**Molecular heterogeneity of cancer associated fibroblasts: implication in cancer pathogenesis, prognosis, and immunotherapy resistance**

Phillip M. Galbo Jr[1,2], Xingxing Zang[2,3], Deyou Zheng[1,4,†]

[1] Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

[2] Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

[3] Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

[4] Departments of Neurology and Neuroscience, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

## Supplementary Materials

I. **Analysis of CAF subclusters in MEL, HNSC and LC cancers**

II. **Supplemental Figures and Legends**

III. **Supplemental Table Captions**

**Molecular characterization of CAF heterogeneity from individual cancer types**

We extracted CAFs from each of the three cancer types using the cell type annotation information by the original authors, and then performed CAF-focused sub-clustering analysis using *Seurat* (1). For the melanoma CAFs, the original study by Tirosh *et al* only defined one cluster and found that CAFs highly expressed genes related to complement immunity (2). Our analysis showed that the MEL CAFs could actually be separated into two distinct subtypes (MEL CAFs 1 and 2) (Supplementary Figure 1A). MEL CAFs 1 comprised of 42 cells (68.9%; red) and was classified as inflammatory-like CAFs (iCAFs), based on elevated expression of several molecules related to complement immunity (*C3*, *C1S*, etc.) and chemokines (*CXCL14*) (Supplementary Figure 1B), and supported by gene ontology (GO) analysis of the genes expressed significantly higher in MEL CAFs 1 (i.e., markers), which showed enrichments for biological processes associated with complement immunity (Complement Cascade, etc.) and immune cell activation (Supplementary Figure 1C). MEL CAFs 2 comprised of 19 cells (31.1%; blue) (Supplementary Figure 1A) and was classified as proliferating CAFs (pCAFs), based on elevated expression of genes associated with cell cycle (*MKI67*, *BIRC5,* etc.) (Supplementary Figure 1B) and enrichment of GO terms associated with cell cycle (Supplementary Figure 1C). Our re-analysis thus defined two subtypes for the MEL CAFs that were originally considered as one group.

For HNSC, Puram *et al* were able to identify three clusters of fibroblasts: CAFs, intermediate (resting), and myofibroblasts; and, through sub-clustering, further distinguish the CAFs into two subpopulations (CAF1 and CAF2) (3). We identified five distinct sub-populations (HNSC CAFs 1-5) (Supplementary Figure 1D). Our HNSC 1 closely resembled the myofibroblast cluster by Purm *et al*. It comprised of 709 cells (49.9%, red) and was classified as myofibroblast-like CAFs (myCAFs), based on elevated expression of marker genes for activated fibroblast markers (ACTA2) and myogenesis (*MYH11*, *MYL9*, *MYLK*) (Supplementary Figure 1E), and enrichment of GO terms for smooth muscle contraction, myofibril assembly and regulation of blood vessel diameter (Supplementary Figure 1F). Our HNSC CAFs 2 and 3 closely resembled Puram *et al's* CAF 1 and CAF 2 clusters, respectively (3). HNSC CAFs 2 comprised of 306 cells (21.5%, dark) (Supplementary Figure 1D) and was classified as desmoplastic CAF (dCAF), based on elevated expression of markers associated with collagen (*COL1A1*, *COL1A2*, and *COL3A1*) and extracellular matrix (ECM) remodeling (*MMP11* and *MMP19*) (Supplementary Figure 1E) and enrichment of GO terms associated with ECM remodeling (Supplementary Figure 1F). Our HNSC CAFs 3 comprised of 252 cells (17.7%, light green) (Supplementary Figure 1D) and was classified as iCAFs, based on elevated expression of complement immunity related molecules (*C3*), CAF chemokine marker *CXCL12* (*SDF-1*), and *CXCL14* (Supplementary Figure 1E), and enrichment of GO terms related to development and signaling (Supplementary Figure 1F). Our HNSC CAF 4 comprised of 91 cells (6.4%,

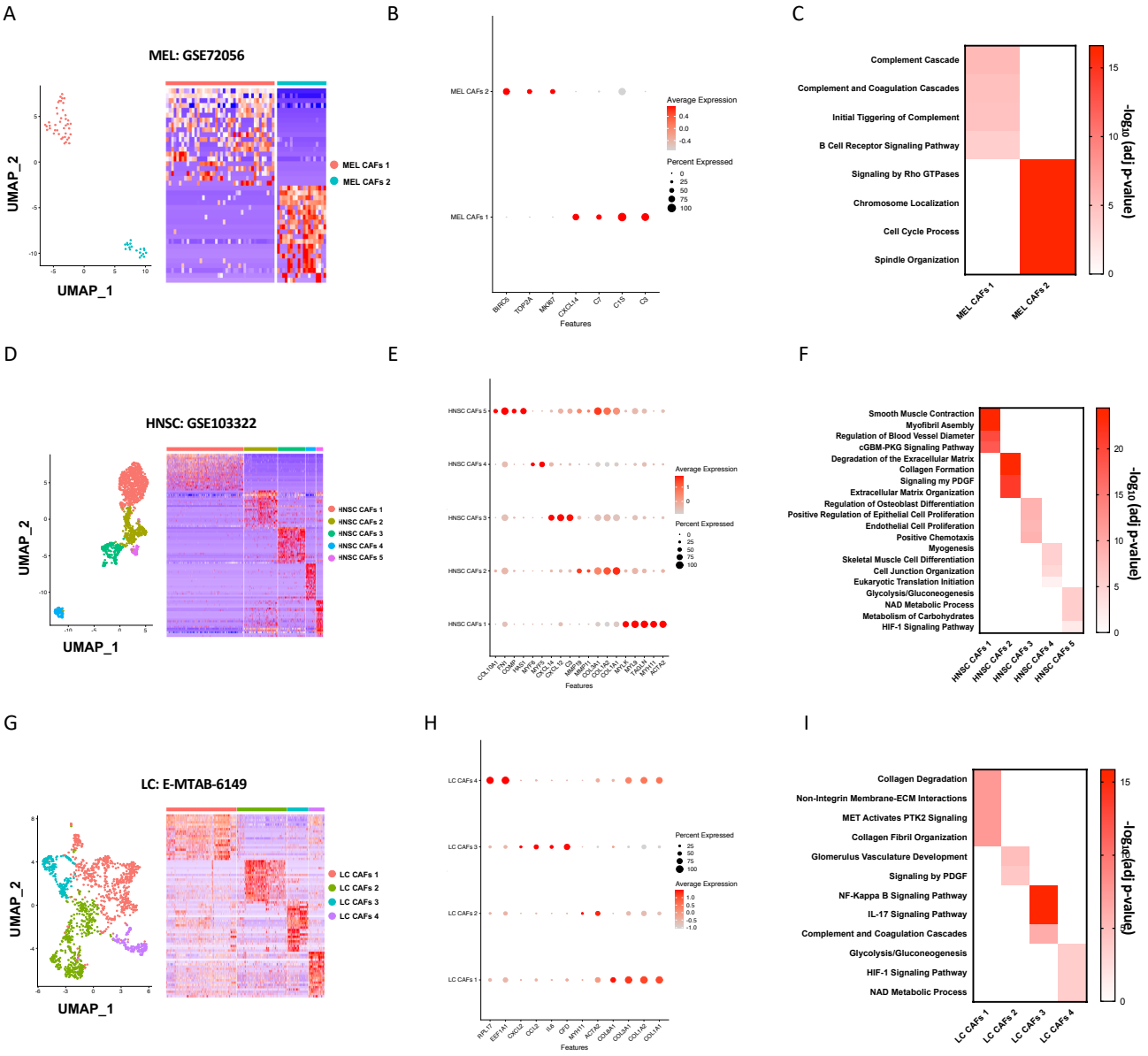
blue) and were clustered far away from other HNSC CAF sub-clusters (Supplementary Figure 1D). They displayed elevated expression of myogenic related genes (*MYF5* and *MYF6*) (Supplementary Figure 1E), and genes expressed significantly higher in them were enriched for myogenesis GO terms (Supplementary Figure 1F). Thus, we classified this cluster as normal myofibroblast (nmyCAF), which corresponded closely to Puram *et al*'s intermediate (resting) fibroblast cluster, as determined by the lack of expression of genes related to both myosin light-chain proteins and extracellular matrix (3). The fifth CAF sub-cluster for HNSC in our analysis comprised of 64 cells (4.5%, purple) (Supplementary Figure 1D) and had elevated expression of ECM related genes (*FN1*, *COL10A1*, etc.) (Supplementary Figure 1E). Its marker genes were enriched for GO terms associated with metabolism (Supplementary Figure 1F), thus we classified this cluster as metabolic CAF (mCAF). This cluster closely resembles Puram *et al*'s CAFs, because they highly expressed genes for extracellular matrix (3).

For LC, Lambrechts *et al* identified several distinct types of fibroblasts with unique gene expression patterns of collagens and extracellular matrix molecules (4). We identified four distinct sub-populations (LC CAFs 1-4) (Supplementary Figure 1G). Our LC CAFs 1 comprised of 623 cells (44.8%, red) (Supplementary Figure 1G) and was classified as dCAFs based on elevated expression of collagen genes (*COL1A1*, *COL1A2*, *COL3A1*, and *COL8A1*) (Supplementary Figure 1H). GO analysis of its marker genes found enrichment for collagen degradation and collagen fibril organization (Supplementary Figure 1I). Our LC CAFs 2 closely resembled the CAFs involved in myogenesis described by Lambrechts *et al*. It comprised of 441 cells (31.7%, dark green) (Supplementary Figure 1G) and was classified as myCAFs, based on elevated expression of activated fibroblast marker, *ACTA2*, and smooth muscle cell marker *MYH11* (Supplementary Figure 1H). This is supported by GO analysis that indicated enrichment for glomerulus vasculature development and signaling by PDGF (Supplementary Figure 1I). LC CAFs 3 comprised of 186 cells (13.4%, blue) (Supplementary Figure 1G) and was classified as iCAFs, based on elevated expression of inflammatory factors including complement factor *CFD* and cytokines *IL-6*, *CCL2*, and *CXCL2* (Supplementary Figure 1H). In addition, GO analysis showed enrichment for inflammation (Supplementary Figure 1I). Our LC CAFs 4 comprised of 140 cells (10.1%, purple) (Supplementary Figure 1G), highly expressed several genes related to translation (*EEF1A1* and *RPL17*) (Supplementary Figure 1H), and its marker genes were enriched for metabolism (Supplementary Figure 1I), thus we classified it as mCAF. Our LC CAFs 1 and 4 were similar to Lambrechts *et al*'s tumor-derived fibroblast populations because they highly express CAF related collagens (i.e. collagens type I, III, and V) (4). LC CAFs 2 closely resembled Lambrechts *et al*'s myofibroblast population due to high expression of myogenesis gene (4). Lastly, LC CAFs 3 was likely non-tumor derived fibroblast in Lambrechts *et al* because the lack expression of CAF related collagens (4).

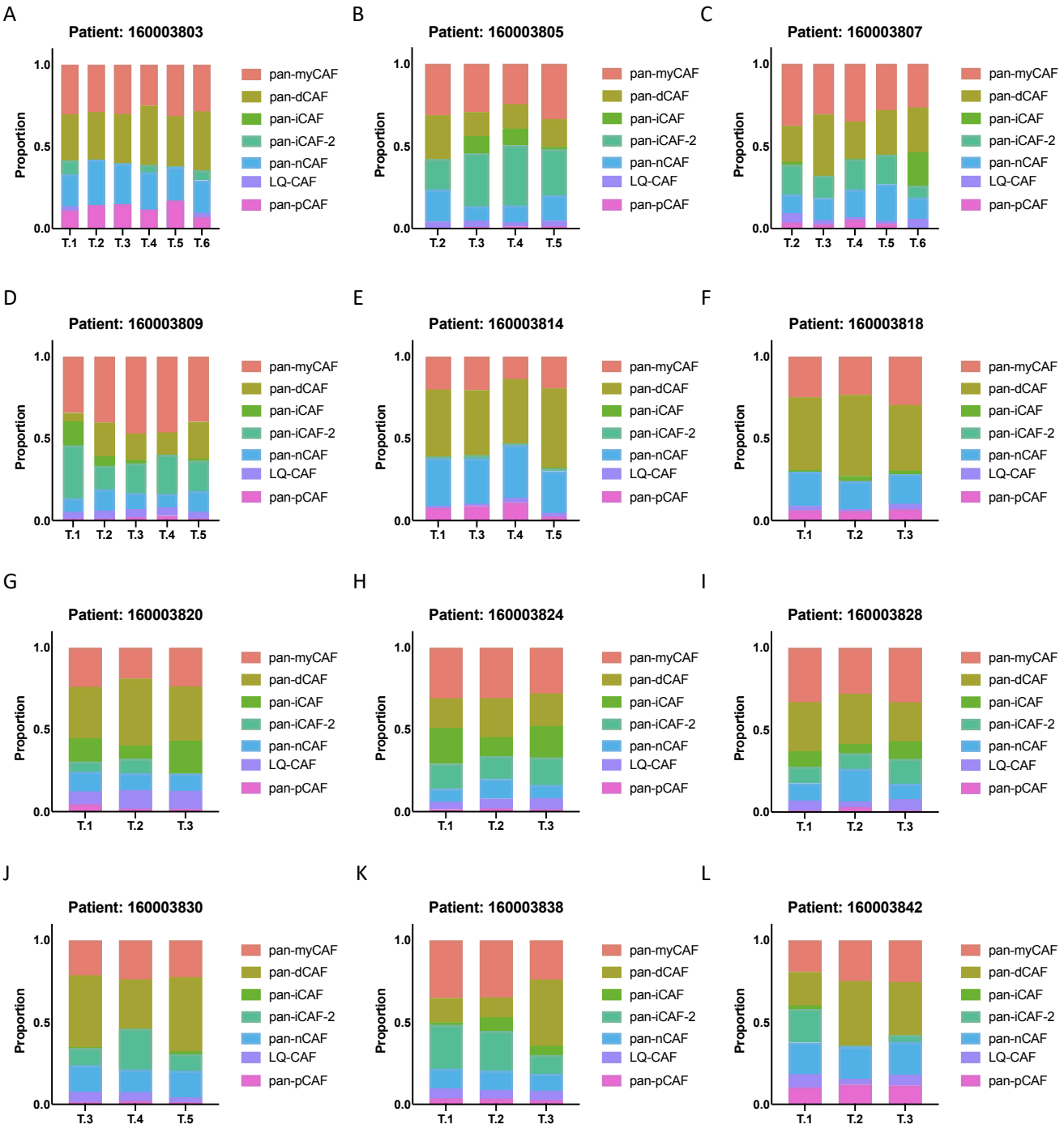
Collectively, these results recapitulate the previous CAF classifications but provide more systematic and consistent transcriptomic characteristics and molecular function of CAF subtypes in MEL, HNSC, and LC cancer types, as well as set up a strong foundation for our subsequent studies.

**References:**

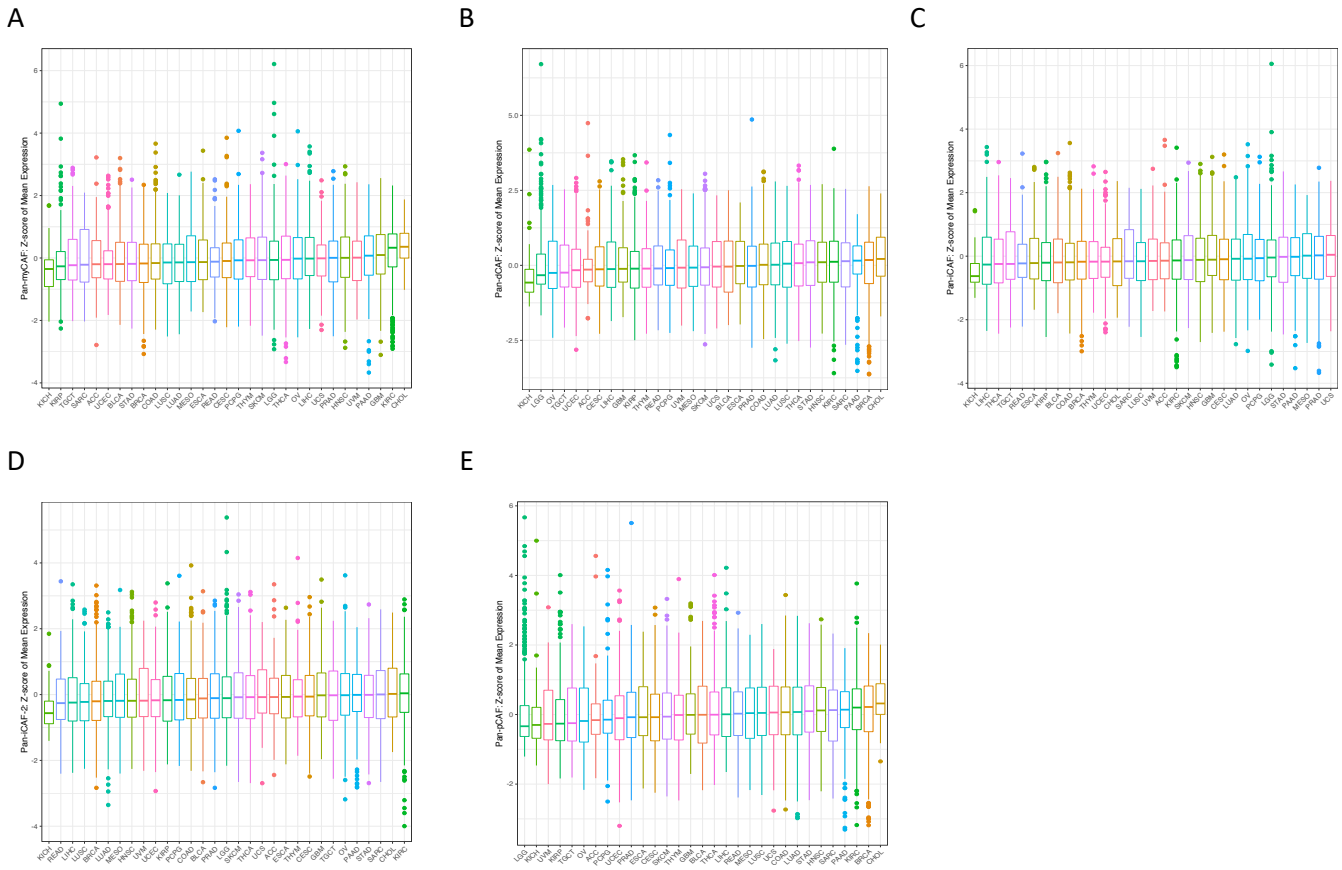

1. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **2018**;36:411-420.
2. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **2016**;352:189-196
3. 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: 1611-1624.
4. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, *et al*. Phenotype molding stromal cells in the lung tumor microenvironment. *Nat Med* **2018**;24:1277-1289.



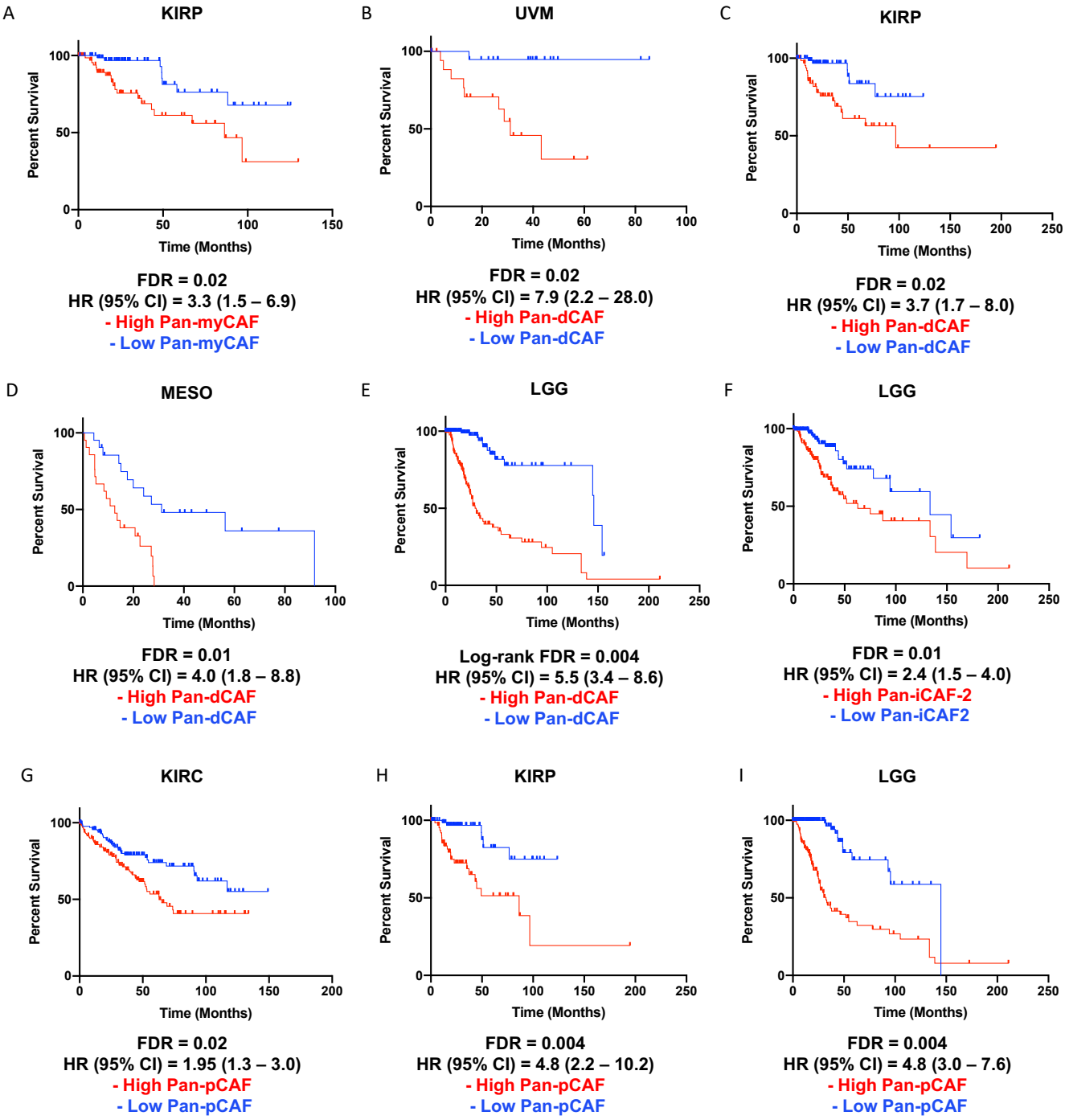

**Supplementary Figure S1. CAF heterogeneity in individual cancer types and their associated marker genes and biological processes. (A)** UMAP (left) depicting CAF subtypes and heatmap (right) depicting marker genes associated with CAF subtypes in melanoma tumors. **(B)** Dot plot of marker gene expression for CAF subtypes in melanoma. **(C)** Enriched gene sets associated with melanoma CAFs. **(D)** UMAP depicting CAF subtypes (left) and heatmap (right) depicting marker genes associated with CAF subtypes in head and neck squamous cell carcinoma tumors. **(E)** Dot plot of marker gene expression for CAF subtypes in head and neck squamous cell carcinoma. **(F)** Enriched gene sets associated with head and neck squamous cell carcinoma CAFs. **(G)** UMAP depicting CAF subtypes (left) and heatmap (right) depicting marker genes associated with CAF subtypes in Lung cancer. **(H)** Dot plot of marker gene expression for CAF subtypes in lung cancer. **(I)** Enriched gene sets associated with identified lung cancer CAFs.




**Supplementary Figure S2. Relative pan-CAF subtype abundances estimated for different tumor topological sections from 12 patients diagnosed with non-small cell lung cancer (NSCLC).**




**Supplementary Figure S3. Distribution of pan-CAF subtype specific gene expression profiles across the TCGA database.** Box plots displaying change in Z-score of mean expression of genes associated with pan-CAF subtype signatures across 31 cancer types available in the TCGA: **(A)** pan-myCAF, **(B)** pan-dCAF, **(C)** pan-iCAF, **(D)** pan-iCAF-2, **(E)** pan-pCAF.




**Supplementary Figure S4. Pan-CAFs are related to a poor clinical outcome in several distinct cancer types.** Kaplan-Meier plots depicting the survival differences among patients with high and low pan-myCAF in **(A)** KIRP. Kaplan-Meier plots depicting the survival differences among patients with high and low pan-dCAF in **(B)** KIRP, **(C)** LGG, **(D)** UVM, **(E)** MESO. Kaplan-Meier plots depicting the survival differences among patients with high and low pan-iCAF-2 in **(F)** LGG. Kaplan-Meier plots depicting the survival differences among patients with high and low pan-pCAF in **(G)** KIRC, **(H)** KIRP, **(I)** LGG.

**Supplementary Tables**

**Supplementary Table 1**. The relationship between CAF subtypes identified form individual cancer types to pan-CAF subtypes identified upon integrated analysis.

**Supplementary Table 2**. Robustness of pan-CAF clustering. Cells were randomly samples from the total number of CAF cells at the indicated percentage (90%, 80%, 75%) and subject to the same clustering analysis used for clustering all cells. The memberships of cells in the subsampling clustering and full data clustering were compared and summed in this table.

**Supplementary Table 3**. Signature genes for individual pan-CAF subtypes

**Supplementary Table 4**. Multivariate cox proportional hazard modeling analysis of individual pan-CAF gen signatures while adjusting for age and gender factors.