Immune-related gene signatures predict the outcome of neoadjuvant chemotherapy

Gautier Stoll1,2, David Enot3, Bernhard Mlecnik1,5,6, Jérôme Galon1,5,6, Laurence Zitvogel7,8,9, and Guido Kroemer1,2,3,9,10,*

1Université Paris Descartes/Paris V, Sorbonne Paris Cité; Paris, France; 2Equipe 11 labellisée Ligue Nationale contre le Cancer: Cordeliers Research Center; INSERM U1138; Paris, France; 3Metabolomics and Cell Biology Platforms; Gustave Roussy Cancer Campus; Villejuif, France; 4Laboratory of Integrative Cancer Immunology; INSERM U1138; Paris, France; 5Cordeliers Research Center; Université Pierre et Marie Curie Paris 6; Paris, France; 6INSERM U1015; Villejuif, France; 7Faculté de Médecine; Université Paris Sud; Le Kremlin Bicêtre, France; 8Center of Clinical Investigations in Biotherapies of Cancer (CICBT) 507; Villejuif, France; 9INSERM, U848; Villejuif, France; 10Pôle de Biologie; Hôpital Européen Georges Pompidou, AP-HP; Paris, France

Keywords: autophagy, breast cancer, colorectal cancer, endoplasmic stress, immunogenic cell death, tumor-infiltrating lymphocytes

Abbreviations: ALOX5AP, arachidonate 5-lipoxygenase-activating protein; ATF4, activating transcription factor 4; ATF6, activating transcription factor 6; CXCL13, chemokine (C-X-C motif) ligand 13; ER, endoplasmic reticulum; FAC, fluorouracil plus doxorubicin plus cyclophosphamide; FEC, fluorouracil plus epirubicin plus cyclophosphamide; HIF1α, hypoxia-inducible factor 1α; LAMP3, lysosomal-associated membrane protein; pCR, pathological complete response; TCGA, The Cancer Genome Atlas; T11, Type 1 T helper; Treg, regulatory T cell; XBP1, x-box binding protein 1

Introduction

Accumulating evidence indicates that pre-therapy anticancer immune responses significantly influence disease progression in breast carcinoma patients.1-3 This corollary has also been perceived in the prognosis of other malignancies,3-5 including colorectal carcinomas.6-7 The infiltration of CD8+ lymphocytes in primary breast cancer lesions is positively associated with both overall and relapse-free survival, especially if the tumors lack immunosuppressive CD68+ macrophages.8-10 Similarly, the absence of regulatory T cells (Tregs) in diagnostic breast cancer biopsies has a positive impact on the long-term fate of patients after adjuvant chemotherapy.11,12 A high frequency of infiltrating lymphocytes detected in diagnostic biopsies also constitutes an independent predictive biomarker for the induction of pathological complete responses (pCRs) in locally advanced breast cancers.13-16 In the neoadjuvant setting, chemotherapeutic responses have been associated with the de novo induction of local immune responses. Thus, an amelioration in the intratumoral T-cell ratio of CD8+ T lymphocytes to FOXP3+ Tregs detected after the first cycle of anthracycline-based chemotherapy is sufficient to predict the occurrence of pCR in response to the full chemotherapeutic regimen of at least 6 cycles.17,18 While most the studies have used immunohistochemical methods to investigate the composition of the immune infiltrate in breast cancer, a few have identified immune-relevant transcripts and gene sets or “metagenes” (composed by several genes expressed in Type 1 T helper (T11) cells and interferon responses) that predict pCRs in each of the breast cancer subtypes investigated in this respect.19-22

One of the mechanisms that could potentially underlie anticancer immune responses post-chemotherapy is that of so-called immunogenic cell death. Anthracyclines, DNA-damaging agents often used in the treatment of breast cancer, stimulate 2 types...
of premortem stress responses possibly linked to immunogenic cell death. Namely, these include the endoplasmic reticulum (ER) stress response (also called the “unfolded stress response”) and autophagy consisting of the sequestration and lysosomal degradation of cytoplasmic components.\textsuperscript{18,23–25} Autophagy is required for the cells to release ATP during the apoptotic blebbing phase, and ATP acts as a potent chemoattractant for myeloid cells, including antigen-presenting dendritic cell precursors.\textsuperscript{26–28} ER stress is required for the exposure of calreticulin thought to act as an eat-me signal for improving the transfer of tumor antigens to antigen-presenting cells.\textsuperscript{29} Tumors that are deficient in ER stress or in autophagy fail to elicit immune responses in mice post-chemotherapy and the malignancy becomes incurable.\textsuperscript{24,25} Similarly, therapeutic agents that fail to induce premortem ER stress and autophagy have limited therapeutic potential.\textsuperscript{30}

Based on these premises, we set out to investigate breast cancer microarray data for signs of ER stress, autophagy, or lysosomal transcriptional perturbations, hoping to establish correlations between parameters of cellular stress and local immune responses. From this analysis, we established a series of gene signatures or “metagenes” reflecting correlated gene expression patterns with special reference to the immune response, the ER stress response, autophagy, or lysosomes. We then interrogated correlated the clinical relevance of these metagenes among each other, in several independent cohorts of breast cancers subjected to neo-adjuvant chemotherapy, and evaluated their impact on patient survival post-chemotherapy. Our results reveal the clinical relevance of several immunity-related metagenes, despite the apparently limited impact on local cellular stress responses.

### Results and Discussion

**Immune- and stress-related metagenes with a high degree of reproducibility across data sets**

Among genes known to relate to the immune system, ER-stress, autophagy, or lysosomes (refer to Materials and Methods for the procedure used in the construction of the gene lists), we determined aggregate patterns of coordinately expressed genes (i.e., metagenes) by hierarchical clustering and principal component analyses of the initial data set of 522 breast cancer microarrays accessible in The Cancer Genome Atlas (TCGA). Thresholds were arbitrarily set so that this operation would yield 18–20 metagenes in each of the 4 categories (Tables S1–4, corresponding to genes associated with immunity, ER-stress, autophagy, or the...
lysosome, respectively). These metagenes were then re-examined within a series of additional microarray data sets profiling cancer types linked to antitumor immunity. Among these validation data sets, 3 that were obtained were composed of breast cancer patient samples (161 samples collected by Bonnefoi et al.,31 184 samples analyzed by Hatzis et al., 32 and 178 advanced, late-stage carcinoma samples in the collection by Tabchy et al. 33). Three other additional validation data sets involved colorectal carcinomas or head and neck cancers. This procedure established a ranking of metagenes based on their reproducibility across multiple distinct data sets and immunity-associated cancer types (Fig. 1).

Seven of the immune-related metagenes were judged to be reproducible (reproducibility $P$ value < 0.05 for all data sets except 1). The salient genetic components of each reproducible metagene are shown in Figure 2. The first (and hence most reproducible) immune metagene “CXCL13,” named according to the highest weighted gene in the signature chemokine (C-X-C motif) ligand 13, mainly reflects the presence of effector T cells within the tumor bed. The second and third metagenes (“CTSS” and “S100A4,” respectively) constitute aggregates of genes typically expressed in myeloid cells. And the subsequent ones (CCL8, CXCL9, CXCL1, CCL19) reflect chemokine-related gene signatures. More generally, most of the genes appearing in the list of salient components contributing to immune-relevant metagenes are related to tumor infiltrates, such as CD8B, 34 CD80, 35 CD86, 35 CTLA4, 36 PRF1, 37 IL-2, 38 IL-10, 39-41 CCL5, 42 CCL2, 43 CCL19, 44 and CXCL9. 45 Four metagenes among each of the 3 stress-related categories (ER-stress, autophagy, and lysosomes) yielded reproducible patterns across all data sets (Fig. 2).

**Correlations among immune- and stress-related metagenes**

In the subsequent step of this analysis, Pearson’s correlations were calculated between the expression levels of all metagenes, in each of the 4 breast-cancer cohorts studied here (Fig. 3). The immune system-related metagenes yielded preponderantly positive correlations, especially among the most reproducible metagenes (plotted in the upper left corner of each graph in Fig. 3A) and demarcated by the solid yellow lines. No significant negative correlations could be identified among the immune-related genes, indicating that the overall gene expression signature in this category reflects inflammatory and immune reactions as an en bloc (unified) process. These results suggest a failure to identify antagonistic relationships between cells and associated genes reputed to play a positive role in immunosurveillance (such as effector T cells) and immunosuppressive elements (such as myeloid-derived...
suppressor cells and regulatory T cells). A similar, though less obvious pattern of mostly positive correlations was obtained for lysosome-related metagenes (Fig. 3D). In sharp contrast, no clear correlations could be established among ER stress-related metagenes (Fig. 3B) and autophagy-related genes (Fig. 3C). Thus, the principal ER-stress metagene (HSPH1) was not correlated with any of the other reproducible metagenes within this category (SLC1A5, ASS1, FKBP5; Fig. 3B). Moreover, the 2 principal autophagy-related metagenes (BCL2A1, CTSD) were neither correlated with each other nor with the 2 other reproducible metagenes (GADD45B, BNIP3) contained within this category (Fig. 3C). Collectively, immune-related and lysosome-related metagenes exhibited more positive internal correlations (higher r-values) than ER stress- and autophagy-related metagenes among the subsets of metagenes that were considered to be reproducible across distinct cancer types (yellow box plots in Fig. 3E). Finally, the reproducibility of the correlations (as evidenced by a lower reproducibility P value) across the 4 distinct cohorts of breast cancer patients was also higher for immune-related and lysosome-related metagenes as compared with ER-stress and autophagy-related metagenes (Fig. 3F). Altogether, these data suggest that immune-related and lysosome-relevant metagenes reflect rather monolithic processes, while ER-stress and autophagy-related metagenes comprise apparently heterogeneous alterations in gene expression. The relatively low level of correlations among the stress-related metagenes relative to those found among immune-relevant metagenes might also reflect a greater reliance of cellular stress responses on post-transcriptional (rather than transcriptional) changes.

Correlations between immune-related and stress-related metagenes

The principal objective of this study was to explore correlations between immune-related gene expression patterns on one hand and cellular stress responses on the other. Cross correlations between immune- and ER stress-related metagenes were generally weak and poorly reproducible among the 4 breast cancer cohorts with the notable exception of the ER stress-related metagene ALOX5AP (Fig. 4A), which regroups 2 pro-inflammatory genes, namely arachidonate 5-lipoxygenase-activating protein (ALOX5AP or FLAP)46 itself and another component of the ALOX5AP metagene G protein-coupled receptor 84 (GPR84),

Figure 3. Pearson's correlations between metagenes within each of the 4 functional categories as determined in 4 distinct breast cancer cohorts. A–D. Reproducibility (P value) calculated based on the variance between metagene data sets within the indicated cohort. E–F. Box plots of correlation matrix (E) elements (or r-values) calculated relative to the T cga cohort and reproducibility P values (F), either for the complete set of metagenes (in white) or focusing on the subset of reproducible metagenes (in yellow) delimited by yellow lines in A–D (as described in Fig. 1 legend). Zero (black) indicates that the pair are uncorrelated, a positive value (red) indicating a correlation and a negative value (green) indicating an inverse relationship.
a receptor expressed by immune cells. However, the ALOX5AP metagene was not ranked as a reproducible gene across distinct data sets (Fig. 1). The correlation between immune- and autophagy-relevant metagenes was weak as well, with the notable exception of the highly reproducible metagene BCL2A1 (Fig. 4B). In contrast, we observed a relatively strong correlation between several immune- and lysosome-related metagenes, especially among the reproducible subsets of metagenes (upper left corner in Fig. 4C). Thus, strong correlations were found between most immune-relevant metagenes and the ACP5 and HPSE metagenes, which both primarily reflect the expression level of lysosomal hydrolases, and the LAMP3 metagene. This may be partly explained by the fact that some lysosomal hydrolases and the lysosomal-associated membrane protein 3 (LAMP3, also called DC-LAMP or CD208) isoform of the LAMP family are strongly expressed by innate immune effectors. Altogether, the correlations between metagenes falling into distinct functional categories (Fig. 4D and E) tended to be as low as those observed among each of the stress-related categories outside of the immune-related category (compare with Fig. 3E and F).
We next turned to another strategy for the identification of possible correlations between immune-related metagenes and cellular stress pathways. Target genes of transcription factors involved in the ER-stress response, including activating transcription factor 4 and 6 (ATF4 and ATF6, respectively), x-box binding protein 1 (XBP1), and a regulator of lysosomal/autophagic biogenesis transcription factor EB (TFEB) were listed together with hypoxia-inducible factor 1α subunit (HIF1α) and the pro-inflammatory nuclear factor-kB (NF-kB) as controls. These lists were then used to interrogate the TCGA breast cancer cohort and 5 metagenes corresponding to each of these transcription factors were identified (Table S5). Again, as previously performed for immune- and stress-related metagenes, these transcription-factor associated metagenes were ranked for reproducibility in 6 distinct cancer cohorts (Fig. 5A–D). Subsequently, as shown in Figure 5D (middle panels), these novel metagenes were correlated with the 19 immune-related metagenes initially identified (Fig. 1; Table S1). Finally, these correlations were evaluated for reproducibility within the 4 breast cancer cohorts (lower panels in Fig. 5A–D). We found that among the ATF4 target metagenes (Fig. 5A), 2 (PSAT1 and PTX3) positively correlated with the immune-related metagenes.
with a high degree of reproducibility, whereas 1 metagene (STC2) reproducibly exhibited a negative correlation with the immune-related metagenes. Among the ATF6 target genes (Fig. 5B), only the IL6 metagene, dominated by the gene coding for the pro-inflammatory cytokine interleukin-6, exhibited a highly reproducible positive correlation with immune-related metagenes, whereas 1 metagene (XBP1) negatively correlated with some immune metagenes, albeit with a low level reproducibility. Among XBP1 target genes (Fig. 5C), again an IL6-centered metagene (dominated by the genes coding for IL-6, cyclin A1,
and RANK ligand) exhibited a positive correlation with immune genes across all 4 breast cancer cohorts. Another XB1 target metagene (Fig. 5C), PO2UAF1, comprising several immune-relevant genes, including vascular cell adhesion molecule-1 (VCAM1) activation-induced cytidine deaminase (AICDA) the cell death receptor FAS/APO-1/CD95, the intercellular adhesion molecule-1 (ICAM1), and FK506-binding protein-11 (FKB11), among others, correlated positively with immune-related metagenes in 3 out of 4 breast cancer cohorts. In contrast, an estrogen receptor-1 (ESR1)-dominated XB1 target metagene negatively correlated with immune-metagenes, confirming the previous observation that estrogen receptor is linked to a reduced infiltration of CD8^+ T cells into tumors.\textsuperscript{30} No significant correlations were detectable between TFEB target metagenes and immune-related metagenes (Fig. 5D). Among our controls, we also found 1 HIF1α target metagene (MMP1) and 2 NFKB1 target metagenes (IL-6 and PTX3) that positively correlated with immune metagenes (Fig. S1), likely reflecting the presence of target genes of these factors that are preponderantly expressed by immune cells (Table S5).

Altogether, it appears that some metagenes of transcription factor targets related to ER stress (ATF4, ATF6, XB1) correlated with some degree with immune-related metagenes (which is not the case for TFEB target metagenes), although these correlations are comparable to those observed with HIF1α and NFKB1 target metagenes (Fig. 5F). Moreover, the reproducibility of ER or lysosome/autophagy-related metagenes is not superior to that observed for HIF1α and NFKB1 target metagenes (Fig. 5G).

**Prognostic values of immune- and stress-related metagenes across distinct data sets**

The ultimate goal of such cancer microarray analyses is their clinical application to prospectively identify biomarkers that can predict the natural progression of the disease or clinical responses to chemotherapy. Therefore, we evaluated the impact of each metagene described in this report on the clinical response of the 3 cohorts for which such data are available. All 3 cohorts received neo-adjuvant chemotherapy, and pathological complete responses (pCR) were evaluated following surgical resection of the tumors. The Bonnefoi collection of breast cancers were locally invasive estrogen receptor-negative carcinomas treated with anthracycline-based neoadjuvant chemotherapy (either epirubicin + fluorouracil + cyclophosphamide or epirubicin plus docetaxel).\textsuperscript{31} The Hatzis cohort comprised ERBB2-negative breast cancers that received neoadjuvant chemotherapy based on anthracycline and taxane combinations (either paclitaxel or docetaxel (followed by endocrine therapy if estrogen receptor-positive).\textsuperscript{32} The Tabchy cohort received neoadjuvant chemotherapy involving fluorouracil plus epirubicin plus cyclophosphamide (FEC) or fluorouracil plus doxorubicin plus cyclophosphamide (FAC), alone or combined with taxanes.\textsuperscript{33} Among the metagenes that have been defined in this report, rather few correlated with the chemotherapeutic response. Included in this rare subset was the immune-related CXCL13 metagene that exhibited a strong positive correlation with pCR (combined \(P \text{ value of 0.005})\), and to some degree to CCL8 (combined \(P \text{ value of 0.02})\) and CXCL9 (combined \(P \text{ value of 0.05})\), both of which also exhibited positive correlations with pCR although with lower reproducibility than that of the CXCL13 metagene (Fig. 6A). Among the ER-stress-related, autophagy-related, and lysosome-related metagenes (Fig. 6B–D) no reproducible positive correlations could be identified with the notable exception of the LAMP3 metagene, (combined \(P \text{ value of 0.0002})\), which most likely reflects the presence of LAMP3-expressing dendritic cells in the tumor bed. Other metagenes could only acquire significant combined \(P \text{ values if their impact on patient survival was measurable in all patient cohorts. This applies to several ER stress-related metagenes (HSPA1A, CLGN, PCK2, NFE2L1) but to only 1 autophagy- (THBS1) and 3 lysosome- (ASAH1, ARSA, GLA) related metagenes. Moreover, no significant correlations were found between the stress-related transcription factor metagenes and patient survival in any of the individual cohorts. Furthermore, it was only after calculation of combined \(P \text{ values for all 4 cohorts that some transcription factor-related metagenes (STC2, XB1, SERPINA1, ESR1, CDH1, GLA) appeared to correlate with the therapeutic response (Fig. 6E–I; Fig. S1). In conclusion, the “best” metagene predicting pCR after neoadjuvant breast cancer chemotherapy is the CXCL13 metagene that was also the most reproducible metagene across different breast cancer cohorts and immune-implicated cancer types.

**Concluding Remarks**

The present study confirms the idea that the expression level of select immune response-related mRNA species may constitute a robust biomarker with the potential to predict chemotherapeutic responses in breast cancer.\textsuperscript{21} This is illustrated in particular by 1 metagene (CXCL13) composed by several genes encoding various chemokines (in decreasing order of importance: CXCL13, CXCL11, XCL1, CL5, XCL1, CCL17, CCL2, CCL23, CXCL10), interferon-γ (IFNγ), components of the T cell receptor/CD3 complex (CD3G, CD3D, CD3E), a cytotoxic T cell marker (CD8B), a general T cell marker (CD7), T cell activation markers (CD25A, CTLA4), the co-activation receptor CD28, several components of cytotoxic granules such as perforin-1 (PRF1) and the granymes B (GZMB) and M (GZMM), as well as the dendritic cell activation marker CD83 (Table S1). Hence, this metagene likely relates to the presence of IFNγ-producing cytotoxic T lymphocytes, a cell type that has previously been reported to play a major role in immuno-surveillance.\textsuperscript{51,52} This particular metagene was the most reproducible one among all immune-related metagenes identified in this study. In addition, constituents of this metagene have been identified as major players in immune cell tumor infiltration in colorectal cancer, such as has been reported for CXCL13,\textsuperscript{33} CXCL10,\textsuperscript{54} and CDB8.\textsuperscript{55} Contrasting with the high degree of clinical relevance and reproducibility of immune-related metagenes, we found that metagenes related to cellular stress responses exhibited a relatively poor degree of reproducibility and failed to predict clinical responses. There are a number of plausible explanations that may underlie this observation. First, stress responses may occur in a transient and spatially limited fashion, despite potentially durable effects on cognate immune responses. Second, although
stress responses do impact the transcriptome due to the activation of stress-elicited transcription factors, post-transcriptional adaptations are known to play a major role in the response to stress, as exemplified by eIF2α phosphorylation and dephosphorylation reactions regulation translation in the unfolded protein response.\textsuperscript{56} Similar post-translational regulatory mechanisms include the acute activation of autophagy by cytoplasmic protein (de)-phosphorylation and (de)-acetylation reactions.\textsuperscript{57} Hence, the transient nature of stress responses, as well as their limited impact on transcription, may explain the absence of robust interactions between immune-related and stress-related metagenes. Indeed, the few interactions that we detected can be explained by the fact that several stress-related metagenes (such as LAMP3) included genes that are preferentially expressed by immune cells.

If microarray analyses constitute a suboptimal tool for exploring stress responses within tissues, it will be important to explore alternative technologies for measuring such stress responses. Thus, a proteomic approach may offer distinct technical advantages, such as immunohistochemical-based monitoring of eIF2α phosphorylation as a parameter of the ER stress response, an event found to correlate with T cell response within breast cancer tissues post-chemotherapy.\textsuperscript{58} Similarly, autophagy can be detected immunohistochemically by measuring the aggregation of LC3B cytosolic protein in autophagosomes.\textsuperscript{59} It will be important to apply such protein-based methods to further establish the chronological, spatial, and clinical correlations between stress responses and anticancer immunosurveillance.

Materials and Methods

Data sets

We used 4 breast cancer data sets for this work. The first one, “TCGA” (downloaded from TCGA website cancergenome.nih.gov), was used as a basis for the construction of meta-genomes. The 3 others were chosen because the information regarding patient’s response to treatment was accessible and included: “Bonnefoi”\textsuperscript{31} GSE6861, “Tabchy”\textsuperscript{31} GSE20271, and “Hatzis”\textsuperscript{32} GSE25065. We add 2 data sets of colorectal cancer (“Smith”\textsuperscript{59,60} GSE17536 and “Grone”\textsuperscript{61} GSE18088) and 1 of head and neck cancer (“Rickman”\textsuperscript{62} E-TABM-302) to study the applicability of metagenes to other cancer types (Fig. 1).

For each data set, normalized data were downloaded (except for the Rickman data set for which RMA normalization was applied). For Affymetrix technology, we considered only the best probesets according to Jetset.\textsuperscript{63}

The entire compliment of patients within each data set was generally considered for metagene construction and metagene reproducibility, correlation matrices and correlation matrix reproducibility calculations. In the analysis of metagene correlation with therapeutic response, distinct subgroups of patients were defined according to their inclusion in a particular therapeutic regimen (Fig. 6).

All computations for data analysis (e.g., normalization, statistical tests and figure drawing) were done within R.\textsuperscript{64}

Genes included in distinct functional groups

For the group of immune-related genes, a series of genes specifically expressed in immune cells was complemented by the inclusion of a complete set of genes encoding interferons, cytokines, chemokines and HLA genes (135 genes in total, as indicated in Table S1).

Groups of stress-related genes were extracted from previous publications on ER stress (81 genes, Table S2),\textsuperscript{65} autophagy (86 genes, Table S3),\textsuperscript{66} and lysosomes (88 genes, Table S4),\textsuperscript{67} following removal of (most of) the genes overlapping with the immune-related group. Genes modulated by the transcription factors ATF4 (177 genes), ATF6 (29 genes), XBP1 (173 genes), TFEB (26 genes), HIF1A (299 genes), and NFKB1 (177 genes) were obtained from the IPA database (Ingenuity® Systems, www.ingenuity.com, Table S5), and again genes listed in the immune-related group were removed during this process.

Metagene construction

A metagene was defined as an aggregate of genes whose expression correlates in a series of samples.\textsuperscript{68} Sets of genes comprising metagenes were obtained by hierarchical clustering of the TCGA data set, fixing the number of metagenes (20 for stress related groups, 5 for transcription factor targets) using the R-functions hclust and cutree to generate the correlation matrix. Gene coefficients inside a meta-gene (Tables S1–4 and Supplementary Material) were obtained by performing principal component analysis on covariance of expression within the set of genes that constituted each metagene using princomp in R. For each metagene, the first principal component was extracted and used as the gene coefficient.

Reproducibility of metagenes

The reproducibility of a metagene (Fig. 1) between several data sets was obtained by comparing TCGA with the other breast cancer data sets. More precisely, to calculate reproducibility for a given metagene, in a given data set, a multistep process was chosen and applied as follows:

1. Construct the set of coefficients within the given data set as the first principal component by applying the same procedure as for the initial TCGA-metagene construction, except that the data set to be used does not originate from TCGA.

2. Measure the angle between the coefficient set constructed in 1) with the one the TCGA-defined metagene. Extract the angle from the scalar product between these 2 sets considered as 2 vectors in the vector space of gene components. If the expression of a given gene was not measured in a data set, remove that component when measuring the angle (reducing the dimension of the vector space). If no components remain, the metagene is removed (this occurred in the lysosome-relevant group of genes).

3. Compute the significance of the angle by comparing the computed value with the null hypothesis distribution obtained by randomly selecting (n = 2000) gene sets within the given group (Immune-related genes, ER stress genes, etc.) of identical length compared with the given metagene and computing the angle as described in steps 1 and 2.

Correlation matrices

The correlation matrices of Figures 3–5 represent the Pearson’s correlation coefficient of metagene expression with zero
Reproducibility of correlation matrices

Because there is no consensus for measuring correlation reproducibility in statistics, we developed our own method by applying a bootstrapping procedure on the variance of each correlation matrix elements. More precisely, for each matrix element, we computed the variance among the different data sets. We associated a P value by constructing a null hypothesis distribution (2000 events) by shuffling metagenes independently between data sets.

Correlation with treatment responses

Correlations with therapeutic responses were calculated for the 3 breast cancer cohorts for which treatment responses were publicly available. In order to accomplish this, expression of each metagene in each sample was first calculated using gene coefficients. Next, for each metagene per data set, unpaired Student’s t tests were performed (one-tailed, for both over- and under-expression alternative hypotheses), comparing the gene coefficient of tumors with or without cPR. For each metagene, the combined P values were then calculated by means of Fisher’s method for combining probability tests. The log(10) of the smallest P value (from the 2 alternative hypotheses) was chosen with positive and negative coefficients. Next, for each metagene per data set, unpaired Student’s t tests were performed (one-tailed, for both over- and under-expression). The log(10) of the smallest P value (from the 2 alternative hypotheses) was chosen with positive and negative coefficients. Then, for each metagene per data set, the combined P values were then calculated by means of Fisher’s method for combining probability tests. The log(10) of the smallest P value (from the 2 alternative hypotheses) was chosen with positive and negative coefficients.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
G.K. is supported by the Ligue contre le Cancer (équipe labellisée); Agence National de la Recherche (ANR); Association pour la recherche sur le cancer (ARC); Cancéropôle Île-de-France; Institut National du Cancer (INCa); Fondation Bettencourt-Schueller; Fondation de France; Fondation pour la Recherche Médicale (FRM); the European Commission (ArtForce); the European Research Council (ERC); the LabEx Immuno-Oncology; the SIRIC Stratified Oncology Cell DNA Repair and Tumor Immune Elimination (SOCRATE); the SIRIC Cancer Research and Personalized Medicine (CARPEM); and the Paris Alliance of Cancer Research Institutes (PACRI). We thank also TCGA Research Network for generating TCGA data sets.

Supplemental Materials
Supplemental materials may be found here:
http://www.landesbioscience.com/journals/oncoimmunology/article/27884/
60. Freeman TJ, Smith JJ, Chen X, Washington MK, Roland JT, Means AL, Eschrich SA, Yeatman TJ, Deane NG, Beauchamp RD. Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of β-catenin. Gastroenterology 2012; 142:562, e2; PMID:22135830; http://dx.doi.org/10.1053/j.gastro.2011.11.026

61. Gröne J, Lenze D, Jurinovic V, Hummel M, Seidel H, Leder G, Beckmann G, Sommer A, Grützmann R, Pilarsky C, et al. Molecular profiles and clinical outcome of stage UICC II colon cancer patients. Int J Colorectal Dis 2011; 26:847-58; PMID:21465190; http://dx.doi.org/10.1007/s00384-011-1176-x

62. Rickman DS, Millon R, De Reynies A, Thomas E, Wasylyk C, Muller D, Abecassis J, Wasylyk B. Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays. Oncogene 2008; 27:6607-22; PMID:18679425; http://dx.doi.org/10.1038/ onc.2008.251

63. Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene. BMC Bioinformatics 2011; 12:474; PMID:22172014; http://dx.doi.org/10.1186/1471-2105-12-474

64. R Core Team. R: A language and environment for statistical computing. (2013). at <http://www.R-project.org/>

65. Dombroski BA, Nayak RR, Ewens KG, Ankener W, Cheung VG, Spielman RS. Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. Am J Hum Genet 2010; 86:719-29; PMID:20398888; http://dx.doi.org/10.1016/j.ajhg.2010.03.017

66. Pietrocola F, Izzo V, Niso-Santano M, Vacchelli E, Galluzzi L, Maiuri MC, Kroemer G. Regulation of autophagy by stress-responsive transcription factors. Semin Cancer Biol 2013; 23:310-22; PMID:23726895; http://dx.doi.org/10.1016/j.semcancer.2013.05.008

67. Palmieri M, Impeny S, Kang H, di Ronza A, Pelz C, Sardiello M, Ballabio A. Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways. Hum Mol Genet 2011; 20:3852-66; PMID:21752829; http://dx.doi.org/10.1093/hmg/ddr306

68. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bld A, Iversen ES, Liao M, Chen CM, et al. Gene expression predictors of breast cancer outcomes. Lancet 2003; 361:1590-6; PMID:12747878; http://dx.doi.org/10.1016/S0140-6736(03)13308-9