

# Single-Cell Atlas of Lineage States, Tumor Microenvironment, and Subtype-Specific Expression Programs in Gastric Cancer



Vikrant Kumar<sup>1</sup>, Kalpana Ramnarayanan<sup>1</sup>, Raghav Sundar<sup>1,2,3,4,5</sup>, Nisha Padmanabhan<sup>1</sup>, Supriya Srivastava<sup>6</sup>, Mayu Koiwa<sup>7</sup>, Tadahito Yasuda<sup>7</sup>, Vivien Koh<sup>2,8</sup>, Kie Kyon Huang<sup>1</sup>, Su Ting Tay<sup>1</sup>, Shamaine Wei Ting Ho<sup>1,8</sup>, Angie Lay Keng Tan<sup>1</sup>, Takatsugu Ishimoto<sup>7</sup>, Guowei Kim<sup>3,9</sup>, Asim Shabbir<sup>3,9</sup>, Qingfeng Chen<sup>10,11</sup>, Biyan Zhang<sup>12</sup>, Shengli Xu<sup>12,13</sup>, Kong-Peng Lam<sup>11,12,14</sup>, Huey Yew Jeffrey Lum<sup>15</sup>, Ming Teh<sup>15</sup>, Wei Peng Yong<sup>2,5,8</sup>, Jimmy Bok Yan So<sup>3,5,9,16</sup>, and Patrick Tan<sup>1,5,8,13,17,18</sup>

**ABSTRACT**

Gastric cancer heterogeneity represents a barrier to disease management. We generated a comprehensive single-cell atlas of gastric cancer (>200,000 cells) comprising 48 samples from 31 patients across clinical stages and histologic subtypes. We identified 34 distinct cell-lineage states including novel rare cell populations. Many lineage states exhibited distinct cancer-associated expression profiles, individually contributing to a combined tumor-wide molecular collage. We observed increased plasma cell proportions in diffuse-type tumors associated with epithelial-resident *KLF2* and stage-wise accrual of cancer-associated fibroblast subpopulations marked by high *INHBA* and *FAP* coexpression. Single-cell comparisons between patient-derived organoids (PDO) and primary tumors highlighted inter- and intralinear similarities and differences, demarcating molecular boundaries of PDOs as experimental models. We complemented these findings by spatial transcriptomics, orthogonal validation in independent bulk RNA-sequencing cohorts, and functional demonstration using *in vitro* and *in vivo* models. Our results provide a high-resolution molecular resource of intra- and interpatient lineage states across distinct gastric cancer subtypes.

**SIGNIFICANCE:** We profiled gastric malignancies at single-cell resolution and identified increased plasma cell proportions as a novel feature of diffuse-type tumors. We also uncovered distinct cancer-associated fibroblast subtypes with *INHBA*-*FAP*-high cell populations as predictors of poor clinical prognosis. Our findings highlight potential origins of deregulated cell states in the gastric tumor ecosystem.

**INTRODUCTION**

Gastric cancer is a leading cause of global cancer morbidity and mortality (1) with particularly high incidence in Asia, Eastern Europe, and Central America (2). Between individual patients, gastric tumors frequently exhibit high levels of histologic, transcriptomic, and (epi)genomic variation, with distinct clinical behaviors and treatment response (“interpatient heterogeneity”). Factoring this heterogeneity into gastric cancer clinical management, and identifying molecular pathways driving hallmarks of gastric cancer variation, represent important goals for improving patient outcomes. Although progress has been made in defining specific molecular subtypes of gastric cancer through consortia such as The Cancer Genome Atlas (TCGA) and Asian Cancer Research Group (ACRG; refs. 3, 4), tangible

improvements in patient outcomes based on these findings have been modest, compounded by the growing recognition that gastric cancers also exhibit high levels of within-patient heterogeneity [“intrapatient heterogeneity” (ITH)]. High-resolution studies probing the molecular extent of gastric cancer ITH across a wide range of patients with gastric cancer are thus required (5) to understand key principles governing gastric cancer evolution, selection, and adaptation and how clinical care pathways should be adapted to manage gastric cancer ITH.

Using “bulk-transcriptome” experimental methods, we and others have previously established that each gastric tumor possesses a personalized expression profile comprising distinct transcriptional programs, contributed by both cancer epithelial cells and other cell types in the tumor microenvironment (TME; refs. 6, 7). However, our understanding

<sup>1</sup>Cancer and Stem Cell Biology Program, Duke-NUS Medical School, Singapore. <sup>2</sup>Department of Haematology-Oncology, National University Cancer Institute, National University Health System, Singapore. <sup>3</sup>Yong Loo Lin School of Medicine, National University of Singapore, Singapore. <sup>4</sup>The N.I Institute for Health, National University of Singapore, Singapore. <sup>5</sup>Singapore Gastric Cancer Consortium, Singapore. <sup>6</sup>Department of Medicine, National University of Singapore, Singapore. <sup>7</sup>Gastrointestinal Cancer Biology, International Research Center for Medical Sciences (IRCMS), Kumamoto University, Kumamoto, Japan. <sup>8</sup>Cancer Science Institute of Singapore, National University of Singapore, Singapore. <sup>9</sup>Department of Surgery, University Surgical Cluster, National University Health System, Singapore. <sup>10</sup>Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Singapore. <sup>11</sup>Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, Singapore. <sup>12</sup>Singapore Immunology Network (SigN), A\*STAR, Singapore. <sup>13</sup>Department of Physiology, National University of Singapore, Singapore. <sup>14</sup>School of Biological Sciences, Nanyang Technological University, Singapore. <sup>15</sup>Department of Pathology, National University Health System, Singapore. <sup>16</sup>Division of Surgical Oncology, National University Cancer Institute, Singapore.

<sup>17</sup>Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore. <sup>18</sup>SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore.

**Note:** Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

V. Kumar, K. Ramnarayanan, R. Sundar, and N. Padmanabhan contributed equally to this article.

**Corresponding Author:** Patrick Tan, Cancer and Stem Cell Biology Program, Duke-NUS Medical School, 8 College Road, Singapore 169857, Singapore. Phone: 65-6516-1783; Fax: 65-6221-2402; E-mail: gmstanp@duke-nus.edu.sg

Cancer Discov 2022;12:670-91

doi: 10.1158/2159-8290.CD-21-0683

This open access article is distributed under the Creative Commons Attribution License 4.0 International (CC BY).

©2021 The Authors; Published by the American Association for Cancer Research

of mechanisms by which TME-resident cell types such as immune cells, fibroblasts, and blood vessels drive gastric cancer phenotypes and clinical trajectories remains nascent (8). Although advanced bioinformatic programs have been designed to decompose bulk sequencing data into lineage-specific constituent programs, these deconvolution algorithms are often not able to discern fine-scale tissue lineages, relationships between lineages, rare cell populations, and cell-cell interactions (9). To tackle these challenges, single-cell RNA sequencing (scRNA-seq) is proving to be a powerful tool for characterizing gene expression across thousands of cells simultaneously, enabling comprehensive profiling of different cell types in tumors in distinct biological states and conditions (10). In gastric cancer, recent scRNA-seq studies have provided unique insights on different aspects of gastric tumor biology. For example, Wang and colleagues performed scRNA-seq on malignant cells from ascitic fluids of patients with gastric cancer to develop a prognostic signature based on features of malignant peritoneal cells (11). Other scRNA-seq studies of primary gastric cancer tumors have been performed, but on limited numbers of samples and cells (8–13 patients, 27,000 to 55,000 cells per study; refs. 12–14). Another limitation of current single-cell sequencing platforms is the requirement for tissue dissociation, which leads to loss of spatial information. To address this, newer “spatial transcriptomic” platforms such as digital spatial profiling (DSP), *in situ* sequencing, and MERFISH have been developed that retain spatial architecture, thereby allowing analysis of tumor–TME interactions at unprecedented depth (15–17).

Here, we performed scRNA-seq on an expanded cohort of more than 30 patients with gastric cancer with different subtypes and stages, across a large number of cells (>200,000 cells), to generate a comprehensive single-cell landscape of gastric cancer encompassing both inter- and intratumoral heterogeneity. We discovered unique and novel features of the gastric TME, including an increased proportion of plasma cells in diffuse-type tumors, and a role for *INHBA*, a subunit of activin–inhibin complexes, in specific subtypes of cancer-associated fibroblasts (CAF). We also performed spatial transcriptomic analysis on these samples to geographically validate predicted intercell and intracellular relationships *in situ*, and further reinforced our findings using functional assays and conventional IHC.

## RESULTS

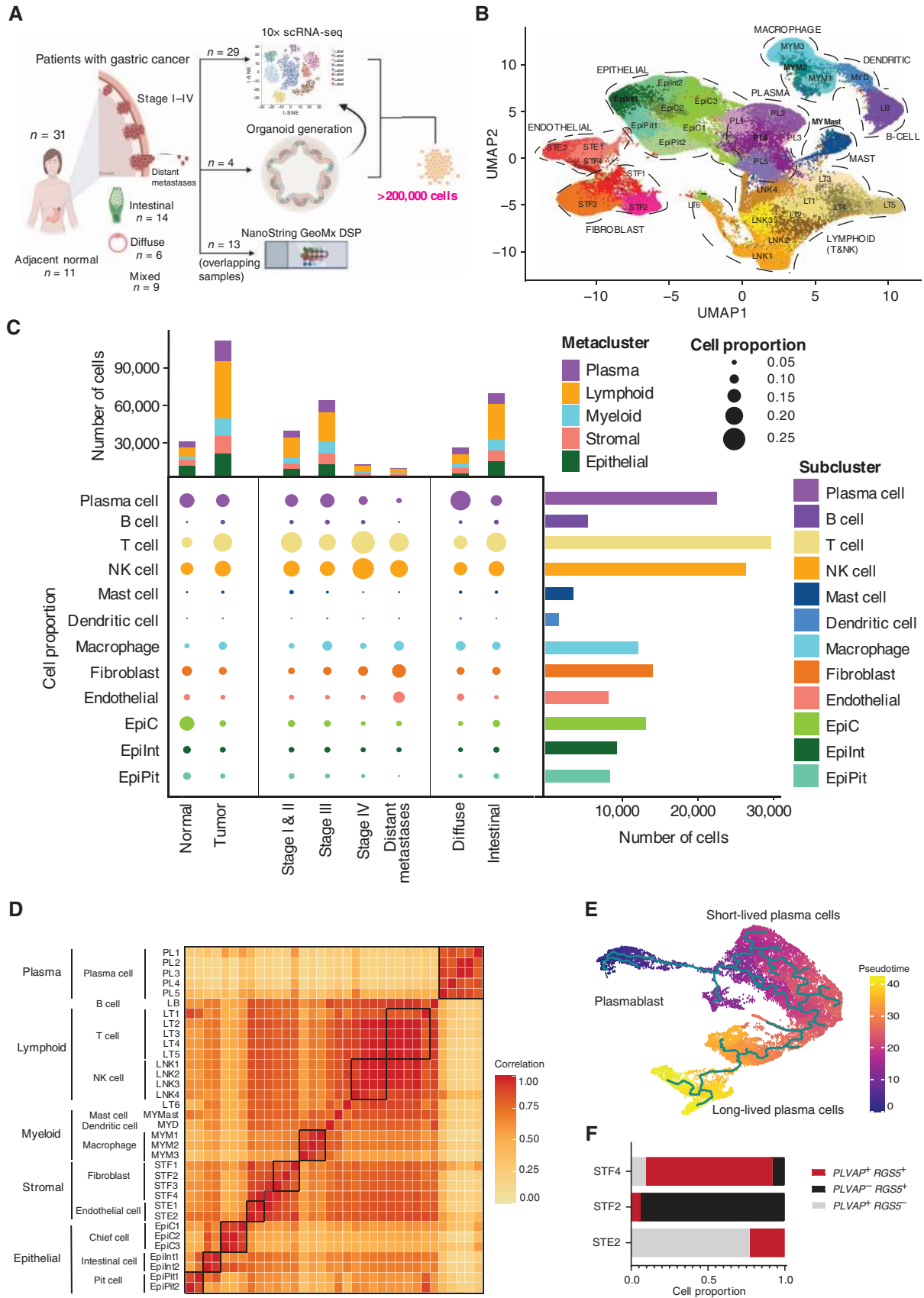
### scRNA-seq of ~200,000 Gastric Cancer Cells Identifies Diverse Tissue-Lineage States, Cell Fate Trajectories, and Rare Cell Populations

Droplet-based scRNA-seq (10× Genomics) was performed on 48 surgical resection and biopsy samples across 31 unique individuals with gastric cancer, ranging from stages I to IV, distinct histologic subtypes (diffuse and intestinal), molecular subtypes (TCGA), primary tumors to distant (peritoneal) metastases, and matched normal gastric tissues (Fig. 1A; Supplementary Table S1). We also performed scRNA-seq on tumor and matched normal gastric cancer patient-derived organoids (PDO) to investigate the effects of *in vitro* culture on cancer-intrinsic and TME signatures (Fig. 1A; Supplementary Table S1). After quality control (QC) and removal of batch effects (18), 200,954 single cells were included in the final data set. For 13 samples [10 tumor and 3 normal; 156 regions of interest (ROI)], we also performed spatial transcriptomics using the DSP platform (NanoString GeoMx) to gain insights into *in situ* geographic and spatial relationships linking discrete cell states and molecular interactions.

We first performed dimensionality reduction on 152,423 cells derived from 40 primary samples (29 tumor and 11 normal, of which 10 were matched). This analysis revealed 34 unique tissue states (Fig. 1B). Using tissue type-specific canonical markers defined in the literature (ref. 14; Supplementary Fig. S1), the cell states were broadly categorized into five major cell types, referred to as “metaclusters” (myeloid, lymphoid, plasma, epithelial, and stromal; Fig. 1C; Supplementary Fig. S2A). Supporting the supervised cell type-specific marker analysis, an unsupervised global clustering similarity matrix also classified the cell states into five metaclusters (Fig. 1D). Alternative clustering algorithms (e.g., integrative nonnegative matrix factorization) confirmed the molecular distinctiveness of each cluster, and comparison of global versus metacluster-specific clustering approaches revealed similar results (Supplementary Fig. S2B–S2E).

Overall, immune-cell populations dominated the cell states (21 of 34 states). These included: (i) a “myeloid metacluster” (five cell states), consisting of a mast cell cluster (*KIT* positive: “MAST”), a dendritic cell cluster (*PLD4* positive: “MYD”), and three macrophage clusters (*CD163* positive: “MYM1–3”); (ii) a “lymphoid metacluster” (11 cell states), consisting of

**Figure 1.** scRNA-seq of gastric tumor and normal samples defines 34 cell states including rare cell populations. **A**, Schematic representation of experimental design and techniques used in this study. Thirty-one unique patients with gastric cancer undergoing surgical resection or endoscopy had tumor samples ( $n = 31$ ) and adjacent normal samples ( $n = 11$ ) harvested for analysis. Tumors ranged from stage I to IV and included samples of both primary tumors, distant (peritoneal) metastases, and matched normal gastric tissues. Twenty-nine tumors had scRNA-seq performed using the 10× platform (along with 11 adjacent normal tissues). Four patients had PDOs generated from their tumors (4 tumors + 4 adjacent normal), which were also sequenced by 10× scRNA-seq. A subset of 13 samples also had DSP performed using the NanoString GeoMx platform (10 tumor + 3 normal). In total, more than 200,000 cells were sequenced in this study. **B**, Uniform Manifold Approximation and Projection (UMAP) of 152,423 cells representing 34 unique cell states color-coded by their corresponding cell lineage or subtype. Each dot in the UMAP represents a single cell. **C**, Cell-lineage compositions of gastric cancer and normal samples inferred by scRNA-seq data. Middle (bubble plot), cell subclusters (rows) by tumor versus normal, stage, and gastric cancer histologic subtype (diffuse vs. intestinal). The size of the circle represents the cell proportion of each specific cell lineage/type. The circles are color-coded by defined cell lineages/types as shown in **B**. The stacked bar graph on the top shows the number of cells in each meta-cluster for each category. The histogram on the right shows the absolute cell numbers in each subcluster. **D**, Cluster-cluster heat map of gene-expression data of all 34 cell states across all samples using Pearson correlation matrix. Darker colors correspond to higher correlation. **E**, Pseudotime analysis of plasma metacluster generated using Monocle. The trajectory was rooted against the plasmablasts. Pseudotime analysis demonstrates different stages of plasma cell differentiation and maturation including plasmablasts, short-lived plasma cells, and long-lived plasma cells. **F**, Expression of *PLVAP* and *RGS5* in endothelial (STE2) and fibroblast (STF2 and STF4) clusters. Doublets were identified and filtered out using DoubletFinder. *PLVAP*<sup>+</sup> *RGS5*<sup>+</sup> cells are predominantly present in the endothelial cluster (STE2). *PLVAP*<sup>+</sup> *RGS5*<sup>+</sup> cells are predominant in the fibroblast cluster (STF2). The STF4 cluster shows cells expressing *PLVAP*<sup>+</sup> *RGS5*<sup>+</sup>, suggestive of a rare mixed-lineage population.



Downloaded from <http://aacrjournals.org/cancerdiscovery/article-pdf/12/3/670/3196075/670.pdf> by guest on 22 August 2022

six T-cell clusters, mapping to effector and naïve CD8 T cells (*CD8A*, *GZMH*, *GZMM*, and *NKG7*; LT1–2), naïve and helper CD4 T cells (*CCR7* and *STAT4*; LT3–4), regulatory T cells (Treg; *IL2RA* and *STAT3*; LT5), and proliferative T cells (expressing cell-cycle genes; LT6), four natural killer (NK) cell clusters (*KLRD1* positive; LNK1–4), and one B-cell cluster (*MS4A1* positive; LB); and (iii) a “plasma metacluster” (five cell states), consisting of five mature B cells/plasma cell clusters (*TNFRSF17* positive; PL1–5; Supplementary Fig. S1; Supplementary Table S2).

The “epithelial” metacluster (seven cell states; *CDH1* positive) contained three distinct sublineages: “EpiPit,” expressing markers of mucous pit cells (*MUC5AC* and *TFF1*; EpiPit1&2); “EpiC,” expressing markers of chief cells (*LIPF* and *PGA3*; EpiC1–3); and “EpiInt,” expressing intestinal-type markers (*REG4* and *TFF3*; EpiInt1&2; Supplementary Fig. S1). EpiInt cells exhibited features of intestinal metaplasia (IM), a premalignant condition recognized histologically by gastric epithelial cells acquiring intestinal cell-type features (12). This was supported by the expression of *CDX1* and *CDX2* master transcriptional regulators, largely in EpiInt cells, which play a key role in the *trans*-differentiation of gastric epithelial cells to intestinal cell type (Supplementary Fig. S2F; ref. 19). The “stromal metacluster” (six cell states; *FNI* positive) consisted of pericytes (STF2; defined by *RGS5* and *NOTCH3*), fibroblasts (STF1 and STF3; defined by *LUM* and *DCN*), and *PLVAP*-positive endothelial cell subclusters. The latter could be further partitioned into *ACKR1*-positive venular endothelial cells (STE1) or nonvenular endothelial cells (STE2; ref. 20; Supplementary Fig. S1). The remaining cluster (STF4) could not be assigned to a single cell type and is described further in the next paragraph. The cell types varied in their relative proportions across the ~150K cells (Fig. 1C; Supplementary Fig. S2G). Specifically, lymphoid cells exhibited the largest proportion (39.7%; range, – 19.1% for T cells to 3.6% for B cells) followed by epithelial cells (20%; range, 8.5% for EpiC to 5.4% for EpiPit cells). Myeloid, stromal, and plasma cells constituted proportions of 11.4%, 14.4%, and 14.5%, respectively.

The large number of cells profiled in our study enabled us to relate distinct cell states to one another across biological state transitions. For example, single-cell trajectory and pseudotime analysis of the plasma cell metacluster demonstrated different stages of plasma cell differentiation and maturation, consistent with the known literature (ref. 21; bootstrapping  $P < 0.05$  on major branching nodes; Fig. 1E). These included plasmablasts (high *XBPI* and low *SDCI*),

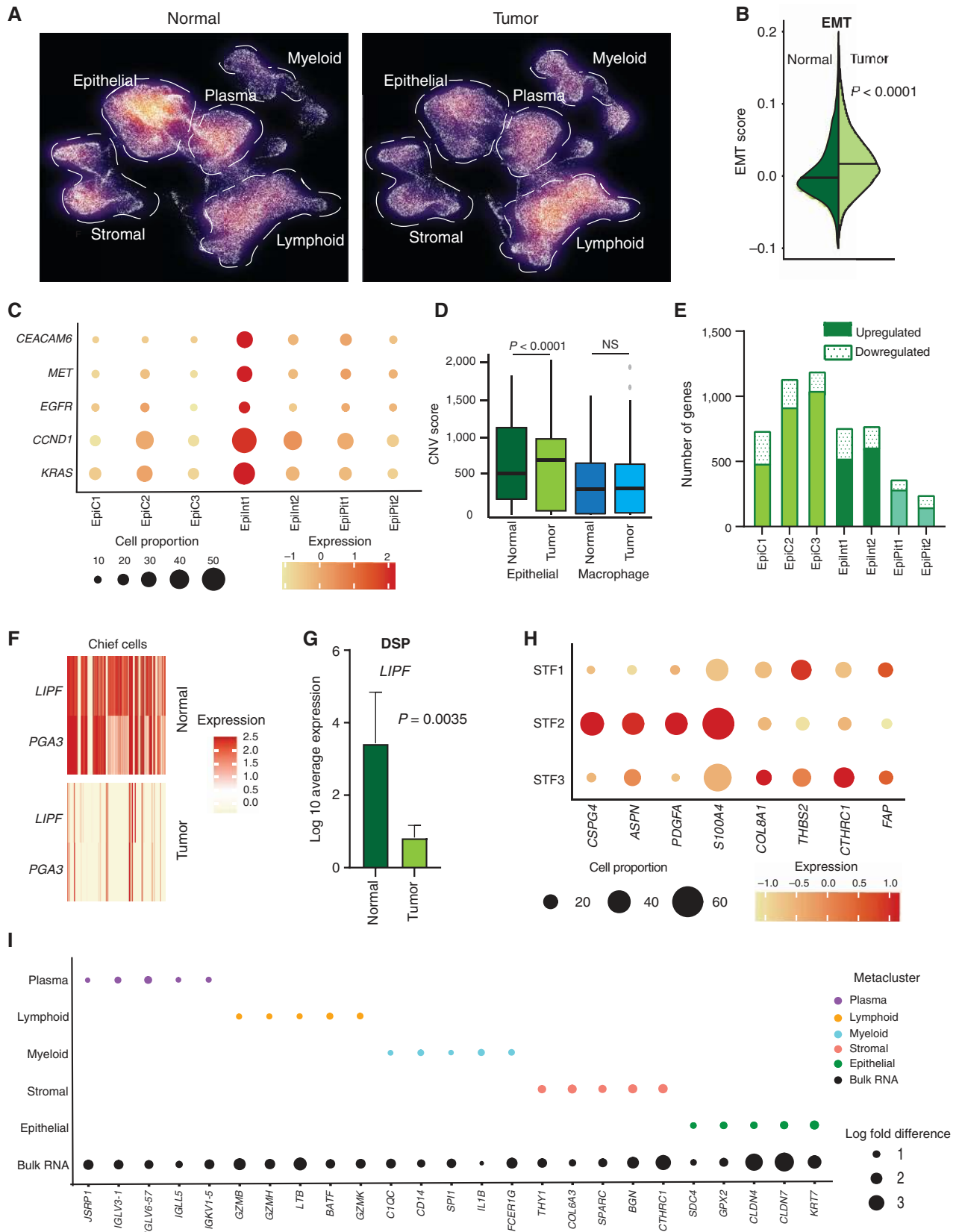
short-lived plasma cells (high *SDCI*), and long-lived plasma cells (high *SDCI*, *STAT3*, and *IKZF3*; refs. 22–25). Short-lived plasma cells predominated the trajectory plot, whereas long-lived plasma cells exhibited relatively higher expression of *IGHA1* (IgA), consistent with increased levels of IgA detected in gastric lamina propria (ref. 26; Supplementary Fig. S2H). Similarly, single-cell trajectory analysis of macrophages revealed two distinct cell states: proinflammatory “M1-like” (high *CD163* and *S100A12*) and “M2-like” tumor-associated macrophages (high *CD163* and *FOLR2*; refs. 27 and 28; Supplementary Fig. S2I). The large number of analyzed cells also enabled characterization of a novel and rare cell type (STF4) within the stromal metacluster (0.4% of all cells;  $n = 821$ ), expressing markers associated with both endothelial (*PLVAP*) and fibroblast (*RGS5*) lineages (Fig. 1F), possibly highlighting cells undergoing endothelial-mesenchymal transition (EndoMT), a process where endothelial cells acquire a mesenchymal or myofibroblastic phenotype. We ruled out the possibility that the double-lineage nature of STF4 cells is caused by potential technical artifacts such as doublet effects (29) and orthogonally validated the existence of the *PLVAP*/*RGS5* double-positive population on formalin-fixed paraffin-embedded (FFPE) sections using dual-color RNAScope (Supplementary Fig. S2J). Taken together, these results showcase the ability of scRNA-seq data, when appropriately powered, to reveal cells exhibiting diverse tissue-lineage states, different state transitions, and rare populations associated with mixed or dual lineages.

### scRNA-seq Discerns Cell State-Specific Transcriptional Patterns Associated with Gastric Cancer

To examine tumor-associated expression programs at the single-cell level, we compared primary gastric tumor samples to normal tissues ( $n = 26$  tumor;  $n = 10$  normal). After random downsampling to ensure statistical equivalency (Fig. 2A), the proportion of epithelial cells in tumors was lower compared with normal tissues ( $P = 0.002$ ), whereas myeloid cells in tumors were higher ( $P = 0.05$ ; Supplementary Table S3). The proportions of lymphoid, plasma, and stromal cells were not statistically different between tumor and normal samples.

Compared with their cognate cell states in normal tissues, tumor-associated epithelial cells exhibited higher expression of multiple oncogenic gene signatures related to epithelial-mesenchymal transition (EMT), cell motility, and cancer signatures derived from previous gastric cancer scRNA-seq

**Figure 2.** scRNA-seq deconvolutes gastric tumor programs associated with distinct cell states. **A**, Density plot of UMAP representation comparing normal and gastric tumor samples after random downsampling to approximately 30,000 cells each to allow statistical equivalence. Each dot represents a single cell. Dashed lines highlight higher proportions of epithelial cells in normal samples and myeloid cells in tumor samples. **B**, Split violin plot of EMT oncogenic gene signature score in normal and tumor cells, showing a significantly higher score in tumor cells. **C**, Bubble plot depicting the expression of gastric cancer oncogenes in tumor epithelial cell clusters. The size of the circle represents the percentage of cells expressing the gene in that specific epithelial cell cluster, whereas the color represents the average expression of the gene. **D**, Box plot depicting CNV scores for epithelial cells (green) and macrophage cells (blue) in normal and tumor samples. CNV scores were computed using InferCNV.  $P$  values were computed using Wilcoxon rank-sum test. **E**, Bar graph depicting differences in transcriptomic profiles between tumor and normal tissue by number of upregulated and downregulated genes in epithelial cell clusters. **F**, Heat map of *LIPF* and *PGA3* gene expression (classic chief cell marker genes) in tumor versus normal samples. Darker color signifies higher expression. **G**, Gene expression of *LIPF* by DSP analysis, in tumor epithelial cells (Pan-CK<sup>+</sup>) compared with normal samples ( $n = 13$ ). **H**, Bubble plot depicting sublineage-specific expression of CAF marker genes *FAP*, *CSPG4*, *PDGFA*, *ASPN*, *S100A4*, *COL8A1*, *THBS2*, and *CTHRC1* in fibroblast clusters. STF1 and STF3 are *LUM*-associated fibroblasts, whereas STF2 comprises proangiogenic pericytes. The size of the circle represents the percentage of cells expressing the gene in that specific fibroblast cell cluster, whereas the color represents the average expression of the gene. **I**, Bubble plot depicting significant log fold differences in expression of genes between tumor and normal by metacluster mapped against the bulk RNA-seq data (five genes per metacluster are shown). The size of the circle represents the log fold change in the expression of specific genes.



Downloaded from <http://aacrjournals.org/cancerdiscovery/article-pdf/12/3/670/3196075/670.pdf> by guest on 22 August 2022

studies (refs. 12–14; Fig. 2B; Supplementary Fig. S3A and S3B; Supplementary Table S4;  $P < 0.0001$ ). When we assessed the expression status of individual gastric cancer oncogenes (e.g., *CEACAM6*, *EGFR*, *MET*, *CCND1*, and *KRAS*) among the tumor epithelial clusters, the EpiInt1 tumor epithelial cluster exhibited the largest increase (Fig. 2C). We orthogonally validated the oncogenic gene signature associations in multiple ways. First, we confirmed a significant correlation between the proportions of inferred scRNA-seq tumor cells and EpiInt1 tumor cellularity proportions measured by hematoxylin and eosin (H&E) staining ( $R = 0.64$ ,  $P < 0.0074$ ). Second, inferring copy-number variations (CNV) from the scRNA-seq data using approaches employed in other cancer single-cell studies (30–32), we confirmed that tumor-associated epithelial cells exhibited significantly higher numbers of inferred aberrant CNVs compared with normal epithelial cells ( $P < 0.0001$ ). In contrast, CNV differences were not observed between macrophage populations in tumor and normal samples (Fig. 2D). Third, further supporting a substantial proportion of epithelial cells in the tumor samples as malignant, CNV profiles inferred from scRNA-seq exhibited significant concordance with CNV profiles of the same tumor samples determined by bulk whole-exome sequencing (WES;  $P$ -value range, 0.022 to  $< 0.0001$ ; Supplementary Fig. S3C). At the subcluster level, similar to their high expression of oncogenic gene signatures, EpiInt1 cells also harbored the highest proportion of CNVs (Supplementary Fig. S3D).

Comparisons in a cell state-specific manner revealed that “chief cells” and “intestinal-type cells” showed the greatest differences in transcriptomic profiles between gastric cancers and their cognate cell state in normal gastric tissues (Fig. 2E; “chief cells”: 1,232 genes/EpiC3 cluster; “intestinal-type cells”: 775 genes/EpiInt2 cluster). Pathway analysis indicated that each tumor epithelial cluster shared both common upregulated modules relevant to cancer, but also cluster-specific tumor-associated modules. For example, all tumor epithelial clusters exhibited upregulation of immune-related pathways such as “antigen presentation: folding, assembly, and peptide loading of class I MHC.” However, only EpiC3 exhibited upregulation of “MHC class II antigen presentation” and downregulation of “citric acid (TCA) cycle and respiratory electron transport” (Supplementary Table S5). Among downregulated genes, the classic chief cell markers *LIPF* and *PGA3* were among the top downregulated genes in tumor epithelial cells (Fig. 2F;  $P < 0.0001$ ). Using spatial DSP, we confirmed the loss of *LIPF* in tumor epithelial cells (Pan-CK<sup>+</sup>) compared with normal samples ( $n = 13$ ,  $P = 0.0035$ ; Fig. 2G). Interestingly, we and others have previously identified *LIPF* as a lineage-specific target of recurrent insertion–deletion mutations in gastric cancer (33, 34).

Among stromal metacluster cell states, absolute proportions of fibroblasts (clusters STF1–4) and endothelial cells (clusters STE1 and STE2) were not significantly altered between tumor and normal samples (Fig. 2A; Supplementary Fig. S3E; Supplementary Table S3). However, in the endothelial lineage, pathway analysis revealed a conserved expression response related to extracellular matrix (ECM) remodeling between tumor and normal samples (Supplementary Table S6). Similarly, in the fibroblast lineage, two fibroblast sublineages comprising *LUM*-associated fibroblasts

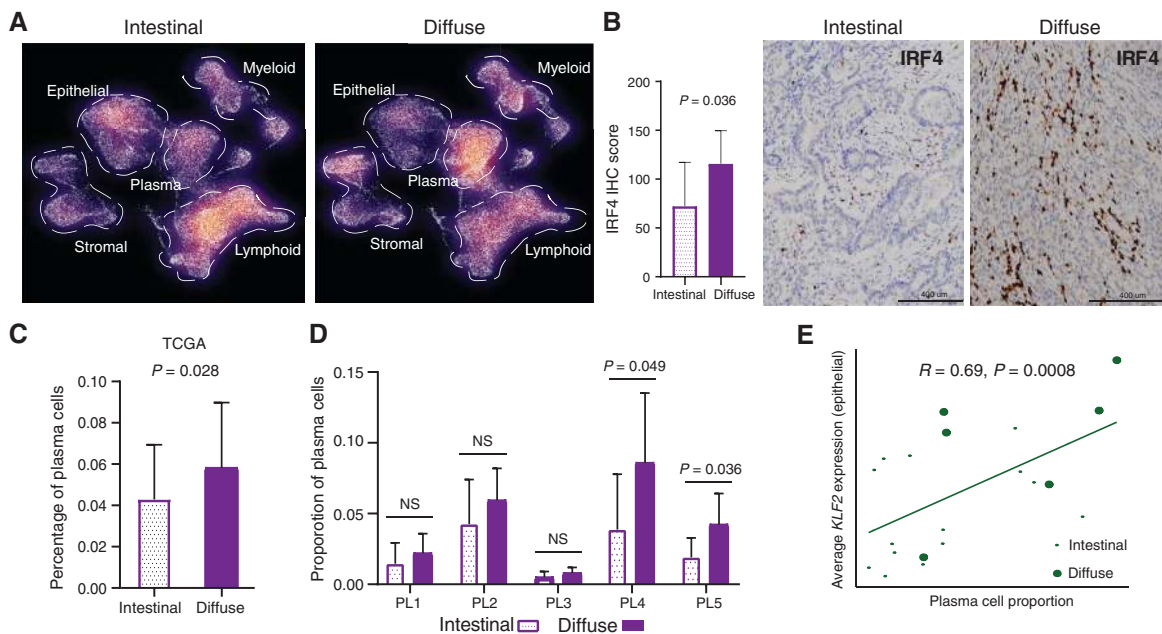
(STF1 and STF3) exhibited upregulation of CAF genes such as *FAP*, *COL8A1*, *THBS2*, and *CTHRC1*, whereas a third sublineage comprising proangiogenic pericytes (STF2) displayed upregulation of another set of CAF marker genes such as *CSPG4*, *PDGFA*, *ASPN*, and *S100A4* (Fig. 2H; Supplementary Fig. S3F). Pathway analysis indicated several common (e.g., “ECM proteoglycans”) and specific pathways individualized to each cluster (Supplementary Table S7). For example, STF2 showed upregulation of “signaling by PDGF,” whereas STF3 had upregulation of “activation of matrix metalloproteinases.”

We also investigated T-cell proportions and immune checkpoints of therapeutic interest between tumor and normal samples (35). Consistent with the prior literature (36), we observed a higher proportion of Tregs ( $P = 0.0048$ ) and naïve CD4 T cells ( $P = 0.048$ ) in tumors compared with normal (Supplementary Fig. S4A and S4B). T-cell receptor (TCR) sequencing revealed TCR diversity ranges of 34.3 to 994, similar to studies in other tumor types (refs. 37, 38; see Supplementary Fig. S4C for further details). Taken collectively, these results suggest that in the gastric cancer tumor ecosystem, distinct cell states from both the epithelial and tumor microenvironment likely express different oncogenic transcriptomic features—each required for different cancer hallmarks and ultimately converging and intermixing to elicit a composite tumor molecular portrait (Fig. 2I; Supplementary Fig. S4D).

### scRNA-seq Reveals a *KLF2*-Associated Plasma Cell Program in Diffuse-Type Gastric Cancer

Lauren’s gastric cancer classification has genomic, clinical, and prognostic value (39). We analyzed our scRNA-seq data set comparing diffuse-type ( $n = 6$ ) and intestinal-type ( $n = 14$ ) tumors, after random downsampling to achieve subtype matching. The relative proportion of plasma cells was higher in diffuse compared with intestinal gastric cancers ( $P = 0.05$ ). In contrast, epithelial cells were lower in diffuse gastric cancers ( $P = 0.04$ ), with intestinal gastric cancers exhibiting higher proportions of epithelial subclusters and inferred CNVs ( $P = 0.029$ ; Supplementary Fig. S5A). There were no statistically significant differences in the lymphoid, myeloid, and stromal cell populations (Supplementary Table S8; Fig. 3A). We confirmed an increased level of plasma cells in diffuse versus intestinal tumors using IRF4 (a plasma cell marker) IHC on a subset of tumors ( $n = 17$ ,  $P = 0.036$ ; Fig. 3B; Supplementary Fig. S5B). We also independently verified the increased level of plasma cells in diffuse-type gastric cancers by reanalyzing bulk RNA-seq gastric cancer TCGA data using CIBERSORTx deconvolution ( $P = 0.028$ ; Fig. 3C). Performing sublineage analysis in single-cell data, we established that the enrichment of plasma cells in diffuse-type gastric cancer was primarily driven by plasma cell clusters PL4 and PL5 ( $P < 0.05$ ; Fig. 3D). Pathway analysis comparing PL4–5 against PL1–3 in tumor samples indicated upregulation of cytokine and interleukin signaling and downregulation of CD22-mediated B-cell receptor (BCR) regulation in PL4–5 (Supplementary Table S9).

*KLF2* is a gene previously shown to regulate homing of plasma cells in multiple myeloma (40, 41). We hypothesized that the increased recruitment of plasma cells in diffuse-type gastric cancers might be mediated by *KLF2* activity. To test this hypothesis, we compared *KLF2* expression between diffuse- and intestinal-subtype gastric cancers across the various



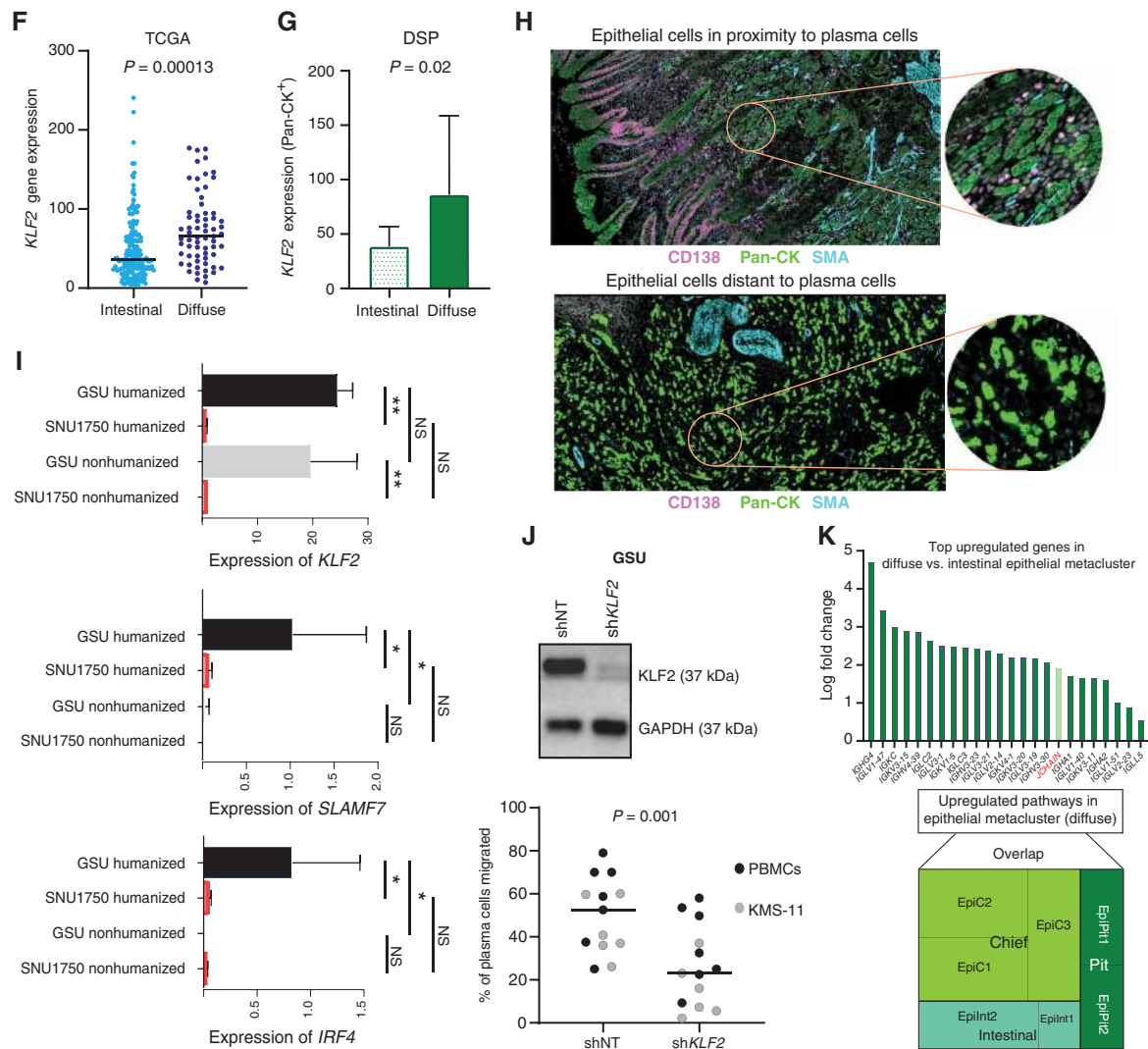
**Figure 3.** Differential TME analysis between histologic subtypes identifies increased plasma cells in diffuse-type tumors. **A**, Density plot of UMAP representation comparing diffuse and intestinal gastric cancer samples after random downsampling to approximately 25,000 cells each. Each dot represents a single cell. Dashed lines highlight a higher proportion of plasma cells in diffuse gastric cancer. **B**, IHC staining of IRF4 expression in diffuse ( $n = 5$ ) and intestinal ( $n = 12$ ) gastric cancer samples (scale bar, 400  $\mu\text{m}$ ). Bar graph showing significantly higher average IRF4 IHC score in diffuse compared with intestinal gastric cancer. **C**, Plasma cell proportions deconvoluted by CIBERSORTX in diffuse and intestinal gastric cancer in the TCGA data set (diffuse  $n = 63$ ; intestinal  $n = 167$ ). **D**, Bar graph showing enrichment of plasma cell proportions in PL4 and PL5 plasma subclusters in diffuse versus intestinal gastric cancer single-cell samples. **E**, Pearson correlation plot showing significant positive correlation of plasma cell proportion to average *KLF2* expression in the epithelial metacluster. (continued on next page)

cell states. Compared with intestinal-type cell states, plasma cells and epithelial cell clusters in diffuse-type gastric cancers expressed increased *KLF2* expression in >50% of cells with log-fold difference >0.5 (Supplementary Fig. S5C). Notably, we observed significant correlations only between plasma cell proportions and *KLF2*-expressing epithelial cells ( $R = 0.69$ ,  $P = 0.0008$ ) and not between plasma cell proportions and *KLF2*-expressing plasma cells ( $R = 0.35$ ,  $P = 0.13$ ; Fig. 3E; Supplementary Fig. S5D). This finding suggests that high-*KLF2*-expressing epithelial cells may be associated with plasma cell recruitment in diffuse-type gastric cancer. Sublineage analysis of diffuse-type epithelial cells revealed the highest *KLF2* expression in EpiC clusters (log fold from 1.7 to 0.8, compared with EpiC intestinal-type cells; Supplementary Table S10). To further confirm *KLF2* expression differences between intestinal-type and diffuse-type gastric cancers, we analyzed epigenomic promoter activity [as measured by H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq)] and gene expression of *KLF2* in an independent bulk RNA data set of 24 gastric cancers (9 diffuse-type and 14 intestinal-type). Both *KLF2* H3K27ac promoter levels and gene expression were higher in diffuse compared with intestinal gastric cancers ( $P < 0.05$ ), with good correlations between promoter signals and gene expression ( $R = 0.55$ ,  $P = 0.005$ ; Supplementary Fig. S5E). In a second independent bulk RNA-seq data set (gastric cancer TCGA), we further confirmed increased *KLF2* expression in diffuse-type compared with intestinal-type tumors ( $P = 0.00013$ ), with boundaries of variance similar to with other credentialed

oncogenes such as *ERBB2* and *HNF4 $\alpha$*  associated with intestinal-type gastric cancer (Fig. 3F; Supplementary Fig. S5F; refs. 42, 43). To explore *KLF2* differences in a spatial context, we then used DSP analysis to study the expression of *KLF2* in epithelial cells (Pan-CK<sup>+</sup>) and confirmed higher *KLF2* transcript levels in diffuse-type epithelial cells compared with intestinal-type epithelial cells ( $P = 0.02$ ;  $n = 10$ ; Fig. 3G). Using the same platform, we next studied *KLF2* expression in epithelial cells (Pan-CK<sup>+</sup>) in the context of proximity to plasma cells (CD138<sup>+</sup>; Epi<sup>prox</sup> vs. Epi<sup>dist</sup>). We observed a trend toward higher *KLF2* expression in epithelial cells in close proximity to plasma cells compared with distal epithelial cells ( $P = 0.096$ ; Fig. 3H; Supplementary Fig. S5G).

To explore dynamic temporal interactions between *KLF2* expression in tumor epithelial cells with host plasma cells, we used a humanized mouse cancer *in vivo* model. Immune-deficient mouse pups were engrafted with human umbilical cord blood CD34<sup>+</sup> cells, and mice with postengraftment human immune-cell reconstitution (termed “humanized mice”) were selected for the study. We investigated *KLF2*<sup>pos</sup> (GSU) and *KLF2*<sup>neg</sup> (SNU1750) diffuse-type gastric cancer tumors grown in humanized mice compared with immune-deficient mice (Fig. 3I). In agreement with our results in primary gastric cancers, *KLF2*<sup>pos</sup> xenografts had high expression of plasma cell markers *IRF4* and *SLAMF7* in the humanized mice only, reflecting increased plasma cell recruitment associated with *KLF2*. To further demonstrate a causative role for epithelial *KLF2* in plasma cell recruitment, we then performed *in vitro*





**Figure 3. (Continued)** **F**, Bee swarm plot showing increased *KLF2* expression in diffuse versus intestinal gastric cancer samples in the bulk RNA-seq TCGA-STAD data set. **G**, Bar graph showing increased *KLF2* expression in Pan-CK<sup>+</sup> epithelial morphologic regions of diffuse versus intestinal gastric cancer by DSP ( $n = 10$ ). **H**, DSP analysis depicting epithelial ROIs proximal (top) and distal to plasma cells (bottom). Analysis is based on immunofluorescence staining for Pan-CK (epithelial, green), CD138 (plasma, pink), smooth muscle actin (SMA; fibroblast, cyan), and DAPI (blue). Each circular ROI is 300  $\mu\text{m}$  in diameter. **I**, Bar graph showing expression of *KLF2*, *IRF4*, and *SLAMF7* genes in GSU humanized and nonhumanized mice against SNU1750 humanized and nonhumanized mice. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . **J**, Western blot (top) showing stable knockdown of *KLF2* in gastric cancer cell line GSU (sh*KLF2*) compared with shNT (nontargeting) control. Loss of *KLF2* in GSU significantly reduces migration of plasma cells derived from peripheral blood mononuclear cells (PBMC) and multiple myeloma cell line KMS-11 ( $N = 17$ ; bottom). **K**, Upregulation of immunoglobulin genes in diffuse versus intestinal epithelial metacluster (top). Tree map shows the overlap of upregulated pathways in epithelial metacluster versus subclusters, with the EpiC cluster showing the greatest overlap (bottom).

migration assays using plasma cells derived from peripheral blood mononuclear cells (PBMC) or KMS-11 (a multiple myeloma cell line) cocultured with *KLF2*-positive gastric cancer cells. Knockdown of *KLF2* expression in two independent *KLF2*-positive gastric cancer cell lines (GSU and LMSU) was sufficient to significantly decrease plasma cell migration ( $P = 0.001$ ; Fig. 3J; Supplementary Fig. S5H). Taken collectively, our results pinpoint specific B-cell sublineages exhibiting increased proportions in diffuse-type gastric cancer and highlight epithelial cell-resident *KLF2* expression as a potential driver of plasma cell recruitment.

We further studied the relationship between epithelial cells from diffuse-type gastric cancers with pathways related

to immune-cell biology. scRNA-seq-driven pathway analysis between diffuse- and intestinal-subtype epithelial cells revealed several upregulated genes in diffuse-type epithelial cells related to immune-signaling pathways, including numerous immunoglobulin (Ig) genes belonging to both light-chain and heavy-chain genes (IgG/IgA) and the Ig linker gene *JCHAIN* (Fig. 3K, top panel; Supplementary Table S11). A subcluster pathway analysis of upregulated genes revealed that most of these immune-related modules were expressed in EpiC cells, similar to *KLF2* (Fig. 3K, bottom panel; Supplementary Table S12). In contrast, downregulated genes did not show pathway enrichments related to immune signaling. Taken together, these data suggest a general transcriptional

cassette expressed in diffuse-type gastric cancer epithelial cells related to engaging the tumor immune microenvironment.

### scRNA-seq Reveals Distinct Fibroblast Populations and *INHBA*-*FAP* Axis as a CAF Regulator

CAFs are known to influence tumor growth, migration, and invasion through the regulation of ECM components in various tumor types (44, 45). However, little is presently known about specific pathways of CAF regulation and heterogeneity in gastric cancer, due to limited *in vitro* models of gastric cancer CAFs and the challenges of deconvoluting bulk RNA-seq data (46). Using our scRNA-seq data, we first compared the tumor fibroblast clusters (STF1–3) according to clinical stage and histologic subtype. Supporting an important role for CAFs in gastric malignancy, we observed a stage-dependent increase in all three STF clusters, with STF3 as the dominant population (Fig. 4A). We next closely assessed a panel of CAF canonical markers, *FAP*, *CSPG4*, *ACTA2*, and *TAGLN* (Supplementary Table S4), in STF1–S3. Although STF1 and STF3 clusters exhibited upregulation of *FAP*, *ACTA2*, and *TAGLN*, the STF2 cluster exhibited upregulation of only *CSPG4* (Fig. 4B), further indicative of distinct CAF sublineages in gastric cancer.

TGF $\beta$  superfamily signaling has been reported to influence CAF function in other cancer types (47). We thus elected to study the activin-inhibin signaling module, a major component of the TGF $\beta$  pathway, comprising nine canonical genes (Supplementary Table S4). Among these nine genes, only *INHBA* exhibited significant upregulation in tumor-associated STF2 and STF3 clusters compared with normal (Fig. 4C). As tumor STF2 (*CSPG4*-high) and STF3 (*FAP*-high) represent distinct CAF populations, we then performed a coexpression correlation analysis between *INHBA* expression and the respective cluster markers. *INHBA* exhibited a significant positive correlation with *FAP* in STF3 ( $R = 0.21$ ,  $P < 0.0001$ ), whereas no correlations were found with *CSPG4* in STF2 ( $R = -0.03$ ; Supplementary Fig. S6A). In TCGA bulk RNA-seq data, *INHBA* was similarly correlated to *FAP* (Pearson  $R = 0.59$ ;  $P < 0.0001$ ) but not to *CSPG4* (Pearson  $R = 0.00$ ;  $P = 0.964$ ; Supplementary Fig. S6B). To orthogonally support the association between *FAP* and *INHBA*, we adopted multiple approaches. First, we analyzed spatial DSP data of *FAP* and *INHBA* in fibroblast regions marked by  $\alpha$ -smooth muscle actin ( $n = 13$  samples; 4 ROI per sample; Fig. 4D). We observed higher expression of both *FAP* and *INHBA* in tumor fibroblasts compared with normal ( $P < 0.05$ ; Fig. 4E and F), and within tumor fibroblasts there was a strong correlation between *FAP* and *INHBA* coexpression levels ( $R = 0.92$ ,  $P < 0.0001$ ; Fig. 4G). Second, in an independent cohort of 10 *in vitro* cultured patient-matched normal and tumor fibroblasts, isolated and purified by fluorescence-activated cell sorting from patients with gastric cancer (46), we confirmed *INHBA* upregulation ( $P = 0.023$ ) in tumor fibroblasts, along with increased *FAP* ( $P = 0.027$ ; Fig. 4H). Third, siRNA-mediated knockdown of *INHBA* in the tumor fibroblasts resulted in significant *FAP* gene downregulation ( $P < 0.001$ ; Fig. 4I; Supplementary Fig. S6C). Conversely, treatment of two normal fibroblast lines with recombinant *INHBA* was sufficient to increase *FAP* expression at 48 ( $P < 0.001$ ) and

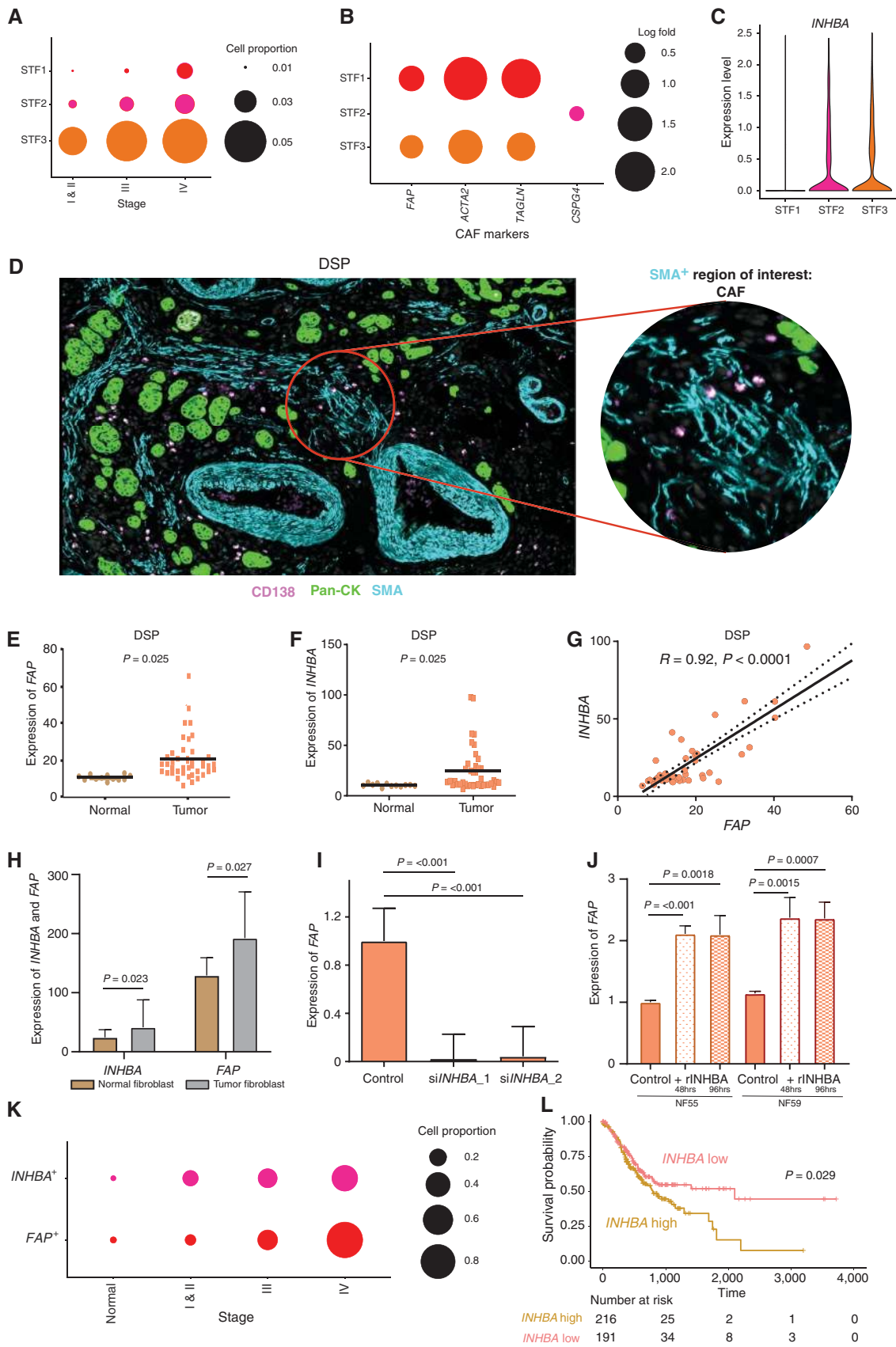
96 ( $P < 0.001$ ) hours (Fig. 4J). Taken collectively, these results highlight *INHBA* as a positive regulator of *FAP* in the gastric cancer STF3 fibroblast population.

We also studied correlations between *INHBA* expression and collagen-related gene expression (48–52), a surrogate for fibrogenic processes regulated by TGF $\beta$  signaling. Of nine collagen genes, five positively correlated with *INHBA*, including *COL1A1*, *COL1A2*, *COL6A3*, *COL8A1*, and *COL12A1* ( $R > 0.2$ ,  $P < 0.0001$ ; Supplementary Fig. S6D). To further verify these findings, we then analyzed DSP data and again confirmed a significant positive correlation of *INHBA* with *COL1A1*, *COL1A2*, and *COL6A3* ( $R > 0.69$ ,  $P < 0.0001$ ; Supplementary Fig. S6E; *COL8A1* and *COL12A1* were not represented on the DSP platform). These results support a positive relationship between *INHBA* activity and collagen gene induction. To explore the potential paracrine capability of *INHBA* (53), we then treated CAFs *in vitro* with recombinant *INHBA* for 48 hours and measured the expression of collagen genes. We observed significant increases in collagen genes, including TGF $\beta$  targets *COL1A1*, *COL1A2*, and *COL6A3* in two CAF lines (Supplementary Fig. S6F), consistent with a cell nonautonomous role for *INHBA*.

To assess the clinical relevance of our findings, we mapped *FAP* and *INHBA* expression levels across the gastric cancer cohort. We found an increasing abundance of *FAP*-positive and *INHBA*-positive cells in STF3, in a stage-wise manner from normal to stage IV (Fig. 4K;  $P = 0.041$ , after normalizing for differences in STF3 proportion). Survival analysis of TCGA samples by *INHBA* levels revealed a significantly poorer survival for tumor samples with high *INHBA* expression (HR: 0.70; 95% CI, 0.51–0.97,  $P = 0.029$ ; Fig. 4L). This difference remained statistically significant even after adjusting for stage (HR: 0.71,  $P = 0.038$ ). Similar findings were also seen in a pooled analysis of several available gastric cancer microarray data (ref. 54; *INHBA* high vs. *INHBA* low, HR: 0.80; 95% CI, 0.68–0.96,  $P = 0.014$ ; Supplementary Fig. S6G).

### scRNA-seq of Gastric Cancer Organoids Supports Increased Cancer Cell Transcriptional Plasticity

Finally, PDOs are increasingly being used as a platform to model gastric cancer tumor behavior and drug responses (55–57). To investigate the extent to which *in vitro* organoid culture conditions affect transcriptional lineage states or overall cellular proportions compared with *in vivo* primary gastric cancers, we derived and examined four pairs of matched normal-tumor PDOs ( $n = 48,531$  cells; <12 passages; Methods; Supplementary Fig. S7A). To perform a comparative cell-state analysis between PDOs and primary gastric cancers, we integrated the scRNA-seq data set from PDOs with the primary tumors. This analysis recapitulated the five major metaclusters (Fig. 5A). We then performed a sublineage-level analysis of epithelial and stromal metaclusters in the PDOs. For the PDO epithelial metacluster, the proportion of the three sublineages was EpiC (71% in PDO vs. 43% in primary), EpiInt (15% vs. 30%), and EpiPit (14% vs. 27%). In the PDO stromal clusters, we also observed pericytes (28% vs. 14% in primary), fibroblasts (32% vs. 45% in primary), and endothelial cells (39% vs. 37%; Supplementary Fig. S7B). These findings suggest that sublineage heterogeneity is indeed present in PDOs, although proportions may differ compared with primary gastric cancers.



Similar to primary gastric cancers, tumor PDO epithelial cells exhibited upregulation of cancer-associated modules and gastric cancer-related genes compared with normal PDO epithelial cells ( $P < 0.0001$ ; Fig. 5B). Interestingly, trajectory plot analysis of epithelial cells in both normal and tumor PDOs demonstrated that epithelial cells from tumor organoids have multiple extended branches relative to normal epithelial cells, consistent with tumor PDO epithelial cells undergoing pervasive and ongoing differentiation/dedifferentiation (Fig. 5C). These findings are noteworthy given recent studies reporting that tumor-associated epithelial cells have increased transcriptional plasticity, which may drive key aspects of ITH (11). Our ability to observe similar phenomena in PDOs suggests that PDOs could be used as an *in vitro* experimental model to investigate molecular pathways governing transcriptional plasticity in gastric cancer.

Our analysis also revealed differences between primary samples and PDOs. For example, epithelial and stromal metaclusters in PDOs were significantly enriched, relative to lymphoid and plasma cell clusters that were depleted (Fig. 5D;  $P < 0.001$  for stromal, epithelial, and lymphoid;  $P = 0.018$  for plasma). A gene-expression comparison between PDO and primary samples indicated plasma cells as showing the greatest differences in gene-expression profile in PDOs, whereas epithelial signatures were relatively more conserved (Fig. 5E). In congruence with these findings, comparison of Reactome programs (58) between the PDOs and primary tumors by metaclusters confirmed a high overlap of Reactome programs for the epithelial [Jaccard similarity index (JSI) = 0.62] and stromal (JSI = 0.43) metaclusters, with poorer overlap for the plasma (JSI = 0.09) and lymphoid (JSI = 0.15) metaclusters. Pathway analysis of genes unique to plasma cells implicated “classical antibody-mediated complement activation” as significantly downregulated (Fig. 5F; Supplementary Table S13). Notably, a significant proportion of genes and pathways were also commonly upregulated in PDOs agnostic of cluster, which included “cellular response to starvation,” highlighting umbrella culture effects (Fig. 5G; Supplementary Table S14). Together, these results raise the possibility that besides epithelial cells, PDO culture may also influence the transcriptional profiles of other cell types associated with tumors.

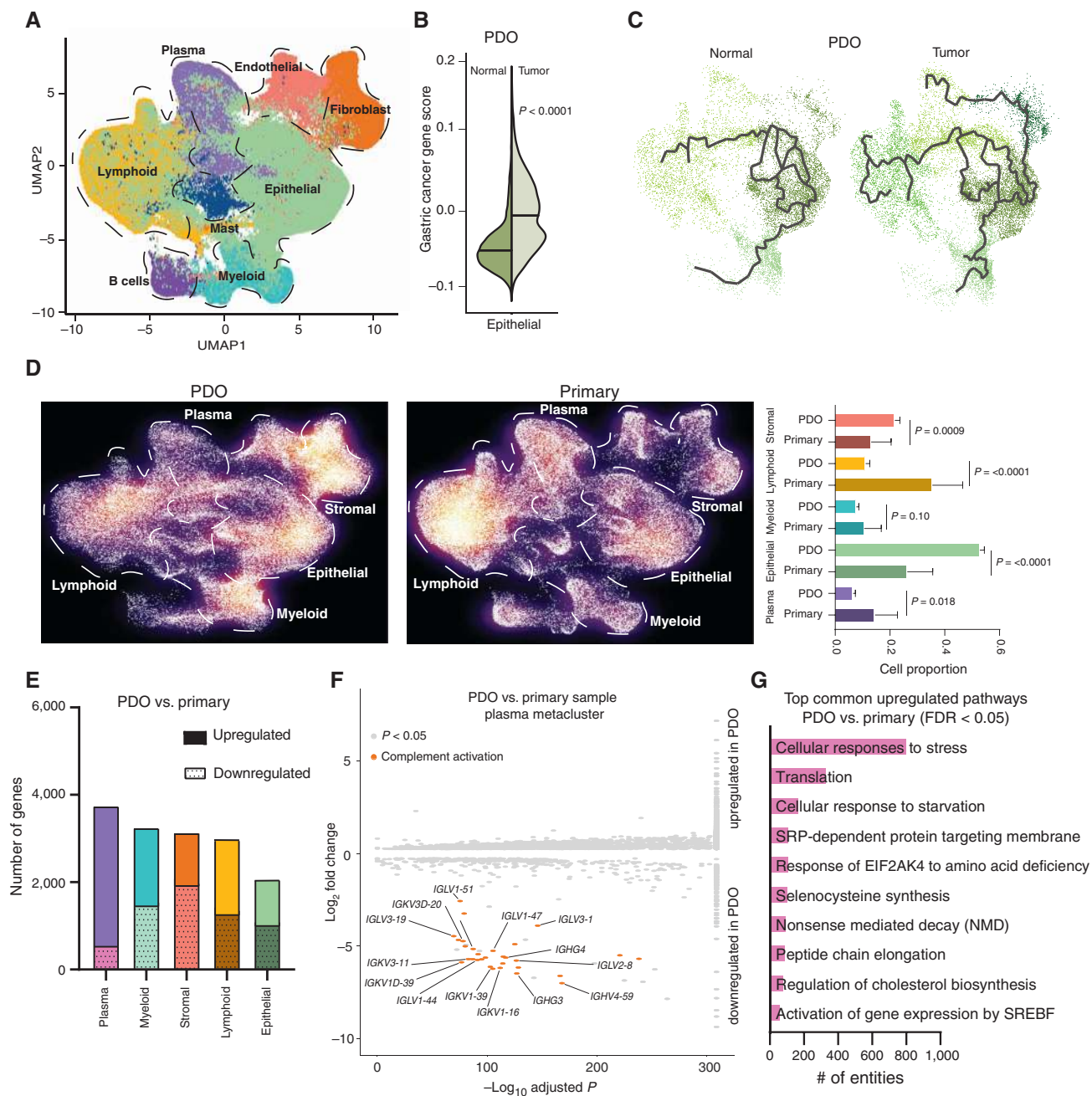
## DISCUSSION

In this study, we applied scRNA-seq across a large number of cells to discover several novel features of gastric cancer,

including rare cell populations undergoing state transitions, cell type-specific expression programs associated with gastric cancer, and distinct plasma cell and CAF sublineages associated with gastric cancer histologic subtypes and clinical stages. Another notable aspect of our study was the application of very recently available spatial transcriptomics technologies (DSP) to orthogonally verify our key findings. Compared with earlier gastric cancer scRNA-seq studies, our experimental design and analyses are differentiated by a large cell number size (>200,000 cells, higher than all prior gastric cancer studies combined), samples reflecting multiple clinical stages and subtypes ( $n = 31$  patients, 48 samples), and the comparative analysis of gastric organoids (Fig. 6). For example, Zhang and colleagues performed scRNA-seq on gastric cancer biopsy samples ( $n = \sim 30,000$  cells) from patients with premalignant lesions (atrophic gastritis and IM), and one early gastric cancer sample, focusing on the evolution of epithelial cells from normal to malignancy (14). At the terminal end of the cancer spectrum, Wang and colleagues performed scRNA-seq only on malignant ascites cells ( $n = \sim 45,000$  cells; ref. 11). Sathe and colleagues studied  $\sim 55,000$  cells from seven patients with gastric cancer and one patient with IM, generating a receptor–ligand network of the TME components, agnostic of tumor subtypes (13).

scRNA-seq analysis of tumor–normal comparisons performed in a cell-lineage-specific manner led to the emergence of a composite tumor profile assembled from distinct faulty signatures expressed by different lineages. Our results suggest that the greatest tumor-associated expression differences in the tumor epithelial component likely originate from chief cells and intestinal-type cells. Interestingly, high gastric cancer oncogene expression was observed in a specific subpopulation of intestinal-type epithelial cells (EpiInt1), suggesting that EpiInt1 cells likely represent a key epithelial cell population important in the transition into malignancy from metaplasia. One potential limitation of our study is the lack of directly inferred single cell-based DNA alterations, as currently available platforms are unable to deliver both single-cell DNA-level and RNA-level alterations from the same cell on a genome-wide scale. Additionally, the global clustering methodology used in our study does not preclude the possibility that more granular cell types may exist, which could be identified using refined local clustering (59, 60). Among nonepithelial cell types, lineage analysis of fibroblasts identified discrete CAF clusters, such as CAFs that were *LUM* positive or pericyte CAFs that were *CSPG* positive.

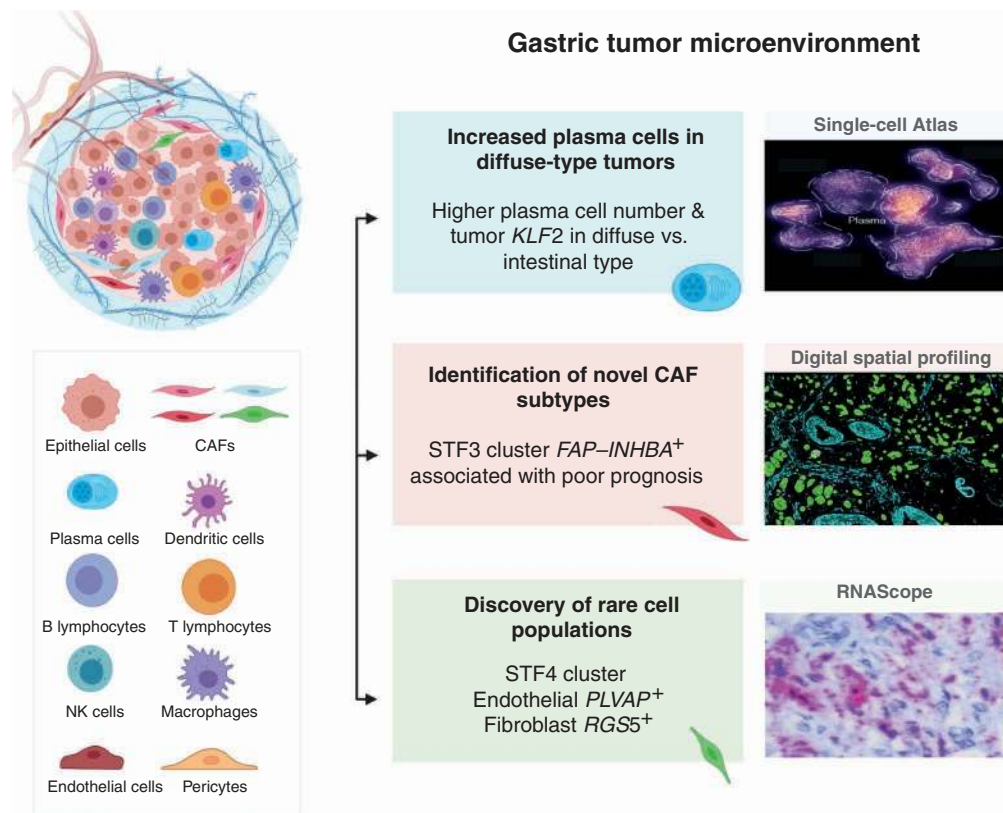
**Figure 4.** scRNA-seq enables identification of distinct gastric cancer fibroblast subtypes and *INHBA*–*FAP* axis as a CAF regulator. **A**, Bubble plots demonstrating stage-dependent increases in the proportion of fibroblast cells with *STF3* being the dominant subcluster. The size of the circle represents the proportion of cells expressing subcluster-specific genes. **B**, Bubble plots showing fibroblast subclusters (*STF1*–*3*) expressing distinct CAF canonical markers (*FAP*, *CSPG4*, *ACTA2*, and *TAGLN*). The size of the circle represents the proportion of cells expressing different genes. **C**, Violin plot showing the expression of *INHBA* in *STF2* and *STF3* fibroblast clusters with negligible expression in the *STF1* fibroblast cluster. **D**, Fibroblast ROIs captured by DSP analysis based on immunofluorescence staining for Pan-CK (epithelial, green), CD138 (plasma, pink), SMA (fibroblast, cyan), and DAPI (blue). The circular ROI is 300  $\mu\text{m}$  in diameter. **E**, Bee swarm plot showing differential expression of *FAP* in fibroblast ROIs of normal and tumor samples by DSP ( $n = 13$ ). **F**, Bee swarm plot showing differential expression of *INHBA* in fibroblast ROIs of normal and tumor samples by DSP ( $n = 13$ ). **G**, Pearson correlation graph demonstrating strong positive correlations between *INHBA* and *FAP* gene expression in fibroblast ROIs using DSP. **H**, Bar graph showing significant expression of *FAP* and *INHBA* genes in flow-sorted tumor fibroblasts compared with matched normal fibroblasts ( $n = 10$  each). **I**, Bar graph showing significant reduction in *FAP* gene expression after siRNA-mediated *INHBA* knockdown in tumor fibroblast lines. Two independent siRNAs were used. **J**, Bar graph showing significant increases in *FAP* gene expression in two normal fibroblast lines after treatment with recombinant *INHBA* (r*INHBA*) for 48 and 96 hours, respectively. **K**, Bubble plot depicting stage-dependent increases of *FAP*<sup>+</sup> and *INHBA*<sup>+</sup> cells in fibroblast cluster *STF3* ( $P = 0.041$ ). The circle sizes represent the relative proportion of cells expressing these genes.  $P$  values were computed using Kendall's  $\tau$  method. **L**, Kaplan–Meier survival curves of TCGA-STAD data showing significant differences in overall survival between *INHBA*-high and *INHBA*-low samples.  $P$  values were computed using log-rank tests.



**Figure 5.** Comparative analysis of primary and organoid samples. **A**, UMAP representation of approximately 200,000 cells (–48,000 cells from tumor PDOs with matched normal PDOs, combined with primary samples; –152,000 cells) recapitulating the major five metaclusters color-coded by their cell types. Each dot in the UMAP represents a single cell. **B**, Violin plot showing the expression of gastric cancer gene module scores in tumor PDOs compared with matched normal samples. **C**, Trajectory plot analysis of epithelial cells from tumor and normal PDOs demonstrating the expression of cellular differentiation gene programs in tumor PDOs depicted by long multiple branches. **D**, Density plot of UMAP representation comparing PDO and primary gastric samples demonstrating enrichment of lymphoid and plasma metaclusters in primary samples compared with PDOs. **E**, Graph showing the number of upregulated and downregulated genes in PDOs versus primary samples in the five metaclusters. The plasma meta-cluster shows the highest number of differentially expressed genes as compared with other metaclusters. **F**, Volcano plot of upregulated and downregulated genes in the plasma metacluster between PDOs and primary samples, showing significant downregulation of antibody-mediated complement factor genes in PDOs. x-axis shows the  $-\log_{10}$  adjusted P value and y-axis  $\log_2$  fold change in gene expression. **G**, Top common upregulated pathways in PDOs versus primary samples across all metaclusters.

Both *LUM* and *CSPG* encode proteoglycans, and pericytes are being increasingly recognized as key players in tumor vessel formation and growth (61). The differences in gene-expression profiles between these two fibroblast clusters may highlight distinct tumor-promoting functions. For example,

*LUM*-associated CAFs may be involved in the proliferation and survival of tumor cells, whereas pericyte CAFs may be involved in neovascularization, thereby underscoring the functional heterogeneity of CAFs in gastric cancer. We also identified a small cluster of stromal cells (*STF4*) undergoing



**Figure 6.** Comprehensive single-cell atlas of gastric cancer. This study included more than 200,000 cells from 31 primary gastric tumor samples. In total, 34 distinct cell-lineage states were identified, related by developmental trajectories and previously unreported rare cell populations. An increase in plasma cell proportions was observed as a feature of diffuse-type tumors associated with epithelial-resident *KLF2*. A stage-wise accrual of novel cancer-associated fibroblast subpopulations was marked by high *INHBA* and *FAP* coexpression. Findings were complemented using digital spatial transcriptomics and RNAScope. Our results provide a high-resolution molecular resource for gastric cancer translational studies, identifying intra- and interpatient lineage states across distinct gastric cancer subtypes.

EndoMT—these cells expressed markers associated with both endothelial (*PLVAP*) and fibroblast (*RGS5*) lineages, confirmed after doublet filtering and by orthogonal RNAScope analysis. EndoMT, an embryonic process required for normal cardiovascular development (62), has been associated with TME plasticity, resistance to antineoplastic therapy, and TGF $\beta$  and BMP signaling (63–65). Together, these scRNA-seq findings suggest a conceptual model wherein individual cell lineages respond distinctly, either directly or indirectly, to malignant transformation, leading to a highly complex tumor ecosystem. It is reasonable to posit that these observations would have been largely obscured in bulk RNA-seq data.

CAFs, representing a predominant stromal cell population, have been shown to play cardinal roles in shaping tumor growth and metastasis in various tumor types (46, 66, 67). CAFs are thought to largely affect tumor behavior via ECM modeling, secretion of soluble factors, and promoting angiogenesis (66). In the case of gastric cancer, although some studies have investigated CAF-associated signaling pathways in tumor cells, molecular events underpinning CAF function and heterogeneity remain poorly defined. In our data set, we found that gastric cancer CAFs have at least three distinct subtypes (STF1–3), each expressing distinct subsets of canonical CAF markers. Moreover, *FAP*-high STF3 cells also exhibited high *INHBA* coexpression, a correlation we confirmed in multiple

orthogonal settings inclusive of a clinical stage-wise gradation. Functionally, silencing of *INHBA* affected *FAP* levels in gastric cancer CAFs, implying a potentially direct role for *INHBA* in regulating *FAP* expression, and consistent with TGF $\beta$  signaling as a mediator of *INHBA* (68). A functional role for *INHBA* in gastric cancer CAFs has also been demonstrated by a recent study in gastric cancer that used bulk RNA-seq of laser-capture microdissection (LCM)-derived CAF samples (69). Clinically, patients with high *INHBA*-expressing tumors exhibited poorer survival outcomes in multiple gastric cancer cohorts, consistent with a tumorigenic function for *INHBA* (70–72). *INHBA* encodes a subunit of activin and inhibin, which have been reported to play opposing roles in many facets of normal biology and disease (73). It is possible that high *INHBA* expression levels may facilitate the formation of *INHBA* homodimers, otherwise known as Activin A, leading to the activation of activin receptors such as *ACVRI*, which has established roles in cancer (74). In other cancer types, aberrant increases in *INHBA* expression have been reported in both the epithelial and CAF components, involving autocrine and paracrine functions (75–78). Our observations put forth the *INHBA* pathway as a potential target to disrupt CAF function and warrant testing of drug modalities in appropriate model systems (79).

In most parts of the world, Lauren's histopathologic subtypes are frequently used as reference standards in gastric

clinical pathology. Evolution of intestinal-type gastric cancers is usually ascribed to the Correa cascade, and the etiology and molecular features of diffuse-type gastric cancer are not well understood (80). In our study, scRNA-seq enabled lineage-based comparisons of the TME between these histologic subtypes, leading to the discovery of increased plasma cells in diffuse-type gastric cancers. Compared with T cells, there are fewer studies on B-cell and plasma cell populations in gastric cancer (81, 82). Derks and colleagues reported a higher proportion of tertiary lymphoid structures in genomically stable (GS) tumors with enrichment of B cells and CD4 T cells (83). One report by Katoh and colleagues indicated a finding of increased B-cell lineages (which includes plasma cells) in diffuse-type and GS gastric cancers using the TCGA data set and identified sulfated glycosaminoglycan as a key functional B-cell antigen in these tumors (84). Notably, B-cell/plasma cell infiltration has been associated with protumor (i.e., immune suppressive) and antitumor (i.e., immune active) growth in other cancers (85). For gastric cancer, we speculate a “protumor” function for plasma cell infiltration, because increased plasma cells are associated with diffuse-type gastric cancers that respond poorly to immune checkpoint inhibitors (86, 87) and exhibit classic “immune-suppressive” features (88). To identify mechanisms responsible for the plasma cell increase, we investigated KLF2, a transcription factor previously shown to modulate multiple myeloma cell adhesion and homing of plasma cells (40). We observed that *KLF2* expression in diffuse-type epithelial cells (EpiC cluster) was positively correlated with plasma cell proportions, and increased *KLF2* expression in diffuse-type and GS gastric cancers (Supplementary Table S15) in the TCGA cohort was consistent with a regulatory role for epithelial KLF2 in shaping plasma cell populations. We functionally explored this possibility in “humanized mice,” by xenografting human-derived diffuse-type gastric cancer cell lines that were *KLF2* positive and negative. We found that only tumors with *KLF2* had human plasma cell markers, implying a more casual role for epithelial *KLF2* expression in plasma cell recruitment. Using DSP, we also gained spatial insights of *KLF2* expression on primary samples and observed a trend for higher *KLF2* expression in epithelial cells proximate to plasma cells, further suggestive of a functional interaction between these cell types. The precise mechanisms underlying the plasma cell enrichment remain undefined. Earlier studies have reported that *KLF2* can modulate proinflammatory gene expression in monocytes and endothelial cells (89–91), and our findings suggest a role for KLF2 in driving tumor epithelial cell programs possibly involving paracrine cell signaling pathways (e.g., exosome-derived miRNAs; refs. 92–95). Interestingly, besides *KLF2*, we also found that epithelial cells in diffuse-type tumors (EpiC cluster) displayed a striking upregulation of multiple immune-signaling pathways, including Ig genes and *JCHAIN*, a gene responsible for the secretory form of Ig, mimicking gene-expression signatures of immune cells. This observation of an “epithelial-immune cell state” has been described in other cancers (96, 97). We report the existence of this phenomenon for the first time in diffuse-type tumors, adding a novel and singular dimension to gastric epithelial cell behavior. Probing the mechanisms underlying epithelial–plasma cell cross-talk in diffuse-type gastric cancers may pave

the way for new intervention strategies to modulate TME for this recalcitrant subtype.

Among the platforms to study gastric cancer, PDOs have recently emerged as a promising system for *ex vivo* testing of therapeutic agents, precision oncology, and assessing driver gene function (98, 99). Our comparison of PDO-primary samples using scRNA-seq revealed that PDOs indeed maintained most major cell types, except for a depletion in lymphoid and plasma cell lineages. However, it should be noted that these PDOs represent a single snapshot, and assessment of organoid cellular compositions over time represents an area of future research. Global gene-expression analyses indicated the greatest changes in plasma cells compared with epithelial state. This suggests that PDOs may serve well to study gastric epithelial biology, whereas immune cells may be more affected by the process of culturing. These factors must be considered when using PDOs for personalized *ex vivo* drug testing, especially those involving immunologic pathways or characterizing TME differences. Interestingly, our data reconfirm the association of *ARID1A* loss with the induction of mucinous phenotype, an observation first made by Lo and colleagues (ref. 99; Supplementary Fig. S7C).

In conclusion, in one of the largest single-cell analyses of gastric cancer performed to date, our study forms a unique resource for generating novel biological insights on tumor cell types, subtype-based TME compositions, and cell–cell interactions in gastric tumors. In terms of future directions, our data support the need for further in-depth studies on plasma cell homing biology guided by epithelial KLF2 and the potential clinical implications of perturbing these interactions. The role of INHBA and the TGF $\beta$  superfamily in the regulation of CAFs also carries potential for therapeutic target and/or predictive biomarker discovery. We anticipate future work to utilize combinatorial single-cell approaches, including epigenetic, genetic, and transcriptional layers and spatial context to enhance our understanding of the gastric tumor architecture.

## METHODS

### Ethics Declaration

The study was approved by the local ethics board (National Healthcare Group, Domain Specific Review Board Ref Nos: 2005/00440 and 2016/00059). All animal experiments and procedures were approved by the Institutional Animal Care and Use Committee (IACUC; IACUC #191440) of A\*STAR in accordance with guidelines of Agri-Food and Veterinary Authority and the National Advisory Committee for Laboratory Animal Research of Singapore. Primary CAFs and normal fibroblasts (NF) were isolated from patients with gastric cancer who underwent gastrectomy without preoperative treatment at Kumamoto University after written informed consent was obtained from each patient. The study was approved by the Medical Ethics Committee of Kumamoto University (Approval Number: 1277). PBMCs were isolated from human blood obtained from healthy volunteers with written informed consent. Study protocols were approved by the SingHealth Centralised Institutional Review Board (reference number: 2017/2806). Protocols were performed in accordance with the Declaration of Helsinki for Human Research.

### Sample Acquisition and Tissue Processing

Patients diagnosed with gastric adenocarcinoma and undergoing surgical resection or endoscopy at the National University

Hospital, Singapore, were enrolled after written informed consent was obtained. On-table endoscopic biopsies or surgical resection samples were harvested. For surgical samples, matched normal gastric tissues from sites displaced at least several centimeters from the tumor were used. Tissues from distant metastases to the peritoneum were taken during diagnostic laparoscopy. Tissues were collected in MACS tissue storage buffer (Miltenyi Biotec) immediately after biopsy or resection and stored on ice. Samples were processed using enzymatic and mechanical dissociation using a human tumor dissociation kit and the Gentle MACS Octodissociator (Miltenyi Biotec) following the manufacturer's "37\_h\_TDK\_2" program. Dissociated cells were passed through a MACS smartstrainer (70  $\mu$ m) and incubated with RBC lysis buffer for 5 minutes followed by PBS neutralization. All centrifugation steps were carried out at 300  $\times$ g for 7 minutes. Dissociated cells were washed twice in PBS + 1% bovine serum albumin (BSA) and filtered through a 40- $\mu$ m smartstrainer. Live-cell counts were obtained by manual cell counting using 1:1 trypan blue dilution. Cells were concentrated to 800–1,200 live cells/ $\mu$ L and then processed for single-cell analysis.

### Single-Cell Sequencing

Samples from each patient were processed in a single batch for library preparation. The Chromium Single-Cell 5' Library and Gel Bead Kit (10 $\times$  Genomics) was used according to the manufacturer's protocols. Briefly, gel bead-based emulsions (GEM) were generated by combining barcoded single-cell 5' Gel Beads, cells, and partitioning oil. Ten times barcoded, full-length cDNAs generated from GEMs were amplified by PCR. Enriched libraries were enzymatically digested, size selected, and adaptor ligated for sequencing. To obtain a TCR repertoire profile, VDJ enrichment was carried out as per the Chromium Single-Cell V(D)J Enrichment Kit, Human T-cell #1000005 (10 $\times$  Genomics) using the same input samples. Sequencing libraries were generated with unique sample indices for each sample and quantified using the Kapa library kit. Quantified libraries were sequenced on an Illumina HiSeq4000 sequencer.

### Bioinformatic QC, Normalization, Clustering, and Differential Gene Expression of Single-Cell Data

Cellranger v3.0 (<https://support.10xgenomics.com/single-cell-gene-expression/software/>) was used to align FASTQ sequencing reads to the hg38 reference transcriptome, generating single-cell feature counts for each sample. Using Seurat version 3.0 (100), each sample was considered for genes/features shared by three or more cells, and cells showing 500 or more features and fewer than 6,000 features. Cells with mitochondrial RNA percentages of >20 were filtered out. SCTransform normalization was performed on each sample data set separately, along with regression of mitochondrial RNA as a variable. Single-cell data sets were combined by first developing an initial reference data set of 18 gastric samples with high cell numbers encompassing different characteristics of the primary and peritoneal metastasis samples for identifying anchors between pairs of samples and subsequently integrating the remaining samples as "query" data sets (18). Subsequently, PrepSCTIntegration was run to select features for downstream integration and FindIntegrationAnchors to identify anchor genes. The integrated data were scaled, and principal component analysis was performed. Data were visualized using Uniform Manifold Approximation and Projection (UMAP). Cell clusters were identified by a shared nearest-neighbor (SNN) modularity optimization-based clustering algorithm set at a resolution of 1. To identify differentially expressed genes for cluster demarcation, the FindAllMarkers module was used, and genes expressed in more than 25% of the cells in each cluster were selected. Differentially expressed genes from each cluster were compared with sets of previously described cell-type markers to assign cell identities, lineages, and sublineages. The AddmoduleScore function was used

to calculate composite module scores for different gene-expression programs in each single cell. These scores were calculated by averaging the expression levels of all the genes in each program at the single-cell level, subtracted by the aggregated expression of randomly selected control feature sets. Supplementary Table S2 lists the genes used for the various gene-expression programs/modules. For organoid analysis, the reference 18 scRNA sample data set (above) was integrated with the remaining primary, peritoneal, and organoid data sets, resolving cell clusters by SNN clustering at a resolution of 1.25. Organoid data were integrated with primary sample data sets using canonical correlation analysis (100). Pathway analysis was performed using the REACTOME database (<https://reactome.org/>) separately for differentially upregulated and downregulated genes (adjusted  $P < 0.05$ ) for each cluster. Pathways with entities  $P$  value and FDR  $< 0.05$  were investigated. Using these pathways, an overlap analysis was performed between epithelial metacluster and sublineages. For individual reclustering of metaclusters, we used LIGER (101), which relies on integrative nonnegative matrix (iNMF) factorization to identify shared and data set-specific factors. Briefly, we first selected variable genes, scaled them without centering followed by iNMF and quantile normalization of factors using the selectGenes, scaleNotCenter, optimizeALS ( $k = 30$ ), and quantileAlignSNF modules of LIGER. These iNMFs were added to Seurat for UMAP and clustering visualization.

### Doublet Analysis

scRNA-seq data are commonly affected by technical artifacts known as "doublets," which limit cell throughput and lead to spurious biological conclusions such as discovery of mixed lineages. DoubletFinder was used to identify doublets using gene-expression data (29). DoubletFinder predicts doublets according to each real cell's proximity in gene-expression space to artificial doublets created by averaging the transcriptional profile of randomly chosen cell pairs. After identifying doublet cells, these were removed from the data set.

### Molecular Subtyping

TCGA molecular subtyping was performed as previously described (102). Briefly, gastric cancers were first classified as Epstein-Barr virus (EBV) or microsatellite instability (MSI) using EBV-encoded RNA *in situ* hybridization (EBER-ISH) and mismatch-repair (MMR) protein IHC.

**MMR IHC.** Tissues were cut into 4- $\mu$ m-thick sections. The MMR panel consisted of monoclonal antibodies—mouse anti-MLH1 (Ventana, M1), mouse anti-PMS2 (Ventana, A16-4), mouse anti-MSH2 (Ventana G219-1129), and rabbit anti-MSH6 (Ventana, SP93). The panel was performed using the Ventana Optiview DAB Detection Kit, the Ventana Optiview Amplification Kit, and ancillaries on the fully automated Roche Ventana Ultra instrument. MMR loss was determined when the tumor showed loss of expression for the examined MMR proteins. Normal tissue adjacent to the tumor was used as a positive internal control.

**EBER-ISH.** Tissues were cut on 4- $\mu$ m-thick sections. The EBER probe is a fluorescein-conjugated oligonucleotide ready-to-use ISH probe from Leica Bond, and with hybridization for 2 hours. Detection was performed with anti-FITC antibody, Leica Bond DAB polymer kit, and ancillaries on the fully automated Leica Bond III Biosystem. EBER positivity was determined by localization of the EBER signal within the nucleus of the tumor cells.

WES of the remaining samples was performed as previously described (103). Briefly, 100 ng of DNA from each sample was sheared, adaptor ligated, exome captured using the Agilent Sureselect Human all exon kit v6, and sequenced using Illumina HiSeq4000 platform. Tumors were then classified as either chromosomal instability (CIN)



or GS based on the copy-number profiles inferred from WES using GISTIC2 (104) or scRNA-seq. Tumors were clustered with a uniformly processed in-house data set of >200 gastric cancers, using Euclidean distance and Ward method, based on thresholded copy numbers from significantly altered sites identified by GISTIC2 ( $q < 0.05$ ). Gastric cancer samples exhibiting elevated copy-number changes were assigned as CIN. We confirmed significant associations between GS tumors with diffuse histologic subtype ( $P = 0.03$ ; Chi-square test), and increased T-cell infiltrates in MSI-positive tumors ( $P = 0.021$ ). Samples with whole-genome doubling predicted by WES (thus classified as CIN) also exhibited high CNVs measured by scRNA-seq.

### Trajectory Analysis

Trajectory analysis was performed with Monocle 3.0 using two approaches (105, 106). For plasma cell pseudotemporal analysis, plasma cells were extracted from the data set. Monocle was used to preprocess, align, perform UMAP directionality reduction and cell clustering, and to develop learned graphs and cell orders. To compare trajectories between normal and tumor cells, the Seurat wrapper for Monocle 3.0 was used to cluster cells and develop learned graphs. To assess the robustness of inferred trajectories, we performed bootstrapping by randomly downsampling samples by 10% and computing binomial  $P$  values for each major branching node.

### CNV Analysis

To identify large-scale chromosomal CNVs, InferCNV (<https://github.com/broadinstitute/inferCNV>) was used to explore tumor scRNA-seq compared with normal scRNA-seq data sets, and between diffuse-type and intestinal-type scRNA-seq data sets. Expression intensities in epithelial cells of genes across positions of the genome were compared with a set of reference T and B cells. Genes with a mean count number of 0.1 or less across cells were excluded. In addition, the hidden Markov method (HMM; “type i6”) was used, applying noise filters and “subcluster” analysis. For each sample, the gene expression of cells was restandardized and values were limited from  $-1$  to  $1$ . As the HMM model uses values that are not centered on one but six states, to make gains and losses symmetrical, we merged all “state 6” states with “state 5” before restandardization. The final CNV score of each cell was calculated as the sum of the absolute values of each CNV region. We also used CONICSmatrix for CNV calling (107). Briefly, after filtering out uninformative genes a normalization factor was calculated for each cell and the average expression in each cell was centered using the calculated normalization factor. The  $z$ -score of the centered gene expression across all cells was calculated. Based on these  $z$ -scores, a Gaussian mixture model was calculated to determine regions showing an average gene-expression bimodal distribution across cells. Only results for regions harboring more than 100 expressed genes were calculated to ensure the predictions were not influenced by a few differentially expressed genes in a small region. To avoid batch effects, we called CNV for each sample separately on the gene-expression matrix on all cells.

### TCR Analysis

Cellranger v6.0 was used (using mkfastq module) to align FASTQ sequencing reads to the hg38 reference transcriptome. To generate single-cell V(D)J sequences and annotation for each sample, we used cellranger v2 module. Basic statistics and diversity index were estimated using immunarch (v0.6.6; <https://immunarch.com/index.html>).

### DSP and Analysis

Gastric cancer and normal FFPE tissues were mounted on super frost slides to validate the spatial profiling of RNA by the NanoString GeoMx DSP platform. For H&E staining, FFPE slides were deparaffinized with histoclear, rehydrated and stained with Hematoxylin Solution. Slides were counterstained with eosin and mounted with

mounting media. FFPE slides were subjected to conventional tissue preprocessing (deparaffinization and rehydration). For RNA profiling, the transcriptomics cancer atlas with a 1,800-plex RNA probe set was used along with an additional set of 12 genes (Supplementary Table S16). Standard fluorescence-labeled morphology marker panel consisting of Pan-CK for epithelial regions, CD138 for plasma cells,  $\alpha$ -smooth muscle actin for fibroblast and nuclear stain were used as ROI selection references. Twelve ROIs (four for each region) measuring 300  $\mu\text{m}$  in diameter on each slide were drawn and selected. All oligos from the selected ROI were collected into 96-well plates using the proprietary UV-guided technology in the DSP approach. Resultant oligos representing individual targets for individual ROIs were sequenced using Illumina sequencing. Data were analyzed by uploading the counts data set from the Illumina run into the GeoMx DSP analysis suite. Biological probe QC was performed using default settings. Scaling was performed using geometric means and normalization using Q3 averages of housekeeping genes.

### Humanized Mouse Model

Details of the humanized mouse model experiment have been described previously (108). One- to three-day-old NSG pups were sublethally irradiated at 1 Gy and engrafted with  $1 \times 10^5$  human CD34<sup>+</sup> cord blood cells (HLA-A24:02; STEMCELL Technologies) via intrahepatic injection. Mice with more than 10% human immune-cell reconstitution (calculated based on the proportion of human CD45 relative to the sum of human and mouse CD45) were included in the study. In total, two diffuse-type gastric cancer cell lines of HLA-A24:02 subtype were selected for the experiment (*KLF2* positive, GSU; *KLF2* negative, SNU1750; refs. 109, 110). For each cell line, five humanized mice and five NSG mice were injected with the tumor cells and observed for one month. Mice were sacrificed at the end of one month, necropsies were performed, and tumors were harvested and sequenced (bulk RNA-seq).

### IHC for IRF4

IHC analysis was performed to evaluate the expression of IRF4 in diffuse and intestinal gastric cancer samples using MUM1 (IRF4) primary antibodies (Clone MUM1p, DAKO, IS64430). IHC stains of 4- $\mu\text{m}$  paraffin sections of patient samples were performed using a Bond Max automated immunostainer (Leica Biosystem) using antibodies for IRF4 (MUM1p, 1/500; Dako). Briefly, sections were deparaffinized and rehydrated followed by antigen retrieval at pH 9.0 (Tris buffer) for 36 minutes. Sections were treated with peroxidase solution followed by incubation with MUM1 primary antibody for 60 minutes according to the manufacturer's protocol. Tissue sections were further incubated with horseradish peroxidase-labeled polymer secondary antibody, and localization was performed with horseradish peroxidase-labeled polymer with DAB using a bond polymer refine detection kit (Leica Bondmax, Leica Biosystems) according to standard protocols. Appropriate positive controls were immunostained in each batch of IHC. Gastric epithelial cells served as an internal negative control. MUM1 expression was predominantly observed in the nuclei of plasma cells along with some weak expression in cytoplasmic regions. Strong nuclear expression was considered as positive staining, whereas standalone cytoplasmic staining was considered negative. For each case, 10 fields (at 20 $\times$  magnification) were selected randomly, and MUM1-positive cells were counted in each field using the ImageJ software. A mean value of positive MUM1 expression was calculated for each case.

### Fibroblast Cell Culture Experiments

Human gastric fibroblast cell lines were derived from surgical gastric tissue of patients with gastric cancer who underwent gastrectomy without preoperative treatment with written informed consent from each patient. CAFs were established from the tumoral gastric wall,

and NFs were from nontumoral gastric wall according to previously established protocols (111, 112). Fibroblast cell lines were cultured in RPMI 1640 containing 10% FBS (normal medium) and incubated at 37°C with 5% CO<sub>2</sub>. *INHBA* expression was downregulated by transfecting cells with predesigned Silencer Select siRNAs directed against *INHBA* (#1, s7434; #2, s7436; catalog no.4427037; Thermo Fisher Scientific) and a nontargeting siRNA (#4390843; Thermo Fisher Scientific) was used as the negative control. CAFs were transfected with annealed siRNAs for 48 hours (5 μmol/L) using Lipofectamine RNAiMAX (#13778-150, Thermo Fisher Scientific). Real-time PCR primers used were:

*INHBA*: F 5' AGCTCAGACAGCTCTTACCACA 3'; R 5' TTTTCC TTCTCCTCTTCAGCA 3'  
*FAP*: F 5' TGGCGATGAACAATATCCTAGA 3'; R 5' ATCCGAAC AACGGGATTCTT 3'

NFs were seeded and incubated at 37°C with 5% CO<sub>2</sub>. After 24-hour seeding, the cultured cells were treated with 100 ng/mL recombinant human Activin A (BioLegend) for 48 and 96 hours, respectively. Collagen genes were measured in recombinant *INHBA*-treated and control fibroblast lines using the following primers:

*COL1A1*: F 5' AGACGAAGACATCCACCA 3'; R 5' GTCATCG CACAACACCTTGC 3'  
*COL1A2*: F 5' CTGGAGAGGCTGGTACTGCT 3'; R 5' GAGCA CCAAGAAGACCCTGA 3'  
*COL6A3*: F 5' ACCGTCCAACAGGTCATCTC 3'; R 5' CTCTTGCC ACCAACACCTGG 3'

All experiments were performed in triplicate. Data are presented as mean ± standard deviation (SD). Statistical significance was determined using a two-tailed Student *t* test.

### KLF2 Knockdown Model

LMSU cells were transfected with pooled siRNAs for *KLF2* [Dharmacon, L-006928-00-0020 ON-TARGETplus Human *KLF2* (10365) siRNA-SMARTpool] or negative control siRNAs (Dharmacon, D-001810-10-20 ON-TARGETplus Nontargeting Control Pool) at 100 nmol/L using Lipofectamine 2000 reagent (Thermo Fischer Scientific, 11668027). To obtain stable knockdown of *KLF2* in GSU cells, cells were plated at a concentration of 500,000 cells in a 10-cm dish and infected with MISSION lentiviral transduction particle encoding PLKO.1-PURO-CMV-tGFP-KLF2 (clone ID TRCN000418423) or GFP (as nontargeting control). After 48 hours, transfected cells were selected with 2 μg/mL puromycin. The transfected cells were then expanded and selected in culture medium plus puromycin (1 μg/mL) for 3 weeks to obtain stable *KLF2*-knockdown cells. Total RNA was extracted using the Qiagen RNAeasy mini kit according to the manufacturer's instructions. RNA was converted to cDNA using Improm-II Reverse Transcriptase (Promega) or iScript CDNA synthesis kit (Bio-Rad). Quantitative PCR was performed in triplicate using Quantifast SYBR Green PCR kit (Qiagen) on an Applied Biosystems HT7900 Real-Time PCR System for 40 cycles using the following primers with an annealing temperature of 55°C:

*KLF2*: F 5'CGGCAAGACCTACACCAAGAGT 3' R 5'CGCACA GATGGCACTGGAATG 3'  
*ACTB*: F 5' CATGTACGTTGCTATCCAGGC 3'; R 5' CTCCTTAA TGTCACGCACGAT 3'

Fold change was calculated using the Delta-Delta Ct method. Total intracellular protein was extracted with RIPA lysis buffer (Thermo Fisher). Protein concentration was determined using a bicinchoninic acid (BCA) assay kit (Thermo Fisher).

### Western Blotting

Equal amounts of protein from each sample were separated on 4% to 12% sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE)

and electrotransferred onto polyvinylidene difluoride (PVDF) membranes. Membranes were blocked with 5% fat-free dry milk in 0.1 M Tris-buffered saline-0.1% Tween-20 buffer (TBST) for 1 hour. The membranes were sequentially incubated with a *KLF2* primary antibody (Affinity Bio; dilution of 1 in 1,000) or GAPDH primary antibody (Proteintech; dilution of 1 in 3,000) overnight and horseradish peroxidase-conjugated secondary antibody (anti-rabbit and anti-mouse, Santa Cruz Biotechnology) for 1 hour. Blots were developed with an enhanced chemiluminescence reagent (Amersham Biosciences) and quantified by densitometric scanning and analyses using a ChemiDoc system (Bio-Rad).

### Transwell Migration Assay

Transwell migration assays were performed with human plasmablasts/plasma cells isolated from primary human PBMCs or KMS-11 cells, a human multiple myeloma cell line. Briefly, human plasmablasts/plasma cells from PBMCs were isolated using the Plasma Cell Isolation Kit II according to the manufacturer's instructions (cat # 130-093-628; Miltenyi Biotec). The purity of isolated plasmablasts/plasma cells was assessed by flow cytometry using CD19 and CD38 antibodies (BD Biosciences), whereby peripheral blood plasmablasts/plasma cells were defined as CD19<sup>+</sup> CD38<sup>+</sup> cells. We loaded 2 × 10<sup>4</sup> cells per well into the top chamber, and for KMS-11 cells, 2.5 × 10<sup>5</sup> cells were resuspended in 200 μL of the migration buffer and loaded onto the upper chamber of transwell inserts (Transwell Permeable Support with a 5.0-μm polycarbonate membrane, 6.5-mm insert; 3421, Corning). Gastric cancer cells were resuspended in 600 μL of migration buffer (0.5% BSA-RPMI 1640) and seeded into the bottom chamber of 24-well transwell plate (catalog no. 3421; Corning) and allowed to settle for 4 to 6 hours. Recombinant human CXCL12/SDF1a (350-NS-010-CF; R&D Systems) resuspended in 600 μL migration buffer (final concentration of 200 ng/mL) or migration buffer alone were used as positive and negative controls, respectively. Twenty to 24 hours later, unmigrated cells in the top chamber were counted. Unmigrated cells were subtracted from total cells seeded, and the proportion of migrated cells was calculated. Pair-wise comparisons were performed using the Mann-Whitney test.

### RNAScope

RNAScope ISH in gastric cancer samples was performed according to the manufacturer's protocols (113). Briefly, FFPE slides were baked at 60°C for 1 hour before being deparaffinized in xylene and 100% ethanol. After drying slides for 5 minutes at room temperature, H<sub>2</sub>O<sub>2</sub> was added for 10 minutes at room temperature. For antigen accessibility, slides were incubated in boiling antigen retrieval solution (<98°C) for 15 minutes, washed in water twice, dehydrated in 100% ethanol, and finally treated with Protease Plus for 30 minutes at 40°C. Probes were then hybridized for 2 hours at 40°C followed by RNAScope amplification and chromogenic detection. RNA scope 2.5 HD Duplex detection kits were used for simultaneous visualization of two RNA targets in single-cell resolution using Hs-PLVAP HRP-GREEN (437461) and Hs-RGS5-C2 AP-FAST RED (533421; ACD Bio). Sections were counterstained with hematoxylin and mounted with Vectamount. Duplex probes targeting CI-PPIB and C2 -POLR2A (322435) were used as positive control probes and dihydrodipicolinate reductase (dapB), a bacterial gene (310043) as a negative control probe.

### Analysis of Bulk RNA-seq Data

To assess cellular abundances in bulk tissue transcriptome profiles, CIBERSORTx was used to estimate cellular abundances of B cells and plasma cells in intestinal and diffuse TCGA-STAD data sets downloaded from FireBrowse (9, 114). For reference signature gene matrices, non-small cell lung cancer (NSCLC) PBMC scRNA-seq data sets provided with the CIBERSORTx suite were used. To compare single-cell RNA data with gastric bulk RNA data, combined TCGA and GTEx

gene-expression data sets were used from the Xena browser (115). For survival analysis of *INHBA*, pooled survival analysis of several available gastric cancer microarray data sets was performed using the Kaplan–Meier plotter tool, gastric cancer subgroup (<https://kmpplot.com/analysis/index.php?p=service&cancer=gastric>). Both *INHBA* probes (Affy ID: 204926\_at and 210511\_s\_at) were pooled, and the mean expression of both probes was used to generate *INHBA* high and low subgroups by dividing samples at the median. For the survival analysis of *INHBA* in TCGA, samples were divided in *INHBA* high and low at the median value, and adjustment of stage was performed using Cox regression.

### Analysis of *KLF2* Cis-Regulation

H3K27ac ChIP-seq and RNA-seq of 24 primary gastric cancer samples were performed and analyzed as previously described (116). To explore the *KLF2* locus in greater detail, we first identified *KLF2* promoter regions using an in-house gastric cancer promoter catalog derived from the H3K27ac ChIP-seq of 26 in-house gastric cancer cell lines. For each primary gastric cancer sample, the input subtracted H3K27ac signal at the *KLF2* promoter region was then computed using bigWigAverageOverBed to yield reads per kilobase per million.

### Generation, Maintenance, and Single-Cell Sequencing of Gastric Cancer PDOs

Human gastric tissues were biopsied from tumor and matched adjacent normal sites of each patient during surgical intervention. Tissues were processed as previously described (117). Briefly, tissues were minced and washed in Dulbecco's phosphate-buffered saline (DPBS; Thermo Fisher Scientific) and digested in DPBS containing 1 mg/mL collagenase (Sigma-Aldrich) and 2 mg/mL BSA (Sigma-Aldrich) for 30 minutes at 37°C. Digested tissues were passed through 30- $\mu$ m filters (Miltenyi Biotec). Filtered cells were pelleted at 300  $\times$  g for 5 minutes, resuspended in Matrigel (Corning Life Sciences), and seeded into multiwell plates (Thermo Fisher Scientific). Cultures were maintained in custom gastric PDO culture medium at 37°C in 5% CO<sub>2</sub> and monitored daily for organoid generation. The culture medium in each well was replaced with fresh medium on alternate days. PDOs were passaged once every 7 to 10 days at a 1:3 ratio. The median establishment time to the respective passage at time of sequencing was 17 weeks (range: 17–30 weeks; passage numbers: 9–11). Gastric organoids were harvested from gel matrices by washing briefly with DPBS, incubating with Trypsin-EDTA at 37°C for up to 30 minutes, and pelleted at 300  $\times$  g for 5 minutes. Supernatants were discarded, and cell pellets were washed twice with 10 mL DPBS each and filtered through cell strainers (mesh size: 30  $\mu$ m). After centrifugation at 300  $\times$  g for 5 minutes, the supernatant was discarded, and cells were washed with 1  $\times$  DPBS and then resuspended at ~1,000 cells/ $\mu$ L in 1  $\times$  PBS containing 0.4% BSA. Organoid scRNA-seq libraries were prepared using the 10 $\times$  Genomics Single-Cell 3' Gel Bead and Library Kit.

### Statistical Analysis

All analyses were done using R (V.4.0.3) with statistical significance set at  $P < 0.05$  adjusted for multiple testing. Wilcoxon rank sum test was used to evaluate associations with continuous variables. Student  $t$  test was used to evaluate associations with parametric continuous variables. Bivariate correlation analysis was performed using Pearson or Kendall's Tau (clinical stage correlation). Significance of overlapping CNVs called by WES and scRNA-seq was assessed using hypergeometric distribution tests by phyper (lower.tail = FALSE) in R (V.4.0.3). Kaplan–Meier curves with log-rank statistics were used to compare overall survival. Pearson correlation analysis for TCGA bulk RNA-seq was performed using cBioPortal. We computed JSI for the top 30 significant Reactome programs (58) between PDOs and primary tumors by their metaclusters.

### Data Availability

scRNA-seq data have been uploaded to the Gene Expression Omnibus repository: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183904>.

### Authors' Disclosures

R. Sundar reports grants from National Medical Research Council (NMRC) during the conduct of the study, as well as other support from Bristol Myers Squibb, Merck, Eisai, Bayer, Taiho, Novartis, Eli Lilly, Roche, AstraZeneca, and DKSH, grants and other support from MSD, and grants from Paxman Coolers outside the submitted work. P. Tan reports other support from Tempus Healthcare outside the submitted work. No disclosures were reported by the other authors.

### Authors' Contributions

**V. Kumar:** Conceptualization, data curation, software, formal analysis, investigation, visualization, methodology, writing—original draft. **K. Ramnarayanan:** Conceptualization, data curation, formal analysis, validation, investigation, visualization, methodology, writing—original draft, and project administration. **R. Sundar:** Formal analysis, funding acquisition, investigation, visualization, writing—original draft. **N. Padmanabhan:** Formal analysis, investigation, visualization, writing—original draft. **S. Srivastava:** Formal analysis and investigation. **M. Koiwa:** Formal analysis and investigation. **T. Yasuda:** Formal analysis and investigation. **V. Koh:** Data curation. **K. Huang:** Formal analysis and investigation. **S. Tay:** Data curation. **S.W. Ho:** Data curation. **A. Tan:** Project administration. **T. Ishimoto:** Formal analysis and investigation. **G. Kim:** Data curation. **A. Shabbir:** Data curation. **Q. Chen:** Resources and data curation. **B. Zhang:** Data curation, formal analysis, and investigation. **S. Xu:** Conceptualization, resources, data curation, and funding acquisition. **K. Lam:** Conceptualization, resources, data curation, and funding acquisition. **H. Lum:** Data curation. **M. Teh:** Data curation, formal analysis, and investigation. **W. Yong:** Conceptualization, resources, data curation, and funding acquisition. **J.B.Y. So:** Conceptualization, resources, data curation, and funding acquisition. **P. Tan:** Conceptualization, data curation, supervision, project administration, writing—review and editing.

### Acknowledgments

We thank NanoString Technologies USA for Digital Spatial Transcriptomics Platform expertise, Dr. Anita Chan from Singapore Eye Research Institute, Dr. Anjali Verma, Dr. Amrita Roy for assistance with RNAScope, and the Duke-NUS Genome Biology Facility for assistance with sequencing. Certain figures were created using BioRender.com, with relevant licenses obtained. R. Sundar is supported by the NMRC, Singapore (NMRC/Fellowship/0059/2018 and MOH-000627). P. Tan is supported by Duke-NUS Medical School and the Biomedical Research Council, Agency for Science, Technology and Research. This work was also supported by NMRC grants NR13NMR111OM, NMRC/STaR/0026/2015, and MOH-OFLCG18May-0003. Funding was also provided by Cancer Science Institute of Singapore, NUS, under the National Research Foundation Singapore and the Singapore Ministry of Education Research Centres of Excellence initiative. The funders of the study had no role in study design, data collection, analysis, interpretation, or writing of the manuscript. P. Tan had full access to all data in the study and as corresponding author had final responsibility for the decision to submit for publication.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 28, 2021; revised September 27, 2021; accepted October 7, 2021; published first October 12, 2021.

## REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- The global, regional, and national burden of stomach cancer in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease study 2017. *Lancet Gastroenterol Hepatol* 2020;5:42–54.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513:202–9.
- Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* 2015;21:449–56.
- Sundar R, Liu DH, Hutchins GG, Slaney HL, Silva AN, Oosting J, et al. Spatial profiling of gastric cancer patient-matched primary and locoregional metastases reveals principles of tumour dissemination. *Gut* 2021;70:1823–32.
- Tan IB, Ivanova T, Lim KH, Ong CW, Deng N, Lee J, et al. Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology* 2011;141:476–85, 85.e1–11.
- Lei Z, Tan IB, Das K, Deng N, Zouridis H, Pattison S, et al. Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology* 2013;145:554–65.
- Maman S, Witz IP. A history of exploring cancer in context. *Nat Rev Cancer* 2018;18:359–76.
- Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37:773–82.
- Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018;50:96.
- Wang R, Dang M, Harada K, Han G, Wang F, Pool Pizzi M, et al. Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. *Nat Med* 2021;27:141–51.
- Zhang M, Hu S, Min M, Ni Y, Lu Z, Sun X, et al. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* 2021;70:464–75.
- Sathe A, Grimes SM, Lau BT, Chen J, Suarez C, Huang RJ, et al. Single-cell genomic characterization reveals the cellular reprogramming of the gastric tumor microenvironment. *Clin Cancer Res* 2020;26:2640–53.
- Zhang P, Yang M, Zhang Y, Xiao S, Lai X, Tan A, et al. Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer. *Cell Rep* 2019;27:1934–47.
- Bassiouni R, Gibbs LD, Craig DW, Carpten JD, McEachron TA. Applicability of spatial transcriptional profiling to cancer research. *Mol Cell* 2021;81:1631–9.
- Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;353:78–82.
- Goh JLL, Chou N, Seow WY, Ha N, Cheng CPP, Chang YC, et al. Highly specific multiplexed RNA imaging in tissues with split-FISH. *Nat Methods* 2020;17:689–93.
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. *Cell* 2019;177:1888–902.
- Fujii Y, Yoshihashi K, Suzuki H, Tsutsumi S, Muroh H, Maeda S, et al. CDX1 confers intestinal phenotype on gastric epithelial cells via induction of stemness-associated reprogramming factors SALL4 and KLF5. *Proc Natl Acad Sci U S A* 2012;109:20584–9.
- Thiriou A, Perdomo C, Cheng G, Novitzky-Basso I, McArdle S, Kishimoto JK, et al. Differential DARC/ACKR1 expression distinguishes venular from non-venular endothelial cells in murine tissues. *BMC Biol* 2017;15:45.
- Radbruch A, Muehlinghaus G, Luger EO, Inamine A, Smith KG, Dörner T, et al. Competence and competition: the challenge of becoming a long-lived plasma cell. *Nat Rev Immunol* 2006;6:741–50.
- Maestre L, Tooze R, Cañamero M, Montes-Moreno S, Ramos R, Doody G, et al. Expression pattern of XBP1(S) in human B-cell lymphomas. *Haematologica* 2009;94:419–22.
- McCarron MJ, Park PW, Fooksman DR. CD138 mediates selection of mature plasma cells by regulating their survival. *Blood* 2017;129:2749–59.
- Rodríguez-Bayona B, Ramos-Amaya A, López-Blanco R, Campos-Caro A, Brieva JA. STAT-3 activation by differential cytokines is critical for human in vivo-generated plasma cell survival and Ig secretion. *J Immunol* 2013;191:4996–5004.
- Cortés M, Georgopoulos K. Aiolos is required for the generation of high affinity bone marrow plasma cells responsible for long-term immunity. *J Exp Med* 2004;199:209–19.
- Gutzeit C, Magri G, Cerutti A. Intestinal IgA production and its role in host-microbe interaction. *Immunol Rev* 2014;260:76–85.
- Puig-Kröger A, Sierra-Filardi E, Domínguez-Soto A, Samaniego R, Corcuera MT, Gómez-Aguado F, et al. Folate receptor beta is expressed by tumor-associated macrophages and constitutes a marker for M2 anti-inflammatory/regulatory macrophages. *Cancer Res* 2009;69:9395–403.
- Mossel DM, Moganti K, Riabov V, Weiss C, Kopf S, Cordero J, et al. Epigenetic regulation of S100A9 and S100A12 expression in monocyte-macrophage system in hyperglycemic conditions. *Front Immunol* 2020;11:1071.
- McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* 2019;8:329–37.
- Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 2017;171:1611–24.
- Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* 2016;539:309–13.
- Petti AA, Williams SR, Miller CA, Fiddes IT, Srivatsan SN, Chen DY, et al. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nat Commun* 2019;10:3660.
- Guo YA, Chang MM, Huang W, Ooi WF, Xing M, Tan P, et al. Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nat Commun* 2018;9:1520.
- Imielinski M, Guo G, Meyerson M. Insertions and deletions target lineage-defining genes in human cancers. *Cell* 2017;168:460–72.
- Sundar R, Smyth EC, Peng S, Yeong JPS, Tan P. Predictive biomarkers of immune checkpoint inhibition in gastroesophageal cancers. *Front Oncol* 2020;10:763.
- Kono K, Kawaida H, Takahashi A, Sugai H, Mimura K, Miyagawa N, et al. CD4(+)/CD25<sup>high</sup> regulatory T cells increase with tumor stage in patients with gastric and esophageal cancers. *Cancer Immunol Immunother* 2006;55:1064–71.
- Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 2017;169:1342–56.
- Zhang L, Zhang Z. Recharacterizing tumor-infiltrating lymphocytes by single-cell RNA sequencing. *Cancer Immunol Res* 2019;7:1040–6.
- Jiménez Fonseca P, Carmona-Bayonas A, Hernández R, Custodio A, Cano JM, Lacalle A, et al. Lauren subtypes of advanced gastric cancer influence survival and response to chemotherapy: real-world data from the AGAMENON National Cancer Registry. *Br J Cancer* 2017;117:775–82.
- Ohguchi H, Hideshima T, Bhasin MK, Gorgun GT, Santo L, Cea M, et al. The KDM3A–KLF2–IRF4 axis maintains myeloma cell survival. *Nat Commun* 2016;7:10258.
- Winkelmann R, Sandrock L, Porstner M, Roth E, Mathews M, Hobeika E, et al. B cell homeostasis and plasma cell homing controlled by Krüppel-like factor 2. *Proc Natl Acad Sci U S A* 2011;108:710–5.
- Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. *Ann Oncol* 2008;19:1523–9.
- Kojima K, Kishimoto T, Nagai Y, Tanizawa T, Nakatani Y, Miyazaki M, et al. The expression of hepatocyte nuclear factor-4alpha, a developmental regulator of visceral endoderm, correlates with the intestinal phenotype of gastric adenocarcinomas. *Pathology* 2006;38:548–54.

44. Ham IH, Lee D, Hur H. Role of cancer-associated fibroblast in gastric cancer progression and resistance to treatments. *J Oncol* 2019;2019:6270784.
45. Liu T, Han C, Wang S, Fang P, Ma Z, Xu L, et al. Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. *J Hematol Oncol* 2019;12:86.
46. Ishimoto T, Miyake K, Nandi T, Yashiro M, Onishi N, Huang KK, et al. Activation of transforming growth factor beta 1 signaling in gastric cancer-associated fibroblasts increases their motility, via expression of rhomboid 5 homolog 2, and ability to induce invasiveness of gastric cancer cells. *Gastroenterology* 2017;153:191–204.
47. Calon A, Tauriello DV, Batlle E. TGF- $\beta$  in CAF-mediated tumor growth and metastasis. *Semin Cancer Biol* 2014;25:15–22.
48. Verrecchia F, Chu ML, Mauviel A. Identification of novel TGF- $\beta$ /Smad gene targets in dermal fibroblasts using a combined cDNA microarray/promoter transactivation approach. *J Biol Chem* 2001;276:17058–62.
49. Castro NE, Kato M, Park JT, Natarajan R. Transforming growth factor  $\beta$ 1 (TGF- $\beta$ 1) enhances expression of profibrotic genes through a novel signaling cascade and microRNAs in renal mesangial cells. *J Biol Chem* 2014;289:29001–13.
50. Hopfer U, Hopfer H, Meyer-Schwesinger C, Loeffler I, Fukai N, Olsen BR, et al. Lack of type VIII collagen in mice ameliorates diabetic nephropathy. *Diabetes* 2009;58:1672–81.
51. Sumi E, Jehara N, Akiyama H, Matsubara T, Mima A, Kanamori H, et al. SRY-related HMG box 9 regulates the expression of Col4a2 through transactivating its enhancer element in mesangial cells. *Am J Pathol* 2007;170:1854–64.
52. Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. *Hepatology* 2008;47:2059–67.
53. Nagaraja AS, Dood RL, Armaiz-Pena G, Kang Y, Wu SY, Allen JK, et al. Adrenergic-mediated increases in INHBA drive CAF phenotype and collagens. *JCI insight* 2017;2:e93076.
54. Nagy Á, Munkácsy G, Györfy B. Pancancer survival analysis of cancer hallmark genes. *Sci Rep* 2021;11:6047.
55. Vlachogiannis G, Hedayat S, Vatsiou A, Jamin Y, Fernández-Mateos J, Khan K, et al. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* 2018;359:920–6.
56. Seidlitz T, Merker SR, Rothe A, Zakrzewski F, von Neubeck C, Grützmann K, et al. Human gastric cancer modelling using organoids. *Gut* 2019;68:207–17.
57. Yan HHN, Siu HC, Law S, Ho SL, Yue SSK, Tsui WY, et al. A comprehensive human gastric cancer organoid biobank captures tumor subtype heterogeneity and enables therapeutic screening. *Cell Stem Cell* 2018;23:882–97.
58. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res* 2020;48:D498–d503.
59. Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* 2020;182:497–514.
60. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* 2020;182:1232–51.
61. Barlow KD, Sanders AM, Soker S, Ergun S, Metheny-Barlow LJ. Pericytes on the tumor vasculature: jekyll or hyde? *Cancer Microenviron* 2013;6:1–17.
62. Cui X, Lu YW, Lee V, Kim D, Dorsey T, Wang Q, et al. Venous endothelial marker COUP-TFII regulates the distinct pathologic potentials of adult arteries and veins. *Sci Rep* 2015;5:16193.
63. Patel V, Faure S, Corre I, Clere N. Endothelial-to-mesenchymal transition (EndoMT): roles in tumorigenesis, metastatic extravasation and therapy resistance. *J Oncol* 2019;2019:8361945.
64. Choi KJ, Nam J-K, Kim J-H, Choi S-H, Lee Y-J. Endothelial-to-mesenchymal transition in anticancer therapy and normal tissue damage. *Exp Mol Med* 2020;52:781–92.
65. van Meeteren LA, ten Dijke P. Regulation of endothelial cell plasticity by TGF- $\beta$ . *Cell Tissue Res* 2012;347:177–86.
66. Sahai E, Astsaturro I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer* 2020;20:174–86.
67. Zhi K, Shen X, Zhang H, Bi J. Cancer-associated fibroblasts are positively correlated with metastatic potential of human gastric cancers. *J Exp Clin Cancer Res* 2010;29:66.
68. Chen H, Yang WW, Wen QT, Xu L, Chen M. TGF- $\beta$  induces fibroblast activation protein expression; fibroblast activation protein expression increases the proliferation, adhesion, and migration of HO-8910PM [corrected]. *Exp Mol Pathol* 2009;87:189–94.
69. Grunberg N, Pevsner-Fischer M, Goshen-Lago T, Diment J, Stein Y, Lavon H, et al. Cancer-associated fibroblasts promote aggressive gastric cancer phenotypes via heat shock factor 1-mediated secretion of extracellular vesicles. *Cancer Res* 2021;81:1639–53.
70. Katayama Y, Oshima T, Sakamaki K, Aoyama T, Sato T, Masudo K, et al. Clinical significance of INHBA gene expression in patients with gastric cancer who receive curative resection followed by adjuvant S-1 chemotherapy. *In Vivo* 2017;31:565–71.
71. Oshima T, Yoshihara K, Aoyama T, Hasegawa S, Sato T, Yamamoto N, et al. Relation of INHBA gene expression to outcomes in gastric cancer after curative surgery. *Anticancer Res* 2014;34:2303–9.
72. Wang Q, Wen Y-G, Li D-P, Xia J, Zhou C-Z, Yan D-W, et al. Upregulated INHBA expression is associated with poor survival in gastric cancer. *Med Oncol* 2012;29:77–83.
73. Namwanje M, Brown CW. Activins and inhibins: roles in development, physiology, and disease. *Cold Spring Harb Perspect Biol* 2016;8:a021881.
74. Valer JA, Sánchez-de-Diego C, Pimenta-Lopes C, Rosa JL, Ventura F. ACVR1 function in health and disease. *Cells* 2019;8:1366.
75. Ohga E, Matsuse T, Teramoto S, Ouchi Y. Activin receptors are expressed on human lung fibroblast and activin A facilitates fibroblast-mediated collagen gel contraction. *Life Sci* 2000;66:1603–13.
76. Costea DE, Hills A, Osman AH, Thurlow J, Kalna G, Huang X, et al. Identification of two distinct carcinoma-associated fibroblast subtypes with differential tumor-promoting abilities in oral squamous cell carcinoma. *Cancer Res* 2013;73:3888–901.
77. Cangkruma M, Wietecha M, Mathis N, Okumura R, Ferrarese L, Al-Nuaimi D, et al. A paracrine activin A-mDia2 axis promotes squamous carcinogenesis via fibroblast reprogramming. *EMBO Mol Med* 2020;12:e11466.
78. Li X, Yang Z, Xu S, Wang Z, Jin P, Yang X, et al. Targeting INHBA in ovarian cancer cells suppresses cancer xenograft growth by attenuating stromal fibroblast activation. *Dis Markers* 2019;2019:7275289.
79. Ries A, Schelch K, Falch D, Pany L, Hoda MA, Grusch M. Activin A: an emerging target for improving cancer treatment? *Expert Opin Ther Targets* 2020;24:985–96.
80. Tan P, Yeoh KG. Genetics and molecular pathogenesis of gastric adenocarcinoma. *Gastroenterology* 2015;149:1153–62.
81. Sakimura C, Tanaka H, Okuno T, Hiramatsu S, Muguruma K, Hirakawa K, et al. B cells in tertiary lymphoid structures are associated with favorable prognosis in gastric cancer. *J Surg Res* 2017;215:74–82.
82. Meier A, Nekolla K, Hewitt LC, Earle S, Yoshikawa T, Oshima T, et al. Hypothesis-free deep survival learning applied to the tumour microenvironment in gastric cancer. *J Pathol Clin Res* 2020;6:273–82.
83. Derks S, de Klerk LK, Xu X, Fleitas T, Liu KX, Liu Y, et al. Characterizing diversity in the tumor-immune microenvironment of distinct subclasses of gastroesophageal adenocarcinomas. *Ann Oncol* 2020;31:1011–20.
84. Katoh H, Komura D, Konishi H, Suzuki R, Yamamoto A, Kakiuchi M, et al. Immunogenetic profiling for gastric cancers identifies sulfated glycosaminoglycans as major and functional B cell antigens in human malignancies. *Cell Rep* 2017;20:1073–87.
85. Sharonov GV, Serebrovskaya EO, Yuzhakova DV, Britanova OV, Chudakov DM. B cells, plasma cells and antibody repertoires in the tumour microenvironment. *Nat Rev Immunol* 2020;20:294–307.
86. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513:202–9.

87. Figueroa-Protti L, Soto-Molinari R, Calderón-Osorno M, Mora J, Alpízar-Alpízar W. Gastric cancer in the era of immune checkpoint blockade. *J Oncol* 2019;2019:1079710.
88. Pernot S, Terme M, Radosevic-Robin N, Castan F, Badoual C, Marcheteau E, et al. Infiltrating and peripheral immune cell analysis in advanced gastric cancer according to the Lauren classification and its prognostic significance. *Gastric Cancer* 2020;23:73–81.
89. Das H, Kumar A, Lin Z, Patino WD, Hwang PM, Feinberg MW, et al. Kruppel-like factor 2 (KLF2) regulates proinflammatory activation of monocytes. *Proc Natl Acad Sci U S A* 2006;103:6653–8.
90. Jha P, Das H. KLF2 in regulation of NF- $\kappa$ B-mediated immune cell function and inflammation. *Int J Mol Sci* 2017;18:2383.
91. SenBanerjee S, Lin Z, Atkins GB, Greif DM, Rao RM, Kumar A, et al. KLF2 Is a novel transcriptional regulator of endothelial proinflammatory activation. *J Exp Med* 2004;199:1305–15.
92. Sindi HA, Russomanno G, Satta S, Abdul-Salam VB, Jo KB, Qazi-Chaudhry B, et al. Therapeutic potential of KLF2-induced exosomal microRNAs in pulmonary hypertension. *Nat Commun* 2020;11:1185.
93. Marrone G, Russo L, Rosado E, Hide D, García-Cardena G, García-Pagán JC, et al. The transcription factor KLF2 mediates hepatic endothelial protection and paracrine endothelial-stellate cell deactivation induced by statins. *J Hepatol* 2013;58:98–103.
94. Weinreich MA, Takada K, Skon C, Reiner SL, Jameson SC, Hogquist KA. KLF2 transcription-factor deficiency in T cells results in unrestrained cytokine production and upregulation of bystander chemokine receptors. *Immunity* 2009;31:122–30.
95. Sebzda E, Zou Z, Lee JS, Wang T, Kahn ML. Transcription factor KLF2 regulates the migration of naive T cells by restricting chemokine receptor expression patterns. *Nat Immunol* 2008;9:292–300.
96. Choi SY, Gout PW, Collins CC, Wang Y. Epithelial immune cell-like transition (EIT): a proposed transdifferentiation process underlying immune-suppressive activity of epithelial cancers. *Differentiation* 2012;83:293–8.
97. Gangoso E, Southgate B, Bradley L, Rus S, Galvez-Cancino F, McGivern N, et al. Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoeediting to elicit immune evasion. *Cell* 2021;184:2454–70.
98. Driehuis E, Kretzschmar K, Clevers H. Establishment of patient-derived cancer organoids for drug-screening applications. *Nat Protoc* 2020;15:3380–409.
99. Lo Y-H, Kolahi KS, Du Y, Chang C-Y, Krokhotin A, Nair A, et al. A CRISPR/Cas9-engineered ARID1A-deficient human gastric cancer organoid model reveals essential and nonessential modes of oncogenic transformation. *Cancer Discov* 2021;11:1562–81.
100. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411–20.
101. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 2019;177:1873–87.
102. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JJ, Kim K, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med* 2018;24:1449–58.
103. Chen R, Im H, Snyder M. Whole-exome enrichment with the agilent SureSelect human all exon platform. *Cold Spring Harb Protoc* 2015;2015:626–33.
104. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;12:R41.
105. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381–6.
106. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;14:979–82.
107. Müller S, Cho A, Liu SJ, Lim DA, Diaz A. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. *Bioinformatics* 2018;34:3217–9.
108. Sundar R, Huang KK, Kumar V, Ramnarayanan K, Demircioglu D, Her Z, et al. Epigenetic promoter alterations in GI tumour immunoeediting and resistance to immune checkpoint inhibition. *Gut* 2021;gutjnl-2021-324420.
109. Ku JL, Kim KH, Choi JS, Kim SH, Shin YK, Chang HJ, et al. Establishment and characterization of six human gastric carcinoma cell lines, including one naturally infected with Epstein-Barr virus. *Cell Oncol* 2012;35:127–36.
110. Miyamoto S, Nagamura Y, Nakabo A, Okabe A, Yanagihara K, Fukami K, et al. Aberrant alternative splicing of RHOA is associated with loss of its expression and activity in diffuse-type gastric carcinoma cells. *Biochem Biophys Res Commun* 2018;495:1942–7.
111. Uchihara T, Miyake K, Yonemura A, Komohara Y, Itoyama R, Koiwa M, et al. Extracellular vesicles from cancer-associated fibroblasts containing annexin A6 induces FAK-YAP activation by stabilizing  $\beta$ 1 integrin, enhancing drug resistance. *Cancer Res* 2020;80:3222–35.
112. Yasuda T, Koiwa M, Yonemura A, Miyake K, Kariya R, Kubota S, et al. Inflammation-driven senescence-associated secretory phenotype in cancer-associated fibroblasts enhances peritoneal dissemination. *Cell Rep* 2021;34:108779.
113. Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* 2012;14:22–9.
114. Deng M, Bragelmann J, Kryukov I, Saraiva-Agostinho N, Perner S. FirebrowseR: an R client to the Broad Institute's Firehose Pipeline. *Database* 2017;2017:baw160.
115. Goldman MJ, Craft B, Hasie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 2020;38:675–8.
116. Ooi WF, Xing M, Xu C, Yao X, Ramlee MK, Lim MC, et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat Commun* 2016;7:12983.
117. Steele NG, Chakrabarti J, Wang J, Biesiada J, Holokai L, Chang J, et al. An organoid-based preclinical model of human gastric cancer. *Cell Mol Gastroenterol Hepatol* 2019;7:161–84.