Title

Discovering dominant tumor immune archetypes in a pan-cancer census

Permalink

https://escholarship.org/uc/item/3j172337

Journal

Cell, 185(1)

ISSN

0092-8674

Authors

Combes, Alexis J
Samad, Bushra
Tsui, Jessica
et al.

Publication Date

2022

DOI

10.1016/j.cell.2021.12.004

Peer reviewed

Discovering Dominant Tumor Immune Archetypes in A Pan-Cancer Census

Alexis J. Combes[1,2,3,4,*,&], **Bushra Samad**[1,2,3,4,*], **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], **Matthew F. Krummel**[1,2,3,&,‡]

[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

[&]Correspondence: matthew.krummel@ucsf.edu and alexis.combes@ucsf.edu.
[*]These Authors contributed equally.
[‡]Lead contact and senior author.

AUTHOR CONTRIBUTIONS

AJC and BS contributed equally to this manuscript. AJC, BS, AB and MFK performed and/or provided supervision of experiments, generated, and analyzed data, and contributed to the manuscript by providing figures, tables and important intellectual contributions. AJC, BS and MFK 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. AJC, BS and MFK wrote and edited the manuscript. AJC, BS, SA, and AAR performed computational analyses of the data. AJB and MA performed the library RNA extraction and libraries construction for all the bulk RNA sequencing present in this manuscript. GKF, ABO and SA provided important intellectual input in genomics and statistical analysis for this study. AJC, JT, NWC, PY, GCR, DK, AS, BD, AB, KT, ES, PJ, JS, JH, JL, TC, KCB, RJA, TH, VC 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. RKK, JSC, CEA, AID, PH, AAD, JRK, and EAC, are part of the UCSF Immunoprofiler clinical team who was actively involved facilitation the access to the tumor specimens included in the IPI dataset. AJC, VC, KCD, DJE, GKF, SA and MFK 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. BS, AP, CC, JZ and KCD 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

MFK is a founder and shareholders of PIONYR immunotherapeutic and FOUNDERY innovations. AID is a shareholder of Trex and Neuvogen, he is also a member of the scientific advisory board of Neuvogen, Bristol-Myers and Squibb, Merck, Roche, Pfizer, Genentech, Incyte, Amgen, Novartis. EAC 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 stock ownership of 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.

[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

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 (TME) and modulating tumor-targeting. To discover the dominant immune system archetypes the 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.

Graphical Abstract
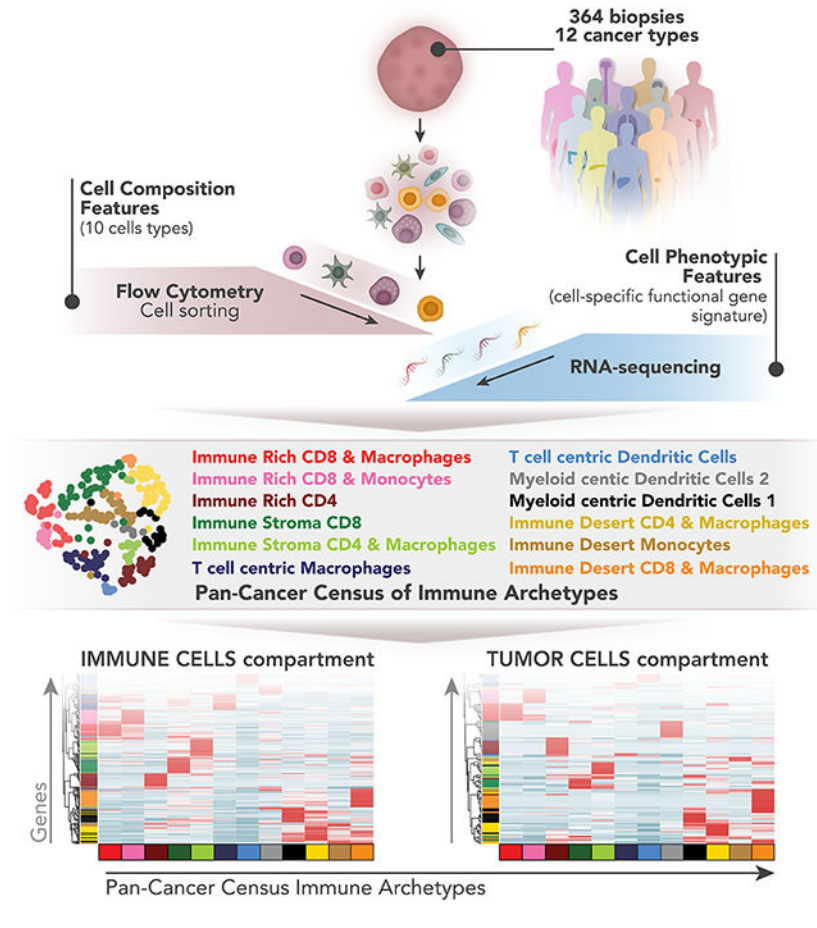
INTRODUCTION

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 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 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) have 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 delineate immune and tumor transcriptomic programs across different tissue type. 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, Supplementary Table 1). 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-sequencing 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 (Figure 1B and S1A left, and see STAR methods for full sort descriptions and potential caveats).This approach was chosen over single-cell RNA sequencing 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 (Tcell feature), myeloid cells (Myeloid feature) and non-immune $CD45^-$, $CD44^+$, $Thy1^+$ stromal cells ($CD44^+$ $CD90^+$ Stroma feature) (Figure S1A left and supplementary table 2). 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, and extended method) 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 which have linked compositional flow cytometry and RNAseq data, we discovered and then validated a set of gene signatures that would allow us to infer and compare compositional data from external RNAseq-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 Tcell, 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 (Figure 1D and Figure S1B). Use of this score, when applied to RNAseq data from the live RNAseq 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 tumor from different tissues by these gene scores showed similar trends, and strong correlations across individuals to those obtained by flow cytometric measurements (Figure 1C/S1D and 1E).

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 4341 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 abundances we describe by tumor type (e.g. 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), 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 type 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, Tcell, Myeloid and CD90+ CD44+ Stromal scores for all 260 samples with a "Live" bulk RNAseq compartment (Supplementary Table S2). The clustering was visualized using UMAP, both on the IPI and TCGA cohorts (Figure 2A/E and 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 (Figure S2A-B and 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 Tcell and Myeloid features and differentiated by Stromal enrichment in IS. This similarly differentiated two Immune desert (ID) clusters (Immune Desert and Immune Stromal Desert) defined by low expression of Tcell and Myeloid features, and again differentiated by enrichment of $CD44^+$ $CD90^+$ stroma within the Immune Desert cluster. Finally, we also identified clusters with enrichment in only one immune feature, namely the Tcell Centric (TC) and Myeloid Centric (MC) features (Figure 2B-D and Figure 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 (Figure 2E-F and Figure S2C-D).

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 immune rich clusters (1-3) had the highest $CD45^+$ infiltration, while immune desert samples had the lowest (Figure 2G-H and Figure S2G-L). Myeloid Centric (MC) clusters were similar to the immune desert clusters for overall CD45+ infiltration as assessed by immunofluorescence (Figure 2G-H and Figure S2G-L). 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.

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 immune rich clusters and a low score in immune desert cluster in both the IPI and TCGA cohorts (Figure 2K-L and Figure S2E). Two exceptions stood out using the pan-chemokine measure: Tcell and Myeloid Centric clusters showed decreased and increased pan-chemokine scores, respectively (Figure 2K-L). We thus examined chemokines individually, using hierarchical clustering of the median chemokine expression amongst the tumors of each archetype, and identified sets of chemokines associated with each cluster (Figure 2M-N). For instance, the three T cell enriched clusters are enriched in either chemokines expressed by T cell (XCL1, XCL2, CCL4) or T cell attracting chemokines (CXCL13, CCL5). Other archetypes had their own pattern of differentially expressed chemokines and these were broadly consistent with their composition. Immune desert 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 type, such as Kidney and Melanoma had significant biases (Figure 3A and Figure 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 immune rich 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 (Figure 3C and Figure S3B-D). While the relative composition of archetypes by tissue of origin were broadly similar in both cohorts, there was some variation e.g. heavily 'Tcell 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 such as melanoma, kidney, bladder, and colorectal appear to stratify survival outcome, 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 as we observed variability within each archetype (Figure 3D-E).

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) (supplementary table 3). Repeating our previous methodology (Figure S1), we used DGE between the Treg RNAseq compartment and the other cell sorted RNAseq 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} RNAseq compartment, we used the flow data to identify samples rich in CD4 or CD8 conventional T cells and performed DGE between them (Figure S4B and STAR methods). Notably, most of the identified genes have been previously associated with CD4+/CD8+ conventional T cell identity (e.g., CD8A, IL7R) or tissue residency (e.g., VCAM1, 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} RNAseq 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 (Figure 4C and Figure S4D-F). 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 Immune Rich and Immune Stromal Rich archetypes, now delineated by CD4 to CD8 ratio (Figure 4D-E and Figure S4E). This increase in feature granularity also resulted in some specimens shifting within the classification. For example, some Tcell 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 'Immune Desert' favored CD4 cells. CD45 densities remain high for the immune rich tumors (Figure S4G). Assessing chemokine gene expression showed that this re-clustering dramatically segregated chemokine gene expression in the Immune Rich archetype and further refined the chemokine expression found amongst the other archetypes (Figure 4F and Figure S4H)

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 and 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 immune desert (ID) archetype (Figure S4L-M) 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. (Figure 4G and Figure 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 scRNAseq sub-study 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 (Figure S5D-E). 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 RNAseq 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 and STAR methods). Again, when we

analyzed the distributions of these five MNP subset features scores by cancer type, a high heterogeneity was apparent (Figure 5A and Figure 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 (Supplementary Table 3). 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 RNAseq (Figure 5A and Figure S5H-J). As 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 (Figure 5B-C). MNP inclusion also resulted in some specimens shifting between MC and ID archetypes, driven by strong monocytes or macrophages 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 (Figure 5D-F and 5O-Q). One archetype, IR CD4 biased, is differentiated from its IR CD8 counterparts by enrichment in cDC2 vs cDC1 (Figure 5D and 5R-T). However, such correlation between CD4/CD8 and cDC2/cDC1 ratios is only observed among IR archetypes and not in archetypes containing high stromal densities (Figure 5G-I and 5R-T).

The relationship between monocytes/macrophages and T cell subsets composition also differed significantly amongst the 12 archetypes (Figure 5D-N). 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 (Figure 5N, 5P and 5T). TC (Tcell centric) archetypes are characterized by a positive correlation between Treg and Macrophage abundance (Figure 5J-L and 5P-Q). 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) (Figure 6A and S6B-C). 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) (Figure 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 RNAseq compartment to explore the level of expression of genes previously associated to 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 to 'M1' and 'M2' macrophage phenotype among archetypes we observed a general enrichment in 'M2' genes in IS archetypes (Figure 6C). On the other hand, 'DC' and 'Co-Stim' associated genes were highly expressed in the Tcell 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 a 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 RNAseq compartments corresponding to their appropriate cell type, namely 'Th1' genes in the T_{conv} RNAseq compartment, 'M1' genes in the myeloid RNAseq compartment and 'Treg homing' in the Treg RNAseq compartment (Figure 6D and Supplementary table 3). Combining the gene sets in this way revealed distinct immune signatures for each archetype (Figure 6D). For instance, while the IR CD8 macrophage bias archetype was enriched in genes associated with the type 1 response (IFNG, TNF, IL1B) both in T_{conv} and MNP compartments, both IS and ID CD8 archetypes were characterized by their unique combination of upregulated gene expression associated to T cell exhaustion (PDCD1, CTLA4, 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 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 identifies 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 (Figure S6A and S6E). A gene signature previously associated to 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 immune rich and immune desert archetypes (Figure S6A and S6F-G). Thus, while 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) (Figure 7A and Figure S7). Regardless of the frequent intra-cancer heterogeneity, we observed a strong relationship between tumor proliferative capacity, measured by the fraction of tumor cells which were Ki67$^+$ as measured by flow, and immune desert 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 immune deserts 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 RNAseq 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 correlated with each archetype (Figure 7E). This identified unique gene signatures for each of the 12 archetypes, within which we observed both immune (e.g. LAYN, CTLA4, CSF1, CCL19) and non-immune (CRHBP, PTGFR, HSPA2, MTCL1) related genes. 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 (Figure S7A-K). Additionally, we observed a similar distribution of archetypes in both cohorts when looking at clinical correlates such as tumor stage, grade, and metastatic status (Figure S7L-N). This suggests that our archetypes can be identified in independent datasets and 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 (Figure S7A-K). 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 to better survival (Median survival:3354 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 cell, TC archetypes characterized by an enrichment of Macrophages over monocytes displayed the higher median survival (2456 days, p-value 1.17E-6) while 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 (Supplementary table 6) 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 specimen from 364 patients. Empowered by complementary profiling assays we were able to discover 12 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 4096. While 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 which 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 subdivides both immune rich and immune desert 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 immune deserts 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 immune rich and immune desert 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 amongst cell densities and chemokine networks. While 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 (Supplementary table 6).

This idea is perhaps also suggested by the distinct patterns of chemokine expression identified for each tumor immune archetype. For example, while IR CD8 archetypes are defined by a CXCL9, CXCL10, CXCL11/CXCR3 axis which resembles the biology of an ongoing chronic viral infection (Metzemaekers et al., 2017), 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, 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 (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 the relationship between 'dominant' and 'reactive' tumor immune archetype across tumor types (Bi et al., 2021; Grünwald 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 (Figure 2H-I) 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 types 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, γδ 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 (Figure 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

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 supplementary 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 cystadenocarcinoma (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 perfomed after tissue enzymatic digestion and allow us to sequence only viable cells after isolation. Therefore it is slighty different than bulk RNAseq from whole tissue prior diggestion as it was done in the TCGA dataset.

Bulk RNAseq compartment	Gating Strategy	Potential Caveats
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 CD25 expression but as we can't use FOXP3 for sorting unpermeabilised 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 down-regulate 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 permessive 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 seconds. 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 & STATISTICAL ANALYSIS

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

Gene Signature: Gene signatures for each of the features (Supplementary Table 4) 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-02 < p <= 1.00e+00
*	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

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 (Supplementary Table 3). 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 naïve, 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 (Supplementary Table 5) 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 (Supplementary Table 4). 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.

Clusterin: 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 (Supplementary Table 4)

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 overlaid 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 (Supplementary Table 5) 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 (B 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, munoz dp et al, wiley cd et al). 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; van Galen et al., 2019) , the Plasma cells gene signature was from (Chen et al., 2020; 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)(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; 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; 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 box plot 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 RESSOURCES

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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

CONSORTIA

The UCSF Immunoprofiler Initiative (IPI) includes: 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, Hideho Okada, Margaret Tempero, Andrew Ko, Kim Kirkwood, Scott Vandenberg, Denise Guevarra, Erica Oropeza, Chris Cyr, Pat Glenn, Jennifer Bolen, Amanda Morton, Walter Eckalbar. Affiliations for Consortia members can be found in supplementary table S7.

REFERENCES

Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, Levenson RM, Lowe JB, Liu SD, Zhao S, et al. (2014). Multiplexed ion beam imaging of human breast tumors. Nat Med 20, 436–442. [PubMed: 24584119]

Aran D, Hu Z, and Butte AJ (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 18, 220. [PubMed: 29141660]

Arce Vargas F, Furness AJS, Litchfield K, Joshi K, Rosenthal R, Ghorani E, Solomon I, Lesko MH, Ruef N, Roddie C, et al. (2018). Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. Cancer Cell 33, 649–663.e4. [PubMed: 29576375]

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. [PubMed: 31835037]

Azizi E, Carr AJ, Plitas G, Cornish AE, 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. [PubMed: 29961579]

B B, T O, O K, M S, S M, S O, Pf G, Rs H, Ac H, Km C, et al. (2018). Epigenomic-Guided Mass Cytometry Profiling Reveals Disease-Specific Features of Exhausted CD8 T Cells (Immunity).

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. [PubMed: 34019806]

Barry KC, Hsu J, Broz ML, Cueto FJ, Binnewies M, Combes AJ, Nelson AE, Loo K, Kumar R, Rosenblum MD, et al. (2018). A natural killer-dendritic cell axis defines checkpoint therapy-responsive tumor microenvironments. Nat Med 24, 1178–1191. [PubMed: 29942093]

Beltra J-C, Manne S, Abdel-Hakeem MS, Kurachi M, Giles JR, Chen Z, Casella V, Ngiow SF, Khan O, Huang YJ, 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. [PubMed: 32396847]

Bengsch B, Ohtani T, Khan O, Setty M, Manne S, O'Brien S, Gherardini PF, Herati RS, Huang AC, 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. [PubMed: 29768164]

Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, Liu Y, Fan H, Shen H, Ravikumar V, et al. (2018). A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell 33, 690–705.e9. [PubMed: 29622464]

Bi K, He MX, Bakouny Z, Kanodia A, Napolitano S, Wu J, Grimaldi G, Braun DA, Cuoco MS, Mayorga A, et al. (2021). Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. Cancer Cell.

Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, 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. [PubMed: 24138885]

Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 24, 541–550. [PubMed: 29686425]

Binnewies M, Mujal AM, Pollack JL, Combes AJ, Hardison EA, Barry KC, Tsui J, Ruhland MK, Kersten K, Abushawish MA, et al. (2019). Unleashing Type-2 Dendritic Cells to Drive Protective Antitumor CD4+ T Cell Immunity. Cell 177, 556–571.e16. [PubMed: 30955881]

Biswas SK, Allavena P, and Mantovani A (2013). Tumor-associated macrophages: functional diversity, clinical significance, and open questions. Semin Immunopathol 35, 585–600. [PubMed: 23657835]

Blank CU, Haining WN, Held W, Hogan PG, Kallies A, Lugli E, Lynn RC, Philip M, Rao A, Restifo NP, et al. (2019). Defining "T cell exhaustion." Nat Rev Immunol 19, 665–674. [PubMed: 31570879]

Blondel VD, Guillaume J-L, Lambiotte R, and Lefebvre E (2008). Fast unfolding of communities in large networks.

Bosteels C, Neyt K, Vanheerswynghels M, van Helden MJ, 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. [PubMed: 32392463]

Böttcher JP, Bonavita E, Chakravarty P, Blees H, Cabeza-Cabrerizo M, Sammicheli S, Rogers NC, 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. [PubMed: 29429633]

Brown SD, Warren RL, Gibb EA, Martin SD, Spinelli JJ, Nelson BH, and Holt RA (2014). Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res 24, 743–750. [PubMed: 24782321]

Broz ML, Binnewies M, Boldajipour B, Nelson AE, Pollack JL, Erle DJ, Barczak A, Rosenblum MD, Daud A, Barber DL, et al. (2014). Dissecting the tumor myeloid compartment reveals rare activating antigen-presenting cells critical for T cell immunity. Cancer Cell 26, 638–652. [PubMed: 25446897]

Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517, 576–582. [PubMed: 25631445]

Cassetta L, Fragkogianni S, Sims AH, Swierczak A, Forrester LM, Zhang H, Soong DYH, Cotechini T, Anur P, Lin EY, 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. [PubMed: 30930117]

Chen DS, and Mellman I (2017). Elements of cancer immunity and the cancer-immune set point. Nature 541, 321–330. [PubMed: 28102259]

Chen B, Khodadoust MS, Liu CL, Newman AM, and Alizadeh AA (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol 1711, 243–259. [PubMed: 29344893]

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. [PubMed: 32580738]

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. [PubMed: 33545035]

Coffelt SB, Kersten K, Doornebal CW, Weiden J, Vrijland K, Hau C-S, Verstegen NJM, Ciampricotti M, Hawinkels LJAC, Jonkers J, et al. (2015). IL-17-producing γδ T cells and neutrophils conspire to promote breast cancer metastasis. Nature 522, 345–348. [PubMed: 25822788]

Combes A, Camosseto V, N'Guessan P, Argüello RJ, 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. [PubMed: 29030552]

Combes AJ, Courau T, Kuhn NF, Hu KH, Ray A, Chen WS, Chew NW, Cleary SJ, Kushnoor D, Reeder GC, et al. (2021). Global absence and targeting of protective immune states in severe COVID-19. Nature.

Davidson-Pilon C, Kalderstam J, Jacobson N, sean-reed, Kuhn B, Zivich P, Williamson M, Abdeali JK, Datta D, Fiore-Gartland A, et al. (2020). CamDavidsonPilon/lifelines: v0.25.6.

Davies DL, and Bouldin DW (1979a). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1, 224–227.

Davies DL, and Bouldin DW (1979b). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1, 224–227.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Dominguez D, Tsai Y-H, Gomez N, Jha DK, 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. [PubMed: 27364684]

Duan Q, Zhang H, Zheng J, and Zhang L (2020). Turning Cold into Hot: Firing up the Tumor Microenvironment. Trends Cancer 6, 605–618. [PubMed: 32610070]

Dvorak HF (1986). Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N Engl J Med 315, 1650–1659. [PubMed: 3537791]

Ephrem A, Epstein AL, Stephens GL, Thornton AM, Glass D, and Shevach EM (2013). Modulation of Treg cells/T effector function by GITR signaling is context-dependent. Eur J Immunol 43, 2421–2429. [PubMed: 23722868]

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

Gajewski TF, 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. [PubMed: 23579075]

van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, Verga JA, Stephansky J, Pastika TJ, Lombardi Story J, et al. (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell 176, 1265–1281.e24. [PubMed: 30827681]

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. [PubMed: 30610226]

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. [PubMed: 17008531]

Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 21, 938–945. [PubMed: 26193342]

Gerlach C, Moseman EA, Loughhead SM, Alvarez D, Zwijnenburg AJ, Waanders L, Garg R, de la Torre JC, and von Andrian UH (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. [PubMed: 27939671]

Ghosh M, Saha S, Bettke J, Nagar R, Parrales A, Iwakuma T, van der Velden AWM, and Martinez LA (2021). Mutant p53 suppresses innate immune signaling to promote tumorigenesis. Cancer Cell.

Glatman Zaretsky A, Konradt C, Dépis F, Wing JB, Goenka R, Atria DG, Silver JS, Cho S, Wolf AI, Quinn WJ, et al. (2017). T Regulatory Cells Support Plasma Cell Populations in the Bone Marrow. Cell Rep 18, 1906–1916. [PubMed: 28228257]

Goldman MJ, Craft B, Hastie M, Repe ka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol 38, 675–678. [PubMed: 32444850]

Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, and Kurzrock R (2017). Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther 16, 2598–2608. [PubMed: 28835386]

Goswami S, Walle T, Cornish AE, Basu S, Anandhan S, Fernandez I, Vence L, Blando J, Zhao H, Yadav SS, et al. (2020). Immune profiling of human tumors identifies CD73 as a combinatorial target in glioblastoma. Nat Med 26, 39–46. [PubMed: 31873309]

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. [PubMed: 28338065]

Gubin MM, Esaulova E, Ward JP, Malkova ON, Runci D, Wong P, Noguchi T, Arthur CD, 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. [PubMed: 30343900]

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.

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. [PubMed: 31870423]

Hanahan D, and Weinberg RA (2011). Hallmarks of cancer: the next generation. Cell 144, 646–674. [PubMed: 21376230]

Hegde PS, 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. [PubMed: 27084740]

Hu KH, Eichorst JP, McGinnis CS, Patterson DM, Chow ED, Kersten K, Jameson SC, Gartner ZJ, Rao AA, and Krummel MF (2020). ZipSeq: barcoding for real-time mapping of single cell transcriptomes. Nat Methods 17, 833–843. [PubMed: 32632238]

Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, 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. [PubMed: 26997480]

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. [PubMed: 15611321]

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. [PubMed: 30193111]

Khan O, Giles JR, McDonald S, Manne S, Ngiow SF, Patel KP, Werner MT, Huang AC, Alexander KA, Wu JE, et al. (2019). TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. Nature 571, 211–218. [PubMed: 31207603]

Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, Warren A, Rodman C, Roth JA, Bender SA, et al. (2019). Pan-cancer single cell RNA-seq uncovers recurring programs of cellular heterogeneity. BioRxiv 807552.

Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, Warren A, Rodman C, Roth JA, Bender SA, et al. (2020). Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. Nat Genet 52, 1208–1218. [PubMed: 33128048]

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. [PubMed: 31740819]

Kumar BV, Ma W, Miron M, Granot T, Guyer RS, Carpenter DJ, 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. [PubMed: 28930685]

Lavin Y, Kobayashi S, Leader A, Amir E-AD, Elefant N, Bigenwald C, Remark R, Sweeney R, Becker CD, Levine JH, et al. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. Cell 169, 750–765.e17. [PubMed: 28475900]

Law CW, Chen Y, Shi W, and Smyth GK (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15, 1–17.

Leach DR, Krummel MF, and Allison JP (1996). Enhancement of antitumor immunity by CTLA-4 blockade. Science 271, 1734–1736. [PubMed: 8596936]

Li B, and Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323. [PubMed: 21816040]

Lopes N, McIntyre C, Martin S, Raverdeau M, Sumaria N, Kohlgruber AC, Fiala GJ, Agudelo LZ, 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. [PubMed: 33462452]

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. [PubMed: 30201786]

Maier B, Leader AM, Chen ST, Tung N, Chang C, LeBerichel J, Chudnovskiy A, Maskey S, Walker L, Finnigan JP, et al. (2020). A conserved dendritic-cell regulatory program limits antitumour immunity. Nature 580, 257–262. [PubMed: 32269339]

Mandal R, enbabao lu Y, Desrichard A, Havel JJ, Dalin MG, Riaz N, Lee K-W, Ganly I, Hakimi AA, Chan TA, et al. (2016). The head and neck cancer immune landscape and its immunotherapeutic implications. JCI Insight 1, e89829. [PubMed: 27777979]

Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE, Koeppen H, Astarita JL, Cubas R, et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 554, 544–548. [PubMed: 29443960]

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, (Cagliari, Italy: ACM Press), pp. 1–5.

Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, Yu EA, Schenk EL, 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. [PubMed: 32822576]

McGinnis CS, Murrow LM, and Gartner ZJ (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst 8, 329–337.e4. [PubMed: 30954475]

McInnes L, Healy J, and Melville J (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

Metzemaekers M, Vanheule V, Janssens R, Struyf S, and Proost P (2017). Overview of the Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-Inducible CXC Chemokine Receptor 3 Ligands. Front Immunol 8, 1970. [PubMed: 29379506]

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. [PubMed: 30013147]

Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D, Church SE, 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. [PubMed: 26982367]

Molgora M, Esaulova E, Vermi W, Hou J, Chen Y, Luo J, Brioschi S, Bugatti M, Omodei AS, Ricci B, et al. (2020). TREM2 Modulation Remodels the Tumor Myeloid Landscape Enhancing Anti-PD-1 Immunotherapy. Cell 182, 886–900.e17. [PubMed: 32783918]

Mondini M, Loyher P-L, Hamon P, Gerbé de Thoré M, Laviron M, Berthelot K, Clémenson C, Salomon BL, Combadière C, Deutsch E, et al. (2019). CCR2-Dependent Recruitment of Tregs and Monocytes Following Radiotherapy Is Associated with TNFα-Mediated Resistance. Cancer Immunol Res 7, 376–387. [PubMed: 30696630]

Mujal AM, and Krummel MF (2019). Immunity as a continuum of archetypes. Science 364, 28–29. [PubMed: 30948539]

Mulder K, Patel AA, Kong WT, Piot C, Halitzki E, Dunsmore G, Khalilnezhad S, Irac SE, 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. [PubMed: 34331874]

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

Nagarsheth N, Wicha MS, and Zou W (2017). Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. Nat Rev Immunol 17, 559–572. [PubMed: 28555670]

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, and Alizadeh AA (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12, 453–457. [PubMed: 25822800]

Oh DY, Kwek SS, Raju SS, Li T, McCarthy E, Chow E, Aran D, Ilano A, Pai CCS, Rancan C, et al. (2020). Intratumoral CD4+ T Cells Mediate Anti-tumor Cytotoxicity in Human Bladder Cancer. Cell 181, 1612–1625.e13. [PubMed: 32497499]

Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, 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. [PubMed: 34290406]

Petitprez F, de Reyniès A, Keung EZ, Chen TW-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. [PubMed: 31942077]

Plitas G, Konopacki C, Wu K, Bos PD, Morrow M, Putintseva EV, Chudakov DM, and Rudensky AY (2016). Regulatory T Cells Exhibit Distinct Features in Human Breast Cancer. Immunity 45, 1122–1134. [PubMed: 27851913]

Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, Foye A, Kothari V, Perry MD, Bailey AM, et al. (2018). Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell 174, 758–769.e9. [PubMed: 30033370]

Reynolds G, Vegh P, Fletcher J, Poyner EFM, Stephenson E, Goh I, Botting RA, Huang N, Olabi B, Dubois A, et al. (2021). Developmental cell programs are co-opted in inflammatory skin disease. Science 371, eaba6500. [PubMed: 33479125]

Richer MJ, Lang ML, and Butler NS (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. [PubMed: 27496973]

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47–e47. [PubMed: 25605792]

Roberts EW, Broz ML, Binnewies M, Headley MB, Nelson AE, Wolf DM, Kaisho T, Bogunovic D, Bhardwaj N, and Krummel MF (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. [PubMed: 27424807]

Robinson MD, and Oshlack A (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 11, R25. [PubMed: 20196867]

Robinson MD, McCarthy DJ, and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. [PubMed: 19910308]

Rooney MS, Shukla SA, Wu CJ, Getz G, and Hacohen N (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell 160, 48–61. [PubMed: 25594174]

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. [PubMed: 27096321]

Sancho D, Joffre OP, Keller AM, Rogers NC, 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. [PubMed: 19219027]

Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, Salgado R, Byrne DJ, Teo ZL, Dushyanthen S, et al. (2018). Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat Med 24, 986–993. [PubMed: 29942092]

Scott AC, Dündar F, Zumbo P, Chandran SS, Klebanoff CA, 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. [PubMed: 31207604]

Slyper M, Porter CBM, 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. [PubMed: 32405060]

Spranger S (2016). Tumor Heterogeneity and Tumor Immunity: A Chicken-and-Egg Problem. Trends Immunol 37, 349–351. [PubMed: 27166403]

Spranger S, Luke JJ, Bao R, Zha Y, Hernandez KM, Li Y, Gajewski AP, Andrade J, and Gajewski TF (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 U S A 113, E7759–E7768. [PubMed: 27837020]

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, and Satija R (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21. [PubMed: 31178118]

Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, et al. (2018). The Immune Landscape of Cancer. Immunity 48, 812–830.e14. [PubMed: 29628290]

Varn FS, Wang Y, Mullins DW, Fiering S, and Cheng C (2017). Systematic Pan-Cancer Analysis Reveals Immune Cell Interactions in the Tumor Microenvironment. Cancer Res 77, 1271–1282. [PubMed: 28126714]

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17, 261–272. [PubMed: 32015543]

Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, Pfeil J, Narkizian J, Deran AD, Musselman-Brown A, et al. (2017). Toil enables reproducible, open source, big biomedical data analyses. Nature Biotechnology 35, 314–316.

Wellenstein MD, Coffelt SB, Duits DEM, van Miltenburg MH, Slagter M, de Rink I, Henneman L, Kas SM, 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. [PubMed: 31367040]

Wiley CD, Flynn JM, Morrissey C, Lebofsky R, Shuga J, Dong X, Unger MA, 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. [PubMed: 28699239]

Wolf FA, Angerer P, and Theis FJ (2018). SCANPY : large-scale single-cell gene expression data analysis. Genome Biol 19, 1–5. [PubMed: 29301551]

Zemmour D, Zilionis R, Kiner E, Klein AM, 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. [PubMed: 29434354]

Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA, 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. [PubMed: 32302573]

Zhang Q, He Y, Luo N, Patel SJ, 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. [PubMed: 31675496]

Zhou L, Chong MMW, and Littman DR (2009). Plasticity of CD4+ T cell lineage differentiation. Immunity 30, 646–655. [PubMed: 19464987]

Zilionis R, Engblom C, Pfirschke C, Savova V, Zemmour D, Saatcioglu HD, Krishnan I, Maroni G, Meyerovitz CV, Kerwin CM, 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. [PubMed: 30979687]

- 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.

Tumors can be extremely heterogenous across multiple parameters. By identifying and classifying recurrent immune features across 12 different tumor types, a framework is provided towards understanding tumor immunity as well as identifying broad as well as common therapeutic targets.
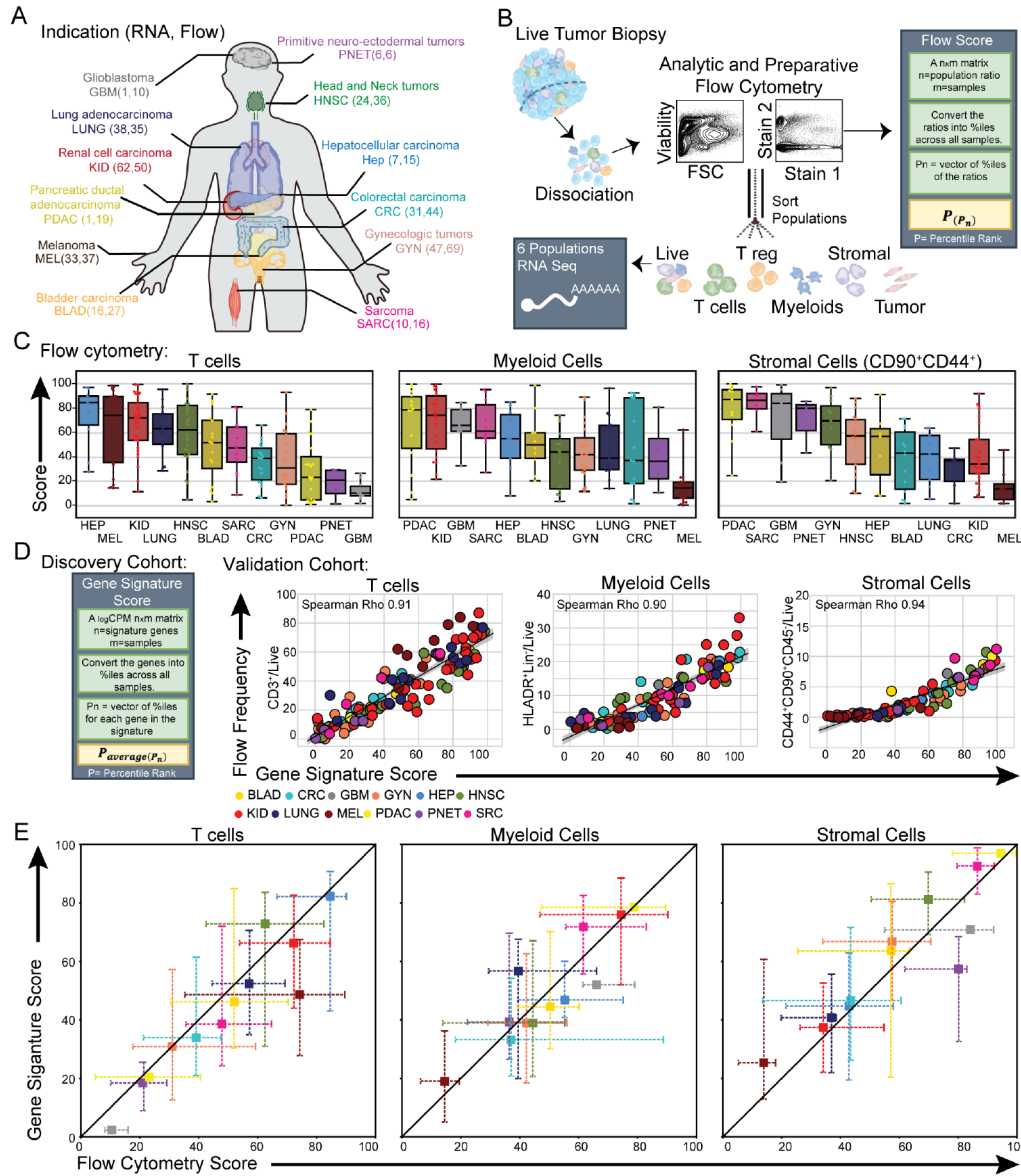
Figure 1: Generation and validation steps of T cells, Myeloid cells and Stromal cells features from solid tumors using flow cytometry and bulk RNA-sequencing.

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). 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 Supplementary 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 (See Fig Supp 1 and STAR protocol) color-coded according to the tumor types shown in 1C.

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 1C, with the interquartile range on both axes.
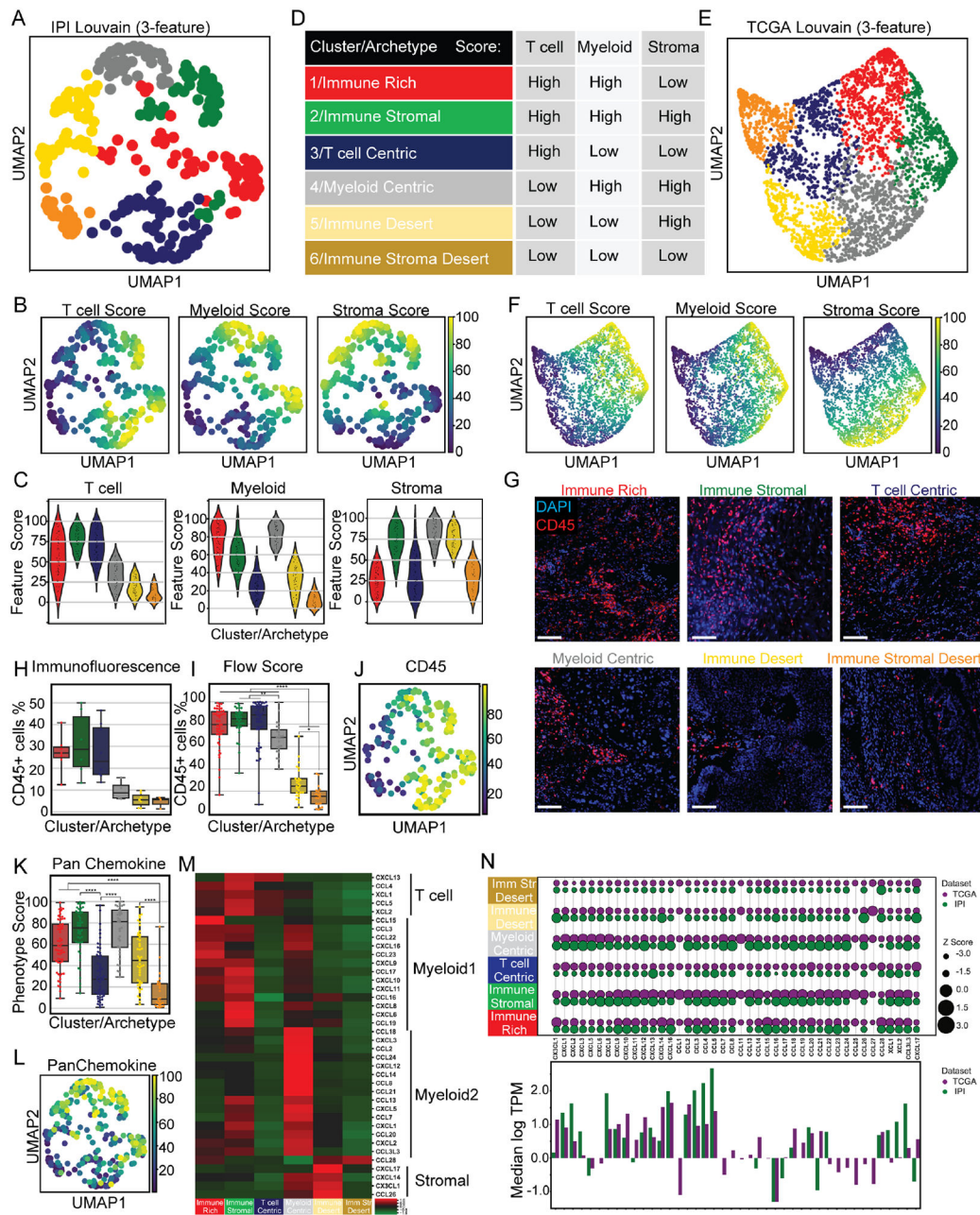
Figure 2: Identification of coarse immune archetypes in solid tumors using Louvain clustering on two independent datasets.

A,E. UMAP display using KNN and Louvain clustering of tumor immune archetypes using Tcell, Myeloid and CD90+ CD44+ Stroma features to cluster patients in the IPI (A) and TCGA (E) cohorts. Each dot represents a single patient. B,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,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). I,J. Box and whisker plot (I) and UMAP overlay (J) of immune cell frequency using flow cytometry. K,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 in Fig. 2D.
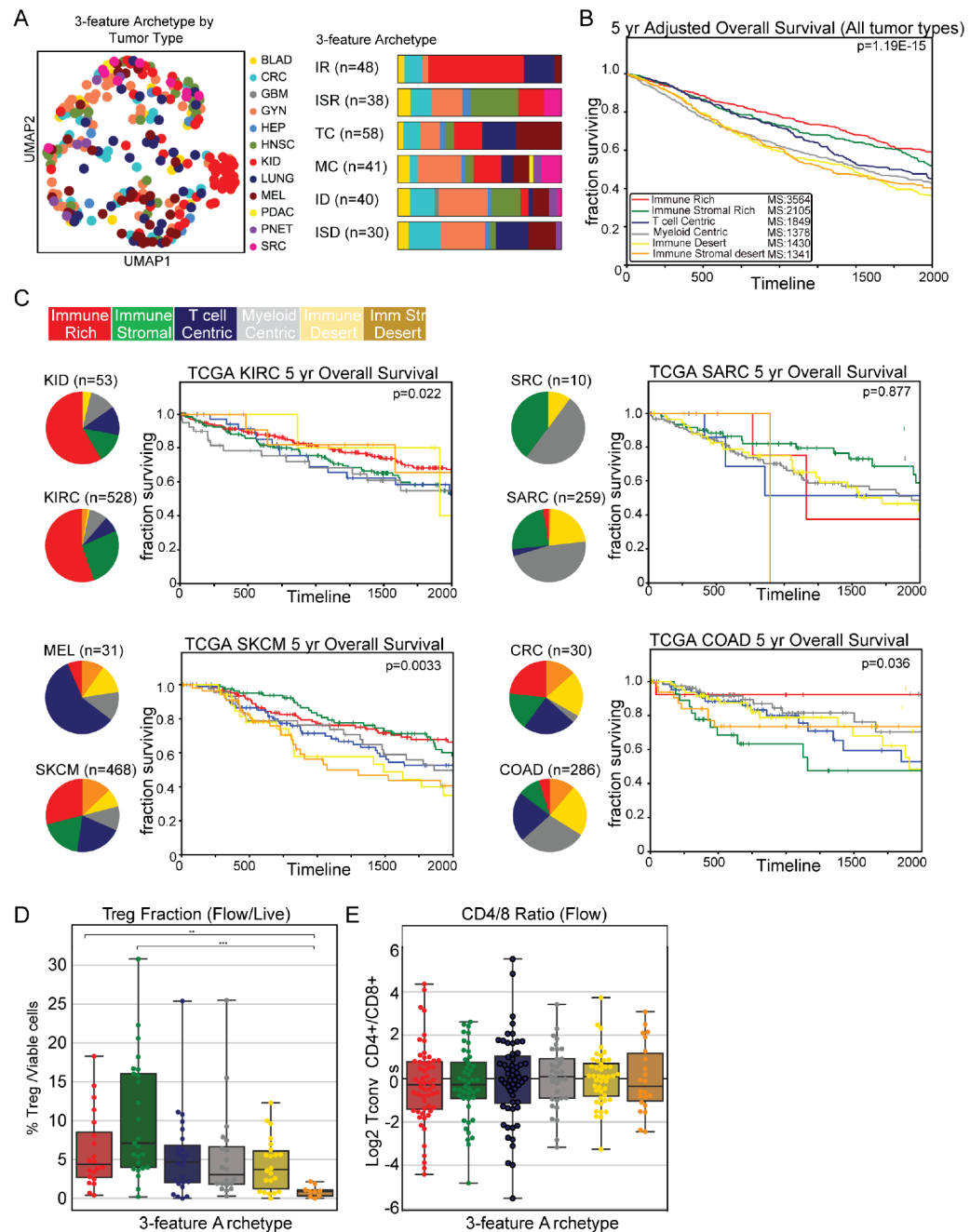
Figure 3: Coarse immune archetypes are independent of tissue origin and associated to overall survival.

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). 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.
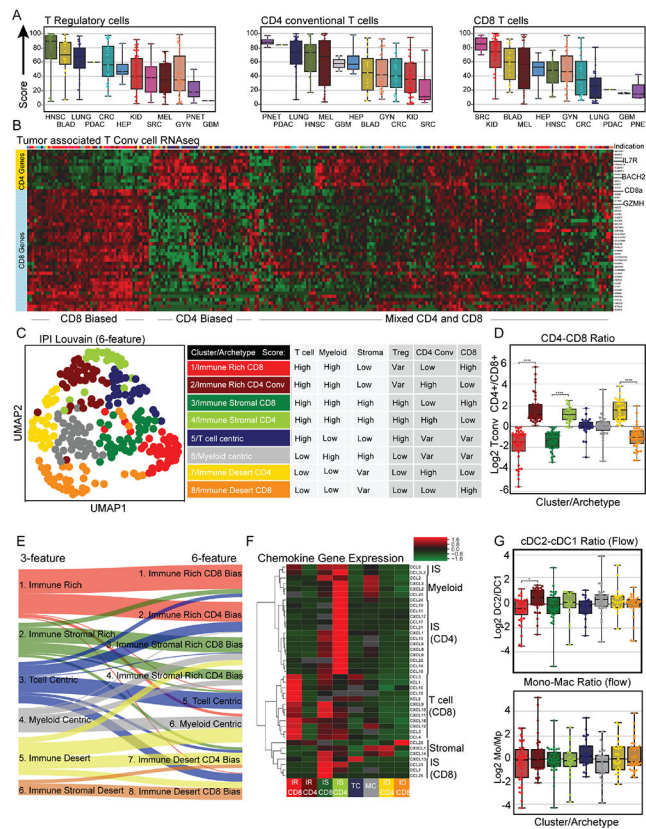
Figure 4: Inclusion of T cell subset features subdivide immune archetypes by CD4 to CD8 ratio.
A. Box and whisker plots of feature gene signature scores for CD4+ regulatory T cells (Treg feature) out of the live compartment, CD4+ and CD8+ conventional T cells (CD4 and CD8 features) out of the tcell compartment 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 tcell compartment. C. (Left) UMAP display and graph-based clustering of tumor immune archetypes using Tcell, 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 Mono to Macs ratio (bottom) measured by flow cytometry for each cluster/archetype identified.
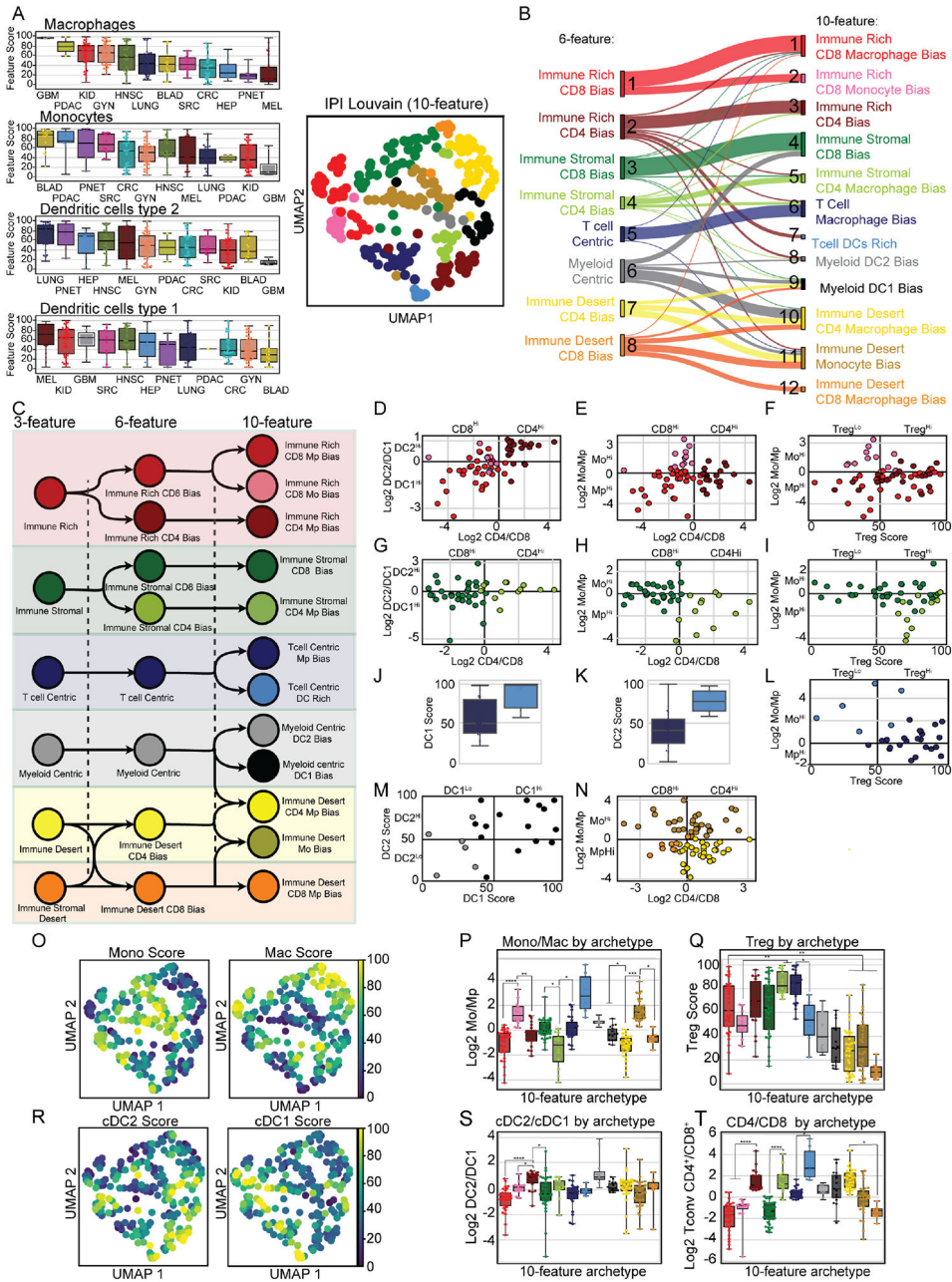
Figure 5: Single-cell RNA sequencing-derived myeloid signatures refines immune archetypes.
A. (Left)Box and whisker plots for the Macrophages, Monocytes, cDC1 and cDC2 features, calculated in the myeloid compartment, from the IPI cohort. (Right) UMAP display and graph-based clustering of tumor immune 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. Scatter plots of different features defining the twelve clusters/archetypes identified in the IPI cohort. O, 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, T. Box and whisker plot of Ln Mono to Macs ratio (P), Treg feature gene score (Q) Ln cDC2 to cDC1 ratio (S) Ln CD4 to CD8 conventional T cells ratio (T) for each cluster/archetype identified.
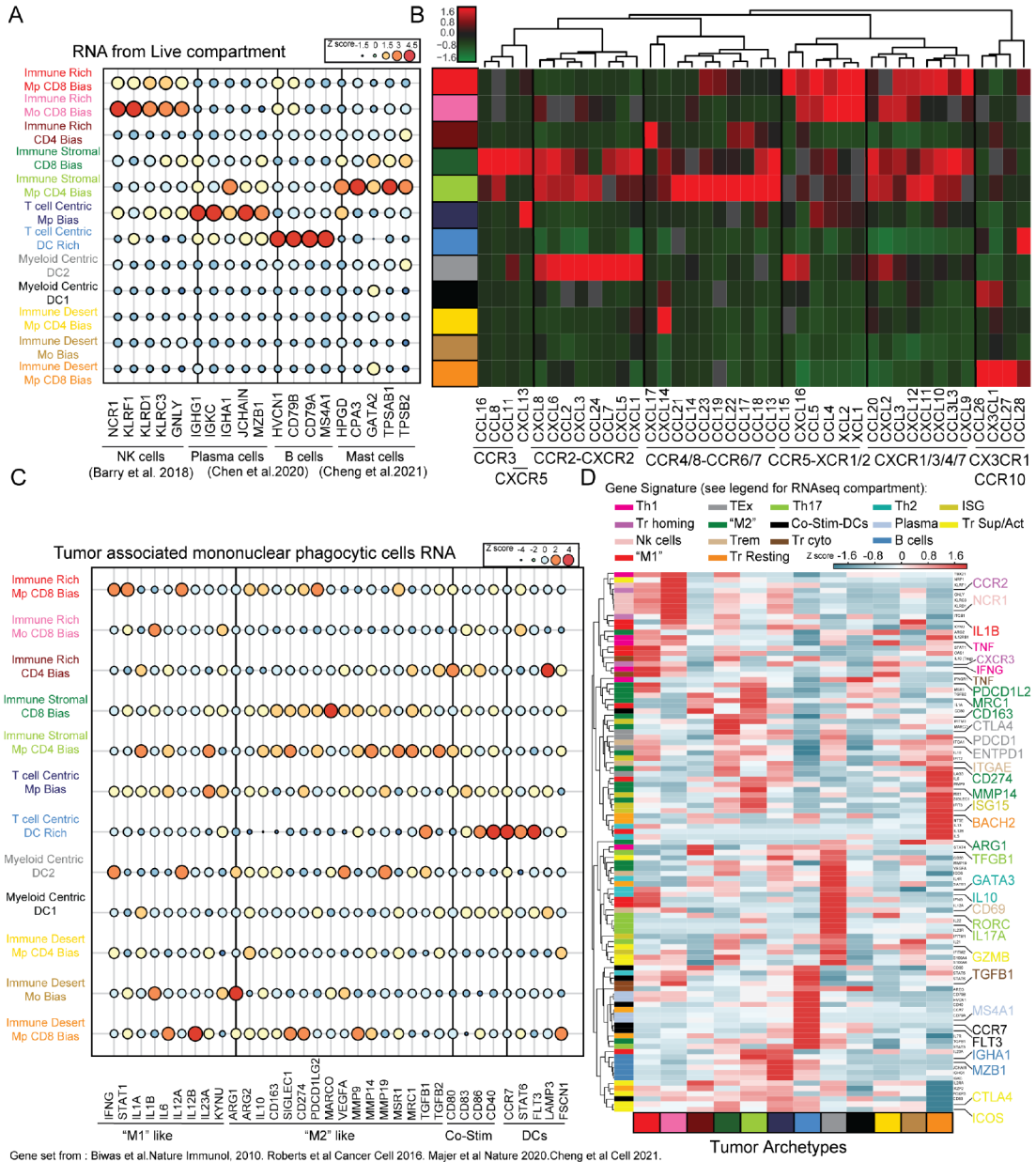
Figure 6: Each tumor archetype is defined by a unique combination of immune gene expression pattern

A. Bubble plot of NK cells (natural killer 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, M2) and Dendritic cells (Co-Stim, DC) function in the myeloid compartment, grouped by cluster/archetypes in the IPI cohort. D. Heatmap and hierarchical clustering of the median gene expression of B cells (all viable RNAseq compartment), NK cells (all viable RNAseq compartment), plasma cells (all viable RNAseq compartment) T cells phenotypes (Tconv

RNAseq compartment, T regs(Treg RNAseq compartment), macrophages and dendritic cells function (Myeloid RN Aseq compartment) grouped by cluster/archetype in IPI cohort.
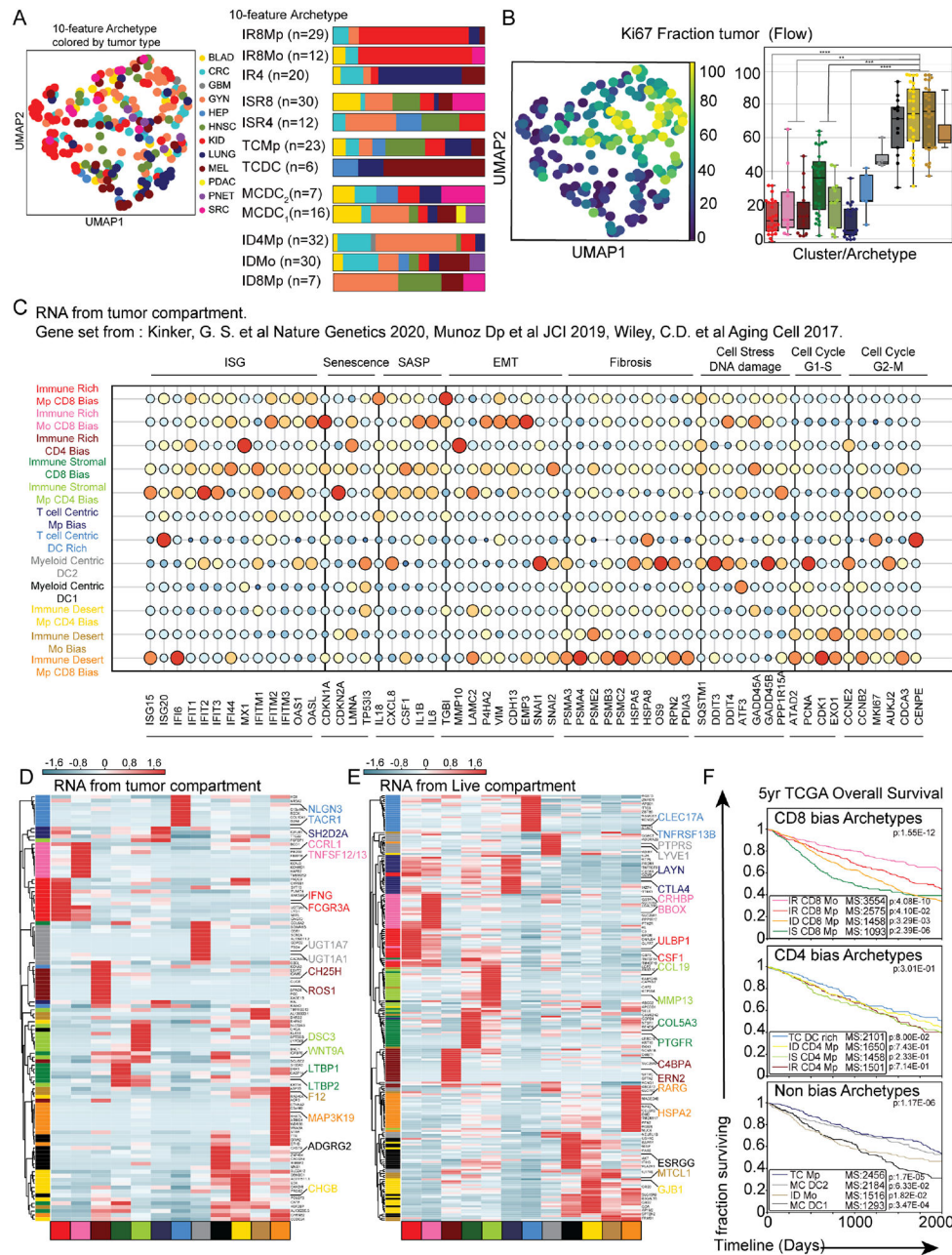
Figure 7: Immune archetypes tie closely to tumor biology and disease outcome.

A. 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,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.

Key Resource Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Streptavidin BV421	Biolegend	405226
anti-human CD45 APC/e780 (clone HI30)	Thermo Fisher	47-0459-42
anti-human CD3e PerCP/e710 (clone OKT3)	Thermo Fisher	46-0037-42
anti-human HLA-DR BUV395 (clone G46-6)	BD Biosciences	564040
anti-human CD56 BUV737 (clone NCAM16.2)	BD Biosciences	564448
anti-human CD4 PE/Dazzle 594 (clone S3.5)	Biolegend	100455
anti-human CD8a BV605 (clone RPA-T8)	Biolegend	301039
anti-human CD127 BV650 (clone HIL-7R-M21)	BD Biosciences	563225
anti-human CD38 AF700 (clone HIT2)	Biolegend	303523
anti-human CD25 APC (clone 2A3)	BD Biosciences	340939
anti-human CD45RO PE (clone UCHL1)	BD Biosciences	561889
anti-human PD-1 BV786 (clone EH12)	BD Biosciences	563789
anti-human ICOS BV711 (clone DX29)	BD Biosciences	563833
anti-human FoxP3 PE/Cy7 (clone 236A/E7)	Thermo Fisher	25-4777-41
anti-human CTLA-4 BV421 (clone BNI3)	BD Biosciences	565931
anti-human/mouse/rat Ki67 AF488 (clone SolA15)	Thermo Fisher	11-5698-82
anti-human CD19 PerCP/e710 (clone H1B19)	Thermo Fisher	45-0199-42
anti-human CD20 PerCP/e710 (clone 2H7)	Thermo Fisher	45-0209-42
anti-human CD56 PerCP/e710 (clone CMSSB)	Thermo Fisher	46-0567-42
anti-human CD64 BUV737 (clone 10.1)	BD Biosciences	564425
anti-human CD11c AF700 (clone 3.9)	Thermo Fisher	56-0116-42
anti-human CD16 BV605 (clone 3G8)	Biolegend	302039
anti-human CD273/PDL2 BV650 (clone MIH18)	BD Biosciences	563844
anti-human/mouse TREM2 APC (clone 237920)	R&D Systems	FAB17291A
anti-human CD304 PE (clone 12C2)	Biolegend	354503
anti-human CD1C/BDCA-1 PE/Cy7 (clone L161)	Biolegend	331515
anti-human CD197 BV421 (clone G043H7)	Biolegend	353207
anti-human BDCA-3 FITC (clone AD5-14H12)	Miltenyi	130-098-843
anti-human PDL1 BV786 (clone MIH1)	BD Biosciences	563739
anti-human CD14 BV711 (clone M5E2)	Biolegend	301837
Bacterial and Virus Strains		
Biological Samples		
Human tumor samples	UC San Francisco	IRB # 20-31740

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
N/A		
Critical Commercial Assays		
N/A		
Deposited Data		
All bulk RNAseq data for the IPI cohort and single cell RNAseq data previously unpublished	This paper	GSE184398
Single cell RNAseq data from melanoma invaded lymph node.	Binnewies et al	GSE125680
Single cell RNAseq data from Renal Clear cell carcinoma	Argüello et al	GSE159913
GitHub	UCSF DSCO_LAB Github	https://github.com/UCSF-DSCOLAB/pan_cancer_immune_archetypes
Experimental Models: Cell Lines		
N/A		
Experimental Models: Organisms/Strains		
N/A		
Oligonucleotides		
N/A		
Recombinant DNA		
N/A		
Software and Algorithms		
Python (2.7.15 & 3.7.2)	(Rossum, 1995)	https://www.python.org
Pandas (0.24.1,1.0.5)	(McKinney, 2010)	https://pandas.pydata.org/
Seaborn (0.9.0, 0.10.1)	(Waskom et al., 2020)	https://seaborn.pydata.org
Matplotlib (2.2.3, 3.2.2)	(Hunter, 2007)	https://matplotlib.org
Lifelines	(Davidson-Pilon et al., 2020)	https://lifelines.readthedocs.io/en/latest/
Scanpy 1.5.1	(Wolf et al., 2018)	https://scanpy.readthedocs.io/en/stable
SciPy	(Virtanen et al., 2020)	https://www.scipy.org/
BWA-mem	(Li, 2013)	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)	
limma	(Ritchie et al., 2015)	https://bioconductor.org/packages/release/bioc/html/limma.html

REAGENT or RESOURCE	SOURCE	IDENTIFIER
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/
Other		