Cell Reports Medicine, Volume *4*

Supplemental information

Desmoplastic stromal signatures predict patient outcomes in pancreatic ductal adenocarcinoma

Shamik Mascharak, Jason L. Guo, Deshka S. Foster, Anum Khan, Michael F. Davitt, Alan T. Nguyen, Austin R. Burcham, Malini S. Chinta, Nicholas J. Guardino, Michelle Griffin, David M. Lopez, Elisabeth Miller, Michael Januszyk, Shyam S. Raghavan, Teri A. Longacre, Daniel J. Delitto, Jeffrey A. Norton, and Michael T. Longaker

Supplemental Information

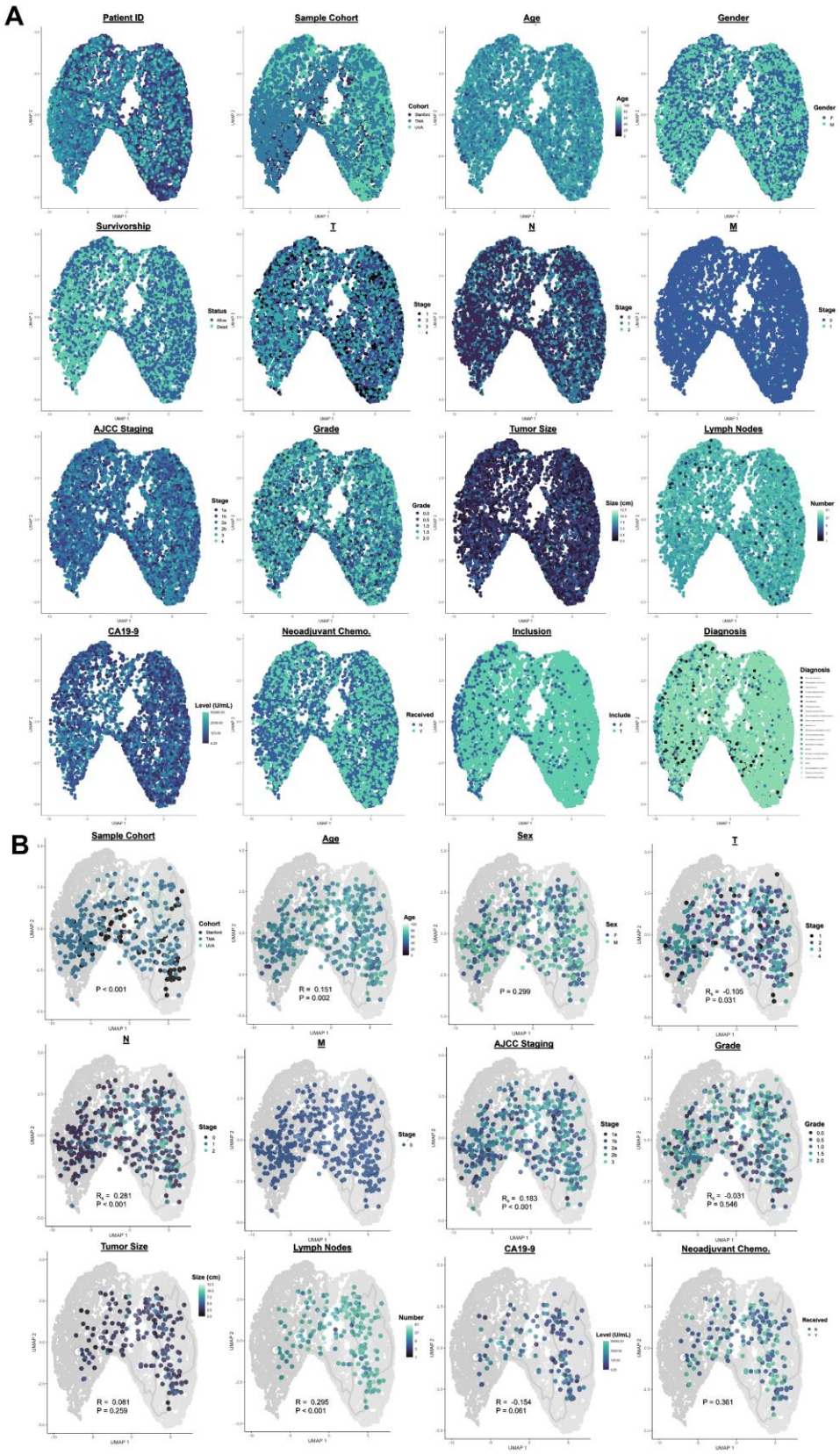


Figure S1: Correlation of matrix architectural data with clinical characteristics, related to Figure 1. (A) Mapping of patient metadata (age, gender, stage, grade, etc.) onto the trajectory. (**B**) Localization of included patient-level centroids. Correlations of relevant metadata with patient-level pseudotime are shown as Pearson coefficients (R) and associated p-values for continuous metadata (age, tumor size, lymph nodes, CA19-9), Spearman coefficients (R_s) and associated p-values for ordinal metadata (T, N, AJCC stage, grade), and ANOVA p-values for categorical metadata (cohort, gender, neoadjuvant chemotherapy).

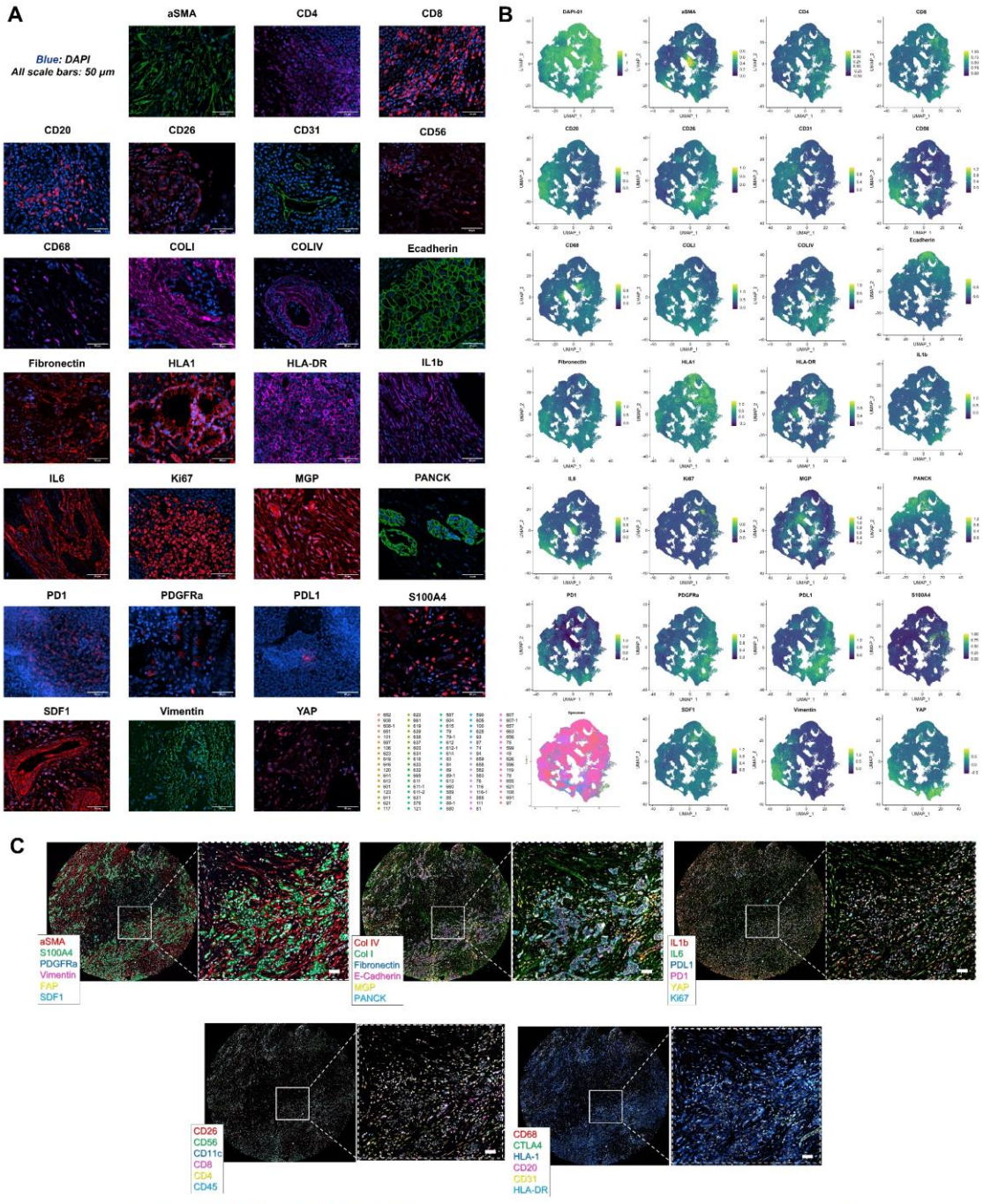

A
Blue: DAPI
All scale bars: 50 μm
aSMA
CD4
CD8
CD20
CD26
CD31
CD56
CD68
COLI
COLIV
Ecadherin
Fibronectin
HLA1
HLA-DR
IL1b
IL6
Ki67
MGP
PANCK
PD1
PDGFRa
PDL1
S100A4
SDF1
Vimentin
YAP
B
C
aSMA
S100A4
PDGFRa
Vimentin
FAP
SDF1
Col IV
Col I
Fibronectin
E-Cadherin
MGP
PANCK
IL1b
IL6
PDL1
PD1
YAP
Ki67
CD26
CD56
CD11c
CD8
CD4
CD45
CD68
CTLA4
HLA-1
CD20
CD31
HLA-DR
D
Cell
Segmentation

Figure S2: CODEX markers and associated validation data, related to Figure 2. (A) Validation of individual CODEX markers used for definition of cell phenotypes. Scale bars represent 50 μm. **(B)** Feature plots of each corresponding CODEX marker, as well as specimen of origin (bottom left). (**C**) Visualizations of multiplexed staining, with zoomed-in areas of interest in white boxes. Scale bars represent 200 μm. (**D**) Example of cell segmentation from DAPI staining in CODEX.

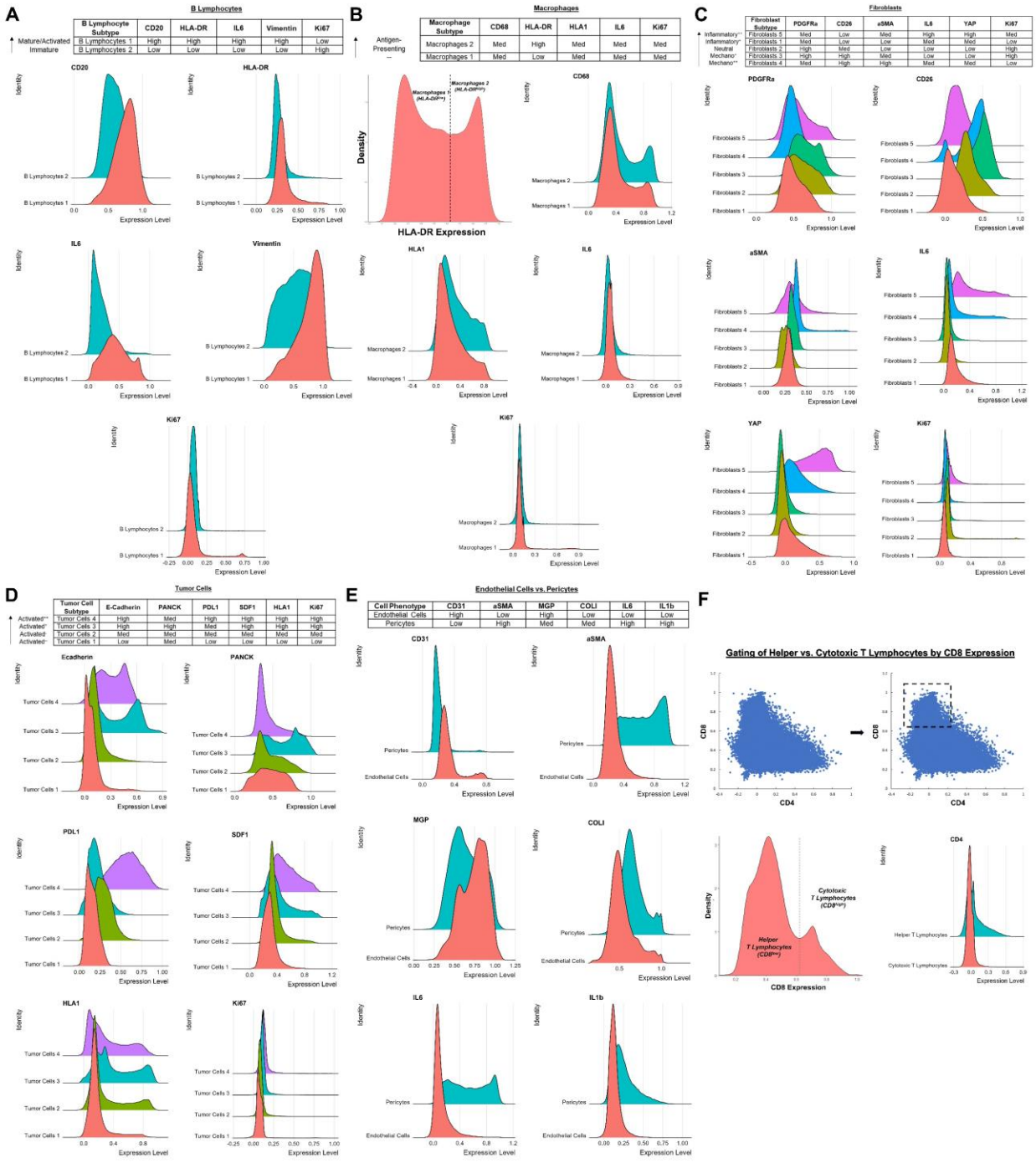


Figure S3: Characteristic protein expression by cell subpopulations, related to Figure 2. (A) Summary of differential protein expression by B lymphocyte subtypes, including profiles for CD20, HLA-DR, IL6, Vimentin, and Ki67. (**B**) Summary of differential protein expression by macrophage subtypes, including splitting of macrophage subtypes by HLA-DR expression and protein expression profiles for CD68, HLA1, IL6, and Ki67. (**C**) Summary of differential protein expression by fibroblast subtypes, including profiles for PDGFRa, CD26, aSMA,

IL6, YAP, and Ki67. (**D**) Summary of differential protein expression by tumor cell subtypes, including profiles for E-cadherin, PANCK, PDL1, SDF1, HLA1, and Ki67. (**E**) Summary of differential protein expression by endothelial cells and pericytes, including profiles for CD31, aSMA, MGP, COLI, IL6, and IL1b. (**F**) Summary of differential protein expression by T lymphocyte subtypes, including gating of T lymphocyte subtypes by CD8 expression in FACS-style scatter plot, CD8 density plot, and expression profile for CD4.

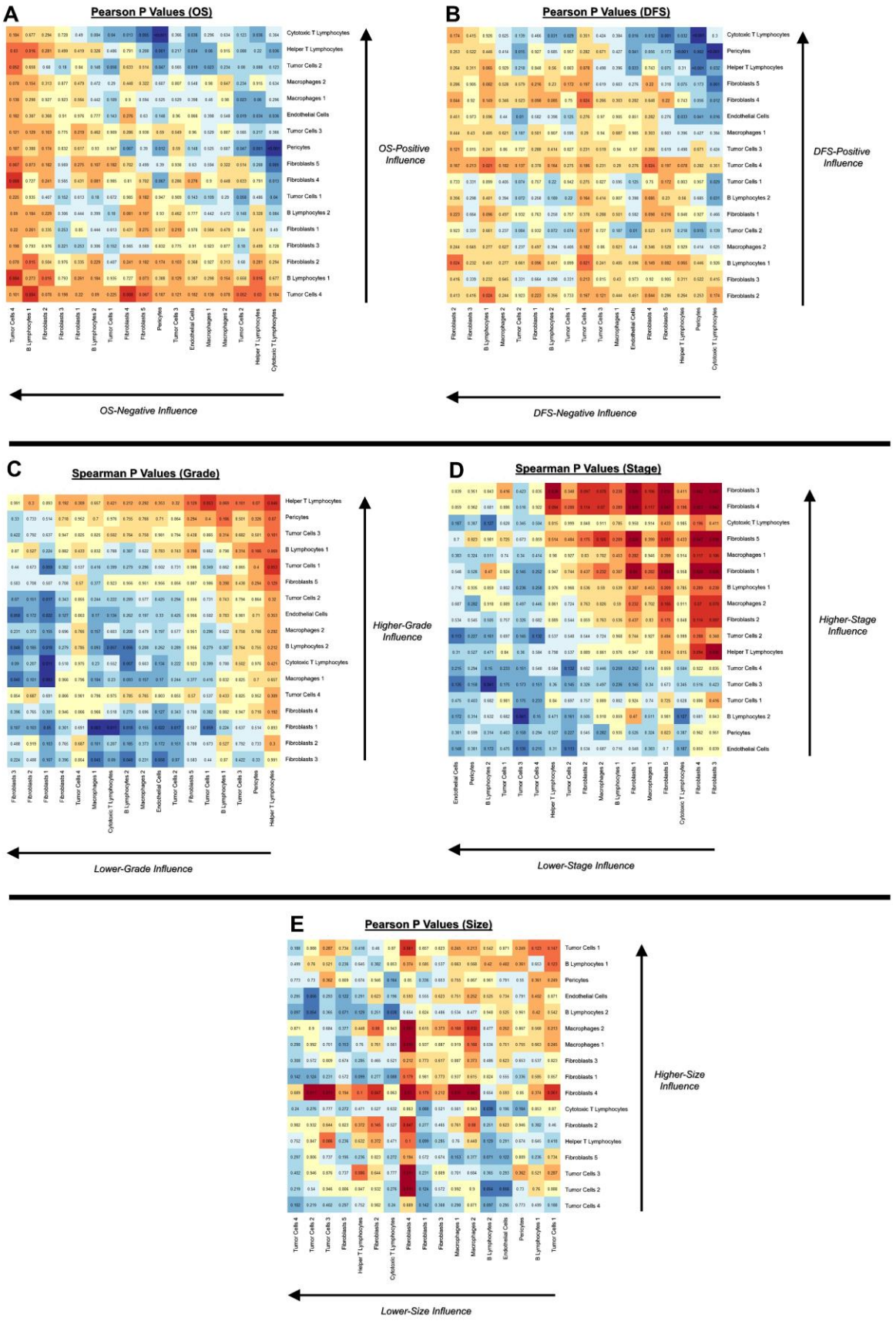
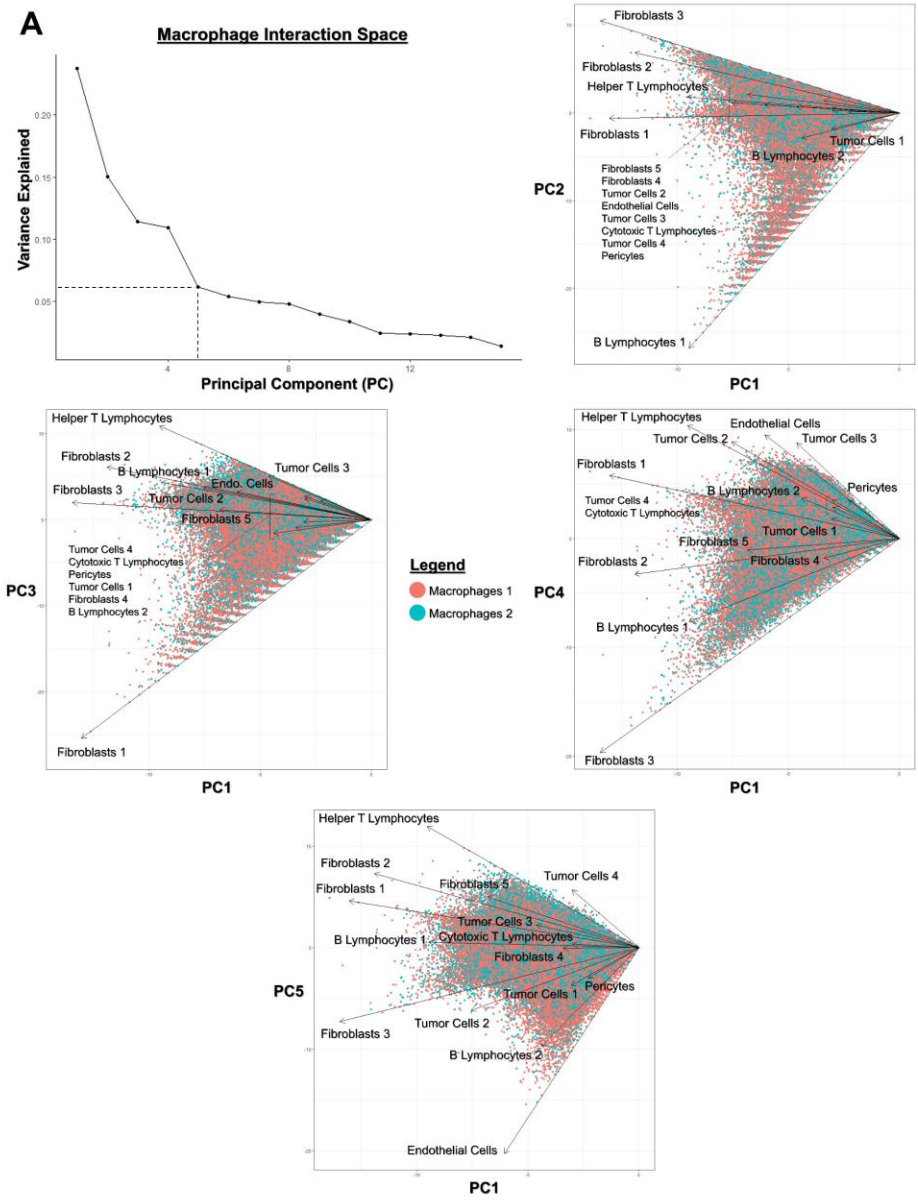

A
Pearson P Values (OS)
Cytotoxic T Lymphocytes
Helper T Lymphocytes
Tumor Cells 2
Macrophages 2
Macrophages 1
Endothelial Cells
Tumor Cells 3
Pericytes
Fibroblasts 5
Fibroblasts 4
Tumor Cells 1
B Lymphocytes 2
Fibroblasts 1
Fibroblasts 3
Fibroblasts 2
B Lymphocytes 1
Tumor Cells 4
OS-Positive Influence
OS-Negative Influence
B
Pearson P Values (DFS)
DFS-Positive Influence
DFS-Negative Influence
C
Spearman P Values (Grade)
Higher-Grade Influence
Lower-Grade Influence
D
Spearman P Values (Stage)
Higher-Stage Influence
Lower-Stage Influence
E
Pearson P Values (Size)
Higher-Size Influence
Lower-Size Influence

Figure S4: P values for correlation between cell-cell spatial interactions and clinically relevant metrics, related to Figures 3 and 5. (A-E) P values for overall survival (A), disease-free survival (B), tumor grade (C), AJCC stage (D), and tumor size (E).


A
Macrophage Interaction Space
Variance Explained
Principal Component (PC)
PC1
PC2
PC3
PC4
PC5
Legend
Macrophages 1
Macrophages 2
Fibroblasts 3
Fibroblasts 2
Helper T Lymphocytes
Fibroblasts 1
Tumor Cells 1
B Lymphocytes 2
Fibroblasts 5
Fibroblasts 4
Tumor Cells 2
Endothelial Cells
Tumor Cells 3
Cytotoxic T Lymphocytes
Tumor Cells 4
Pericytes
B Lymphocytes 1
B Lymphocytes 1
Endo. Cells
Fibroblasts 5
Tumor Cells 4
Cytotoxic T Lymphocytes
Pericytes
Tumor Cells 1
Fibroblasts 4
B Lymphocytes 2
Endothelial Cells
Tumor Cells 3
Tumor Cells 2
Pericytes
Tumor Cells 1
Fibroblasts 5
Fibroblasts 4
Fibroblasts 3
Tumor Cells 4
Tumor Cells 1
Fibroblasts 5
Tumor Cells 3
Cytotoxic T Lymphocytes
Fibroblasts 4
Pericytes
Tumor Cells 2
B Lymphocytes 2
Endothelial Cells

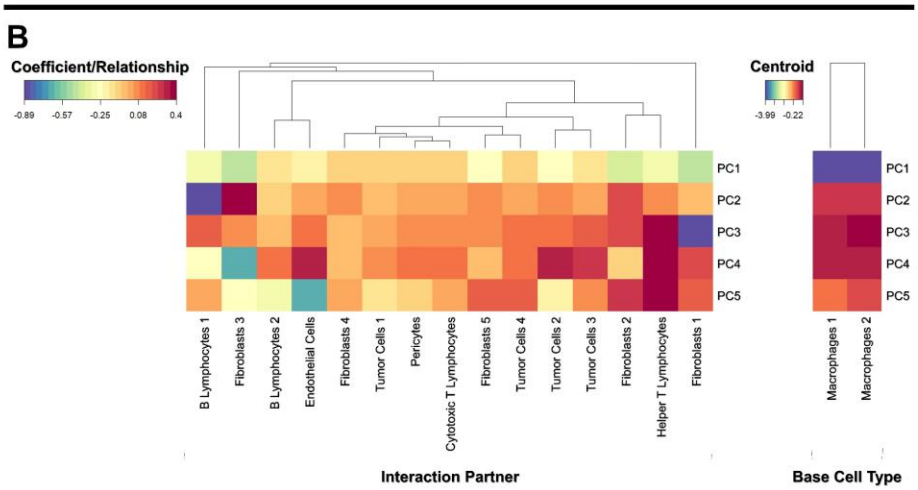

B
Coefficient/Relationship
-0.89
-0.57
-0.25
0.08
0.4
Centroid
-3.99
-0.22
PC1
PC2
PC3
PC4
PC5
B Lymphocytes 1
Fibroblasts 3
B Lymphocytes 2
Endothelial Cells
Fibroblasts 4
Tumor Cells 1
Pericytes
Cytotoxic T Lymphocytes
Fibroblasts 5
Tumor Cells 4
Tumor Cells 2
Tumor Cells 3
Fibroblasts 2
Helper T Lymphocytes
Fibroblasts 1
Macrophages 1
Macrophages 2
Interaction Partner
Base Cell Type

Figure S5: The macrophage interactome and its variation with protein phenotype, related to Figure 6. (**A**) Principal component analysis (PCA) of interaction space up to 5 principal components, as determined by vertex of scree plot. Macrophages exhibit largely similar interaction patterns, albeit with some differences in adjacency with helper T lymphocytes based on protein phenotype. (**B**) Heatmaps of PCA coefficients and centroids for each macrophage subtype.

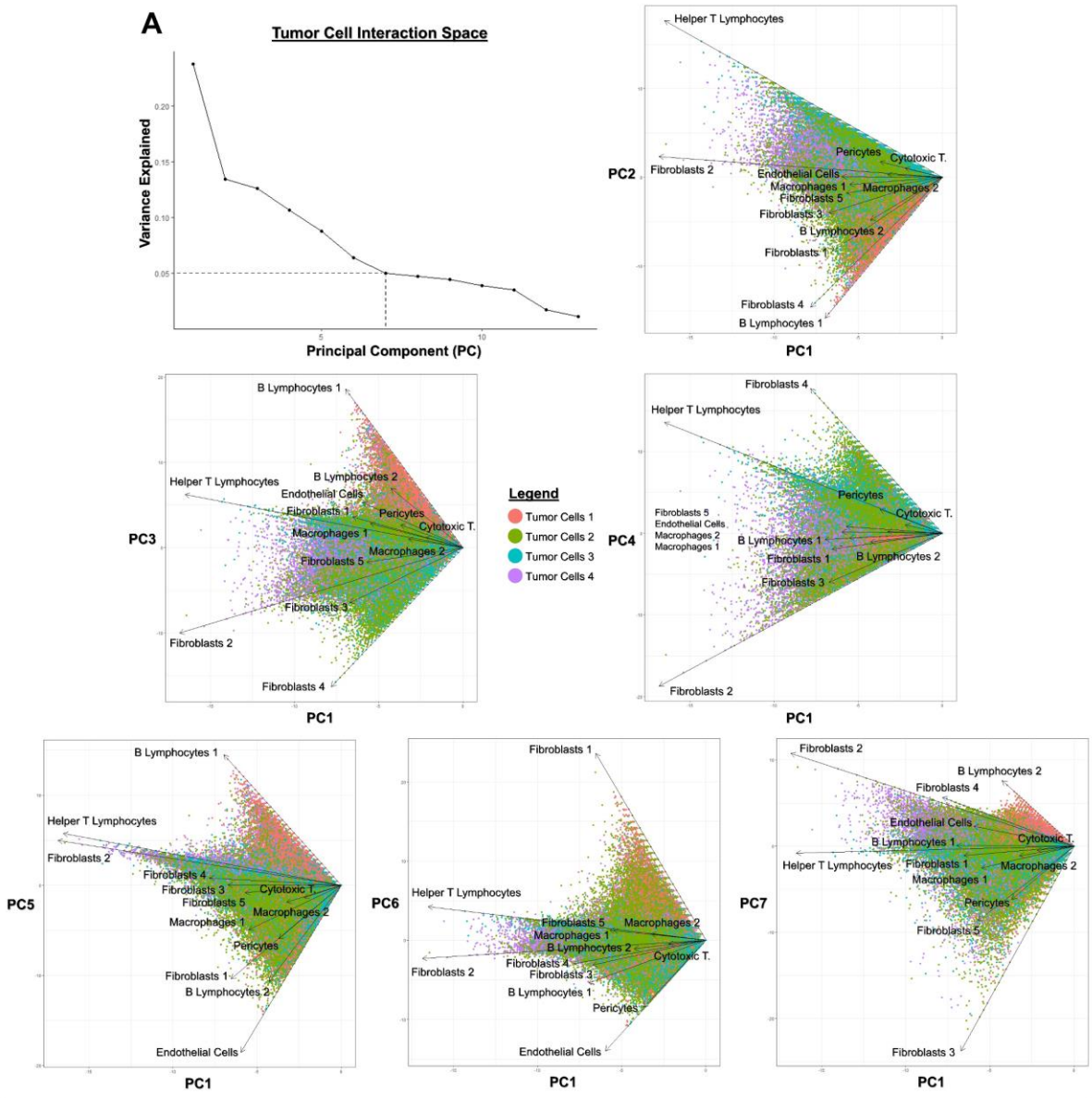

A
Tumor Cell Interaction Space
Variance Explained
Principal Component (PC)
PC1
PC2
PC3
PC4
PC5
PC6
PC7
Legend
Tumor Cells 1
Tumor Cells 2
Tumor Cells 3
Tumor Cells 4
Helper T Lymphocytes
Fibroblasts 2
Pericytes
Cytotoxic T.
Endothelial Cells
Macrophages 1
Macrophages 2
Fibroblasts 5
Fibroblasts 3
B Lymphocytes 2
Fibroblasts 1
Fibroblasts 4
B Lymphocytes 1
B
Coefficient/Relationship
-0.79
-0.4
0
0.39
0.79
Centroid
-2.76
-0.11
Interaction Partner
Base Cell Type

Figure S6: The tumor cell interactome and its variation with protein phenotype, related to Figure 6. (**A**) Principal component analysis (PCA) of interaction space up to 7 principal components, as determined by vertex of scree plot. Tumor cells exhibit highly heterogeneous interaction patterns. (**B**) Heatmaps of PCA coefficients and centroids for each fibroblast subtype.

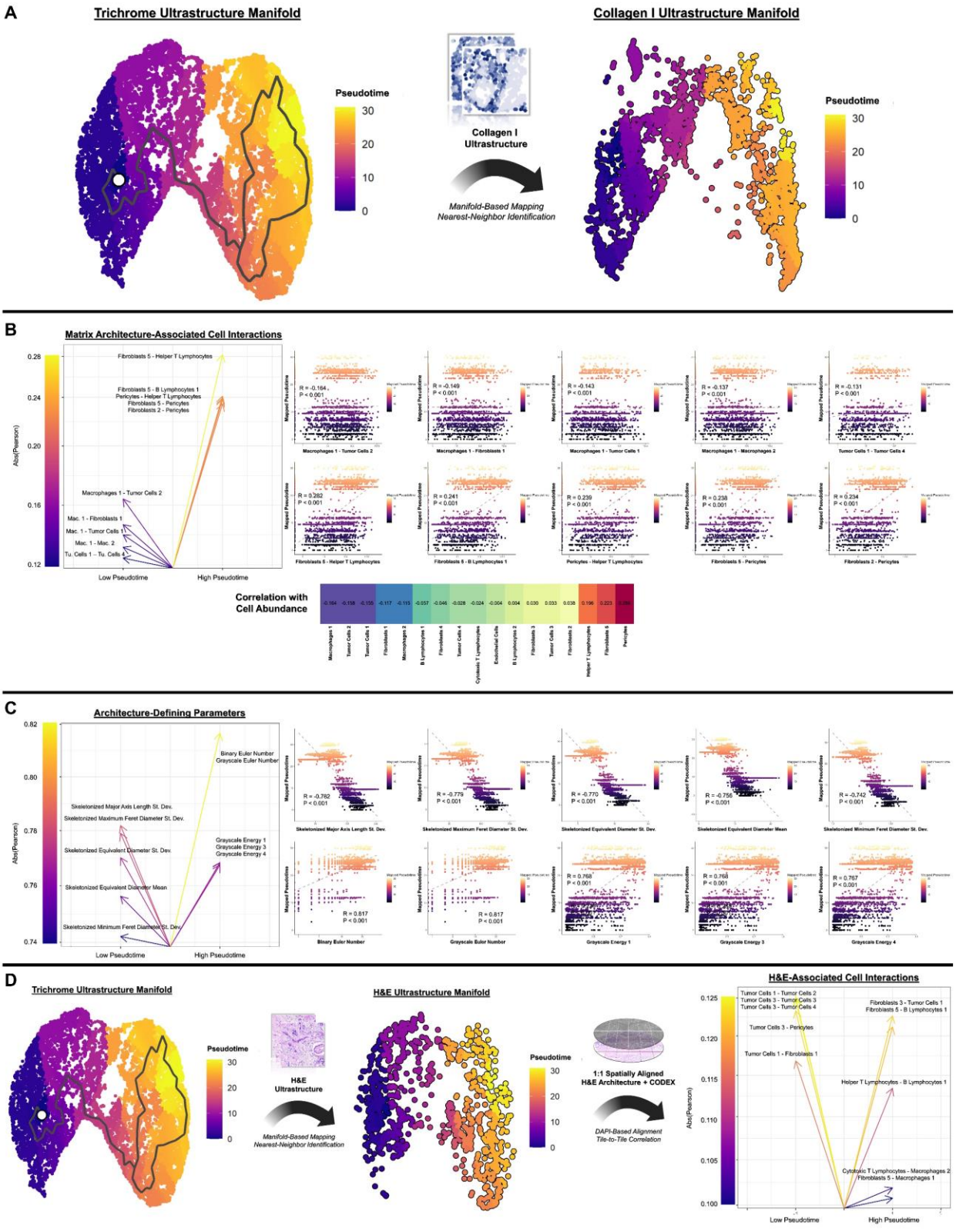

A
Trichrome Ultrastructure Manifold
Pseudotime
30
20
10
0
Collagen I Ultrastructure
Manifold-Based Mapping
Nearest-Neighbor Identification
Collagen I Ultrastructure Manifold
Pseudotime
30
20
10
0
B
Matrix Architecture-Associated Cell Interactions
Abs(Pearson)
Low Pseudotime
High Pseudotime
Correlation with Cell Abundance
C
Architecture-Defining Parameters
Abs(Pearson)
Low Pseudotime
High Pseudotime
D
Trichrome Ultrastructure Manifold
Pseudotime
H&E Ultrastructure
Manifold-Based Mapping
Nearest-Neighbor Identification
H&E Ultrastructure Manifold
Pseudotime
1:1 Spatially Aligned H&E Architecture + CODEX
DAPI-Based Alignment
Tile-to-Tile Correlation
H&E-Associated Cell Interactions
Abs(Pearson)
Low Pseudotime
High Pseudotime

Figure S7: Integration of matrix ultrastructural analysis with CODEX, related to Figures 1 and 2. (**A**) Mapping of collagen I ultrastructure and trichrome ultrastructure using model transformation in *Monocle3* and nearest neighbor identification in *RANN*. (**B**) Identification of top 5 CODEX cell interactions correlated with low and high matrix pseudotime (top), as well as cell type abundances correlated with low and high matrix pseudotime (bottom). (**C**) Identification of top 5 ultrastructural parameters correlated with low and high matrix pseudotime. (**D**) Mapping of H&E to trichrome, followed by one-to-one spatial alignment to CODEX using DAPI mask and identification of top cell interactions correlated with H&E matrix architecture.

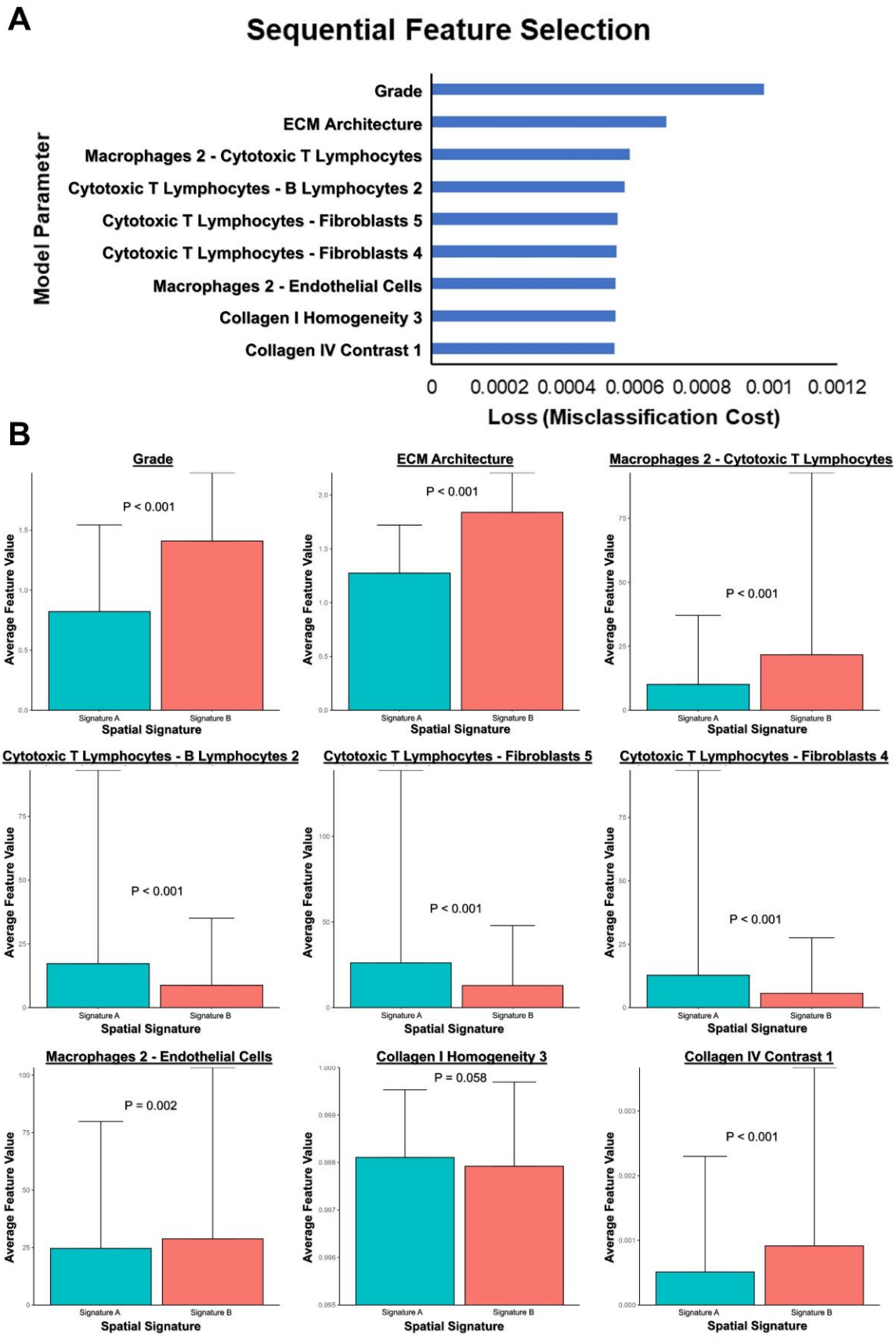

A
Sequential Feature Selection
Model Parameter
Grade
ECM Architecture
Macrophages 2 - Cytotoxic T Lymphocytes
Cytotoxic T Lymphocytes - B Lymphocytes 2
Cytotoxic T Lymphocytes - Fibroblasts 5
Cytotoxic T Lymphocytes - Fibroblasts 4
Macrophages 2 - Endothelial Cells
Collagen I Homogeneity 3
Collagen IV Contrast 1
0 0.0002 0.0004 0.0006 0.0008 0.001 0.0012
Loss (Misclassification Cost)
B
Grade
P < 0.001
ECM Architecture
P < 0.001
Macrophages 2 - Cytotoxic T Lymphocytes
P < 0.001
Cytotoxic T Lymphocytes - B Lymphocytes 2
P < 0.001
Cytotoxic T Lymphocytes - Fibroblasts 5
P < 0.001
Cytotoxic T Lymphocytes - Fibroblasts 4
P < 0.001
Macrophages 2 - Endothelial Cells
P = 0.002
Collagen I Homogeneity 3
P = 0.058
Collagen IV Contrast 1
P < 0.001
Average Feature Value
Signature A
Signature B
Spatial Signature

Figure S8: Generation of compact feature set for machine learning-based spatial signature using sequential feature selection in MATLAB, related to Figure 7. (**A**) Progression of lower misclassification costs produced by forward selection of features. Additional features beyond this selection generated no further improvement to misclassification cost on the training dataset. (**B**) Average feature values for the subsample-level training data.

	Characteristic		TMA	Stanford	UVA
All specimens			n = 359	n = 159	n = 148
	Median age (range) - yr		55 (23-80)	71 (26-90)	66 (40-90)
	Gender - no. (%)				
		Male	143 (40)	83 (52)	70 (47)
		Female	100 (28)	68 (43)	78 (53)
		NA	116 (33)	8 (5)	0 (0)
	Pathology - no. (%)				
		Normal	36 (10)	0 (0)	0 (0)
		PDAC	245 (68)	138 (87)	147 (99)
		Other neoplasm	78 (22)	21 (13)	1 (1)
PDAC only (included in analysis)			n = 239	n = 128	n = 146
	Disease stage - no. (%)				
		1a-1b	86 (36)	24 (19)	35 (24)
		2a-2b	138 (58)	70 (55)	92 (63)
		3	5 (2)	34 (27)	19 (13)
		NA	10 (4)	0 (0)	0 (0)
	Grade - no. (%)				
		0	42 (18)	24 (19)	12 (8)
		0.5-1	116 (49)	58 (45)	51 (35)
		1.5-2	63 (26)	29 (23)	69 (47)
		NA	18 (8)	17 (13)	14 (10)
	Median tumor size (range) - cm		-	3 (0.25-13)	3 (0.5-8.5)
	Median lymph nodes (range) - no.		-	24 (10-86)	22 (0-53)
	Median pre-op CA19-9 (range)		-	166 (0-53704)	63 (2-22720)
	Neoadjuvant chemotherapy - no. (%)		-	28 (22)	63 (43)
	Median disease-free survival (range) - days		-	386 (8-2034)	347 (8-2582)
	Median overall survival (range) - days		-	503 (8-2034)	439 (8-2661)

Table S1: Patient characteristics by cohort origin (third-party tissue microarray, Stanford Hospital, and University of Virginia University Hospital), related to Figure 1.

Cluster/Cell Annotation	Rationale
B Lymphocytes 1	CD20^{high}, HLA-DR^{high}, IL6^{high}, Vimentin^{high}, FAP, YAP^{low}, CD56, CTLA4
B Lymphocytes 2	CD20^{low}, HLA-DR^{low}, IL6^{low}, Vimentin^{low}, FAP^{low}, CD56, CTLA4^{low}
Endothelial Cells	CD31, MGP^{high}, COLI^{low}
Fibroblasts 1	PDGFRa^{med}, CD26^{low}, aSMA^{low}, , IL6^{med}, PDL1^{low}, YAP^{low}, COLI^{low}, COLIV^{low}
Fibroblasts 2	PDGFRa^{high}, CD26^{med}, aSMA^{low}, IL6^{low}, PDL1^{low}, Fibronectin^{low}
Fibroblasts 3	PDGFRa^{high}, CD26^{high}, aSMA^{med}, IL6^{low}, PDL1, COLIV, Fibronectin^{low}
Fibroblasts 4	PDGFRa^{med}, CD26^{high}, aSMA^{high}, IL6^{med}, PDL1, COLI^{low}, COLIV, Fibronectin^{low}
Fibroblasts 5	PDGFRa^{med}, CD26^{low}, aSMA^{med}, IL6^{high}, PDL1, YAP, COLI^{low}, COLIV, Fibronectin^{low}
Macrophages 1	CD68, HLA-DR^{low}, HLA1, IL6
Macrophages 2	CD68, HLA-DR^{high}, HLA1, IL6
Pericytes	aSMA^{high}, MGP, COLI^{med}, IL6^{high}, IL1b^{high}
Helper T Lymphocytes	CD4^{high}, CD8^{low}, S100A4, HLA-DR, HLA1, Ki67
Cytotoxic T Lymphocytes	CD4^{low}, CD8^{high}, S100A4, HLA-DR, HLA1, Ki67
Tumor Cells 1	E-cadherin^{low}, PANCK^{med}, PDL1^{low}, SDF1^{low}, HLA1^{low}, Vimentin^{low}, FAP^{low}
Tumor Cells 2	E-cadherin^{med}, PANCK^{med}, PDL1^{med}, SDF1^{med}, HLA1^{med}
Tumor Cells 3	E-cadherin^{high}, PANCK^{high}, PDL1^{med}, SDF1^{high}, HLA1^{high}
Tumor Cells 4	E-cadherin^{high}, PANCK^{med}, PDL1^{high}, SDF1^{high}, HLA1^{high}

Table S2: Characteristic protein expression by CODEX-defined cell populations, related to Figure 2.

	Artificial Neural Network (ANN)	Generalized Additive Model (GAM)	K-Nearest Neighbor (KNN) Model	Linear Discriminant Analysis (LDA)	Random Forest (RF)	Support Vector Machine (SVM)
AUC	0.902617	0.872331	0.899993	0.881663	0.885204	0.880519
Sensitivity	0.916409	0.934985	0.897833	0.894737	0.934985	0.928793
Specificity	0.815331	0.804878	0.80662	0.815331	0.783972	0.797909
Balanced Accuracy	0.865869	0.869931	0.852226	0.855033	0.859478	0.863350

Table S3: Performance metrics for machine learning models on independent, blinded testing dataset, related to Figure 7.