

Deep learning on image-omics data in identifying prognostic immune biomarkers for ovarian cancer

Ying Zhu

Houston Methodist Research Institute

Sammy Ferri-Borgogno

The University of Texas MD Anderson Cancer Center <https://orcid.org/0000-0002-8270-8963>

Jianting Sheng

Houston Methodist Research Institute,

Tsz-Lun Yeung

University of Texas MD Anderson Cancer Center

Jared Burks

The University of Texas MD Anderson Cancer Center <https://orcid.org/0000-0002-6173-9074>

Paola Cappello

University of Turin <https://orcid.org/0000-0002-5321-7794>

Amir Jazaeri

The University of Texas MD Anderson Cancer Center

Jae-Hoon Kim

Gangnam Severance Hospital, Yonsei University College of Medicine,

Gwan Hee Han

Yonsei University College of Medicine

Michael Birrer

the University of Arkansas for Medical Sciences

Samuel Mok

The University of Texas MD Anderson Cancer Center <https://orcid.org/0000-0001-7013-1805>

Stephen Wong (✉ st Wong@houstonmethodist.org)

Houston Methodist <https://orcid.org/0000-0001-9188-6502>

Article

Keywords: Deep learning, immune biomarkers, ovarian cancer, prognostic features

Posted Date: September 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-67036/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Although stromal and immune cells in the tumor microenvironment have been shown to directly affect tumor growth and chemoresistance, how the interactions among stromal and immune cells and their spatially resolved cell heterogeneity would influence ovarian patients' survival remains largely unknown. To fill this gap, we developed a new imageomics method that (i) incorporates an artificial intelligence–based analytics pipeline for imaging mass cytometry (IMC) to increase the accuracy of cell segmentation and spatial information extraction in order to identify immune biomarkers and their interactions that can predict overall survival rates in patients with treatment-naïve ovarian cancer, and (ii) integrates quantitated spatial IMC image data with microdissected tumor and stromal transcriptomic data from the same patients as well as single-cell RNA sequencing data to detect genes correlated with the prognostic features and postulate novel mechanisms by which these genes contribute to the prognostic features.

Introduction

Advanced high-grade serous ovarian cancer (HGSC) is the most lethal gynecologic malignancy, causing more than 13,000 deaths annually in the United States¹. HGSC is notable for initial sensitivity (75% response rate) to platinum and taxane neoadjuvant chemotherapy or chemotherapy following debulking surgery^{2,3}. However, most tumors (> 75–80%) recur within 12 to 24 months after treatment, and many patients die of progressively chemotherapy-resistant disease^{4–6}. A critically important component that influences the patient survival is the tumor microenvironment^{7,8}, which is primarily composed of fibroblasts, extracellular matrix proteins, endothelial cells, lymphocytic infiltrates, and cancer cells. The tumor microenvironment has been shown to directly affect cancer cell growth, migration, invasion, chemoresistance, cell-cell interactions, and matrix remodeling^{9,10}. However, spatially resolved, single-cell analysis that identifies tumor and stromal cell phenotypes and characterizes their organization and heterogeneity, as well as biomarkers at a single-cell level for predicting survival of HGSC patients, are lacking.

Imaging mass cytometry (IMC) is an imaging-based mass cytometry (CyTOF) that couples immunohistochemical and immunocytochemical methods with high-resolution laser ablation¹¹ to allow the imaging of more than 30 proteins and protein modifications simultaneously at subcellular resolution. This enables researchers to uncover the heterogeneity of cellular phenotypes and cell-cell interactions (phenograph¹²; histocat¹³). Cell segmentation is the first key step of IMC analysis. However, IMC images of biological tissues, particularly those of solid tumors, are extremely challenging for conventional cell segmentation methods, such as the watershed algorithm¹¹ and pixel-based classification¹³, owing to great variations in image intensities and cell shapes, overlapping cells, dense cell clusters, blurred edge information, missing object borders, and low signal-to-noise ratios. In addition, the highly multiplexed IMC data generate rich information of cell phenotypes, spatial organization, and heterogeneity. The way to quantify and integrate of various data to identify reliable prognostic biomarkers remains largely unexplored.

In the current study, we built a novel analytic and modeling pipeline of IMC images using deep learning and applied it to predict patient survival rates using IMC data generated from patient samples of treatment-naïve HGSC tumor tissues. We compared cell density and nearest-neighbor interactions in tumor-enriched regions between long-term and short-term survivors. We also used the cell density or nearest-neighbor interactions as features to build a machine learning model to identify salient prognostic features for predicting patient survival. Finally, we combined the quantitated IMC images with cancer and stromal gene expression data of microdissected tissue specimens from the same HGSC patients to detect genes that are significantly correlated with prognostic features and postulate the mechanisms by which these genes contribute to the prognostic features.

Results

Image analysis pipeline

To comprehensively quantify the cellular heterogeneity and spatial organization of HGSC tissue and find biomarkers that predict patient survival, we used IMC to detect 34 different proteins in 41 tumor samples from treatment-naïve HGSC patients (Table S1). Tissue sections were stained with a panel of metal-tagged antibodies followed by laser ablation coupled to mass spectrometry to generate high-dimensional images as previously described¹¹ (Fig. 1). Among the 34 analyzed proteins, we selected 21 markers with relatively high signal-to-noise ratios for further analysis (Fig. S1A). Our selected panel consisted of markers of proliferation; immune cell regulators, and markers of epithelial, stroma, immune and endothelial lineages (Fig. S2). Resulting data were then analyzed using a novel IMC-adapted image analysis pipeline. Briefly, cell segmentation was performed using a deep learning method, followed by three rounds of clustering to identify and annotate different cell subtypes used for cell density and neighborhood analyses. These two features were then used for survival prediction analysis and correlating IMC phenotype with gene expression profile (Fig. 1).

Cell segmentation and annotation by deep learning-based IMC data analysis

We employed a deep learning method, mask-R-CNN (MRCNN)¹⁴ for IMC cell segmentation based on the computerized outputs of watershed segmentation (Fig. 2A). The mean dice similarity of MRCNN was significantly higher than that of watershed (mean of mean dice similarity of MRCNN = 0.59, mean of mean dice similarity of watershed = 0.56, $p = 0.0023$, $n = 10$ samples, paired t test; Fig. S1B), and the standard deviation of dice similarity of MRCNN was significantly lower than that of watershed (mean standard deviation of dice similarity of MRCNN = 0.20, mean standard deviation of dice similarity of watershed = 0.24, $p = 1.33 \times 10^{-5}$, $n = 10$ samples, paired t test; Fig. S1C). These results suggest that the segmentation by MRCNN is more similar to the manual segmentation than the watershed segmentation and that it generates less oversegmentation than watershed (Fig. 2A).

After three iterations of phenograph clustering to identify cell subtypes, we identified 40 cell subtypes of T and B cell, macrophage, endothelial, and fibroblastic cell populations, as well as tumor cell subtypes from 162,869 cells in 41 images (Fig. 2B) and quantified the normalized expression of all markers across various cell subtypes (Fig. 2C). Specifically, based on the normalized expression of cell subtype-specific markers, we identified nine macrophage/monocyte subtypes (Fig. 3A), four CD8⁺ T cell subtypes and 6 CD4⁺ T cell subtypes (Fig. 3B), nine tumor subtypes (Fig. 3C), and many other fibroblastic and immune cell subtypes (Fig. 3D). Because the normalization step converts the expression of all cells in a sample to a z-score, the normalized expression reflects the relative expression across all cells. For example, although tu_1, tu_2, tu_3, tu_5, tu_6, tu_8, and tu_9 have median normalized Keratin 8/18 expression around zero (Fig. 3C), they exhibit significantly higher Keratin 8_18 expression levels than non-tumor cells (Fig. S3). Based on the results of Fig. 2C and Fig. 3, the phenotypes of all cell subtypes are summarized in Table S2.

Spatially resolved cell density and cell-cell nearest-neighbor interactions analyses of the ovarian tumor microenvironment

We automatically calculated the tumor-enriched region for each image (Methods). By computing the cell density as the cell count in the tumor-enriched region per total tumor cells, we found several cell subtypes exhibiting significant differences between long-term survivors (LTS; overall survival ≥ 60 months, $n = 21$) and short-term survivors (STS; overall survival ≤ 20 months, $n = 20$; Fig. 4A). Among different T cell subtypes, granzyme B⁺ CD8⁺ cytotoxic T cell (CD8_4) density was significantly higher in LTS ($p = 0.019$, Fig. 4A) and CD45RO⁺ CD44⁺ CD8⁺ memory T cell (CD8_3) density had a trend of declining in LTS ($p = 0.084$, Fig. S4) than in STS. In addition, CD45RO⁺ CD4⁺ memory T cell (CD4_4) density was significantly higher in LTS than in STS ($p = 0.024$, Fig. 4A). Among different CD73⁺ cell subtypes, CD73⁺ cell (CD73_1) density and CD73^{mid} cell (CD73_2) density were significantly lower in the LTS than in STS ($p = 0.015$ and $p = 0.002$, respectively; Fig. 4A). CD31⁺ CD73^{mid} endothelial cell (CD31) density was significantly lower in LTS than in STS ($p = 0.007$, Fig. 4A). Among tumor cell subtypes, B7H4⁺ Keratin⁺ tumor cell (tu_9) density was significantly lower in LTS than in STS ($p = 0.018$, Fig. 4A). A comparison of cell densities of all cell subtypes between LTS and STS is shown in Fig. S4.

To determine whether the appearance of one cell subtype is associated with the appearance of another cell subtype, a Spearman correlation matrix between various cell subtype densities was generated. The results demonstrated that the granzyme B⁺ CD8⁺ cytotoxic T cell (CD8_4) density was negatively correlated with cell densities of B7H4⁺ tumor cells (tu_7; $r = -0.47$, $p = 0.0006$) and tu_9 ($r = -0.38$, $p = 0.006$; Fig. 4B), suggesting that these CD8⁺ cytotoxic cells are infiltrating the tumor mass and actively depleting B7H4⁺ ovarian cancer cells in LTS. CD163⁺ CD68⁺ CD14⁺ macrophage (ma_3) density was positively correlated with CD45RO⁺ CD44⁺ CD4⁺ memory T cell (CD4_1) density ($r = 0.56$, $p = 2.8 \times 10^{-5}$), CD45RO^{mid} CD44⁺ CD4⁺ memory T cell (CD4_3) density ($r = 0.56$, $p = 2.4 \times 10^{-5}$), and CD45RO^{mid} CD44^{mid}

CD4⁺ memory T cell (CD4_5) density ($r = 0.45$, $p = 0.001$) (Fig. 4B). CD163, as an M2 marker, and CD68 as a pan-macrophage or M1 marker, have been widely used to classify macrophages^{15,16}. In addition, previous studies demonstrated that tissue-associated macrophages are very diverse and heterogeneous, are subjected to microenvironmental factors and do not have restricted M1 (CD68⁺) or M2 (CD163⁺) phenotypes^{17,18}. Based on these findings, we postulated that a positive correlation between CD163⁺ CD68⁺ CD14⁺ cell type (ma_3), which represents atypical M1-macrophage population and CD4⁺ memory T cells (CD4_1, CD_3, CD_5) could suppress the tumor progression in LTS.

Next, we examined the prognostic significance of cell-cell nearest-neighbor interactions by computing the average cell count of cell subtype X in the nearest neighborhood of cell subtype Y (distance between the center of two cells less than 20 μm) in the tumor-enriched region of every LTS and STS patient sample. We computed the cell-cell nearest-neighbor interactions that were significantly higher or lower (Benjamini-Hochberg adjusted p value < 0.05) in LTS ($n = 21$) than in STS ($n = 20$). We identified 120 cell-cell nearest-neighbor interactions that were significantly different between LTS and STS (Fig. 5A). Among them, granzyme B⁺ CD8⁺ cytotoxic T cells (CD8_4) had significantly more interactions with multiple tumor cell subtypes (tu_1, tu_2, tu_3, and tu_5) in LTS than in STS (Fig. 5A), suggesting increased interactions between CD8 cytotoxic cells and multiple subtypes of tumor cells in LTS. An example of the interaction between CD8_4 and tu_1 is shown in Fig. S5A.

In contrast to more CD8_4–tumor cell interactions in LTS than in STS, CD73^{mid} cell (CD73_2) had significantly fewer interactions with 17 cell subtypes, including CD73_1, CD31, macrophages and monocytes (ma_1, ma_2, ma_4, ma_5, ma_8, ma_9), stromal cells (s_1, s_2), T cells (CD4_5, CD8_2), and tumor cells (tu_1, tu_3, tu_4, tu_5, tu_6), than in STS, suggesting that when the number of CD73_2 cells is high, they are mainly surrounded by macrophages and tumor cells in STS. The interaction between CD73_2 and CD163⁺ CD68⁺ CD14⁺ macrophages (ma_9) is shown in Fig. S5B. CD4_4 cells had significantly more interactions with CD163⁺ CD68⁺ Vista^{mid} CD14⁺ macrophages (ma_1; Fig. S5C) and CD14⁺ monocytes (ma_2) in LTS than in STS (Fig. 5A), suggesting increased interactions between CD4⁺ memory T cells and certain subtypes of macrophages in LTS. CD45RO⁺ CD44⁺ CD8⁺ memory T cells (CD8_3) had significantly fewer interaction with CD163⁺ CD14^{mid} macrophages (ma_8; Fig. S5D, Fig. 5A) in LTS than in STS, suggesting decreased interactions between CD8⁺ memory T cells and this subtype of macrophages in LTS.

Feature selection for overall survival prediction by logistic regression

We used a machine learning method, logistic regression to predict survival, with recursive elimination of features after filtering out highly correlated ones (see Methods). Our first approach included using only the cell density detected in tumor-enriched regions and patient age as features (Fig. 5B). The optimal number of selected features was seven, because both training accuracy (0.947) and validation accuracy (0.938) were high (Fig. 5B a). The test accuracy was 0.78 (test sensitivity = 0.6, test specificity = 1) and

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

the area under the curve (AUC) was 0.8 (Fig. 5B.b). Among the seven prognostic features selected by the model, CD73^{mid} cells (CD73_2), CD31⁺ CD73^{mid} endothelial cells (CD31), CD163⁺ CD68⁺ Vista^{mid} CD14⁺ macrophages (ma_1), CD45RO⁺ CD44⁺ CD8⁺ memory T cells (CD8_3), and age had negative coefficients, suggesting that they were inversely correlated with patient survival. Granzyme B⁺ CD8⁺ cytotoxic T cells (CD8_4) and CD45RO⁺ CD4⁺ memory T cells (CD4_4) had positive coefficients, suggesting that they were positively correlated with patient survival (Fig. 5B.c, Fig. 4A).

Next, we used the cell-cell nearest-neighbor interactions, which are related to the seven prognostic cell density features in the tumor-enriched region, and patient age as features (Fig. 5C). The optimal number of selected features was 11, because both training accuracy (1) and validation accuracy (1) were the highest for that number of features (Fig. 5C.a). The test accuracy was 0.89 (test sensitivity = 1, test specificity = 0.75) and the AUC was 1 (Fig. 5C.b). Among the 11 features that were selected by the logistic regression model, CD73_2 neighboring CD73_1, ma_9, or s_2; CD31 neighboring tu_4; CD8_3 neighboring ma_8; and age were negatively correlated with patient survival. In contrast, CD4_4 neighboring s_2, s_1, ma_1, or CD44 and CD8_4 neighboring tu_1 were positively correlated with survival (Fig. 5C.c). Some of the prognostic nearest-neighbor interaction features, such as CD8_4 neighboring tu_1, CD73_2 neighboring ma_9, CD4_4 neighboring ma_1, and CD8_3 neighboring ma_8, can be visualized in Fig. S5. These results indicate that multiple neighboring interactions may have similar predictive power for survival prediction. The complex cell-cell interaction patterns of ovarian cancer with various immune cell and other stromal subtypes led to divergence in the tumor microenvironment between LTS and STS.

Taken together, our results indicated that using cell-cell nearest-neighbor interactions and age as features allowed a more accurate prediction of patient survival than using cell densities and age as features. The Spearman correlation of any two neighbor interaction features that both had relatively high correlation with patient survival (Spearman correlation coefficient > 0.2) and were related to the seven prognostic cell density features is visualized in Fig. 5D. The Spearman correlation study demonstrated that certain features were highly correlated. For example, we found that the average cell count of CD8_4 neighboring with a tu_1 cell was highly correlated with the average cell count of CD8_4 neighboring with a tu_2 cell ($r = 0.77$, $p = 5 \times 10^{-9}$) and the average cell count of CD8_4 neighboring with a tu_3 cell ($r = 0.72$, $p = 1 \times 10^{-7}$). The feature elimination process of our logistic regression model first filtered out the highly correlated features and kept one feature of each highly correlated feature pair (see Methods). Owing to this elimination process, features that are highly correlated with any feature in the prognostic list identified by the logistic regression model may also have prognostic value. For example, because CD8_4 neighboring tu_1 was a prognostic feature, CD8_4 neighboring tu_2 or CD8_4 neighboring tu_3 may have similar prognostic values.

Correlations between cell subtype density and transcriptomic profiles from microdissected fibroblastic and epithelial compartments of HGSC

To provide mechanistic insights by which certain cell phenotypes identified by IMC modulate survival in HGSC patients, we performed correlation studies to identify genes in the epithelial compartment (Fig. 6A) or fibroblastic compartment (Fig. 6B) of HGSCs that had a significant positive or negative correlation with the IMC cell density in the tumor tissue (Table S3 and 4) and that were significantly different between LTS and STS (Fig. S6 and 7). A total of 26 samples with whole transcriptome data available were used. We focused our attention on the cell densities of the six IMC features that were significantly different between LTS and STS (Fig. 4A), four of which were also selected by our machine learning model for survival prediction (Fig. 5B.c). We discovered relationships consistent with known cancer biology, and we also made several unexpected observations.

In the epithelial compartment of HGSC tumors, we identified expression levels of several genes involved in the migration and invasion of cancer cells correlated with CD73₁, CD73₂, and CD31 densities and with poor survival, suggesting that these genes regulate the metastatic potential of cancer cells in a paracrine manner; these genes included *PARD6B*, *S100A10*, *SLURP1*, and *SPINT1*. In addition, we demonstrated that ovarian cancer cell-derived *ITIH5*, *TACSTD2*, and *WFDC2* expression levels were negatively correlated with granzyme B⁺ CD8⁺ cytotoxic T cell (CD8₄) densities. In contrast, *BPIFB1* and *SLURP1* expression levels were positively correlated, and *CD9* and *ITIH5* were negatively correlated with CD45RO⁺ CD4⁺ memory T cell (CD4₄) densities (Fig. 6A, Table S3, Supp. Figure 6). These findings suggest that these ovarian cancer cell-derived genes, which are known to code for extracellular matrix and metabolism modulators^{19–25}, could facilitate or block T cell infiltration in the tumor microenvironment as well as interfere with T cell activation and facilitate immune system evasion. Moreover, cancer cell-derived *FST* and *PARD6B* were positively correlated with CD31⁺ CD73^{mid} endothelial cell (CD31) density, indicating that these genes modulate endothelial cell activity and subsequently angiogenesis. The prognostic significance of these genes was determined by analyzing 530 samples of optimally debulked advanced HGSC in the KM Plotter, and the results are summarized in Fig. S6.

To identify cancer-associated fibroblast (CAF)-derived mediators that confer the prognostic phenotypes identified by IMC, we analyzed correlations of the expression levels of genes in the transcriptomes generated from the microdissected fibroblastic compartment of HGSC with the prognostic phenotypes identified. Survival analyses of the significantly correlated genes were also performed. The list of CAF-derived genes that were positively correlated with CD73⁺ cell (CD73₁) and CD73^{mid} cell (CD73₂) densities and also associated with STS is shown in Table S4, Fig. 6B, and Fig. S7. These genes are likely expressed by CD73₁ and CD73₂, two different CAF subtypes. Among these genes, *CCDC85B*, *DDAH1*, *EFEMP2*, *F2RL1*, *ITGB1*, *LOX*, *LDLR*, *MFG8*, *MICAL2*, *MKL1*, *MSRB3*, *NCAM1*, *NPTX*, *PLAT*, *SLC2A3*, *SPSB1*, and *VASN* have been described as promoters of tumor cell growth, invasion, and migration, as well as angiogenesis^{26–41} (Table S4, Fig. S7). *HMOX2*, *ICMT*, *MICU1*, and *TSPAN9* also positively correlated with CD73₁ and CD73₂ densities and short survival and have been shown to be involved in conferring chemoresistance in ovarian cancer^{34,42–44} (Table S4, Fig. S7). CAF-derived *ANGPTL2*, *WWTR1*, and *MYO10* regulates CAF rigidity^{45–50}. These

findings indicate that these mediators produced by CD73_1 and CD73_2 CAF subtypes modulate the malignant phenotypes of HGSC.

CAF-derived genes involved in the modulation of the immune response and also associated with patient survival were also examined, and the results are summarized in Table S4 and Fig. S7. Among these genes, *BMPR1B*, a gene encoding the receptor of the bone morphogenetic protein (BMP), was negatively correlated with CD4_4 cells and associated with STS. This finding suggests that CAFs expressing high levels of *BMPR1B* may be more responsive to BMP signaling, which subsequently modulates the rigidity of CAFs, stiffness of the tumor microenvironment, and CD4_4 T cell trafficking. Besides *BMPR1B*, *VSTM4*, a secreted protein that can reduce IFN- γ , IL-2, and IL-17 cytokine production by human T cells and cause a profound decrease in T cell activation⁵¹, was negatively correlated with CD4_4 density but positively associated with STS, suggesting that CAF-derived *VSTM4* modulates CD4_4 activity and subsequently leads to poor survival rates in patients with HGSC.

In addition to these correlations between expression of CAF-derived genes and prognostic immune cell phenotypes, expression of several genes with prognostic significance was also associated with the two CAF phenotypes CD73_1 and CD73_2. *FN1*, *TGFBI*, *TNC*, and *LRRC32* were positively correlated with both CD73_1 and CD73_2 and negatively correlated with survival. *LRRC32* is a key regulator of TGF- β activation⁵². Together with increased TGF β 1 secreted by CD73_1 and CD73_2 CAFs, *LRRC32* may generate an immune-suppressive and pro-tumorigenic microenvironment to support the malignant phenotype of ovarian cancer cells, as we previously described^{53,54}. Both *FN1* and *TNC* encode extracellular proteins that have previously been shown to be associated with short progression-free survival and increased migration-inducing potential in HGSC^{55,56}.

Several fibroblastic genes were associated with density of endothelial cells (CD31). Among them, *MFGE8*, which was also positively correlated with CD73_1 and CD73_2 density, demonstrated the strongest correlation with increased CD31 density and poor patient survival (Table S4 and Fig. S7). Our findings are supported by published studies reporting that *MFGE8* increases tumor angiogenesis by increasing VEGF and ET-1 expression in stromal cells and by enhancing M2 polarization of macrophages³². Moreover, *MFGE8* proteins accumulated around CD31⁺ blood vessels have been shown to promote angiogenesis by enhancing PDGF-PDGFR β signaling mediated by integrin-growth factor receptor crosstalk^{33,57}.

Taken together, the transcriptome analysis revealed multiple key genes in both the cancer cell and fibroblastic compartments of HGSC that correlate with IMC features positively or negatively, confer either a tumor-promoting or an immune-suppressive microenvironment, and subsequently lead to short-term survival in HGSC patients (Fig. 6C).

Discussion

In the present study, we applied a new analytic pipeline to IMC data based on deep machine learning that

markers in each sample, focusing our attention on the characterization of the immune milieu in both tumoral and stromal compartments. MRCNN, a deep learning method that was initially designed for object detection and instance segmentation of natural images¹⁴, outperformed all existing single-model entries on every task in the recent Microsoft Common Objects in Context (COCO) challenge, one of the most authoritative competitions in object detection and segmentation⁵⁸. MRCNN was adapted to perform nuclei segmentation in histologic microscopic images⁵⁹. In the current study, we demonstrated that the MRCNN model can also be adapted for cell segmentation in noisy, densely packed, structurally complicated, and widely varying IMC images with minor modification. We also showed that it performed better than the traditional watershed method for IMC cell segmentation using the computerized outputs of watershed segmentation as training sets.

Analysis of spatially resolved cell densities enabled us to identify several cell subtypes exhibiting significant differences between LTS and STS. In particular, we discovered that the density of granzyme B⁺ CD8⁺ cytotoxic T cells (CD8_4) was significantly higher in LTS than in STS, expanding the prior knowledge that CD3⁺, CD4⁺, and CD8⁺ tumor-infiltrating lymphocytes are associated with positive outcomes in ovarian cancer^{60–63} and highlighting the fact that only a specific subtype of activated T cells (granzyme B⁺) is associated with good clinical outcomes in ovarian cancer. In addition, we found that the density of CD45RO⁺ CD4⁺ memory T cell (CD4_4) was significantly higher in LTS than in STS. Memory CD4⁺ T cells could provide a protective response against cancer cells by making effector cytokines respond early and by enhancing CD8⁺ T and B cell responses, as well as by secreting cytokines that can induce other cells in the tumor microenvironment to mount antitumor immunity⁶¹.

Among different CD73⁺ cell subtypes, densities of CD73_1 and CD73_2, two different CAF subtypes, were significantly lower in LTS than in STS. CD73 is a GPI-anchored nucleotidase that catabolizes the production of extracellular adenosine and promotes tumor immune escape and thereby tumor growth. Indeed, CD73 expression has been shown to be associated with shorter disease-free and overall survival in HGSC patients and decreased CD8⁺ tumor-infiltrating lymphocytes^{64,65}. A recent study reported that CAF-derived CD73 enforces an immune checkpoint⁶⁶. These findings reinforce the hypothesis that CAFs play a role in shaping the immune landscape of the tumor microenvironment and modulating patient survival rates.

Among different tumor cell subtypes, B7H4⁺ Keratin⁺ tumor cell (tu_9) density was significantly lower in LTS than in STS. B7-H4 overexpression in cancer cells has been previously identified in high-grade ovarian tumors⁶⁷, but to the best of our knowledge, this is the first report showing that increased density of tu_9, a subtype of ovarian cancer cells expressing high levels of B7-H4, is associated with poor patient overall survival rates.

Finally, we found CD31⁺ CD73^{mid} endothelial cell (CD31) density was significantly lower in LTS than in STS. Although CD31 expression has shown no prognostic survival value, high CD31 expression was

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js⁶⁸.

Communication between heterogeneous tumor cells and various types of stromal cells, including infiltrating T cells, macrophages, CAFs, endothelial cells, and others, has been shown to be able to shape the ovarian cancer ecosystem, which subsequently modulates disease progression and clinical outcome^{3,62,69}. Our deep machine learning analytic pipeline identified 10 cell-cell nearest-neighbor interactions together with age as features that can achieve the best survival prediction accuracy, resulting in an AUC of 1. These findings indicate that interaction between different cell subtypes in the tumor microenvironment generates better prognostic features in predicting patient survival rates than cell density alone. For example, CD31 neighboring tu_4 was negatively correlated with patient survival, confirming that increased angiogenesis supports tumor cell growth and subsequently leads to poor patient survival rates. In contrast, CD8_4 neighboring tu_1 was positively correlated with survival, suggesting that intratumoral activated CD8⁺ CTL closely interacting with tumor cells represent an attempt of an anti-tumor response, which subsequently leads to improved patient survival rates, as has previously been reported^{60–63}. Increased interaction between heterogeneous populations of cell subtypes in the tumor microenvironment likely involves ligand-receptor crosstalk among the different cell subtypes. Further experiments using spatially resolved single cell transcriptomes on one cell subtype with its nearest neighboring partner will be needed to validate the cell-cell interactions as well as to understand how these interactions and the crosstalk signaling networks contribute to malignant phenotypes and their correlation with patient survival rates.

Several studies have coupled IMC data to multiplatform genomics to understand how the genome shapes the composition and architecture of tumor ecosystems^{70,71}. To delineate the molecular mechanisms by which certain cell phenotypes identified by IMC modulate survival rates in HGSC patients, correlation studies have used genes in the microdissected epithelial compartment or fibroblastic compartment of HGSCs associated with patient survival and prognostic cell phenotypes identified by IMC. We found that cancer cell-derived *WFDC2* negatively correlated with granzyme B⁺ CD8⁺ cytotoxic T cell (CD8_4) density and was associated with poor HGSC patient survival. In fact, *WFDC2*, which encodes the protein HE4, has been shown to correlate with poor survival in HGSC patients and promote tumor growth and confer chemoresistance in ovarian cancer⁷². Moreover, HE4 has been described as a driver of immune failure in ovarian tumors by compromising cytotoxic CD8⁺ T cells through upregulation of self-produced dual-specificity phosphatase 6 (DUSP6)⁷³. These findings suggest that our deep machine learning pipeline showed robust performance in identifying prognostic biomarkers associated with immune cell phenotypes described in previous studies. In addition to *WFDC2*, we found that other interesting genes involved in the regulation of T cell activity, apoptosis, and infiltration were overexpressed in STS; these genes include *ITM2B*, *ITIH5*, *CD9*, and *TACSTD2*^{21,22,74}.

CAFs have been shown to facilitate cancer progression by supporting tumor cell growth, extracellular matrix remodeling, angiogenesis, and formation of an immunosuppressive microenvironment⁷¹. These results are supported by our research showing that CAF subtypes CD73_1 and CD73_2 express genes, such as *ANGPTL2*, *TNC*, *TGFB1*, *FN1*, *BMPR1B*, and *LRRC32*, that have been shown to promote tumor

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js | eling^{45,46,52,55,56,75,76}. Some of these genes,

including *TGFB1*, *BMPR1B*, and *LRRC32*, have been shown to modulate TGF- β signaling, which suppresses infiltration of anticancer immune cells such as cytotoxic T cells and natural killer cells and promotes the function of pro-cancer immune cells, such as regulatory T cells and M2 macrophages, in the tumor microenvironment^{45,77,78}, leading to poor patient survival rates. Using additional antibodies targeting these immune cell phenotypes in IMC analysis will further validate our observations.

In conclusion, our novel deep machine learning-based IMC analytic pipeline combined with transcriptomes generated from microdissected epithelial and fibroblastic compartments of HGSC patient specimens demonstrates the heterogeneity of both tumor and stromal cell subtypes in HGSC. The image-omics analysis also identified cellular features and phenotypes with prognostic significance and helped delineate the molecular mechanism by which these features modulate the tumor-promoting and immune-suppressive microenvironment (Fig. 6C). Integrating quantitative spatial features, IMC-derived cell phenotypes, and transcriptomic data provides a unique opportunity to characterize ovarian cancer in an unprecedented way, generating multiparametric tissue biomarkers to predict clinical outcomes.

Materials And Methods

Patient samples

A total of 41 paraffin-embedded tumor tissue samples obtained from patients with advanced stage (stage IIIB-IV) high-grade serous ovarian cancer (HGSC) were used in the current study. Tissue samples were obtained from the ovarian cancer repositories at The University of Texas MD Anderson Cancer Center and Gangnam Severance Hospital, Yonsei University College of Medicine. They were collected from previously untreated patients undergoing primary cytoreductive surgery for ovarian cancer. After surgery, patients received platinum-based combination chemotherapy. Optimal surgical cytoreduction was defined by a residual tumor no more than 1 cm in diameter. The overall survival duration was measured from the date of diagnosis to the date of death or censored at the date of the last follow-up examination. Long-term survivors were those with an overall survival time ≥ 60 months, and short-term survivors were those with an overall survival time ≤ 20 months. Clinical data, including age, cytoreduction status (optimal vs. suboptimal), and overall survival, were obtained from the records of the patients with HGSC. All samples and clinical data were collected with the approval of the Institutional Review Boards of MD Anderson and Gangnam Severance Hospital.

Preparation and staining

Tissue slides were deparaffinized in xylene followed by rehydration in a graded alcohol series. Antigen retrieval was performed with citrate buffer (pH 6) at 95 °C in a decloaking chamber (Biocare Medical) for 25 minutes. Slides were then blocked with 3% bovine serum albumin in phosphate-buffered saline for 30 minutes and incubated for 2 hours at room temperature with metal-tagged antibodies described in Fig. S1. Following incubation, tissue slides were washed with phosphate-buffered saline and incubated with 0.5 μ M Cell-ID Intercalator-Ir (Fluidigm) for the detection of nuclear DNA. Slides were then rinsed in

Imaging mass cytometry

Imaging mass cytometry (IMC) data were acquired by a Fluidigm Helios CyTOF instrument equipped with a Hyperion System laser ablation module in the Flow Cytometry and Cellular Imaging Facility at MD Anderson. A total of 41 images of 1 mm^3 each were acquired and used for the current study, including 20 images from short-term survivors (overall survival ≤ 20 months) and 21 images from long-term survivors (overall survival ≥ 60 months). Each 1 mm^2 region of interest on the tissue section was selected based on the image from the corresponding hematoxylin and eosin stained serial tissue section, which demonstrated representative of tumor regions surrounded by stromal cells.

Microdissection and microarray analysis of tissue samples

RNA was extracted from microdissected frozen HGSC samples, which included tumor epithelial components and stromal components from 16 HGSC patients, from whom IMC data were available. Extensive details of specimen handling, RNA extraction and amplification, microarray hybridization, and quality-control procedures have been described previously⁵⁴.

Data preprocessing and cell segmentation

Data were converted to TIFF format by MCD viewer (Fluidigm). Channel spillover was compensated using the nonnegative least square approach^{79,80}. Watershed segmentation was performed on the maximal projection of normalized images of H3 and nucleic acid intercalator (¹⁹¹Ir and ¹⁹³Ir). The best Watershed segmentation results were achieved with prior median filtering (2×2 pixels) followed by Gaussian blurring (kernel width of 2 pixels), and standard parameters for watersheds. These steps were performed by an in-house-developed Matlab script with the Matlab image processing toolbox. Mask-R-CNN (MRCNN) segmentation was trained on the outputs of Watershed segmentation. Thirty-one images were employed for training, with each image segmented into 16 small pieces to increase the training speed (each small image had a size of $\sim 250 \times 250$ pixels, image resolution $1 \mu\text{m}/\text{pixel}$). A total of 496 small images were used as the training set. Each image was converted from a grayscale image into an RGB image in which the R and G channels were median-filtered grayscale images with 2×2 and 3×3 neighborhoods, respectively, and the B channel was the raw grayscale image. We used an MRCNN model with a feature pyramid network and a convolutional neural network ResNet-101 backbone based on an existing implementation (https://github.com/matterport/Mask_RCNN) by Matterport Inc. (released under an MIT License) that employed the python open-source libraries Keras and Tensorflow. We initiated the model using weights obtained from pretraining on the MSCOCO dataset⁵⁸. We started with a learning rate of 0.001 and trained with 50 epochs and decreased the learning rate to 0.0001 and trained for 50 epochs. The training was performed using one NVIDIA V100 GPU (Amazon AWS p3.2xlarge instance). To compare the results of Watershed and MRCNN segmentation, we manually segmented 10 testing images (~ 300 cells per image, each image had a size of $\sim 250 \mu\text{m} \times 250 \mu\text{m}$) to be used as a reference segmentation. To quantify the performance of different segmentation methods, we computed the dice

similarity coefficient $\frac{2|x \cap y|}{|x \cup y|}$ between each cell in Watershed segmentation and its maximum

overlapping cell in reference segmentation, and between each cell in MRCNN segmentation and its maximum overlapping cell in reference segmentation. We compared the mean and standard deviation of the dice similarity coefficient of all the cells in each sample between Watershed and MRCNN segmentations (Fig. S1B, C).

Signal-to-noise ratio evaluation and data normalization

After spillover compensation of all channels, the marker expression of a cell was computed as the average of the pixel values of each marker within a segmented cell. The signal level of a marker was calculated as the mean of marker expression in cells with marker expression higher than the 99.5 percentile. The noise level was calculated as the standard deviation of the pixel values of an image. Signal-to-noise ratio larger than 1 for most of the samples was a criterion for keeping good channels for the entire data analysis. The pixel values of each marker were censored at the 99.9 percentile to remove outliers, and the marker expression of cells in each image was z-scored before Phenograph clustering.

Analysis workflow

MRCNN segmentation, survival prediction, and correlation of cell density with gene expression were implemented in python 3.5. Watershed segmentation, image analysis, and the single-cell analysis pipeline were performed by Matlab R2016a. Phenograph clustering and heatmap figures for cell subtype annotation were generated by R 3.6.

Clustering analysis

Phenograph clustering was performed using the Matlab cyt package. In the first step, the data were under-clustered to detect and separate the major cell populations using 100 nearest neighbors. Sixteen markers were used: SMA, CD14, CD163, CD11b, CD45, CD44, CD4, CD73, CD68, CD20, CD8a, granzyme B, Ki67, Coll-I, CD45RO, and Keratin8/18. Of the nineteen clusters generated, 9 tumor or stroma clusters were kept and 10 non-tumor and non-stroma clusters were pooled together and underwent a second round of clustering using 80 nearest neighbors and 15 markers: CD14, CD163, CD11b, CD31, CD45, CD44, CD4, CD73, CD68, CD20, CD8a, CD196, granzyme B, Ki67, and CD45RO. Of the eighteen clusters generated, three macrophage clusters were pooled together and were clustered at a third round by markers: CD44, CD14, CD163, CD68, CD4, CD45RO, and CD11b. Similarly, CD8⁺ T cell cluster, CD4⁺ T cell cluster, and clusters that had mixed tumor and immune cells underwent a third round of clustering using their related markers. Finally, the clusters generated by all three rounds were gathered together and visualized in Barnes-Hut t-SNE⁸¹, a two-dimensional representation of high dimensional data.

Cell density and nearest-neighbor interactions in tumor-enriched regions

Tumor-enriched regions were calculated as a thresholded two-dimensional Gaussian convolved image ($\sigma = 15$ pixels, threshold = 0.0005) of the density map of the centroid of tumor cells. The regions outside of tumor-enriched regions were defined as tumor-unenriched regions. Cell density in tumor-enriched regions

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

ion per total tumor cells. For the cell-cell nearest-

neighbor interactions, the average cell count of cell subtype X in the nearest neighborhood of cell subtype Y (distance between the centroids of X and Y < 20 μm) in the tumor-enriched region was computed. An unpaired t test was used to determine if there was a significant difference (adjusted $p < 0.05$) between the means of cell-cell nearest-neighbor interactions of long-term and short-term survivors. Adjustment for multiple testing was conducted using the Benjamini-Hochberg method⁸².

Survival prediction

The training and test data sets were randomly split at 80% and 20% of the total 41 samples that contained 21 long-term survivors and 20 short-term survivors, respectively. The ratio of the long-term to short-term survivors in the training data set was 1 and in the test data set was 1.25. For survival prediction using only cell density and age, the 41 cell density and age features were first subjected to Spearman correlation with survival, and 22 features that had an absolute correlation coefficient larger than 0.2 were kept. Because logistic regression assumes independence between features, features could not be highly correlated. Only one of the highly correlated features (absolute Spearman correlation coefficient ≥ 0.65) was kept and the rest were dropped, leaving 17 features. Each feature was normalized to a 0 to 1 scale. Recursive feature elimination and logistic regression (python's sklearn package) were used to rank the features according to their importance. Owing to the small sample size, leave-one-out cross validation was used to evaluate performance during the training. The optimal feature number was selected at the highest validation accuracy.

To narrow down the nearest-neighbor interaction features that were related to the prognostic cell density features, we selected 46 features of the nearest-neighbor interaction that contained any of the seven prognostic cell density features and were also significantly different between long-term and short-term survivors and combined these features with age to predict survival. To filter the features, we performed Spearman correlation between survival and these 47 features. Forty-four features that had an absolute correlation coefficient larger than 0.2 were kept. Because features of logistic regression cannot be highly correlated, only one of the highly correlated features (absolute Spearman correlation coefficient ≥ 0.65) was kept, leaving 26 features. Each feature was normalized to a 0 to 1 scale. Recursive feature elimination and logistic regression were used to rank the features according to their importance. Owing to the small sample size, leave-one-out cross validation was used to evaluate performance during the training. The optimal feature number was selected at the highest validation accuracy.

Correlation of cell density with gene expression

Gene expression was normalized by Robust Multi-array Average⁸³. Spearman correlation between gene expressions in microdissected fibroblastic or epithelial components and cell densities in tumor-enriched regions was performed and genes of interest were selected from those that had positive correlation coefficients ($p < 0.05$, absolute correlation coefficient > 0.4) with each cell subtype that showed a significant difference between long-term survivors and short-term survivors. For genes that had multiple probe IDs, we used only the probe ID that had the largest variance. The genes of interest of the microdissected fibroblastic components were filtered by the single-cell RNA sequencing (scRNAseq)

data⁸⁴, and only the genes that were expressed in fibroblasts or stromal cells were kept. Moreover, to understand the molecular mechanism that might explain the correlation between IMC and HGSC patient survival, we included in the heatmap in Fig. 6B only genes encoding for secreted or receptor proteins for all listed IMC features, except for CD73_1 and CD73_2 (for which we included all genes). The genes of interest of the microdissected epithelial components were filtered by the scRNAseq data and only genes that were expressed in epithelial cells were kept. Among these, only genes encoding for secreted or receptor proteins for all listed IMC features, except for tu_9 (for which we included all genes), were included in the heatmap in Fig. 6A. Kaplan-Meier analysis was performed on the filtered genes of interest, and only genes that had significant prognostic values ($p < 0.05$) were kept.

Kaplan-Meier analysis

Kaplan-Meier analysis was performed differently for genes in fibroblastic and epithelial components. For genes in fibroblastic components, Kaplan-Meier analysis was performed on our microdissected gene expression data in the fibroblastic component (70 patients, overall survival < 150 months). For genes in epithelial components, Kaplan-Meier analysis was performed using an online tool, KMplotter⁸⁵, which employs a database of gene expression data and survival information of 530 HGSC patients (stage II/III/IV; grade 2/3; optimal debulking), downloaded from Gene Expression Omnibus and the Cancer Genome Atlas. To analyze the prognostic value of each selected gene, we divided the patients into two groups according to various quantile expressions of the gene, and the best performing threshold was used as a cutoff. Statistical comparison of the two groups was performed using the log-rank test.

Declarations

Acknowledgments: We are grateful for the generous donation of tissue samples by patients undergoing surgery. We thank Siu Fee Rita for excellent technical support with IMC slide staining. We also thank Duncan Mak and Angelique Lin, from Flow Cytometry and Cellular Imaging Facility at MD Anderson Cancer Center, for the exceptional work in assisting with IMC samples acquisition and technical support. We are also grateful to Erica Goodoff from the Scientific Publications, Research Medical Library at MD Anderson Cancer Center for help editing the manuscript.

Funding: This study was supported in part by W81XWH-17-1-0126 and W81XWH-16-1-0038 from the Ovarian Cancer Research Program, US Department of Defense; the MD Anderson Cancer Center Support Grant P30CA016672 from the National Institutes of Health; the Sister Institution Network Fund from the University of Texas MD Anderson Cancer Center; the US Department of Health and Human Services; the T.T. and W.F. Chao Foundation, Stephanie C. Stelter Foundation, John S. Dunn Research Foundation; and Carole Walter Looke Fund. Y.Z. was supported by a training fellowship from the Gulf Coast Consortia, on the Computational Cancer Biology Training Program (CPRIT Grant No. RP170593).

Author contributions: Y.Z, S.F-B., T-L.Y and S.C.M. planned the experiments. Y.Z and S.F-B conducted the majority of the experiments and analyzed data. A.A.J. provided the clinical samples. T-L.Y., P.C., J.S.,

M.J.B., and J.K.B. provided intellectual contributions to experimental design and data analysis. Y.Z, S.F-B., S.T.C.W. and S.C.M. prepared and revised the manuscript.

Competing interests: The authors have no competing interests to declare.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA. Cancer J. Clin.* **70**, 7–30 (2020).
2. Mikuła-Pietrasik, J. *et al.* Comprehensive review on how platinum- and taxane-based chemotherapy of ovarian cancer affects biology of normal cells. *Cell. Mol. Life Sci.* **76**, 681–697 (2019).
3. Sato, S. & Itamochi, H. Neoadjuvant chemotherapy in advanced ovarian cancer: latest results and place in therapy. *Ther. Adv. Med. Oncol.* **6**, 293–304 (2014).
4. Ushijima, K. Treatment for Recurrent Ovarian Cancer—At First Relapse. *J. Oncol.* 2010, (2010).
5. Lisio, M.-A., Fu, L., Goyeneche, A., Gao, Z. & Telleria, C. High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints. *Int. J. Mol. Sci.* **20**, (2019).
6. Pokhriyal, R., Hariprasad, R., Kumar, L. & Hariprasad, G. Chemotherapy Resistance in Advanced Ovarian Cancer Patients. *Biomark. Cancer* **11**, (2019).
7. Son, B. *et al.* The role of tumor microenvironment in therapeutic resistance. *Oncotarget* **8**, 3933–3945 (2016).
8. Senthebane, D. A. *et al.* The Role of Tumor Microenvironment in Chemoresistance: To Survive, Keep Your Enemies Closer. *Int. J. Mol. Sci.* **18**, (2017).
9. Plava, J. *et al.* Recent advances in understanding tumor stroma-mediated chemoresistance in breast cancer. *Mol. Cancer* **18**, 67 (2019).
10. Ungefroren, H., Sebens, S., Seidl, D., Lehnert, H. & Hass, R. Interaction of tumor cells with the microenvironment. *Cell Commun. Signal. CCS* **9**, 18 (2011).
11. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
12. Levine, J. H. *et al.* Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
13. Schapiro, D. *et al.* miCAT: A toolbox for analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873–876 (2017).
14. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 386–397 (2020).
15. Elliott, L. A., Doherty, G. A., Sheahan, K. & Ryan, E. J. Human Tumor-Infiltrating Myeloid Cells: Phenotypic and Functional Diversity. *Front. Immunol.* **8**, (2017).

16. Yamaguchi, T. *et al.* Tumor-associated macrophages of the M2 phenotype contribute to progression in gastric cancer with peritoneal dissemination. *Gastric Cancer* **19**, 1052–1065 (2016).
17. Koelzer, V. H. *et al.* Phenotyping of tumor-associated macrophages in colorectal cancer: Impact on single cell invasion (tumor budding) and clinicopathological outcome. *Oncoimmunology* **5**, (2015).
18. Minami, K. *et al.* Prognostic significance of CD68, CD163 and Folate receptor- β positive macrophages in hepatocellular carcinoma. *Exp. Ther. Med.* **15**, 4465–4476 (2018).
19. Wei, F. *et al.* BPIFB1 (LPLUNC1) inhibits migration and invasion of nasopharyngeal carcinoma by interacting with VTN and VIM. *Br. J. Cancer* **118**, 233–247 (2018).
20. Suppression of cell motility and metastasis by transfection with human motility-related protein (MRP-1/CD9) DNA. *J. Exp. Med.* **177**, 1231–1237 (1993).
21. Hwang, J. R. *et al.* Upregulation of CD9 in ovarian cancer is related to the induction of TNF- α gene expression and constitutive NF- κ B activation. *Carcinogenesis* **33**, 77–83 (2012).
22. Veeck, J. *et al.* The extracellular matrix protein ITIH5 is a novel prognostic marker in invasive node-negative breast cancer and its aberrant expression is caused by promoter hypermethylation. *Oncogene* **27**, 865–876 (2008).
23. Ridge, R. J. & Sloane, N. H. Partial N-terminal amino acid sequence of the anti-neoplastic urinary protein (ANUP) and the anti-tumour effect of the N-terminal nonapeptide of the unique cytokine present in human granulocytes. *Cytokine* **8**, 1–5 (1996).
24. Moriwaki, Y. *et al.* Immune system expression of SLURP-1 and SLURP-2, two endogenous nicotinic acetylcholine receptor ligands. *Life Sci.* **80**, 2365–2368 (2007).
25. Swamynathan, S. & Swamynathan, S. K. SLURP-1 Modulates Corneal Homeostasis by Serving as a Soluble Scavenger of Urokinase-Type Plasminogen Activator. *Invest. Ophthalmol. Vis. Sci.* **55**, 6251–6261 (2014).
26. Feng, Y. *et al.* CCDC85B promotes non-small cell lung cancer cell proliferation and invasion. *Mol. Carcinog.* **58**, 126–134 (2019).
27. Hulin, J.-A., Tommasi, S., Elliot, D. & Mangoni, A. A. Small molecule inhibition of DDAH1 significantly attenuates triple negative breast cancer cell vasculogenic mimicry in vitro. *Biomed. Pharmacother.* **111**, 602–612 (2019).
28. Chen, J. *et al.* Fibulin-4 is associated with tumor progression and a poor prognosis in ovarian carcinomas. *BMC Cancer* **15**, 91 (2015).
29. Ungefroren, H. *et al.* Proteinase-Activated Receptor 2 May Drive Cancer Progression by Facilitating TGF- β Signaling. *Int. J. Mol. Sci.* **18**, (2017).
30. ZHANG, L. & ZOU, W. Inhibition of integrin β 1 decreases the malignancy of ovarian cancer cells and potentiates anticancer therapy via the FAK/STAT1 signaling pathway. *Mol. Med. Rep.* **12**, 7869–7876 (2015).
31. Ji, F. *et al.* Hypoxia inducible factor 1 α -mediated LOX expression correlates with migration and invasion in epithelial ovarian cancer. *Int. J. Oncol.* **42**, 1578–1588 (2013).

32. Yamada, K. *et al.* MFG-E8 drives melanoma growth by stimulating mesenchymal stromal cell-induced angiogenesis and M2 polarization of tumor-associated macrophages. *Cancer Res.* **76**, 4283–4292 (2016).
33. Uchiyama, A. *et al.* MFG-E8 Regulates Angiogenesis in Cutaneous Wound Healing. *Am. J. Pathol.* **184**, 1981–1990 (2014).
34. Zhu, L.-Y. *et al.* Silencing of MICAL-L2 suppresses malignancy of ovarian cancer by inducing mesenchymal–epithelial transition. *Cancer Lett.* **363**, 71–82 (2015).
35. Hernandez, L. *et al.* Characterization of ovarian cancer cell lines as in vivo models for preclinical studies. *Gynecol. Oncol.* **142**, 332–340 (2016).
36. Zecchini, S. *et al.* The adhesion molecule NCAM promotes ovarian cancer progression via FGFR signalling. *EMBO Mol. Med.* **3**, 480–494 (2011).
37. Xu, C. *et al.* NPTX2 promotes colorectal cancer growth and liver metastasis by the activation of the canonical Wnt/ β -catenin pathway via FZD6. *Cell Death Dis.* **10**, (2019).
38. Teliga-Czajkowska, J., Sienko, J., Jalinik, K., Smolarczyk, R. & Czajkowski, K. Prognostic value of tissue plasminogen activator (tPA) in patients with epithelial ovarian cancer undergoing chemotherapy. *Ginekol. Pol.* **90**, 235–241 (2019).
39. Ancey, P.-B., Contat, C. & Meylan, E. Glucose transporters in cancer – from tumor cells to the tumor microenvironment. *FEBS J.* **285**, 2926–2943 (2018).
40. Kim, H.-J. *et al.* SPSB1 enhances ovarian cancer cell survival by destabilizing p21. *Biochem. Biophys. Res. Commun.* **510**, 364–369 (2019).
41. Liang, W. *et al.* Vasorin stimulates malignant progression and angiogenesis in glioma. *Cancer Sci.* **110**, 2558–2572 (2019).
42. Podkalicka, P., Mucha, O., Józkwicz, A., Dulak, J. & Łoboda, A. Heme oxygenase inhibition in cancers: possible tools and targets. *Contemp. Oncol.* **22**, 23–32 (2018).
43. Liu, Q. *et al.* Isoprenylcysteine carboxylmethyltransferase regulates ovarian cancer cell response to chemotherapy and Ras activation. *Biochem. Biophys. Res. Commun.* **501**, 556–562 (2018).
44. Qi, Y. *et al.* TSPAN9 suppresses the chemosensitivity of gastric cancer to 5-fluorouracil by promoting autophagy. *Cancer Cell Int.* **20**, (2020).
45. Crawford, Y. *et al.* PDGF-C Mediates the Angiogenic and Tumorigenic Properties of Fibroblasts Associated with Tumors Refractory to Anti-VEGF Treatment. *Cancer Cell* **15**, 21–34 (2009).
46. Aoi, J. *et al.* Angiopoietin-like Protein 2 Is an Important Facilitator of Inflammatory Carcinogenesis and Metastasis. *Cancer Res.* **71**, 7502–7512 (2011).
47. Zanconato, F., Cordenonsi, M. & Piccolo, S. YAP and TAZ: a signalling hub of the tumour microenvironment. *Nat. Rev. Cancer* **19**, 454–464 (2019).
48. Pankova, D. *et al.* Cancer-associated Fibroblasts Induce a Collagen Cross-link Switch in Tumor Stroma. *Mol. Cancer Res. MCR* **14**, 287–295 (2016).

49. Cao, R. *et al.* Elevated expression of myosin X in tumours contributes to breast cancer aggressiveness and metastasis. *Br. J. Cancer* **111**, 539–550 (2014).
50. Zhang, X. *et al.* Protease activated receptor 2 mediates tryptase-induced cell migration through MYO10 in colorectal cancer. *Am. J. Cancer Res.* **9**, 1995–2006 (2019).
51. Wang, J. *et al.* VSTM4 is a novel negative regulator of T cell activation. *J. Immunol.* **202**, 124.4-124.4 (2019).
52. Liu, T. *et al.* Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. *J. Hematol. Oncol.* **12**, 86 (2019).
53. Yeung, T.-L. *et al.* TGF- β Modulates Ovarian Cancer Invasion by Upregulating CAF-Derived Versican in the Tumor Microenvironment. *Cancer Res.* **73**, 5016–5028 (2013).
54. Yeung, T.-L. *et al.* Systematic Identification of Druggable Epithelial-Stromal Crosstalk Signaling Networks in Ovarian Cancer. *J. Natl. Cancer Inst.* **111**, 272–282 (2019).
55. Didem, T. *et al.* Clinical significance of serum tenascin-c levels in epithelial ovarian cancer. *Tumor Biol.* **35**, 6777–6782 (2014).
56. Yousif, N. G. Fibronectin promotes migration and invasion of ovarian cancer cells through up-regulation of FAK–PI3K/Akt pathway. *Cell Biol. Int.* **38**, 85–91 (2014).
57. Raymond, A., Ensslin, M. A. & Shur, B. D. SED1/MFG-E8: a bi-motif protein that orchestrates diverse cellular interactions. *J. Cell. Biochem.* **106**, 957–966 (2009).
58. 10.1007/978-3-319-10602-1_48
Lin, T.-Y. *et al.* Microsoft COCO: Common Objects in Context. in *Computer Vision – ECCV 2014* (eds. Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) 740–755 (Springer International Publishing, 2014). doi:10.1007/978-3-319-10602-1_48.
59. Johnson, J. W. Adapting Mask-RCNN for Automatic Nucleus Segmentation. *ArXiv180500500 Cs* **944**, (2020).
60. Hwang, W.-T., Adams, S. F., Tahirovic, E., Hagemann, I. S. & Coukos, G. Prognostic significance of tumor-infiltrating T cells in ovarian cancer: A meta-analysis. *Gynecol. Oncol.* **124**, 192–198 (2012).
61. Santoiemma, P. P. & Powell, D. J. Tumor infiltrating lymphocytes in ovarian cancer. *Cancer Biol. Ther.* **16**, 807–820 (2015).
62. Zhang, L. *et al.* Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N. Engl. J. Med.* **348**, 203–213 (2003).
63. Hamanishi, J. *et al.* The comprehensive assessment of local immune status of ovarian cancer by the clustering of multiple immune factors. *Clin. Immunol.* **141**, 338–347 (2011).
64. Wang, L. *et al.* CD73 has distinct roles in nonhematopoietic and hematopoietic cells to promote tumor growth in mice. *J. Clin. Invest.* **121**, 2371–2382 (2011).
65. Turcotte, M. *et al.* CD73 Is Associated with Poor Prognosis in High-Grade Serous Ovarian Cancer. *Cancer Res.* **75**, 4494–4503 (2015).

66. Yu, M. *et al.* CD73 on cancer-associated fibroblasts enhanced by the A 2B -mediated feedforward circuit enforces an immune checkpoint. *Nat. Commun.* **11**, 515 (2020).
67. Liang, L. *et al.* B7-H4 Expression in Ovarian Serous Carcinoma: A Study of 306 Cases. *Hum. Pathol.* **57**, 1–6 (2016).
68. Rask, L. *et al.* Association of CD31 and p53 with survival of ovarian cancer patients. *Anticancer Res.* **39**, 567–576 (2019).
69. Kroeger, D. R., Milne, K. & Nelson, B. H. Tumor-Infiltrating Plasma Cells Are Associated with Tertiary Lymphoid Structures, Cytolytic T-Cell Responses, and Superior Prognosis in Ovarian Cancer. *Clin. Cancer Res.* **22**, 3005–3015 (2016).
70. Ali, H. R. *et al.* Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nat. Cancer* **1**, 163–175 (2020).
71. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* **177**, 1330–1345.e18 (2019).
72. Moore, R. G. *et al.* HE4 (WFDC2) gene overexpression promotes ovarian tumor growth. *Sci. Rep.* **4**, 3574 (2014).
73. James, N. E. *et al.* Human Epididymis Secretory Protein 4 (HE4) Compromises Cytotoxic Mononuclear Cells via Inducing Dual Specificity Phosphatase 6. *Front. Pharmacol.* **10**, (2019).
74. Fleischer, A., Ayllón, V., Dumoutier, L., Renauld, J.-C. & Rebollo, A. Proapoptotic activity of ITM2B s, a BH3-only protein induced upon IL-2-deprivation which interacts with Bcl-2. *Oncogene* **21**, 3181–3189 (2002).
75. Steitz, A. M. *et al.* Tumor-associated macrophages promote ovarian cancer cell migration by secreting transforming growth factor beta induced (TGFB1) and tenascin C. *Cell Death Dis.* **11**, 1–15 (2020).
76. Wang, R. *et al.* GARP regulates the bioavailability and activation of TGF β . *Mol. Biol. Cell* **23**, 1129–1139 (2012).
77. Katsuta, E., Maawy, A. A., Yan, L. & Takabe, K. High expression of bone morphogenetic protein (BMP) 6 and BMP7 are associated with higher immune cell infiltration and better survival in estrogen receptor-positive breast cancer. *Oncol. Rep.* **42**, 1413–1421 (2019).
78. Yang, L., Pang, Y. & Moses, H. L. TGF- β and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends Immunol.* **31**, 220–227 (2010).
79. Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst.* **6**, 612–620.e5 (2018).
80. Novo, D., Grégori, G. & Rajwa, B. Generalized Unmixing Model for Multispectral Flow Cytometry Utilizing Nonsquare Compensation Matrices. *Cytom. Part J. Int. Soc. Anal. Cytol.* **83**, 508–520 (2013).
81. van der Maaten, L. Barnes-Hut-SNE. *ArXiv13013342 Cs Stat* (2013).

82. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
83. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
84. Shih, A. J. *et al.* Identification of grade and origin specific cell populations in serous epithelial ovarian cancer by single cell RNA-seq. *PLOS ONE* **13**, e0206785 (2018).
85. Györfy, B., Lániczky, A. & Szállási, Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr. Relat. Cancer* **19**, 197–208 (2012).

Figures

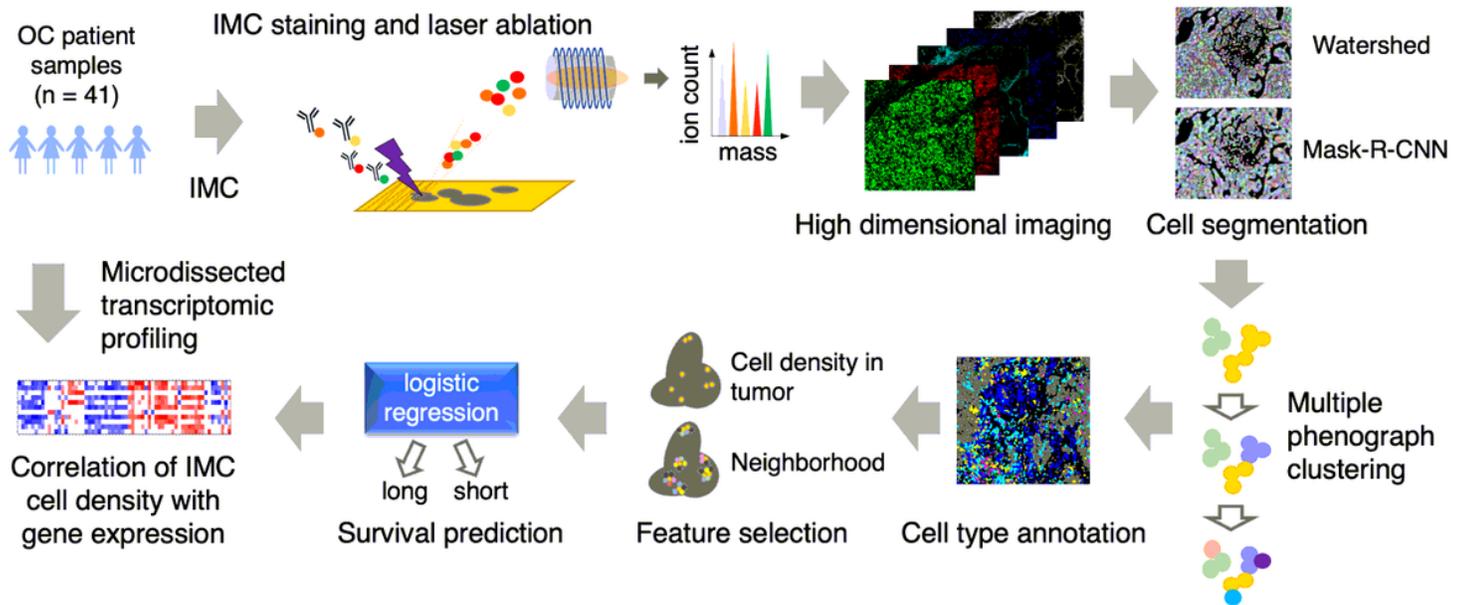


Figure 1

Schematic of imaging mass cytometry (IMC) acquisition of multiplexed images from 41 ovarian cancer patient samples and the data analytic pipeline. The pipeline includes cell segmentation by a mix of the mask-R-CNN method trained on the outputs of the watershed method, multiple rounds of phenograph clustering, cell subtype annotation and visualizations, cell density and nearest-neighborhood feature selection, survival prediction by logistic regression, and integration of IMC cell density with transcriptomes from microdissected epithelial and fibroblastic compartments of HGSC.

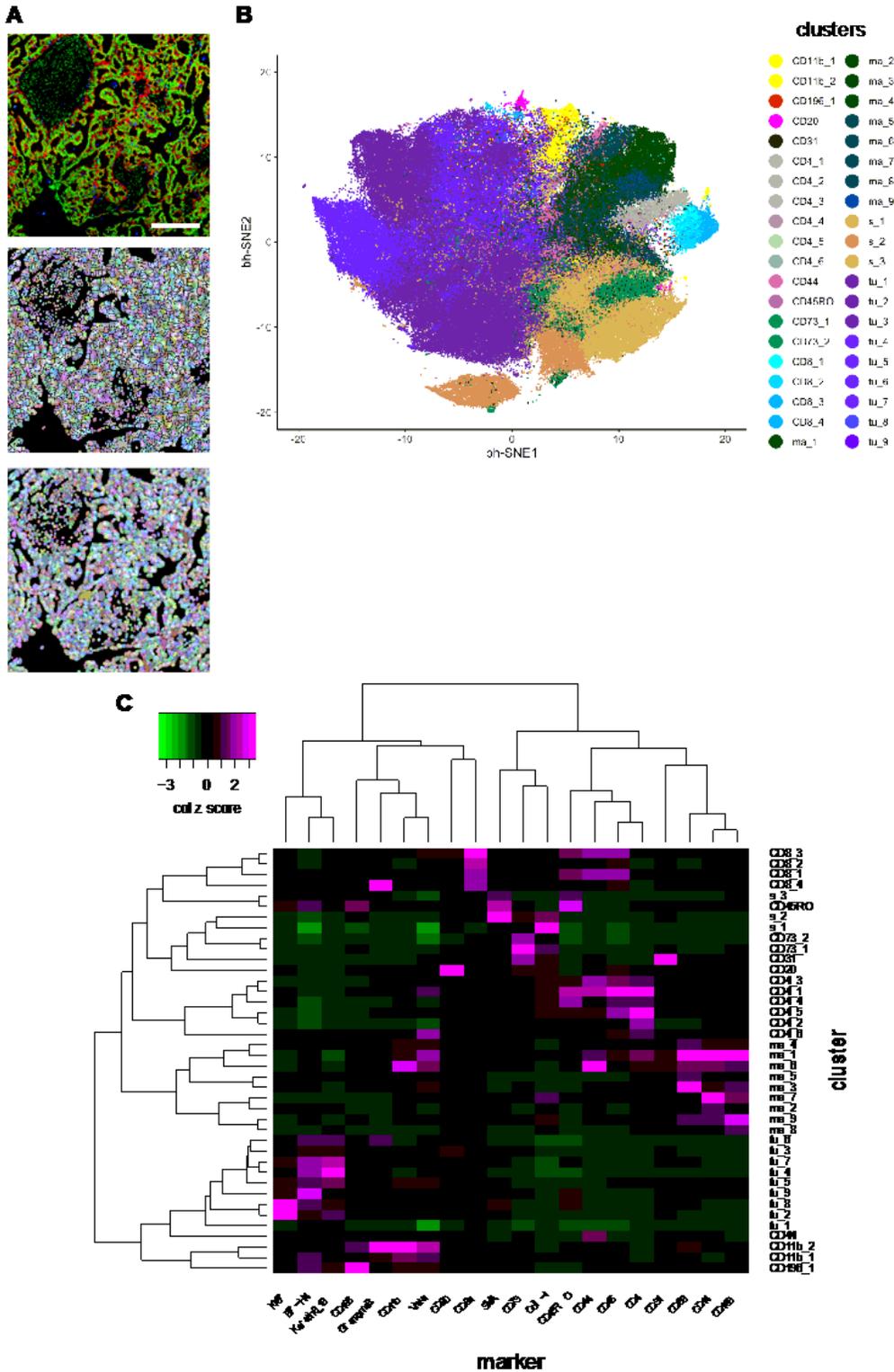


Figure 2

Cell segmentation by mask-R-CNN and cell subtype annotation based on marker expression. A. Comparison of watershed segmentation and mask-R-CNN segmentation. Top, imaging mass cytometry (IMC) image; red: Keratin8/18, green: histone H3, blue: CD68. Middle, watershed segmentation. Bottom, mask-R-CNN segmentation. White scale bar, 200 μ m. B. Two-dimensional bh-SNE representation of

ted by multiple rounds of phenograph

clustering. Each dot represents one cell. C. Heatmap showing the median marker expression (z-scored by column) of each cluster (cell subtype). Twenty markers and 40 clusters were ordered by hierarchical clustering with the distance based on Pearson correlation.

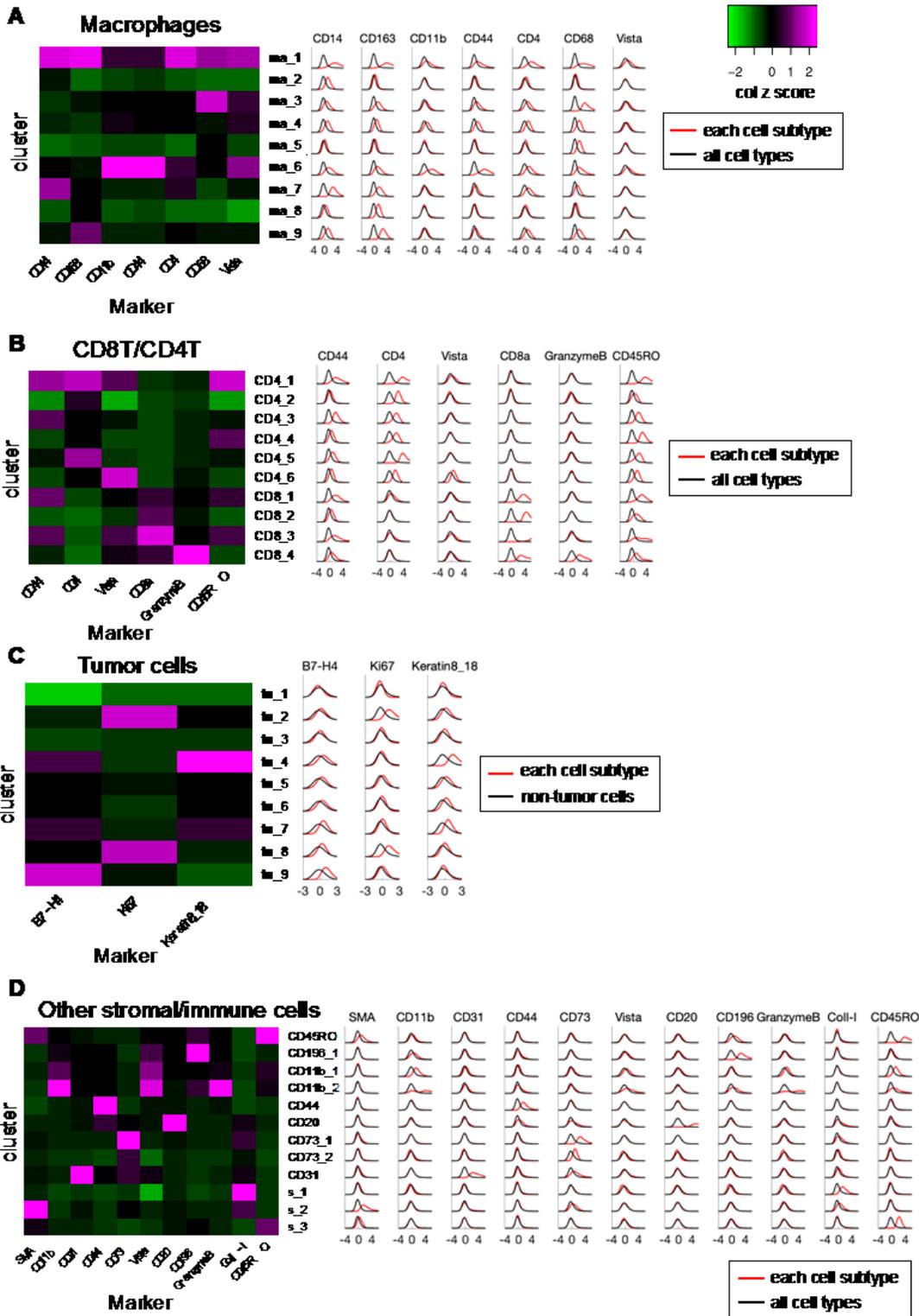


Figure 3

Marker expression levels of each major cell subtypes. Left, heatmaps of median marker expression level
 Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js ght, histograms of Gaussian-smoothed ($\sigma = 0.5$)

normalized marker expression levels of the cells of each cell subtype across all samples (red line) and normalized marker expression levels of the cells of all cell subtypes across all samples (black line, A, B, D) or normalized marker expression levels of the cells of non-tumor cells across all samples (black line, C). A. Macrophage subtypes. B. CD8+/CD4+ T cell subtypes. C. Tumor subtypes. D. Other immune cell subtypes.

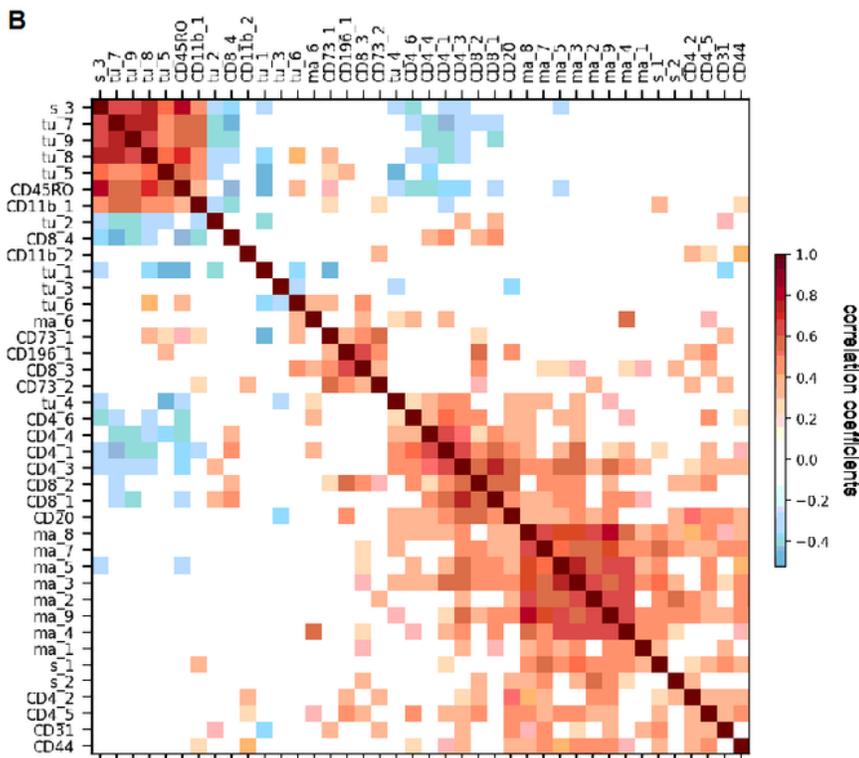
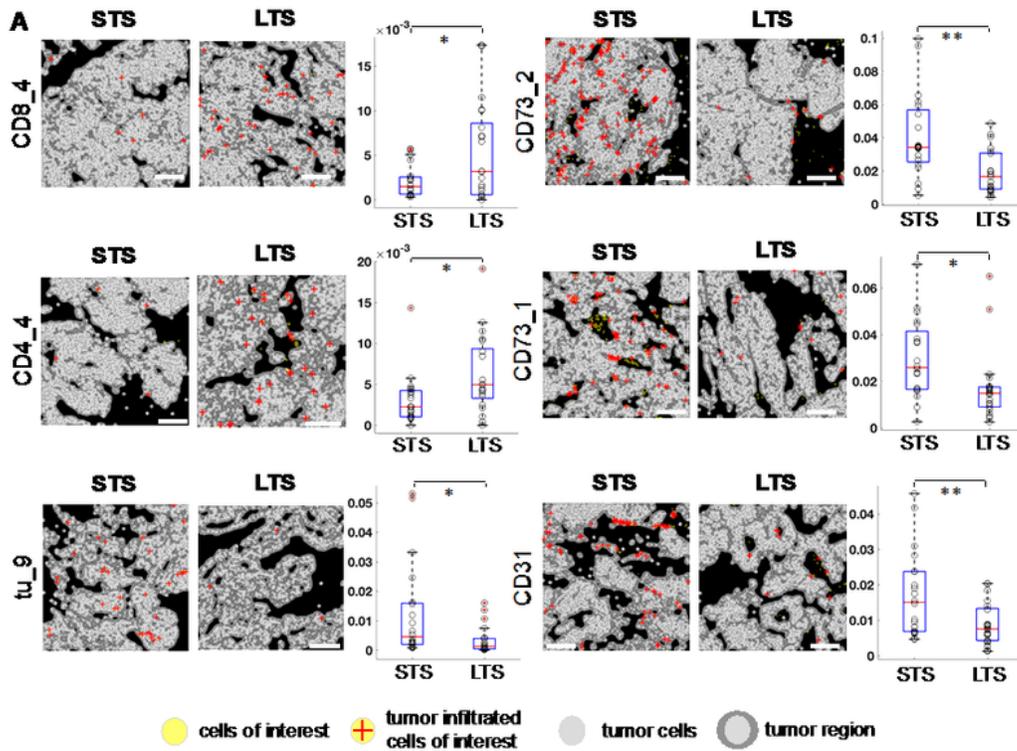


Figure 4

Comparison of nearest-neighbor cell-cell interactions between long-term survivors (LTS) and short-term survivors (STS) and feature selection for patient survival prediction. A. Map of significant increases and decreases in nearest-neighbor interactions, computed as the average number of cell subtype X in the nearest neighbor of cell subtype Y (distance between the center of X and Y less than 20 μm), that are significantly (Benjamini-Hochberg adjusted p value < 0.05) increased (magenta) or decreased (green) in LTS compared with STS. B. (a) Number of cell density features selected by recursive feature elimination as a function of training (red) or validation (blue) accuracy. The optimal number of features is indicated by the green dashed line. Validation was done by leave-one-out cross validation. (b) Receiver operating characteristic (ROC) curve for the test set. (c) Logistic regression coefficients of the features selected by the model. C. (a) Number of features of nearest-neighbor cell-cell interactions selected by recursive feature elimination as a function of training (red) or validation (blue) accuracy. The optimal number of features is indicated by the green dashed line. Validation was done by leave-one-out cross validation. (b) Receiver operating characteristic curve for the test set. (c) Logistic regression coefficients of the features selected by the model. D. Spearman correlation between the features that both correlate with patient survival (absolute correlation coefficient > 0.2). White color indicates correlation coefficient = 0 or $p \geq 0.05$. Neighborhood features were ordered by hierarchical clustering with the Ward method.

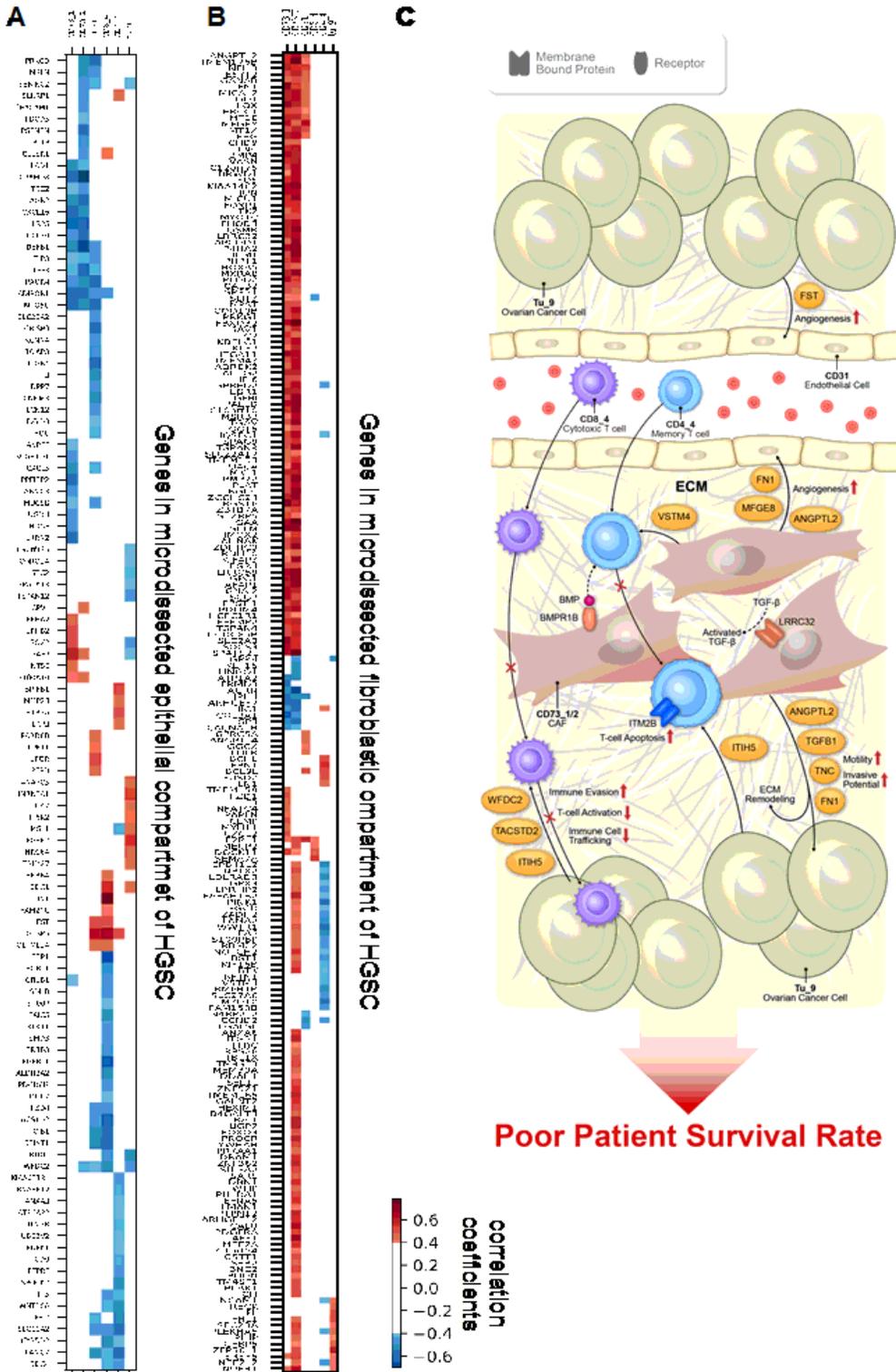


Figure 6

Gene correlation studies. Shown are correlations of gene expression in microdissected epithelial (A) and fibroblastic (B) components of high-grade serous ovarian cancer (HSGC) samples with the imaging mass cytometry cell density in tumor-enriched regions (n = 26 samples, Spearman correlation, p < 0.05). White color indicates absolute correlation coefficient ≤ 0.4 or p ≥ 0.05 . Genes were ordered by hierarchical

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js ing the genes that are significantly correlated

with prognostic IMC features to postulate the mechanisms by which these genes contribute to the prognostic features. Cancer cell-derived WFDC2, TACSTD2 and ITIH5 drive immune surveillance failure by compromising cytotoxic CD8+ T cells activity and infiltration. CAF-derived VSTM4 can reduce cytokine production by CD4 T cells and cause a decrease in T cell activation. CAFs expressing BMPR1B are more responsive to BMP signaling, which subsequently modulates the rigidity of CAFs, and reduces CD4_4 T cell trafficking. CAF-derived LRRC32 suppresses anticancer immune cell infiltration by modulating TGF- β signaling networks. ANGPTL2, TNC, TGFB1, FN1, BMPR1B, and LRRC32 promote tumor progression, angiogenesis, and ECM remodeling.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [IMCsupplementarydata.docx](#)