Natural killer cell and stroma abundance are independently prognostic and predict gastric cancer chemotherapy benefit

Bailiang Li, … , George A. Fisher, Ruijiang Li


In-Press Preview

Graphical abstract

Find the latest version:

https://jci.me/136570/pdf
Natural Killer Cell and Stroma Abundance are Independently Prognostic and Predict Gastric Cancer Chemotherapy Benefit

Bailiang Li¹, PhD, Yuming Jiang¹, MD PhD, Guoxin Li², MD PhD, George Fisher³, MD PhD, Ruijiang Li¹, PhD

1. Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA, USA
2. Department of General Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China
3. Department of Medicine, Division of Medical Oncology, Stanford University School of Medicine, Stanford, CA, USA

Bailiang Li and Yuming Jiang contributed equally to this work.

Financial support: This project is partially supported by the NIH grant 1R01 CA222512.

Correspondence to:

Ruijiang Li, PhD
Department of Radiation Oncology
Stanford University School of Medicine,
1070 Arastradero Rd, Palo Alto, CA 94304
Email: rli2@stanford.edu
Phone: 1-650-2063961

The authors declare no conflict of interest.
Background: Specific features of the tumor microenvironment (TME) may provide useful prognostic information. We conducted a systematic investigation of the cellular composition and prognostic landscape of TME in gastric cancer.

Methods: We evaluated the prognostic significance of major stromal and immune cells within TME. We proposed a composite TME-based risk score and tested it in six independent cohorts of 1,678 patients with gene expression or immunohistochemistry measurements. Further, we devised a new patient classification system based on TME characteristics.

Results: We identified natural killer cells, fibroblasts, and endothelial cells as the most robust prognostic markers. The TME risk score combining these cell types was an independent prognostic factor when adjusted for clinicopathologic variables (gene expression: HR [95% CI]: 1.42 [1.22–1.66]; immunohistochemistry: 1.34 [1.24–1.45], P<0.0001). Higher TME risk scores consistently associated with worse survival within every pathologic stage (HR range: 2.18-3.11, P<0.02) and among patients who received surgery only. The TME risk score provided additional prognostic value beyond stage, and combination of the two improved prognostication accuracy (likelihood-ratio test $\chi^2 = 235.4$ vs. 187.6, P<0.0001; net reclassification index: 23%). The TME risk score can predict the survival benefit of adjuvant chemotherapy in non-metastatic patients (stage I-III) (interaction test P<0.02). Patients were divided into four TME subtypes that demonstrated distinct genetic and molecular patterns and complemented established genomic and molecular subtypes.

Conclusion: We developed and validated a TME-based risk score as an independent prognostic and predictive factor, which has the potential to guide personalized management of gastric cancer.
**Introduction**

Gastric cancer (GC) is a heterogeneous disease with diverse clinical, histological, and molecular characteristics (1). When diagnosed at early stages, gastric cancer can be effectively treated with endoscopic or surgical resection with or without adjuvant therapy. However, survival outcomes can vary widely among patients receiving the same treatment for disease of the same stage (2, 3). A critical unmet need is accurate prognostication tools beyond the current staging system to better guide adjuvant treatment. On the other hand, the prognosis for advanced gastric cancer is extremely poor with a median survival less than one year. New, effective treatments are needed to prolong survival for patients with metastatic disease.

Recent whole-genome and transcriptome studies have significantly improved our understanding of the pathobiology and molecular alterations of gastric cancer. TCGA divided gastric cancer into four subtypes based on genomic characteristics: Epstein-Barr virus (EBV), microsatellite instability (MSI), genomically stable (GS), and chromosomal instability (CIN) (4). The Asian Cancer Research Group (ACRG) proposed another molecular classification system of four subtypes that demonstrated distinct prognoses, most notably between MSI and MSS/EMT subtypes (5). Subsequent studies defined alternative patient clusters that largely overlapped with ACRG subtypes and in part validated these findings (6, 7). The existing subtypes are focused on cancer cell-intrinsic molecular characteristics of the tumor (8-10).

Solid tumors are a complex ecosystem that does not only consist of neoplastic cells, but also includes a variety of non-malignant cell types such as stromal and immune cells, which together with soluble factors and extracellular matrix constitute the tumor microenvironment (TME) (11). TME plays an important role in cancer progression, metastasis, therapeutic response and resistance (12, 13). Specific features of TME have been shown to provide useful prognostic and predictive information in multiple cancers (14-17). More importantly, various components of TME represent potential therapeutic targets (18). Immunotherapy, in particular, immune check
inhibitors, has been the most promising treatment strategy in the past decade (19, 20). However, the response rates remain relatively low at around 15% in the phase III KEYNOTE-062 trial (21). Novel strategies and combination therapies are needed to improve response and survival. This will require a deeper understanding of the TME and its clinical significance.

In this work, we conducted a systematic evaluation of the cellular composition and prognostic landscape of the TME in gastric cancer. We proposed a TME-based risk score and validated its prognostic and predictive significance in multiple independent cohorts using both gene expression profiles (GEP) and immunohistochemistry (IHC) measurements. Further, we devised a novel patient classification system based on TME characteristics and showed that these subtypes were associated with distinct molecular, genomic, and cytokine profiles. Critically, our proposed TME subtypes demonstrated complementary prognostic value to established molecular subtypes.
Results

Development of the TME risk score

TME plays a critical role in cancer progression and metastasis, and may be targeted to improve therapeutic response and survival. Here, we conducted a systematic investigation into the prognostic landscape and therapeutic relevance of major stromal and immune cells in the TME of gastric cancer. Details about the study design and patient cohorts can be found in the methods, supplementary materials, Fig. S1, and Table 1. In the ACRG cohort, the abundance levels of three cell types, namely, natural killer (NK) cells, endothelial cells, and fibroblasts, were significantly associated with overall survival in Cox regression analysis ($P<0.0005$, Fig. S2A). Their prognostic value was independent of pathological stage and adjuvant chemotherapy ($P<0.03$, Fig. S2B). Consistently, the same three cell types were the most important variables among TME cell types for predicting overall survival using the random survival forest algorithm (Fig. S2C). Thus, among major cellular components in the TME, NK cells, fibroblasts, and endothelial cells were identified as the most robust prognostic markers in GC.

There was a high positive correlation (Pearson’s $r = 0.73$) between the abundance of fibroblasts and endothelial cells in the TME (Fig. S2D). Given the collinearity and similar adverse effects on prognosis, we combined the endothelial cells and fibroblasts into a Stroma score by taking the square root of their product, to reflect the overall stroma status (Fig. 1A). As expected, the Stroma score was highly correlated with the endothelial cells and fibroblasts (both Pearson’s $r > 0.91$), but did not correlate with the level of NK cells (Pearson’s $r = -0.28$, Fig. S2D). We further explored the correlation of NK cell abundance and the stroma score with other cell types or established signatures. The abundance of NK cells was weakly or moderately correlated to T cell and CD8 T cell abundance, and the T cell-inflamed signature (22) (Fig. S3A and B). On the other hand, the proposed stroma score was highly correlated with the EMT score (5, 23), fibroblast signatures (24) and the StromalScore estimated from the ESTIMATE algorithm (25) in all GEP
cohorts (Fig. S3C). Bivariate analysis revealed independent, opposite prognostic effects of the NK cells (HR [95% CI]: 0.42 [0.27 – 0.65], \( P = 0.00011 \)) and Stroma score (HR [95% CI]: 1.37 [1.08 – 1.73], \( P = 0.009 \)). Based on these results, we constructed a continuous TME risk score by taking the ratio of Stroma score to NK cells, which summarizes the overall prognostic effects of the TME based on the expression of 50 marker genes (Table S1 and Fig. 1A).

**Validation of the TME risk score as an independent prognostic factor**

In multivariable Cox regression analysis, the continuous TME risk score was an independent prognostic factor for overall survival when adjusted for clinicopathologic factors including age, gender, stage, Lauren histology type, and use of adjuvant chemotherapy in the ACRG cohort (Table 2). We independently confirmed the prognostic value of the TME risk score in two additional microarray gene expression cohorts (Fig. 1B-E). In both GEP validation cohorts, the TME risk score was significantly correlated with survival (Fig. 1B) adjusting for clinicopathologic factors as well as microsatellite status (Table S2). A meta-analysis of the three GEP cohorts indicated that the TME risk score was a strong prognostic factor (HR [95% CI]: 1.42 [1.30 - 1.55], \( P < 0.0001 \), Fig. 1B). In addition, we confirmed the independent prognostic effects of NK cell abundance and Stroma score in the combined GEP validation cohorts (Fig. S4A). To stratify patients into low vs. high risk, we defined a cutoff value for the TME risk score (cutoff\(_{\text{GEP}}\) = 1.78) based on the ACRG cohort while controlling for major confounding factors, i.e. treatment and stage (Fig. S5A). In all three GEP cohorts, we observed significantly worse survival in the high TME risk group (HR ranges from 2.21 to 3.47, all \( P < 0.01 \), Fig. 1C-E).

To further validate its prognostic value, we retrospectively analyzed data in 3 IHC cohorts and assessed the association of our TME risk score with overall survival. Consistent with the results in GEP cohorts, the continuous TME risk score was significantly correlated with OS in all 3 IHC cohorts with an overall HR [95% CI] of 1.32 [1.22 – 1.42] (\( P < 0.0001 \), Fig. 1F). In addition, we confirmed the independent prognostic effects of NK cell abundance and Stroma score in the
combined IHC validation cohorts (Fig. S4B). Similarly as the gene expression analysis, we defined an optimal cutoff value for the TME risk score (cutoff$_{\text{IHC}}$=0.59, Fig S5B) based on the SMU1 cohort. Again, patients in the high TME risk group had significantly worse OS in all three IHC cohorts (HR ranges from 2.18 to 3.17; all $P<0.0005$, Fig. 1G-I). In multivariable Cox analysis, the prognostic value of the continuous TME risk score was independent of major clinicopathologic factors in the IHC cohorts (Table 2).

We investigated whether the proposed TME risk score would provide any additional prognostic information to clinical risk factors. In the combined IHC cohort, we compared the goodness-of-fit for the composite Cox regression model (TME risk score and stage) with the model only including stage by the likelihood-ratio test, which was statistically significant (Composite model: $\chi^2 = 235.4$ vs. Stage only: $\chi^2 = 187.6$, $P=4.67\times10^{-12}$), indicating the additive value of the TME risk score for prognostication. We also quantified the relative improvement in accuracy for prediction of 5-year overall survival between the two models, which showed a continuous net reclassification index of 23.3%.

Finally, we examined the prognostic significance of the TME risk score in different stage and treatment subgroups. Within every pathologic stage (I-IV), a higher TME risk score was consistently associated with worse survival in the combined GEP (Fig. S6) as well as IHC cohort (Fig. 2). In addition, the prognostic value of the TME risk score was confirmed in patients who underwent surgery alone without adjuvant chemotherapy in the ACRG and each IHC validation cohort (Fig. S7).

**Association between the TME risk score and benefit from chemotherapy**

We tested the ability of the TME risk score to predict the benefit of adjuvant chemotherapy in non-metastatic GC (stage I-III) patients by merging the ACRG and 3 IHC cohorts. Before conducting the statistical test, the patients with or without adjuvant chemotherapy were matched
regarding five clinicopathologic factors to mitigate the potential selection bias in retrospective cohorts. As shown in Fig. 3A, patients with a high TME risk score derived a significant survival benefit from adjuvant chemotherapy (HR [95% CI]: 0.53 [0.41 – 0.69], P < 0.0001). On the other hand, patients at a low TME risk score did not benefit from adjuvant chemotherapy (HR [95% CI]: 1.11 [0.65 – 1.91], P=0.705, Fig. 3B). A formal interaction test between the TME risk group and treatment was statistically significant (P = 0.0148), indicating a predictive effect for the benefit from adjuvant chemotherapy. Similar results were observed using all the patients without matching (Fig. S8).

**Identification of TME subtypes**

To elucidate the molecular and genomic characteristics related to the prognostic cell types in TME, we divided patients into four TME subtypes based on the abundance level of NK cells and Stroma score (Fig. 4A). The TME subtypes were associated with distinct prognoses, with a similar pattern between the combined GEP cohort and combined IHC cohort (both P<0.0001, Fig. 4B and C). Specifically, patients in the NK high & Stroma low subgroup had the best prognosis; patients in the NK low & Stroma high subgroup had the worst prognosis; the remaining two subtypes were associated with intermediate prognoses. Some representative examples of IHC images for 4 TME subtypes are shown in Fig. 4D.

**Complementary value of TME subtypes to ACRG molecular subtypes**

We compared the proposed TME subtypes with the established ACRG subtypes for GC, which were defined by cancer cell-intrinsic molecular features. While the ACRG MSI and MSS/EMT subtypes were enriched with patients predicted to be NK high and Stroma high respectively, the distribution of the TME subtypes within the ACRG MSS/non-EMT subtypes was relatively balanced (Fig. 5A). Importantly, we found that the TME subtypes provided complementary prognostic value to the ACRG subtypes (Fig. 5B-D). Within the ACRG MSI
subtype, which is known to have favorable outcomes, our TME subtypes can still distinguish subgroups of patients with distinct prognoses ($P<0.0001$, Fig. 5B). Similarly, within the ACRG MSS/non-EMT subtypes, patients were divided into four different prognostic groups by the TME subtypes ($P=0.0024$, Fig. 5C).

**Complementary value of TME subtypes to intrinsic subtypes for gastric cancer**

We found that two TME subtypes with a high stroma score were enriched for the G-DIF intrinsic molecular subtype, and vice versa (Fig. S9A). This relation is expected given our current understanding of gastric cancer. Within the G-INT subtype, the prognostic effect of TME subtypes was well preserved (Fig. S9B). For the G-DIF subtype, the NK high & Stroma low subtype confers a better prognosis compared with others (Fig. S9C). Again, these results showed that the proposed TME-based subtypes provide additional information to tumor intrinsic classification systems.

**Relation between the TME subtypes and the TCGA genomic subtypes**

We compared our TME subtypes with the five genomic subtypes defined by TCGA research group (Fig. 6). Similar to the comparison with ACRG subtypes, the TCGA MSI subtype was enriched with patients predicted to be NK high (67%), and the NK low & Stroma high subtype only accounted for 6% of the MSI subtype. The majority (90%) of patients in the TCGA GS subtype had a Stroma high phenotype (58% NK high, 32% NK low), while the TCGA EBV subtype mainly (80%) consisted of patients with an NK high phenotype (47% Stroma high, 33% Stroma low). On the other hand, the distribution of the TME subtypes was relatively balanced within the dominant TCGA subtype, CIN, with a slight bias toward patients with NK low relative to those with NK high (62% vs. 38%). Therefore, the proposed TME subtypes and TCGA genomic subtypes were related but largely non-redundant classification systems.

**Genetic alterations associated with TME subtypes**
We evaluated the genetic characteristics of the TME subtypes by leveraging the comprehensive genomic data available in the TCGA STAD cohort. This evaluation was done separately for the two key features (NK cell abundance and stroma score) because they were largely independent of each other (Fig. 4A). The features only correlated with NK cell status were presented in Fig. S10. The most striking observation was that tumors with an NK low status are more likely to have an instable genome and a higher level of aneuploidy (Fig. S10A). What’s more, these tumors demonstrated a higher clonal deletion score, frequency of CIN focal and increased arm-level copy number variations (Fig. S10B-D). The overall ploidy and whole genome doubling events were also significantly associated with low NK level (Fig. S10E and F). In addition, we observed that the low NK status was associated with increased rates of HER2 amplification and TP53 mutation (Fig. S10G and H). On the other hand, tumors with a high NK status were characterized by PIK3CA mutation (Fig. S10I) and elevated rates of epigenetic silencing of CDKN2A (Fig. S10J). What’s more, we have reassessed the distribution of PIK3CA and TP53 mutations across TME subtypes by focusing on the subgroups of patients as defined by either MSI or T-cell inflammation status. We found that the differential mutation status for both genes still hold in the MSS group as well as T-cell inflammation high and low groups (Fig. S11). The only exception is the MSI-H group, which may partly be due to a limited number of MSI patients (n = 96). Three genomic features were correlated with varying stroma status, i.e. SNV density, Indel density and mutation density, the higher level of which were all significantly correlated with stroma low status (Fig. S12).

Four genomic features, i.e. percent of hypermutation, the number of focal CNV events, focal CNV amplification and focal CNV deletion events, were significantly correlated with both NK cell and stroma status (Fig. S13). Specifically, higher level of NK infiltration or lower level of stroma status associated with higher rates of hypermutation (Fig. S13A). Increased focal CNV events (both deletion and amplification) were observed enriched in patients with low NK or low stroma...
status (Fig. S13B-D). Of note, we observed similar patterns when specifically focusing on the CIN subgroup, which indicated the above observation was not confounded by the MSI or EBV status (Fig. S14).

**Gene expression, molecular pathways, and cytokines correlated with Stroma and NK cell infiltration**

Meta-analysis of four GEP cohorts identified robust correlation between NK cell abundance and expression of specific genes and molecular pathway activity. At the individual gene level, we found that NKG2A, an NK cell inhibitory receptor, had the highest correlation with the amount of NK cell infiltration in the TME (meta Pearson’s $r = 0.73$, Table S3). Several other NK cell activating or inhibitory receptors were also highly correlated with NK cell infiltration, including KIR3DL2, NKG2C/E, CD94, and CD244, along with genes related to immune-mediated cytolytic activity such as FASLG, PRF1, and GZMA/B (Table S3).

At the pathway level, we found that the interferon-γ signaling pathway was correlated with an increased level of NK cell infiltration, which is consistent with a preexisting antitumor immune response. By contrast, the hedgehog and Wnt/β catenin signaling pathways were correlated with decreased NK cell infiltration (Fig. S15, Table S4). On the other hand, the TME Stroma score was positively correlated with the EMT and angiogenesis pathways, while it was negatively correlated with cell cycle and proliferation such as MYC and E2F pathways (Fig. S16, Table S4).

Since cytokines are key regulators of cellular migration and composition in the TME, we next identified differentially expressed cytokines related to the NK and Stroma status. A total of 17 cytokines were differentially expressed (all upregulated) in the NK high group, including IL1, IFNγ and FasL, which is consistent with an active inflammatory response and cytolytic activity (Fig. S17A). The chemokines CXCL9, -10, -11 were also upregulated in the NK high group, which mediate immune cell migration, differentiation, and activation (26). On the other hand, cytokines
upregulated in the stroma high group were mostly related to an immunosuppressive function such as IL6, TGFB3, and BMP6 in the TGFβ signaling pathway (Fig. S17B).

In summary, these TME subtypes demonstrated distinct genetic and molecular patterns including aneuploidy, somatic mutation load, interferon-γ, hedgehog, Wnt/β catenin signaling pathway activation, and distinct cytokine profiles.
Discussion

In this study, we developed a tumor microenvironment (TME)-based risk score by integrating immune and stromal signatures, and validated it as an independent prognostic factor in multiple gene expression and IHC cohorts of 1,678 gastric cancer patients. The TME risk score provided additional information beyond current staging system for improved risk stratification, and could potentially guide personalized management of resected gastric cancer regarding adjuvant chemotherapy. Further, we proposed four TME subtypes as defined by the NK cell abundance and Stroma score, which were associated with distinct molecular, genomic, and cytokine profiles. The new TME subtypes reflect differing aspects of the tumor biology and complement established genomic subtypes, which may be used to refine molecular classification of gastric cancer.

Our work adds to a growing body of literature supporting the critical role of TME in cancer progression and its therapeutic relevance (13, 27, 28). Specifically, we demonstrate that Stroma score and NK cells are independent prognostic factors in gastric cancer. Stromal gene signatures have been consistently shown to correlate with a poor prognosis in gastrointestinal cancers (29, 30) including gastric cancer (31). Cancer-associated fibroblasts (CAFs) are one of the most abundant stromal cells in TME and are known to be a driver of cancer progression and therapeutic resistance (32). Cheong et al. recently developed a predictive biomarker for chemotherapy response in gastric cancer based on 4 genes related to immune, stem-like and epithelial signatures (13). Tumor angiogenesis is induced by various growth factors such as the vascular endothelial growth factor (VEGF) in the TME, and significantly contributes to tumor progression and metastasis (18, 33). One recent study (34) investigated stromal gene expression signatures of gastric cancer and confirmed the prognostic value of the vascular signature driven by angiogenesis. Our composite Stroma score combines both fibroblasts and endothelial cells, which may serve as a more robust prognostic biomarker.
One novel finding of our study is the discovery of NK cells as an orthogonal prognostic factor with an opposite effect compared with tumor stroma. NK cells are innate lymphoid cells widely known for their ability to exert robust cytotoxic function against viral infection, and play a prominent role in the control of cancer metastasis (35). However, the relevance of NK cells in the immunosurveillance of primary solid tumors remains controversial. Our data show that higher abundance of NK cells in the TME consistently confers a favorable prognosis independent of the Stroma score across multiple cohorts of gastric cancer patients. In addition, we found that tumors with an NK low phenotype had an increased level of aneuploidy, which is consistent with the fact that aneuploidy correlates with immune evasion and is a marker of aggressive disease and unfavorable outcome (36). Beyond genomic factors, we showed that hedgehog and Wnt/β catenin signaling pathway activation was correlated with reduced NK cell infiltration in gastric cancer, also consistent with recent reports that correlated these pathways to immune evasion across human cancers (37, 38).

There is a growing interest in developing therapeutic strategies to increase the infiltration and improve the function of NK cells in the TME (39), which may synergize with current immunotherapies primarily targeting T cells. It has recently been shown that NK cells are a cytolytic effector against PD-L1 negative tumors treated with anti-PD-L1 antibody (40), and NK cells stimulate the recruitment of type I conventional dendritic cells and CD8+ T cells into the tumor (41). Further, an increased NK cell infiltration in the TME is associated with an improved response to immune checkpoint blockade (42). In our study, we found several inhibitory receptors, including NKG2A/C/E, KIR3DL2, and CD94, were highly expressed on NK cells in gastric cancer TME (Table S3). As the activity of tumor-infiltrating NK cells is strongly suppressed, these inhibitory molecules may be targeted to restore NK cell function (43). Indeed, two recent studies demonstrated that NKG2A blockade in combination with a tumor-specific vaccine or antibody improved immunotherapy response and outcomes (44, 45).
In our study, we did not find the amount of CD8+ T lymphocytes represents a robust prognostic factor, which is in line with previous reports of inconsistent results (46, 47). This could be due to a lower tumor mutational burden in gastric cancer compared with other more immunogenic tumors, leading to an immune-cold or immune-excluded phenotype (48). Alternatively, potent factors in the TME such as TGFβ could also suppress immune function, leading to T cell exhaustion (49).

Strengths of our study include independent validation of the results in multiple patient cohorts, and the use of two different methodologies including gene expression profiles and clinically applicable IHC assays in FFPE tissue samples, which can reduce potential biases and enhance reproducibility of the findings. In future, gene expression assays of our TME risk score using qPCR or nCounter systems may be developed to facilitate its practical use. Our study is mainly limited by its retrospective nature and use of patient cohorts with non-randomized treatment, which makes it challenging to assess the predictive value in a therapeutic setting. The TME risk score only included the major immune and stromal cell types. Incorporating specific phenotypic or functional subsets (such as M2 macrophages) might improve the prognostic value albeit with an increased complexity. Finally, our study included patients primarily of the Asian origin who received adjuvant chemotherapy. The generalizability of our findings in the neoadjuvant setting among non-Asian populations (50) should be tested in future studies.

In summary, we developed and validated a TME-based risk score as an independent prognostic factor in gastric cancer and proposed novel TME subtypes with distinct molecular, genomic, and cytokine profiles. Our findings provide new insight into the prognostic landscape of TME in gastric cancer, which warrants further validation in prospective studies.
Methods

Patients and data

We retrospectively analyzed data for patients who were diagnosed with primary gastric cancer and treated with surgical resection. For discovery purposes, we used the ACRG cohort of 300 patients for whom microarray GEP as well as detailed patient-level clinical and treatment information is publicly available (accession number: GSE62254) (5). For validation purposes, we used two additional microarray GEP cohorts measuring fresh-frozen tissue of the largest sample sizes (GSE84437 (13): n=433 and GSE15459 (51): n=192) with survival. For further validation, we used three independent IHC cohorts (denoted as SMU1, SMU2, and SYSU) of 753 patients who were consecutively treated between 2005 and 2009 at two medical centers. Finally, we used TCGA stomach adenocarcinoma (STAD) data to evaluate the genetic and molecular characteristics of our TME subtypes.

Discovery and validation of a TME-based risk score

Given the bulk gene expression data, we computed the absolute abundance levels of the major cell types in the TME, including 8 immune and 2 stroma cell types (fibroblasts and endothelial cells), by averaging the expression of a set of carefully selected marker genes (Table S1) provided by the Microenvironment Cell Populations-counter (MCP-counter) algorithm (52). The main reason to use the MCP-counter algorithm is that it provides an estimation of the absolute abundance of cellular components in the TME, and therefore meaningful comparisons can be made across different samples. By contrast, several other algorithms such as ESTIMATE only provide the fraction of stromal and immune cells as a whole. On the other hand, CIBERSORT generates the relative fractions of 22 immune cells. However, these fractions are normalized within a sample and thus not comparable across samples.
We independently assessed the prognostic effects of these cell types using two different approaches (Cox regression analysis and random survival forest model) in the ACRG cohort. Those cell types that demonstrated a consistent prognostic significance in both models were selected for subsequent analyses. After identifying cell types with the most robust prognostic value, we integrated them into a composite TME risk score, by taking consideration of their pairwise correlation and impact on survival. The prognostic significance of the TME risk score was independently tested in two additional GEP cohorts and three IHC cohorts. The IHC-based TME risk score was derived using the same formula, where semi-quantitative evaluation of stains for established cell markers was performed. Details about these analyses can be found in the supplementary materials.

**TME-based subtypes and molecular characterization**

Based on the cellular composition of the TME, we stratified patients into 4 distinct groups, i.e., TME subtypes. We compared our TME subtypes with the existing ACRG molecular subtypes (5), the intrinsic subtypes for gastric cancer (8) as well as TCGA genomic subtypes (53), in terms of patient assignment and prognostic stratification. We evaluated the genetic and molecular characteristics of our TME subtypes, including genetic or epigenetic alterations of specific driver genes, genome/chromosome instability, and mutational burden. In addition, we reassessed the differential patterns of certain genomic features across TME subtypes by focusing on subgroups of patients defined by MSI, T-cell inflamed or CIN status. Finally, we investigated gene expression, cancer hallmark pathway activity, and soluble factors correlated with infiltration of prognostic relevant cell types in the TME. Details are presented in the supplementary materials.

**Statistics**

The Cox regression model was used to assess the prognostic effect of continuous variables. A fixed-effect model was used to summarize the prognostic effect of the TME risk score
in the GEP and IHC cohorts. We performed multivariable Cox regression analyses to assess the independent prognostic value of the TME risk score by adjusting for clinicopathologic factors. The log-rank test was used to evaluate the survival difference among different patient groups. We used the likelihood-ratio test and net reclassification index (54) to assess the additive prognostic value of the TME risk score to pathological stage.

We used a multivariable Cox regression model with the main effects (TME risk group and chemotherapy) and the interaction effect (TME risk group × chemotherapy) to test the ability of the TME risk score to predict the chemotherapy benefit. To minimize selection bias and confounding effects, we used a matching strategy to balance patients within the GEP (ACRG) and IHC cohorts, respectively. We performed exact 1:1 matching of non-metastatic patients (stage I-III) who received chemotherapy versus those who did not according to five clinicopathologic factors including stage, age (> 50 years), gender, Lauren histology type, and tumor location. We also tested the predictive effect of the TME risk group in the merged unmatched cohorts.

Chi-squared and Mann-Whitney tests were used to assess the difference among TME subgroups regarding categorical and continuous genomic features, respectively. Fisher's z transformation of correlation was used to assess the overall correlation of single gene and pathway activity with NK or Stroma score in different cohorts. Differentially expressed cytokine genes were identified via limma (55). The Benjamini-Hochberg method was used to compute the false discovery rate (FDR) to adjust for multiple testing. More details are presented in the supplementary materials. All statistical analyses were performed in R version 3.5.3.

**Study approval**

This study was approved by the institutional review board (IRB) and conducted in accordance with ethical guidelines such as the Declaration of Helsinki. The informed consent for
patients in the GEP cohort was waived given the use of existing, de-identified public datasets.

Written-informed consent was obtained from all the enrolled patients in the IHC cohort.

**Author contributions**

BL, YJ, GF and RL designed the study. BL, YJ, and GL conducted the experiments and gathered the data. GL, GF and RL supervised the design, analysis, and study interpretation. BL, JY and RL wrote the manuscript.
References:


Figure 1. Prognostic significance of the TME risk score in the GEP and IHC cohorts. (A) The formula to define the TME risk score. The abundance level of each cell type is calculated by taking the average expression of preselected marker genes listed in table S2. (B) Increased TME risk score was significantly correlated with inferior overall survival in all three GEP cohorts (ACRG: n = 299, GSE15459: n = 192, and GSE84437: n = 433). A fixed-effect model indicated a strong overall prognostic effect of the TME risk score. Cox regression was used to measure the prognostic effects of TME risk score. (C-E) High TME risk group was associated with worse overall survival in these cohorts (ACRG: n = 299, GSE15459: n = 192, and GSE84437: n = 433). The GEP cutoff value for TME risk score was defined by optimizing the Cox regression P value in the ACRG cohort. (F-I) Same as above except for three IHC cohorts (SMU1: n = 247, SMU2: n = 234, and SYSU: n = 272). The IHC cutoff value was defined by optimizing the Cox regression P value in the SMU1 cohort. Hazard ratios (HR) and confidence intervals (CI) were estimated by Cox regression. P values were generated by log-rank test.
Figure 2. The prognostic effects of the TME risk score in patients within each pathological stage in the combined IHC cohorts. A high TME risk score was consistently associated with worse overall survival in patients with stage I (n = 113, A), stage II (n = 141, B), stage III (n = 401, C), and stage IV (n = 98, D) disease. The cutoff value for TME risk score was the same as in Fig. 1. Hazard ratios (HR) and confidence intervals (CI) were estimated by Cox regression. P values were generated by log-rank test.
Figure 3. Predictive relevance of the TME risk score for the benefit of chemotherapy in stage I-III gastric cancer. (A) Patients with a high TME risk score (n = 419) derived a significant survival benefit from adjuvant chemotherapy at 5 years. However, patients with a low TME risk score (n = 175) did not benefit from adjuvant chemotherapy (B). The patients treated with or without chemotherapy were matched according to five clinicopathologic factors. Hazard ratios (HR) and confidence intervals (CI) were estimated by Cox regression. P values were generated by log-rank test. The P value for the interaction between the TME risk group and adjuvant chemotherapy was 0.0148.
Figure 4. The definition of TME subtypes and their prognostic significance. (A) Patients were divided into 4 TME subtypes, based on the distribution of NK cell abundance and Stroma scores in the merged GEP cohorts (n = 1,340). In the merged GEP cohorts (except TCGA) (n = 925, B) and IHC cohorts (n = 753, C), the NK high & Stroma low and NK low & Stroma high groups correspond to the best and the worst prognosis, respectively. The NK high & Stroma high and NK low & Stroma low subtypes were associated with an intermediate prognosis. (D) Representative examples of IHC images for 4 TME subtypes. CD57, CD34, and αSMA are stains for NK cells, endothelial cells, and fibroblasts. P values were generated by log-rank test.
Figure 5. Complementary prognostic value of the TME subtypes to the ACRG subtypes. (A) The correspondence between patients classified according to the TME subtypes and ACRG subtypes in the merged GEP cohorts (n = 1,340). (B-C) The TME subtypes can further stratify patients within the ACRG MSI (n = 222) and MSS/TP53+/− (n = 562) subtypes subtype into groups with distinct prognoses. The survival difference within the ACRG MSS/EMT (n = 141) subgroup showed a trend due to a smaller number of patients (D). P values were generated by log-rank test.
Figure 6. The correspondence between patients classified according to the TCGA subtypes and the TME subtypes in the TCGA cohort. Genomic features which were significantly enriched in certain TME subtypes in the TCGA STAD cohort (n = 415) were also presented.
Table 1. Clinicopathologic and treatment information for patients in the GEP and IHC cohorts

<table>
<thead>
<tr>
<th></th>
<th>GEP cohorts</th>
<th></th>
<th>TCGA STAD</th>
<th></th>
<th>IHC cohorts</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACRG</td>
<td>GSE15459</td>
<td>GSE84437</td>
<td>TCGA STAD</td>
<td>SMU1</td>
<td>SMU2</td>
<td>SYSU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients</td>
<td>300</td>
<td>192</td>
<td>433</td>
<td>415</td>
<td>247</td>
<td>234</td>
<td>272</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median age (range)</td>
<td>64 (24-86)</td>
<td>67 (23-92)</td>
<td>62 (27-86)</td>
<td>67 (30-90)</td>
<td>57 (23-90)</td>
<td>55 (21-77)</td>
<td>58 (24-82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>199 (66)</td>
<td>125 (65)</td>
<td>296 (68)</td>
<td>268 (65)</td>
<td>177 (72)</td>
<td>143 (61)</td>
<td>175 (64)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>30 (10)</td>
<td>31 (16)</td>
<td>-</td>
<td>57 (15)</td>
<td>31 (12)</td>
<td>37 (16)</td>
<td>45 (17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>96 (32)</td>
<td>29 (15)</td>
<td>-</td>
<td>123 (32)</td>
<td>56 (23)</td>
<td>35 (15)</td>
<td>50 (18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>95 (32)</td>
<td>72 (38)</td>
<td>-</td>
<td>169 (43)</td>
<td>135 (55)</td>
<td>129 (55)</td>
<td>137 (50)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>77 (26)</td>
<td>60 (31)</td>
<td>-</td>
<td>41 (11)</td>
<td>25 (10)</td>
<td>33 (14)</td>
<td>40 (15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>0</td>
<td></td>
<td>11 (3)</td>
<td>22 (5)</td>
<td>21 (9)</td>
<td>32 (14)</td>
<td>34 (13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>186 (62)</td>
<td></td>
<td>38 (9)</td>
<td>88 (22)</td>
<td>36 (15)</td>
<td>19 (8)</td>
<td>22 (8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>91 (31)</td>
<td></td>
<td>92 (21)</td>
<td>181 (45)</td>
<td>6 (2)</td>
<td>31 (13)</td>
<td>8 (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>21 (7)</td>
<td></td>
<td>292 (67)</td>
<td>115 (28)</td>
<td>184 (74)</td>
<td>152 (65)</td>
<td>208 (76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>38 (13)</td>
<td></td>
<td>80 (18)</td>
<td>123 (31)</td>
<td>73 (30)</td>
<td>63 (27)</td>
<td>91 (33)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>131 (44)</td>
<td></td>
<td>188 (43)</td>
<td>112 (28)</td>
<td>40 (16)</td>
<td>53 (23)</td>
<td>110 (40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>80 (27)</td>
<td></td>
<td>132 (30)</td>
<td>79 (20)</td>
<td>82 (33)</td>
<td>62 (26)</td>
<td>45 (17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>51 (17)</td>
<td></td>
<td>33 (8)</td>
<td>82 (21)</td>
<td>52 (21)</td>
<td>56 (24)</td>
<td>26 (10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>273 (91)</td>
<td></td>
<td>-</td>
<td>367 (93)</td>
<td>222 (90)</td>
<td>201 (86)</td>
<td>232 (85)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>27 (9)</td>
<td></td>
<td>-</td>
<td>27 (7)</td>
<td>25 (10)</td>
<td>33 (14)</td>
<td>40 (15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antrum</td>
<td>155 (52)</td>
<td></td>
<td>-</td>
<td>155 (39)</td>
<td>124 (50)</td>
<td>116 (50)</td>
<td>103 (38)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body</td>
<td>107 (36)</td>
<td></td>
<td>-</td>
<td>143 (36)</td>
<td>40 (16)</td>
<td>45 (19)</td>
<td>80 (29)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardia</td>
<td>32 (11)</td>
<td></td>
<td>-</td>
<td>97 (24)</td>
<td>49 (20)</td>
<td>40 (17)</td>
<td>76 (28)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole</td>
<td>6 (2)</td>
<td></td>
<td>-</td>
<td>2 (1)</td>
<td>34 (14)</td>
<td>33 (14)</td>
<td>13 (5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lauren (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intestinal</td>
<td>146 (49)</td>
<td>99 (52)</td>
<td>-</td>
<td>176 (72)</td>
<td>199 (81)</td>
<td>158 (68)</td>
<td>100 (37)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diffuse/Mixed</td>
<td>151 (51)</td>
<td>93 (48)</td>
<td>-</td>
<td>69 (28)</td>
<td>48 (19)</td>
<td>76 (32)</td>
<td>172 (63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy (%)</td>
<td>144 (48)</td>
<td></td>
<td>-</td>
<td>-</td>
<td>104 (42)</td>
<td>148 (63)</td>
<td>126 (46)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median follow-up (mo)</td>
<td>80</td>
<td>74</td>
<td>116</td>
<td>22</td>
<td>83</td>
<td>77</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* denotes data is not available

*: The version of TNM staging system for ACRG, GSE15459 and GSE84437 was AJCC 6th edition.

The version of TNM staging system for IHC cohorts was AJCC 8th edition. The version of TNM staging system for TCGA cohort was 6th or 7th edition.
Table 2. Multivariable Cox regression analysis of overall survival using the TME risk score and clinicopathologic factors

<table>
<thead>
<tr>
<th>Variable</th>
<th>ACRG cohort</th>
<th>Combined IHC cohort</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P value</td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td>TME risk score</td>
<td>1.42 (1.22 – 1.66)</td>
<td>7.6×10⁻⁶ ***</td>
<td>1.34 (1.24 – 1.45)</td>
</tr>
<tr>
<td>Age</td>
<td>1.03 (1.01 – 1.04)</td>
<td>0.0022 **</td>
<td>1.00 (0.99 – 1.01)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Male</td>
<td>1.26 (0.89 - 1.79)</td>
<td>0.2</td>
<td>0.92 (0.75 – 1.13)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>II</td>
<td>1.56 (0.60 – 4.07)</td>
<td>0.36</td>
<td>1.96 (1.24 – 3.10)</td>
</tr>
<tr>
<td>III</td>
<td>3.18 (1.24 – 8.17)</td>
<td>0.016 *</td>
<td>4.54 (3.04 – 6.76)</td>
</tr>
<tr>
<td>IV</td>
<td>7.48 (2.96 – 18.94)</td>
<td>2.2×10⁻⁵ ***</td>
<td>14.09 (9.05 – 21.95)</td>
</tr>
<tr>
<td>Lauren classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diffuse/Mixed</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Intestinal</td>
<td>0.69 (0.48 – 0.98)</td>
<td>0.039 *</td>
<td>0.98 (0.80 – 1.18)</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Yes</td>
<td>0.49 (0.34 – 0.71)</td>
<td>0.00012 ***</td>
<td>0.61 (0.50 – 0.73)</td>
</tr>
</tbody>
</table>

The TME risk score and age were continuous variables.

Stage, chemotherapy, Lauren classification, and gender were categorical variables.

* P < 0.05; ** P < 0.01; *** P < 0.001