related metagenes [LK (25) and Th (29)] verifying its strong association with T-cell related signature (Supplementary Fig. S19); we also evaluated the association of the CTM with cell types (Treg, macrophages) and immune signaling (co-inhibitory molecules expressed on T cell or antigen-presenting cells), which are expected to be involved in immune tolerance and escape (34). A significant positive correlation was found for all but Treg signatures, for which only a similar nonsignificant trend was described (Supplementary Fig. S19).

**Prognostic and predictive value of MBRPs in ER/^HER2/^- patients treated with chemotherapy and endocrine treatment**

In ER/^HER2/- patients treated with neoadjuvant chemotherapy (n = 357, CHEMO collection), we evaluated the likelihood of achieving a pCR according to the three risk groups previously defined based on the proliferation and ER-related metagenes. The high-risk group (high-proliferation/low ER-related metagenes) had the highest pCR rate (18.9%) compared with the low-risk group [low proliferation/high ER-related metagenes, 4.4%; OR = 5.01 (1.76–17.99), P = 0.005; Fig. 2C]. This association was driven by one of the two data sets included in the analysis (Supplementary Fig. S20).

The prediction of risk of relapse according to the three risk groups was evaluated in patients treated with neoadjuvant or adjuvant chemotherapy followed by endocrine treatment (n = 350, CHEMO collection; Fig. 2D and Supplementary Table S10). Compared with the low-risk group, the high-risk group showed the poorest prognosis despite chemotherapy administration [74.1% 5-year DMFS, HR = 3.73 (1.63–8.51), P = 0.0018]. This association was robust and consistent across the two data sets included in the analysis (Supplementary Fig. S21 and Supplementary Table S10). Notably, the low-risk group demonstrated an excellent prognosis even in tumors with residual disease after neoadjuvant CT (96.1% 5-year DMFS), although the high-risk group had poor prognosis (69.2% 5-year DMFS; P = 0.01; Supplementary Fig. S22).

**Prediction in the overall HER2^- group**

Because nowadays patients with HER2^- tumors will receive also trastuzumab as standard treatment, we evaluated the overall MBRPs’ performance in the HER2^- group in which chemotherapy and endocrine treatments are the actual standard of care. Risk groups were identified separately in ER/^HER2/- and HER2/^HER2/- subtypes and then combined. In all HER2^- patients, the low-, intermediate-, and high-risk groups had 91%, 83%, and 72% 5-year DMFS, respectively (P = 1.5E-06; Supplementary Fig. S23). In a multivariable analysis, considering only patients treated with adjuvant chemotherapy to allow adjusting for clinical variables (grade, age, and nodal status; n = 371, CHEMO collection), high- and intermediate-risk groups had a significantly higher risk of relapse [HR = 3.53 (1.57–7.92), P = 0.002 and HR = 2.54 (1.19–5.43), P = 0.016, respectively; Table 3]. Similarly, high- and intermediate-risk groups had a significantly higher risk of relapse in ER/^HER2/- and ER/^HER2/+ groups separately (Table 3).

**Comparison between prediction performances of unrefined and refined metagenes**

In this study, instead of simply calculating an average expression value of all the genes belonging to FFPE-adapted clusters, we introduced a cross-validated feature selection step to define refined metagenes (Fig. 1 and Table 2). We assessed whether this step leads to improved prognostic performances using an independent validation cohort of chemotherapy-treated patients. The refined metagenes assessed as continuous variable performed always better (lower P value and higher c indices) than the...
unrefined metagenes, confirming the usefulness of the feature selection step (Table 4).

### Discussion

In early breast cancer, the add-on drug development strategy led to a remarkable improvement of patient outcome over the last two decades (5), but these improvements came at the price of an increasing overtreatment. Such drawback is inherent in the one-fits-all approach, where average instead of individual benefit is the leading goal. In this context, biomarkers able to refine residual risk after standard treatment would be extremely useful. For instance, identifying those patients who do already well with a treatment will exclude them from overtreatment with additional therapies.

---

**A**

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Comparison</th>
<th>OR</th>
<th>CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER⁺ or HER2⁺ (n = 260)</td>
<td>Intermediate vs. low CTM expression</td>
<td>4.13 (1.93–9.52)</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High vs. low CTM expression</td>
<td>3.87 (1.79–8.95)</td>
<td>0.0009</td>
<td></td>
</tr>
<tr>
<td>ER–HER2⁻ (n = 210)</td>
<td>Intermediate vs. low CTM expression</td>
<td>4.24 (1.86–10.36)</td>
<td>0.0008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High vs. low CTM expression</td>
<td>2.72 (1.17–6.79)</td>
<td>0.0245</td>
<td></td>
</tr>
<tr>
<td>HER2⁺ (n = 50)</td>
<td>Intermediate vs. low CTM expression</td>
<td>3.41 (0.61–27.04)</td>
<td>0.1854</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High vs. low CTM expression</td>
<td>4.09 (0.77–31.74)</td>
<td>0.1207</td>
<td></td>
</tr>
</tbody>
</table>

---

**B**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>OR</th>
<th>CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate vs. low risk</td>
<td>1.85 (0.64–6.67)</td>
<td>0.2922</td>
<td></td>
</tr>
<tr>
<td>High vs. low risk</td>
<td>5.01 (1.76–17.99)</td>
<td>0.0053</td>
<td></td>
</tr>
</tbody>
</table>

---

**C**

**D**

Figure 2.

Prognostic and predictive role of MBPs in ER⁺HER2⁻, HER2⁺, and ER⁺HER2⁺ treated breast cancer. A, logistic regression analysis of the consensus T cell–related metagene (CTM) expression and pCR after neoadjuvant chemotherapy. Metagene expression was categorized in low, intermediate, and high by tertiles. B, Kaplan–Meier analysis for the association of the CTM expression by tertiles with 5-year DMFS in chemotherapy-treated patients for the subgroup of ER⁺HER2⁻ and HER2⁺ (left), ER⁺HER2⁻ (middle), and HER2⁺ (right). Survival differences were evaluated by log-rank test. C, logistic regression analysis of groups defined by combining proliferation and ER-related metagenes with pCR after neoadjuvant chemotherapy in ER⁺HER2⁻ tumors (low-risk = low proliferation and high ER-related metagenes; intermediate-risk = high proliferation and high ER-related or low proliferation and low ER-related metagenes; high-risk = high proliferation and low ER-related metagenes). D, Kaplan–Meier analysis for the association of the three risk groups with 5-year DMFS in ER⁺HER2⁻ patients treated with chemoendocrine therapy. Survival differences were evaluated by log-rank test.
Breast Cancer Outcome Prediction after Standard Treatment

Pointing in this direction, in the last decades several markers have been developed to define which ER\(^+\) HER2\(^+\) patients do not need additional chemotherapy to adjuvant endocrine treatment (9, 11, 35). However, this approach has not been applied extensively for evaluating residual risk after standard chemotherapy. In this study, we demonstrated that, using a metagene-based risk prediction approach, it is possible to identify HER2\(^+\) patients at different risk of recurrences despite receiving neoadjuvant or adjuvant chemotherapy (ER\(^-\) HER2\(^-\)) or chemoendocrine therapy (ER\(^+\) HER2\(^+\)). High-risk patients despite systemic treatments are the ideal candidates for clinical trials testing new combinations of available drugs or investigational compounds, as they represent an unmet clinical need. Clinical trials designed to enroll these patients would have an increased chance of demonstrating a clinical meaningful benefit reducing false-negative results (36), at the same time requiring a smaller sample size, thus reducing costs (37).

In the ER\(^+\) HER2\(^-\) subtype, patients with high proliferation and low ER-related metagene expression were at the highest risk of distant metastasis (43.9% 5-year DMFS and lower pCR rate (11.6%). A similar association has been described for tumor-infiltrating lymphocytes (TIL) in triple negative patients treated with adjuvant chemotherapy (41, 42). However, in these studies the lymphocyte-predominant breast cancer subgroup, which has the best prognosis, correspond to only 10.5% (41) and 4.4% (42) of the study population, whereas our data suggest that at least one third of TN tumors had such good prognosis. Moreover, our immune metagene identified one third of patients with a dismal prognosis, resulting from both a higher baseline risk of recurrences and lack of benefit from standard therapy. These patients deserve a priority enrollment in trials testing investigational compounds. Our data demonstrated a very heterogeneous prognosis by immune marker value in ER\(^+\) HER2\(^+\) patients treated with neoadjuvant chemotherapy but not achieving a pCR. These data could represent the result of both a different baseline prognosis and long-term benefit from chemotherapy, which did not result in a pCR.

A recently presented "proof of concept" study demonstrating the clinical activity of pembrolizumab, an immune-checkpoint inhibitor, in patients with advanced TN breast cancer (43). In line with these results, our findings reinforce the rational for testing a combination of immunomodulating agents and chemotherapy in TN breast cancer, also considering the positive association between our immune metagene and co-inhibitory immune molecules, which are likely to be induced as negative regulatory feedbacks to dampen an otherwise actively engaged immune system. In the HER2\(^+\) group, we confirmed that higher immune metagene score was associated with higher likelihood of achieving a pCR, as previously reported (20, 24). However, the association with the risk of recurrences was not significant. This could be a false-negative result due to the smaller sample size but also a true clinical observation. The endocrine treatment received by ER\(^+\) HER2\(^-\) patients could have generated unexpected interactions. Moreover, similarly to our finding, in HER2\(^+\) patients TILs were not associated with a different risk of recurrence in two clinical trials considering patients treated with chemotherapy only (41, 44). This lack of association with outcome warrants further confirmation and investigation.

The use of metagenes linked to defined biologic processes as prognostic/predictive markers instead of mixture of genes related with several functions has the advantage of easily interpreting the meaning of the associations with the outcomes. The six genes of our immune metagene (CXCCL13, PRF1, IRF1, IKZF1, GZMB, and HLA-E) are mainly associated with key adaptive immune cells (21, 24, 25, 28, 29), we developed a robust six-gene-based immune metagene significantly prognostic in both groups. This metagene was applied in patients treated with neoadjuvant or adjuvant chemotherapy. In the ER\(^+\) HER2\(^-\) subtype, higher tertile of expression identified patients at lower risk of recurrences (85% 5-year DMFS) and higher rate of achieving a pCR, although lower tertile of expression was associated with very high risk of distant metastasis (43.9% 5-year DMFS and lower pCR rate (11.6%).

### Table 4. Comparison of refined and unrefined metagenes in the CHEMO data set

<table>
<thead>
<tr>
<th>Metagene</th>
<th>HR (95% CI)</th>
<th>P</th>
<th>c Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proliferation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refined</td>
<td>0.54 (0.33-0.90)</td>
<td>0.092</td>
<td>0.61</td>
</tr>
<tr>
<td>Unrefined</td>
<td>0.54 (0.53-0.90)</td>
<td>0.0206</td>
<td>0.63</td>
</tr>
<tr>
<td>ER-related</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refined</td>
<td>0.46 (0.31-0.70)</td>
<td>0.0003</td>
<td>0.67</td>
</tr>
<tr>
<td>Unrefined</td>
<td>0.54 (0.33-0.90)</td>
<td>0.092</td>
<td>0.61</td>
</tr>
<tr>
<td>CTM (ER(^+) HER2(^+), n = 205)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refined</td>
<td>0.57 (0.41-0.77)</td>
<td>0.0004</td>
<td>0.65</td>
</tr>
<tr>
<td>Unrefined</td>
<td>0.57 (0.39-0.83)</td>
<td>0.0036</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Abbreviation: CTM, consensus T cell-related metagene.
functions. CXCL13-producing CD4+ follicular helper T cells were associated with tertiary lymphoid structure that can contribute to generate effective long-term antitumor immunity (29). Perforin (PRF1) and granzyme B (GZMB) are two key cytotoxic effectors, which are upregulated upon cytotoxic T-cell activation (45). IRF1 is an activator of type I IFNs and IFN-inducible genes (46). Overall, our data confirm that similar immune functions are involved in tumor spread control (47–49) and in the cooperation with chemotherapy activity. This link is not surprising and has also been described by others (29). Indeed, some chemotherapies (i.e., anthracyclines and oxaliplatin) are able to induce an immunogenic cell death that can lead to an optimal activation of adaptive immunity (50–52). However, our data suggest that such immune system engagement is more likely to be effective if the baseline immune microenvironment is already at least partially activated.

The strategy adopted in this study to develop metagene predictors includes some elements of interest and novelty. We started from clusters of correlated genes with known biologic and prognostic relevance (proliferation, ER-related genes, and immune function), and then we optimized such metagenes by removing noninformative probes. This feature selection step (metagene refinement) resulted in a significant increase of the prediction value in the CHEMO validation cohort compared with unrefined metagenes. At the same time, reducing the number of genes needed could facilitate the transfer of derived signatures to non–microarray-based platforms, characterized by a lower throughput, but higher accuracy. This overall strategy represents a model that can be successfully applied in other tumor contexts.

As potential limitations of our study, chemotherapy administered in our CHEMO collection was not homogeneous and the short available follow-up does not allow assessing for late relapse, which could be relevant in the ER/HER2+ group. A validation in homogeneous cohorts of patients enrolled in clinical trials represent the ideal subsequent step (13). To improve its feasibility, we developed our MBRPs in a way to be suitable for application in FFPE-derived GEPs (i) by applying a processing method specifically optimized for FFPE data (19); (ii) by removing poor performing probes in FFPE-derived GEPs; and (iii) calculating the metagene value avoiding assigning weights for each gene.

Disclosure of Potential Conflicts of Interest
G. Bianchini is a consultant/advisory board member for Genomic Health. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions
Conception and design: M. Callari, V. Cappelletti, M.G. Daidone, L. Gianni, G. Bianchini
Development of methodology: M. Callari, V. Musella, A. Lembo, P. Provero, G. Bianchini
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): V. Musella, F. Petel, M.G. Daidone, L. Gianni, G. Bianchini
Writing, review, and/or revision of the manuscript: M. Callari, V. Cappelletti, V. Musella, F. Petel, T. Karn, T. Iwamoto, L. Gianni, G. Bianchini
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Callari, G. Bianchini
Study supervision: M.G. Daidone, L. Gianni, G. Bianchini
Other (data integration, clinical annotation curation, and data analysis of some the data subsets): F. Petel

Grant Support
This study was supported by the Italian Association for Cancer Research (AIRC) grants (IG4915 to L. Gianni, IG 10611 to M.G. Daidone, and MIAF 13428 to G. Bianchini) and Fondazione Michelangelo for the Advancement of the Study and Treatment of Cancer grant to G. Bianchini.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertised in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 31, 2015; revised August 31, 2015; accepted September 9, 2015; published OnlineFirst September 30, 2015.

References
Breast Cancer Outcome Prediction after Standard Treatment


