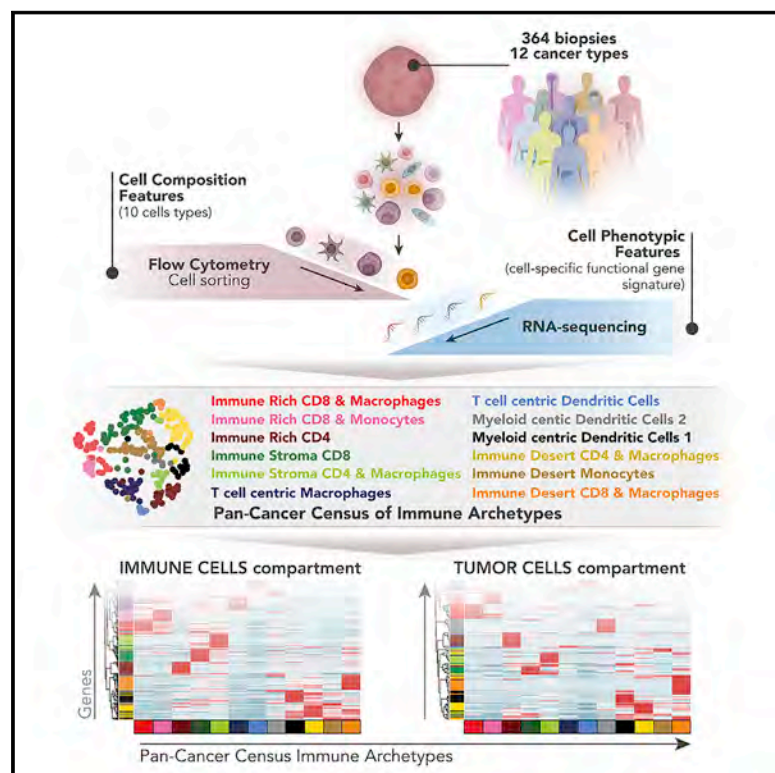# Discovering dominant tumor immune archetypes in a pan-cancer census

Graphical abstract

Authors

Alexis J. Combes, Bushra Samad, Jessica Tsui, ..., Saurabh Asthana, Vincent Chan, Matthew F. Krummel

Correspondence

alexis.combes@ucsf.edu (A.J.C.), matthew.krummel@ucsf.edu (M.F.K.)

In brief

Tumors can be extremely heterogenous across multiple parameters. By identifying and classifying recurrent immune features across 12 different tumor types, a framework is provided toward understanding tumor immunity as well as identifying broad and common therapeutic targets.

Highlights

- UCSF Immunoprofiler Initiative is a set of human tumors with multimodal linked data

- Clustering upon 10 features identifies 12 unique tumor archetypes spanning cancer types

- Each archetype concentrates similarities in additional immune and tumor features

- Dominant archetypes aid in tumor classification and identifying therapeutic targets

CellPress

Cell

Resource

Discovering dominant tumor immune archetypes in a pan-cancer census

Alexis J. Combes,[1,2,3,4,19,*] Bushra Samad,[1,2,3,4,19] Jessica Tsui,[1,2,3,4] Nayvin W. Chew,[1,2,3,4] Peter Yan,[1,2,3,4] Gabriella C. Reeder,[1,2,3,4] Divyashree Kushnoor,[1,2,3,4] Alan Shen,[1,2,3,4] Brittany Davidson,[2,3,4] Andrea J. Barczak,[3,4] Michael Adkisson,[3,4] Austin Edwards,[1,2,3,4] Mohammad Naser,[1,2,3,4] Kevin C. Barry,[5] Tristan Courau,[1,2,3,4] Taymour Hammoudi,[1,2,3,4] Rafael J. Argüello,[6] Arjun Arkal Rao,[1,2,3,4] Adam B. Olshen,[3,7,8] The Immunoprofiler Consortium, Cathy Cai,[1,2,3,4] Jenny Zhan,[1,2,3,4] Katelyn C. Davis,[1,2,3,4] Robin K. Kelley,[7] Jocelyn S. Chapman,[3,9] Chloe E. Atreya,[3,7,10] Amar Patel,[3,10] Adil I. Daud,[10,11] Patrick Ha,[12] Aaron A. Diaz,[13] Johannes R. Kratz,[2,14] Eric A. Collisson,[3,7,10] Gabriela K. Fragiadakis,[2,3,4,15] David J. Erle,[2,3,4,16] Alexandre Boissonnas,[17] Saurabh Asthana,[3,8] Vincent Chan,[2,3,18] and Matthew F. Krummel[1,2,3,20,21,*]

[1]Department of Pathology, University of California, San Francisco, San Francisco, CA 94143, USA
[2]ImmunoX Initiative, University of California, San Francisco, San Francisco, CA 94143, USA
[3]UCSF Immunoprofiler Initiative, University of California, San Francisco, San Francisco, CA 94143, USA
[4]UCSF CoLabs, University of California, San Francisco, San Francisco, CA 94143, USA
[5]Translational Research Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[6]Aix Marseille University, CNRS, INSERM, CIML, Centre d'Immunologie de Marseille-Luminy, Marseille, France
[7]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA
[8]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94143, USA
[9]Departments of Obstetrics, Gynecology, and Reproductive Sciences, Medicine, University of California, San Francisco, San Francisco, CA 94143, USA
[10]Department of Medicine, Division of Hematology and Oncology, University of California, San Francisco, San Francisco, CA 94143, USA
[11]Department of Dermatology, University of California, San Francisco, San Francisco, CA 94143, USA
[12]Department of Otolaryngology, University of California, San Francisco, San Francisco, CA 94143, USA
[13]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA 94143, USA
[14]Department of Surgery, University of California, San Francisco, San Francisco, CA 94143, USA
[15]Department of Medicine Division of Rheumatology, University of California, San Francisco, San Francisco, CA 94143, USA
[16]Lung Biology Center, Department of Medicine and the Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA 94143, USA
[17]Sorbonne Université, INSERM, CNRS, Centre d'Immunologie et des Maladies Infectieuses-CIMI, Paris, France
[18]Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA 94143, USA
[19]These authors contributed equally
[20]Senior author
[21]Lead contact
*Correspondence: alexis.combes@ucsf.edu (A.J.C.), matthew.krummel@ucsf.edu (M.F.K.)
https://doi.org/10.1016/j.cell.2021.12.004

SUMMARY

Cancers display significant heterogeneity with respect to tissue of origin, driver mutations, and other features of the surrounding tissue. It is likely that individual tumors engage common patterns of the immune system—here "archetypes"—creating prototypical non-destructive tumor immune microenvironments (TMEs) and modulating tumor-targeting. To discover the dominant immune system archetypes, the University of California, San Francisco (UCSF) Immunoprofiler Initiative (IPI) processed 364 individual tumors across 12 cancer types using standardized protocols. Computational clustering of flow cytometry and transcriptomic data obtained from cell sub-compartments uncovered dominant patterns of immune composition across cancers. These archetypes were profound insofar as they also differentiated tumors based upon unique immune and tumor gene-expression patterns. They also partitioned well-established classifications of tumor biology. The IPI resource provides a template for understanding cancer immunity as a collection of dominant patterns of immune organization and provides a rational path forward to learn how to modulate these to improve therapy.

INTRODUCTION

Pathologists have long recognized that tumors are infiltrated by cells of both the innate and adaptive arms of the immune system and thereby mirror inflammatory conditions arising in non-neoplastic environments (Dvorak, 1986). Indeed, tumors are complex environments where malignant cells interact with both immune and nonimmune cells to form the complex cellular network of the tumor microenvironment (TME) (Hanahan and Weinberg, 2011). Cancer immunotherapy has revolutionized
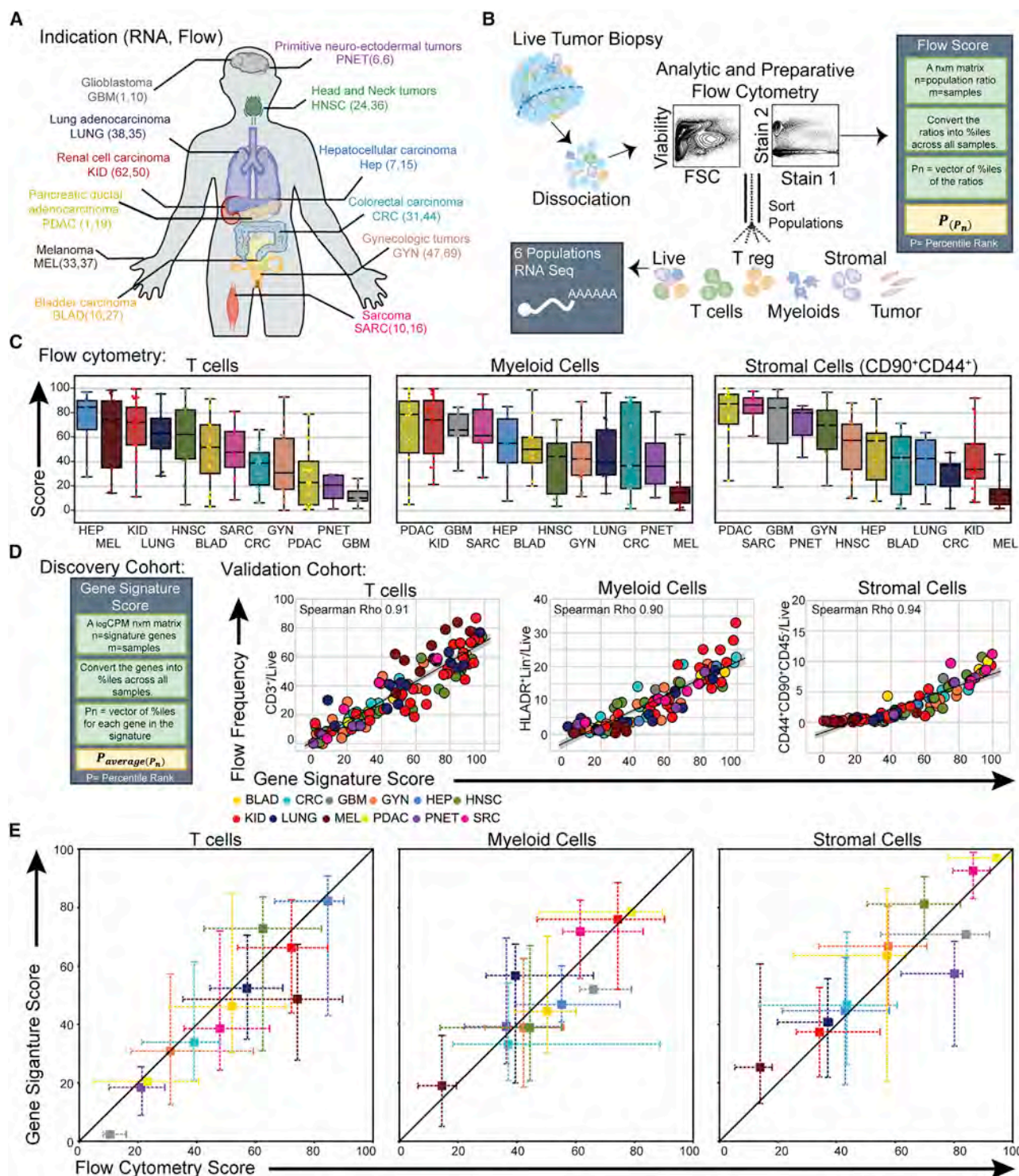
Figure 1. Generation and validation steps of T cells, myeloid cells, and stromal cells features from solid tumors using flow cytometry and bulk RNA-seq

(A) Details of the Immunoprofiler initiative (IPI) cohort tumor samples collection, color-coded by anatomical region and annotated with case numbers of bulk-RNA sequencing of viable cells sorted from fresh surgical tumor specimens and total number of samples with flow cytometry data.

(B) Description of the processing pipeline for digesting fresh tumor specimens into single cell suspension, submitting to multi-parametric flow cytometry for immune phenotyping and cell sorting into six different cell population compartments (live [viable cells], Tcell [conventional T cells], Treg [T regulatory cells], myeloid [myeloid cells], stroma [CD90$^+$ CD44$^+$ stromal cells], and tumor [tumor cells]).

(legend continued on next page)

cancer care by acting directly on the TME and re-engaging the anti-tumor immune response (Iwai et al., 2005; Leach et al., 1996). However, the biology of the immune microenvironment opposing these therapies is incompletely understood (Hugo et al., 2016; Spranger, 2016), and many patients experienced minimal or no clinical benefit from immunotherapies. A deeper understanding of the diversity of the immune microenvironment across human malignancies is critical to the improvement of immunotherapy treatment strategies (Binnewies et al., 2018; Gajewski et al., 2013; Gotwals et al., 2017; Hegde et al., 2016).

To date, the combination of The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Network, 2015) and deconvolution technique base on gene expression from bulk tissue such as CIBERSORT (Aran et al., 2017; Chen et al., 2018; Newman et al., 2015) have established a foundational but low resolution landscape of the TME across human tumors (Bindea et al., 2013; Gentles et al., 2015; Mlecnik et al., 2016; Rooney et al., 2015; Thorsson et al., 2018). More recently, single-cell RNA sequencing (scRNA-seq) technologies have been increasingly applied to define the diversity of TME in a single cancer type (Azizi et al., 2018; Goswami et al., 2020; Lavin et al., 2017; Zhang et al., 2019) or focus on a single immune cell compartment (Cheng et al., 2021; Gueguen et al., 2021; Oh et al., 2020; Zhang et al., 2020; Zilionis et al., 2019). The identification of recurrent motifs in the immune system at a higher resolution and spanning multiple cancer types is still lacking.

It is now well understood that complex coordination of immune cell states is required to achieve important tissue functions such as wound healing, tissue homeostasis, or viral clearance (Mujal and Krummel, 2019). Rather than focusing on the state of a single cell type, a given immune response can be conceived as a collection of cell subsets and specific immune cell pairings linked with function allowing us to define "immune archetypes." For response to therapy, a strong archetypal relationship between immune cell subsets such as cDC1, CD8 T cells, and natural killer (NK) cells (Barry et al., 2018; Böttcher et al., 2018) or cDC2 and CD4 conventional and regulatory T cells (Binnewies et al., 2019; Bosteels et al., 2020) has been previously identified in specific tumor types and viral infection. Tumors have also been described as "hot" (immune infiltrated) or "cold" (sparsely infiltrated) with variations in stromal content (Bagaev et al., 2021). However, these are likely incomplete delineations, and whether they are associated with tumor biology and granular immune cell composition is still unclear (Chen and Mellman, 2017).

In this study, we leveraged a unique dataset composed of both cell type compositional and transcriptomic data from 364 fresh surgical specimens across 12 tumor types to identify conserved tumor immune archetypes. We used an unsupervised clustering approach based on tumor specific immune gene signatures, benchmarked against "ground truth" cell-type composition

data from flow cytometry, to identify and validate 12 unique tumor immune archetypes that were also identified in the TCGA dataset. These archetypes, discovered with only ten features, differentially aggregated other cell types that were not part of the discovery features, and delineated immune and tumor transcriptomic programs across different tissue types. They and these data thus provide an unprecedented resource to study cancer immunity and cancer targets to improve response to immunotherapy.

RESULTS

A pan-cancer high-dimensional study of dominant immune composition

The UCSF Immunoprofiler Initiative (IPI) collected fresh surgical specimens from 12 distinct tissues of origin using an unbiased approach (i.e., agnostic of tumor type, stage, and grade) (Figure 1A; Table S1). We performed standardized processing of 364 individual tumor specimens including rapid digestion into single-cell suspension for immune phenotyping using multi-parametric flow cytometry (flow panels, see STAR Methods). To identify patterns of gene expression within six broadly defined cell populations, we also performed cell sorting for bulk RNA-seq including compartments denoted: (1) "live": all viable cells at the time of sorting; (2) "T_{conv}": sorted conventional CD4$^+$ and CD8$^+$ T cells; (3) "T_{reg}": CD25$^+$ CD4$^+$ (enriched for regulatory) T cells; (4) "myeloid": lymphocyte-negative HLA-DR$^+$ (enriched for myeloid) cells; (5) "stromal": CD45$^-$ CD44$^+$Thy1$^+$ cells; and (6) "tumor": all other CD45$^-$ cells (Figures 1B and S1A, left; see STAR Methods for full sort descriptions and potential caveats). This approach was chosen over single-cell RNA-seq because greater read depths capture weaker transcripts and because this technology was an immature technology at the start of the program.

Because they are the basis for many previous single-cancer type studies, we initially focused upon the abundance of three major cell types in the TME by flow cytometry (hereafter "3-feature" or "3-f"), namely total α/β T cells (T cell feature), myeloid cells (myeloid feature), and non-immune CD45$^-$, CD44$^+$, and Thy1$^+$ stromal cells (CD44$^+$ CD90$^+$ stroma feature) (Figure S1A, left; Table S2). Consistent with previous descriptions (Bagaev et al., 2021; Mandal et al., 2016; Thorsson et al., 2018; Varn et al., 2017), the cell abundances for these three cell types, plotted as a score (Figure 1B; STAR Methods) for each tumor type show tremendous heterogeneity across and within tumors from different tissues, suggesting the need for a tumor classification that goes beyond tissue of origin (Figure 1C).

Taking advantage of a distinguishing feature of our cohort that has linked compositional flow cytometry and RNA-seq data, we discovered and then validated a set of gene signatures that

(C) Box and whisker plots of flow score for T cells (n = 200), mononuclear phagocytes (n = 159), and stromal cells (n = 121) based on population percent in tumor specimens measured by flow cytometry (see Table S2 for details by cancer type).

(D) Left: gene score calculation method (see STAR Methods). Right: correlation plots of Tcell, myeloid, and CD90$^+$ CD44$^+$ stroma gene signature scores, for each tumor specimens, against their corresponding flow score (Figure S1; STAR Methods protocol) color-coded according to the tumor types shown in (C).

(E) Cross-whisker plots comparing median Tcell, myeloid, and CD90$^+$ CD44$^+$ stroma gene signature scores by tumor type to the median flow score, color-coded according to the tumor types shown in (C), with the interquartile range on both axes.

See also Tables S1 and S2.
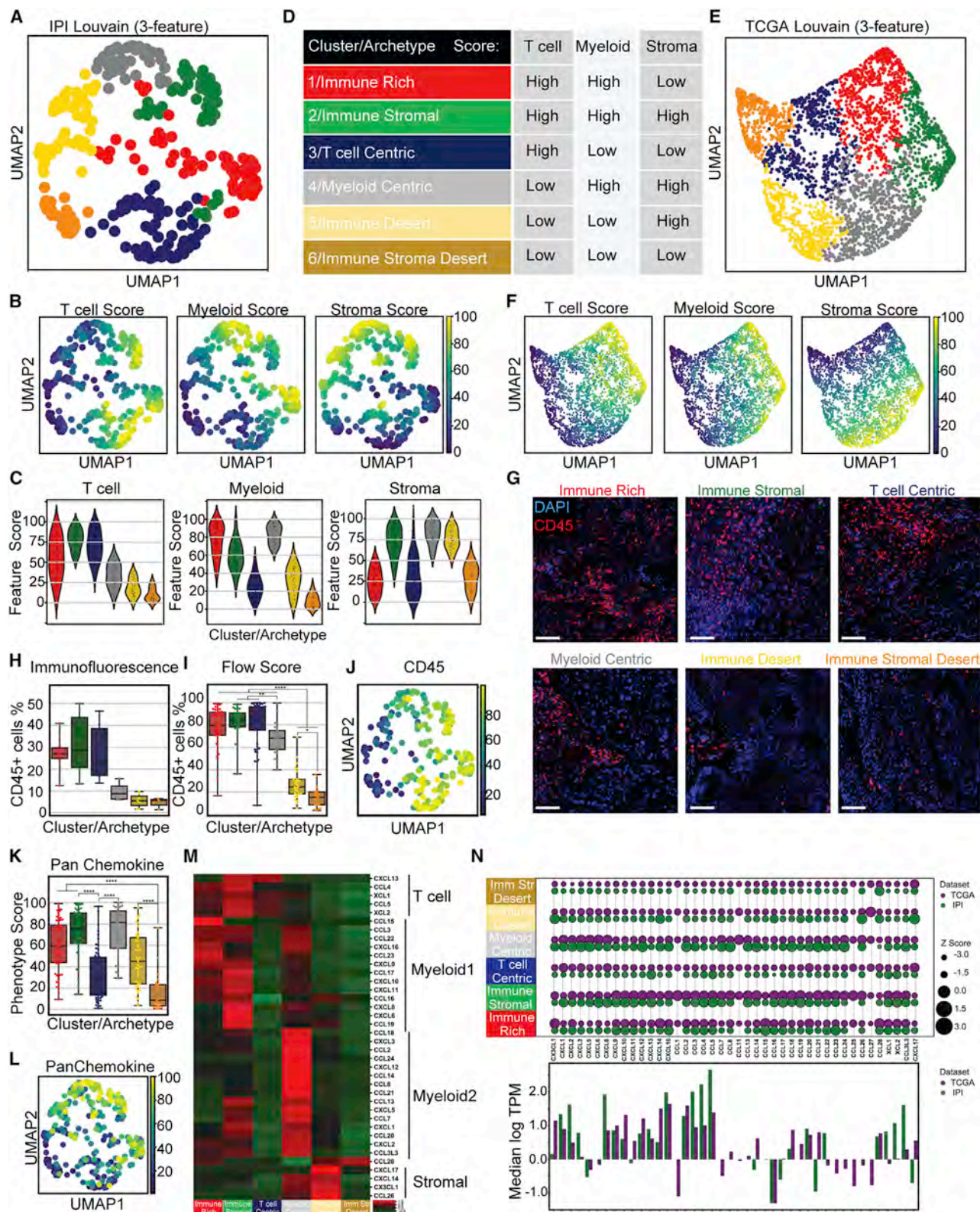
would allow us to infer and compare compositional data from external RNA-seq-only datasets such as TCGA. We used differential gene expression (DGE) analysis on high quality RNA-seq data from sorted tumor-associated T cell, myeloid, and stromal cells characterized by well-separated clusters of transcriptomic profiles revealed by K-means clustering (Figure S1A, right). We defined unique gene signatures for the T cell, myeloid, and CD44$^+$ CD90$^+$ stroma features, comprised of 25, 29, and 21 genes, respectively, and used this to generate a compositional score where a high value corresponded with high cell abundancy in the tumor specimen analyzed (Figures 1D and S1B). Use of this score, when applied to RNA-seq data from the live RNA-seq compartment in a validation cohort showed high concordance with the cell type frequencies obtained via flow cytometry, independent of tissue of origin (Spearman correlation 0.91, 0.90, and 0.94 of T cell, myeloid, and stromal, respectively) (Figure 1D). Moreover, rank ordering of the tumor from different tissues by these gene scores showed similar trends and strong correlations across individuals to those obtained by flow cytometric measurements (Figures 1C, 1E, and S1D).

Our signatures for intra-tumoral composition significantly outperformed signatures derived from cell lines (e.g., CIBERSORT) (Figure S1C) when compared to the "ground-truth" of flow cytometric composition data. We presume this improvement results from our data being taken from cells directly collected from fresh tumor tissue as opposed to cell lines or isolated samples (Aran et al., 2017; Newman et al., 2015). When we applied our gene signatures to RNA-seq data from 4,341 TCGA tumor specimens RNA-seq data, we found that the median score for each tumor type between datasets for the majority of the tissues surveyed was similar, suggesting that the abundancies we describe by tumor type (Figure 1C) extend beyond our sample processing protocol (Figure S1E). The exceptions to this were for lung (LUNG), liver (HEP), glioblastoma (GBM), and pancreatic (PDAC) tumors (Figure S1E), and this may be due to incomplete or different methods of sampling or patient-selection variation between the cohorts; thus, in the remainder of this analysis, we did not consider these cancer types when making additional comparisons.

Unsupervised clustering of 3 features independent of tissue of origin

We performed unsupervised clustering, using the Louvain community detection algorithm on a K nearest-neighbor (KNN)

weighted graph, on 3 features (T cell, myeloid, and CD90$^+$ CD44$^+$ stromal scores) for all 260 samples with a "live" bulk RNA-seq compartment (Table S2). The clustering was visualized using UMAP, both on the IPI and TCGA cohorts (Figures 2A and 2E; STAR Methods). The optimal clustering parameters were evaluated by minimizing the Davies-Bouldin index (DBI) (Davies and Bouldin, 1979), a metric that assesses the ratio of the intra-cluster distance to inter-cluster distance (Figures S2A and S2B; STAR Methods), and in both cohorts, this resulted in six clusters. In the IPI cohort, this separated two immune rich clusters (termed immune rich [IR] and immune stromal rich [IS]) defined by high expression of T cell and myeloid features and differentiated by stromal enrichment in IS. This similarly differentiated two immune desert (ID) clusters (ID and immune stromal desert) defined by low expression of T cell and myeloid features and again differentiated by enrichment of CD44$^+$ CD90$^+$ stroma within the ID cluster. Finally, we also identified clusters with enrichment in only one immune feature, namely the T cell centric (TC) and myeloid centric (MC) features (Figures 2B–2D and S2C, left). Application of our gene scores to the TCGA cohort also produced six clusters when minimizing the DBI, and these cluster corresponded to the six clusters found in the IPI cohort (Figures 2E, 2F, S2C, and S2D).

To provide validation of the flow cytometry data, we ran immunofluorescence assays on tissues collected from a subset of the IPI tumor surgical specimens. Consistent with expectation, we observed that samples taken from IR clusters (1–3) had the highest CD45$^+$ infiltration, whereas ID samples had the lowest (Figures 2G, 2H, and S2G–S2L). Myeloid centric (MC) clusters were similar to the ID clusters for overall CD45$^+$ infiltration as assessed by immunofluorescence (Figures 2G, 2H, and S2G–S2L). However, MC clusters presented an intermediate CD45$^+$ cells frequency when measured by flow cytometry. Overall, we observed significant correlation between the frequency of immune cells among all cells versus among viable cells after tissue dissociation (Figure S2F). The discrepancy between flow cytometry and immunofluorescence observed in MC archetype may represent a modest preferential recovery of immune cells during tissue dissociation for MC tumors. Future analyses of spatial dimensions may reveal paired infiltration patterns (i.e., T cell plus or minus myeloid or "excluded" versus "infiltrated") (Galon et al., 2006; Mlecnik et al., 2016). This will involve considerable orthogonal analyses due to the intrinsic complexity of diverse tissue morphologies (e.g., lung versus colon).

Figure 2. Identification of coarse immune archetypes in solid tumors using Louvain clustering on two independent datasets
(A and E) UMAP display using KNN and Louvain clustering of tumor immune archetypes using T cell, myeloid, and CD90$^+$ CD44$^+$ stroma features to cluster patients in the IPI (A) and TCGA (E) cohorts. Each dot represents a single patient.
(B and F) UMAP overlays of the Tcell, myeloid, and CD90$^+$ CD44$^+$ stroma features in the IPI (B) and TCGA (F) cohorts.
(C) Violin plots of Tcell, myeloid, and CD90$^+$ CD44$^+$ stroma features for each cluster/archetype in IPI cohort.
(D) Table summarizing the six cluster/archetypes with descriptions based on the level of the Tcell, myeloid, and CD90$^+$ CD44$^+$ stroma features.
(G and H) Representative Immunofluorescence of tumor specimens using CD45 (red) and DAPI (blue) staining for each cluster/archetype (G) and respective quantification of immune cell frequency (H). (G) Scale bars, 100 uM.
(I and J) Box and whisker plot (I) and UMAP overlay (J) of immune cell frequency using flow cytometry.
(K and L) Box and whisker plot (K) and UMAP overlay (L) of a pan chemokine phenotype gene signature score.
(M) Heatmap and hierarchical clustering of median chemokine gene expression per cluster/archetype identified in IPI cohort.
(N) Top: bubble plot of median chemokine gene expression by cluster/archetype identified in the IPI (green) and TCGA (violet) cohorts. Bottom: bar plots of median Log TPM gene expression of each chemokine in the IPI (green) and TCGA (violet) cohorts. The colors used correspond to the archetypes presented in (D).
*1.00e−02 < p ≤ 5.00e−02; **1.00e−03 < p ≤ 1.00e−02; ***1.00e−04 < p ≤ 1.00e−03. See also Figure S2 and Tables S1, S2, S4, and S5.
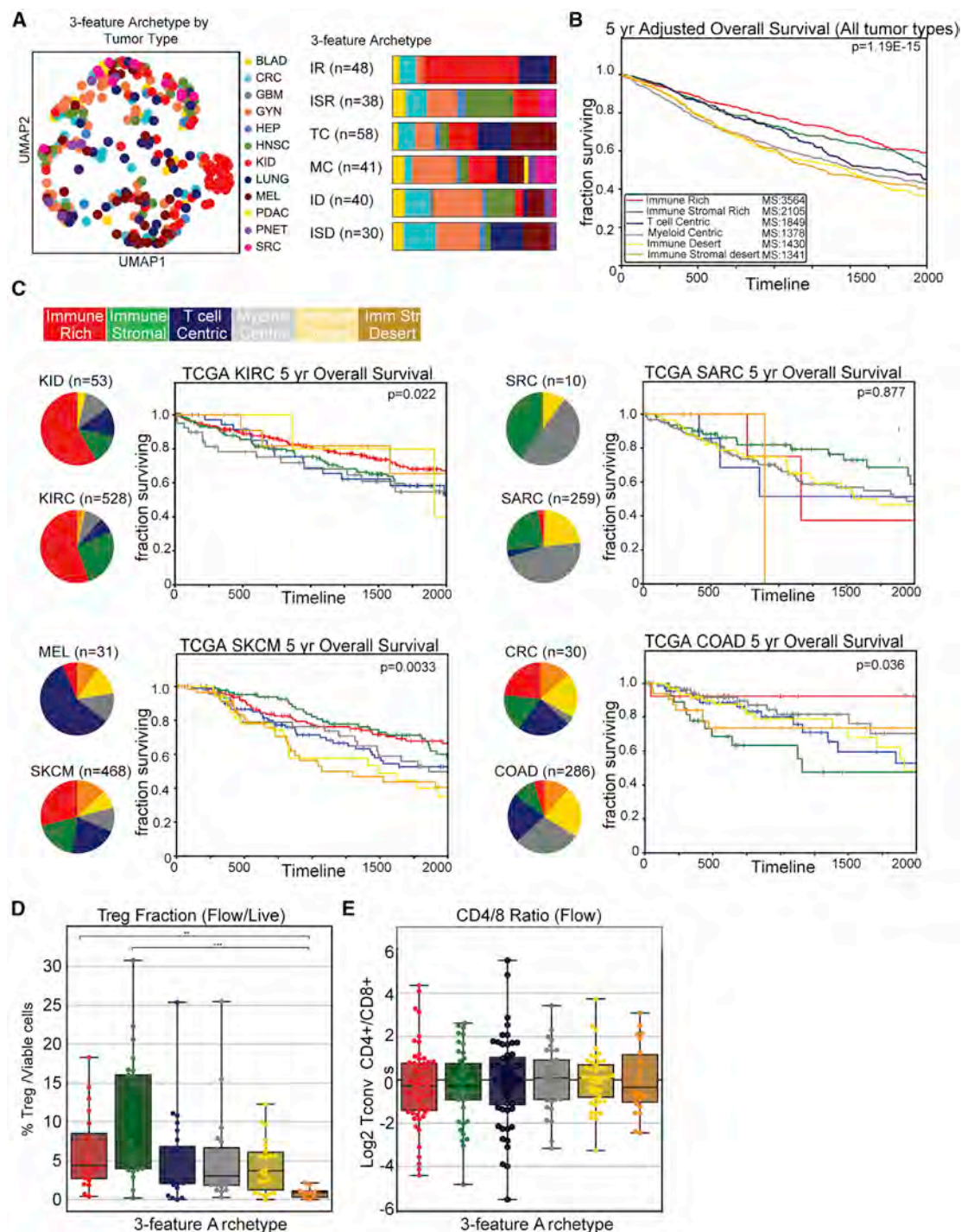
Figure 3. Coarse immune archetypes are independent of tissue origin and associated to overall survival

(A) Left: UMAP display, and graph-based clustering of 3-feature tumor immune archetypes color-coded by tumor type. Right: stacked bar plot of the tumor type distribution for 3-feature archetypes.

(B) Kaplan-Meier overall survival curves for each immune tumor archetype identified in the TCGA cohort.

(C) Left: pie charts representing distributions of each archetype by cancer type in the IPI (top) and TCGA (bottom) cohorts. Right: Kaplan-Meier overall survival curve for each immune tumor archetype identified in the TCGA cohort for kidney renal clear cell carcinoma (KIRC), skin cutaneous melanoma (SKCM), sarcoma (SARC), and colon adenocarcinoma (COAD).

(legend continued on next page)

Given the importance of chemokines in recruitment of immune cells, we sought to determine whether these archetypes had unique chemokine gene-expression signatures that might corroborate their classification. We first derived a single gene signature score applied to the live compartment, based on 39 chemokines (Nagarsheth et al., 2017) and found a high score in IR clusters and a low score in ID cluster in both the IPI and TCGA cohorts (Figures 2K, 2L, and S2E). Two exceptions stood out using the pan-chemokine measure: T cell and MC clusters showed decreased and increased pan-chemokine scores, respectively (Figures 2K and 2L). We thus examined chemokines individually, using hierarchical clustering of the median chemokine expression among the tumors of each archetype, and identified sets of chemokines associated with each cluster (Figures 2M and 2N). For instance, the three T cell-enriched clusters are enriched in either chemokines expressed by T cell (XCL1, XCL2, and CCL4) or T cell attracting chemokines (CXCL13 and CCL5). Other archetypes had their own pattern of differentially expressed chemokines, and these were broadly consistent with their composition. ID clusters typically had a few unique and specific chemokines expressed. The overall patterns observed on a chemokine-by-chemokine basis for the six archetypes was similar between IPI and TCGA (Figure 2N), with a few exceptions such as CCL1, CCL11, and CCL27. These results demonstrate that a 3-feature scoring of tumor tissue identifies six unique clusters defined by a different degree of immune infiltration and expression of distinct sets of chemokines.

Distribution of 3-feature archetypes by cancer type and outcome

3-feature-based clustering demonstrated a heterogenous distribution of tumor types among the different clusters, although some cancer types (e.g., kidney and melanoma) had significant biases (Figures 3A and S3A). Taking advantage of the large clinical dataset available in the TCGA cohort, we sought to broadly assess whether these simple archetypal classifications had a relationship to prognosis. Agnostic to the tissue of origin, the overall survival at 5 years analyzed by multivariate regression is significantly better for the IR archetype compared to all the others (p value $3.9E^{-11}$) with a general trend in outcome that tracks with overall immune infiltration (Figure 3B). To extend this analysis and focus within tumor types, we analyzed distributions of archetypes in both IPI and TCGA and assessed the outcome in individual cancer (Figures 3C and S3B–S3D). Although the relative composition of archetypes by tissue of origin was broadly similar in both cohorts, there was some variation (e.g., heavily "T cell centric" bias in melanoma in IPI cohort can appear as a combination of IR or IS in TCGA). Furthermore, archetypical classification in some tumor types (e.g., melanoma, kidney, bladder, and colorectal) appear to stratify survival outcome, and trends were weaker within other individual cancer type. This variability prompted us to consider that these coarse-

grained 3-feature archetypes may not capture the full range of immune cell heterogeneity in tumors. Measurement by flow cytometry of Treg density and CD4$^+$/CD8$^+$ conventional T cell ratio confirmed our assumption because we observed variability within each archetype (Figures 3D and 3E).
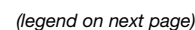
Developing a 6-feature archetype definition

Using flow cytometry data, we again grouped tumors by cancer type to assess variation in regulatory T cells (Treg) in live cells and CD4$^+$ or CD8$^+$ conventional T cell frequencies in T cells, both between and within tumor types (Figure 4A) (Table S3). Repeating our previous methodology (Figure S1), we used DGE between the Treg RNA-seq compartment and the other cell sorted RNA-seq compartments to generate Treg gene signature composed of 9 genes (Figure S4A) that was highly similar to other published signatures (Arce Vargas et al., 2018; Plitas et al., 2016; Zemmour et al., 2018). To isolate a signature for CD8 versus CD4 within our T$_{conv}$ RNA-seq compartment, we used the flow data to identify samples rich in CD4 or CD8 conventional T cells and performed DGE between them (Figure S4B; STAR Methods). Notably, most of the identified genes have been previously associated with CD4$^+$/CD8$^+$ conventional T cell identity (e.g., CD8A and IL7R) or tissue residency (e.g., VCAM1 and BACH2) (Richer et al., 2016; Savas et al., 2018) (Figure S4B). Assessment of a validation cohort (samples that were not used to generate the features scores) showed very high correlations (Spearman correlation Treg:0.86, CD4:0.97, and CD8:0.98, respectively), with population abundances measured by flow cytometry. Again, these correlations were significantly better than those obtained using CIBERSORT (Figure S4C). Unsupervised clustering of all samples using the CD4 and CD8 feature signature genes within our T$_{conv}$ RNA-seq revealed the existence of at least 3 distinct groups of tumors across cancer type: CD8-biased tumors, CD4-biased tumors, and a large population that was mixed for both these signatures (Figure 4B).

Using these scores, we next performed 6-feature clustering (Figures 4C and S4D–S4F). DBI optimization yielded eight clusters, and analysis by alluvial plot revealed that the two new clusters formed are mostly a subdivision of the previous coarse IR and IS archetypes, now delineated by CD4 to CD8 ratio (Figures 4D, 4E, and S4E). This increase in feature granularity also resulted in some specimens shifting within the classification. For example, some T cell centric samples now shifted to being considered "immune-rich:CD4" because of their profound high CD4 score, Furthermore, this analysis revealed that the 3-feature cluster of immune stromal desert was relatively CD8 rich whereas the ID favored CD4 cells. CD45 densities remain high for the IR tumors (Figure S4G). Assessing chemokine gene expression showed that this re-clustering dramatically segregated chemokine gene expression in the IR archetype and further refined the chemokine expression found among the other archetypes (Figures 4F and S4H).

(D) Box and whisker plot of CD4$^+$ regulatory T cells frequency in tumor measured by flow cytometry for each cluster/archetype identified in 3-feature archetypes.
(E) Box and whisker plot of log2 CD4$^+$ to CD8$^+$ conventional T cell frequency ratio in tumor measured by flow cytometry for each cluster/archetype in 3-feature archetypes.
*1.00e−02 < p ≤ 5.00e−02; **1.00e−03 < p ≤ 1.00e−02; ***1.00e−04 < p ≤ 1.00e−03. See also Figure S3.

A

T Regulatory cells

CD4 conventional T cells

CD8 T cells

B

Tumor associated T Conv cell RNAseq

— CD8 Biased — — CD4 Biased — Mixed CD4 and CD8 —

C

IPI Louvain (6-feature)

Cluster/Archetype Score:	T cell	Myeloid	Stroma	Treg	CD4 Conv	CD8
1/Immune Rich CD8	High	High	Low	Var	Low	High
2/Immune Rich CD4 Conv	High	High	Low	Var	High	Low
3/Immune Stromal CD8	High	High	High	High	Low	High
4/Immune Stromal CD4	High	High	High	High	High	Low
5/T cell centric	High	Low	Low	High	Var	Var
6/Myeloid centric	Low	High	High	Low	Var	Var
7/Immune Desert CD4	Low	Low	Var	Low	High	Low
8/Immune Desert CD8	Low	Low	Var	Low	Low	High

D

CD4-CD8 Ratio

E

3-feature → 6-feature

F

Chemokine Gene Expression

G

cDC2-cDC1 Ratio (Flow)

Mono-Mac Ratio (flow)

(legend on next page)

Mapping T cell exhaustion and myeloid subset heterogeneity in 6-feature archetypes

Exhaustion in T cells (T_{ex}) represents a transcriptional state for T cells that arises in cancers and chronic viral infections and is characterized by progressive loss of effector functions, high and sustained inhibitory receptor expression, and acquisition of a distinct transcriptional program (Blank et al., 2019). We used the T_{conv} RNA-seq compartment to identify genes that showed the highest correlation with CTLA4, PDCD1, HAVCR2, CD38, and LAG3, previously identified exhaustion markers. We identified 11 such genes, which included TOX, a transcription factor recently described as key driver of T cell exhaustion (Beltra et al., 2020; Khan et al., 2019; Scott et al., 2019) (Figure S4I; STAR Methods). We then benchmarked this gene signature score against the abundance of CD4$^+$, CD8$^+$, and the sum of both CD4$^+$ and CD8$^+$ T cells co-expressing CTLA4, PD-1, and CD38 as markers of T_{ex} (Figure S4J). This T_{ex} gene score best mirrored the flow-cytometry-based abundance of CTLA-4$^+$/PD-1$^+$ cells within the combined CD8$^+$ and CD4$^+$ compartments (Figure S4K). Using this T_{ex} gene signature score, we observed enrichment of exhaustion in CD8-biased archetypes including the ID archetype (Figures S4L and S4M) and together with low MHC I expression by the tumor cells in these archetypes (Figure S4N).

In addition to T cell exhaustion, we also sought to assess myeloid heterogeneity in our tumor landscape, widely known to be variable in tissues. We thus probed the abundance of mononuclear phagocytes (MNP) subsets including monocytes (Mo), macrophages (Mp), classical dendritic cells (cDC2, cDC1), and plasmacytoid dendritic cells (pDCs) for each archetype using flow cytometry (Figure S4O). This revealed that, despite slight enrichment between archetypes (e.g., of cDC2 over cDC1 in IR CD4 bias), the cDC1/cDC2 ratio, monocyte, and macrophage densities were highly variable within each of the 6-feature archetypes (Figures 4G and S4P).

10-feature archetype definition

The complexity of the MNP identity and plasticity has complicated efforts to determine which populations are beneficial or subversive to the anti-tumor response (Broz et al., 2014; Etzerodt et al., 2020; Gubin et al., 2018). To assess myeloid diversity and density across all samples, we generated a scRNA-seq substudy of sorted tumor-associated MNP to identify specific gene signatures for the principal MNP subsets (Figure S5A).

Following removal of non-MNP cellular contaminants, we found 5 unique clusters from 3,880 input cells (Figure S5B). Using DGE between each cluster, we identified unique 5 gene signatures for each subset (Figure S5C). Most of the genes identified have been previously associated with their respective MNP subsets such as VCAN, TREM2, CD1C, CLEC9A, and LAMP5 in monocytes, macrophages, cDC2, cDC1, and pDCs, respectively (Binnewies et al., 2018; Cheng et al., 2021; Combes et al., 2017; Molgora et al., 2020; Sancho et al., 2009). To validate these signatures across individuals and tissue types, we performed a similar analysis on 11 fresh resected tumors across 3 tissue types (melanoma, head and neck, and kidney) (Figure S5D). We confirmed that the same 5 gene signatures that we observed in our initial melanoma sample uniquely identified the MNP subsets independent of the tumor type (Figures S5D and S5E). A sixth cluster characterized by high expression of FSCN1, CCR7, and LAMP3 was also detected in this dataset (Figure S5E). This specific gene expression signature has been recently associated with a distinct cDC transcriptomic/activation state (provisionally named "cDC3") characterized by high expression of regulatory molecules such as PD-L1 and highly correlated with Treg abundance in the tumor (Cheng et al., 2021; Maier et al., 2020; Mulder et al., 2021; Zhang et al., 2019). However, due to the absence of specific markers in our antibody panel to validate the presence of this subset by flow cytometry, we could not include this population subset in our MNP features scores.

Gene signature scores for these MNP subsets were generated in bulk RNA-seq for tumor-associated myeloid cells and validated against cellular abundance by flow cytometry. The correlation of each MNP subset feature score across all cancer type were highly correlated (Spearman correlation of 0.86, 0.91, 0.93, 0.94, and 0.94 for monocytes, macrophages, cDC2, cDC1, and pDC, respectively) (Figure S5F; STAR Methods). Again, when we analyzed the distributions of these five MNP subset features scores by cancer type, a high heterogeneity was apparent (Figures 5A and S5G).

Thus, we repeated the unsupervised clustering after adding four MNP subset features including macrophages, monocytes, and the two types of classical dendritic cells to the previous six features (Table S3). The resulting clustering (hereafter "10-feature" or "10-f") produced 12 unique tumor immune archetypes after DBI minimization, with no bias toward specific archetypes found between scores calculated from flow and scores calculated directly from RNA-seq (Figures 5A and S5H–S5J). As

---

Figure 4. Inclusion of T cell subset features subdivide immune archetypes by CD4 to CD8 ratio

(A) Box and whisker plots of feature gene signature scores for CD4$^+$ regulatory T cells (Treg feature) out of the live compartment, and CD4$^+$ and CD8$^+$ conventional T cells (CD4 and CD8 features) out of the T cell compartment of patients in the IPI cohort.

(B) Heatmap and hierarchical clustering of CD4 (yellow) and CD8 (blue) feature genes' normalized expression, for patients in the T cell compartment.

(C) Left: UMAP display and graph-based clustering of tumor immune archetypes using T cell, myeloid, CD90$^+$ CD44$^+$ stroma, CD4, CD8, and Treg features to cluster patients in the IPI cohort. Each dot represents a single patient. Right: table summarizing the eight cluster/archetypes with descriptions based on the abundance of the Tcell, myeloid, CD90$^+$ CD44$^+$ stroma, CD4, CD8, and Treg features.
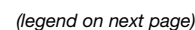
(D) Box and whisker plot of log2 CD4$^+$ to CD8$^+$ conventional T cell feature gene signature score ratio in tumor for each of the clusters/archetypes identified with 6-feature clustering.

(E) Alluvial plot depicting how cluster/archetype membership perpetuates or subdivides from 3- to 6-feature clustering.

(F) Heatmap and hierarchical clustering of the median chemokine gene expression for each cluster/archetype identified in the 6-feature clustering.

(G) Box and whisker plot of log2 cDC2 to cDC1 ratio (top) and monocyte to macrophage ratio (bottom) measured by flow cytometry for each cluster/archetype identified.

*1.00e−02 < p ≤ 5.00e−02; **1.00e−03 < p ≤ 1.00e−02; ***1.00e−04 < p ≤ 1.00e−03; ****p ≤ 1.00e−04. See also Figure S4 and Tables S4 and S5.

(legend on next page)

previously observed when adding T cell subsets, the addition of the four MNP measures subdivided preexisting immune archetypes but did not increase those by 16-fold as could occur if these were randomly assorted (Figures 5B and 5C). MNP inclusion also resulted in some specimens shifting between MC and ID archetypes, driven by strong monocyte or macrophage enrichment.

Analyzing the 12 archetypes produced in 10-f clustering, we found that generally both IR and ID sub-archetypes were distinguished by distinct pairings of T cell and myeloid subsets but that these pairings varied. Specifically, the monocyte/macrophage ratio demarcates two different IR CD8 bias archetypes where Treg abundance is generally higher in macrophage-enriched tumor specimen (Figures 5D–5F and 5O–5Q). One archetype, IR CD4 biased, is differentiated from its IR CD8 counterparts by enrichment in cDC2 versus cDC1 (Figures 5D and 5R–5T). However, such correlation between CD4/CD8 and cDC2/cDC1 ratios is only observed among IR archetypes and not in archetypes containing high stromal densities (Figures 5G–5I and 5R–5T).

The relationship between monocytes/macrophages and T cell subsets composition also differed significantly among the 12 archetypes (Figures 5D-5N). In IS (Figure 5H), CD8 abundance is opposed to macrophage abundance, but in ID, both CD4- and CD8-rich sub-archetypes are equivalently enriched in macrophages and monocytes (Figures 5N, 5P, and 5T). TC (T cell centric) archetypes are characterized by a positive correlation between Treg and macrophage abundance (Figures 5J–5L and 5P–5Q). This discordance between the abundances of macrophages and other immune cells in different archetypes could suggest that a generalized macrophage score does not capture critical heterogeneity in macrophage phenotypes or that macrophages and T cells may engage in additional as-yet-undiscovered relationships that regulate their numbers.

Defining immune gene expression pattern across 10-feature archetypes

Focusing on the 12 tumor archetypes identified by 10-f clustering, we examined associations with cell abundances and gene sets that were not used for clustering. Significant archetype-specific enrichment was found for gene sets that define intra-tumor NK cells in IR based on published signatures (Barry et al., 2018) and these were confirmed by flow cytometry measurement, while ISR were enriched in gene sets defining Mast cells (Cheng et al., 2021) (Figures 6A, S6B, and S6C). Plasma cells, defined by enrichments of IgG genes, were enriched in T cell centric Macrophage bias tumors whereas B cell measured

by proteomic and transcriptomic expression (MS4A1; CD20) were most prevalent in T cell centric DC rich (Chen et al., 2020; van Galen et al., 2019) (Figures 6A and S6D). These may correspond to a prevalence for tertiary lymph nodes in those archetypes although we did not observe higher DC frequencies in those tumor specimens (Figure S5J). An archetype-specific immune composition pattern again correlated with specific enrichment of groups of chemokine transcripts (Figure 6B). Notably, chemokines with specificity for families of chemokine receptors now frequently co-clustered within the 10-f archetypes.

We next used the myeloid RNA-seq compartment to explore the level of expression of genes previously associated with tumor-infiltrating MNP, namely "M1," "M2," "DC," and "co-stimulatory molecules" (Biswas et al., 2013; Cassetta et al., 2019; Cheng et al., 2021; Maier et al., 2020; Roberts et al., 2016). Despite the heterogeneity in individual gene expression associated with M1 and M2 macrophage phenotypes among archetypes, we observed a general enrichment in M2 genes in IS archetypes (Figure 6C). On the other hand, DC- and co-stimassociated genes were highly expressed in the T cell centric DC rich archetype but LAMP3, FCSN1, and CCR7 expression (typically indicative of the cDC3 state) did not appear coordinated across archetypes (Figure 6C). This may indicate heterogeneity of dendritic cell transcriptomic states between archetypes.

To further elucidate the heterogeneity found in myeloid function, we combined the myeloid "signature" gene sets with a selection of 12 other gene sets highly linked to subsets of Treg, T_{conv}, B cell, plasma cells, and NK cells and performed hierarchical clustering of these genes sets based on their median expression in the archetypes. The gene sets were evaluated in different RNA-seq compartments corresponding to their appropriate cell type, namely "Th1" genes in the T_{conv} RNA-seq compartment, M1 genes in the myeloid RNA-seq compartment, and "Treg homing" in the Treg RNA-seq compartment (Figure 6D; Table S3). Combining the gene sets in this way revealed distinct immune signatures for each archetype (Figure 6D). For instance, although the IR CD8 macrophage bias archetype was enriched in genes associated with the type 1 response (interferon gamma [IFNG], tumor necrosis factor [TNF], interleukin [IL]-1B) both in T_{conv} and MNP compartments, both IS and ID CD8 archetypes were characterized by their unique combination of upregulated gene expression associated with T cell exhaustion (PDCD1, CTLA4, and ENTPD1) together with M2-like macrophages (PDCD1L2, CD163, CD274, and MRC1). This analysis also identified putative gene expression interactions present in the same archetype across cell types, such as high CCR2 expression by

Figure 5. Single-cell RNA-seq-derived myeloid signatures refine immune archetypes
(A) Left: box and whisker plots for the macrophages, monocytes, cDC1, and cDC2 features, calculated in the myeloid compartment, from the IPI cohort. Right: UMAP display and graph-based clustering of tumor immune archetypes using Tcell, myeloid, CD90+ CD44+ stroma, Treg, CD4, CD8, macrophages, monocytes, cDC1, and cDC2 features to cluster patients in the IPI cohort.
(B) Alluvial plot depicting how cluster/archetype membership perpetuates or subdivides from 6- to 10-feature clustering.
(C) Left: schematic of a "phylogeny" of the cluster/archetypes as they progressed from 3-feature to 6-feature to 10-feature clustering.
(D–N) Scatterplots of different features defining the twelve clusters/archetypes identified in the IPI cohort.
(O and R) UMAP overlay of macrophages (macs) and monocytes (mono) (O) and classical dendritic cell type 1 (cDC1) and 2 (cDC2) feature scores (R) for each cluster/archetype identified by 10-feature clustering.
(P, Q, S, and T) Box and whisker plot of Ln mono to macs ratio (P), Treg feature gene score (Q), Ln cDC2 to cDC1 ratio (S), and Ln CD4 to CD8 conventional T cells ratio (T) for each cluster/archetype identified.
*$1.00e-02 < p \leq 5.00e-02$; **$1.00e-03 < p \leq 1.00e-02$; ***$1.00e-04 < p \leq 1.00e-03$; ****$p \leq 1.00e-04$. See also Figure S5 and Tables S3 and S4.
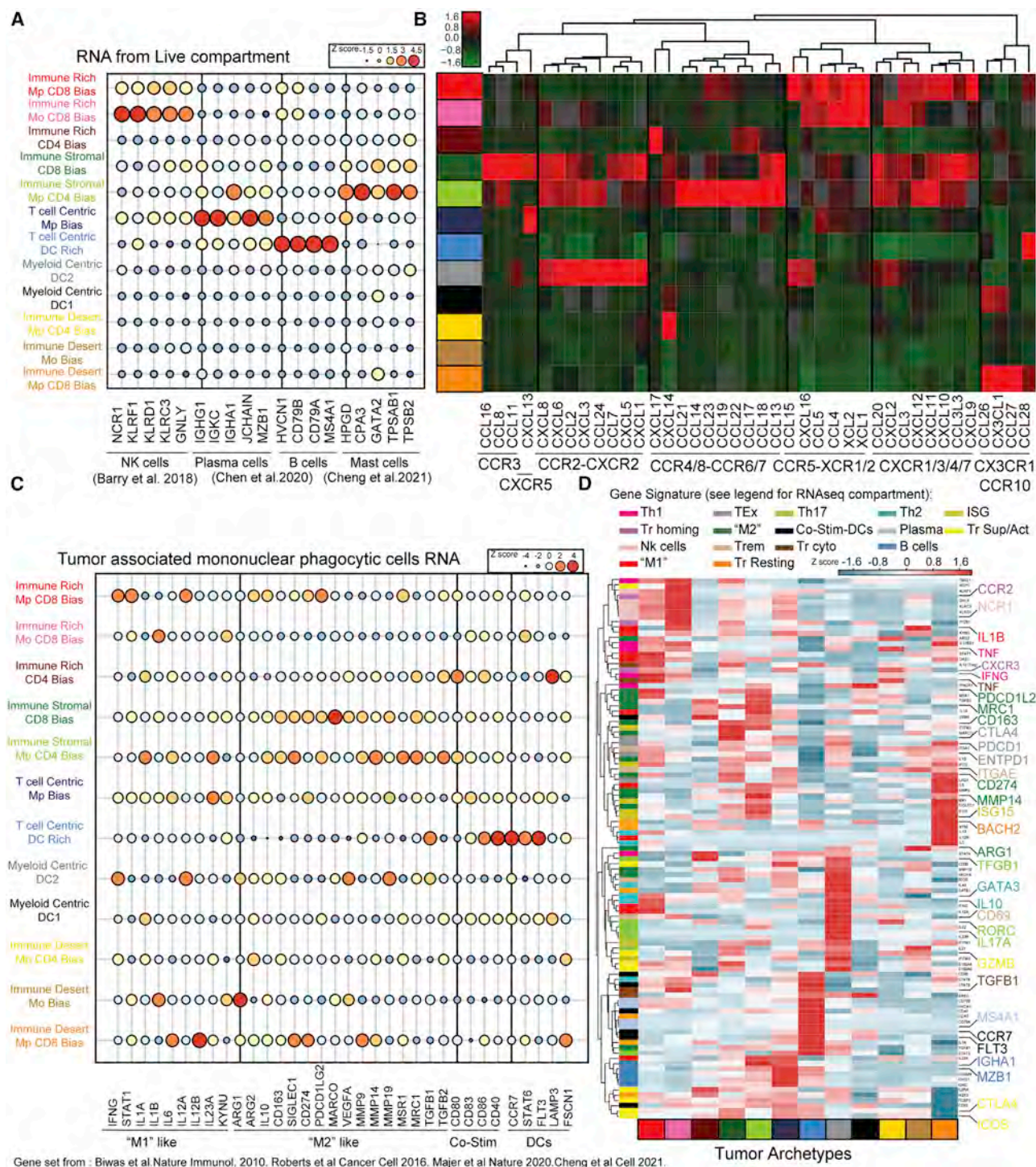
Figure 6. Each tumor archetype is defined by a unique combination of immune gene expression pattern

(A) Bubble plot of natural killer (NK) cells, B cells, plasma cells, and mast cells associated gene expression in the live compartment grouped by clusters/archetypes identified in the IPI cohort using 10-feature clustering.

(B) Heatmap and hierarchical clustering of the median chemokine gene expression of all chemokines in the chemokine phenotype signature, grouped by cluster/archetype.

(C) Bubble plot of gene expression in macrophages (M1 and M2) and dendritic cells (co-stim and DC) function in the myeloid compartment, grouped by cluster/archetypes in the IPI cohort.

(legend continued on next page)

Tregs in an archetype rich in monocytes, which are known producers of the CCR2 ligand, CCL2 (Loyher et al., 2018; Mondini et al., 2019) (Figure 6D). Taken together, this demonstrated that the 12 tumor immune archetypes found in 10-f analysis also identify a unique combination of cell composition and transcriptomic profiles.

Notably, certain immune populations remained more variable across the archetypes. For instance, neutrophils were mostly uncorrelated although there was a slight and statistically insignificant rise in neutrophil infiltration in archetypes with low immune infiltration (Figures S6A and S6E). A gene signature previously associated with stimulatory dendritic cells abundance and better outcome in melanoma patients is slightly enriched in archetypes with high total immune infiltration (Barry et al., 2018). Exhaustion signature is enriched in both IR and ID archetypes (Figures S6A, S6F, and S6G). Thus, although the 12 immune archetypes from 10-f analysis identify dominant combinations of cell types and transcriptomic profiles present in tumors, other cell types may also be independently layered within and across this apparently conserved biology.

10-feature tumor archetypes tie closely to tumor biology and disease outcome

Finally, we sought to examine the relationship of 10-f archetypes to the phenotype of the tumor cells themselves. As previously shown for the coarse archetypes, 10-f archetypes are diverse in their tissue of origin; however, some tumor types remain highly represented within an archetype (e.g., kidney, immune rich CD8; melanoma, T cell centric archetypes) (Figures 7A and S7). Regardless of the frequent intra-cancer heterogeneity, we observed a strong relationship between tumor proliferative capacity, measured by the fraction of tumor cells that were Ki67$^+$ as measured by flow, and ID and myeloid centric archetypes (Figure 7B).

Based on this, we assembled gene signatures for various aspects of tumor biology to determine whether immune features mapped to tumor biology. Expression of cell-cycle-associated genes indicated enhanced tumor cell proliferative capacity in ID archetypes, consistent with the Ki67 data (Figure 7C). Conversely, IR and ISR tumors were enriched in other transcriptomic programs such as interferon-stimulated genes (ISG), senescence, or epithelial-mesenchymal transition (EMT)-related genes (Kinker et al., 2020; Muñoz et al., 2019; Wiley et al., 2017), and only some IDs were highly enriched in fibrosis-associated genes (Figure 7C).

Without pre-selection of genes from previous literature, we performed DGE and then hierarchical clustering using the most differentially expressed genes between archetypes on the sorted tumor RNA-seq compartment (Figure 7D). Within the 12 sets of differential genes, we noted some discernable patterns, for example increased IFNG expression in tumor cells from the IR CD8 macrophage archetype, concordant with increased type 1 response in immune cells from the same archetype.

To define gene signatures that we could assess in TCGA, we queried our sorted live compartment to find genes that corre-
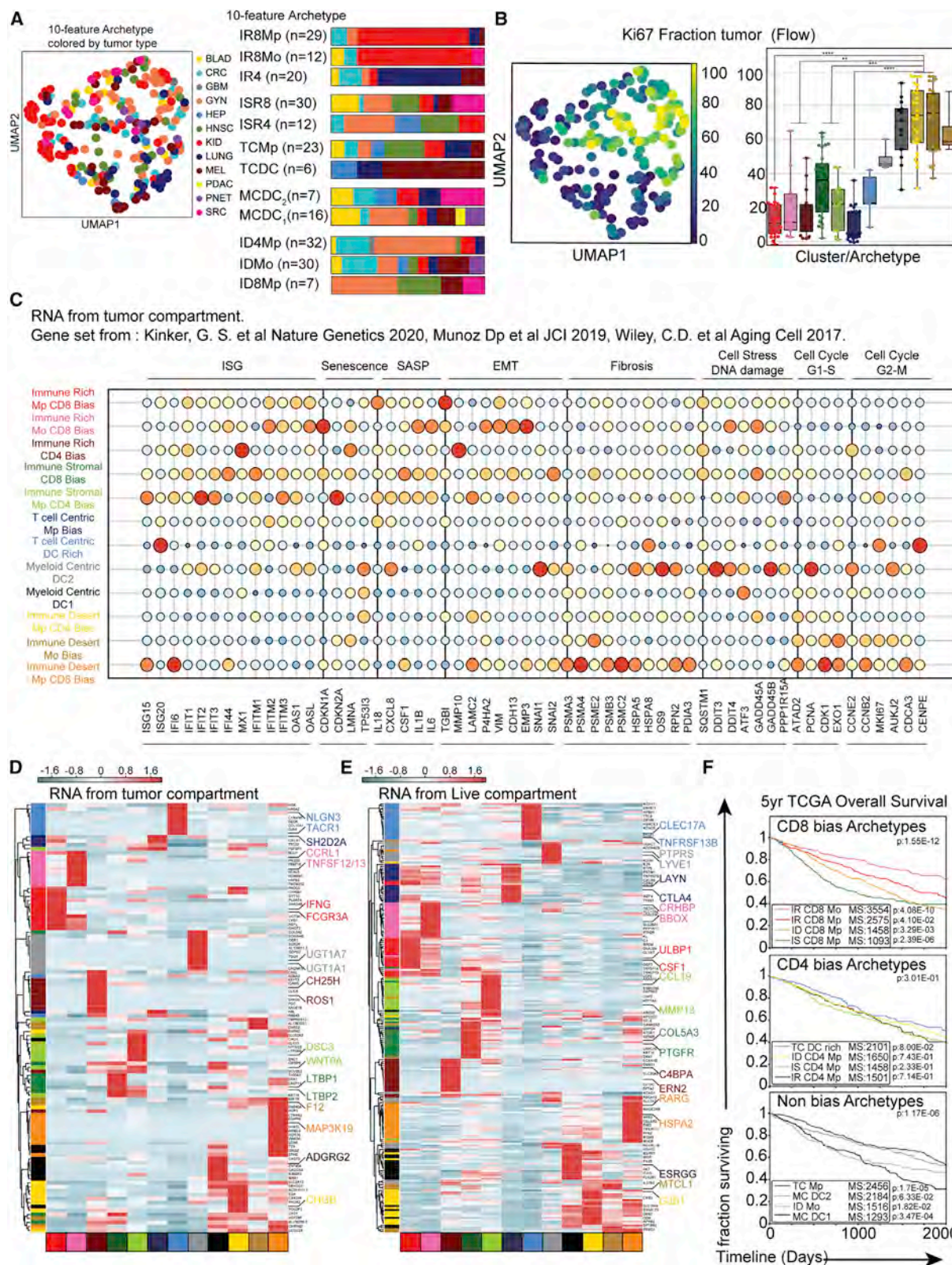
lated with each archetype (Figure 7E). This identified unique gene signatures for each of the 12 archetypes, within which we observed both immune-related genes (e.g., LAYN, CTLA4, CSF1, and CCL19) and non-immune-related genes (e.g., CRHBP, PTGFR, HSPA2, and MTCL1). Applying these signatures to the TCGA dataset, we found that the tumor archetypes identified in our IPI dataset were retrieved with a similar relative composition of archetypes by cancer type in KIRC, CRC, LUAD, HNSC, BLCA, or with a slight shift between archetypes from the same coarse archetype (i.e., ID CD4 bias to IR CD8 bias) in GYN and SKCM (Figures S7A–S7K). Additionally, we observed a similar distribution of archetypes in both cohorts when looking at clinical correlates such as tumor stage, grade, and metastatic status (Figures S7L–S7N). This suggests that our archetypes can be identified in independent datasets and in the predictive potential of these signatures for rapid classification of patients in both primary and metastatic tumors independent of tumor grade and stage.

Finally, survival analysis of the different tumor archetypes identified in the TCGA dataset for each cancer type separately showed that better outcomes may be cancer-type-specific, and each cancer type may have a different type of TME promoting the best immune response (Figures S7A–S7K). However, when survival analysis was performed across tumors using a multivariate survival regression, we detected significant outcome differences between archetypes that have similar T cell subset enrichment regardless of the tissue of origin (Figure 7F). For instance, in IR CD8 archetypes, the apparent enrichment in monocytes over macrophages (pink archetype versus red) is associated with better survival (median survival, 3,354 days) and more so when compared to ID and IS CD8 biased archetypes (p value 1.55E−12) (Figure 7F, top). Conversely, in archetypes with no significant bias between CD4 and CD8 T cells, TC archetypes characterized by an enrichment of macrophages over monocytes displayed the higher median survival (2,456 days, p value 1.17E−6), whereas in archetypes with CD4 T cell bias, no significant outcome difference between archetypes was detected regardless of myeloid biases (p value 3.01E−1) (Figure 7F, middle and bottom). This indicates that tumor archetypes provide a template to study different subtypes of anti-tumor immune response in a variety of primary human cancers. Finally, our study represents an initial framework—one certain to be refined—to understand how tumor archetypes relate to other biological contexts (Table S6) and how to best modulate these archetypes depending on their specific immune context.

DISCUSSION

In this study, we present a holistic survey of dominant immune archetypes across 12 cancers from different tissues using fresh surgical tumor specimens from 364 patients. Empowered by complementary profiling assays, we were able to discover 12

(D) Heatmap and hierarchical clustering of the median gene expression of B cells (all viable RNA-seq compartment), NK cells (all viable RNA-seq compartment), plasma cells (all viable RNA-seq compartment), T cells phenotypes (Tconv RNA-seq compartment), T regs (Treg RNA-seq compartment), macrophages, and dendritic cells function (myeloid RNA-seq compartment) grouped by cluster/archetype in IPI cohort.
See also Figure S6 and Tables S5 and S6.

A 10-feature Archetype colored by tumor type

10-feature Archetype
IR8Mp (n=29)
IR8Mo (n=12)
IR4 (n=20)
ISR8 (n=30)
ISR4 (n=12)
TCMp (n=23)
TCDC (n=6)
MCDC₂(n=7)
MCDC₁(n=16)
ID4Mp (n=32)
IDMo (n=30)
ID8Mp (n=7)

Tumor types: BLAD, CRC, GBM, GYN, HEP, HNSC, KID, LUNG, MEL, PDAC, PNET, SRC

B Ki67 Fraction tumor (Flow)

C RNA from tumor compartment.
Gene set from : Kinker, G. S. et al Nature Genetics 2020, Munoz Dp et al JCI 2019, Wiley, C.D. et al Aging Cell 2017.

D RNA from tumor compartment

E RNA from Live compartment

F 5yr TCGA Overall Survival

CD8 bias Archetypes p:1.55E-12
IR CD8 Mo MS:3554 p:4.08E-10
IR CD8 Mp MS:2575 p:4.10E-02
ID CD8 Mp MS:1458 p:3.29E-02
IS CD8 Mp MS:1093 p:2.39E-06

CD4 bias Archetypes p:3.01E-01
TC DC rich MS:2101 p:8.00E-02
ID CD4 Mp MS:1650 p:7.43E-01
IS CD4 Mp MS:1458 p:2.33E-01
IR CD4 Mp MS:1501 p:7.14E-01

Non bias Archetypes p:1.17E-06
TC Mp MS:2456 p:1.17E-05
MC DC2 MS:2184 p:6.33E-02
ID Mo MS:1516 p:1.82E-02
MC DC1 MS:1293 p:3.47E-04

(legend on next page)

distinct immune archetypes that span cancer types. Each archetype is made up of a unique combination of cell composition—some used to derive the cluster, but many that are learned from it—and immune and tumoral transcriptomic phenotypes.

Starting with just 10 independent cell compositional features, unsupervised clustering revealed only 12 distinct clusters whereas 10 independently assorting binary variables might have produced 2^{10} or 4,096. Although more work will need to be done to ensure that this is not due to a lack of sampling, this implies that some combinations of cell densities may not exist in the TME of solid tumors. This is partly in line with previous work that identified between four and six immune subtypes spanning multiple tumor types in TCGA using mainly immune cell fractions from deconvolution analysis of all tissue bulk RNA-seq data and knowledge-based immune gene expression signatures (Bagaev et al., 2021; Thorsson et al., 2018). Notably, similar dominant immune pathways are also revealed in our analysis such as the IFN pathway or the imbalance between T cell and MNP subsets. However, these similarities go beyond immune-related features. We also identified stromal frequency as a key distinctive feature in the tumor classification across tissues that subdivide both IR and ID archetypes. Additional features may partly explain our higher number of archetypes but it may also lie in the power of a dataset that has large numbers of members and spans a diverse spectrum of cancers. Nevertheless, our progressive analysis from three coarse features through more granular definitions of T cells and MNP subsets suggest that it is the intersection of these 10 features that seem to arrive at a stable view of the major archetypes. However, further refinement and discovery is required, especially in IDs where the paucity of overall immune cells likely decreases our ability to resolve differences using only these 10 measurements.

As previously shown by others, in a single cancer type or in tumor mouse models, the TME can be coarsely categorized into IR and ID areas (Bagaev et al., 2021; Duan et al., 2020; Galon and Bruni, 2019; Mariathasan et al., 2018). Our analysis revealed that the tumor immune archetypes subdivide this broad classification and identify distinct immune networks for each archetype with unique relationships among cell densities and chemokine networks. Although IR archetypes are characterized by patterns previously described in melanoma, head and neck, or breast malignancies—such as a correlation between CD8$^+$ T cells, cDC1 ratio, and NK cell abundance (Barry et al., 2018; Salmon et al., 2016) or between CD4$^+$ T cell and cDC2 (Binnewies et al., 2019; Michea et al., 2018)—other archetypes implicate different immune cell networks such as the apparent enrichment of plasma cells and Tregs in TC archetype. That particular co-association was previously shown to support the plasma cell residency in bone marrow, and the presence of both cell types together associated with better response to immune checkpoint blockade therapy in sarcoma (Glatman Zaretsky et al., 2017; Petitprez et al., 2020). It is tempting to propose that tumors hijack specific immune archetypes that originate for use in a completely different setting (Table S6).

This idea is perhaps also suggested by the distinct patterns of chemokine expression identified for each tumor immune archetype. For example, although IR CD8 archetypes are defined by a CXCL9, CXCL10, CXCL11/CXCR3 axis that resembles the biology of an ongoing chronic viral infection (Metzemaekers et al., 2018), ID CD8 macrophage bias archetype is defined by a strong expression of chemokines binding CX3CR1, which is an axis essential in settings of immune surveillance and homeostasis (Gerlach et al., 2016). Further work is certainly needed to profile the vast array of other immune niches in homeostasis and immune responses and to test how these linked cell states are conserved across tissue and disease; the increasing availability of multi-omics resources will be an important step forward (Mulder et al., 2021; Reynolds et al., 2021).

Tumor-specific genetics and mutational burden have been proposed to be key for anti-tumor immunity in multiple cancers (Brown et al., 2014; Ghosh et al., 2021; Goodman et al., 2017). However, further investigation on extrapolating immune gene signature across TCGA cohort showed no correlation between gene expression and mutational burden in any cancer type (Bagaev et al., 2021; Spranger et al., 2016). Conversely, we demonstrated that immune tumor archetypes are associated with tumor proliferation, diverse tumor transcriptomic programs, and overall survival. Studying the synergy between immune archetypes and tumor mutational profile may further elucidate the importance of tumor mutational burden in anti-tumor immune response (Berger et al., 2018; Quigley et al., 2018).

In summary, our comprehensive characterization of the TME across many human solid cancer types reveals the existence of common and reproducible immune archetypes defined by distinct cell networks. To facilitate usage of our data for the wide research community, we will provide access to transcriptomic and compositional data for each archetype (https://datalibrary.ucsf.edu/public-resources). It is still unclear whether improving tumor cure rates will be based on the enhancement of a specific archetype, and the expansion of this classification in metastatic tumors as well as in animal models may help to test this (Maynard et al., 2020). Moreover, it will be important to expand this analysis to the peripheral immune system as well as biopsies before and after immunotherapies in order to define

Figure 7. Immune archetypes tie closely to tumor biology and disease outcome
(A) Left: UMAP display and graph-based clustering of tumor immune archetypes using 10-feature clustering color-coded by tumor type. Each dot represents a single patient. Right: stacked bar plot of the tumor type distribution for 10-feature immune archetypes.
(B) UMAP overlay (left) and box and whisker plot (right) of tumor proliferation measured by frequency of Ki67$^+$ CD45$^-$ cells by flow cytometry in the IPI cohort.
(C) Bubble plot of the median gene expression, in the tumor compartment, of gene sets associated with previously identified tumor transcriptional programs grouped by cluster/archetypes in the IPI cohort.
(D and E) Heatmap and hierarchical clustering of immune archetype gene signatures median expression in the tumor compartment (D) and the live compartment (E) in the IPI cohort.
(F) Multivariate survival regression of overall survival in the TCGA cohort for each immune archetype after multivariate analysis using gene signatures in (E), split by T conventional subset enrichment. Median survival (MS) and p value associated with each survival curve are noted.
*$1.00e-02 < p \leq 5.00e-02$; **$1.00e-03 < p \leq 1.00e-02$; ***$1.00e-04 < p \leq 1.00e-03$; ****$p \leq 1.00e-04$. See also Figure S7 and Tables S5 and S6.

the relationship between "dominant" and "reactive" tumor immune archetype across tumor types (Grünwald et al., 2021; Bi et al., 2021). Furthermore, there will undoubtedly be many additional ways to explore this dataset and discover patterns of immune and cancer biology. In this sense, the IPI dataset and the accompanying curation by these archetypal assignments can now serve as a rich resource to gain deeper understanding of cancer immunity in patients and therefore serve as a framework to direct immunotherapies to the most relevant biology.

Limitation of the study

There are a few caveats to note that impose limitations on this study that endeavors to define the dominant tumor immune archetypes across many tumor types. First, we used a single standardized processing approach regardless of the tumor type which was not ideal but necessary due to the limited quantity of available tumor tissue. Nevertheless, both simple immunofluorescence and identification of archetype-specific chemokine expression program would appear to support that we reliably recovered cells across tissues, but some archetypes such as the myeloid centric appear to be more impacted (Figures 2H and 2I), and further investigation of the spatial landscape of the tumor immune archetypes would be needed. In the future, a combination of multiplexed imaging technologies such ion beam imaging (MIBI) (Angelo et al., 2014; Keren et al., 2018) and single-cell spatial transcriptomics (Asp et al., 2019; Hu et al., 2020) will help mitigate this issue and enable analysis. In this way, archetype-discovery will play an important role in selecting tissues containing similar biology for studies.

Second, we focused our study on 10 major tumor-associated immune cell proportions across a wide variety of human cancer types, and this potentially biases analyses to dominant tumor archetypes because we only examine pre-defined cell populations. Interestingly, we demonstrated that the archetypes defined by these 10 features often correlate with enrichment of immune cell types that were not part of the clustering (Figure 6A). However, some other populations previously described as important for the tumor immune response, such as neutrophils, $\gamma\delta$ T cells (Coffelt et al., 2015; Wellenstein et al., 2019), or specific transcriptomic states of cDCs (Barry et al., 2018; Cheng et al., 2021; Lopes et al., 2021; Maier et al., 2020) were unable to be analyzed or appeared to be variable within archetypes (Figures 6C and S6). Future work using single-cell omics across tumor types—not only transcriptomic based due to the sensitivity of some of those cell types—will be critical. The combination of single-cell profiling in the context of immune archetypes would determine if the 12 tumor archetypes identified in our study can be further subdivided or if the cell subsets identified using these technologies generally align with our current landscape (Cheng et al., 2021; Oliveira et al., 2021; Slyper et al., 2020; Zhang et al., 2020).

Finally, using the IPI cohort, we developed archetype-specific gene signatures that enabled us to explore the association with disease outcome on a larger scale in the TCGA cohort. The two cohorts present important differences in term of demographics, tissue histology, and patient clinical history. Therefore, an extension of this work on a bigger cohort with the addition of clinical trials designed to demonstrate the benefits of targeting tumor archetypes would be incredibly important.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human tumor collection of the UCSF Immunoprofiler Initiative (IPI)
  - Assembling Cohorts
- METHOD DETAILS
  - Human tissue digestion and Multiparametric Flow Cytometry staining and sorting
  - Cell sorting for bulk RNA sequencing
  - Cell sorting for single-cell RNA sequencing of the tumor associated mononuclear phagocytic cells
  - Imaging Staining and Acquisition
  - Bulk RNA-Sequencing
  - Bioinformatic Data Processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis using 3 Features – Tcell, Myeloid, CD90+ CD44+ Stroma
  - Analysis using 6 Features – Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8, Tregs
  - Analysis using 10 Features – Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8, Tregs, Macrophages, Monocytes, cDC1, cDC2
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.12.004.

CONSORTIA

The members of the UCSF Immunoprofiler Initiative (IPI) are Matthew Spitzer, Lawrence Fong, Amanda Nelson, Raj Kumar, Justin Lee, Arun Burra, Joy Hsu, Caroline Hackett, Karen Tolentino, Jasmine Sjarif, Peter Johnson, Evans Shao, Darrell Abrau, Leonard Lupin, Cole Shaw, Zachary Collins, Tasha Lea, Carlos Corvera, Eric Nakakura, Julia Carnevale, Michael Alvarado, Kimberley Loo, Lawrence Chen, Melissa Chow, Jennifer Grandis, Will Ryan, Ivan El-Sayed, David Jablons, Gavitt Woodard, Maxwell W. Meng, Sima P. Porten, Hideho Okada, Margaret Tempero, Andrew Ko, Kim Kirkwood, Scott Vandenberg, Denise Guevarra, Erica Oropeza, Chris Cyr, Pat Glenn, Jennifer Bolen, Amanda Morton, and Walter Eckalbar. The affiliations for the Consortia members can be found in Table S7.

ACKNOWLEDGMENTS

we thank all patients and their families, for placing their trust in us. Acquisition and analysis of certain human samples described in this study was partially funded by contributions from AbbVie, Amgen, Bristol-Myers Squibb, and Pfizer as part of the UCSF Immunoprofiler Initiative. Further support came from the NIH (R01CA197363 and U01CA217864 to M.F.K.), NCI (R01 CA178015, CA222862, CA227807, CA239604, and CA230263 to E.A.C.), U24 (CA210974 to E.A.C.), U54 (CA224081 to E.A.C.), and P30 (P30CA082103 to A.B.O.).

AUTHOR CONTRIBUTIONS

A.J.C. and B.S. contributed equally to this manuscript. A.J.C., B.S., A.B., and M.F.K. performed and/or provided supervision of experiments; generated and analyzed data; and contributed to the manuscript by providing figures, tables, and important intellectual contributions. A.J.C., B.S., and M.F.K. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the analyses. A.J.C., B.S., and M.F.K. wrote and edited the manuscript. A.J.C., B.S., S.A., and A.A.R. performed computational analyses of the data. A.J.B. and M.A. performed the library RNA extraction and libraries construction for all the bulk RNA sequencing present in this manuscript. G.K.F., A.B.O., and S.A. provided important intellectual input in genomics and statistical analysis for this study. A.J.C., J.T., N.W.C., P.Y., G.C.R., D.K., A.S., B.D., A.B., K.T., E.S., P.J., J.S., J.H., J.L., T.C., K.C.B., R.J.A., T.H., and V.C. were part of the UCSF Immunoprofiler team who performed digestion, staining, sorting, and initial quality control analysis on the different tumor specimens included in the IPI dataset. R.K.K., J.S.C., C.E.A., A.I.D., P.H., A.A.D., J.R.K., and E.A.C. are part of the UCSF Immunoprofiler clinical team that were actively involved in facilitation of the access to the tumor specimens included in the IPI dataset. A.J.C., V.C., K.C.D., D.J.E., G.K.F., S.A., and M.F.K. are leadership members of the UCSF Immunoprofiler consortium and were actively involved in the establishment of the pipeline and patient consent and the direction of projects. B.S., A.P., C.C., J.Z., and K.C.D. were actively participated in patient enrollment and clinical data collection. The UCSF Immunoprofiler Initiative Consortium included all other scientist involved in the generation of data included in this study. All authors edited and critically revised the manuscript for important intellectual content and gave final approval for the version to be published.

DECLARATION OF INTERESTS

M.F.K. is a founder and shareholder of PIONYR immunotherapeutic and FOUNDERY innovations. A.I.D. is a shareholder of Trex and Neuvogen, and he is also a member of the scientific advisory board of Neuvogen, Bristol-Myers and Squibb, Merck, Roche, Pfizer, Genentech, Incyte, Amgen, and Novartis. E.A.C. is a consultant at IHR Therapeutics, Valar, and Pear Diagnostics, reports receiving commercial research grants from Astra Zeneca, Ferro Therapeutics, Senti Biosciences, Merck KgA, and Bayer, and has stock ownership in Tatara Therapeutics, Clara Health, BloodQ, and Guardant Health.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as living with a disability.

REFERENCES

Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson, R.M., Lowe, J.B., Liu, S.D., Zhao, S., et al. (2014). Multiplexed ion beam imaging of human breast tumors. Nat. Med. 20, 436–442.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 18, 220.

Arce Vargas, F., Furness, A.J.S., Litchfield, K., Joshi, K., Rosenthal, R., Ghorani, E., Solomon, I., Lesko, M.H., Ruef, N., Roddie, C., et al.; TRACERx Melanoma; TRACERx Renal; TRACERx Lung consortia (2018). Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. Cancer Cell 33, 649–663.e4.

Argüello, R.J., Combes, A.J., Char, R., Gigan, J.-P., Baaziz, A.I., Bousiquot, E., Camosseto, V., Samad, B., Tsui, J., Yan, P., et al. (2020). SCENITH: A Flow Cytometry-Based Method to Functionally Profile Energy Metabolism with Single-Cell Resolution. Cell Metab. 32, 1063–1075.e7. https://doi.org/10.1016/j.cmet.2020.11.007.

Asp, M., Giacomello, S., Larsson, L., Wu, C., Fürth, D., Qian, X., Wärdell, E., Custodio, J., Reimegård, J., Salmén, F., et al. (2019). A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. Cell 179, 1647–1660.e19.

Azizi, E., Carr, A.J., Plitas, G., Cornish, A.E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V., Setty, M., et al. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell 174, 1293–1308.e36.

Bengsch, B., Ohtani, T., Khan, O., Setty, M., Manne, S., O'Brien, S., Gherardini, P.F., Herati, R.S., Huang, A.C., Chang, K.-M., et al. (2018). Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 T cells. Immunity 48, 1029–1045.e5.

Bagaev, A., Kotlov, N., Nomie, K., Svekolkin, V., Gafurov, A., Isaeva, O., Osokin, N., Kozlov, I., Frenkel, F., Gancharova, O., et al. (2021). Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell 39, 845–865.e7.

Barry, K.C., Hsu, J., Broz, M.L., Cueto, F.J., Binnewies, M., Combes, A.J., Nelson, A.E., Loo, K., Kumar, R., Rosenblum, M.D., et al. (2018). A natural killer-dendritic cell axis defines checkpoint therapy-responsive tumor microenvironments. Nat. Med. 24, 1178–1191.

Beltra, J.-C., Manne, S., Abdel-Hakeem, M.S., Kurachi, M., Giles, J.R., Chen, Z., Casella, V., Ngiow, S.F., Khan, O., Huang, Y.J., et al. (2020). Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms. Immunity 52, 825–841.e8.

Berger, A.C., Korkut, A., Kanchi, R.S., Hegde, A.M., Lenoir, W., Liu, W., Liu, Y., Fan, H., Shen, H., Ravikumar, V., et al.; Cancer Genome Atlas Research Network (2018). A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell 33, 690–705.e9.

Bi, K., He, M.X., Bakouny, Z., Kanodia, A., Napolitano, S., Wu, J., Grimaldi, G., Braun, D.A., Cuoco, M.S., Mayorga, A., et al. (2021). Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. Cancer Cell 39, 649–661.e5.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 39, 782–795.

Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat. Med. 24, 541–550.

Binnewies, M., Mujal, A.M., Pollack, J.L., Combes, A.J., Hardison, E.A., Barry, K.C., Tsui, J., Ruhland, M.K., Kersten, K., Abushawish, M.A., et al. (2019). Unleashing Type-2 Dendritic Cells to Drive Protective Antitumor CD4+ T Cell Immunity. Cell 177, 556–571.e16.

Biswas, S.K., Allavena, P., and Mantovani, A. (2013). Tumor-associated macrophages: functional diversity, clinical significance, and open questions. Semin. Immunopathol. 35, 585–600.

Blank, C.U., Haining, W.N., Held, W., Hogan, P.G., Kallies, A., Lugli, E., Lynn, R.C., Philip, M., Rao, A., Restifo, N.P., et al. (2019). Defining 'T cell exhaustion'. Nat. Rev. Immunol. 19, 665–674.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. arXiv, 0803.0476. https://arxiv.org/abs/0803.0476.

Bosteels, C., Neyt, K., Vanheerswynghels, M., van Helden, M.J., Sichien, D., Debeuf, N., De Prijck, S., Bosteels, V., Vandamme, N., Martens, L., et al. (2020). Inflammatory Type 2 cDCs Acquire Features of cDC1s and Macrophages to Orchestrate Immunity to Respiratory Virus Infection. Immunity 52, 1039–1056.e9.

Böttcher, J.P., Bonavita, E., Chakravarty, P., Blees, H., Cabeza-Cabrerizo, M., Sammicheli, S., Rogers, N.C., Sahai, E., Zelenay, S., and Reis e Sousa, C. (2018). NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. Cell 172, 1022–1037.e14.

Brown, S.D., Warren, R.L., Gibb, E.A., Martin, S.D., Spinelli, J.J., Nelson, B.H., and Holt, R.A. (2014). Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res. 24, 743–750.

Broz, M.L., Binnewies, M., Boldajipour, B., Nelson, A.E., Pollack, J.L., Erle, D.J., Barczak, A., Rosenblum, M.D., Daud, A., Barber, D.L., et al. (2014). Dissecting the tumor myeloid compartment reveals rare activating antigen-presenting cells critical for T cell immunity. Cancer Cell 26, 638–652.

Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517, 576–582.

Cassetta, L., Fragkogianni, S., Sims, A.H., Swierczak, A., Forrester, L.M., Zhang, H., Soong, D.Y.H., Cotechini, T., Anur, P., Lin, E.Y., et al. (2019). Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets. Cancer Cell 35, 588–602.e10.

Chen, D.S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. Nature 541, 321–330.

Chen, B., Khodadoust, M.S., Liu, C.L., Newman, A.M., and Alizadeh, A.A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol. Biol. 1711, 243–259.

Chen, J., Tan, Y., Sun, F., Hou, L., Zhang, C., Ge, T., Yu, H., Wu, C., Zhu, Y., Duan, L., et al. (2020). Single-cell transcriptome and antigen-immunoglobin analysis reveals the diversity of B cells in non-small cell lung cancer. Genome Biol. 21, 152.

Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., Qin, S., Zhang, L., Ouyang, H., Du, P., et al. (2021). A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell 184, 792–809.e23.

Coffelt, S.B., Kersten, K., Doornebal, C.W., Weiden, J., Vrijland, K., Hau, C.-S., Verstegen, N.J.M., Ciampricotti, M., Hawinkels, L.J.A.C., Jonkers, J., and de Visser, K.E. (2015). IL-17-producing γδ T cells and neutrophils conspire to promote breast cancer metastasis. Nature 522, 345–348.

Combes, A., Camosseto, V., N'Guessan, P., Argüello, R.J., Mussard, J., Caux, C., Bendriss-Vermare, N., Pierre, P., and Gatti, E. (2017). BAD-LAMP controls TLR9 trafficking and signalling in human plasmacytoid dendritic cells. Nat. Commun. 8, 913.

Combes, A.J., Courau, T., Kuhn, N.F., Hu, K.H., Ray, A., Chen, W.S., Chew, N.W., Cleary, S.J., Kushnoor, D., Reeder, G.C., et al. (2021). Global absence and targeting of protective immune states in severe COVID-19. Nature 591, 124–130.

Davidson-Pilon, C., Kalderstam, J., Jacobson, N., sean-reed, Kuhn, B., Zivich, P., Williamson, M., Abdeali, J.K., Datta, D., Fiore-Gartland, A., et al. (2020). CamDavidsonPilon/lifelines: v0.25.6. Zenodo. https://doi.org/10.5281/zenodo.4136578.

Davies, D.L., and Bouldin, D.W. (1979). A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1, 224–227.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Dominguez, D., Tsai, Y.-H., Gomez, N., Jha, D.K., Davis, I., and Wang, Z. (2016). A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. Cell Res. 26, 946–962.

Duan, Q., Zhang, H., Zheng, J., and Zhang, L. (2020). Turning Cold into Hot: Firing up the Tumor Microenvironment. Trends Cancer 6, 605–618.

Dvorak, H.F. (1986). Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N. Engl. J. Med. 315, 1650–1659.

Ephrem, A., Epstein, A.L., Stephens, G.L., Thornton, A.M., Glass, D., and Shevach, E.M. (2013). Modulation of Treg cells/T effector function by GITR signaling is context-dependent. Eur. J. Immunol. 43, 2421–2429.

Etzerodt, A., Moulin, M., Doktor, T.K., Delfini, M., Mossadegh-Keller, N., Bajenoff, M., Sieweke, M.H., Moestrup, S.K., Auphan-Anezin, N., and Lawrence, T. (2020). Tissue-resident macrophages in omentum promote metastatic spread of ovarian cancer. J. Exp. Med. 217, e20191869.

Gajewski, T.F., Woo, S.-R., Zha, Y., Spaapen, R., Zheng, Y., Corrales, L., and Spranger, S. (2013). Cancer immunotherapy strategies based on overcoming barriers within the tumor microenvironment. Curr. Opin. Immunol. 25, 268–276.

Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat. Rev. Drug Discov. 18, 197–218.

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 313, 1960–1964.

Gentles, A.J., Newman, A.M., Liu, C.L., Bratman, S.V., Feng, W., Kim, D., Nair, V.S., Xu, Y., Khuong, A., Hoang, C.D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. 21, 938–945.

Gerlach, C., Moseman, E.A., Loughhead, S.M., Alvarez, D., Zwijnenburg, A.J., Waanders, L., Garg, R., de la Torre, J.C., and von Andrian, U.H. (2016). The Chemokine Receptor CX3CR1 Defines Three Antigen-Experienced CD8 T Cell Subsets with Distinct Roles in Immune Surveillance and Homeostasis. Immunity 45, 1270–1284.

Ghosh, M., Saha, S., Bettke, J., Nagar, R., Parrales, A., Iwakuma, T., van der Velden, A.W.M., and Martinez, L.A. (2021). Mutant p53 suppresses innate immune signaling to promote tumorigenesis. Cancer Cell 39, 494–508.e5.

Glatman Zaretsky, A., Konradt, C., Dépis, F., Wing, J.B., Goenka, R., Atria, D.G., Silver, J.S., Cho, S., Wolf, A.I., Quinn, W.J., et al. (2017). T Regulatory Cells Support Plasma Cell Populations in the Bone Marrow. Cell Rep. 18, 1906–1916.

Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. 38, 675–678.

Goodman, A.M., Kato, S., Bazhenova, L., Patel, S.P., Frampton, G.M., Miller, V., Stephens, P.J., Daniels, G.A., and Kurzrock, R. (2017). Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol. Cancer Ther. 16, 2598–2608.

Goswami, S., Walle, T., Cornish, A.E., Basu, S., Anandhan, S., Fernandez, I., Vence, L., Blando, J., Zhao, H., Yadav, S.S., et al. (2020). Immune profiling of human tumors identifies CD73 as a combinatorial target in glioblastoma. Nat. Med. 26, 39–46.

Gotwals, P., Cameron, S., Cipolletta, D., Cremasco, V., Crystal, A., Hewes, B., Mueller, B., Quaratino, S., Sabatos-Peyton, C., Petruzzelli, L., et al. (2017). Prospects for combining targeted and conventional cancer therapy with immunotherapy. Nat. Rev. Cancer 17, 286–301.

Grünwald, B.T., Devisme, A., Andrieux, G., Vyas, F., Aliar, K., McCloskey, C.W., Macklin, A., Jang, G.H., Denroche, R., Romero, J.M., et al. (2021). Spatially confined sub-tumor microenvironments in pancreatic cancer. Cell 184, 5577–5592. https://doi.org/10.1016/j.cell.2021.09.022.

Gubin, M.M., Esaulova, E., Ward, J.P., Malkova, O.N., Runci, D., Wong, P., Noguchi, T., Arthur, C.D., Meng, W., Alspach, E., et al. (2018). High-Dimensional Analysis Delineates Myeloid and Lymphoid Compartment Remodeling during Successful Immune-Checkpoint Cancer Therapy. Cell 175, 1014–1030.e19.

Gueguen, P., Metoikidou, C., Dupic, T., Lawand, M., Goudot, C., Baulande, S., Lameiras, S., Lantz, O., Girard, N., Seguin-Givelet, A., et al. (2021). Contribution of resident and circulating precursors to tumor-infiltrating CD8+ T cell populations in lung cancer. Sci. Immunol. 6, eabd5778.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell 144, 646–674.

Hegde, P.S., Karanikas, V., and Evers, S. (2016). The Where, the When, and the How of Immune Monitoring for Cancer Immunotherapies in the Era of Checkpoint Inhibition. Clin. Cancer Res. 22, 1865–1874.

Hu, K.H., Eichorst, J.P., McGinnis, C.S., Patterson, D.M., Chow, E.D., Kersten, K., Jameson, S.C., Gartner, Z.J., Rao, A.A., and Krummel, M.F. (2020). ZipSeq: barcoding for real-time mapping of single cell transcriptomes. Nat. Methods 17, 833–843.

Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. Cell 165, 35–44.

Iwai, Y., Terawaki, S., and Honjo, T. (2005). PD-1 blockade inhibits hematogenous spread of poorly immunogenic tumor cells by enhanced recruitment of effector T cells. Int. Immunol. 17, 133–144.

Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S.-R., Kurian, A., Van Valen, D., West, R., et al. (2018). A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell 174, 1373–1387.e19.

Khan, O., Giles, J.R., McDonald, S., Manne, S., Ngiow, S.F., Patel, K.P., Werner, M.T., Huang, A.C., Alexander, K.A., Wu, J.E., et al. (2019). TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. Nature 571, 211–218.

Kinker, G.S., Greenwald, A.C., Tal, R., Orlova, Z., Cuoco, M.S., McFarland, J.M., Warren, A., Rodman, C., Roth, J.A., Bender, S.A., et al. (2019). Pan-cancer single cell RNA-seq uncovers recurring programs of cellular heterogeneity. bioRxiv. https://doi.org/10.1101/807552.

Kinker, G.S., Greenwald, A.C., Tal, R., Orlova, Z., Cuoco, M.S., McFarland, J.M., Warren, A., Rodman, C., Roth, J.A., Bender, S.A., et al. (2020). Pan-cancer single cell RNA-seq identifies recurring programs of cellular heterogeneity. Nat. Genet. 52, 1208–1218.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296.

Kumar, B.V., Ma, W., Miron, M., Granot, T., Guyer, R.S., Carpenter, D.J., Senda, T., Sun, X., Ho, S.-H., Lerner, H., et al. (2017). Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. Cell Rep. 20, 2921–2934.

Lavin, Y., Kobayashi, S., Leader, A., Amir, E.D., Elefant, N., Bigenwald, C., Remark, R., Sweeney, R., Becker, C.D., Levine, J.H., et al. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. Cell 169, 750–765.e17.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29.

Leach, D.R., Krummel, M.F., and Allison, J.P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. Science 271, 1734–1736.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Lopes, N., McIntyre, C., Martin, S., Raverdeau, M., Sumaria, N., Kohlgruber, A.C., Fiala, G.J., Agudelo, L.Z., Dyck, L., Kane, H., et al. (2021). Distinct metabolic programs established in the thymus control effector functions of γδ T cell subsets in tumor microenvironments. Nat. Immunol. 22, 179–192.

Loyher, P.-L., Hamon, P., Laviron, M., Meghraoui-Kheddar, A., Goncalves, E., Deng, Z., Torstensson, S., Bercovici, N., Baudesson de Chanville, C., Combadière, B., et al. (2018). Macrophages of distinct origins contribute to tumor development in the lung. J. Exp. Med. 215, 2536–2553.

Maier, B., Leader, A.M., Chen, S.T., Tung, N., Chang, C., LeBerichel, J., Chudnovskiy, A., Maskey, S., Walker, L., Finnigan, J.P., et al. (2020). A conserved dendritic-cell regulatory program limits antitumour immunity. Nature 580, 257–262.

Mandal, R., Şenbabaoğlu, Y., Desrichard, A., Havel, J.J., Dalin, M.G., Riaz, N., Lee, K.-W., Ganly, I., Hakimi, A.A., Chan, T.A., and Morris, L.G. (2016). The head and neck cancer immune landscape and its immunotherapeutic implications. JCI Insight 1, e89829.

Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., III, Koeppen, H., Astarita, J.L., Cubas, R., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 554, 544–548.

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., and Azzi, M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. In Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter - CHItaly '17 (ACM Press), pp. 1–5.

Maynard, A., McCoach, C.E., Rotow, J.K., Harris, L., Haderk, F., Kerr, D.L., Yu, E.A., Schenk, E.L., Tan, W., Zee, A., et al. (2020). Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell 182, 1232–1251.e22.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 8, 329–337.e4.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv, 1802.03426. https://arxiv.org/abs/1802.03426.

Metzemaekers, M., Vanheule, V., Janssens, R., Struyf, S., and Proost, P. (2018). Overview of the Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-Inducible CXC Chemokine Receptor 3 Ligands. Front. Immunol. 8, 1970.

Michea, P., Noël, F., Zakine, E., Czerwinska, U., Sirven, P., Abouzid, O., Goudot, C., Scholer-Dahirel, A., Vincent-Salomon, A., Reyal, F., et al. (2018). Adjustment of dendritic cells to the breast-cancer microenvironment is subset specific. Nat. Immunol. 19, 885–897.

Mlecnik, B., Bindea, G., Angell, H.K., Maby, P., Angelova, M., Tougeron, D., Church, S.E., Lafontaine, L., Fischer, M., Fredriksen, T., et al. (2016). Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. Immunity 44, 698–711.

Molgora, M., Esaulova, E., Vermi, W., Hou, J., Chen, Y., Luo, J., Brioschi, S., Bugatti, M., Omodei, A.S., Ricci, B., et al. (2020). TREM2 Modulation Remodels the Tumor Myeloid Landscape Enhancing Anti-PD-1 Immunotherapy. Cell 182, 886–900.e17.

Mondini, M., Loyher, P.-L., Hamon, P., Gerbé de Thoré, M., Laviron, M., Berthelot, K., Clémenson, C., Salomon, B.L., Combadière, C., Deutsch, E., and Boissonnas, A. (2019). CCR2-Dependent Recruitment of Tregs and Monocytes Following Radiotherapy Is Associated with TNFα-Mediated Resistance. Cancer Immunol. Res. 7, 376–387.

Mujal, A.M., and Krummel, M.F. (2019). Immunity as a continuum of archetypes. Science 364, 28–29.

Mulder, K., Patel, A.A., Kong, W.T., Piot, C., Halitzki, E., Dunsmore, G., Khalilnezhad, S., Irac, S.E., Dubuisson, A., Chevrier, M., et al. (2021). Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. Immunity 54, 1883–1900.e5.

Muñoz, D.P., Yannone, S.M., Daemen, A., Sun, Y., Vakar-Lopez, F., Kawahara, M., Freund, A.M., Rodier, F., Wu, J.D., Desprez, P.-Y., et al. (2019). Targetable mechanisms driving immunoevasion of persistent senescent cells link chemotherapy-resistant cancer to aging. JCI Insight 5, e124716.

Nagarsheth, N., Wicha, M.S., and Zou, W. (2017). Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. Nat. Rev. Immunol. *17*, 559–572.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Oh, D.Y., Kwek, S.S., Raju, S.S., Li, T., McCarthy, E., Chow, E., Aran, D., Ilano, A., Pai, C.S., Rancan, C., et al. (2020). Intratumoral CD4⁺ T Cells Mediate Antitumor Cytotoxicity in Human Bladder Cancer. Cell *181*, 1612–1625.e13.

Oliveira, G., Stromhaug, K., Klaeger, S., Kula, T., Frederick, D.T., Le, P.M., Forman, J., Huang, T., Li, S., Zhang, W., et al. (2021). Phenotype, specificity and avidity of antitumour CD8⁺ T cells in melanoma. Nature *596*, 119–125.

Petitprez, F., de Reyniès, A., Keung, E.Z., Chen, T.W.-W., Sun, C.-M., Calderaro, J., Jeng, Y.-M., Hsiao, L.-P., Lacroix, L., Bougoüin, A., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. Nature *577*, 556–560.

Plitas, G., Konopacki, C., Wu, K., Bos, P.D., Morrow, M., Putintseva, E.V., Chudakov, D.M., and Rudensky, A.Y. (2016). Regulatory T Cells Exhibit Distinct Features in Human Breast Cancer. Immunity *45*, 1122–1134.

Quigley, D.A., Dang, H.X., Zhao, S.G., Lloyd, P., Aggarwal, R., Alumkal, J.J., Foye, A., Kothari, V., Perry, M.D., Bailey, A.M., et al. (2018). Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell *174*, 758–769.e9.

Reynolds, G., Vegh, P., Fletcher, J., Poyner, E.F.M., Stephenson, E., Goh, I., Botting, R.A., Huang, N., Olabi, B., Dubois, A., et al. (2021). Developmental cell programs are co-opted in inflammatory skin disease. Science *371*, eaba6500.

Richer, M.J., Lang, M.L., and Butler, N.S. (2016). T Cell Fates Zipped Up: How the Bach2 Basic Leucine Zipper Transcriptional Repressor Directs T Cell Differentiation and Function. J. Immunol. *197*, 1009–1015.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Roberts, E.W., Broz, M.L., Binnewies, M., Headley, M.B., Nelson, A.E., Wolf, D.M., Kaisho, T., Bogunovic, D., Bhardwaj, N., and Krummel, M.F. (2016). Critical Role for CD103(+)/CD141(+) Dendritic Cells Bearing CCR7 for Tumor Antigen Trafficking and Priming of T Cell Immunity in Melanoma. Cancer Cell *30*, 324–336.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.

Salmon, H., Idoyaga, J., Rahman, A., Leboeuf, M., Remark, R., Jordan, S., Casanova-Acebes, M., Khudoynazarova, M., Agudo, J., Tung, N., et al. (2016). Expansion and Activation of CD103(+) Dendritic Cell Progenitors at the Tumor Site Enhances Tumor Responses to Therapeutic PD-L1 and BRAF Inhibition. Immunity *44*, 924–938.

Sancho, D., Joffre, O.P., Keller, A.M., Rogers, N.C., Martínez, D., Hernanz-Falcón, P., Rosewell, I., and Reis e Sousa, C. (2009). Identification of a dendritic cell receptor that couples sensing of necrosis to immunity. Nature *458*, 899–903.

Savas, P., Virassamy, B., Ye, C., Salim, A., Mintoff, C.P., Caramia, F., Salgado, R., Byrne, D.J., Teo, Z.L., Dushyanthen, S., et al.; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab) (2018). Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat. Med. *24*, 986–993.

Scott, A.C., Dündar, F., Zumbo, P., Chandran, S.S., Klebanoff, C.A., Shakiba, M., Trivedi, P., Menocal, L., Appleby, H., Camara, S., et al. (2019). TOX is a critical regulator of tumour-specific T cell differentiation. Nature *571*, 270–274.

Slyper, M., Porter, C.B.M., Ashenberg, O., Waldman, J., Drokhlyansky, E., Wakiro, I., Smillie, C., Smith-Rosario, G., Wu, J., Dionne, D., et al. (2020). A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat. Med. *26*, 792–802.

Spranger, S. (2016). Tumor Heterogeneity and Tumor Immunity: A Chicken-and-Egg Problem. Trends Immunol. *37*, 349–351.

Spranger, S., Luke, J.J., Bao, R., Zha, Y., Hernandez, K.M., Li, Y., Gajewski, A.P., Andrade, J., and Gajewski, T.F. (2016). Density of immunogenic antigens does not explain the presence or absence of the T-cell-inflamed tumor microenvironment in melanoma. Proc. Natl. Acad. Sci. USA *113*, E7759–E7768.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.-H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al.; Cancer Genome Atlas Research Network (2018). The Immune Landscape of Cancer. Immunity *48*, 812–830.e14.

van Galen, P., Hovestadt, V., Wadsworth Ii, M.H., Hughes, T.K., Griffin, G.K., Battaglia, S., Verga, J.A., Stephansky, J., Pastika, T.J., Lombardi Story, J., et al. (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell *176*, 1265–1281.e24.

Varn, F.S., Wang, Y., Mullins, D.W., Fiering, S., and Cheng, C. (2017). Systematic Pan-Cancer Analysis Reveals Immune Cell Interactions in the Tumor Microenvironment. Cancer Res. *77*, 1271–1282.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.; SciPy 1.0 Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Vivian, J., Rao, A.A., Nothaft, F.A., Ketchum, C., Armstrong, J., Novak, A., Pfeil, J., Narkizian, J., Deran, A.D., Musselman-Brown, A., et al. (2017). Toil enables reproducible, open source, big biomedical data analyses. Nat. Biotechnol. *35*, 314–316.

Wellenstein, M.D., Coffelt, S.B., Duits, D.E.M., van Miltenburg, M.H., Slagter, M., de Rink, I., Henneman, L., Kas, S.M., Prekovic, S., Hau, C.-S., et al. (2019). Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. Nature *572*, 538–542.

Wiley, C.D., Flynn, J.M., Morrissey, C., Lebofsky, R., Shuga, J., Dong, X., Unger, M.A., Vijg, J., Melov, S., and Campisi, J. (2017). Analysis of individual cells identifies cell-to-cell variability following induction of cellular senescence. Aging Cell *16*, 1043–1050.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

Zemmour, D., Zilionis, R., Kiner, E., Klein, A.M., Mathis, D., and Benoist, C. (2018). Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. Nat. Immunol. *19*, 291–301.

Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., et al. (2019). Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. Cell *179*, 829–845.e20.

Zhang, L., Li, Z., Skrzypczynska, K.M., Fang, Q., Zhang, W., O'Brien, S.A., He, Y., Wang, L., Zhang, Q., Kim, A., et al. (2020). Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell *181*, 442–459.e29.

Zhou, L., Chong, M.M.W., and Littman, D.R. (2009). Plasticity of CD4+ T cell lineage differentiation. Immunity *30*, 646–655.

Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H.D., Krishnan, I., Maroni, G., Meyerovitz, C.V., Kerwin, C.M., et al. (2019). Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. Immunity *50*, 1317–1334.e10.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-human CD45 APC/e780 (clone HI30)	Thermo Fisher	Cat# 47-0459-42; RRID:AB_1944368
anti-human CD3e PerCP/e710 (clone OKT3)	Thermo Fisher	Cat# 46-0037-42; RRID:AB_1834395
anti-human HLA-DR BUV395 (clone G46-6)	BD Biosciences	Cat# 564040; RRID:AB_2738558
anti-human CD56 BUV737 (clone NCAM16.2)	BD Biosciences	Cat# 564448; RRID:AB_2744432
anti-human CD4 PE/Dazzle 594 (clone S3.5)	Biolegend	Cat# 100455; RRID:AB_2565844
anti-human CD8a BV605 (clone RPA-T8)	Biolegend	Cat# 301039; RRID:AB_11126985
anti-human CD127 BV650 (clone HIL-7R-M21)	BD Biosciences	Cat# 563225; RRID:AB_2738081
anti-human CD38 AF700 (clone HIT2)	Biolegend	Cat# 303523; RRID:AB_2228785
anti-human CD25 APC (clone 2A3)	BD Biosciences	Cat# 340939; RRID:AB_400551
anti-human PD-1 BV786 (clone EH12)	BD Biosciences	Cat# 563789; RRID:AB_2738425
anti-human FoxP3 PE/Cy7 (clone 236A/E7)	Thermo Fisher	Cat# 25-4777-41; RRID:AB_2573449
anti-human CTLA-4 BV421 (clone BNI3)	BD Biosciences	Cat# 565931; RRID:AB_2739395
anti-human/mouse/rat Ki67 AF488 (clone SolA15)	Thermo Fisher	Cat# 11-5698-82; RRID:AB_11151330
anti-human CD19 PerCP/e710 (clone H1B19)	Thermo Fisher	Cat# 45-0199-42; RRID:AB_2043821
anti-human CD20 PerCP/e710 (clone 2H7)	Thermo Fisher	Cat# 45-0209-42; RRID:AB_10717086
anti-human CD56 PerCP/e710 (clone CMSSB)	Thermo Fisher	Cat# 46-0567-42; RRID:AB_10548939
anti-human CD64 BUV737 (clone 10.1)	BD Biosciences	Cat# 564425; RRID:AB_2744446
anti-human CD11c AF700 (clone 3.9)	Thermo Fisher	Cat# 56-0116-42, RRID:AB_10547281)
anti-human CD16 BV605 (clone 3G8)	Biolegend	Cat# 302039; RRID:AB_2561354)
anti-human CD304 PE (clone 12C2)	Biolegend	Cat# 354503; RRID:AB_11219200
anti-human CD1C/BDCA-1 PE/Cy7 (clone L161)	Biolegend	Cat# 331515; RRID:AB_1953227
anti-human BDCA-3 FITC (clone AD5-14H12)	Miltenyi	Cat# 130-098-843; RRID:AB_2661177
anti-human CD14 BV711 (clone M5E2)	Biolegend	Cat# 301837; RRID:AB_11218986
Biological samples		
Human tumor samples	UC San Francisco	IRB # 20-31740
Deposited data		
All bulk RNaseq data for the IPI cohort and single cell RNaseq data previously unpublished	This paper	GEO: GSE184398
Single cell RNaseq data from melanoma invaded lymph node.	(Binnewies et al., 2019)	GEO: GSE125680
Single cell RNaseq data from Renal Clear cell carcinoma	(Argüello et al., 2020)	GEO: GSE159913
GitHub	UCSF DSCO_LAB Github	https://github.com/UCSF-DSCOLAB/pan_cancer_immune_archetypes
Software and algorithms		
Python (2.7.15 & 3.7.2)		https://www.python.org
Pandas (0.24.1,1.0.5)		https://pandas.pydata.org/
Seaborn (0.9.0, 0.10.1)		https://seaborn.pydata.org
Matplotlib (2.2.3, 3.2.2)		https://matplotlib.org
Lifelines	Davidson-Pilon et al., 2020	https://lifelines.readthedocs.io/en/latest/

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Scanpy 1.5.1	Wolf et al., 2018	http://scanpy.readthedocs.io/en/stable
SciPy	Virtanen et al., 2020	https://scipy.org/
BWA-mem		http://bio-bwa.sourceforge.net/
RSEM	Li and Dewey, 2011	http://deweylab.biostat.wisc.edu/rsem/README.html
Imaris v9.5		https://imaris.oxinst.com/
STAR	Dobin et al., 2013	N/A
limma	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
edgeR	Robinson et al., 2010	https://bioconductor.org/packages/release/bioc/html/edgeR.html
voom	Law et al., 2014	https://www.rdocumentation.org/packages/limma/versions/3.28.14/topics/voom
R	R Development Core Team, 2010	https://www.r-project.org/

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Matthew F. Krummel (matthew.krummel@ucsf.edu).

Materials availability
This study did not generate new unique reagents.

Data and code availability
Single-cell RNA-seq and bulk RNaseq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resource table. Microscopy and flow data reported in this paper will be shared by the lead contact upon request.

All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the Key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human tumor collection of the UCSF Immunoprofiler Initiative (IPI)
Tumor samples for the Immunoprofiler was transported from various cancer operating rooms (ORs) as well as from outpatient clinics. All patients consented by the UCSF IPI clinical coordinator group for tissue collection under a UCSF IRB approved protocol (UCSF IRB# 20-31740). Samples were obtained after surgical excision with biopsies taken by Pathology Assistants to confirm the presence of tumor cells. Patients were selected without regard to prior treatment. Freshly resected samples were placed in ice-cold PBS or Leibovitz's L-15 medium in a 50 mL conical tube and immediately transported to the laboratory for sample labeling and prepare either the whole tissue for digestion into single-cell suspension or a part of the tissue was sliced and preserved for imaging analysis. All Clinical information on the different patient of the cohort can be found in Table S1.

Assembling Cohorts
UCSF Immunoprofiler initiative (IPI)
An initial set of 427 bulk RNaseq patient samples in the live compartment were evaluated based on an in-house metric, the EHK score, that serves as a measure of data quality. Each sample is given a score of 0 through 10 depending on the number of EHK genes that are expressed above a precalculated minimum threshold. The threshold was learned from our data by examining the expression distributions of EHK genes and validated using the corresponding distributions in TCGA. A score of 10 represents the highest quality

data where 10 out of 10 EHK genes are expressed above the minimum threshold. Filtering for samples with an EHK score of EHK8, EHK9 and EHK10 reduced the sample set to 298 patient samples. The sample set was then filtered to remove all adjacent normal samples and all biological replicates, reducing the sample set further to 260 samples. These 260 patient samples are the IPI cohort along with 199 overlapping tcell compartment samples and 189 myeloid compartment samples.

TCGA

Tumor RNaseq counts and TPM along with curated clinical data for 13 cancer types (Bladder urothelial carcinoma (BLCA), Colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), Head and Neck squamous cell carcinoma (HNSC), Kidney renal clear cell carcinoma (KIRC), Liver hepatocellular carcinoma (LIHC), Lung adenocarcinoma (LUAD), Ovarian serous cysta-denocarcinoma (OV), Pancreatic adenocarcinoma (PAAD), Sarcoma (SARC), Skin Cutaneous Melanoma (SKCM), Uterine Corpus Endometrial Carcinoma (UCEC), Uterine Carcinosarcoma (UCS)), from the Toil recompute (Vivian et al., 2017) data in the TCGA Pan-Cancer (PANCAN) cohort, were downloaded from the UCSC Xena browser (Goldman et al., 2020). The initial set of 4677 tumor samples was filtered down to include primary solid tumors and metastatic sample types (sample type codes = 01 & 06) only, to parallel the IPI cohort sample types as accurately as possible. This reduced the patient sample set to 4341 tumor samples.

METHOD DETAILS

Human tissue digestion and Multiparametric Flow Cytometry staining and sorting

Tumor or metastatic tissue was thoroughly chopped with surgical scissors and transferred to GentleMACs C Tubes (Miltenyi Biotec) containing 20 uL/mL Liberase TL (5 mg/ml, Roche) and 50 U/ml DNase I (Roche) in RPMI 1640 per 0.3 g tissue. GentleMACs C Tubes were then installed onto the GentleMACs Octo Dissociator (Miltenyi Biotec) and incubated for 45min according to the manufacturer's instructions. Samples were then quenched with 15 mL of sort buffer (PBS/2% FCS/2mM EDTA), filtered through 100 um filters and spun down. Red blood cell lysis was performed with 175 mM ammonium chloride if needed.

Cells were then incubated with Human FcX (Biolegend) to prevent non-specific antibody binding. Cells were then washed in DPBS and incubated with Zombie Aqua Fixable Viability Dye (Thermo). Following viability dye, cells were washed with sort buffer and incubated with cell surface antibodies mix diluted in the BV stain buffer (BD Biosciences) following manufacturer instruction for 30 minutes on ice in the dark and subsequently fixed in either Fixation Buffer (BD Biosciences) or in Foxp3/Transcription Factor Staining Buffer Set (eBioscience) if intracellular staining was required.

Cell sorting for bulk RNA sequencing

Bulk RNaseq compartment	Gating Strategy	Potential Caveats
All viable cells 'Live' compartment	All cells Viability dye negative LIVE/DEAD Fixable Aqua Dead Cell Stain Cat:L34957	This was performed after tissue enzymatic digestion and allow us to sequence only viable cells after isolation. Therefore it is slightly different than bulk RNaseq from whole tissue prior digestion as it was done in the TCGA dataset.
Conventional CD4+ and CD8+ T cells compartment	CD4+ and CD8+ Tconv: CD45+, CD3+, CD4+, CD8+, CD25- CD19/20-, CD56-	Using this gating strategy we aimed to enrich for both Conventional CD4+ and CD8+ T cells but we don't separate CD4+ and CD8+
Regulatory CD4+ T cells compartment	Treg: CD45+, CD3+, CD4+, CD8-, CD25+, CD19/20-, CD56-	Using this gating strategy we aimed to enrich for Tregulatory T cells using CD 25 expression but as we can't use FOXP3 for sorting unpermeabilized viable cell this compartment may contain some activated CD4+ T cells.
Myeloid cells compartment	CD45+, CD3+, CD19-, CD20- CD56- HLA-DR+	Using this gating strategy we aimed to enrich for tumor associated myeloid cell however this compartment may contain plasma cells which would have downregulate CD19/20 but still express HLA-DR
Cd44+ CD90 stromal cells compartment	CD45-, CD44+, CD90+	Using this gating strategy we aimed to enrich for one type of stromal cells however this gating strategy doesn't encompass all stromal cell subtype present in tumor
CD45- Tumor cell compartment	CD45-, CD44-, CD90-	Using this gating strategy we aimed to enrich for tumor cells however this gating strategy is permissive and this compartment may contain other tumor associated non immune cell types

For cell sorting single cell suspension was stained with cell surface antibodies mix diluted in BV stain buffer (BD Biosciences) following manufacturer instruction for 30 minutes on ice in the dark and subsequently washed three time and resuspended in of sort buffer (PBS/2% FCS/2mM EDTA) after filtering in 100um filter (thermo). When possible, a maximum of 6 different tumor associated cell subsets were sorted with a BD FACSAria Fusion into 1.5mL Eppendorf tubes containing 150ul of lysis buffer used for mRNA extraction (Invitrogen). The 6 different tumor associated cell types between 5000 and 50000 cells were sorted based their protein expression profile as describe here: 'live cells' all cells negative for the viability dye; Tconv: negative for viability dye CD45+ CD3+ CD19- CD56- CD25-; Treg: negative for viability dye CD45+ CD3+ CD19- CD56- CD4+ CD25+; Myeloid: negative for viability dye CD45+ CD3- CD19- CD56- HLADR+; CD90+ CD44+ Stroma: negative for viability dye CD45- CD44+ CD90+; Tumor: negative for viability dye CD45- CD44- CD90-. Names of the different sorted populations were given based on the presumptive populations, and we accepted that both due to slight sorting contamination and due to the limits of antibodies to completely 'define' a population, that other populations would occasionally infiltrate these e.g., some 'activated' CD4 cells express CD25 and might infiltrate this RNA sample.

Cell sorting for single-cell RNA sequencing of the tumor associated mononuclear phagocytic cells

For the discovery cohort scRNA-seq, live CD3-CD19/20-CD56- SSC-A dim CD16 dim (to exclude neutrophils) cells were sorted from a melanoma involved draining LN on a BD FACSAria Fusion. For the discovery cohort the same sorting strategy was applied to melanoma, kidney and head and Neck primary tumors. After sorting, cells were pelleted and resuspended at 1x10e3 cells/ul in 0.04% BSA/PBA and loaded onto the Chromium Controller (10X Genomics). Samples were processed for single-cell encapsulation and cDNA library generation using the Chromium Single Cell 3′ v2 Reagent Kits (10X Genomics). The library was subsequently sequenced on an Illumina HiSeq 4000 (Illumina). All samples were sequenced at 25,000 reads per cell.

Imaging Staining and Acquisition

H&E slides were prepared in Leica Autostainer XL; slides stained in hematoxylin (Thermo Scientific Shandon Instant Hematoxylin cat. 6765015) for 7 minutes and in Eosin (Thermo Scientific Shandon Instant Eosin-Y Alcoholic cat. 531946) for 20 s. Immunofluorescence (IF) staining was performed on the Ventana Discovery Ultra autostainer using Discovery reagents (Ventana Medical Systems) according to the manufacturer's instructions, except as noted. After deparaffinization, antigen retrieval was performed with Cell Conditioning 1 (CC1) solution (cat. 950-124) for 64 minutes at 95°C. Primary antibody CD45 (D9M8I by CST #13917), was incubated for 20 minutes at 36°C. Secondary antibodies (cat. 760-149) were incubated for 12 minutes. Endogenous peroxidase was inhibited by Discovery Inhibitor (cat. 760-4840) for 8 minutes and non-specific binding blocked with Goat block (cat. 760-6008) for 4 minutes. The primary antibody was visualized with Discovery Rhodamine 6G Kit (cat. 760-244). Finally, slides were counterstained with DAPI (Akoya cat. FP1490) for 8 minutes. IF slides were scanned in a whole slide scanner AxioScan.Z1 (Zeiss) with Plan-Apochromat 20x/0.8 M27 objective and images were captured by Orca-Flash 4.0 v2 CMOS camera (Hamamatsu). Filters used for specific fluorophores are: Spectral Gold (Semrock) filter was used for Rhodamine 6G and 87 HE DAPI (Zeiss) for DAPI. H&E slides were scanned in brightfield with the same objective and a HV-F202SCL CCD camera (Hitachi).

Bulk RNA-Sequencing
Library Preparation

The mRNA was isolated via DynaBeads Direct, then converted into amplified cDNA using Tecan Ovation RNA-Seq System V2 kit, following the manufacturer protocol. The dsDNA went through tagmentation, amplification and clean up with AMPure XP beads steps, using the Illumina Nextera XT DNA Library Prep Kit. Quality control analysis was performed on the resulting pooled libraries with the Agilent Bioanalyzer HS DNA chip to assess fragment size distribution and concentration.

Sequencing

The pooled libraries were then sequenced via single-read MiSeqMiniSeq to check if more than 10 percent of the reads aligned to coding regions and contained more than 1,000 unique reads in total. Sequencing efficiency was determined by protein-coding read fractions. The protein-coding read fractions were then used to normalize libraries for high depth sequencing with the goal of obtaining an average of 10M protein-coding reads per sample. These libraries were submitted the UCSF Center for Advanced Technology for 100bp paired end (PE100) sequencing on the HiSeq4000.

Bioinformatic Data Processing
Bulk RNaseq

The RNaseq reads were first aligned to reference of rRNA and mitochondrial rRNA using BWA-mem to deplete the dataset of any remaining rRNA. The remaining reads were aligned to the Ensembl GRCh38.85 transcriptome build using STAR (Dobin et al., 2013). Gene expression was computed from the alignments in counts and TPM (transcripts per million) using RSEM (Li and Dewey, 2011).

QC of Bulk RNaseq

QC of the of the RNaseq was done to identify any swapped samples. DEG was performed on all pairs of the tcell, myeloid, treg, stroma and tumor compartments. The top 500 genes by p value were identified for each DEG pair giving 5000 genes which were reduced to 2171 genes after removing redundant genes. Principal component analysis (PCA) was done on the mean-centered

and scaled logCPM (log counts per million) of the counts of these genes. K-means clustering was performed on the first 4 principal components with K = 5. This yielded 5 clusters that were mostly made up of one compartment. Any samples that clustered with a different compartment were considers swaps except for the tumor compartment which we assumed would be promiscuous.

Data pre-processing of 10x Genomics Chromium scRNA-seq data:

Feature-barcode matrices were obtained for sample by aligning the raw fastqs to GRCh38 reference genome (annotated with Ensembl v85) using the Cellranger count. Raw feature-barcode matrices were loaded into Seurat 3.1.5 (Stuart et al., 2019) and genes with fewer than 3 UMIs were dropped from the analyses. Matrices were further filtered to remove events with greater than 20% percent mitochondrial content, events with greater than 50% ribosomal content, or events with fewer than 100 total genes. The cell cycle state of each cell was assessed using a published set of genes associated with various stages of human mitosis (Dominguez et al., 2016).

Data quality control and Normalization for scRNA-seq data:

The filtered count matrices were normalized, and variance stabilized using negative binomial regression via the scTransform method offered by Seurat (Hafemeister and Satija, 2019). The effects of mitochondrial content, ribosomal content, and cell cycle state were regressed out of the normalized data to prevent any confounding signal. The normalized matrices were reduced to a lower dimension using Principal Component Analyses (PCA) and the first 30 principal coordinates per sample were subjected to a non-linear dimensionality reduction using UMAP reduction. Clusters of cells sharing similar transcriptomic signal were identified using the Louvain algorithm, and clustering resolutions varied between 0. 6 and 1.2 based on the number and variety of cells obtained in the datasets.

Data integration and Batch correction for the scRNA seq validation dataset

The individual processed objects per library were normalized, and variance stabilized using negative binomial regression via the scTransform method offered by Seurat (Hafemeister and Satija, 2019). Counts matrices were merged into a single Seurat object and the batch (or library) of origin was stored in the metadata of the object. The log-normalized counts were reduced to a lower dimension using PCA and the individual libraries were aligned in the shared PCA space in a batch-aware manner (Each individual library was considered a batch) using the Harmony algorithm (Korsunsky et al., 2019). The resulting Harmony components were used to generate a batch corrected UMAP, and to identify clusters of transcriptionally similar cells across each of the 11 samples. The 5 Genes signature identified in the discovery dataset from the melanoma tumor sample was used to identify monocytes, macrophages, cDC2 and cDC1 respectively.

Single Cell RNAseq intra-sample heterotypic doublet detection:

All libraries were further processed to identify heterotypic doublets arising from the 10X sample loading. Processed, annotated Seurat objects were processed using the DoubletFinder package (McGinnis et al., 2019). Briefly, the cells from the object are modified to generate artificial duplicates, and true doublets in the dataset are identified based on similarity to the artificial doublets in the modified gene space. The prior doublet rate per library was approximated using the information provided in the 10x knowledgebase (https://kb.10xgenomics.com/hc/en-us/articles/360001378811) and this was corrected to account for homotypic doublets using the per-cluster numbers in each dataset.

Imaging bioinformatics

Each image was processed and analyzed in Imaris v9.5 (Bitplane Inc.). We first applied a median filter with size 3x3x1 to every channel of the image. The background of each channel was then subtracted using a filter width of 100um. Next, we used the Spots function to detect cell nuclei in the DAPI channel. We then applied a binary threshold on the max intensity of the CD45 channel to classify each spot as positive or negative for CD45.

We then cropped out a 650um x 650um section containing the tumor border from each histological image. This same section of tissue was found in the immunofluorescence image by looking for tissue features shared between the Immunofluorescence (IF) and histological images. Using the CD45 counts described above, two representative images from each immune class were selected.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis using 3 Features – Tcell, Myeloid, CD90+ CD44+ Stroma

In this part we will describe how we calculated the three first features used to perform the KNN clustering namely: Tcell, Myeloid, CD90+ CD44+ Stroma.

Gene Signature

Gene signatures for each of the features (Table S4) were generated by performing differential gene expression (DEG) analysis, between the different sorted RNAseq compartments, using edgeR (Robinson et al., 2010), limma (Ritchie et al., 2015) and Voom (Law et al., 2014), a function of limma that modifies RNAseq data for use with limma. Expression counts for all EHK10 samples in each of the compartments (tcell, myeloid, stroma, treg and tumor) were selected to ensure the DEG analysis was done using the highest quality data. The median adjusted p value for these DEGs were values much less than 1.00×10^{-55} across all comparisons. For the Tcell feature gene signature, DEG analysis was done between the tcell compartment and the stroma, myeloid, and tumor compartments separately. The intersection of the top 50 genes by log fold change (logFC) for each of these three comparisons was designated the Tcell feature gene signature. Similarly, a CD90+ CD44+ Stroma feature gene signature was designated as the intersection between DEG comparisons of the stroma compartment versus the myeloid, tumor and tcell compartments. For the Myeloid feature gene signature, it was necessary to use the top 100 genes by logFC from the DEG between the myeloid compartment and tcell

compartment to compensate for the fact that myeloid cells are more auto fluorescent and thus more likely to contaminate the tcell compartment sort. The stroma and tumor compartments are sorted from CD45- and are less likely to experience this issue. The process of feature gene signature selection for each of the three features was visualized with volcano plots that showed the DEG results and Venn diagrams that showed the intersection of the top genes from each DEG and the resulting feature gene signature.

Flow Score

The flow scores for the flow cytometry population ratios were calculated by scaling the ratios from 1-100. This was done by calculating the percentile rank of each population ratio relative to the other population ratios. For example, a percentile rank of 65 means that that 65% of the population ratios are below that ratio. This array of percentile ranks was then further transformed to percentile ranks to increase the spread in the scores. Boxplots were generated to visualize the scores by cancer type using Seaborn a python plotting module.

Gene Signature Score

The feature gene signature scores were calculated using an m x n matrix where m represented the TMM (Robinson and Oshlack, 2010) normalized logCPM (log2 counts per million) expression of the feature signature genes and n represented the selected sample set. TMM normalization is a method that adjusts library size so samples can be compared, since sequencing depth can differ between samples.

The expression of each gene was converted to percentile ranks across the samples using the SciPy (Virtanen et al., 2020) Python module.

$$E \rightarrow P$$

where: E = m x n matrix of gene expression
p = m x n matrix of gene expression percentile ranks across samples
A score was generated for each sample n follows:

$$Score_n = percentile(Average(P_n))$$

where: P_n = column n of P corresponding to sample n
(The second percentile transformation was employed to increase the spread between the scores).

Boxplots were generated to visualize the distribution of feature scores by cancer type with statistical analysis between archetypes performed using the statannot python package. We used the Mann-Whitney test between two archetypes, as we could not assume normal distributions, with the Bonferroni correction for multiple comparisons. The significance levels were annotated on the boxplots using the P value annotation legend below.

Annotation	P value Range
ns	$5.00e\text{-}02 < p \leq 1.00e\text{+}00$
*	$1.00e\text{-}02 < p \leq 5.00e\text{-}02$
**	$1.00e\text{-}03 < p \leq 1.00e\text{-}02$
***	$1.00e\text{-}04 < p \leq 1.00e\text{-}03$
****	$p \leq 1.00e\text{-}04$

For the IPI gene signature scores for the Tcell, Myeloid, CD90+ CD44+ Stroma were calculated using the live compartment. For the TCGA cohort the gene signature scores for the Tcell, Myeloid and CD90+ CD44+ Stroma features were calculated for all tumor type. Cross-Whisker plots, that show the median value, the interquartile range (IQR) in both the x and y directions and the identity line, were made to compare the median Tcell, Myeloid and CD90+ CD44+ Stroma feature scores in the IPI cohort with the TCGA cohort. Based on the Cross-Whisker plots only the BLCA, COAD, UCS, UCEC, OV, HNSC, KIRC, SARC and SKCM, cancer type were used for further analysis as their, median scores per tumor type appeared to correlate well with the IPI patient samples. The rationale behind including only tumor type that correlated well with IPI was to ensure that further analysis in TCGA would have extrapolated relevance in the IPI patient samples. The excluded cancer type lack of correlation with IPI patient samples could be attributed to variety of reasons that cannot be easily controlled, e.g., technical variability in upstream processing steps and cohort race differences due to local demographic biases.

Gene Signature Score Validation

Based on the assumption that the flow cytometry population ratios represent the true feature abundance in the patient sample, correlation between flow cytometry population ratios and feature gene signatures were used to validate the feature gene signatures. On average 9 to 10 samples per cancer type, that had both flow cytometry data and RNaseq data in the relevant compartments, were selected (Table S3). There were fewer samples selected in PDAC (Pancreatic), PNET(Neuroendocrine) and GBM(Glioblastoma) cancer type as there were not many representative samples for theses cancer in the IPI cohort. Correlation plots were generated using the scores and their corresponding flow population ratios and the Spearman's rho was calculated for each of the correlations.

Additionally, the Tcell and Myeloid feature samples were submitted to CIBERSORT (Newman et al., 2015). The TPMs of all protein coding genes of the samples were input as the mixture file, the built in LM22 gene signatures were used and quantile normalization was disabled as this was the recommended setting for RNaseq data. The CIBERSORT score used for the Tcell feature was the summation of the T cells CD8, T cells CD4 naive, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs) and T cells gamma delta cell type relative fractions. The CIBERSORT score used for the Myeloid feature was the summation of the Macrophages M0, Macrophages M1, Macrophages M2, Dendritic cells resting, Dendritic cells activated, Mast cells resting, Mast cells activated, Eosinophils and Neutrophils cell type relative fractions. Correlation plots were generated using these CIBERSORT scores and their corresponding flow population ratios and the Spearman's rho was calculated for each of the correlations.

Clustering

The feature gene signature scores in the IPI cohort were clustered using SCANPY (Wolf et al., 2018) a Python-based toolkit. The feature clustering was done in three steps. In the first step, a neighborhood graph was constructed using the k-nearest neighbors (KNN) algorithm with Euclidean distances. In the second step, the neighborhood graph was clustered using the Louvain method (Blondel et al., 2008), a community detection algorithm that maximizes network modularity. In the third step, the resulting clustering was evaluated using the Davies Bouldin Index (DBI) (Davies and Bouldin, 1979), a metric that assesses the ratio of the intra-cluster distance to inter-cluster distance. The lower the DBI the better the separation between clusters, the more compact the samples within clusters and hence the better the clustering. The clustering steps were repeated incrementally over values of n between 3 to 300 for the nearest neighbors for KNN. At each iteration of n, the values for the resolution parameter for Louvain clustering were varied between 0.3 to 2 in 0.1 increments. The formula below was used to optimize the nearest neighbor and resolution parameters for final cluster selection.

$$min[DBI(Cluster_{n,res})]$$

Where: n = [3,4,5...,300] (KNN nearest neighbor)
  res = [03,0.4,0.5...,2] (resolution)

In the TCGA cohort the feature signature score clustering was done as above except for the values for the range of n for KNN which was 3-1,500 as the number of samples in the TCGA cohort were greater.

The final clustering selections were given archetype labels based on the prevalence and distributions of the features within the cluster which were visualized as violin plots of all clusters in each feature. Both the IPI and TCGA final clustering selections had six clusters and hence six archetypes. The final clustering selection (IPI-was visualized as UMAP (McInnes et al., 2018), a dimensionality reduced projection of the clustering. The min-dist UMAP parameter, which is the minimum distance between points, controls how close together points are placed in the low-dimensional space of the UMAP was set as of 0.25. Population flow proportions and phenotype gene signatures were overlayed on the UMAP clustering.

Univariate survival analysis on the TCGA data was performed per cancer type and multivariate survival regression was performed across all cancer types using cancer types as a covariates using lifelines. Pie charts were made to compare the archetype abundance per tumor type between IPI and TCGA. Additionally, 3D plots of IPI score and TCGA score visualized the distribution of the unclustered scores.

Phenotype Gene Signatures

The Chemokine gene signature (Table S5) was a selection of 39 chemokines (Nagarsheth et al., 2017) of interest and was evaluated in the live compartment to generate a score. Hierarchical clustered heatmaps were made to assess chemokine expression in archetypes. The median TPM expression per archetype of each of the 39 chemokines in the Chemokine gene signature were clustered using the Euclidean distance metric and the Ward linkage method. A bubble plot was made to compare the median TPM expression of Chemokine gene signature genes in IPI and TCGA in each archetype. The bubble size represented values of the median TPM in each archetype transformed to a z-score.

Analysis using 6 Features – Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8, Tregs
Gene Signature – CD4, CD8 and Tregs

CD4 high and CD8 high samples were identified based on the flow proportion of (CD4+, CD25-, FOXP3-) / CD3+ and (CD8+, CD4-) / CD3+ respectively. There were nine samples that were both EHK10 in the tcell compartment of RNaseq and had a (CD4+, CD25-, FOXP3-) / CD3+ flow proportion greater than 60%, these were labeled CD4 high. Similarly, there were 12 samples that were EHK10 in the tcell compartment of RNaseq that had a (CD8+, CD4-) / CD3+ flow proportion greater than 60%, these were labeled CD8 high. The CD4 gene signatures were generated by performing DEG analyses using edgeR and voom on the tcell compartment counts of CD4 high samples versus CD8 high samples. The top 50 genes by adjusted P value were selected and all genes within this gene set that had a positive logFC were designated the CD4 feature gene signature. The same was done for the CD8 feature gene signature except the DEG was CD8 high samples versus CD4 high samples (Table S4). A hierarchical clustered heatmap was made using the CD4 and CD8 gene signature logCPM values for all tcell compartment patient samples. The logCPM values were mean-centered and scaled prior to clustering using the Euclidean distance metric and the Ward linkage method.

The Treg feature gene signature followed all the same steps as the Tcell feature gene signature but with four DEG comparisons instead of three, the treg compartment versus the tcell, myeloid, stroma and tumor compartments. The resulting intersection of these four comparisons was designated the Treg feature gene signature.

Gene Signature Score

Gene signature scores for the three new features were calculated as described above for three features. The CD4 and CD8 gene signatures were evaluated in the tcell compartment and the Treg gene signature was evaluated in the live compartment. Correlation between flow cytometry population ratios and feature gene signatures were plotted used to validate the feature gene signatures.

Calculated Score for Missing Data

The IPI cohort consists of 260 patient samples that all have a live compartment sequenced. However, there are only 199 overlapping patient samples with the tcell compartment sequenced. Since the CD4 and CS8 gene signature score were evaluated in the tcell compartment a need for a method to calculate scores for missing data was warranted. Gene signature scores were calculated for missing patient samples by leveraging the fact that there was very high correlation between gene signatures scores and their corresponding flow population ratios. For example, the correlation between the Tcell gene signature score and the CD3+ / live population ratio from flow cytometry had a Spearman's rho of 0.91. The missing scores were calculated by modeling the correlation using linear regression and using the regression equation and the flow population ratios to calculate the score. Of the 61 missing tcell compartment samples 5 did not have flow data.

Clustering

The feature gene signature scores for the six features were clustered as described above. The final clustering selections had eight clusters and hence eight archetypes. The final clustering was visualized as a UMAP. Population flow proportions and phenotype gene signatures were overlayed on the UMAP clustering.

Phenotype Gene Signatures

Hierarchical clustered heatmaps were made to assess chemokine expression in archetypes. The median TPM expression per archetype of each of the 39 chemokines in the Chemokine gene signature were clustered using the Euclidean distance metric and the Ward linkage method.

Analysis using 10 Features – Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8, Tregs, Macrophages, Monocytes, cDC1, cDC2

Gene Signature – Macrophages, Monocytes, cDC1 and cDC2

Gene signatures were generated from the discovery single cell RNaseq on tumor associated mononuclear phagocytes (MNP) dataset using a melanoma tumor sample. Gene signatures were generated using the FindMarkers/FindAllMarkers functions in Seurat. The genes that were considered significant had a log-fold change greater than 0.4, an adjusted p value less than or equal to 0.05 (based on Bonferroni correction) and were expressed in at least 35% of the groups. Cluster marker genes were identified as the top upregulated genes of the DEGs between clusters and curated based on their log-fold change as well as low expression by the other clusters. The different subsets of mononuclear phagocytic cells were identified by comparing cluster marker genes with public sources referenced in the text. The list of the different genes for each subset is listed in Table S4.

Gene Signature Score

Gene signature scores for the four new features were calculated as described above for three features. The gene signatures were evaluated in the myeloid compartment Correlation between flow cytometry population ratios and feature gene signatures were plotted used to validate the feature gene signatures.

Calculated Score for Missing Data

Of the of 260 patient samples that all have a live compartment sequenced there were only 189 overlapping patient samples with the myeloid compartment sequenced. The same method used to calculate missing data scores for the CD4 and CD8 gene signatures was used for missing data in the Macrophage, Monocyte, cDC1 and cDC2 gene signature. Of the 71 missing myeloid compartment samples 15 did not have flow data.

Clustering

The feature gene signature scores for the 10 features were clustered as described above. The final clustering selections had 12 clusters and hence 12 archetypes. The final clustering was visualized as a UMAP. Population flow proportions and phenotype gene signatures were overlayed on the UMAP clustering. Alluvial plots were made using RAWgraphs (Mauri et al., 2017) to visualize the stability of clusters as the clustering progressed from six features to ten features.

Phenotype Gene Signatures

Phenotype gene signatures (Table S5) were used to further characterize the immune archetypes obtained by clustering feature scores generated from the feature gene signatures. The phenotype gene signatures were assembled from standard gene signatures, derived from analysis of IPI samples or curated from gene sets found in literature.

The Tcell Exhaustion gene signature was generated by correlating the expression of all protein coding genes in EHK10 samples from the tcell compartment with the expression of five genes, CTLA4, PDCD1, CD38, HAVCR2 and LAG3, a subset of a published T cell exhaustion gene signature (Bengsch et al., 2018). The top 50 genes by Spearman's rho for each of the five correlations were selected as gene sets and genes that fell in the intersection of at least four of these gene sets were designated as the Tcell Exhaustion feature gene signature. The Spearman's rho of the CTLA4, PDCD1, CD38, HAVCR2 and LAG3 against the Tcell Exhaustion gene

signature genes with additional known exhaustion genes were plotted on heatmap to illustrate the gene signature selection criteria.

The ISG, Senescence, EMT, Fibrosis, Cell Stress DNA damage, Cell Cycle G1-S and Cell Cycle G2-M genes signatures were from literature (Kinker et al., 2019; Muñoz et al., 2019; Wiley et al., 2017). A bubble plot was made to visualize the median TPM expression per archetype in the tumor compartment. The bubble size and color represented values of the median TPM in each archetype, transformed to a z-score.

The NK cell gene signature was from literature (Barry et al., 2018), the B cells gene signature was from Chen et al. (2020) and van Galen et al. (2019), the Plasma cells gene signature was from Chen et al. (2020) and van Galen et al. (2019) and the Mast cell signature was from Cheng et al. (2021). A bubble plot was made to visualize the median TPM expression per archetype in the live compartment. The bubble size and color represented values of the median TPM in each archetype, transformed to a z-score.

The Th1, Th2, Th17, Trm and Tex gene signatures were from Kumar et al. (2017), Savas et al. (2018), Zhou et al. (2009), and Bengsch et al. (2018).

The Resting Tregs, Suppressive Tregs, Tissue Homing Tregs, Treg Activation, Treg Cytokines, Treg Metabolism were from Arce Vargas et al. (2018), Ephrem et al. (2013), Plitas et al. (2016), and Zemmour et al. (2018).

The M1, M2 and Costimulatory molecule gene signatures were from Biswas et al. (2013), Cassetta et al. (2019), Maier et al. (2020), and Roberts et al. (2016). A bubble plot was made to visualize the median TPM expression per archetype in the myeloid. The bubble size and color represented values of the median TPM in each archetype, transformed to a z-score.

The MHC Class I gene signature was from HGNC (https://www.genenames.org). The ISG gene signature was from Combes et al. (2021). A boxplot was made to visualize the scores per archetype that were calculated in the live compartment.

A hierarchical clustered heatmap was made to visualize the Th1, Th2, Th17, Trm, Tex, Resting Tregs, Suppressive Tregs, Tissue Homing Tregs, Treg Cytokines, Treg Activation, Treg Metabolism, NK cells, Plasma cells, B cells, M1, M2, Costimulatory molecule, and ISG gene signature gene expression. These genes for each of the phenotype gene signatures were assessed different RNaseq compartments as described above. The median TPM expression per archetype of each of the genes in the gene signatures were clustered using correlation distance metric and the average linkage method.

Immune Archetype Gene Signatures

Live. Archetype gene signatures were generated by DEG analysis, between the live compartment counts of samples in each of the 12 archetypes, using limma and Voom. The intersection between the top 3000 genes by logFC of each of 11 DEGs per archetype was assigned as an initial gene signature. If the initial gene signatures had less than 20 genes it was designated as the archetype gene signature. Otherwise, the archetype gene signature was designated as the top 20 genes with both the lowest coefficient of variation (CV) of the log10(TPM+0.001) expression and non-zero expression in at least 80% of the samples in the archetype. A hierarchical clustered heatmap was made to visualize archetype gene signature gene expression. The median TPM expression per archetype of each of the genes in the gene signatures were clustered using correlation distance metric and the average linkage method.

Tumor. Archetype gene signatures were generated by DEG analysis, between the combined tumor and epcam compartment counts of samples in each of the 12 archetypes, using edgeR and Voom. The intersection between the top 100 genes by logFC of each of 11 DEGs per archetype was assigned as an initial gene signature. If the initial gene signatures had less than 10 genes it was designated as the archetype gene signature. Otherwise, the archetype gene signature was designated as the top 10 genes with both the lowest coefficient of variation (CV) of the log10(TPM+0.001) expression and non-zero expression in at least 80% of the samples in the archetype. A hierarchical clustered heatmap using correlation distance metric and the average linkage method was made for the median TPM expression per archetype of each of the genes in the gene signatures.

Live Archetype Gene Signature Score

Using TCGA data, a gene signature score for each archetype was calculated using the gene signature score method above. Each sample has 12 scores for each archetype. Each sample was assigned to the archetype for which it had the highest score and if the highest score was tied between archetypes, the sample was excluded from the analysis.

Univariate survival analysis was performed per cancer type and multivariate survival regression was performed across all cancer types using cancer types as a covariates. The analysis was done using lifelines (Davidson-Pilon), a survival analysis Python library. If an archetype had less than fifteen representative samples, it was excluded from analysis. Pie charts were made to compare the archetype abundance per cancer type between IPI and TCGA.

ADDITIONAL RESOURCES

To facilitate usage of our data for the wide research community, we will provide access to transcriptomic and compositional data for each archetype: https://datalibrary.ucsf.edu/public-resources.

Supplemental figures

(legend on next page)

A.-Left- Sorting strategy for the 6 tumor associated cell compartments including All viable cells (black), T cells (Green), Tregs (Yellow),myeloid cells (Blue), CD90+,CD44+ Stromal cells (Red) and Tumor (Pink)-Right- 3D projection using K-means clustering on PC-1,2 and 3 of the top 500 genes expression per DEG pair (5000 genes total) (k = 5). B. Volcano plots, Venn diagrams and gene names showing the method of feature gene signature discovery for the Tcell, Myeloid and CD90+ CD44+ Stroma features using differential gene expression in tumor associated sorted compartments (see STAR Methods). C. Correlation plots of Tcell and Myeloid gene signature score against their corresponding flow population fraction (top) and CIBERSORT derived fraction (bottom) color-coded by tumor type. D. Box and whisker plots of Tcell, Myeloid and CD90+ CD44 Stroma features gene scores in the IPI cohort. E. Cross-whisker plots comparing median Tcell, Myeloid and CD90+ CD44+ Stroma gene signature scores by tumor type between the IPI and TCGA cohorts with interquartile range on both axes. Statistical difference between IPI and TCGA dataset within cancer type has been assessed by Wilcoxon-Mann-Whitney *p value < 0.05.
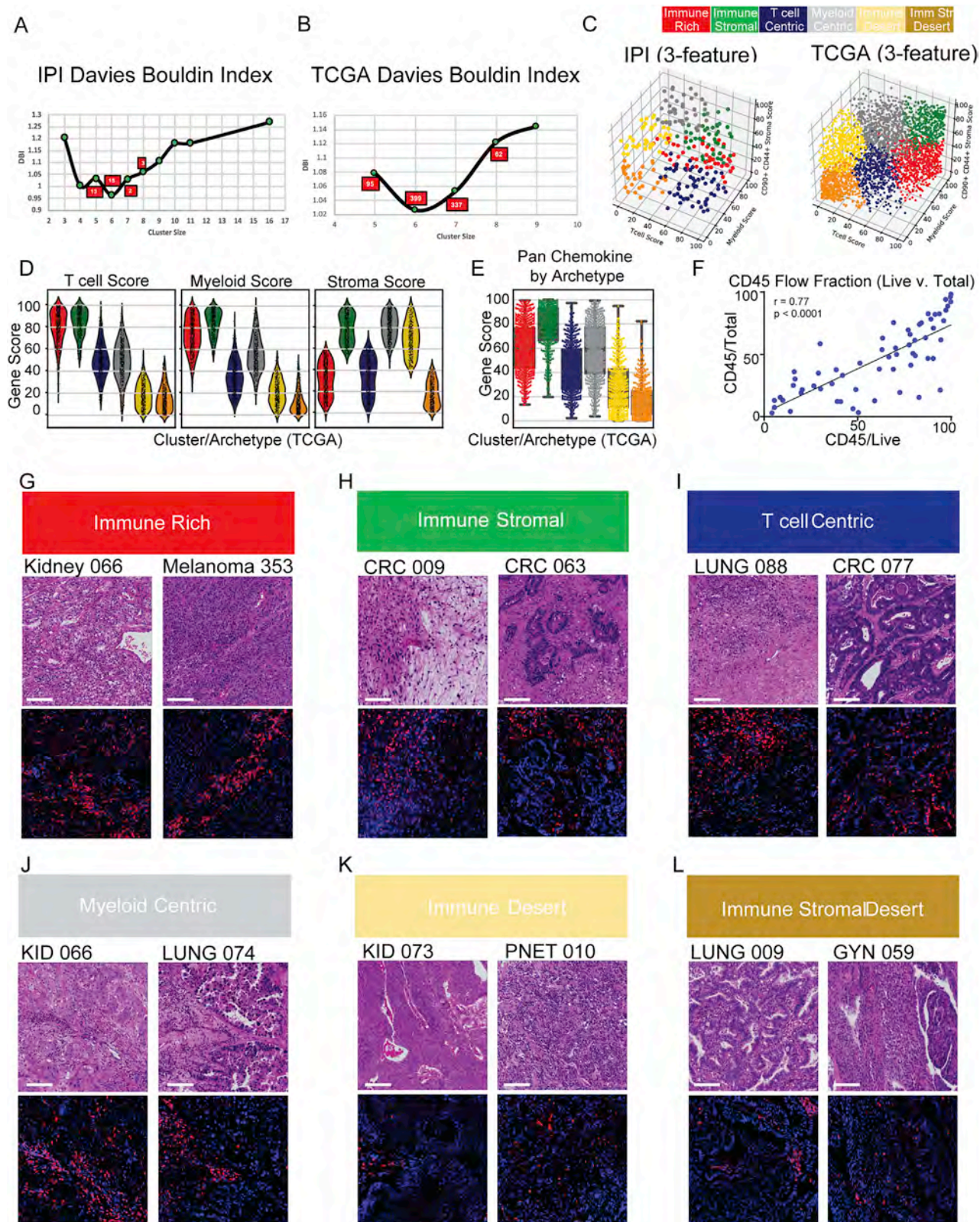
A IPI Davies Bouldin Index

B TCGA Davies Bouldin Index

C
| Immune Rich | Immune Stromal | T cell Centric | Myeloid Centric | Immune Desert | Imm Str Desert |

IPI (3-feature)   TCGA (3-feature)

D T cell Score   Myeloid Score   Stroma Score

Gene Score

Cluster/Archetype (TCGA)

E Pan Chemokine by Archetype

Gene Score

Cluster/Archetype (TCGA)

F CD45 Flow Fraction (Live v. Total)

r = 0.77
p < 0.0001

CD45/Total

CD45/Live

G Immune Rich

Kidney 066   Melanoma 353

H Immune Stromal

CRC 009   CRC 063

I T cell Centric

LUNG 088   CRC 077

J Myeloid Centric

KID 066   LUNG 074

K Immune Desert

KID 073   PNET 010

L Immune Stromal Desert

LUNG 009   GYN 059

(legend on next page)

A, B. Scatterplot of the Davies-Bouldin index and cluster size over multiple iterations of Louvain clustering and varying parameters using 3 features in the IPI (A) or TCGA (B) cohort. C. 3D plot of Tcell, Myeloid and CD90+ CD44+ Stroma scores color-coded by their cluster assignment from Louvain clustering these features in the IPI (left) and TCGA (right) cohorts. D. Violin plots of the Tcell, Myeloid and CD90+ CD44+ Stroma features for each cluster/archetype in TCGA cohort. E. Box and whisker plot of a pan chemokine gene score by cluster/archetype identified in TCGA cohort. F. Scatterplot of immune cell population fraction using only viable cells or total cells as denominator. G, H, I, J, K, L. Representative H&E (top) and immunofluorescence (bottom) images of tumor biopsies from lung, kidney skin, pancreas, uterus and colorectal tumor tissues using CD45 (red) and DAPI (blue) staining for each cluster/archetype identified in IPI cohort.
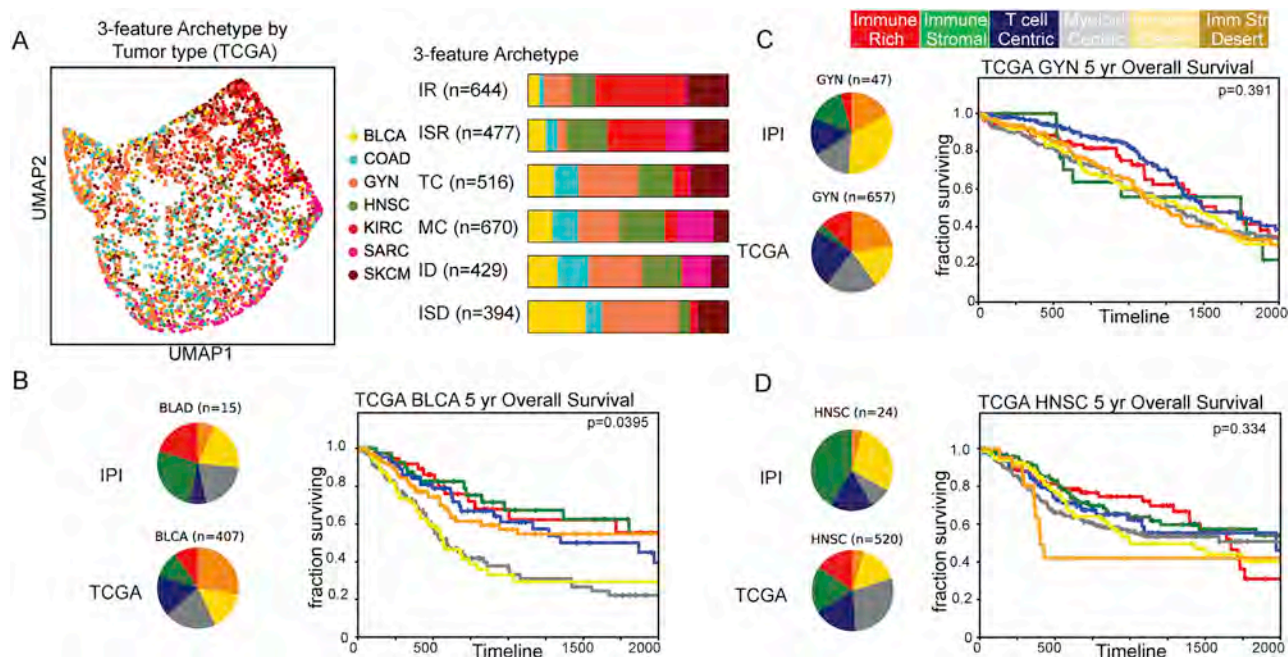
Figure S3. Coarse immune archetypes identified in TCGA are independent of tissue origin and associated to overall survival, related to Figure 3

A. Left-UMAP display and graph-based clustering of immune archetypes using 3-feature clustering within TCGA cohort color-coded by tumor type. Right stacked bar plot of the tumor type distribution for 3-feature archetypes in the TCGA cohort - B, C, D. (Left) Pie charts representing distribution of each archetype by cancer type from IPI (top) and TCGA (bottom) cohorts using 3 features. (Right) Kaplan-Meier overall survival curves for each immune archetype identified on TCGA dataset for Bladder urothelial carcinoma (B, BLCA), Gynecologic tumors (C,UCS + UCEC +OV),) and Head and Neck squamous cell carcinoma (D, HNSC).
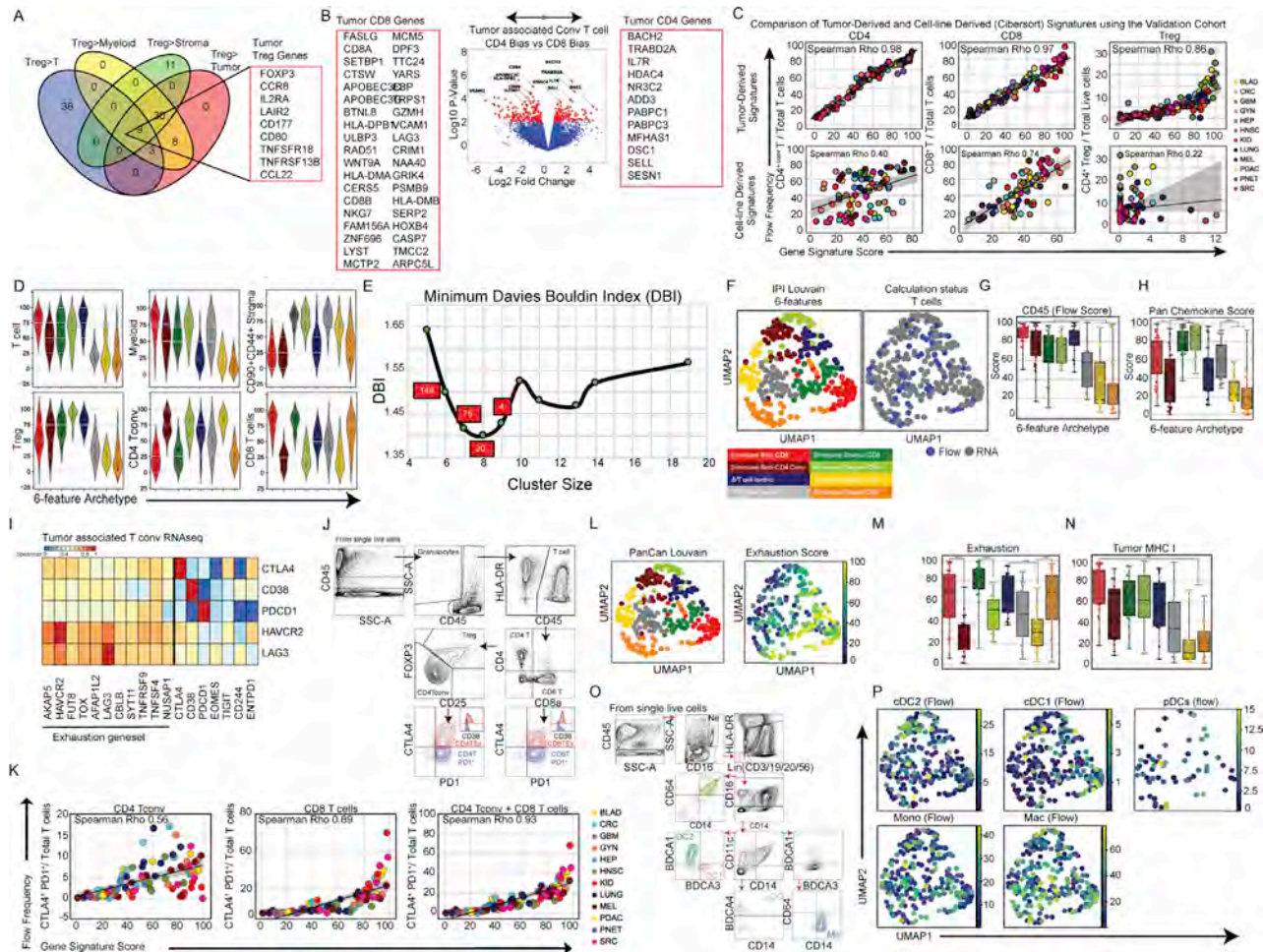
Figure S4. Generation and validation of CD4 CD8 and Treg signatures and further exploration of 6-feature archetypes, related to Figure 4

A. Venn diagram and gene names of the Treg (CD4+ regulatory T cell) feature gene signature score (see STAR Methods). B. Volcano plot visualizing differential gene expression between CD4+ or CD8+ T cells in the tcell compartment. Feature gene signatures are listed in red squares. C. Correlation plots of CD4, CD8 and Treg features versus their corresponding flow population fractions (top) and CIBERSORT derived fractions (bottom) color-coded by tumor type. D. Violin plots of the 6 features distributions in each cluster/archetype in the IPI cohort. E. Scatterplot of the Davies-Bouldin index and cluster size over multiple iterations of Louvain clustering and varying parameters using 6-feature clustering in the IPI cohort. F UMAP of immune archetypes using 6-feature clustering in the IPI cohort. (Right) UMAP overlay of samples that needed to calculate their CD4 and CD8 feature score flow cytometry population fractions (see STAR Methods).G, H. Box and whisker plot of total immune cell population fraction (G) and pan chemokine gene score (H) by 6-feature cluster/archetypes in the IPI cohort. I. Heatmap of the Spearman correlation coefficient of gene expression between CTLA4, CD38, PDCD1, HAVCR2 and LAG3 and our Exhaustion phenotype gene signature with additional genes associated with exhaustion used as a control. J. Gating strategy of tumor associated conventional CD4+ and CD8+ T cells and the markers used to define exhaustion. K. Correlation plots of individual and combined CD4+ PD1+ CTLA4+ and CD8+ PD1+ CTLA4+ conventional T cells population fractions against the Exhaustion phenotype gene signature score color-coded by tumor type. L. (Left) UMAP display and graph-based clustering of tumor immune archetypes using 6-feature clustering in the IPI cohort. (Right) UMAP overlay of the Exhaustion phenotype score calculated in the tcell compartment. M, N. Box and whisker plots of Exhaustion phenotype score (M) and MHC Class I phenotype score calculated in the tumor compartment (N) for each cluster/archetype identified in IPI cohort using 6-feature clustering. O. Gating Strategy for the quantification of the mononuclear phagocytic subsets (monocytes "Mo," Macrophages "Mp," Classical dendritic cell type 2 "DC2" and Type 1 "DC1," Plasmacytoid dendritic cells "pDCs" and Neutrophils "NE." P. UMAP overlay on the 6-feature clustering of classical dendritic cell type 1 (cDC1) and 2 (cDC2), plasmacytoid dendritic cells (pDCs), monocytes (Mono) and macrophages (Mac) frequencies measured by flow cytometry.
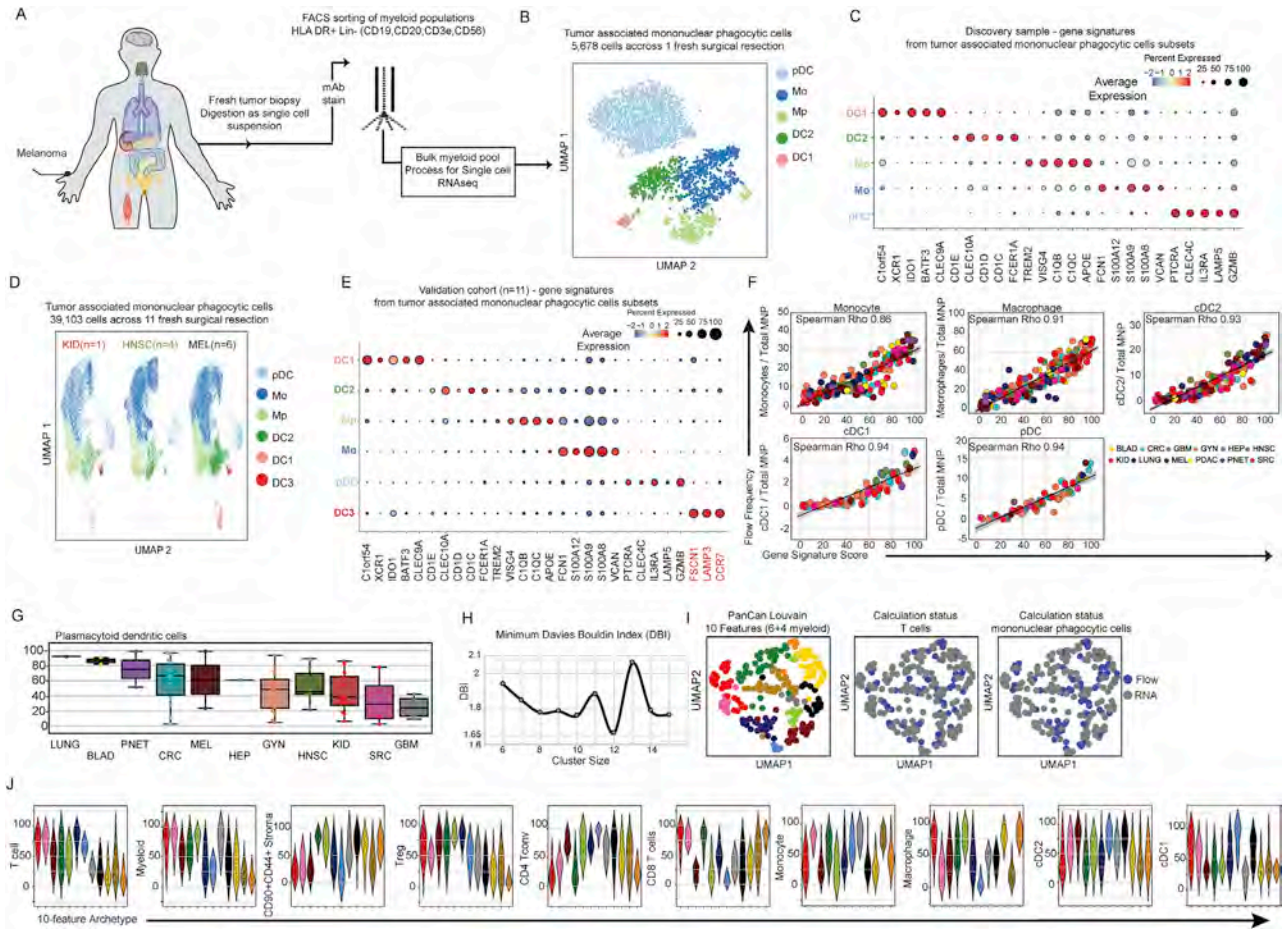
Figure S5. Single-cell RNA-seq-derived myeloid gene signatures refine immune archetypes, related to Figure 5

A. Details of the processing pipeline for digesting fresh tumor biopsies into single cell suspension, submitting to multi-parametric flow cell sorting of tumor associated myeloid population and encapsulation for single-cell RNA sequencing. B. UMAP and graph-based clustering of tumor associated myeloid cells from Melanoma processed for single-cell RNA sequencing. Each dot represents a cell. C. Dot plot of the top differentially-expressed-genes (DEG) between clusters identified in tumor associated mononuclear phagocytic cell (MNP) subsets in the discovery melanoma sample. D. UMAP and graph-based clustering of tumor associated myeloid cells from 11 different tumor resected tissues coming from 3 different cancer types (Kidney, Melanoma and Head and Neck) processed for single-cell RNA sequencing as depicted in A. Each dot represents a cell. E. Dot plot of the top differentially-expressed-genes (DEG) identified in C and top (DEG for DC3 cluster are colored in Red. F. Correlation plots of Macrophages, Monocytes, cDC1, cDC2 and pDCs gene scores against their corresponding flow population fractions in the myeloid compartment, color-coded by tumor type. G. Box and whisker plots gene score for pDCs (out of myeloid compartment) in IPI cohort. H. Scatterplot of the Davies-Bouldin index and cluster size over multiple iterations of Louvain clustering and varying parameters using 10-feature clustering in the IPI cohort. I. (Left) UMAP of tumor immune archetypes using 10-feature clustering in the IPI cohort. (Center-Right) UMAP overlay of samples that needed to calculate their CD4 and CD8 feature score and their Macrophages, Monocytes, cDC1 and cDC2 feature scores using flow cytometry population fractions (see STAR Methods). J. Violin plots of the 10 features in each cluster/archetype in the IPI cohort.
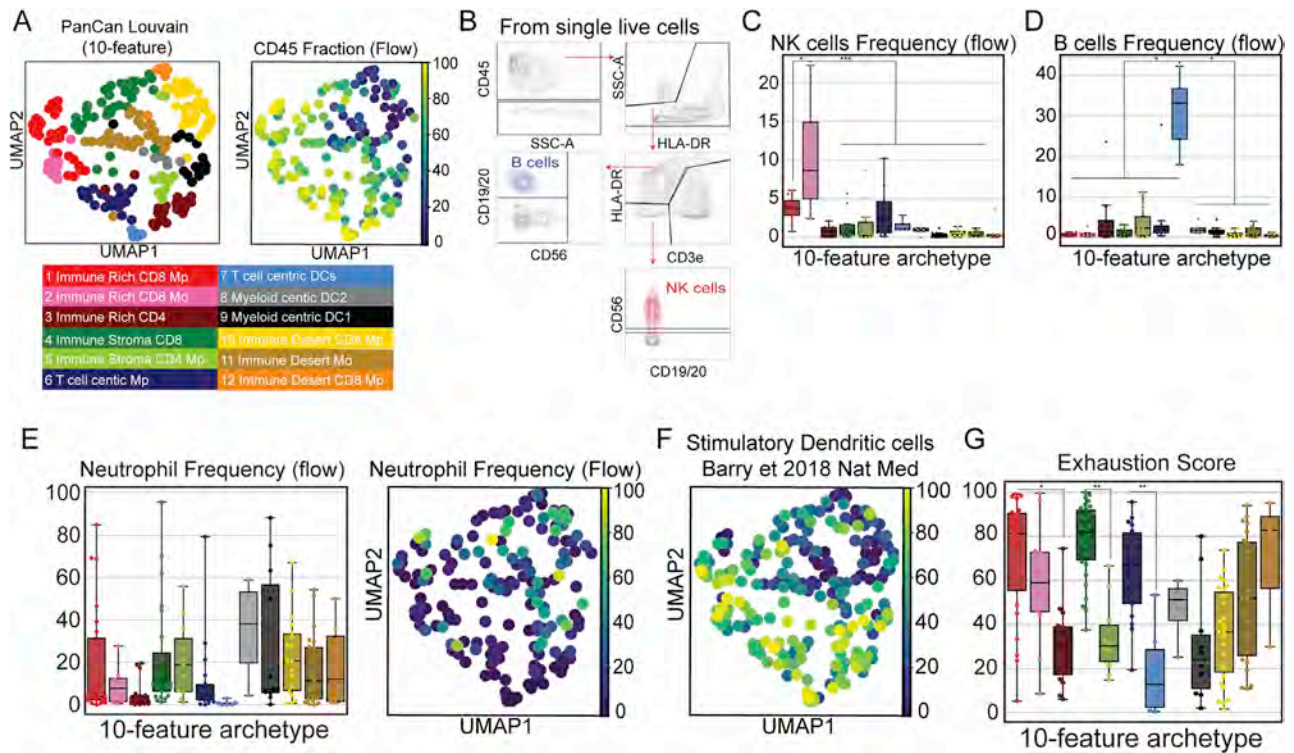
Figure S6. Immune cell frequency and transcriptomic profile pattern associated to each tumor archetype, related to Figure 6
A. (Left) UMAP display and graph-based clustering of tumor immune archetypes using 10-feature clustering in the IPI cohort. Each dot represents a single patient summarized by the 10 features. (Right) UMAP overlay of immune cell frequency measured by flow cytometry. B. Gating Strategy for the quantification of B cells (blue) and NK cells (red). C-D. Box and whisker plots of NK cells (C) and B cells (D) frequency out of all viable cells quantified by flow cytometry for each cluster/ archetype identified in IPI cohort. E. UMAP overlay of neutrophils frequency out the total immune cells measured by flow cytometry. F. UMAP overlay of a Stimulatory DC gene signature score measure on 'all viable cells' RNaseq compartment. G. Box and whisker plots of Exhaustion phenotpye score across 10-feature archetypes.

Legend (top color key):
1 Immune Rich CD8 Mp | 2 Immune Rich CD8 Mo | 3 Immune Rich CD4 | 4 Immune Stroma CD8 | 5 Immune Stroma CD4 Mp | 6 T cell centic Mp | 7 T cell centric DCs
8 Myeloid centic DC2 | 9 Myeloid centric DC1 | 10 Immune Desert Stroma | 11 Immune Desert Mo | 12 Immune Desert CD8 Mp

A — KID (n=47), KIRC (n=528); TCGA KIRC 5 yr Overall Survival (p:4.00E-3)
IR CD8 Mp MS:3554; IR CD8 Mo MS:4551; TC Mp MS:1238; TC DC MS:2090; MC DC2 MS:1913

B — CRC(n=28), COAD(n=286); TCGA COAD 5 yr Overall Survival (p:5.57E-1)
MC DC1 MS:1849; ID CD4 Mp MS:2475; ID Mo MS:2821

C — HEP(n=6), LIHC(n=369); TCGA LIHC 5 yr Overall Survival (p:4.13E-4)
IR CD8 Mp MS:1372; IR CD8 Mo MS:NA*; MC DC2 MS:2456; ID Mo MS:837

D — LUNG(n=27), LUAD(n=504); TCGA LUAD 5 yr Overall Survival (p:1.54E-2)
IR CD4 Mp MS:1501; IS CD4 Mp MS:2174; TC Mp MS:2617; TC DC MS:NA*; ID Mo MS:777

E — BLAD(n=14), BLCA(n=407); TCGA BLCA 5 yr Overall Survival (p:3.64E-1)
IS CD8 Mp MS:864; IS CD4 Mp MS:706; TC Mp MS:NA*; TC DC MS:1367; MC DC2 MS:1008; ID CD4 Mp MS:1348; ID Mo MS:2641

F — HNSC(n=21), HNSC(n=519); TCGA HNSC 5 yr Overall Survival (p:2.95E-1)
IS CD8 Mp MS:1430; IS CD4 Mo MS:1398; TC Mp MS:2083; ID CD4 Mp MS:2070; ID Mo MS:1134; ID CD8 Mp MS:915

G — GYN(n=43), GYN(n=656); TCGA GYN 5 yr Overall Survival (p:2.17E-2)
TC DC MS:737; MC DC1 MS:1499; ID CD4 Mp MS:1579; ID CD8 Mp MS:1384

H — MEL (n=22), SKCM (n=468); TCGA SKCM 5 yr Overall Survival (p:1.19E-6)
TC Mp MS:3379; TC DC MS:4930; ID CD8 Mo MS:1766

I — SRC(n=9), SARC(n=258); TCGA SARC 5 yr Overall Survival (p:7.28E-1)
IR CD8 Mp MS:2034; IS CD8 Mp MS:1552; TC Mp MS:2448; ID CD8 Mp MS:1941

J — PDAC(n=1), PAAD(n=177); TCGA PAAD 5 yr Overall Survival (p:8.79E-2)
IS CD8 Mp MS:498; TC DC MS:1332; MC DC1 MS:598

K — GBM(n=1), GBM(n=152); TCGA GBM 5 yr Overall Survival (p:8.68E-1)
IR CD8 Mp MS:454; IR CD8 Mo MS:623; ID CD8 Mp MS:442

(10-feature Archetypes) — UMAP1 / UMAP2

L — Immune tumor Archetype by Tumor Stage
IPI: T1 (n=50), T2 (n=46), T3 (n=77), T4 (n=36)
TCGA: T1 (n=875), T2 (n=858), T3 (n=762), T4 (n=579)

M — Immune tumor Archetype by Tumor Grade
IPI: G1 (n=32), G2 (n=58), G3 (n=96), G4 (n=12)
TCGA: G1 (n=177), G2 (n=872), G3 (n=999), G4 (n=93)

N — Immune tumor Archetype by Tumor Type
IPI: Primary (n=188), Metastatic (n=18)
TCGA: Primary (n=3953), Metastatic (n=371)

(legend on next page)