Impact of chemotactic factors and receptors on the cancer immune infiltrate: a bioinformatics study revealing homogeneity and heterogeneity among patient cohorts

Gautier Stoll\textsuperscript{a,b,c,d,e}, Jonathan Polo\textsuperscript{a,b,c,d,e}, Vassili Soumelis\textsuperscript{f,g,h,i}, Laurence Zitvogel\textsuperscript{j,k,l,m}, and Guido Kroemer\textsuperscript{o,a,b,c,d,e,n,o}

\textsuperscript{a}Université Paris Descartes, Sorbonne Paris Cité, Paris, France; \textsuperscript{b}Equipe 11 labellisée Ligue Nationale contre le Cancer, Centre de Recherche des Cordeliers, Paris, France; \textsuperscript{c}Institut National de la Santé et de la Recherche Médicale, U1138, Paris, France; \textsuperscript{d}Université Pierre et Marie Curie, Paris, France; \textsuperscript{e}Metabolomics and Cell Biology Platforms, Gustave Roussy Cancer Campus, Villejuif, France; \textsuperscript{f}pôle de biopathologie, Institut Curie, Paris, France; \textsuperscript{g}Institut National de la Santé et de la Recherche Médicale, U932, Paris, France; \textsuperscript{h}CIC IGR-Curie 1428, Paris, France; \textsuperscript{i}PSL*, Paris, France; \textsuperscript{j}Équipe labellisée Ligue Nationale Contre le Cancer, Institut National de la Santé et de la Recherche Médicale, U1015, Villejuif, France; \textsuperscript{k}Institut Gustave Roussy Cancer Campus, Villejuif, France; \textsuperscript{l}Faculty of Medicine, University of Paris Sud, Kremlin-Bicêtre, France; \textsuperscript{m}Center of Clinical Investigations in Biotherapies of Cancer (CICBT), Villejuif, France; \textsuperscript{n}Pôle de Biologie, Hôpital Européen Georges Pompidou, AP-HP, Paris, France; \textsuperscript{o}Department of Women’s and Children’s Health, Karolinska University Hospital, Stockholm, Sweden

ABSTRACT

Multiple soluble factors including proteins (in particular chemokines), non-proteinaceous factors released by dead cells, as well as receptors for such factors (in particular chemokine receptors, formyl peptide receptors and purinergic receptors), influence the recruitment of distinct cell subsets into the tumor microenvironment. We performed an extensive bioinformatic analysis on tumor specimens from 5953 cancer patients to correlate the mRNA expression levels of chemotactic factors/receptors with the density of immune cell types infiltrating the malignant lesions. This meta-analysis, which included specimens from breast, colorectal, lung, ovary and head and neck carcinomas as well as melanomas, revealed that a subset of chemotactic factors/receptors exhibited a positive and reproducible correlation with several infiltrating cell types across various solid cancers, revealing a universal pattern of association. Hence, this meta-analysis distinguishes between homogeneous associations that occur across different cancer types and heterogeneous correlations, that are specific of one organ. Importantly, in four out of five breast cancer cohorts for which clinical data were available, the levels of expression of chemotactic factors/receptors that exhibited universal (rather than organ-specific) positive correlations with the immune infiltrate had a positive impact on the response to neoadjuvant chemotherapy. These results support the notion that general (rather than organ-specific) rules governing the recruitment of immune cells into the tumor bed are particularly important in determining local immunosurveillance and response to therapy.

Introduction

Although cancer has been viewed for a long time as a cell-autonomous, genetic and epigenetic disease, there is little doubt that non-transformed cells, in particular leukocytes have a major influence on the development, progression, therapeutic response and fatal relapse of malignant disease.\textsuperscript{1–4} In essence, the relationship between cancer pathogenesis and leukocytes can be defined as ambiguous. On one hand, chronic subclinical or manifest inflammation can contribute to oncogenesis by favoring local generation of reactive oxygen species, inducing genomic instability, by increasing the turnover of parenchymal or stromal cells, increasing the probability of genetic and chromosomal aberrations, by disrupting homeostatic regulation in the affected tissue, favoring the emergence of more ‘competitive’ cells, and by inducing local immunosuppression, hence subverting immunosurveillance.\textsuperscript{5,6} On the other hand, specific leukocyte subpopulations may contribute to the recognition of premalignant and fully transformed cells, causing their elimination (the best-case scenario) or engaging in a precarious equilibrium state between malignant cells and immune effectors. In most cases, it is only when cancer cells manage to suppress the anticancer immune response (or to hide from recognition) that the disease becomes clinically detectable.\textsuperscript{2,4,7–11}

The attraction of leukocytes into premalignant and malignant lesions is dictated by multiple soluble factors including so-called danger-associated molecular patterns (DAMPs) that include extracellular nucleotides and nucleosides (with ATP as a prominent chemotactic factor for myeloid cells), proteins that are usually confined in intracellular compartments yet are released from stressed, dying and dead cells (such as annexin-1, ANXA1; calreticulin, CALR, F-actin, and high molecular group protein-1, HMGB1)\textsuperscript{12–17} and proteins that are actively secreted by cancer or stromal cells, in particular chemokines (C-C motif chemokine ligand 1 to 9 and 11 to 28, CCL1-9,
These factors act on a series of receptors, such as the adenosine receptor family AdoR (ADOR A1, A2A, A2B and A3), metabotropic purinergic receptors (P2RY1, 2, 4, 6, 8 and 10 to 14), ionotropic purinergic receptors (P2RX1 to 7), formylpeptide receptors (FPR1 to FPR3) and chemokine receptors (XCR1, CXCR1 to 7, CX2CR1, CCR1 to 10, CCRL1 and 2) to induce the chemotaxis of different leukocyte subtypes into the tumor.18-21 The expression of such chemotactic factors and their receptors determines the composition of the tumor immune infiltrate and has a major impact on the therapeutic response and prognosis of cancer.9,22-25

Most of the knowledge about the immune infiltrate that can be retrieved from large data sets is based on microarray or RNAseq data obtained on whole tumor specimens and biopsies, requiring bioinformatic deconvolution of the data.26 Methods for estimating the abundance of leukocyte subsets in tumor, based on gene expression data (microarray or RNAseq), have been developed.27-30

Here, we report a meta-analysis of the relationship between the relative abundance of distinct leukocyte subsets and the level of expression of individual chemotactic factors and their receptors in a large set of human tumors.

**Results and discussion**

**Datasets and expression variability**

We first assembled a series of public microarray data into one single database. Five different datasets were included for several major cancer types, in particular breast carcinoma (TCGA consortium),31-34 colorectal carcinoma (http://www.intgen.org or TCGA consortium),35,36 non-small cell lung carcinoma (TCGA consortium)37-40 and melanoma,41-45 to study the correlation between the expression level of metagenes indicating the presence of immune cell types and a variety of chemotactic factors and receptors. An extra dataset concerning breast carcinoma46 was used for studying expression variability and treatment response. We also used two datasets for ovarian (TCGA consortium)47 and head and neck carcinoma48,49 each. The total number of tumors analyzed in this study amounts to 5953 (Table 1). We first examined the expression variability of each of the metagenes relevant to tumor-infiltrating immune cell types (determined by using the MCP counter method)27,30 and mRNAs encoding annexins (ANXA), chemokines (CCL, CX3CL, CXCL), chemokine receptors (CCR, CXCR, XCR, CCRL), formyl peptide receptors (FPR), purinergic receptors (ADOR, P2RX, P2RY), compared to housekeeping genes (ACTB, GADPH, TUB), as this is typically done when protein expression is measured by immunoblot analysis. However the variability in the expression of housekeeping genes was not expected to be smaller than that of the genes of interest (including immune metagenes) due to the normalization methods of microarrays. Nevertheless, a systematic low variance of genes of interest (including immune metagenes) compared to housekeeping genes, might produce correlation coefficients (computed below) that are overestimated. Therefore, genes that have a systematic low variance should be removed from the analysis. We analyzed the relative variance of each gene of interest, expressed in a log_{10} scale for all datasets (Figure 1). In addition, variance difference tests were performed, between housekeeping genes and genes of interest (Supplementary Figure 1). As a general tendency, the selected genes and metagenes corresponding to distinct immune types exhibited a normalized variance around 1 (i.e. 0 on a log_{10} scale), with no major difference between immune-relevant and housekeeping genes. Nevertheless, some genes have a smaller variance compared to housekeeping genes, especially in the two breast cancer datasets (METABRIC and TCGA). As a result, we decided to include these close-to-invariant genes in the analyses, because the possible overestimation of correlation coefficients is not systematic.

**General correlation patterns across cancer types**

We computed the Spearman correlation coefficients between immune cell type-relevant metagenes and our selection of chemotaxis-relevant genes. We used Fisher’s method (or Fisher’s combined probability test, for multiple testing) to identify reproducible correlations for five cancer types, namely breast cancer (Figure 2), colorectal cancer (Figure 3a), non-small cell lung cancer (Figure 3b) and melanoma (Figure 3c), while performing a meta-analysis of five different data sets for each cancer (Table 1). This procedure was not applicable to ovarian and head & neck cancers because it was not possible to collect 2 × 5 sufficiently large datasets. Correlations that were reproducible across the five data sets were color-coded to indicate positive (red) or negative (green) associations between the expression levels of individual chemotactic factors/receptors and leukocyte subset-specific metagenes, while the absence of reproducible correlations were indicated by white boxes. As to be expected, we detected a mostly positive correlation between leukocyte subsets and chemotactic factors/receptors in all four cancer types (Figures 2 and 3). The reproducibility of correlations was less pronounced in melanoma (Figure 3C), probably because the datasets available for this cancer type were comparatively small (Table 1).

When all these figures were grouped together, (Supplementary Figure 2), one could observe that the variability seemed larger among cancer types than among distinct immune cell types. For this reason, we performed separate analyses of distinct cancer types in different (sub-)figures and then grouped immune cell types. We applied Fisher’s method to identify correlations that were reproducible across the aforementioned five cancer types (Figure 4). Unsupervised hierarchical clustering of these results identified two types of genes coding for chemotactic factors/receptors, namely (i) genes that exhibit reproducible positive correlations with most of the leukocyte subtypes (in the lower part of Figure 4) and (ii) genes that have almost no reproducible correlations with immune cell types (in the upper part of Figure 4). The interpretation of this separation in two groups is not obvious. The second group of chemotactic factors/receptors could be interpreted as a cancer type-specific immune system activation. To visualize this possibility, we removed the reproducible correlations across cancer types.
| Cancer type | Cohort name | Number of samples | Characteristics of the cohort | Treatment & outcome | Reference | Platform |
|-------------|-------------|-------------------|-----------------------------|---------------------|-----------|----------|
| Melanoma    | Xu          | 83                | Primary and metastatic tumors |                     | GSE8401   | Affymetrix Human Genome U133A Array |
| Melanoma    | Harlin      | 44                | Metastatic tumors           |                     | GSE12627  | Affymetrix Human Genome U133A Array |
| Melanoma    | Bogunovic   | 44                | Metastatic tumors           |                     | GSE19234  | Affymetrix Human Genome U133 Plus 2.0 Array |
| Melanoma    | RikerMel    | 56                | Primary and metastatic tumors |                     | GSE7553   | Affymetrix Human Genome U133 Plus 2.0 Array |
| Melanoma    | Talantov    | 45                | Primary tumors              |                     | GSE3189   | Affymetrix Human Genome U133 Plus 2.0 Array |
| Colon       | BittColon   | 307               | Various colon tumors        |                     | GSE2109   | Affymetrix Human Genome U133A Array |
| Colorectal  | Smith       | 177               | Various colorectal tumors   |                     | GSE17536  | Affymetrix Human Genome U133 Plus 2.0 Array |
| Colon       | Vilar1      | 155               | Colon tumors                |                     | GSE26682, 1st set | Affymetrix Human Genome U133 Plus 2.0 Array |
| Colon       | Vilar2      | 176               | Colon tumors                |                     | GSE26682, 2nd set | Affymetrix Human Genome U133A Array |
| Colon       | TCGA        | 174               | Various colon tumors        |                     | TCGA consortium |                         |
| Breast      | METABRIC    | 1781              | Various breast tumors       |                     | METABRIC Study v3 |                         |
| Breast      | TCGA        | 522               | Various breast tumors       |                     | TCGA consortium |                         |
| Breast      | Bonnefoi    | 161               | Locally advance or large operable breast tumors, estrogen receptor negative | FEC or ET treatment. Pathological complete response (complete disappearance of the tumour with no more than a few scattered tumour cells) vs no pathological complete response | GSE6861   | Affymetrix Human X3P Array |
| Breast      | Hatzis      | 198               | HER2 negative breast tumors | Taxane-anthracycline chemotherapy preoperatively and endocrine therapy if ER-positive. Pathological complete response (no invasive or metastatic breast cancer identified) vs rapid development | GSE25065  | Affymetrix Human Genome U133A Array |
| Breast      | Tabchy      | 178               | Various type of breast tumors before treatment | FCE or FAC neo-adjuvant chemotherapy. Pathological complete response vs residual disease (clinical or radiological progression) | GSE20271  | Affymetrix Human Genome U133A Array |
| Breast      | Korde       | 61                | Various type of breast tumors, stage 2 or 3 breast cancer with tumor size ≥2cm at patients selection, prior to AC treatment | 4 cycles of TX, 4 cycles of adriamycin, cyclophosphamide on day 1 and 21 (neoadjuvant) and AC (neo-adjuvant or adjuvant). Response vs no response (change in tumor size by clinical exam and pathological response). | GSE18728  | Affymetrix Human Genome U133 Plus 2.0 Array |
| Lung        | AdenoConsortium | 462        | Various type of Adenocarcinomas |                     |                         |                     |
(Figure 4) from each organ-specific correlation pattern (Figures 2 and 3) with the scope of identifying cancer type-specific correlation patterns (Supplementary Figure 3–6). Again, this separation between fully reproducible and cancer specific-reproducible correlations supported the idea that part of the chemotaxis of different leukocyte subtypes may occur in an organ-specific fashion. We tested the robustness of this two-group separation of chemotactic factors/receptors, by changing the p-value threshold used for reproducibility tests from 0.05 to 0.01. Changing the threshold of the p-value did not affect the final result (Supplementary Fig. 7B) as in Figure 4.

Next, we investigated the reproducible correlation pattern identified in Figure 4 on two other cancer types, namely, ovarian cancer and head and neck carcinoma. Because the number of large datasets for these two cancers types is rather limited (two datasets for each of them, Table 1), it was not possible to determine reproducible correlation patterns within each of these two cancer types. For each dataset, we computed the correlation matrix between immune cell types and the selection of genes. Then, we removed the reproducible correlations identified in Figure 4. The remaining “specific” correlations are illustrated in Supplementary Fig. 8–11. The number of such specific correlations was rather small, exhibiting no obvious pattern and hence resembled those obtained for breast, colorectal, non-small cell lung cancer and melanoma (Supplementary Figure 3–6). Therefore, ovarian cancer and head and neck carcinoma exhibited a similar global correlation pattern between chemotactic factors/receptors and the density of the immune infiltrate, supporting the universal validity of the common correlation pattern (Figure 4) across distinct malignancies.

One question that can be raised concerns the cancer specificity of these reproducible correlations (fully reproducible or organ-specific reproducible) compared to normal tissue. Although a full analysis, comparing normal and cancer tissue is beyond the scope of this work, we produced a similar Figure than Figure 2, using normal breast samples available from the METABRIC dataset (Supplementary Fig. 12). The global pattern is not strongly different from Figure 2. Therefore, it seems that the mechanism involving chemotactic factors/receptors in immune infiltrate in cancer in not fundamentally different from a normal immune activation.
Figure 1. Heatmap representation of gene expression variances and immune cell type activity variances, in log$_{10}$, for the different datasets. For each gene, in each dataset, gene variance was normalized by the variance all gene expressions pooled together. For each immune cell type activities, in each dataset, immune cell type activity variance is normalized by the variance all immune cell type activities pooled together. Housekeeping genes are highlighted.
The reproducible correlation pattern across cancers (Figure 4) has been established without any knowledge about the abundance of the chemotactic factors/receptors in different immune cell types or in distinct cancer types. This information is not directly accessible from microarray data, because of normalization methods. Nevertheless, we explored the possibility that the abundance of chemotactic factors/receptors that have a reproducible

**Figure 2.** Heatmap representation of reproducible Spearman’s correlation coefficients, in breast carcinomas.
correlation with immune cell types might be higher or lower than that of those failing to exhibit reproducible correlations, while investigating absolute expression in our microarray datasets and the mRNA expression data from the Human Protein Atlas.\textsuperscript{50,51} Chemotactic factors/receptors were removed from the analyses if their expression was systemically low in the microarray data or in the spleen and lymph node (see Materials and Methods for the exact procedure). Among the 66 chemotactic factors/receptors that were classed as non-reproducible in their correlation with immune metagenes (in Figure 4),\textsuperscript{1,2} exhibited a low expression. Among the remaining 39 chemotactic factors/receptors reproducibly correlating with immune metagenes (in Figure 4), only one exhibited a low expression. Hence, correlation reproducibility is anti-correlated with low expression (p-value = 0.029, exact Fisher’s test). Indeed, lowly expressed genes should not yield reproducible correlations with immune cell infiltrates. Because the choice for filtering thresholds was quite arbitrary (see Material and Methods), we applied the same procedure by a) doubling threshold on our mRNA datasets, b) doubling threshold on mRNA data from Human Protein Atlas.\textsuperscript{50,51} We still observed the same tendency; in the case a) Fisher’s p-value = 0.155 (filtering 2 genes from the reproducible list and 14 from the non-reproducible one), in case b), Fisher’s p-value = 0.028 (filtering 2 genes from the reproducible list and 14 from the non-reproducible one).

Figure 4 reveals 13 chemotactic receptors and chemokines that positively correlated with all identified cell subsets (irrespective of their stromal and leukocytic nature) contained in the tumor environment, regardless of the cancer type. This applies to CCR1,2,5; CXCR4; FPR1; ADORA2A; P2RY6,13,14; CCL18,19,21; and CXCL12. In other words, these factors appear intrinsically linked to the neoplastic process. Accordingly, CXCL12 and its receptor CXCR4 have been involved in promoting cancer proliferation, survival, invasion, metastasis, stemness and angiogenesis.\textsuperscript{52–57} CCR1 and its ligand CCL5 have been associated with cancer cell invasion and metastasis.\textsuperscript{58–60} Similar observations have been reported for the CCR2-CCL2, CCR5-CCL2,3,4,5, CCR7-CCL19,21 and CCR8-CCL18 axes.\textsuperscript{61–69} The role of FPR1 in tumor progression remains controversial as it has been associated with tumor invasion in colorectal cancer but with tumor suppression in gastric cancer.\textsuperscript{70,71} As far as P2RY6 is concerned, Placet et al. recently suggested its role in cancer cell survival.\textsuperscript{72} Overall, these 13 chemotactic receptors and chemokines may constitute universal targets for cancer therapy. Further consolidating the relevance of our results, additional correlations between tumor-infiltrating immune cell subtypes and chemokine/receptors across cancer types previously described in separate preclinical/clinical studies were found (Figure 4). Among these, we can mention the association between the infiltration of T cells and the requirement of the receptor ADORA2\textsuperscript{73} or the role of CXCR3 and its ligands CXCL9 to 11 in the recruitment of CD8\textsuperscript{+} T cells, Th1 cells and NK cells.\textsuperscript{18} However, Figure 4 also revealed original interactions between immune cell subtypes. For instance, neutrophil infiltrates appeared negatively

\textbf{Figure 3.} Heatmap representation of reproducible Spearman’s correlation coefficients, in colorectal carcinomas (A), in non-small cell lung cancers (B) and in melanomas (C).
correlated with the chemokines CXCL9-11. Based on these results, the recruitment of T lymphocytes and neutrophils into the tumor microenvironment could be mutually exclusive. Indeed, a recent work demonstrating an inhibitory activity of CXCL9 peptides on neutrophil migration in murine models supporting such a hypothesis.⁷⁴

Figure 4. Heatmap representation of reproducible Spearman’s correlation coefficients, across different cancer types, integrating the results shown in Figure 2 to Fig. 5 on breast, colorectal, non-small cell lung cancer and melanoma.
**Impact of chemotactic factors and receptors on breast cancer prognosis**

It is well established that breast cancer is under immunosurveillance and that neoadjuvant chemotherapy is particularly successful if the cancer is massively infiltrated by immune effectors including T lymphocytes. Driven by the dichotomy between mRNAs coding for chemotactic factors/receptors that either (i) positively correlated with leukocyte infiltration across distinct cancer types or (ii) lacked a clear pattern of correlation (Figure 4), we investigated the impact of these two groups of genes on the survival of breast cancer patients. We subdivided the second group into two subgroups (see Supplementary Table 1): a) genes that positively correlate with leukocyte infiltration in breast cancer (but fail to do so across the entire spectrum of cancers analyzed in this study), b) remaining chemotactic factors/receptors (that fail to correlate with immune metagenes in breast cancer). We also considered the group of chemotactic factors/receptors that positively correlated with leukocyte infiltration across distinct cancer types, without the single genes that are characterized by low expression (see Supplementary Table 1).

We explored the link between chemotactic factors/receptors and response to treatment in the four out of the five Breast Carcinoma datasets for which the treatment response was annotated (as pathological complete response or simple response, see Table 2) after neo-adjuvant chemotherapy: Bonnefoi, Hatzis, Tabchy, and Korde. We also included the METABRIC dataset in this analysis, because we had access to the information whether patients were alive or deceased, a status that is related to disease prognosis and treatment response. In four out of these five datasets (all except the Korde cohort, probably due to its small size), expression of chemotactic factors/receptors that positively correlated with leukocyte infiltration across distinct cancer types (as defined in Figure 4) was positively associated with treatment response (Bonnefoi, Hatzis and Tabchy datasets) or overall survival (METABRIC dataset) (Table 2). METABRIC results differed for non-reproducible chemotactic factors/receptors, probably because treatment response is measured indirectly. In contrast, no correlation was observed between the remaining chemotactic factors/receptors and prognostic features (Table 2). This is also illustrated by the Volcano Plots of different groups of chemotactic factors/receptors (Supplementary Figure 13). Because METABRIC dataset has the clinical information about survival, we attempted to build simple biomarkers (see Material and Methods) from the different groups of chemotactic factors/receptors (Supplementary Figure 14). Only the group of chemotactic factors/receptors reproducibly detected across different cancer types tended to separate the cohort into two groups based on the expression level of the biomarkers, with a difference between survival outcomes that happened to be almost significant (p = 0.081). Altogether, these data support the idea that chemotactic factors and receptors that positively and reproducibly correlate with the immune infiltrate of different cancer types play a positive role in local immunosurveillance, hence improving the probability of complete therapeutic responses in the context of neoadjuvant chemotherapy. The global function of this group of reproducible chemotactic factors/receptors is not well established, and therefore this group cannot be considered as a single metagene, like for immune cell types. Nevertheless it may be interesting to compare the differential abundance of these chemotactic factors/receptors with the differential expression of the immune cell metagenes (in Supplementary Table 2, one sided t-tests of differential expression of treatment response). Although the combined t-tests produce significant differential expression of immune cell metagenes, only one cohort (Bonnefoi) has significant p-values on its own (thresholds 0.05), which appears less significant than the tests for chemotactic factors/receptors of Table 2. It is always difficult to interpret such variation between cohorts. The main reason could reside in the treatment regimens and in the clinical definition of treatment response that were not the same for all cohorts.

**Concluding remarks**

The present bioinformatic study has been conducted with the explicit aim of identifying universal (rather than organ-specific) correlations between the expression level of mRNAs coding for chemotactic factors and receptors, and metagene signatures reflecting distinct leukocyte subtypes infiltrating a variety of cancer types. The specific strategy that we have chosen for this

---

**Table 2. Combined p-values are obtained by applying Fisher’s method.**

| Breast cancer dataset | Criterion of positive prognosis | P-value of exact Fisher test, for positively correlated, reproducible mRNAs coding for chemotactic factors/receptors | P-value of exact Fisher test, for positively correlated, reproducible mRNAs coding for chemotactic factors/receptors, with no low expression | P-value of exact Fisher test, for breast cancer-specific mRNAs coding for chemotactic factors/receptors | P-value of exact Fisher test, for other chemotactic factors/receptors |
|-----------------------|---------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------|---------------------------------------------------------------|
| Bonnefoi              | Pathological complete response   | 5.65e-16                                                      | 3.95e-5                                                      | 0.0258                          | 1                                                            |
| Hatzis                | Pathological complete response   | 0.00118                                                       | 0.00104                                                      | 0.443                           | 1                                                            |
| Tabchy                | Pathological complete response   | 1.24e-6                                                       | 1.08e-6                                                      | 0.0200                          | 1                                                            |
| Korde                 | Treatment response               | 1                                                            | 1                                                            | 1                               | 0.227                                                        |
| METABRIC              | Alive                            | 0.0141                                                        | 0.0137                                                       | 1                               | 0.00183                                                      |
| Combined              |                                  | 6.56e-21                                                      | 3.61e-21                                                      | 0.0797                          | 0.247                                                        |
meta-analysis addresses important aspects with respect to the homogeneity and heterogeneity of patient cohorts. Undoubtedly, there are major interindividual variations in the density, composition, architecture and functional state of the tumor infiltrate.\textsuperscript{9,11,82,83} Notwithstanding this fact, the variations in metagene expression across different cancer types were not higher than those observed for a well known ‘house-keeping’ gene (ACTB, GAPDH, TUB), suggesting that the heterogeneity of the immune infiltrate is not substantially higher than that of the expression of mRNAs coding for essential components of the cytoskeleton (ACTB) or glycolytic metabolism (GAPDH).

A second notion that emerges from this meta-analysis concerns the regulation of the immune infiltrate by chemotactic factors and receptors occurring in a universal (homogeneous) pattern across distinct cancer types, irrespective of their localization in the body, versus organ-specific (heterogeneous) patterns. Surprisingly, it appears that most of the (positive) correlations between the chemotactic factors and receptors analyzed herein and different infiltrating leukocyte subtypes are highly reproducible across different cancer types, while few of them are organ-specific. This points to the importance of common rules dictating the recruitment and permanence of immune cells into the tumor bed, irrespective of the precise cancer type that is concerned. Such universal rules appear to supersede in importance those that may influence organ-specific circuitries. Because this analysis is based on correlation, the causal interpretation is not obvious (are chemotactic factors/receptors attracted by cell types, or is it the contrary?). A better understanding of these circuitries would include a description/modeling of signaling networks at the microenvironment level, which is beyond the scope of the present work.

Most importantly, it appears that the mRNAs coding for chemotactic factors and receptors that reproducibly and positively correlate with the local presence of distinct immune cell types seem to have a positive impact on breast cancer prognosis. This contrasts with chemotactic factors and receptors that exhibit heterogeneous and organ-specific patterns of correlation with the immune infiltrate that have no such prognostic impact. On a positive note, this implies that, at least in the realm of chemotaxis, insights that have been gained in one cancer type can be extrapolated to other malignancies, irrespective of their cell of origin.

**Material and methods**

**Datasets and immune infiltrate estimation**

We explored public datasets of transcriptome microarrays. The complete list of these datasets is shown in Table 1. We use data that were already normalized, as provided from the different repository websites. The MCPcounter\textsuperscript{22} R-package was used for estimating immune cell type activities. This method uses groups of genes to construct metagenes, whose expressions measure indirectly immune cell abundances. The identification of gene groups have been done in a strongly reproducible manner, learned from datasets where immune cell type abundances are known.

**Test of variance difference**

One sided F-test was used, to test if variance of genes of interest (including immune metagenes) is smaller than variance of housekeeping genes.

**Patterns of reproducible correlation**

For each dataset, we computed the correlation coefficients between immune cell activities and expression of individual mRNAs coding for chemotactic factors/receptors. We decided to use the Spearman’s correlation method, because the normal hypothesis necessary for using Pearson’s approach is not guaranteed for immune activities estimated by MCP counter method. For each cancer type, we used the Fisher’s method in the following way. We computed one-sided Spearman correlation test for both cases (H1: positive correlation and H1: negative correlation). For each case (H1 positive and H1 negative), we combined the p-value by using Fisher’s method (Chi-square test on sum of p-value logarithms). If both p-values were higher than 0.05, the correlation was considered as non-reproducible. In any other case, we considered that the correlation was reproducible. We then used the coefficient that has the smallest p-value and whose sign is compatible with the hypothesis H1 that had a significant combined p-value. These correlation coefficients are illustrated in Figure 2–3.

For the fully reproducible correlation coefficients listed in Figure 6, we applied the same procedure as above. Thus, we considered the correlations coefficients of Figure 2–3 and their associated p-values, as computed above (combined one-sided correlation tests by Fisher’s method). These p-values were one-sided ones, related to the sign of the associated correlation coefficient. We computed the p-values associated to one-sided correlation test of opposite sign (‘one sided p-value’ → ‘1-one sided p-value’). Therefore, we could apply Fisher’s method for both H1 hypothesis (positive and negative correlation) to a common set of cancer types. Again, if both p-values were higher than 0.05, the correlation was interpreted as non-reproducible. In the other case, we considered that correlation was reproducible; we then used the coefficient that has the smallest p-value and whose sign is compatible with the hypothesis H1 that has a significant combined p-value. These correlation coefficients are shown in Figure 4, Suppl. Fig 7 is produced by using the same analysis, replacing the p-value threshold of 0.05 (Figure 4) by 0.01 (Suppl. Fig. 7).

**Filtering for low expression**

The first filter is based on microarray expression among the five cancer types. For each dataset, we computed the average (among different patients) of the expression rank, for each chemotactic factors/receptors. Because no obvious threshold emerged from these mean rank expressions, we choose it arbitrarily (value = 0.2) to filter out a reasonable amount of these chemotactic factors/receptors that are expressed at a low level.

The second filter is based on mRNA expression from the Human Protein Atlas\textsuperscript{50,51,84} from the three methods (HPA,
GTEX, FANTOM5), in the spleen and lymph node. Again, no obvious threshold emerged from these data. Therefore we arbitrarily considered that a gene has a low expression if its mRNA abundance is lower than (TPM/RPKM = 1), as determined by all three methods (HPA, GTEX, FANTOM5) and in both spleen and lymph node.

We applied these two filters to the list of chemotactic factors/receptors. We also considered a threshold of 2 for (TPM/RKM), applying the same analysis.

**Fisher’s exact test for treatment response**

We considered genes that were overexpressed in responsive tumors (or alive patients in METABRIC dataset), using a p-value threshold of 5%, a logFC threshold of log₂(1.2) for the METABRIC, Bonnefoi and Korde datasets, a logFC threshold of log₂(1.5) for Hatzis’ and Tabchy’s datasets (these thresholds varied in order to obtain non-empty intersections with the reproducible correlation gene list of Figure 4), using limma R-package.89 For each dataset, we computed the contingency tables regarding overexpressed genes and reproducible correlation gene list of Figure 4 and applied Fisher’s exact tests. Significant results are shown in Table 2.

**Acknowledgments**

We thank Prof. Catherine Sautès-Fridman and Dr. Florent Petitprez (Centre de Recherche des Cordeliers, Paris, France) for critical comments. GK is supported by the Ligue contre le Cancer Comité de Charente-Maritime (équipe labellisée); Agence National de la Recherche (ANR) – Projets blancs; ANR under the frame of E-Rare-2, the ERA-Net for Research on Rare Diseases; Association pour la recherche sur le cancer (ARC); Cancéropôle Ile-de-France; Chancellerie des universités de Paris (Legs Poix), Fondation pour la Recherche Médicale (FRM); the European Research Council (ERC); the LabEx Immuno-Oncology; the RHU Torino Lumière, the Seerave Foundation; 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).

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**ORCID**

Guido Kroemer [http://orcid.org/0000-0002-9334-4405](http://orcid.org/0000-0002-9334-4405)

**References**

1. Kroemer G, Poussegur J. Tumor cell metabolism: cancer’s Achilles’ heel. Cancer Cell. 2008;13:472–482. doi:10.1016/j.ccr.2008.05.005.
2. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity’s roles in cancer suppression and promotion. Science. 2011;331:1565–1570. doi:10.1126/science.1203486.
3. Hanahan D, Weinberg RA. Hallmarks of cancer; the next generation. Cell. 2011;144:646–674. doi:10.1016/j.cell.2011.02.013.
4. Palucka AK, Coussens LM. The basis of oncoimmunology. Cell. 2016;164:1233–1247. doi:10.1016/j.cell.2016.01.049.
5. Balkwill F. Cancer and the chemokine network. Nat Rev Cancer. 2004;4:540–550. doi:10.1038/nrc1388.
6. Lin WW, Karin M. A cytokine-mediated link between innate immunity, inflammation, and cancer. J Clin Invest. 2007;117:1173–1183. doi:10.1172/JCI31537.
7. Zitvogel L, Apetoh L, Ghiringhelli F, Kroemer G. Immunological aspects of cancer chemotherapy. Nat Rev Immunol. 2008;8:59–73. doi:10.1038/nri2216.
8. Zitvogel L, Kepp O, Kroemer G. Immune parameters affecting the efficacy of chemotherapeutic regimens. Nat Rev Clin Oncol. 2011;8:151–160. doi:10.1038/nrcc3245.
9. Fridman WH, Pages F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. 2012;12:298–306. doi:10.1038/nrc3349.
10. Kroemer G, Senovilla L, Galluzzi L, Andre F, Zitvogel L. Natural and therapy-induced immunosurveillance in breast cancer. Nat Med. 2015;21:1128–1138. doi:10.1038/nm.3944.
11. Fridman WH, Zitvogel L, Sautès-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol. 2017;14:717–734. doi:10.1038/nrcc3381.
12. Obeid M, Tesniere A, Papametaxia T, Tufo R, Joza N, van Endert P, Lionel G, Nathalie A, Caroline C, Evelyn F, et al. Ecto-calreticulin in immunogenic chemotherapy. Immunol Rev. 2007;220:22–34. doi:10.1111/j.1600-065X.2007.00567.x.
13. Apetoh L, Ghiringhelli F, Tesniere A, Criollo A, Ortiz C, Lidereau R, Nathalie M, Jean C, Mira P, Delaloge S, André F, et al. The interaction between HMG1 and TLR4 dictates the outcome of antitumor chemotherapy and radiotherapy. Immunol Rev. 2007;220:47–59. doi:10.1111/j.1600-065X.2007.00573.x.
14. Vacchelli E, Ma Y, Baracco EE, Sistigu A, Enot DP, Pietrocola F, Yang H, Adjemian S, Chabà K, Semararo M, et al. Chemotherapy-induced antitumor immunity requires formyl peptide receptor 1. Science. 2015;350:972–978. doi:10.1126/science.aad0779.
15. Ahrens S, Zelenay S, Sancho D, Hanc P, Kjaer S, Feest C, Fletcher G, Durck C, Postigo A, Skehel M, et al. F-actin is an evolutionarily conserved damage-associated molecular pattern recognized by DNLR-1, a receptor for dead cells. Immunity. 2012;36:635–645. doi:10.1016/j.immuni.2012.03.008.
16. Stoll G, Zitvogel L, Kroemer G. Differences in the composition of the immune infiltrate in breast cancer, colorectal carcinoma, melanoma and non-small cell lung cancer: A microarray-based meta-analysis. Oncoimmunology. 2016;5:e1067746. doi:10.1080/2162042X.2015.1067746.
17. Stoll G, Iribarren K, Michels J, Leary A, Zitvogel L, Cremer I, Kroemer G. Calreticulin expression: interaction with the immune infiltrate and impact on survival in patients with ovarian and non-small cell lung cancer. Oncoimmunology. 2016;5:e1177692. doi:10.1080/2162042X.2016.1177692.
18. Nagarsheth N, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. Nat Rev Immunol. 2017;17:559–572. doi:10.1038/nri.2017.49.
19. Schulz O, Hammerschmidt SI, Muschovakis GL, Forster R. Chemokines and chemokine receptors in lymphoid tissue dynamics. Annu Rev Immunol. 2016;34:203–242. doi:10.1146/annurev-immunol-041015-055649.
20. Di Virgilio F, Adinolfi E. Extracellular purines, purinergic receptors and tumor growth. Oncogene. 2017;36:293–303. doi:10.1038/onc.2016.206.
21. Di Virgilio F, Falzoni S, Giuliani AL, Adinolfi E. P2 receptors in cancer progression and metastatic spreading. Curr Opin Pharmacol. 2016;29:17–25. doi:10.1016/j.coph.2016.05.001.
22. Qian BZ, Li J, Zhang H, Kitamura T, Zhang J, Campion LR, Kaiser EA, Snyder LA, Pollard JW. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature. 2011;475:222–225. doi:10.1038/nature10138.
23. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obernau AC, Angell H, Fredriksen T, Lafontaine L, Berger A, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity. 2013;39:782–795. doi:10.1016/j.immuni.2013.10.003.
24. Highfill SL, Cui Y, Giles AJ, Smith JP, Zhang H, Morse E, Kaplan RN, Mackall CL. Disruption of CXCR2-mediated MDSC tumor...
trafficking enhances anti-PDI efficacy. Sci Transl Med. 2014;6:237ra67. doi:10.1126/scitranslmed.3007974.

25. Peng D, Kryczek I, Nagarsheth N, Zhao L, Wei S, Wang W, Sun Y, Zhao E, Vatan L, Szeliga W, et al. Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. Nature. 2015;527:249–253. doi:10.1038/nature15520.

26. Hackl H, Charoenpong P, Finotello F, Trajanoski Z. Computational genomics tools for dissecting tumour-immune cell interactions. Nat Rev Genet. 2016;17:441–458. doi:10.1038/nrrgen.2016.67.

27. Becht E, Giraldo NA, Lacroix B, Buttard B, Elarouci N, Petitprez F, SELves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:218. doi:10.1186/s13059-016-1070-5.

28. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–457. doi:10.1038/nmeth.3337.

29. Gajoux R, Seoige C. CellMix: a comprehensive toolbox for gene expression deconvolution. Bioinformatics (Oxford, England). 2013;29:2211–2212. doi:10.1093/bioinformatics/btt351.

30. Petitprez F, Vano YA, Becht E, Giraldo NA, de Reynies A, Sautès-Fridman C, Fridman WH. Transcriptomic analysis of the tumour microenvironment to guide prognosis and immunotherapies. Cancer Immunol Immunother. 2017.

31. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486:346–352. doi:10.1038/nature10983.

32. Bonnefoi H, Piccart M, Bogaerts J, Mauriac L, Fumoleau P, Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–457. doi:10.1038/nmeth.3337.

33. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Thomas D, Moskaluk C, Wang Y, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res. 2006;66:7466–7472. doi:10.1158/0008-5472.CAN-06-1191.

34. Xu L, Shen SS, Hoshida Y, Subramanian A, Ross K, Brunet JP, Wagner SN, Ramaswamy S, Mesirov JP, Hynes RO. Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. Mol Cancer Res. 2008;6:760–769. doi:10.1158/1541-7786.MCR-07-0344.

35. Harlin H, Meng Y, Peterson AC, Zha Y, Tretiakova M, Sluling C, McKee M, Gajewski TF. Chemokine expression in melanoma metastases associated with CD8+ T-cell recruitment. Cancer Res. 2009;69:3077–3085. doi:10.1158/0008-5472.CAN-08-2281.

36. Bogunovic D, O’Neill DW, Belitskaya-Letty I, Vacic V, Yu YL, Adams S, Dravisvian F, Berman R, Shapiro R, Pavlick AC, et al. Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival. Proc Natl Acad Sci U S A. 2009;106:20429–20434. doi:10.1073/pnas.0905190106.

37. Riker AI, Enkemann SA, Fodstad O, Liu S, Ren S, Morris C, Xi Y, Howell P, Metge B, Samant RS, et al. The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. BMC Med Genomics. 2008;1:13. doi:10.1186/1755-8794-1-13.

38. Talantov D, Mazumder A, Yu JX, Briggs T, Jiang Y, Backus J, Atkins D, Wang Y. Novel genes associated with malignant melanoma but not benign melanocytic lesions. Clin Cancer Res. 2005;11:7234–7242. doi:10.1158/1078-0432.CCR-05-0683.

39. Korde LA, Lusa L, McShane L, Lebowitz PF, Lukes L, Camphausen K, Parker JS, Swain SM, Hunter K, Zujewski J. Gene expression pathway analysis to predict response to neoadjuvant doxorubicin and capcitabine for breast cancer. Breast Cancer Res Treat. 2011;119:685–699. doi:10.1007/s10549-009-0651-3.

40. Yoshihara K, Tajima A, Yahata T, Kodama S, Fujitawa H, Suzuki K, Onishi Y, Hatae M, Sueyoshi K, Fujitawa H, et al. Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. PLoS One. 2010;5:e9615. doi:10.1371/journal.pone.0009615.

41. Peng CH, Liao CT, Peng SC, Chen YJ, Cheng AJ, Juan JL, Tsai Y-C, Chen T-C, Chuang Y-J, Teng C-Y, et al. A novel molecular signature identified by systems genetics approach predicts prognosis in oral squamous cell carcinoma. PLoS One. 2011;6:e23452. doi:10.1371/journal.pone.0023452.

42. Rickman DS, Millon R, De Reynies A, Thomas E, Wasylyk C, Muller D, Abecasis 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;27:6607–6622. doi:10.1038/onc.2008.251.

43. Uhlen M, Börjling E, Agaton C, Szigiartyo CA, Amini B, Andersen E, Andersson A-C, Angelidou P, Asplund A, Asplund C, et al. A knowledge-based human protein atlas. Nat Biotechnol. 2010;28:1248–1250. doi:10.1038/nbt.1210.1248.

44. Kryczek I, Wei S, Keller E, Liu R, Zou W. Stroma-derived factor (SDF-1/CXCL12) and human tumor pathogenesis. Am J Physiol Cell Physiol. 2005;292:C987–995. doi:10.1152/ajpcell.00406.2006.

45. Murakami T, Maki W, Cardones AR, Fang H, Tun Kyi A, Nestle FO, Wasylyk B. Stroma-derived factor (SDF-1/CXCL12) and human tumor pathogenesis. Am J Physiol Cell Physiol. 2005;292:C987–995. doi:10.1152/ajpcell.00406.2006.
