CANCER

Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing

Xudong Xing[1,2]*, Fan Yang[3]*, Qi Huang[3,4†], Haifa Guo[3], Jiawei Li[3], Mantang Qiu[3‡], Fan Bai[5,6‡], Jun Wang[3‡]

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Lung adenocarcinomas (LUAD) that radiologically display as subsolid nodules (SSNs) exhibit more indolent biological behavior than solid LUAD. The transcriptomic features and tumor microenvironment (TME) of SSN remain poorly understood. Here, we performed single-cell RNA sequencing analyses of 16 SSN samples, 6 adjacent normal lung tissues (nLung), and 9 primary LUAD with lymph node metastasis (mLUAD). Approximately 0.6 billion unique transcripts were obtained from 118,293 cells. We found that cytotoxic natural killer/T cells were dominant in the TME of SSN, and malignant cells in SSN undergo a strong metabolic reprogram and immune stress. In SSN, the subtype composition of endothelial cells was similar to that in mLUAD, while the subtype distribution of fibroblasts was more like that in nLung. Our study provides single-cell transcriptomic profiling of SSN and their TME. This resource provides deeper insight into the indolent nature of SSN and will be helpful in advancing lung cancer immunotherapy.

INTRODUCTION

The application of low-dose computed tomography (LDCT) screening has substantially increased the detection rate of early-stage lung adenocarcinoma (LUAD) that manifests as radiological subsolid nodules (SSNs) (*1*). In a screening study in Shanghai, 84.87% lung cancer cases detected at baseline LDCT screening were SSNs (*2*). It has been well recognized that LUAD radiologically manifested as SSN is a static state of relatively indolent tumor, which has good survival (*3*–*7*). In a recent large-scale retrospective study of patients with stage I LUAD after surgery, the 5-year overall survival was 94.9% for part-solid nodules (*3*). Because of the indolent growth pattern and good prognosis, a consensus has been reached that tumors with subsolid features should be treated less "aggressively," and longer-term follow-up is recommended (*8*–*10*).

Much effort has been devoted to decoding the indolent nature of SSN. Our team has found that the genome of SSN has a relatively low mutation burden and few copy number alterations (*11*). The tumor lesion is a complex ecosystem composed of malignant cells, various types of immune cells, and stromal cells (*12*). The heterogeneity of tumor cells and different types of tumor microenvironment (TME) plays a vital role in shaping tumor behavior (*13*–*17*). Therefore, it is critically important to decode the complex interplay between tumor cells and the TME in SSN.

In this study, we conducted single-cell RNA sequencing (scRNA-seq) of 16 LUAD samples manifested as SSN. By comparing SSN with nine samples of primary LUAD with lymph node metastasis (mLUAD) and six samples of adjacent normal lung tissues (nLung), we comprehensively characterized the transcriptome features of malignant cells, immune cells, and stromal cells of SSN, and we decoded dynamic changes in cell percentage, the heterogeneity of cell subtypes, and intercellular interactions, providing new knowledge regarding the biological basis of SSN and LUAD development.

RESULTS

Single-cell transcriptomic profiling of the multicellular ecosystem of SSN

Droplet-based scRNA-seq (10X Genomics) was performed on a total of 16 SSN samples from 16 treatment naïve patients (fig. S1A and table S1). In parallel, the scRNA-seq data of nine mLUAD and six nLung (*17*) samples were downloaded for integrative analyses (Fig. 1A). Approximately ~0.6 billion unique transcripts were obtained from 118,293 cells. Among these cells, 70,461 cells (59.56%) were from SSN, 26,344 cells (22.27%) were from mLUAD, and 21,488 cells (18.17%) were from nLung. All high-quality cells were integrated into an unbatched and comparable dataset and subjected to principal components analysis (PCA) after correction for read depth and mitochondrial read counts (fig. S1, B to D). By graph-based uniform manifold approximation and projection (UMAP), 26 high-confidence cell clusters (fig. S1B) were identified, which could be assigned to known cell lineages (fig. S1E). We identified 10 major cell types (Fig. 1B) according to the expression of canonical gene markers (Fig. 1C and fig. S1F): T cells, natural killer (NK) cells, myeloid cells, B cells, plasma cells, mast cells, fibroblasts, endothelial cells (ECs), EPCAM$^+$ epithelial cells, and erythroblasts. The relative abundance of T cells increased stepwise from nLung to SSN and mLUAD (Fig. 1, D and E; fig. S1, G and H; and table S3). The relative abundance of NK cells in SSN was comparable to that in nLung but higher than that in mLUAD. In addition, mast cells were significantly enriched in SSN. In both SSN and mLUAD, the relative abundance of mononuclear phagocytes decreased in comparison with nLung, while that of B and plasma cells was increased. Seven-plex immunohistochemistry

[1]School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]Peking University–Tsinghua University–National Institute of Biological Sciences Joint Graduate Program, Tsinghua University, Beijing 100084, China. [3]Department of Thoracic Surgery, Peking University People's Hospital, Beijing 100044, China. [4]Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450003, China. [5]Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China. [6]Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing 100871, China.
*These authors contributed equally to this work.
†Present address: Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450003, China.
‡Corresponding author. Email: jwangmd@pku.edu.cn (J.W.); fbai@pku.edu.cn (F.B.); qiumantang@163.com (M.Q.)

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

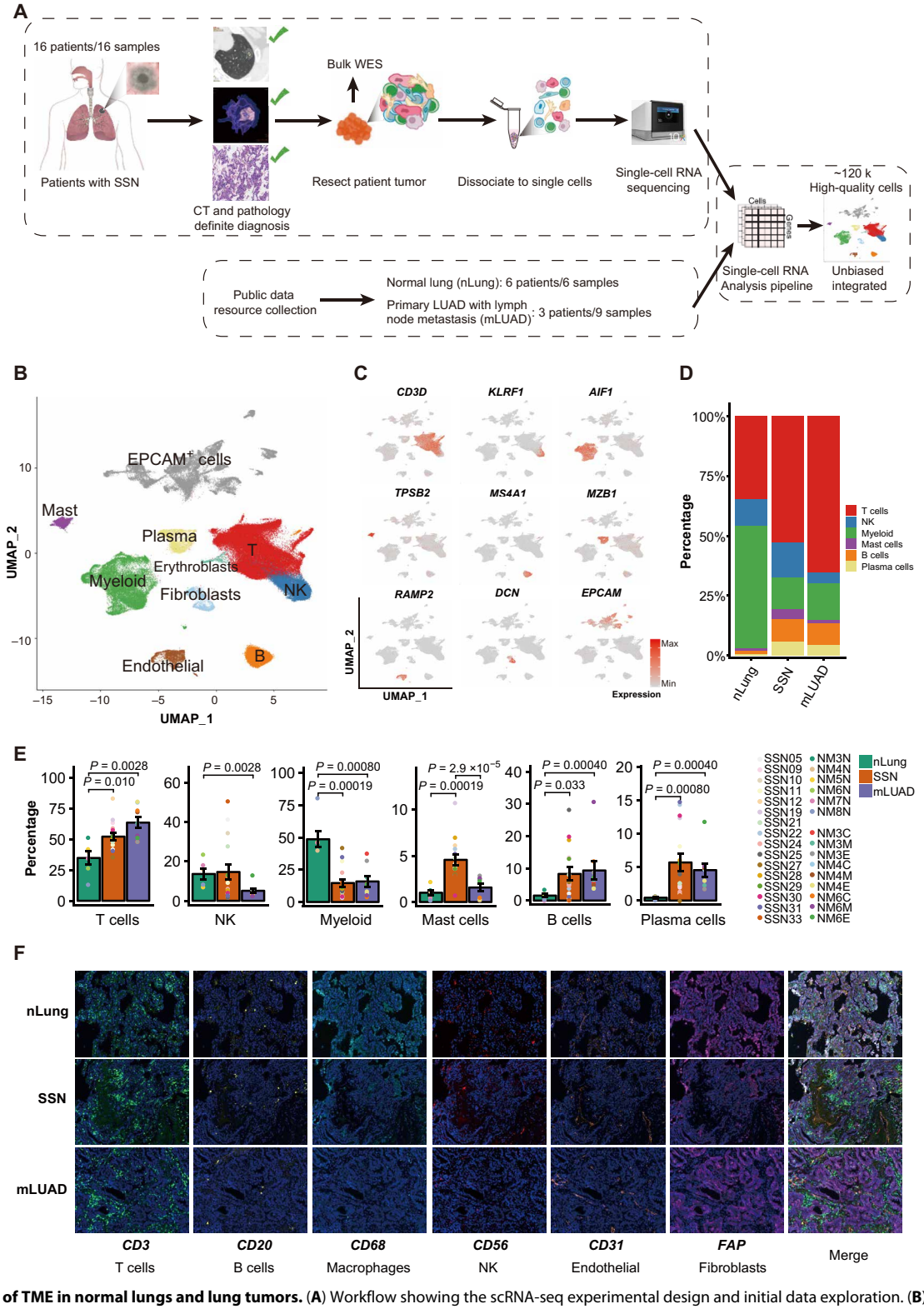

Fig. 1. Overview of TME in normal lungs and lung tumors. (**A**) Workflow showing the scRNA-seq experimental design and initial data exploration. (**B**) Cellular populations identified. The UMAP projection of 118,293 single cells from nLung ($n = 6$), SSN ($n = 16$), and mLUAD ($n = 9$) samples shows the formation of 10 main clusters with label names. Each dot corresponds to a single cell, colored according to cell type. (**C**) Canonical cell markers were used to label clusters by cell identity as represented in the UMAP plot. (**D**) Average proportion of six main types of CD45$^+$ immune cells among nLung, SSN, and mLUAD samples. (**E**) Percentages of the six types CD45$^+$ immune cells among three groups. *Y* axis: Average percent of samples across the three groups. Groups are shown in different colors. Each bar plot represents one cell cluster. Error bars represent ± SEM for normal and tumor samples. Colored dots represent different samples. All differences with $P < 0.05$ are indicated; two-sided unpaired Wilcoxon rank sum test was used for analysis. (**F**) Seven-plex staining panel showing the cellular components of nLung, SSN, and mLUAD tissues.

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

staining was further conducted to provide an overview of the multicellular ecosystems of nLung, SSN, and mLUAD (Fig. 1F and fig. S1I). These results suggest that SSN represent a multicellular ecosystem distinct from those of nLung and mLUAD.

Hallmark signatures and metabolism disturbance in malignant cells of SSN

Next, we focused on the transcriptomic features of each major cell type. A total of 1997 normal epithelial cells were obtained from nLung samples and further clustered as alveolar type I cell (AT1; *AGER*$^+$), alveolar type II cell (AT2; *SFTPA1*$^+$), secretory club cell (Club; *SCGB1A1*$^+$), basal airway epithelial cells (Basal; *KRT17*$^+$), and ciliated airway epithelial cells (Ciliated; *TPPP3*$^+$) (Fig. 2A) based on canonical markers (Fig. 2B, fig. S2A, and table S2) as described previously (*18*), independent of patient origin (Fig. 2C) and other features (fig. S2B).

Malignant cells were identified by inferring large-scale copy number variations (CNVs) with immune and stromal cells as references (*13*–*15*). The CNV patterns inferred in malignant cells were consistent with the CNVs calculated from paired bulk whole-exome sequencing (WES) data (Fig. 2D and fig. S2C). We found that malignant cells formed clusters according to patient origin, indicating a high degree of intertumor heterogeneity (Fig. 2E and fig. S2D) (*13*–*15*).

Carcinogenesis has been described as the acquisition of advantageous biological capabilities by malignant cells (*19*, *20*). Gene set variation analyses (GSVA) (*17*, *21*) comparing mLUAD with SSN revealed that E2F targets, MYC targets, the interferon-α (IFN-α) and IFN-γ response, and PI3K-AKT and hypoxia pathways were up-regulated in malignant cells in mLUAD (Fig. 2F, top, and fig. S2E) (*19*, *20*, *22*). Next, a comparison of SSN with nLung revealed that many metabolic pathways were significantly up-regulated in malignant cells of SSN—including glycolysis; oxidative phosphorylation; and fatty acid, xenobiotic, and HEME metabolism (Fig. 2F, bottom). Further comprehensive dissection of metabolic profiles (*23*) suggested that malignant cells in SSN had distinct metabolic patterns (Fig. 2G and fig. S2F). Specifically, some metabolic pathways were highly expressed in SSN malignant cells, including oxidative phosphorylation; arginine and proline metabolism; histidine metabolism; and metabolism of alanine, aspartate, and glutamate.

Last, we explored differences in the expression programs in malignant cells from SSN and mLUAD samples using non-negative matrix factorization (*15*). Hierarchical clustering identified five common expression programs that varied within mLUAD (Fig. 2H, top), including cell cycle, mitochondrial signaling, epithelial-mesenchymal transition (EMT), epithelial differentiation, and hypoxia. In SSN, the identification of an expression program including *JUN*, *FOS*, *IER2*, and immediate early genes was indicative of cellular activation and stress responses (Fig. 2H, bottom). In summary, at single-cell resolution, we found that malignant cells in SSN showed a strong metabolic reprogram and immune stress.

Cytotoxic dominant T and NK cells in the TME of SSN

Subclustering 57,301 T and NK cells revealed 12 subtypes (Fig. 3, A to D, and fig. S3, A to E): 5 subtypes of CD4$^+$ T cells (CD4-C1 to C5; *CD3D*$^+$*CD4*$^+$) including 1 regulatory CD4$^+$ T cell subtype (CD4-C4; *FOXP3*$^+$), 5 subtypes of CD8$^+$ T cells (CD8-C1 to C5; *CD3D*$^+$*CD8A*$^+$), and 2 subtypes of NK cells (NK-C1 and C2, *CD3D*$^-$*CD56*$^+$*TYROBP*$^+$). Specifically, 35,185 T/NK cells were obtained from SSN (Fig. 3B).

For CD4$^+$ T cells, we identified memory (CD4-C1; *ANXA1*$^+$), effector memory (CD4-C2; *ANXA1*$^+$*GZMA*$^+$), naïve (CD4-C3; *CCR7*$^+$*SELL*$^+$*LEF1*$^+$), regulator (CD4-C4; *FOXP3*$^+$*IL2RA*$^+$), and exhausted (CD4-C5; *CXCL13*$^+$*PDCD1*$^+$*BTLA*$^+$*TOX*$^+$) CD4$^+$ T cells (Fig. 3, C and D, and table S2) (*24*, *25*). Furthermore, we found that naïve CD4$^+$ T cell cluster (CD4-C3) also encompassed with little proportion of naïve CD8$^+$ T cells (fig. S3F). The inferred developmental trajectory of conventional CD4$^+$ T cells exhibited a branched structure, with the highest naïve state CD4-C3 (Fig. 3E) as the root and with effector CD4-C2 and exhausted CD4-C5 as the ending clusters (fig. S3G). In SSN, the relative percentage of effector CD4$^+$ T (CD4-C2) was reduced in comparison with that of nLung, but higher than that of mLUAD, while the relative percentages of suppressive regulator (CD4-C4) and exhausted CD4$^+$ T (CD4-C5) cells in SSN were comparable to those in nLung but significantly less than those in mLUAD (Fig. 3, F and G, fig. S3H, and table S3).

For CD8$^+$ T cells, CD8-C1 with high expression of *GZMK* but low expression of other cytotoxic effectors represents pre-effector CD8$^+$ T cells (Fig. 3, C and D, and table S2). Meanwhile, CD8-C1 shows the low expression of *TCF7* (also *TCF-1*), *PDCD1*, and *EOMES* but lacks the expression of *GZMB* and *HAVCR2* (also *TIM-3*), which was consistent with the "precursor exhausted" T cells defined by the previous study (*26*). CD8-C2 was identified as memory CD8$^+$ T cells based on *ZNF683* expression. CD8-C4 corresponded to effector T cells due to high cytotoxic marker expression, such as *NKG7*, *PRF1*, and *CX3CR1*. CD8-C3 and CD8-C5 were assigned to terminal exhausted CD8$^+$ cytotoxic T cells, which was characterized by the expression of cytotoxic effectors (*IFNG*, *RPF1*, *NKG7*, *GZMA*, *GZMB*, and *GNLY*), inhibitory markers (*PDCD1*, *HAVCR2*, *LAG3*, *TIGIT*, and *CTLA4*), and T cell exhaustion–associated transcription factors *TOX* and *EOMES* (Fig. 3, C to E, and table S2). Compared with CD8-C3 cells, CD8-C5 cells showed higher expression levels of proliferative genes, such as *MKI67* (Fig. 3, C and D, and table S2). High expression levels of signature genes of CD8-C3 and CD8-C5 cells were both significantly associated with poor survival of patients with LUAD according to The Cancer Genome Atlas (TCGA) (Fig. 3H and fig. S3I). The developmental trajectory of CD8$^+$ T cells also suggested a binary branched structure (Fig. 3I), which was consistent with previous studies (*24*, *25*): naïve T cells as the root, CD8-C3/C5 as the end state of exhausted T cells, and CD8-C4 as the end state of cytotoxic T cells; CD8-C1 and CD8-C2 cells were located between these two end states. The subtype composition of CD8$^+$ T cells in SSN was similar to that in nLung but opposite to that observed in mLUAD (Fig. 3, F and G, fig. S3H, and table S3). In SSN, the percentage of the most cytotoxic CD8$^+$ T cells (CD8-C4) (Fig. 3E) was higher than that in mLUAD. However, the percentage of exhausted CD8$^+$ T (CD8-C3 and CD8-C5) cells in SSN was much less than that in mLUAD. In addition, the curve for the "Cytotoxic/Exhausted score" of SSN was located between those of nLung and mLUAD (Fig. 3J).

NK-C1 (*CD3*$^-$*CD56*^{dim}*CD16*$^+$) cells were characterized by *FCGR3A* (*CD16*) expression in comparison with NK-C2 (*CD3*$^-$*CD56*^{bright}*CD16*$^-$) cells and represented the most cytotoxic cluster (Fig. 3E and table S2). The percentage of NK cells in SSN was comparable with that in nLung (Fig. 3F), and NK-C1 cells were enriched in SSN compared with mLUAD (Fig. 3G, fig. S3H, and table S3). The multiplex immunohistochemistry staining also demonstrated enriched NK cells in SSN in comparison with mLUAD (fig. S3J).

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

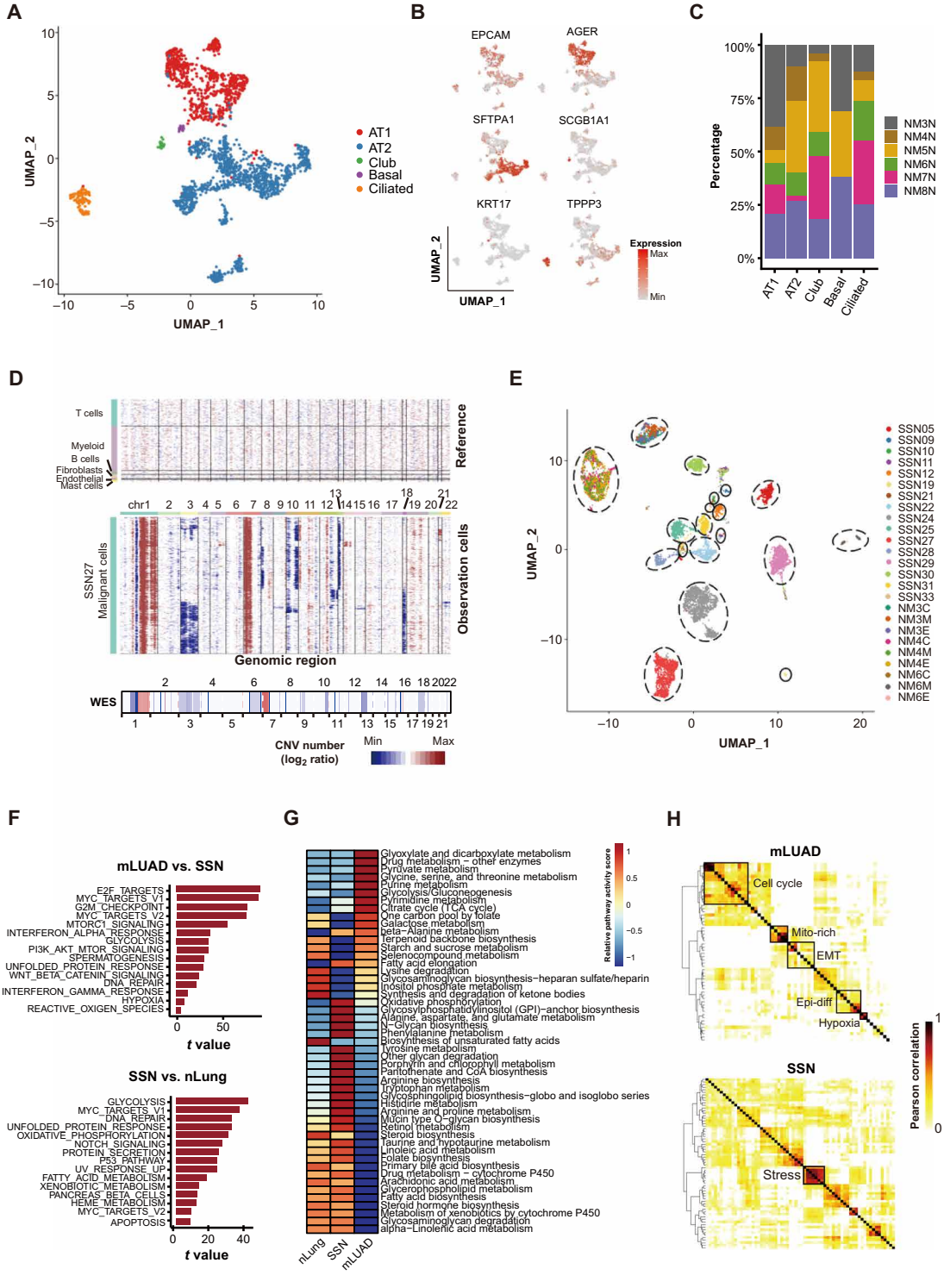

Fig. 2. Identification and characterization of malignant cells in SSN. (**A**) Clustering of 1997 epithelial cells from nLung ($n = 6$). Each dot corresponds to a single cell, colored according to cell type. (**B**) Canonical cell markers were used to label epithelial subtypes as represented in the UMAP plot. (**C**) Sample distribution in each cluster. Each bar corresponds to one cell type cluster, colored according to the samples. (**D**) Heatmap showing large-scale CNVs for individual cells (rows) from one SSN sample (SSN27) with WES paired data. Nonmalignant cells were treated as references (top), and large-scale CNVs were observed in malignant cells (middle). The CNVs of the sample were validated by WES analysis (bottom). The color shows the $\log_2$ CNV ratio. Red: amplifications; blue: deletions. (**E**) UMAP projection of 9281 malignant cells from SSN ($n = 16$) and mLUAD ($n = 9$). Each dot corresponds to a single cell, colored according to the samples. (**F**) Top 15 up-regulated hallmark pathways in malignant cells. Top: mLUAD versus SSN. Bottom: SSN versus nLung. (**G**) Heatmap showing differences in metabolic pathways scored per cell by GSVA between normal epithelial cells in nLung and malignant cells in SSN and mLUAD. (**H**) Heatmap depicting pairwise correlations of intratumoral programs derived from mLUAD (top) and SSN (bottom). Coherent expression programs are identified and labeled.

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

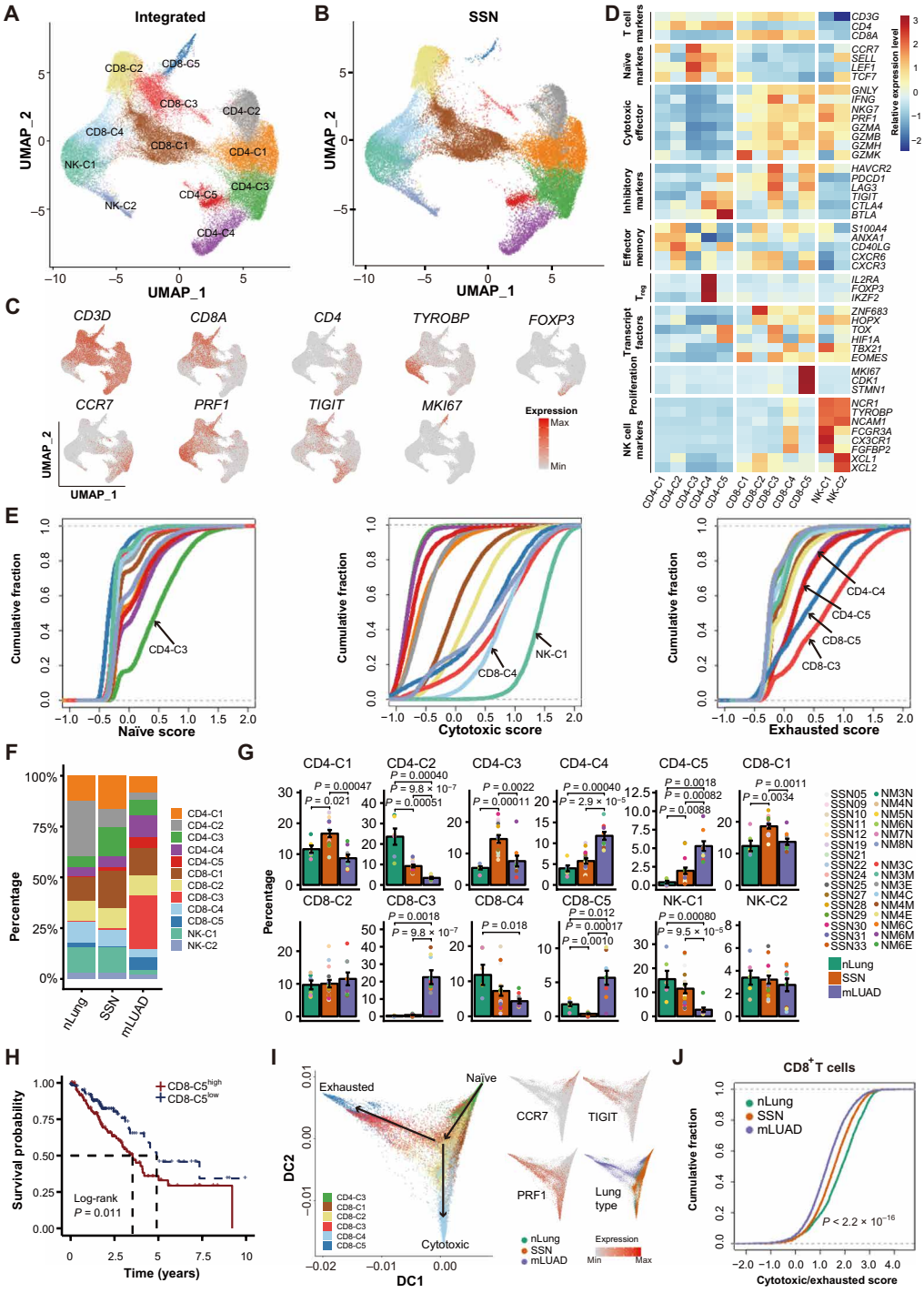

Fig. 3. Cytotoxic dominant T and NK cells in SSN. (**A**) UMAP projection of 57,301 T and NK cells, showing the composition of 12 main subtypes. (**B**) UMAP projection of 35,185 T and NK cells derived from SSN. (**C**) Canonical cell markers were used to identify T/NK cell subtypes. (**D**) Heatmap of functional gene sets in T and NK clusters. T_{reg}, regulatory T cell. (**E**) Cumulative distribution function showing the distribution of naïve (left), cytotoxic (middle), and exhausted (right) state scores in each T/NK subtype. A rightward shift of the curve indicates increased state scores. (**F**) Average proportion of each subtype between nLung, SSN, and mLUAD. (**G**) Percentages of each T/NK cell subtype among nLung, SSN, and mLUAD. *Y* axis: Average percent of samples across the three groups. Groups are shown in different colors. Each bar plot represents one cell cluster. Error bars represent ± SEM for normal and tumor samples. Colored dots represent different samples. All differences with $P < 0.05$ are indicated; two-sided unpaired Wilcoxon rank sum test was used for analysis. (**H**) Kaplan-Meier plot showing that patients with LUAD in the TCGA dataset with high expression of CD8-C5 cluster markers have shorter overall survival. The high and low groups are divided by the 75% quantile value of the mean expression of the above gene set. (**I**) Development trajectory of $CD8^+$ T cells inferred by diffusion map, colored by cell subtype and expression of example genes. (**J**) As in (E), but for "cytotoxic/exhausted score" defined as the average expression level of cytotoxic genes divided by the average expression level of exhausted genes to measure the functional state of $CD8^+$ T cells in nLung, SSN, and mLUAD. *P* value was calculated by two-sided unpaired Kruskal-Wallis rank sum test.

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

SSNs are enriched with DCs and lack a subtype of TAM

Subclustering of 18,380 myeloid cells revealed 17 subsets (Fig. 4, A to D; fig. S4, A to E; and table S2): 9 subtypes of macrophage (Macro-C1-C9; *APOE*$^+$), 2 subtypes of monocytes (Mono-C1/C2; *FCN1*$^+$), 1 subtype of monocyte-derived dendritic cells (Mono-DCs; *MRC1*$^+$ *CD14*$^+$), 3 subtypes of conventional DCs (cDCs) (DCs-C1; *CLEC9A*$^+$, DCs-C2; *CD207*$^+$ and DCs-C3; *LAMP3*$^+$), 1 subtype of pDCs (*LILRA4*$^+$), and 1 subtype of granulocytes (*G0S2*$^+$). In detail, for nine macrophage and two monocyte subsets, we identified alveolar resident macrophages (Macro-C1 to C3; *PPARG*$^+$), perivascular resident macrophages (Macro-C4; *LYVE1*$^+$), anti-inflammatory macrophages (Macro-C5; *CHI3L1*$^+$, Macro-C6; *TNF*$^+$*AXL*$^+$), tumor-associated macrophage (TAM) (Macro-C7; *VEGFA*$^+$), proliferating macrophage (Macro-C8; *PCNA*$^+$), early-stage macrophage (Macro-C9; *CXCL10*$^+$), classical monocytes (Mono-C1; *CD14*$^+$), and nonclassical monocytes (Mono-C2; *FCGR3A*$^+$), which were consistent with previous studies (*25*, *27*). Specifically, 6655 (36.21%) myeloid cells were obtained from SSN (Fig. 4B).

DCs are professional antigen-presenting cells that play a key role in CD8$^+$ T cell activation (*28*–*30*). All three subtypes of cDCs were significantly enriched in SSN (Fig. 4, E and F, fig. S4F, and table S3). DCs-C1 corresponded to cDC type 1 (cDC1; *CLEC9A*$^+$) and DCs-C2 to cDC type 2 (cDC2; *CD1C*$^+$*CD207*$^+$), while DCs-C3 represented migratory cDCs (*CCR7*$^+$*CCL19*$^+$) (Fig. 4G and table S2). The DCs-C3 subtype—which has also high expression levels of *BIRC3*, *CCL22*, *CD80*, *CD86*, *CD40*, and *FSCN1* (Fig. 4G and fig. S4G)—represents cDCs in an activated state (fig. S4H). In particular, DCs in SSN have a low expression level of indoleamine 2,3-dioxygenase 1 (*IDO1*), a gene that inhibits T cell activation and induces T cell differentiation into suppressive regulatory T cells (Fig. 4H) (*12*).

Macrophages are usually classified into the canonical proinflammatory M1 and anti-inflammatory M2 classes (*31*, *32*). We found that Macro-C4 and Macro-C6 exhibited an M2-dominant gene signature, but no macrophage subtype exhibited only an M1-like phenotype (fig. S4I) (*27*). Conventional M1 (*CD86* and *TLR2*) and M2 (*MSR1* and *MRC1*) genes were not subtype specific (Fig. 4I), suggesting that macrophage activation in the TME does not follow the polarization model (*16*). In particular, Macro-C7, which expressed TAM markers and genes related to glycolysis and angiogenesis (Fig. 4, I and J) (*33*), was absent in SSN but enriched in mLUAD (Fig. 4F, fig. S4F, and table S3). Signature genes from Macro-7 were significantly associated with poor overall survival in TCGA LUAD patients (Fig. 4K).

SSNs have a similar distribution of endothelial subtypes to that of mLUAD

Subclustering of 3381 ECs revealed six subtypes (Fig. 5, A and B; fig. S5, A to E; and table S2): extra-alveolar capillary ECs (cECs) (Endo-C1; *EDN1*$^+$*SLC6A4*$^+$, Endo-C5; *EDN1*$^+$*CCL2*$^+$), alveolar cECs (Endo-C2; *EDNRB*$^+$ *IL1RL1*$^+$), tumor ECs (Endo-C3; *IGFBP7*$^+$*PLVAP*$^+$), arterial ECs (Endo-C4; *GJA5*$^+$*FBLN5*$^+$), and lymphatic ECs (Endo-C6; *PDPN*$^+$ *CCL21*$^+$). Furthermore, tumor ECs (Endo-C3) also contained tip ECs (*ESM1* and *NID2*) and high endothelial venules (HEVs) and venous ECs (*ACKR1* and *SELP*), which was consistent with the previous observations that tip ECs and HEVs mostly resided in malignant tissue (*25*). Specifically, 911 (26.94%) ECs were obtained from SSN (fig. S5D). SSN and mLUAD had similar distribution of endothelial subtype, characterized by expansion of Endo-C3 and depletion of Endo-C5 (Fig. 5, C and D, fig. S5F, and table S3). A direct comparison of ECs between nLung and SSN/mLUAD revealed that Myc targets and inflammatory response were the most enriched gene sets in the ECs of SSN/mLUAD and nLung, respectively (fig. S5, G and H).

Endo-C5 cells highly expressed genes associated with immune activation and lymphocyte homing, such as *BIRC3*, *CCL2*, *CD44*, and *ICAM1*, whereas Endo-C3 cells highly expressed genes related to angiogenesis, like *HSPG2* and *POSTN* (fig. S5, E and I). Hallmark pathway analyses showed that Endo-C5 cells were enriched in immune activation–related pathways [inflammatory response, tumor necrosis factor–α (TNFA) signaling pathway, and IFN-γ response], while Endo-C3 cells were enriched in biosynthetic and metabolic pathways related to angiogenesis and extracellular remodeling (Fig. 5E). In contrast to mLUAD, Endo-C2 cells were abundant in SSN (Fig. 5, C and D). Endo-C2 cells had also high expression levels of immune activation–related genes, including *ICAM1/2*, *IL32*, and major histocompatibility complex (MHC) II molecules (fig. S5I).

Immunomodulatory fibroblasts are highly abundant in SSN

We found five distinct subtypes by reclustering 2257 fibroblasts (Fig. 5, F and G; fig. S6, A to E; and table S2): normal fibroblasts (Fibro-C1; *RGCC*$^+$*MACF1*$^+$), myofibroblasts (Fibro-C2; *ACTA2*$^+$ *PTN*$^+$), cancer-associated fibroblasts (CAFs) (Fibro-C3; *FAP*$^+$*TGFB1*$^+$), immune-modulatory fibroblasts (Fibro-C4; *CFD*$^+$*CXCL14*$^+$), and pericytes (Fibro-C5; *RGS5*$^+$*PDGFRB*$^+$). Specifically, 1196 (52.99%) fibroblasts were obtained from SSN (fig. S6D). The distribution of fibroblast subtypes in SSN was similar to that in nLung; both groups were characterized by abundant Fibro-C1/C2/C4 cells and depleted Fibro-C3/C5 cells in comparison with mLUAD (Fig. 5, H and I, fig. S6F, and table S3). A direct comparison of fibroblasts from mLUAD and SSN/nLung revealed that fibroblasts in SSN/nLung were enriched in immunomodulatory pathways, including TNFA signaling via NFκB and IL6-JAK-STAT3 (interleukin-6–Janus kinase–signal transducer and activator of transcription 3) signaling, whereas tumor support hallmarks were enriched in fibroblasts from mLUAD, including the oxidative phosphorylation, angiogenesis, EMT, and active transcript pathways (fig. S6, G and H).

Notably, Fibro-C4 cells, the most enriched subtype in SSN (Fig. 5I), expressed high levels of cytokines/chemokines, like *CXCL14* and *CXCL12*, indicating immunomodulatory features (Fig. 5J). Intriguingly, further analyses revealed that Fibro-C4 cells also expressed MHC II and *CD74*, which was consistent with the recently defined "antigen-presenting CAFs" (*34*). Fibro-C3 cells were specifically enriched in mLUAD and expressed genes indicative of CAFs, including *FAP*, *PDPN*, and *TGFB1* (Fig. 5J and fig. S6E). CAFs act as synthetic machines that produce various extracellular components that promote carcinogenesis (*12*, *35*). In addition to collagens broadly expressed in all clusters, Fibro-C3 cells uniquely expressed collagens V, VIII, and XII (fig. S6I), suggesting functional specialization of tumor-supported collagens. Hallmark pathway analysis further confirmed that the IFN-γ response and IL2-STAT5 signaling were highly activated in the Fibro-C4 cluster. However, Fibro-C3 cells were enriched in pathways that support tumor progression, including EMT, TGFB signaling, and angiogenesis (Fig. 5K).

B and plasma cells are strongly enriched in the lung TME

There is notable evidence supporting the critical role of B cells in antitumor immunology (*36*). A total of 10,903 B cells were analyzed, and seven subtypes were identified (fig. S7, A to D, and table S2): two subtypes of follicular B cells [follicular B-C1: memory cells (*CD20*$^+$

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

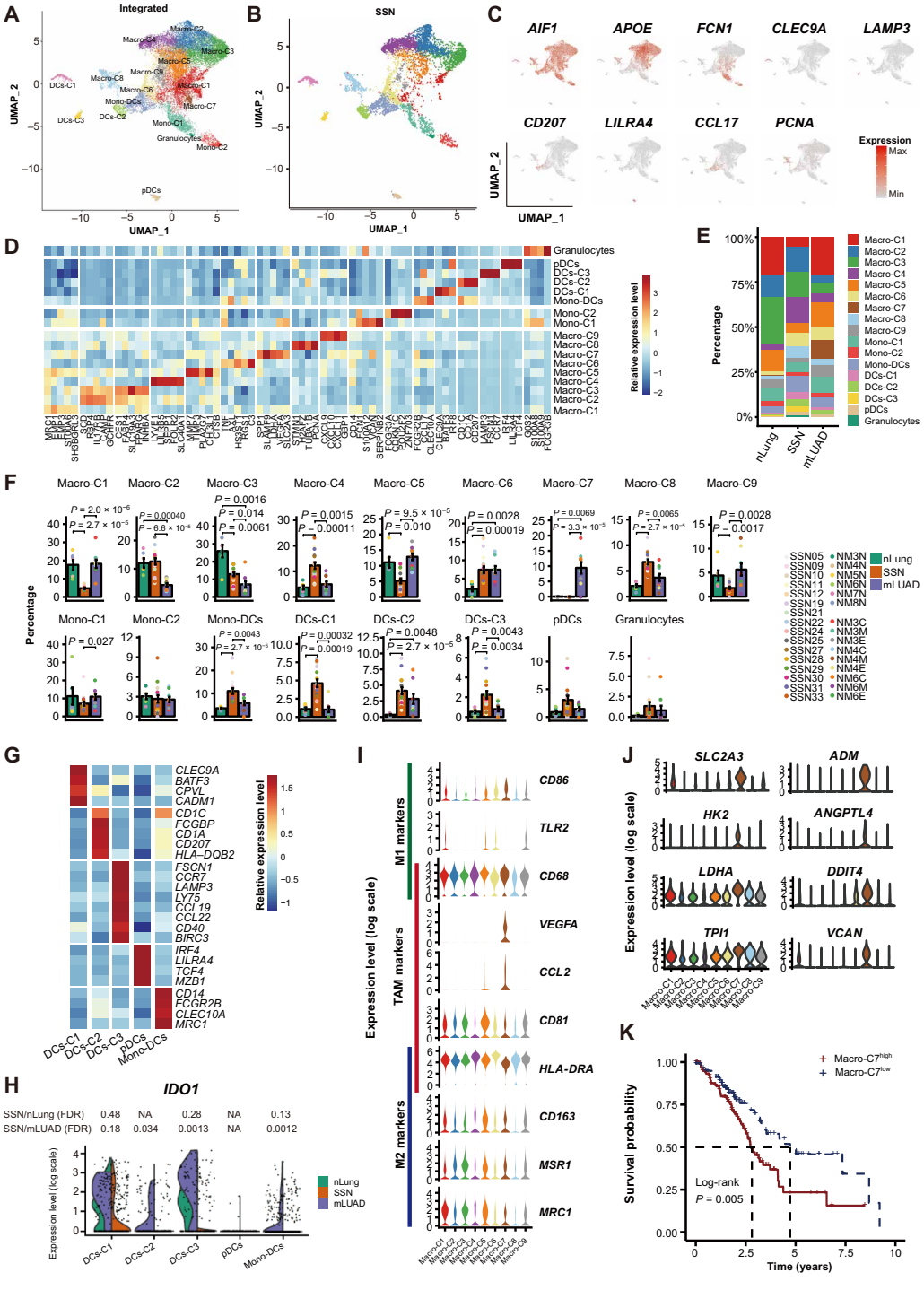

Fig. 4. Detailed characterization of myeloid cells. (**A**) UMAP projection of 18,380 myeloid cells, showing the composition of 17 main subtypes. (**B**) UMAP projection of 6655 myeloid cells derived from SSN. (**C**) Canonical cell markers were used to identity myeloid cell subtypes. (**D**) Heatmap of marker gene expression in myeloid clusters. (**E**) Average proportion of each myeloid subtype among nLung, SSN, and mLUAD. (**F**) Percentages of each myeloid cell subtype among nLung, SSN, and mLUAD. *Y* axis: Average percent of samples across the three groups. Groups are shown in different colors. Each bar plot represents one cell cluster. Error bars represent ± SEM for normal and tumor samples. Colored dots represent different samples. All differences with $P < 0.05$ are indicated; two-sided unpaired Wilcoxon rank sum test was used for analysis. (**G**) Heatmap showing the markers of different DC subtypes. (**H**) Violin plots showing the expression of *IDO1* in DC subtypes, split by sample origin. *P* values were calculated by differential expression test (DE test) using a pseudo-bulk method with Benjamini-Hochberg–corrected value. NA: *P* values cannot be calculated because of low expression. FDR, false discovery rate; NA, not applicable. (**I**) Violin plots showing the expression of M1, M2, and TAM markers in macrophage subtypes. (**J**) Violin plots showing the expression of example Macro-C7 markers involved in glycolysis and angiogenesis. (**K**) Kaplan-Meier plot showing that patients with LUAD in TCGA dataset with high expression of Macro-C7 cluster markers have shorter overall survival. The high and low groups were divided by the 75% quantile value of the mean expression level of the Macro-C7 gene set.

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

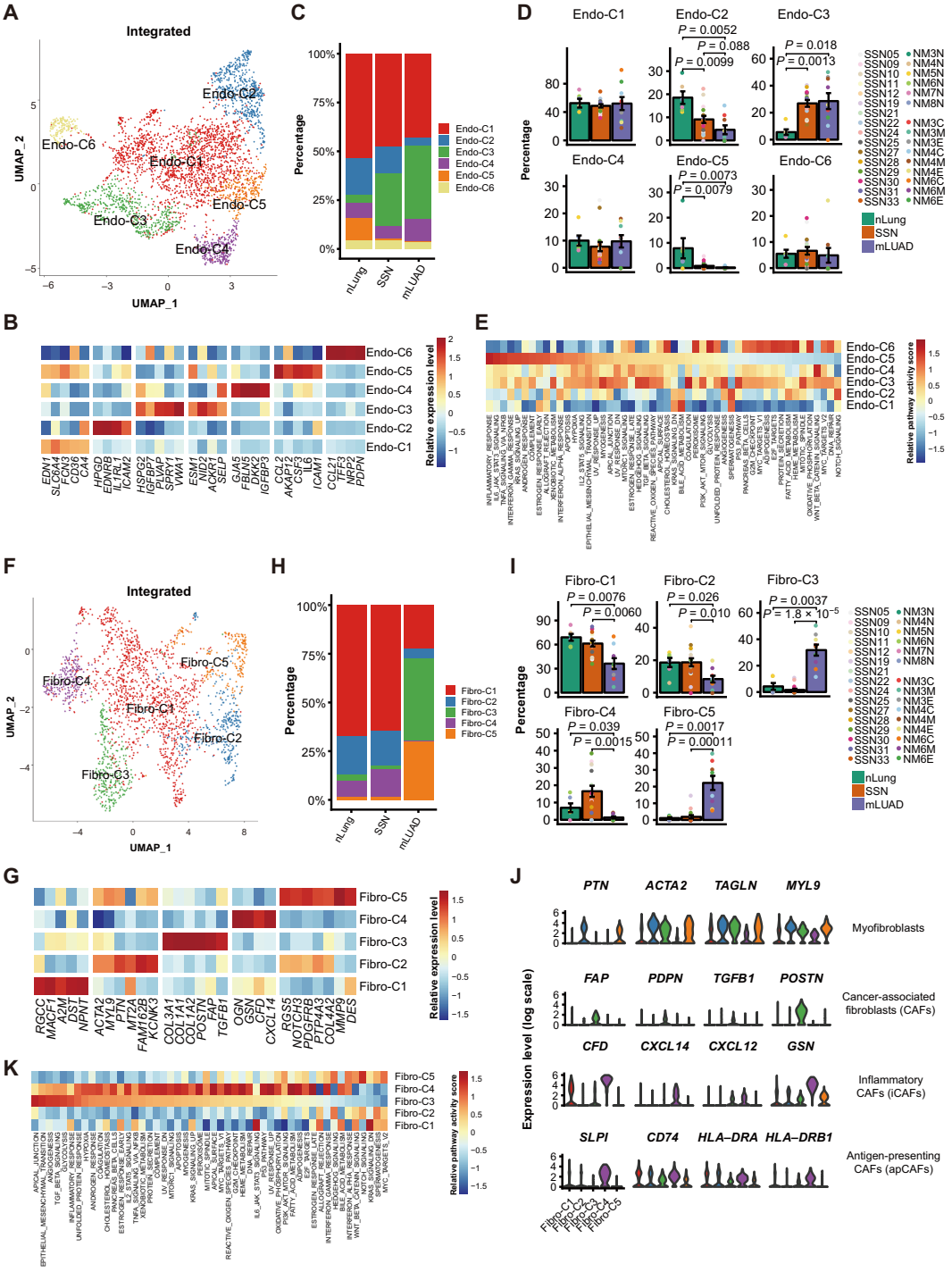

Fig. 5. Distinct EC and fibroblast subtype distribution in SSN. (**A**) UMAP projection of 3381 ECs, showing the composition of six main subtypes. (**B**) Heatmap of marker gene expression in endothelial clusters. (**C**) Average proportion of each subtype between nLung, SSN, and mLUAD. (**D**) Percentages of each EC subtype among nLung, SSN, and mLUAD. *Y* axis: Average percent of samples across the three groups. Groups are shown in different colors. Each bar plot represents one subtype. Error bars represent ± SEM for normal and tumor samples. Colored dots represent different samples. All differences with $P < 0.05$ are indicated; two-sided unpaired Wilcoxon rank sum test was used for analysis. (**E**) Differentially expressed pathways are scored per cell by GSVA between six endothelial subtypes. The relative activity scores were obtained from a linear model by limma and sorted by pathway activity in Endo-C5 cells. (**F**) UMAP projection of 2257 fibroblasts, showing the composition of five main subtypes. (**G**) Heatmap of marker gene expression in fibroblast clusters. (**H**) Average proportion of each fibroblast subtype among nLung, SSN, and mLUAD. (**I**) Percentages of each fibroblast subtype in nLung, SSN, and mLUAD. *Y* axis: Average percent of samples across the three groups. Groups are shown in different colors. Each bar plot represents one subtype. Error bars represent ± SEM. Colored dots represent different samples. All differences with $P < 0.05$ are indicated; two-sided unpaired Wilcoxon rank sum test was used for analysis. (**J**) Violin plots showing the expression of selected marker genes of different fibroblast subtypes. (**K**) Differentially expressed pathways are scored per cell by GSVA between five fibroblast subtypes. The relative activity scores were obtained from a linear model by limma and sorted by pathway activity in Fibro-C3 cells.

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

CD27$^+$*IGHD*$^-$); follicular B-C2: naïve cells (*CD20*$^+$*CD27*$^-$*IGHD*$^+$)], mucosa-associated lymphoid tissue–derived B cells (*IGHA*$^+$), germinal center B cells (*NEIL1*$^+$), two subtypes of plasma B cells [plasma B-C1: immunoglobulin G (IgG) mature (*IGHG*$^+$*PRDM1*$^+$); plasma B-C2: IgG immature (*IGHG*$^+$*PRDM1*$^-$)], and proliferating B cells (*PCNA*$^+$). Specifically, 7839 (71.90%) B/plasma cells were obtained from SSN (fig. S7B).

Compared with mLUAD, SSN showed increased abundance of follicular B-C2 cells, while the abundance of follicular B-C1 cells was decreased (fig. S7, E to G). Direct comparison of the follicular B cells of SSN and mLUAD revealed strong activation of KRAS (V-KI-RAS2 Kirsten rat sarcoma viral oncogene homolog) signaling, TNF-α–induced proliferation, and inflammatory responses among B cells in SSN, indicating an inflammatory state; whereas IFN response (types I and II), energy supply (oxidative phosphorylation and glycolysis), and biomass production [mTOR (mammalian target of rapamycin) and fatty acid metabolism] pathways were highly activated in B cells in mLUAD (fig. S7, H and I). In line with these results, global transcript abundance in B cells in mLUAD was significantly higher than that in B cells in SSN (fig. S7J).

Plasma B-C2 cells were enriched in SSN, but plasma B-C1 cells were enriched in mLUAD (fig. S7, E and F). Of note, *XBP1*, which plays a central role in the induction of the secretory phenotype of plasma B cells, was highly expressed in the plasma B-C1 cluster, and high expression levels of *IGHGs* were also detected (fig. S7K). These results suggest that B cells in SSN exhibit an inflammation-dominant gene expression pattern, while B cells in mLUAD transcribe more actively and have a stronger secretory-like phenotype.

Characterization of cell-to-cell interactions involved in SSN

To characterize intercellular interactions in SSN, we inferred putative cell-to-cell interactions based on ligand-receptor signaling inferred from our high-resolution scRNA-seq data. If one cell expresses a receptor or ligand, then this "ligand-receptor" interaction is defined as incoming or outgoing, respectively, for this cell (*15*). Fibroblasts and ECs had the most outgoing interactions across the three groups. In comparison with nLung, ECs and macrophages in SSN and mLUAD had more outgoing interactions, while CD8$^+$ T cells had more incoming interactions. Of note, NK cells, CD8$^+$ T cells, and CD4$^+$ T cells in SSN had more incoming interactions than those in nLung and mLUAD, and these incoming connections mainly originated from fibroblasts, monocytes, ECs, and macrophages (Fig. 6A).

Compared with NK cells in mLUAD and nLung, those in SSN had more interactions (Fig. 6A). To investigate how NK cells were recruited, we screened all receptors expressed on NK cells (fig. S8A), yielding two receptors of high expression: *CXCR4* and *CX3CR1*. The ligand of *CXCR4* is *CXCL12*, which was expressed by fibroblasts and ECs (Fig. 6B). In detail, *CXCL12* was exclusively expressed by Endo-C4 cells in SSN (Fig. 6C and fig. S8B). In addition, high expression of *CX3CL1*, the only ligand of *CX3CR1*, was mainly observed in Endo-C2 and Endo-C5 cells in SSN (Fig. 6C). Therefore, we propose that ECs may be responsible for the recruitment of NK cells in SSN. Furthermore, immunostaining confirmed that ECs and NK cells are more colocalized in SSN (fig. S8C).

Analysis of the biological functions revealed that interactions related to extracellular matrix remodeling and EMT (collagen-integrin, *TGF7-NRP1*, *CXCL12*/*TGFB1-CXCR4*, and *CSF1-CSFR1*) were more abundant in mLUAD (Fig. 6D). Interactions related to angiogenesis signaling were more abundant in SSN and mLUAD in comparison with nLung, and interactions related to lymphocyte recruitment and homing (*HAS2*/*MMP7-CD44* and *ICAM1/2*-integrins) were more abundant in SSN compared with mLUAD.

DISCUSSION

In this study, we have comprehensively characterized the heterogeneity of tumor cells, immune cells, and stromal cells in SSN lesions, as well as dynamic changes in cell subtype composition and intercellular interactions across nLung, SSN, and mLUAD. Compared with mLUAD, cytotoxic NK and T cells were enriched in SSN, indicating well-functioning immunosurveillance in SSN. Consistent with this finding, enrichment of DCs and mast cells in SSN conferred enhanced immune activation and recruitment of antitumor effector cells (*28*, *30*, *37*). Malignant cells in SSN were enriched in various metabolic pathways. Dynamic metabolic reprogramming during tumor progression warrants further investigation to discover how tumor metabolism shapes the TME (*38*–*40*).

The EMT is a potential driver of invasion and metastasis by human epithelial tumors (*41*). Consistent with the clinical observation that metastasis rarely occurs in SSN, the EMT program was not found in malignant cells of SSN, but it was identified in mLUAD. Moreover, at single-cell resolution, we found that fibroblasts in mLUAD interact closely with tumor cells by highly expressing ligands that promote the EMT (*35*, *42*). EMT and TGFB signaling pathways were enriched in Fibro-C3, a subtype of CAF that was specifically enriched in mLUAD. These data support the notion that fibroblasts contribute to the EMT in mLUAD and subsequent metastasis (*35*). Notably, we found a subtype of TAM (Macro-C7) characterized by high expression of *VEGFA* that was only present in mLUAD. The markers of Macro-C7 such as *SLC2A*, *HK2*, *ANGPTL4*, and *VEGFA* are hypoxia-inducible genes, and mLUAD core samples comprised most of the Macro-C7 cells. This TAM subtype was similar to a recently reported TAM in colon cancer that promotes tumor angiogenesis and metastasis (*43*).

Lavin *et al.* (*44*) reported significantly reduced abundance of CD16$^+$ NK cells in early-stage lung cancer lesions compared with nLung. However, we observed that the percentage of CD16$^+$ NK cells was not altered in SSN, but it was significantly decreased in mLUAD compared with nLung. Intercellular interactions suggest that ECs in SSN might be responsible for the recruitment of NK cells by specific ligands.

We found that the endothelial subtype composition in SSN was similar to that of mLUAD, whereas the fibroblast subtype composition in SSN was similar to that of nLung, indicating that ECs play a critical role during the early stage of tumorigenesis (*45*, *46*). Thus, we hypothesize that ECs are reprogrammed at a very early stage in LUAD TME, while fibroblasts are reprogrammed at a late stage. Notably, we found enrichment of "antigen-presenting fibroblast" cells (Fibro-C4) in SSN, supporting the immune-modulatory role of fibroblasts in SSN (*34*, *47*, *48*).

There are several limitations of our study. First, with the current scRNA-seq strategy, it cannot be determined whether a sample is from solid component or subsolid component of a given tumor. Second, the clonal relationship of T cells and B cells was not investigated. Therefore, future studies with new scRNA-seq technologies, such as immune profiling of B cell and T cell receptors and spatial transcriptomics, may help to further investigate the indolent nature of SSN. In summary, we have comprehensively decoded the multicellular ecosystem of a unique type of LUAD, which radiologically presented as SSN.

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

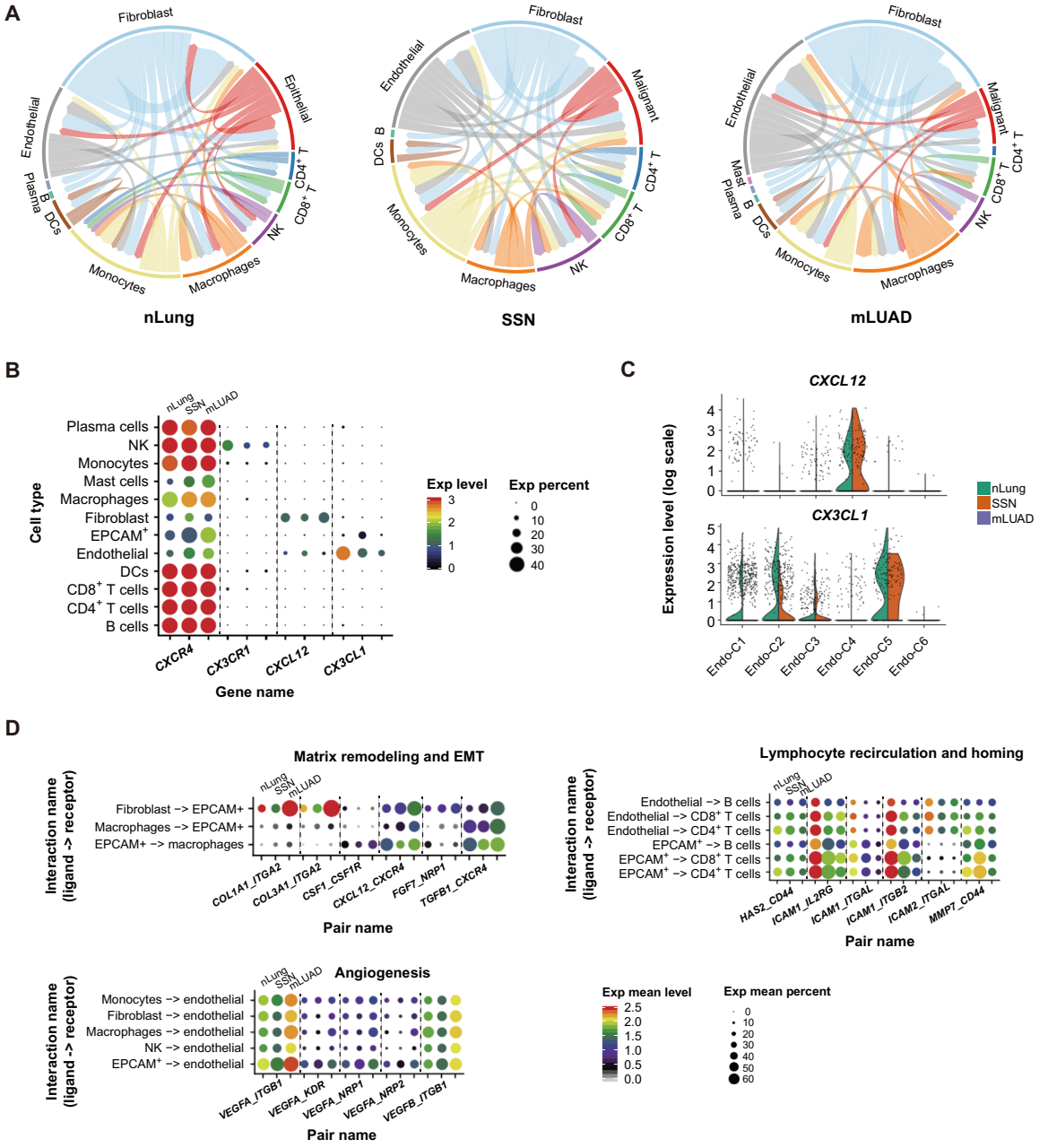

Fig. 6. Intercellular interactions in normal lungs and lung tumors. (**A**) Circos plot showing the intercellular interactions among different cell types in nLung, SSN, and mLUAD. The strings are directional and represent interactions determined on the basis of expression of a ligand by one cell type and expression of a corresponding receptor by another cell type. The thickness of each string corresponds to the amount of different interaction pairs, colored according to cell type. (**B**) Dot plot showing the expression level and percentage of selected genes in different cell types among nLung, SSN, and mLUAD. (**C**) Violin plots showing the expression of *CXCL12* and *CX3CL1* in different EC subtypes, split by sample sources. (**D**) Dot plot showing the mean expression level and percentage of selected interaction pairs involved in EMT, lymphocyte homing, and angiogenesis. The expression of each gene was considered separately for each sample source.

MATERIALS AND METHODS

Patients and sample collection

Patients with pulmonary SSNs who underwent surgery at the Department of Thoracic Surgery of Peaking University People's Hospital were enrolled with the following criteria: (i) pulmonary SSNs in CT images; (ii) pathologically diagnosed LUAD, adenocarcinoma in situ, and atypical adenomatous hyperplasia; (iii) no history of other malignancies; and (iv) no anticancer treatment (chemotherapy, radiotherapy, targeted therapy, etc.) before surgery. Chest CT images of enrolled patients were reviewed by two experienced thoracic surgeons independently. After SSN samples were resected, tumor tissues were cut into two pieces along the long axis: One half was used for scRNA-seq, and the other half was used for pathological diagnosis and WES. Blood samples were collected before surgery, and white blood cells were isolated and stored at −80°C until WES. The clinical information of these patients is summarized in table S1. This study

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

was approved by the Ethics Committee Board of Peking University People's Hospital, and written informed consent was obtained from all participants included in this study.

Preparation of single-cell suspensions

Primary tumor tissue samples were transported in ice-cold H1640 (Gibco, Life Technologies) immediately after surgical resection. The primary tumor tissues were rinsed with phosphate-buffered saline (PBS; Thermo Fisher Scientific), minced into ~1-mm cubic piece, and ground with a UTTD (ULTRA-TURRAX® Tube Drive) disperser (IKA, Germany). The ground tumor tissue samples were digested by 0.25% trypsin (Gibco, Life Technologies), terminated by H1640 supplemented with 10% fetal bovine serum (Gibco, Life Technologies), and then transferred to 10 ml of digestion medium containing collagenase IV (100 U/ml; Gibco, Life Technologies) and dispase (0.6 U/ml; Gibco, Life Technologies). The digested samples were filtered through a 70-μm nylon mesh. After centrifuging, the pelleted cells were suspended with ice-cold red blood cell lysis buffer (Solarbio) and filtered with a 40-μm nylon mesh. Last, the pelleted cells were suspended with 1 ml of Dulbecco's PBS (Solarbio), and the concentrations of live cells and clumped cells were determined using an automated cell counter (Countstar).

Droplet-based single-cell sequencing

Using the Single Cell 3′ Library and Gel Bead Kit V2 (10X Genomics) and Chromium Single Cell A Chip Kit (10X Genomics), the cell suspension was loaded onto the Chromium single-cell controller (10X Genomics) to generate single-cell gel beads in the emulsion (GEMs) according to the manufacturer's protocol. Briefly, single cells were suspended in PBS containing 0.04% bovine serum albumin. Approximately 10,000 cells were added to each channel, and approximately 6000 cells were recovered. The captured cells were lysed, and the released RNA was barcoded via reverse transcription in individual GEMs. Reverse transcription was performed at 53°C for 45 min, followed by 85°C for 5 min, after which the temperature was held at 4°C. Complementary DNA was generated and amplified, after which its quality was assessed using an Agilent 4200 (performed by CapitalBio Technology, Beijing) according to the manufacturer's instructions. scRNA-seq libraries were constructed using the Single Cell 3′ Library Gel Bead Kit V2. The libraries were lastly sequenced using an Illumina NovaSeq 6000 with a paired-end 150–base pair (PE150) reading strategy (performed by CapitalBio Technology, Beijing).

Multiplex immunohistochemistry

Formalin-fixed/paraffin-embedded samples from the analyzed patients were collected from Peking University People's Hospital. The specimens were treated as previously described (*49*). The antibodies used in this section were anti-CD3 (ZM0417), anti-CD20 (TA800385), anti-CD68 (ZM0060), anti-CD56 (ZM0057), anti-CD16 (16559-1-AP), anti-CD31 (ZM0044), and anti-FAP (ab53066). The antigenic binding sites were visualized using the Opal 7-Color Manual IHC Kit (PerkinElmer, NEL811001KT) according to the manufacturer's protocol. Multicolor immunohistochemistry data were collected using a Mantra Quantitative Pathology Workstation (PerkinElmer, CLS140089) and analyzed by InForm (version 2.2.1).

scRNA-seq data processing

Raw gene expression matrices were generated for each sample using the Cell Ranger (version 2.2.0) Pipeline coupled with human reference version GRCh38. After removal of empty droplets using the DropletUtils (*50*) package (version 1.2.2), the output-filtered gene expression matrices were analyzed by R software (version 3.5.3) with the Seurat (*51*) package (version 3.0.0). In brief, genes expressed at a proportion >0.1% of the data and cells with >200 genes detected were selected for further analyses. Low-quality cells were removed if they met the following criteria: (i) <500 UMIs (unique molecular identifiers), (ii) >6000 or <200 genes, or (iii) >10% UMIs derived from the mitochondrial genome. After removal of low-quality cells, the gene expression matrices were normalized by the NormalizeData function, and 2000 features with high cell-to-cell variation were calculated using the FindVariableFeatures function. To reduce the dimensionality of the datasets, the RunPCA function was conducted with default parameters on linear transformation scaled data generated by the ScaleData function. Next, the ElbowPlot, DimHeatmap, and JackStrawPlot functions were used to identify the true dimensionality of each dataset, as recommended by the Seurat developers. Last, we clustered cells using the FindNeighbors and FindClusters functions and performed nonlinear dimensional reduction with the RunUMAP function with default settings. All details regarding the Seurat analyses performed in this work can be found in the website tutorial (https://satijalab.org/seurat/v3.0/pbmc3k_tutorial.html).

Multiple dataset integration

To compare cell types and proportions across three conditions, we used the integration methods described at https://satijalab.org/seurat/v3.0/integration.html (*52*). The Seurat package (version 3.0.0) was used to assemble multiple distinct scRNA-seq datasets into an integrated and unbatched dataset. In brief, we identified 2000 features with high cell-to-cell variation as described above. Next, we identified "anchors" between individual datasets with the FindIntegrationAnchors function and inputted these anchors into the IntegrateData function to create a "batch-corrected" expression matrix of all cells, which allowed cells from different datasets to be integrated and analyzed together.

Cell type annotation and cluster marker identification

After nonlinear dimensional reduction and projection of all cells into two-dimensional space by UMAP, cells clustered together according to common features. The FindAllMarkers function in Seurat was used to find markers for each of the identified clusters. Clusters were then classified and annotated on the basis of expression of canonical markers of particular cell types. Clusters that expressed two or more canonical cell type markers were classified as doublet cells, and clusters that expressed no canonical cell type markers were classified as low-quality cells. Both doublet cell clusters and low-quality cell clusters were excluded from further analyses.

Subclustering of major cell types

For each major cell type, cells were extracted from the overview integrated dataset first. Next, these major cell types were integrated for further subclustering. After integration, genes were scaled to unit variance. Scaling, PCA, and clustering were performed as described above.

Defining cell state scores

We used cell scores to evaluate the degree to which individual cells expressed a certain predefined expression gene set. The cell scores were initially based on the average expression of the genes from the

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

predefined gene set in the respective cell (*15*). The AddModuleScore function in Seurat was used to implement the method with default settings. We used four well-defined naïve markers (*CCR7*, *TCF7*, *LEF1*, and *SELL*), 12 cytotoxicity-associated genes (*PRF1*, *IFNG*, *GNLY*, *NKG7*, *GZMB*, *GZMA*, *GZMH*, *KLRK1*, *KLRB1*, *KLRD1*, *CTSW*, and *CST7*), and five exhausted markers (*LAG3*, *TIGIT*, *PDCD1*, *CTLA4*, and *HAVCR2*) to define naïve, cytotoxicity, and exhaustion scores. The resting and active scores of DCs were measured on the basis of the top 30 genes of "LM22" (*53*).

WES and analysis

Paired-end read sequences were aligned to human genome hg19 [University of California, Santa Cruz (UCSC)] using the Burrows-Wheeler Aligner (version 0.7.17) (*54*) with default parameters and sorted using the SortSam function embedded in Picard (http://Picard.Sourceforge.net; version 2.18.7). The MarkDuplicates function was used to mark and discard duplicates. A base quality recalibration was carried out using the Genome Analysis Toolkit (GATK version 3.8.0) (*55*). To portray copy number states across the whole genome based on WES data, Sequenza (R package, version 2.1.2) (*56*) with default parameters was applied to model copy numbers to integers with consideration of both ploidy and cellularity. Last, the results were displayed by copynumber (R package, version 1.22.0).

CNV estimation and identification of malignant cells

To infer CNVs from the scRNA-seq data, we used an approach described previously with the R code provided in https://github.com/broadinstitute/inferCNV with the default parameters. Immune cells and stromal cells were considered as putative nonmalignant cells, and their CNV estimates were used to define a baseline (*13*). The calculated CNV signal (x axis) was defined as the mean square of the CNV estimates across all genomic locations. The calculated CNV R-scores (y axis) were defined as the Pearson correlation coefficient between each cell's CNV pattern and the average CNV pattern of the top 5% of cells from the same tumor with respect to CNV signal. All $EPCAM^+$ epithelial cells in the lung tumor samples were inputted, and those with CNV R-scores of ≥ 0.3 were defined as malignant cells.

Expression programs of intratumor heterogeneity

Malignant cells from each lung cancer sample (samples with less than 100 malignant cells were excluded from analysis) were first normalized as described above and then center-scaled for each gene. After transformation of all negative values to zero, non-negative matrix factorization was performed using the nmf function in the NMF R package (version 0.21.0), and the top 10 ranks of each sample were calculated. Genes were ranked by their average scores, and the top 30 genes for each cluster were defined as a meta-signature and used to calculate the coefficient of variation (CV). The CV for each program was defined as the SD divided by the mean of the program signature value and multiplied by 100 to give a percentage value (%). To remove duplicates, the Pearson correlations for any 2 of the 10 programs were calculated within each sample. We only retained programs that had a CV greater than 25% and a Pearson correlation less than 0.4 within the corresponding sample.

Putative interactions between cell types

We identified putative interactions between different cell types based on the expression of a receptor by one cell type and the expression of an interacting ligand by another cell type as described previously (*15*). The set of potential receptor-ligand interactions was obtained from Ramilowski *et al.* (*57*). A ligand or receptor transcript was defined as "expressed" by a given cell type if its average expression in that cell type was above the threshold of 0.5 in $\log_2$ scale, and it was expressed in at least 10% of cells of that type. If one cell type expressed the receptor or the ligand, then the corresponding interaction was defined as incoming or outgoing, respectively, for that cell type. The edges between pairs of cell types above the third quantile of the maximum number of ligand-receptor pairs were displayed.

Gene set variation analysis

Pathway analyses were performed on the 50 hallmark pathways described in the molecular signature database (*58*). We also assessed metabolic pathway activities using a previously described curated dataset with 85 metabolic pathways (*23*). To assign pathway activity estimates to individual cells, we applied GSVA using standard settings, as implemented in the GSVA (*21*) R package (version 1.30.0) as described previously (*17*). The differential activities of pathways between clusters or conditions were calculated using Limma (*59*) R package (version 3.38.3). Significantly disturbed pathways were identified with Benjamini-Hochberg–corrected P value of ≤ 0.01.

Differential abundance and expression test

Differential abundance test (DA test) was used to detect changes in cell type abundance between conditions and summarized in table S3. Differential expression test (DE test) using a pseudo-bulk method for comparing differential gene expression was also performed. Both analyses using methods from the edgeR (*60*) R package (version 3.24.3) and codes were adapted from a Bioconductor tutorial (https://osca.bioconductor.org/multi-sample-comparisons.html). Significant changes were identified with Benjamini-Hochberg–corrected P value of ≤ 0.05.

Diffusion map and pseudotime analysis

Single cells assigned to $CD8^+$ T cell and $CD4^+$ T cells clusters were subjected to diffusion map and pseudotime analysis. First, differentially expressed genes for each cluster were identified. Data were then imported to the DiffusionMap function from the destiny R package (version 2.12.0) for diffusion map and pseudotime analysis with default settings. A naïve T cell was used as the root cell for diffusion pseudotime computation using the first two diffusion components to present results.

Correlation to TCGA data

We downloaded preprocessed gene expression data [$\log_2(x + 1)$-transformed RSEM (RNA-seq by expectation-maximization)-normalized count] and clinical data for the TCGA LUAD (TCGA-LUAD) RNA-seq gene expression dataset from UCSC Xena (http://xena.ucsc.edu/). Kaplan-Meier analysis and log-rank testing were performed to investigate the effect of the top 30 specific gene signatures on patient survival. All analyses were implemented using R packages survival (version 2.43.3) and survminer (version 0.4.4).

Statistics

The statistical tools, methods, and threshold for each analysis are explicitly described with the results or detailed in the figure legends or Materials and Methods.

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/5/eabd9738/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. J. E. Walter, M. A. Heuvelmans, U. Yousaf-Khan, M. D. Dorrius, E. Thunnissen, A. Schermann, H. J. M. Groen, C. M. van der Aalst, K. Nackaerts, R. Vliegenthart, H. J. de Koning, M. Oudkerk, New subsolid pulmonary nodules in lung cancer screening: The NELSON trial. *J. Thorac. Oncol.* **13**, 1410–1414 (2018).
2. L. Fan, Y. Wang, Y. Zhou, Q. Li, W. Yang, S. Wang, F. Shan, X. Zhang, J. Shi, W. Chen, S. Y. Liu, Lung cancer screening with low-dose CT: Baseline screening results in shanghai. *Acad. Radiol.* **26**, 1283–1291 (2019).
3. F. Fu, Y. Zhang, Z. Wen, D. Zheng, Z. Gao, H. Han, L. Deng, S. Wang, Q. Liu, Y. Li, L. Shen, X. Shen, Y. Zhao, Z. Zhao, T. Ye, J. Xiang, Y. Zhang, Y. Sun, H. Hu, H. Chen, Distinct prognostic factors in patients with stage I non-small cell lung cancer with radiologic part-solid or solid lesions. *J. Thorac. Oncol.* **14**, 2133–2142 (2019).
4. T. Ye, L. Deng, S. Wang, J. Xiang, Y. Zhang, H. Hu, Y. Sun, Y. Li, L. Shen, L. Xie, W. Gu, Y. Zhao, F. Fu, W. Peng, H. Chen, Lung adenocarcinomas manifesting as radiological part-solid nodules define a special clinical subtype. *J. Thorac. Oncol.* **14**, 617–627 (2019).
5. A. Hattori, T. Matsunaga, K. Takamochi, S. Oh, K. Suzuki, Importance of ground glass opacity component in clinical stage IA radiologic invasive lung cancer. *Ann. Thorac. Surg.* **104**, 313–320 (2017).
6. K. Aokage, T. Miyoshi, G. Ishii, M. Kusumoto, S. Nomura, S. Katsumata, K. Sekihara, K. Tane, M. Tsuboi, Influence of ground glass opacity and the corresponding pathological findings on survival in patients with clinical stage I non-small cell lung cancer. *J. Thorac. Oncol.* **13**, 533–542 (2018).
7. R. Yip, K. Li, L. Liu, D. Xu, K. Tam, D. F. Yankelevitz, E. Taioli, B. Becker, C. I. Henschke, Controversies on lung cancers manifesting as part-solid nodules. *Eur. Radiol.* **28**, 747–759 (2018).
8. J. A. Howington, M. G. Blum, A. C. Chang, A. A. Balekian, S. C. Murthy, Treatment of stage I and II non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* **143**, e278S–e313S (2013).
9. D. P. Naidich, A. A. Bankier, H. MacMahon, C. M. Schaefer-Prokop, M. Pistolesi, J. M. Goo, P. Macchiarini, J. D. Crapo, C. J. Herold, J. H. Austin, W. D. Travis, Recommendations for the management of subsolid pulmonary nodules detected at CT: A statement from the Fleischner Society. *Radiology* **266**, 304–317 (2013).
10. H. MacMahon, D. P. Naidich, J. M. Goo, K. S. Lee, A. N. C. Leung, J. R. Mayo, A. C. Mehta, Y. Ohno, C. A. Powell, M. Prokop, G. D. Rubin, C. M. Schaefer-Prokop, W. D. Travis, P. E. Van Schil, A. A. Bankier, Guidelines for management of incidental pulmonary nodules detected on CT images: From the fleischner society 2017. *Radiology* **284**, 228–243 (2017).
11. Y. Li, X. Li, H. Li, Y. Zhao, Z. Liu, K. Sun, X. Zhu, Q. Qi, B. An, D. Shen, R. Li, T. Liu, J. Mi, L. Wang, F. Yang, F. Bai, J. Wang, Genomic characterisation of pulmonary subsolid nodules: Mutational landscape and radiological features. *Eur. Respir. J.* **55**, 1901409 (2020).
12. N. K. Altorki, G. J. Markowitz, D. Gao, J. L. Port, A. Saxena, B. Stiles, T. McGraw, V. Mittal, The lung microenvironment: An important regulator of tumour growth and metastasis. *Nat. Rev. Cancer* **19**, 9–31 (2019).
13. A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suva, A. Regev, B. E. Bernstein, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
14. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth II, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jane-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
15. S. V. Puram, I. Tirosh, A. S. Parikh, A. P. Patel, K. Yizhak, S. Gillespie, C. Rodman, C. L. Luo, E. A. Mroz, K. S. Emerick, D. G. Deschler, M. A. Varvares, R. Mylvaganam, O. Rozenblatt-Rosen, J. W. Rocco, W. C. Faquin, D. T. Lin, A. Regev, B. E. Bernstein, Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
16. E. Azizi, A. J. Carr, G. Plitas, A. E. Cornish, C. Konopacki, S. Prabhakaran, J. Nainys, K. Wu, V. Kiseliovas, M. Setty, K. Choi, R. M. Fromme, P. Dao, P. T. McKenney, R. C. Wasti, K. Kadaveru, L. Mazutis, A. Y. Rudensky, D. Pe'er, Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308.e36 (2018).
17. D. Lambrechts, E. Wauters, B. Boeckx, S. Aibar, D. Nittner, O. Burton, A. Bassez, H. Decaluwe, A. Pircher, K. Van den Eynde, B. Weynand, E. Verbeken, P. De Leyn, A. Liston, J. Vansteenkiste, P. Carmeliet, S. Aerts, B. Thienpont, Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
18. M. Deprez, L. E. Zaragosi, M. Truchi, C. Becavin, S. Ruiz García, M.-J. Arguel, M. Plaisant, V. Magnone, K. Lebrigand, S. Abelanet, F. Brau, A. Paquet, D. Pe'er, C.-H. Marquette, S. Leroy, P. Barbry, A single-cell atlas of the human healthy airways. *Am. J. Respir. Crit. Care Med.* **202**, 1636–1645 (2020).
19. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
20. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
21. S. Hanzelmann, R. Castelo, J. Guinney, GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
22. Cancer Genome Atlas Research Network, Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
23. Z. Xiao, Z. Dai, J. W. Locasale, Metabolic landscape of the tumor microenvironment at single cell resolution. *Nat. Commun.* **10**, 3763 (2019).
24. X. Guo, Y. Zhang, L. Zheng, C. Zheng, J. Song, Q. Zhang, B. Kang, Z. Liu, L. Jin, R. Xing, R. Gao, L. Zhang, M. Dong, X. Hu, X. Ren, D. Kirchhoff, H. G. Roider, T. Yan, Z. Zhang, Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
25. J. Qian, S. Olbrecht, B. Boeckx, H. Vos, D. Laoui, E. Etlioglu, E. Wauters, V. Pomella, S. Verbandt, P. Busschaert, A. Bassez, A. Franken, M. V. Bempt, J. Xiong, B. Weynand, Y. van Herck, A. Antoranz, F. M. Bosisio, B. Thienpont, G. Floris, I. Vergote, A. Smeets, S. Tejpar, D. Lambrechts, A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020).
26. A. Kallies, D. Zehn, D. T. Utzschneider, Precursor exhausted T cells: Key to successful immunotherapy? *Nat. Rev. Immunol.* **20**, 128–136 (2020).
27. R. Zilionis, C. Engblom, C. Pfirschke, V. Savova, D. Zemmour, H. D. Saatcioglu, I. Krishnan, G. Maroni, C. V. Meyerovitz, C. M. Kerwin, S. Choi, W. G. Richards, A. De Rienzo, D. G. Tenen, R. Bueno, E. Levantini, M. J. Pittet, A. M. Klein, Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity* **50**, 1317–1334.e10 (2019).
28. A. Gardner, B. Ruffell, Dendritic cells and cancer immunity. *Trends Immunol.* **37**, 855–865 (2016).
29. J. P. Bottcher, E. Bonavita, P. Chakravarty, H. Blees, M. Cabeza-Cabrerizo, S. Sammicheli, N. C. Rogers, E. Sahai, S. Zelenay, C. Reis e Sousa, NK cells stimulate recruitment of cDC1 into the tumor microenvironment promoting cancer immune control. *Cell* **172**, 1022–1037.e14 (2018).
30. K. Palucka, J. Banchereau, Cancer immunotherapy via dendritic cells. *Nat. Rev. Cancer* **12**, 265–277 (2012).
31. C. Engblom, C. Pfirschke, M. J. Pittet, The role of myeloid cells in cancer therapies. *Nat. Rev. Cancer* **16**, 447–462 (2016).
32. Y. Okabe, R. Medzhitov, Tissue biology perspective on macrophages. *Nat. Immunol.* **17**, 9–17 (2016).
33. C. Murdoch, M. Muthana, S. B. Coffelt, C. E. Lewis, The role of myeloid cells in the promotion of tumour angiogenesis. *Nat. Rev. Cancer* **8**, 618–631 (2008).
34. E. Elyada, M. Bolisetty, P. Laise, W. F. Flynn, E. T. Courtois, R. A. Burkhart, J. A. Teinor, P. Belleau, G. Biffi, M. S. Lucito, S. Sivajothi, T. D. Armstrong, D. D. Engle, K. H. Yu, Y. Hao, C. L. Wolfgang, Y. Park, J. Preall, E. M. Jaffee, A. Califano, P. Robson, D. A. Tuveson, Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).
35. R. Kalluri, The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* **16**, 582–598 (2016).
36. G. J. Yuen, E. Demissie, S. Pillai, B lymphocytes and cancer: A love-hate relationship. *Trends Cancer* **2**, 747–757 (2016).
37. S. J. Galli, M. Grimbaldeston, M. Tsai, Immunomodulatory mast cells: Negative, as well as positive, regulators of immunity. *Nat. Rev. Immunol.* **8**, 478–486 (2008).
38. N. N. Pavlova, C. B. Thompson, The emerging hallmarks of cancer metabolism. *Cell Metab.* **23**, 27–47 (2016).
39. M. Reina-Campos, J. Moscat, M. Diaz-Meco, Metabolism shapes the tumor microenvironment. *Curr. Opin. Cell Biol.* **48**, 47–53 (2017).
40. C. A. Lyssiotis, A. C. Kimmelman, Metabolic interactions in the tumor microenvironment. *Trends Cell Biol.* **27**, 863–875 (2017).
41. I. Pastushenko, C. Blanpain, EMT transition states during tumor progression and metastasis. *Trends Cell Biol.* **29**, 212–226 (2019).
42. L. Bu, H. Baba, N. Yoshida, K. Miyake, T. Yasuda, T. Uchihara, P. Tan, T. Ishimoto, Biological heterogeneity and versatility of cancer-associated fibroblasts in the tumor microenvironment. *Oncogene* **38**, 4887–4901 (2019).
43. L. Zhang, Z. Li, K. M. Skrzypczynska, Q. Fang, W. Zhang, S. A. O'Brien, Y. He, L. Wang, Q. Zhang, A. Kim, R. Gao, J. Orf, T. Wang, D. Sawant, J. Kang, D. Bhatt, D. Lu, C. M. Li,

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

A. S. Rapaport, K. Perez, Y. Ye, S. Wang, X. Hu, X. Ren, W. Ouyang, Z. Shen, J. G. Egen, Z. Zhang, X. Yu, Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* **181**, 442–459.e29 (2020).
44. Y. Lavin, S. Kobayashi, A. Leader, E. D. Amir, N. Elefant, C. Bigenwald, R. Remark, R. Sweeney, C. D. Becker, J. H. Levine, K. Meinhof, A. Chow, S. Kim-Shulze, A. Wolf, C. Medaglia, H. Li, J. A. Rytlewski, R. O. Emerson, A. Solovyov, B. D. Greenbaum, C. Sanders, M. Vignali, M. B. Beasley, R. Flores, S. Gnjatic, D. Pe'er, A. Rahman, I. Amit, M. Merad, Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* **169**, 750–765.e17 (2017).
45. B. Ghesquiere, B. W. Wong, A. Kuchnio, P. Carmeliet, Metabolism of stromal and immune cells in health and disease. *Nature* **511**, 167–176 (2014).
46. N. Maishi, K. Hida, Tumor endothelial cells accelerate tumor metastasis. *Cancer Sci.* **108**, 1921–1926 (2017).
47. H. Kobayashi, A. Enomoto, S. L. Woods, A. D. Burt, M. Takahashi, D. L. Worthley, Cancer-associated fibroblasts in gastrointestinal cancer. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 282–295 (2019).
48. M. B. Buechler, S. J. Turley, A short field guide to fibroblast function in immunity. *Semin. Immunol.* **35**, 48–58 (2018).
49. K. He, J. Zhou, F. Yang, C. Chi, H. Li, Y. Mao, B. Hui, K. Wang, J. Tian, J. Wang, Near-infrared intraoperative imaging of thoracic sympathetic nerves: From preclinical study to clinical trial. *Theranostics* **8**, 304–313 (2018).
50. A. T. L. Lun, S. Riesenfeld, T. Andrews, T. P. Dao, T. Gomes; Participants in the 1st Human Cell Atlas Jamboree, J. C. Marioni, EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019).
51. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
52. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
53. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
54. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
56. F. Favero, T. Joshi, A. M. Marquard, N. J. Birkbak, M. Krzystanek, Q. Li, Z. Szallasi, A. C. Eklund, Sequenza: Allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
57. J. A. Ramilowski, T. Goldberg, J. Harshbarger, E. Kloppmann, M. Lizio, V. P. Satagopam, M. Itoh, H. Kawaji, P. Carninci, B. Rost, A. R. Forrest, A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
58. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
59. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
60. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

Acknowledgments: We thank D. Shen (Department of Pathology, Peking University People's Hospital) for assistance with pathological diagnosis. We thank Y. Hou and Y. Wang (Thoracic Laboratory of Peking University People's Hospital) for sample collection. We thank Q. Qi (Department of Radiology, Peking University People's Hospital) for the revision of this paper. **Funding:** This study was supported by the National Natural Science Foundation of China (81702256 to M.Q., 81871879 to J.W., and 81772469 to F.Y.), the Beijing Municipal Natural Science Foundation (7182169 to M.Q.), the National Science and Technology Major Project (2018ZX10302205 and 2019YFC1315702) to F.B., and the Guangdong Province Key Research and Development Program (2019B020226002) to F.B. **Author contributions:** J.W., F.B., M.Q., X.X., and F.Y. conceived and designed the study. M.Q., F.Y., Q.H., H.G., and J.L. performed the experiments. X.X. performed the bioinformatic analyses. X.X. and M.Q. wrote the manuscript. F.B., F.Y., and J.W. edited the manuscript. All authors read and approved the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive of the BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number HRA000154 and are publicly accessible at http://bigd.big.ac.cn/gsa-human. The codes to process and analyze data are publicly available at GitHub repository (https://github.com/SeadonXing/SSN_scRNA-seq). Detailed information will be available from the corresponding authors upon reasonable request.

Submitted 23 July 2020
Accepted 8 December 2020
Published 27 January 2021
10.1126/sciadv.abd9738

Citation: X. Xing, F. Yang, Q. Huang, H. Guo, J. Li, M. Qiu, F. Bai, J. Wang, Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. *Sci. Adv.* **7**, eabd9738 (2021).

Downloaded from http://advances.sciencemag.org/ on June 14, 2021

Science Advances

Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing

Xudong Xing, Fan Yang, Qi Huang, Haifa Guo, Jiawei Li, Mantang Qiu, Fan Bai and Jun Wang

Sci Adv **7** (5), eabd9738.
DOI: 10.1126/sciadv.abd9738

ARTICLE TOOLS	http://advances.sciencemag.org/content/7/5/eabd9738
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2021/01/25/7.5.eabd9738.DC1
REFERENCES	This article cites 60 articles, 7 of which you can access for free http://advances.sciencemag.org/content/7/5/eabd9738#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Downloaded from http://advances.sciencemag.org/ on June 14, 2021