



OPEN

Cancer-associated fibroblasts are the main contributors to epithelial-to-mesenchymal signatures in the tumor microenvironment

Peter M. Szabo^{1,2}, Amir Vajdi^{3,4}, Namit Kumar⁵✉, Michael Y. Tolstorukov³, Benjamin J. Chen⁶, Robin Edwards^{1,7}, Keith L. Ligon³, Scott D. Chasalow¹, Kin-Hoe Chow³, Aniket Shetty³, Mohan Bolisetty¹, James L. Holloway⁸, Ryan Golhar¹, Brian A. Kidd⁹, Philip Ansumana Hull⁹, Jeff Houser⁹, Logan Vlach^{9,10}, Nathan O. Siemers^{1,11} & Saurabh Saha^{1,12}

Epithelial-to-mesenchymal transition (EMT) is associated with tumor initiation, metastasis, and drug resistance. However, the mechanisms underlying these associations are largely unknown. We studied several tumor types to identify the source of EMT gene expression signals and a potential mechanism of resistance to immuno-oncology treatment. Across tumor types, EMT-related gene expression was strongly associated with expression of stroma-related genes. Based on RNA sequencing of multiple patient-derived xenograft models, EMT-related gene expression was enriched in the stroma versus parenchyma. EMT-related markers were predominantly expressed by cancer-associated fibroblasts (CAFs), cells of mesenchymal origin which produce a variety of matrix proteins and growth factors. Scores derived from a 3-gene CAF transcriptional signature (*COL1A1*, *COL1A2*, *COL3A1*) were sufficient to reproduce association between EMT-related markers and disease prognosis. Our results suggest that CAFs are the primary source of EMT signaling and have potential roles as biomarkers and targets for immuno-oncology therapies.

Epithelial-to-mesenchymal transition (EMT) is a process of tissue dedifferentiation occurring during development and tissue repair that results in epithelial cells transitioning into cells with a more mesenchymal phenotype^{1–4}. Depending on the contextual signals received by the cell within a tissue, a series of intermediate phenotypic states may be generated along the epithelial-mesenchymal spectrum up to a fully mesenchymal cell state⁴. Tumor cells may induce this transition to drive tumor growth and initiate metastasis⁵. Given the important role of this process in cancer progression, a variety of gene expression signatures that characterize EMT have been identified^{6–11}.

EMT-related gene expression has been shown to impact T-cell function and infiltration into the tumor parenchyma¹⁰. The extent of lymphocyte infiltration is influenced by stromal factors within the tumor microenvironment (TME), such as transforming growth factor beta (TGF β). TGF β signaling has been shown to inhibit T-cell proliferation and regulate T-cell effector functions, and it has been implicated as a contributor to the acquisition of a cancer-associated fibroblast (CAF) phenotype in stromal cells^{12–15}. CAFs have been suggested to originate from mesenchymal stromal cells, among others¹⁶, and shown to exert diverse tumor-promoting functions in the context of the TME, notably tumorigenesis, cancer cell proliferation and metastasis, extracellular matrix remodeling, drug resistance, immune evasion, and immunosuppression^{14,17–21}. Additionally, TGF β signaling and collagen production by CAFs²² may contribute to the exclusion of T cells from the tumor parenchyma²³.

¹Bristol Myers Squibb, Princeton, NJ, USA. ²Present address: Fate Therapeutics, San Diego, CA, USA. ³Dana-Farber Cancer Institute, Boston, MA, USA. ⁴Present address: Merck & Co., Inc., Kenilworth, NJ, USA. ⁵Bristol Myers Squibb, San Diego, CA, USA. ⁶Bristol Myers Squibb, Cambridge, MA, USA. ⁷Present address: Daiichi Sankyo, Inc., Princeton, NJ, USA. ⁸Bristol Myers Squibb, Seattle, WA, USA. ⁹Bristol Myers Squibb, Redwood City, CA, USA. ¹⁰Present address: Vanderbilt University, Nashville, TN, USA. ¹¹Present address: Fiveprime Group, Monterey, CA, USA. ¹²Present address: Centessa Pharmaceuticals, Cambridge, MA, USA. ✉email: Namit.kumar@bms.com

Previous studies have found that tumors with high baseline expression levels of immune-related genes in the TME were more likely to respond favorably to immuno-oncology (I-O) therapy^{24–26}. In tumors with high levels of inflammation, gene expression signature scores indicative of lower EMT and stromal signaling have been shown to be associated with increased response rates and prolonged survival in patients with urothelial carcinoma (UC) treated with the programmed death 1 (PD-1) inhibitor nivolumab^{10,27}. In patient-derived xenograft (PDX) models of UC, EMT-related gene signatures were shown to be strongly correlated with stroma-related gene signatures, implying that stromal cells may be a key source of EMT-related gene expression¹⁰. Of note, these models have been shown to have many advantages compared with other models, such as retaining the characteristics of the donor tumor, including tissue structure and gene expression profiles, and therefore are potentially predictive of human tumor biology and patient response to treatment^{28–33}. Separate work done in colorectal cancer (CRC) using single-cell RNA sequencing (scRNA-seq) demonstrated that tumors with a strong epithelial signature and weak fibroblast and myeloid signatures had the best survival rate³⁴. Notably, in this study, expression of EMT-related genes was significantly increased in CAFs³⁴.

The role of EMT signaling has been extensively studied using preclinical models of individual tumor types. While previous evidence suggests that EMT signaling is emanating from cells within the stromal compartment¹⁰, there is ongoing research into the role of specific cell types, such as CAFs, in the EMT process and their associations with factors that can influence response to I-O therapy. In the current study, bulk and single-cell RNA sequencing analysis and immunohistochemistry (IHC)-based methods were used to investigate the source of EMT- and stroma-related gene expression across multiple tumor types and to determine factors in the stromal compartment of the TME that may associate with prognosis or response to treatment.

Results

Correlation between EMT- and TME-related gene expression signatures. In RNA-seq data for 17 tumor types in The Cancer Genome Atlas (TCGA), scores for 7 of 8 previously published EMT gene signatures (EMT_Carretero_2010, IPRES_Hugo_2016, PATHWAY_Carretero_2010, EMT_STROMA_Wang_2018, GENERIC_EMT_MES_Tan_2014, HALLMARK_EMT, and PAN_EMT_Mak_2016)^{6–11} were negatively correlated with scores for a tumor purity signature and positively correlated with scores for immune and stroma signatures (Fig. 1a, Supplementary Fig. 1, and Supplementary Table 1). Correlations tended to be smaller and in the opposite direction for the GENERIC_EMT_EPI_Tan_2014⁷ gene set, which is characteristic of epithelial features rather than mesenchymal features (Fig. 1a). A strong positive correlation was observed between EMT gene set signature scores and estimates of stromal content in samples from TCGA across all tumor types with epithelial origin, including colon, stomach, urothelial, and breast (Fig. 1b, Supplementary Fig. 1, and Supplementary Table 1). In tumors with a mesenchymal origin, such as sarcoma, these correlations were weaker (Supplementary Fig. 1n).

EMT-related genes enriched in the stromal compartment of PDX samples. Given the above findings, we sought to elucidate the source of EMT-related gene expression in the tumor parenchyma or stromal cells. Expression of genes in each of the 8 selected EMT gene sets and the ESTIMATE_STROMA gene set^{6–11} were analyzed in RNA-seq data from PDX samples representing individual tumor types and containing tissue derived from the stromal (mouse) and tumor (human) compartments of the TME. While variability was observed across tumor types in the proportion of EMT-related gene expression derived from the stroma, in many tumor types, a substantial proportion of transcripts of EMT-related genes were derived from stromal tissue (Fig. 2, Supplementary Fig. 2, and Supplementary Table 2). The exception was the GENERIC_EMT_EPI_Tan_2014 gene set (Supplementary Fig. 2f.), which represents epithelial cells before transition to mesenchymal cells; it is expected that these genes would originate from the tumor parenchymal tissue. Evidence of stromal origin for most of the transcripts in the EMT gene sets was not observed for ovarian carcinosarcoma or uterine carcinoma (Fig. 2c and Supplementary Fig. 2), likely due to their mesenchymal origin or presence of intrinsic mesenchymal components. Despite this, we detected significant enrichment of stromal transcripts in the majority of EMT gene sets (Supplementary Table 2), suggesting measurable contributions from stromal cells in gene expression estimates in these cancer types. As a positive control, a large majority of transcripts of genes in the ESTIMATE_STROMA gene set were derived from stromal tissue for all tumor types, including ovarian carcinosarcoma (Supplementary Fig. 2e).

EMT-related genes expressed by CAFs, endothelial cells, and myeloid progenitor cells. The types of cells in samples from 14 treatment-naïve patients with colorectal or renal carcinoma were identified using canonical correlation analysis of scRNA-seq profiles, shown in Fig. 3a as a two-dimensional projection. After filtering for poor quality, a total of 56,667 cells across 32 resected tumor and adjacent normal tissue samples were selected for downstream analysis. EMT signature gene expression was scored for each scRNA-seq dataset and applied to the projection to identify cell types with enriched expression of EMT signatures. The greatest fold enrichment was observed in fibroblasts; other enrichments were observed to a lesser extent in adipocytes, epithelial, endothelial, and myeloid progenitor cells (example shown in Fig. 3b,c for the HALLMARK_EMT gene set).

Median scores for the EMT-related gene sets used in this study were determined in the scRNA-seq datasets identified as fibroblasts or as epithelial tumor parenchymal cells. Across a select range of EMT-related gene sets, the median scores for the EMT signatures (with the exception of those in the GENERIC_EMT_EPI_Tan_2014 gene set) were greater ($p < 0.01$) in fibroblasts found in the stroma than in tumor parenchymal cells (Supplementary Fig. 3).

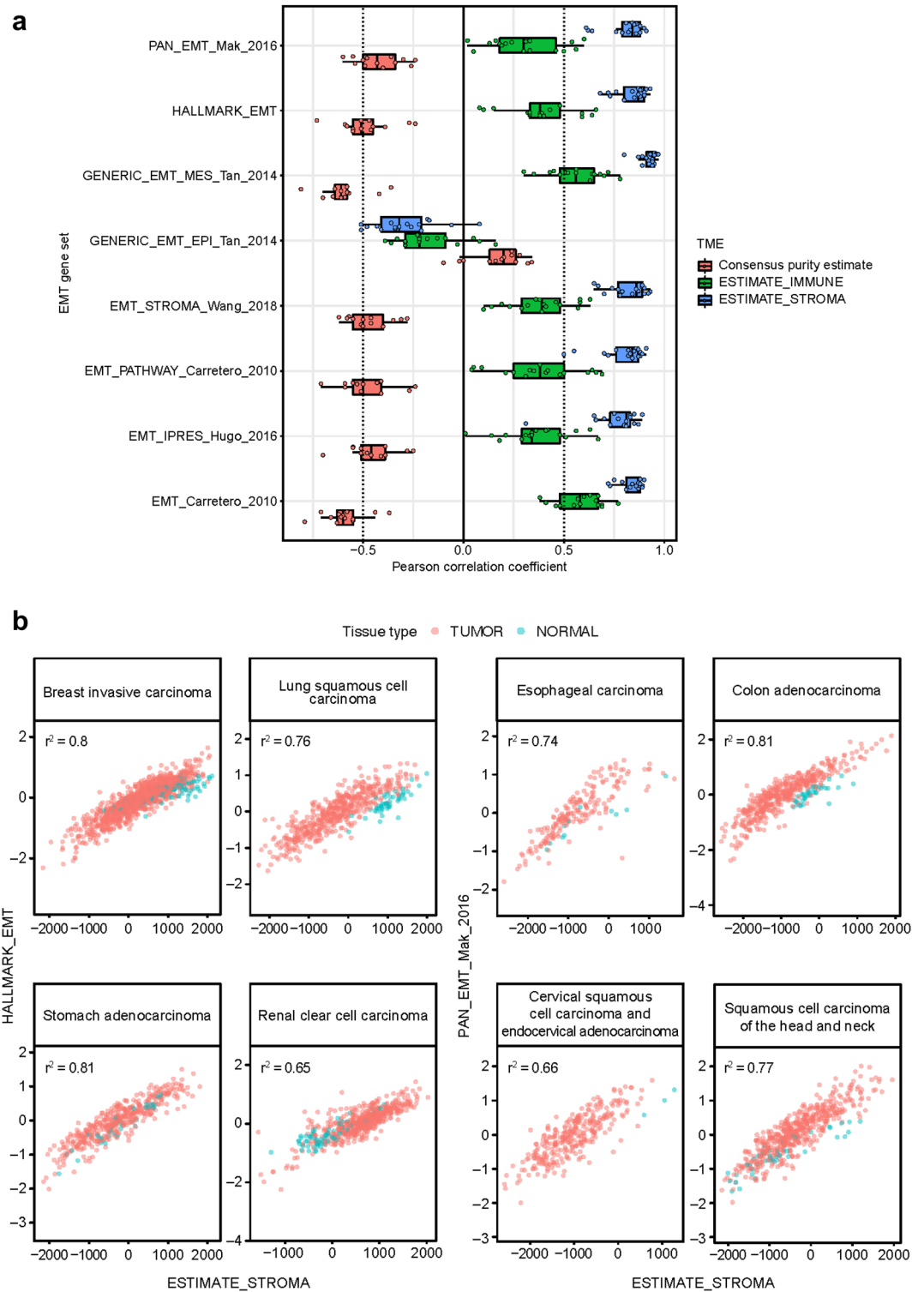


Figure 1. EMT signatures and the TME. (a) Distribution over tumor types in TCGA of estimated correlation (Pearson coefficients) between scores for EMT-related gene sets and tumor purity, an immune gene expression signature, and a stroma signature. Boxes extend from the first to third quartiles, the middle line shows the median, and the whiskers extend to the most extreme data point that is no more than 1.5 times the IQR from the box. Consensus purity estimates were not available for esophageal carcinoma, pancreatic and stomach adenocarcinoma, and sarcoma. (b) Scatterplots of scores for two EMT-related gene signatures versus a stroma signature, for selected tumor types in TCGA. Pink plotting symbols represent tumor tissue. Aqua plotting symbols represent samples collected from non-involved healthy tissue of patients with cancer in TCGA.

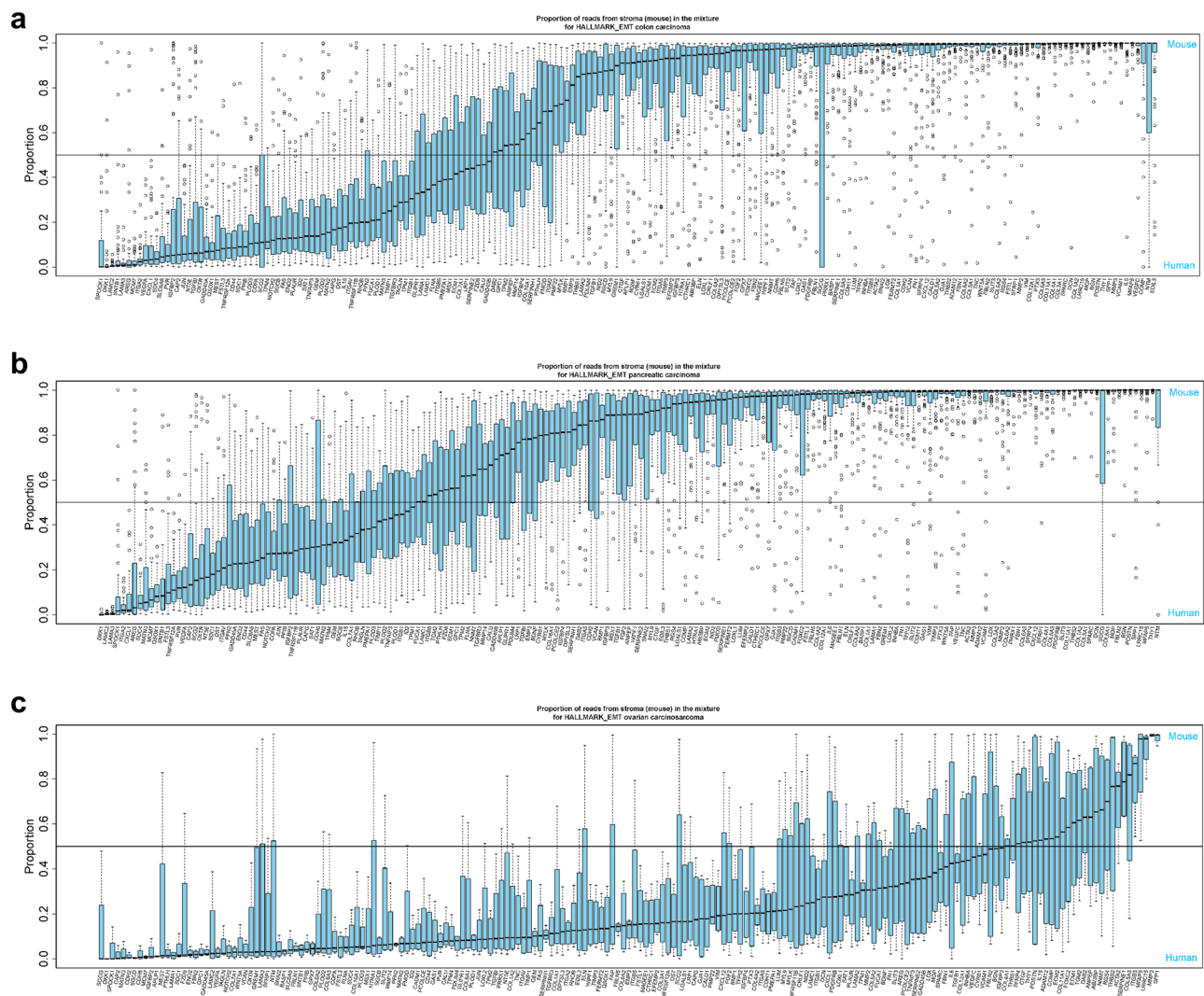


Figure 2. Expression of HALLMARK_EMT-related genes in mouse (stromal) and human (tumor parenchymal) tissues using PDX samples. Horizontal reference line indicates proportion of reads from stroma = 0.5 (signifying equal contribution from stromal and cancer cells to the estimated gene expression). Boxes extend from the first to third quartiles, the middle line shows the median, and the whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range from the box, open circles show individual values that are more than 1.5 times the interquartile range from the box. (a) Colon ($n = 70$). (b) Pancreatic ($n = 57$). (c) Ovarian carcinosarcoma ($n = 4$). Additional tumor types and gene sets are shown in Supplementary Fig. 2. These figures demonstrate that, for multiple tumor types, a greater proportion of transcripts of EMT-related genes were derived from stromal tissue than from tumor tissue, with the exception being the GENERIC_EMT_EPI_Tan_2014 gene set (shown in Supplementary Fig. 2f.).

Strong positive correlation between EMT and CAF gene expression signature scores and association with overall survival (OS). To further support the hypothesis that CAFs are a major source of EMT gene expression, we developed a “CAF signature” derived from the expression of 3 collagen genes (*COL1A1*, *COL1A2*, and *COL3A1*), as described in the Methods, and assessed the correlation between EMT-related and CAF gene expression in samples from TCGA. Across multiple tumor types, a strong positive correlation was observed between scores for all the EMT gene sets (except GENERIC_EMT_EPI_Tan_2014) and CAF signature scores (Fig. 4, Supplementary Fig. 4, Supplementary Table 3). As expected, the positive correlations were somewhat weaker for sarcoma (Supplementary Fig. 4n, Supplementary Table 3).

In these same tumor types, lower CAF signature scores tended to be associated with prolonged survival (Fig. 5). As expected, given the strong positive correlations observed, associations between OS and CAF signature scores were similar in magnitude and direction to associations between OS and EMT signature scores (Supplementary Table 4). This is consistent with previous literature regarding associations between OS and EMT gene expression¹⁰.

Collagen scores inversely associated with parenchymal T-cell counts in squamous cell carcinoma of the head and neck (SCCHN). We then sought to investigate possible mechanisms behind the

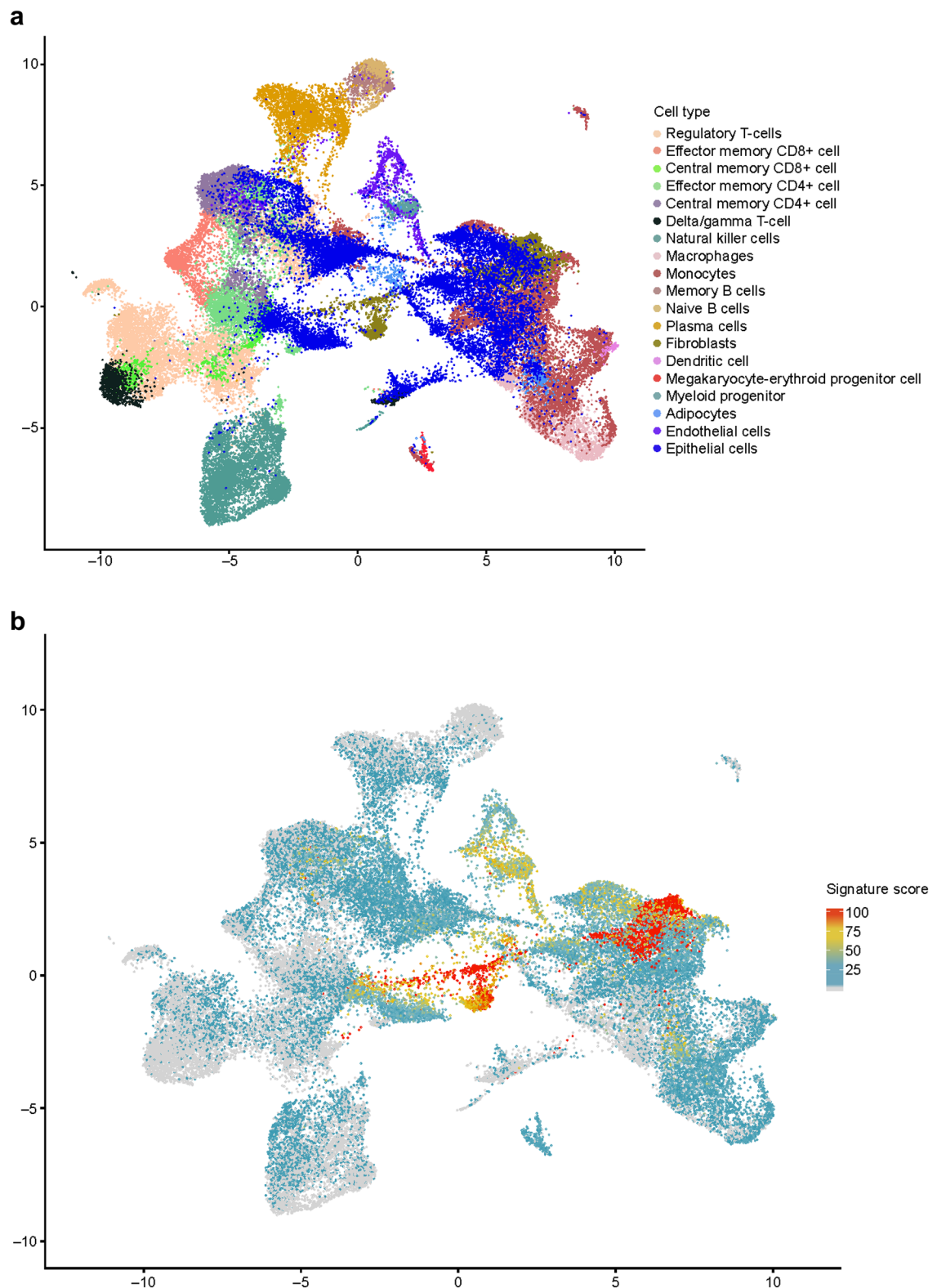


Figure 3. scRNA-seq of tumor samples. **(a)** Identification of cell type in individual scRNA-seq datasets displayed as a 2-dimensional projection. **(b)** Location and expression level of HALLMARK_EMT signature genes in scRNA-seq datasets displayed as a 2-dimensional projection. **(c)** HALLMARK_EMT-related gene expression by cell type. In **(b)** and **(c)**, the gene signature scores (0–100) represent the gene expression level for the EMT signatures in the scRNA-seq data. Analysis of additional EMT gene set signatures showed greater fold enrichment in fibroblasts and other non-malignant stromal cells than from tumor tissue, with the exception being the GENERIC_EMT_EPI_Tan_2014 gene set (data not shown).

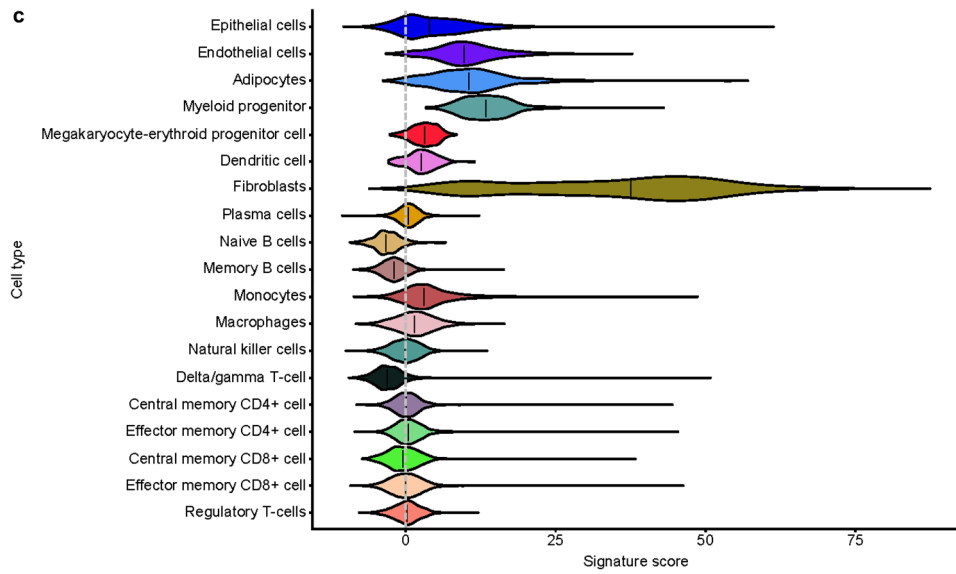


Figure 3. (continued)

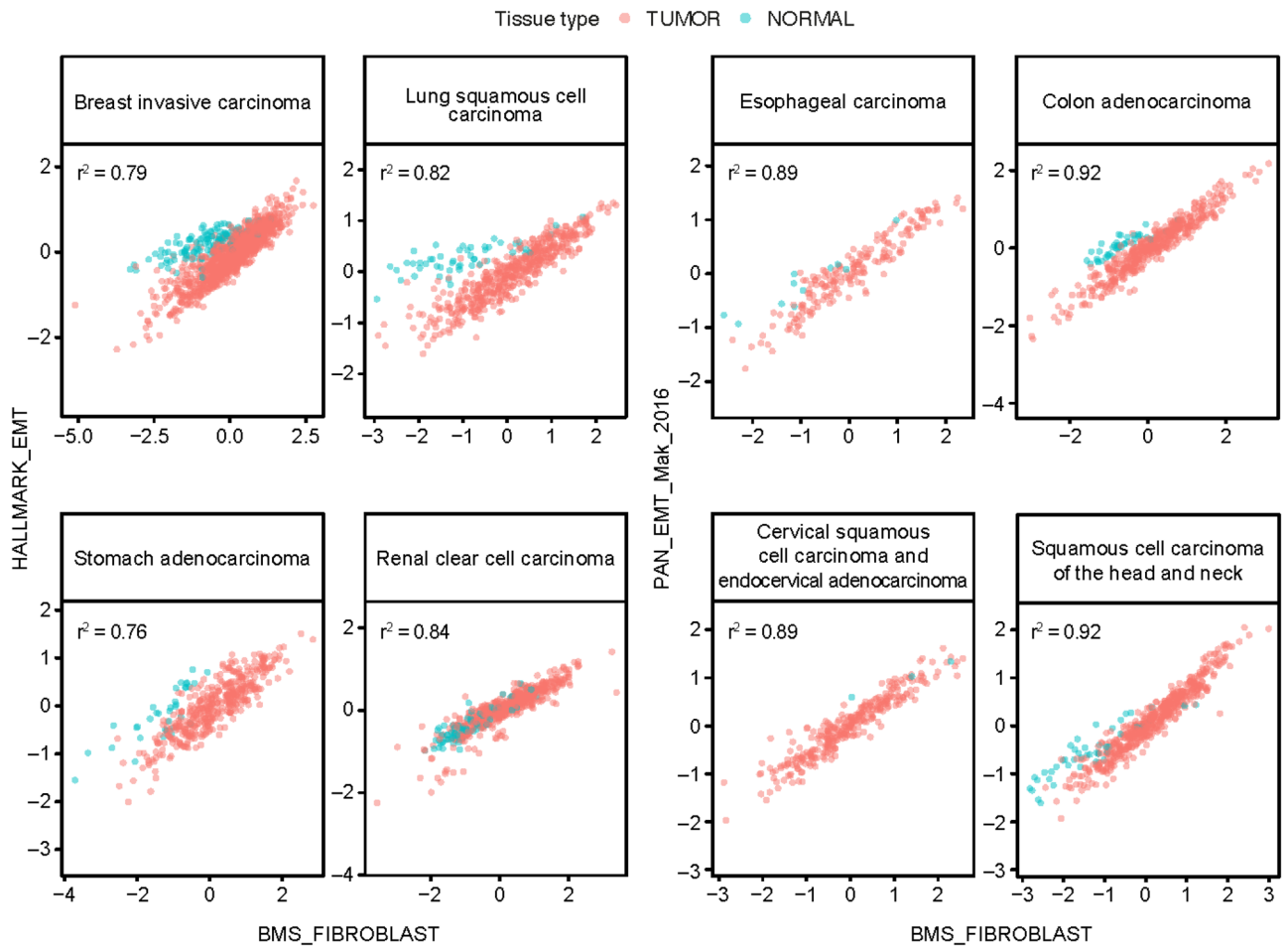


Figure 4. Correlation between EMT-related and CAF (BMS_FIBROBLAST) signature scores in tumor datasets from TCGA. Pink plotting symbols represent tumor tissue. Aqua plotting symbols represent samples collected from non-involved healthy tissue of patients with cancer from TCGA.

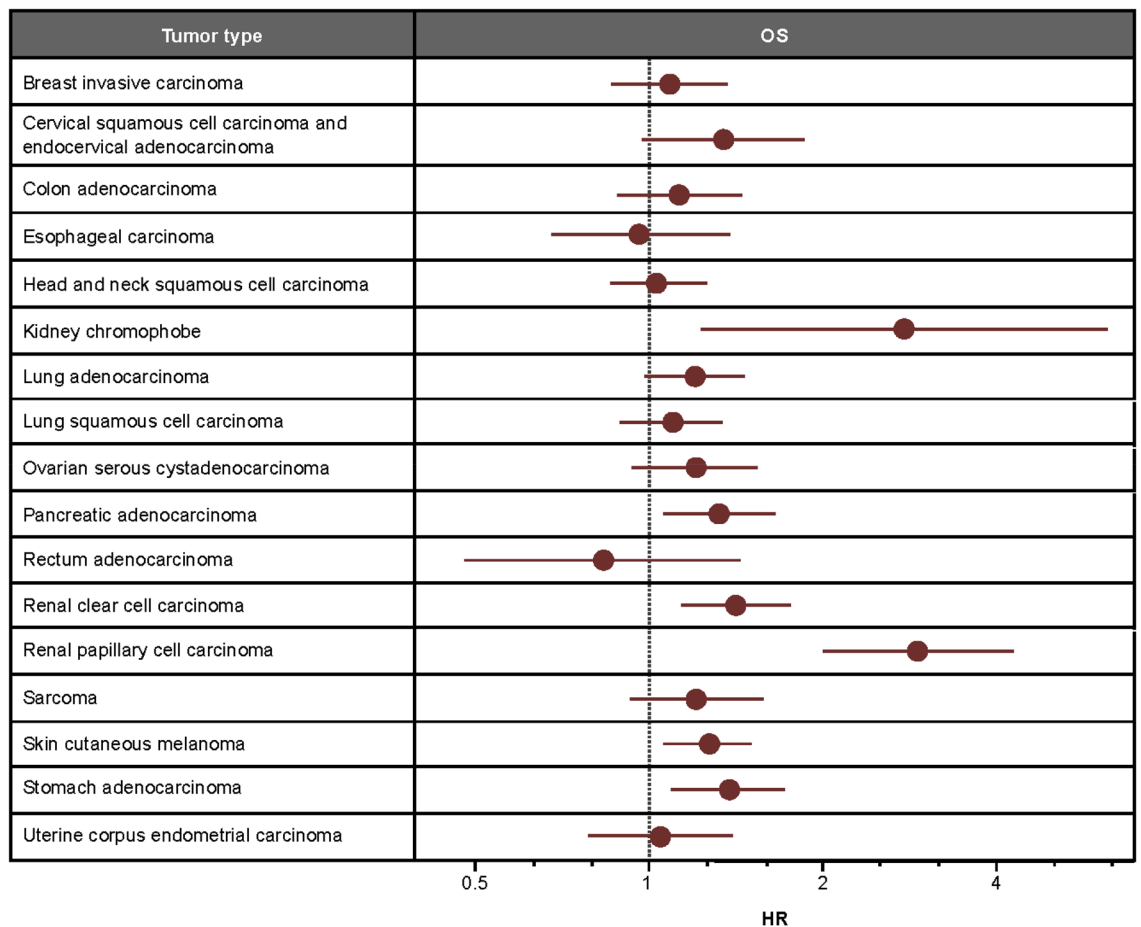


Figure 5. Association between CAF signature scores and OS. HRs were estimated from TCGA datasets using Cox proportional hazards models assuming a linear effect of signature scores. Estimated HRs (high signature score/low signature score) were scaled to compare hazards for scores differing by the IQR for each tumor type.

link between CAF scores and prognosis. Using IHC, we studied the relationship between collagen abundance in the tumor stroma and exclusion of CD8+ T cells from the tumor parenchyma in SCCHN tumor specimens. Low stromal collagen scores were observed in samples with high parenchymal CD8+ T-cell counts, while high collagen scores were observed in samples with low levels of parenchymal CD8+ T cells. Samples with low levels of both stroma collagen and parenchymal CD8+ T cells were also observed; however, no sample had high levels of both stroma collagen and parenchymal CD8+ T cells (Fig. 6). Representative IHC images are shown in Supplementary Fig. 5.

Correlation between TGF β - and TME-related gene expression signatures. Given the known role of TGF β signaling in the acquisition of a CAF phenotype¹⁴, we analyzed various TGF β -related signatures, including the Hallmark_TGF_BETA_SIGNALING and GO_TRANSFORMING_GROWTH_FACTOR_BETA_BINDING gene sets using RNA-seq data from TCGA. These signatures were obtained from previously published literature^{11,35–38} or retrieved from an online database as described in the methods. These signatures were negatively correlated with tumor purity and positively correlated with stromal signature scores (Supplementary Fig. 6), similar to the patterns observed for the EMT signature scores. A smaller positive correlation between TGF β -related signatures and immune signature scores was also observed (Supplementary Fig. 6).

To further investigate the cellular source of TGF β signaling, expression scores were determined for the TGF β -binding protein gene signatures and were applied to our scRNA-seq projections to identify which cell types were enriched for expression of TGF β -related genes (Fig. 3a). Expression of TGF β -related genes was strongly enriched in myeloid progenitor cells, endothelial cells, and CAFs, relative to other cell types (Supplementary Fig. 7a, b). Similar to our results regarding EMT-related gene expression, TGF β -related gene expression was greater on average in stromal cells than in parenchymal cells.

Discussion

EMT is a critical step for cancer cell metastasis, drug resistance, and immune evasion². In this study, we sought to determine the source of EMT-related gene expression within the TME. A graphical representation of our overall findings is presented in Fig. 7. According to prior literature supporting tumor cells as the source of

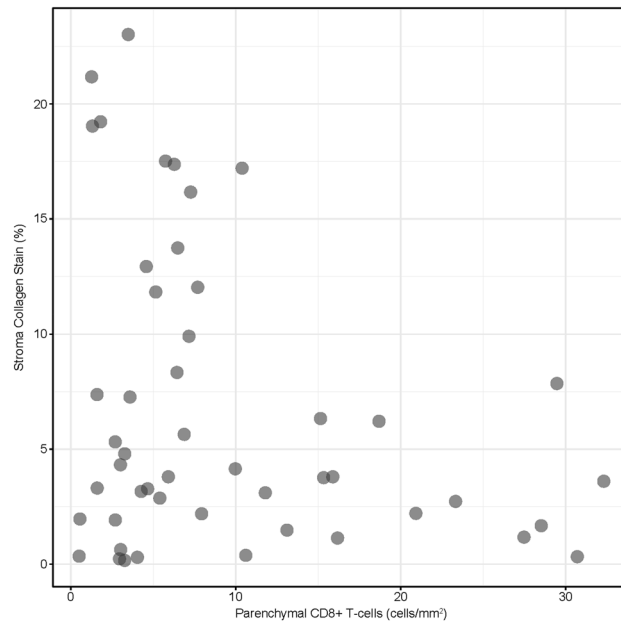


Figure 6. Association between collagen and parenchymal CD8+ T-cell abundance in SCCHN tumor samples. Collagen staining was assessed by image analysis to identify tumor parenchymal and stromal compartments and calculate the % area of collagen staining relative to the area of the total stromal compartment. An AI-powered image analysis algorithm was used to classify the tumor parenchymal and stromal regions and identify and quantify the density (cells/mm²) of CD8+ T cells in each compartment.

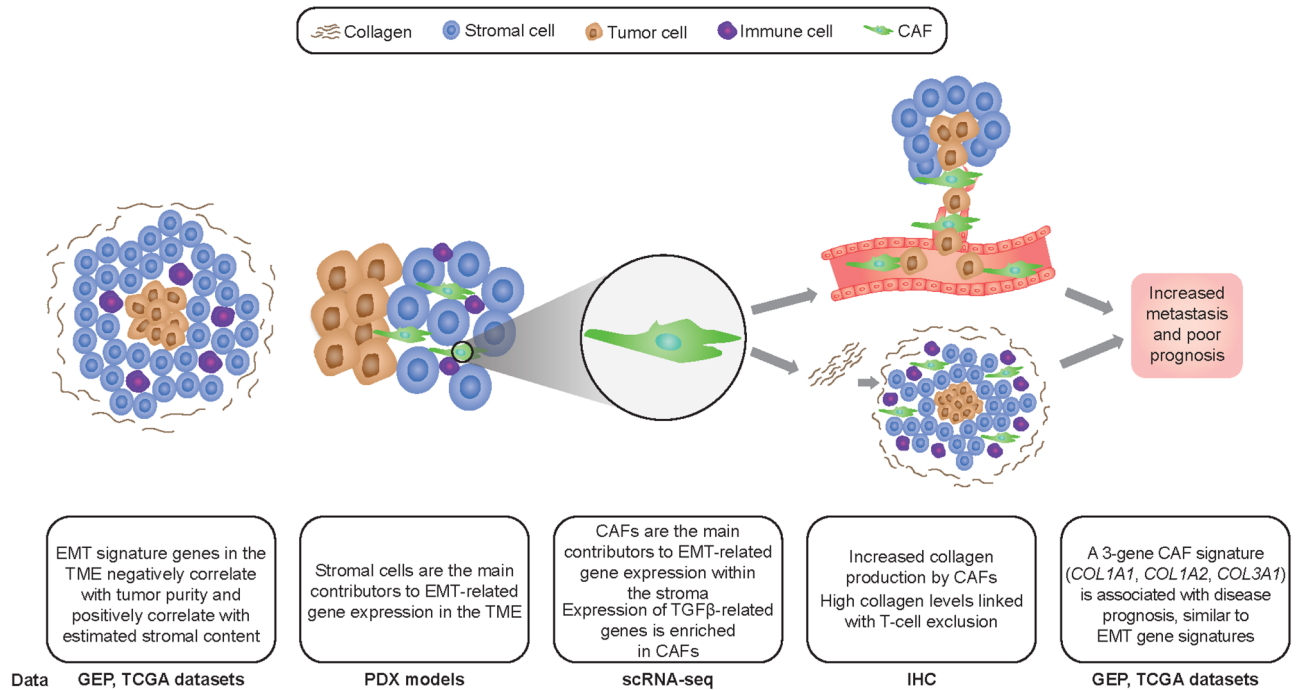


Figure 7. Overview of the main findings from this study. Top images provide graphical representation of the TME. Bottom text describes the main results and corresponding experimental datasets and methods. GEP, gene expression profiling.

EMT-related gene expression³⁹, one would assume tumors become more resistant to immune cell infiltration as they go through the EMT process and begin to express mesenchymal genes. One would therefore expect expression of EMT genes to be positively correlated with the number of cancer cells present in the tumor and negatively correlated with immune and other stromal cell content. Instead, using gene expression profiling of bulk tumor tissues, the opposite was observed across a variety of tumor types: a strong negative correlation between multiple EMT gene set signatures and proportion of cancer cells, and a strong positive correlation between these same gene set signatures and stromal content. This suggests that EMT gene signatures largely are not measuring tumor cell-related mechanisms, but rather the amount of stromal content in the TME. The exception was the *GENERIC_EMT_EPI_Tan_2014* gene set, which did not show strong correlations with signatures of tumor purity, stroma, or immune components of the TME. This could be because, in our study, most of the analyzed EMT gene sets measured mesenchymal features of the tumor cells, whereas this signature was chosen by Tan et al. to represent polarized epithelial cells before the transition to mesenchymal cells⁷. Tumor cells with epithelial origins, such as carcinomas, already express epithelial markers and could therefore display positive correlations with the *GENERIC_EMT_EPI_Tan_2014* gene set and tumor purity.

Based on these findings, we hypothesized that stromal elements are the key contributors to EMT expression in the TME. Using a species-specific approach to parse mouse and human reads in a pan-cancer collection of gene expression data from xenograft samples, we demonstrated that, in the majority of tumor types, a greater proportion of transcripts from EMT-related genes originated from mouse stromal cells than from human tumor cells. In cancers known to have mesenchymal lineage components, such as ovarian carcinosarcomas⁴⁰, expression of EMT-related genes did not originate predominantly from mouse stromal regions. Analysis of individual cell transcriptomes from tumor samples by scRNA-seq demonstrated that these EMT-related genes were highly expressed by CAFs. Our results are consistent with other recent studies examining EMT and CAFs using similar methods. In a study by Puram et al., scRNA-seq was used to generate profiles for 18 SCCHN primary tumors⁴¹. Similar to our study, genes used to identify a mesenchymal phenotype were found to be expressed by CAFs. This was further supported by Zhang et al., who used two scRNA-seq platforms to classify immune and stromal cell populations and evaluate cell-cell interactions in CRC tumors and adjacent normal tissues. In addition to CAFs being enriched in the stromal tissue compared with normal tissue, certain subsets of tumor-associated macrophages were found to preferentially bind to CAFs and endothelial cells leading to activation of *MMP2*, a gene associated with tumor growth and metastasis⁴². These interactions highlight a potential mechanism by which CAFs can be regulated to drive tumor growth and metastasis.

To further validate CAFs as the source of EMT-related gene expression, we used fibroblast activation protein (previously established as a CAF-specific marker gene) as a sentinel marker and collagen genes, representing one of the fundamental functions of CAFs (extracellular matrix production), to generate a 3-gene CAF signature. Using collagen genes in preference to other genes associated with the CAF phenotype, such as α -smooth muscle actin, platelet-derived growth factor receptor β , and secreted protein acidic and rich in cysteine (SPARC)¹⁵, allowed us to use histological techniques (e.g. collagen quantitation) to investigate the biological mechanisms underlying the expression of EMT-related genes by CAFs. High correlation was observed between the EMT gene sets and our CAF signature when compared in bulk tumor tissue from TCGA, providing additional evidence that CAFs, rather than epithelial cells, are the primary source of EMT-related gene expression. Given the role of EMT in metastasis, drug resistance, and immune evasion, it is important to determine any clinical implications of EMT signature scores on patient survival and possible biological mechanisms behind these findings. Having observed a correlation between EMT and CAF signatures, we performed a retrospective analysis of patient data in TCGA. We found that CAF signature scores, similar to EMT signature scores, were negatively associated with overall survival, demonstrating the prognostic value of the CAF signature.

Preliminary exploration of the biological mechanisms involved in the association of EMT signature scores with immune cell infiltrates and signatures have shown that accumulation of CAFs in the tumor stroma is associated with decreased CD8+ T-cell infiltration, while the depletion of CAFs increases intratumoral T-cell concentrations^{43–45}. Our results in SCCHN showed that collagen, produced by CAFs in the TME²², was also associated with T-cell exclusion and may thereby contribute to resistance to I-O therapy. Similar to these results, high expression of CAF and EMT-related genes was associated with T-cell exclusion in UC¹⁰, ovarian cancer⁴⁶, and hepatocellular carcinoma⁴⁷. These findings support the possibility of using trichrome staining for collagen quantitation as a surrogate for CAF content.

Given that TGF β signaling is the most well-characterized pathway known to induce EMT, acts through various intracellular messengers to drive immunosuppression, and is one of the main regulators of collagen synthesis, fibrosis, and conversion of normal fibroblasts to a CAF phenotype^{15,48–50}, we investigated TGF β -related gene expression in the stromal and immune compartments of the TME. We demonstrated that genes coding for TGF β -binding proteins are expressed by multiple cells within the stromal region, including endothelial cells, CAFs, and myeloid progenitors. These results are supported by previous studies examining TGF β signaling in stromal cells. In a study using UC tumors from a large cohort of patients treated with the anti-PD-1 agent atezolizumab, response was associated with CD8+ T-cell prevalence, while lack of response was associated with TGF β signaling in fibroblasts and T-cell exclusion²³. Along with our results, this suggests a possible role for combining TGF β inhibitors with immune checkpoint inhibitors to increase clinical benefit in patients with solid tumors.

There are some potential limitations to our study. Although our robust xenograft model provides further evidence of EMT signaling in the stromal region, interpretation is limited by the potential lack of a normal complement of immune cells in immunocompromised murine hosts. However, our scRNA-seq results demonstrated that EMT gene expression was greater in fibroblasts than in immune cells, suggesting that this is not a major issue. The lack of a comparator arm in our survival analysis using samples from TCGA limits the ability to determine a predictive versus prognostic nature of the CAF signature score, and prospective validation of these results will be required. It should also be noted that prior studies in solid tumors have shown that tumor cells at the invasive

front also display mesenchymal features and may potentially be another source of EMT expression^{41,51}. This type of signaling may be diminished when using bulk RNA-seq methods but may explain the observed moderate increase in EMT expression in a subset of epithelial cells in our scRNA-seq analysis.

The results of our study provide comprehensive evidence that CAFs and other stromal cells, rather than tumor parenchymal cells as previously thought, are the predominant sources of EMT-related gene expression in the TME and are major contributors to assessments of EMT gene signatures derived from bulk tumor tissue across a broad range of tumor types. These novel findings align with previous literature showing that upregulation of EMT-related transcription factors in CAFs leads to promotion of tumor growth and metastasis^{52–54}. These results underscore the potential of studying CAFs as biomarkers for predicting response and resistance to immune checkpoint inhibitors. Given the important impact of these cells on tumor growth and immune cell infiltration, further study of these stromal cells may lead to the identification of novel targets for I-O therapies.

Methods

Analysis of gene expression data in TCGA. All tumor and normal tissue samples that were available from TCGA were used. To avoid bias no samples were excluded from the analysis, therefore the number of data points was not consistent between tumor types. RNA-seq data were analyzed using several previously published EMT- and TGF β -related gene sets^{6–11,35–38,55}. Additional TGF β gene sets were retrieved from the KEGG PATHWAY (<https://www.genome.jp/kegg/pathway/hsa/hsa04350.html>), Mouse Genome Informatics (<http://www.informatics.jax.org/go/term/GO:0005160>), and Gene Set Enrichment Analysis (https://www.gsea-msigdb.org/gsea/msigdb/cards/BIOCARTA_TGFB_PATHWAY) databases. Signature scores were derived by z-scoring (within each tumor type) the expression values for each gene separately, and then for each sample taking the median of these z-scored expression values over the genes in a set.

The presence of stroma in tumor tissue and infiltration of T cells into the parenchyma were assessed using the ESTIMATE gene signatures. Consensus tumor purity estimates were used to assess tumor cell content in the TME. Details of both methods have been published previously⁵⁵.

A 3-gene “CAF signature” was created as a transcriptional marker for fibroblasts. This set was derived by first evaluating co-expression of candidate sentinel markers that displayed selectivity of RNA expression for the target cell type with all transcripts across TCGA solid-tumor cohorts, details of which have been published previously⁵⁶. Fibroblast activation protein is selectively expressed by CAFs in the TME and was therefore chosen as the sentinel marker^{21,43,57}. We then used a stringent method of mutual rank distance to identify gene neighbors for the sentinel markers⁵⁸. Based on these methods, *COL1A1*, *COL1A2*, and *COL3A1* were chosen as a gene set for the CAF signature.

Gene expression profiling of PDX samples. Our study utilized previously published RNA-seq data on 333 PDX samples from 11 tumor types (colon, $n = 70$; pancreatic, $n = 57$; breast, $n = 54$; melanoma, $n = 42$; lung, $n = 40$; ovarian, $n = 32$; renal, $n = 19$; endometrial, $n = 9$; uterine, $n = 4$; sarcoma, $n = 4$)⁵⁹. Bioinformatic analyses were then performed to distinguish transcripts from patient donor (tumor parenchyma) and mouse recipient (tumor stroma) as previously described⁶⁰. For each tumor type, to assess whether expression of genes in a given EMT gene set^{6–11} was enriched in stroma-derived mouse transcripts, we first calculated the fraction of transcripts originating from mouse (the “mouse fraction”) for each gene in each sample. To account for potential differences in mouse content among individual samples, the mouse fraction for each gene was adjusted by subtracting the mean (over the genes) within each sample. Then, for each gene, we calculated the median (over the samples) of these adjusted mouse fractions. The genes were ranked by the medians, and we used these rankings as the input for gene set enrichment analysis⁶¹. Benjamini–Hochberg false discovery rates (FDRs) were used to adjust for multiple testing. The FDR adjustment was performed for each tumor type separately.

scRNA-seq. Tissue procurement. Resected tumors and adjacent normal tissues were procured from BioIVT (BioIVT, Westbury, NY) from 14 treatment-naïve patients with colorectal or renal cell carcinoma. After pathological analysis, all remaining tissue resections (>0.6 g) were used to ensure tissue type representation.

Sorting. Tissue samples were enzymatically dissociated using a human Tumor Dissociation Kit (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer’s recommendations, and red blood cells were lysed using 1X RBC solution (eBioscience, Thermo Fisher Scientific, Waltham, MA, USA) for 5 min at room temperature before quenching with cold phosphate-buffered saline. Subsequently, single cell suspensions were Fc-blocked using TruStain FcX (1:125 dilution; BioLegend, San Diego, CA, USA) for 5 min on ice. Surface CD45 was stained with anti-CD45-FITC (1:50; #REA747, Miltenyi Biotec) for 20 min at 4 °C. Washed cells were stained with propidium iodide (PI; BD Biosciences, Franklin Lakes, NJ, cat: 556463) and with DRAQ5 (#62251, Thermo Fisher Scientific, Waltham, MA, USA) diluted 1:5000 in cold 1% bovine serum albumin in phosphate-buffered saline and stained for 10 min at room temperature. Single cells were sorted for live, nucleated CD45+ and CD45– cells to enrich for immune and non-immune cells.

Sequencing. Consecutive single cell libraries were prepared using the Chromium Single Cell Controller, version 3.16 (10 \times Genomics, Pleasanton, CA) according to the respective protocol. Target cell number was 5000 cells per sample, and sequencing depth was 80,000 reads per cell. All data was merged and filtered for high-quality sequencing data using thresholds ($nFeature_RNA > 200$ & $nFeature_RNA < 6000$ & $mito.read.fraction < 0.15$ & $nCount_RNA < 15,000$). Expression counts were normalized using SCTransform⁶² and aligned using canonical correlation analysis, with uniform manifold approximation and projection (UMAP)⁶³ used for visualization. After clustering, cell types were identified (SingleR)⁶⁴, combined, batch-corrected, and aligned using canonical

correlation analysis (Seurat v3, custom R script)⁶⁵. UMAP was then used to project individual cells across a two-dimensional plane based on similarity of gene expression characteristics⁶³.

CD8 IHC and trichrome staining. CD8 IHC was performed on commercially procured SCCHN samples ($n = 50$) by Mosaic Laboratories (Lake Forest, CA) using a monoclonal CD8 (clone C8/144B) antibody (Dako, an Agilent Technologies Co, Santa Clara, CA). Sample details can be found in Supplementary Table 5. A PathAI-developed artificial intelligence (AI)-powered image analysis algorithm (PathAI research platform, Boston, MA) was used to classify the tumor parenchymal and stromal regions and identify and quantify the density (cells/mm²) of CD8+ T cells in each compartment²⁷.

Collagen was visualized on adjacent sections using a trichrome staining method consisting of Weigert's hematoxylin, Biebrich scarlet-acid fuchsin, phosphomolybdic-phosphotungstic acid, and aniline blue, as described previously⁶⁶. Collagen staining was assessed by image analysis (Halo software, Indica Labs, Albuquerque, NM). Tumor parenchymal and stromal compartments were identified, and the percent area of collagen (blue) staining relative to the area of the total stromal compartment was calculated.

Statistical analyses. Associations between EMT signature scores and other TME signature scores were summarized by Pearson's correlation coefficient (r) and linear regression models for each TCGA tumor type. For the scRNA-seq data, the strength of evidence for differences in median EMT signature scores between stroma fibroblasts and tumor parenchymal cells was assessed by Wilcoxon signed-rank tests. Cox proportional hazards regression models were used to explore the association between EMT and CAF signature scores and OS. Estimated hazard ratios (HRs; high signature score/low signature score) were scaled to compare hazards for scores differing by the interquartile range for each tumor type.

Data availability

The single cell RNA-seq data generated in this study have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE213912. The mouse PDX model RNA-seq count tables and metadata generated in this study have been deposited in GEO and are accessible through GEO Series accession number GSE215307. The raw FASTQ files may be obtained from a repository at the Wasabi cloud platform upon approval by the Dana Farber Cancer Institute Center for Derived Patient Models. More information on Bristol Myers Squibb's data sharing policy can be found here: <https://www.bms.com/researchers-and-partners/clinical-trials-and-research/disclosure-commitment.html>. Any additional data not included in the manuscript or supplementary files that support the findings of this study are available from the corresponding author N.K.

Received: 5 April 2022; Accepted: 19 January 2023

Published online: 21 February 2023

References

- Chen, T., You, Y., Jiang, H. & Wang, Z. Z. Epithelial-mesenchymal transition (EMT): A biological process in the development, stem cell differentiation and tumorigenesis. *J. Cell Physiol.* **232**, 3261–3272 (2017).
- Jolly, M. K. & Celia-Terrassa, T. Dynamics of phenotypic heterogeneity associated with EMT and stemness during cancer progression. *J. Clin. Med.* **8**, 1542 (2019).
- Lu, W. & Kang, Y. Epithelial-mesenchymal plasticity in cancer progression and metastasis. *Dev. Cell* **49**, 361–374 (2019).
- Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).
- Ribatti, D., Tamma, R. & Annesse, T. Epithelial-mesenchymal transition in cancer: A historical overview. *Transl. Oncol.* **13**, 100773 (2020).
- Mak, M. P. *et al.* A patient-derived, pan-cancer EMT signature identifies global molecular alterations and immune target enrichment following epithelial-to-mesenchymal transition. *Clin. Cancer Res.* **22**, 609–620 (2016).
- Tan, T. Z. *et al.* Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* **6**, 1279–1293 (2014).
- Carretero, J. *et al.* Integrative genomic and proteomic analyses identify targets for Lkb1-deficient metastatic lung tumors. *Cancer Cell* **17**, 547–559 (2010).
- Hugo, W. *et al.* Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44 (2016).
- Wang, L. *et al.* EMT- and stroma-related gene expression and resistance to PD-1 blockade in urothelial cancer. *Nat. Commun.* **9**, 3503 (2018).
- Liberzon, A. *et al.* The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
- Ahmadzadeh, M. & Rosenberg, S. A. TGF- β 1 attenuates the acquisition and expression of effector function by tumor antigen-specific human memory CD8 T cells. *J. Immunol.* **174**, 5215–5223 (2005).
- Wrzesinski, S. H., Wan, Y. Y. & Flavell, R. Transforming growth factor-beta and the immune response: Implications for anticancer therapy. *Clin. Cancer Res.* **13**, 5262–5270 (2007).
- Liu, T. *et al.* Cancer-associated fibroblasts: An emerging target of anti-cancer immunotherapy. *J. Hematol. Oncol.* **12**, 86 (2019).
- Chandra Jena, B. *et al.* Paracrine TGF- β 1 from breast cancer contributes to chemoresistance in cancer associated fibroblasts via upregulation of the p44/42 MAPK signaling pathway. *Biochem. Pharmacol.* **186**, 114474 (2021).
- Borriello, L. *et al.* Cancer-associated fibroblasts share characteristics and protumorigenic activity with mesenchymal stromal cells. *Cancer Res.* **77**, 5142–5157 (2017).
- Straussman, R. *et al.* Tumor microenvironment induces innate RAF-inhibitor resistance through HGF secretion. *Nature* **487**, 500–504 (2012).
- Feig, C. *et al.* Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc. Natl. Acad. Sci. U S A* **110**, 20212–20217 (2013).
- Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580–584 (2013).

20. Ruhland, M. K. *et al.* Stromal senescence establishes an immunosuppressive microenvironment that drives tumorigenesis. *Nat. Commun.* **7**, 11762 (2016).
21. Yang, X. *et al.* FAP promotes immunosuppression by cancer-associated fibroblasts in the tumor microenvironment via STAT3–CCL2 signaling. *Cancer Res.* **76**, 4124–4135 (2016).
22. Hanley, C. J. *et al.* A subset of myofibroblastic cancer-associated fibroblasts regulate collagen fiber elongation, which is prognostic in multiple cancers. *Oncotarget* **7**, 6159–6174 (2016).
23. Mariathasan, S. *et al.* TGF- β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
24. Hamid, O. *et al.* A prospective phase II trial exploring the association between tumor microenvironment biomarkers and clinical activity of ipilimumab in advanced melanoma. *J. Transl. Med.* **9**, 204 (2011).
25. Ji, R.-R. *et al.* An immune-active tumor microenvironment favors clinical response to ipilimumab. *Cancer Immunol. Immunother.* **61**, 1019–1031 (2012).
26. Ayers, M. *et al.* IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
27. Szabo, P. M. *et al.* P110 Tumor CD8+ T-cell infiltration assessed by gene expression profiling alone or by immunohistochemistry plus epithelial-mesenchymal transition gene expression in urothelial carcinoma in CheckMate 275. *J. Immunother. Cancer* **7**, 282 (2019).
28. Bertotti, A. *et al.* A molecularly annotated platform of patient-derived xenografts (“xenopatiens”) identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov.* **1**, 508–523 (2011).
29. Julien, S. *et al.* Characterization of a large panel of patient-derived tumor xenografts representing the clinical heterogeneity of human colorectal cancer. *Clin. Cancer Res.* **18**, 5314–5328 (2012).
30. DeRose, Y. S. *et al.* Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat. Med.* **17**, 1514–1520 (2011).
31. Zhang, X. *et al.* A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. *Cancer Res.* **73**, 4885–4897 (2013).
32. Keysar, S. B. *et al.* A patient tumor transplant model of squamous cell cancer identifies PI3K inhibitors as candidate therapeutics in defined molecular bins. *Mol. Oncol.* **7**, 776–790 (2013).
33. Sivanand, S. *et al.* A validated tumorigraft model reveals activity of dovitinib against renal cell carcinoma. *Sci. Transl. Med.* **4**, 137ra175 (2012).
34. Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
35. Verrecchia, F., Chu, M.-L. & Mauviel, A. Identification of novel TGF-beta /Smad gene targets in dermal fibroblasts using a combined cDNA microarray/promoter transactivation approach. *J. Biol. Chem.* **276**, 17058–17062 (2001).
36. Karakas, B. *et al.* Interleukin-1 alpha mediates the growth proliferative effects of transforming growth factor-beta in p21 null MCF-10A human mammary epithelial cells. *Oncogene* **25**, 5561–5569 (2006).
37. Coulouarn, C., Factor, V. M. & Thorgeirsson, S. S. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. *Hepatology* **47**, 2059–2067 (2008).
38. Plasari, G. *et al.* Nuclear factor I-C links platelet-derived growth factor and transforming growth factor beta1 signaling to skin wound healing progression. *Mol. Cell Biol.* **29**, 6006–6017 (2009).
39. Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nature* **18**, 128–134 (2018).
40. Hollis, R. L. *et al.* Ovarian carcinosarcoma is a distinct form of ovarian cancer with poorer survival compared to tubo-ovarian high-grade serous carcinoma. *Br. J. Cancer* **134**, 46 (2022).
41. Puram, S. V. *et al.* Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624 (2017).
42. Zhang, L. *et al.* Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* **181**, 442–459 (2020).
43. Gunderson, A. J. *et al.* Blockade of fibroblast activation protein in combination with radiation treatment in murine models of pancreatic adenocarcinoma. *PLoS ONE* **14**, e0211117 (2019).
44. Lo, A. *et al.* Tumor-promoting desmoplasia is disrupted by depleting FAP-expressing stromal cells. *Cancer Res.* **75**, 2800–2810 (2015).
45. Zhang, Y. & Ertl, H. C. J. Depletion of FAP+ cells reduces immunosuppressive cells and improves metabolism and functions CD8+T cells within tumors. *Oncotarget* **7**, 23282–23299 (2016).
46. Desbois, M. *et al.* Integrated digital pathology and transcriptome analysis identifies molecular mediators of T-cell exclusion in ovarian cancer. *Nat. Commun.* **11**, 5583 (2020).
47. Okrah, K. *et al.* Transcriptomic analysis of hepatocellular carcinoma reveals molecular features of disease progression and tumor immune biology. *NPJ Precis. Oncol.* **2**, 25 (2018).
48. Hua, W., Dijke, P. T., Kostidis, S., Giera, M. & Hornsveld, M. TGF β -induced metabolic reprogramming during epithelial-to-mesenchymal transition in cancer. *Cell Mol. Life Sci.* **77**, 2103–2123 (2020).
49. Gonzalez, D. M. & Medici, D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci. Signal* **7**, re8 (2015).
50. Wu, F. *et al.* Signaling pathways in cancer-associated fibroblasts and targeted therapy for cancer. *Signal Transduct. Target. Ther.* **6**, 218 (2021).
51. Bonde, A.-K., Tischler, V., Kumar, S., Soltermann, A. & Schwendener, R. A. Intratumoral macrophages contribute to epithelial-mesenchymal transition in solid tumors. *BMC Cancer* **12**, 35 (2012).
52. Baulida, J. Epithelial-to-mesenchymal transition transcription factors in cancer-associated fibroblasts. *Mol. Oncol.* **11**, 847–859 (2017).
53. Sasaki, K. *et al.* Analysis of cancer-associated fibroblasts and the epithelial-mesenchymal transition in cutaneous basal cell carcinoma, squamous cell carcinoma, and malignant melanoma. *Hum. Pathol.* **79**, 1–8 (2018).
54. You, J. *et al.* Snail1-dependent cancer-associated fibroblasts induce epithelial-mesenchymal transition in lung cancer cells via exosomes. *QJM* **112**, 581–590 (2019).
55. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
56. Siemers, N. O. *et al.* Genome-wide association analysis identifies genetic correlates of immune infiltrates in solid tumors. *PLoS ONE* **12**, e0179726 (2017).
57. Denton, A. E., Roberts, E. W. & Linterman, M. A. Fibroblastic reticular cells of the lymph node are required for retention of resting but not activated CD8+ T cells. *Proc. Natl. Acad. Sci. U S A* **111**, 12139–12144 (2014).
58. Huttenhower, C. *et al.* Nearest Neighbor Networks: Clustering expression data based on gene neighborhoods. *BMC Bioinformatics* **8**, 250 (2007).
59. Gao, H. *et al.* High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21**, 1318–1325 (2015).
60. Liu, Y. *et al.* Gene expression differences between matched pairs of ovarian cancer patient tumors and patient derived xenografts. *Sci. Rep.* **9**, 6314 (2018).
61. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* **102**, 15545–15550 (2005).

62. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
63. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv* <https://arxiv.org/abs/1802.03426> (2020).
64. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
65. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
66. Prophet, E., Mills, B., Arrington, J. & Sobin, L. *Laboratory Methods in Histotechnology*. American Registry of Pathology (1992).

Acknowledgments

This study was supported by Bristol Myers Squibb (Princeton, NJ, USA). The results shown here are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. The authors would like to thank Ruth Lan for her contributions to this study. The authors would also like to thank Dako, an Agilent Technologies, Inc. company, for collaborative development of the PD-L1 IHC 28-8 pharmDx assay and PathAI for development and use of the image analysis algorithm used to classify CD8 T cells. Writing and editorial assistance were provided by Emily Motola, PharmD, and Jay Rathi, MA, of Spark Medica Inc., funded by Bristol Myers Squibb, according to Good Publication Practice guidelines.

Author contributions

K.L.L. contributed to the conceptualization, data curation, supervision, investigation, methodology, and project administration. S.D.C. contributed to the software generation, visualization, and methodology. B.J.C. was involved in the investigation and methodology. J.L.H. aided in the data curation and investigation. A.V., M.Y.T., N.K., and K.-H.C. contributed to data curation, formal analysis, visualization, methodology, and investigation. M.Y.T. also provided supervision, N.K. assisted in validation, and K.-H.C. aided in project administration. J.H. contributed to the methodology. S.S. assisted in conceptualization, resourcing, funding acquisition, project administration, and supervision. N.O.S. contributed to the methodology. A.S. assisted in resourcing, data curation, and methodology. R.E. aided in conceptualization and provided consultation. P.M.S. contributed to the conceptualization, resourcing, data curation, formal analysis, supervision, funding acquisition, investigation, visualization, and methodology. All authors contributed to the writing, review, and editing of this manuscript.

Competing interests

P.M.S. is an employee of Fate Therapeutics. A.V., M.Y.T., K.-H.C., and A.S. have nothing to disclose. N.K., B.J.C., S.D.C., M.B., J.L.H., R.G., B.A.K., P.A.H., and J.H. are employees of and own stock in Bristol Myers Squibb. R.E. is a previous employee of Bristol Myers Squibb and owns stock in the company; and is an employee of Daiichi Sankyo, Inc. and owns stock in the company. K.L.L. has received financial support payable to Dana Farber Cancer Institute related to the submitted work. Outside the submitted work, K.L.L. has received personal fees from IntegraGen, RareCyte, and Bristol Myers Squibb and grants from Amgen, Lilly, and Bristol Myers Squibb. K.L.L. is also a cofounder of and owns equity in Travera LLC. L.V. is an employee of Vanderbilt University. N.O.S. is an employee of and owns stock in Abiosciences, Inc. S.S. is the CEO and Board Director of Centessa Pharmaceuticals.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-28480-9>.

Correspondence and requests for materials should be addressed to N.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023