A Pan-Cancer Census of Dominant Tumor Immune Archetypes

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. Barczac⁴, Michael Adkisson⁴, Austin Edwards^{1,2,3,4}, Mohammad Naser^{1,2,3,4}, Kevin C. Barry⁵, Tristan Courau^{1,2,3,4}, Taymour Hammoudi^{1,2,3,4}, Rafael J Arguëllo⁶, 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. Kelley⁷, Jocelyn S. Chapman^{3,9}, Chloe E. Attreya^{3,7,10}, Amar Patel^{3,10}, Adil I. Daud¹¹, Patrick Ha¹², Aaron A. Diaz¹³, 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 Boissonnas¹⁷, Saurabh Asthana^{3,8}, Vincent Chan^{2,3,18}, & Matthew F. Krummel^{1,2,3†}

Affiliations:

¹Department of Pathology, ²ImmunoX Initiative, ³UCSF Immunoprofiler Initiative, ⁴UCSF CoLabs, ⁷Helen Diller Family Comprehensive Cancer Center, ⁸Department of Epidemiology and Biostatistics, ⁹Departments of Obstetrics, Gynecology, and Reproductive Sciences, Medicine, ¹⁰Department of Medicine, Division of Hematology and Oncology, ¹¹Department of Dermatology, ¹²Department of Otolaryngology, ¹³Department of Neurological Surgery, ¹⁴Department of Surgery, ¹⁵Department of Medicine Division of Rheumatology, ¹⁶Division of Pulmonary and Critical Care Medicine, Department of Medicine and the Cardiovascular Research Institute, ¹⁸Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA 94143, USA; ⁵Translational Research Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁶Aix Marseille University, CNRS, INSERM, CIML, Centre d'Immunologie de Marseille-Luminy, Marseille, FRANCE. ¹⁷Sorbonne Université, INSERM, CNRS, Centre d'Immunologie et des Maladies Infectieuses - CIMI, Paris, France.

To whom correspondence should be addressed. Mail: matthew.krummel@ucsf.edu

Submitted as a 'Resource' to Cell, April 2021.

Combes et al.

SUMMARY

Cancers display significant heterogeneity with respect to tissue of origin, driver mutations and other features of the surrounding tissue. It is likely that persistent tumors differentially engage inherent patterns—here ‘Archetypes’—of the immune system, to both benefit from a tumor immune microenvironment (TIME) and to disengage tumor-targeting. To discover dominant immune system archetypes, the 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 archetypes that exist across indications. These Immune composition-based archetypes differentiate tumors based upon unique immune and tumor gene-expression patterns. Archetypes discovered this way also tie closely to well-established classifications of tumor biology. The IPI resource provides a template for understanding cancer immunity as a collection of dominant patterns of immune infiltration and provides a rational path forward to learn how to modulate these patterns to improve therapy.

Combes et al.

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 tissues (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 driving these responses 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;

Combes et al.

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). Nonetheless, the identification of recurrent motifs in the immune system that spans 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). A given immune response can thus be conceived as a collection of cell subsets and specific immune cell pairings linked with function allowing us to define “immune archetypes. For responses to therapy, a strong archetypal relationship between cDC1 and NK cells (Barry et al., 2018; Böttcher et al., 2018) or cDC2 and CD4 conventional and regulatory T cells (Binnewies et al., 2018; Bosteels et al., 2020) have been previously identified in specific tumor types and viral infection. However, the association of dominant tumor-promoting archetypes with tumor biology 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 tumors 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 could then be identified and validated in the TCGA dataset. These

Combes et al.

archetypes, discovered with only ten measurements, are well correlated to immune and tumor transcriptomic programs that span different tissues of origin and provide an unprecedented resource to classify and study cancer immunity and cancer targets to improve response to immunotherapy.

Combes et al.

RESULTS

A Pan-Cancer High-Dimensional Study of Dominant Immune Composition.

The UCSF Immunoprofiler (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 extended method). 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 see Extended Methods for full sort descriptions and potential caveats). This approach was chosen over single-cell RNA sequencing (scRNA-seq) because deep sequencing is needed to fully capture transcriptomic differences in sorted cell subsets populations and because scRNA-seq was an immature technology at the start of the program.

We initially used flow cytometry and focused upon the abundance of three major cell types in the TME, namely total α/β T cells (Tcell feature), myeloid cells (Myeloid feature) and non-immune CD45⁻, CD44⁺, Thy1⁺ stromal cells (CD44⁺ CD90⁺ Stroma feature) (supplementary table 2). Consistent with previous descriptions (Mandal et al., 2016; Thorsson et al., 2018;

Combes et al.

Varn et al., 2017) the cell abundances for these three cell types, plotted as a score (**Figure 1B, and extended method**) for each tumor indication show tremendous heterogeneity across and within tumor indications suggesting the need for a tumor classification that goes beyond tissue of origin (**Figure 1C**).

A distinguishing feature of our cohort is that it has linked compositional data via flow cytometry and RNAseq, which we leveraged to discover and then validate 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 to define 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 will correspond to a high cell abundance in the tumor specimen analyzed. (**Figure S1A**). Use of this score, when applied to RNAseq data from the 'Live' RNAseq compartment in a validation cohort showed a high degree of concordance with the cell type frequencies obtained via flow cytometry, independent of tissue of origin (spearman correlation 0.91, 0.90 and 0.94) (T, Myeloid, and Stromal, respectively) (**Figure 1D**). In addition, rank ordering of tumor indications by these gene scores shows similar trends, and strong correlations across individuals to those obtained by flow cytometric measurements (**Figure 1C/S1C and 1E**).

Our signatures for intra-tumoral composition significantly outperformed signatures derived from cell lines (e.g. CIBERSORT **Figure S1B**) when compared to the 'ground-truth' of flow

Combes et al.

cytometric composition data. We presume this improvement results from our data being taken from cells directly collected from fresh tumors 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 indication between datasets for the majority of the tissues surveyed was similar, suggesting that the abundances we describe by indication (e.g. **Figure 1C**) extend beyond our sample processing protocol (**Figure S1D**). The exceptions to this were for lung (LUNG), liver (HEP) and pancreatic (PDAC) tumors and this may be due to sampling or patient-selection variation between the cohorts; thus, in the remainder of this analysis, we did not consider these indications when making additional comparisons.

Unsupervised Clustering of 3-Features, Independent of Tissue of Origin

We next performed unsupervised clustering, using the Louvain community detection algorithm on a K nearest-neighbor (KNN) weighted graph, on these 3 features for all 260 samples with a “Live” bulk RNAseq compartment (**Supplementary Table S2**). The clustering was visualized using UMAP, a dimensionality reduction technique (Blondel et al., 2008; Zhu et al., 2020) (see extended method section), both on the IPI and TCGA cohorts (**Figure 2A/E**). 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**; see Feature Clustering in Extended Methods section) 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))

Combes et al. Figure 2

Combes et al.

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 of surgical specimens 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 corresponded to the six clusters found in the IPI cohort (**Figure 2E-F and Figure S2C-D**).

To both evaluate the histological descriptions in our archetypes and provide validation of the flow cytometry data, we ran immunohistochemistry 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 clusters were more similar to the immune desert clusters for overall CD45⁺ infiltration as assessed by imaging (**Figure 2G-H and Figure S2I-J**). However, MC clusters were intermediate when considering flow cytometry percentages which may represent a modest preferential recovery of immune cells by dissociation when compared to non-immune (**Figure 2I-J**). 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), although this will

Combes et al.

involve considerable orthogonal analyses due to the intrinsic complexity of diverse tissue morphologies (e.g. in 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” bulk RNAseq 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) (Nagarsheth et al., 2017). 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 patterns observed were similar on a chemokine-by-chemokine basis for archetypes as defined by composition in TCGA (**Figure 2N**), using the gene-signature archetype delineation (**Figure 2E**). Taken together, these results demonstrate that a three-

Combes et al.

feature scoring of tumor tissue identifies six unique clusters strongly correlated with immune infiltration and expression of distinct sets of chemokines.

Distribution of 3-feature Archetypes by Indication and outcome

Assessing archetypes based on just these three features demonstrated a heterogeneous distribution of tumor indications among the different clusters, although many indications, 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 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 indications, we analyzed distributions of archetypes in both IPI and TCGA and also assessed the outcome in individual indications (**Figure 3C and Figure S3B-D**). While the relative composition of archetypes by indication were broadly similar in both cohorts, there was some variation e.g. heavily '3 Tcell Centric' bias in Melanoma in IPI cohort manifests as a combination of these and '1:IR' or 'IS' in TCGA. Furthermore, while some indications such as colorectal cancer appears to have slightly better prognosis for 'IR' archetypes, trends were weaker within most individual indications. This variability prompted us to consider that these coarse-grained three-feature archetypes may not capture the full range of immune cell heterogeneity in tumors. Heterogeneity was especially evident in features such as Neutrophil

Combes et al.

density, Treg and CD4⁺/CD8⁺ T cell ratio which all showed high variability within 3-feature archetypes (**Figure 3D-F**).

Developing A 6-Feature Archetype Definition

Using flow cytometry data, we again grouped tumors by indication to assess variation in regulatory (Treg), CD4⁺ and CD8⁺ conventional T cell frequencies amongst T cells, both between and within indications (**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 extended method). 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 S4A-B**). The Treg, CD4, and CD8 feature scores, assessed in a selected validation cohort, showed very high correlation (Spearman correlation 0.86, 0.97 and 0.98, respectively), with their respective population abundances measured by flow cytometry. Again, these correlations were significantly better than correlations of cell-line derived signatures on the same tumor specimens (**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

Combes et al.

indications: CD8-biased tumors, CD4-biased tumors and a large population that was mixed for both these signatures (**Figure 4B**).

We next combined the CD4, CD8 and Treg gene signature scores with the Tcell, Myeloid and CD44⁺ CD90⁺ Stroma gene signature scores to perform 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 partly 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 and neutrophils remain generally variable within and between archetypes (**Figure S4G-H**). Assessing chemokine gene expression showed that this re-clustering dramatically segregated chemokine gene expression in the Immune Rich archetype and also further refined the chemokine expression found amongst the other archetypes (**Figure 4F and Figure S4I**)

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 metabolic and

Combes et al. Figure 4

Combes et al.

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 S4J** and Extended Method). 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 S4K**). This T_{ex} gene score best mirrored the flow-cytometry-based abundance of CTLA-4+/PD-1+ within the combined $CD8^+$ and $CD4^+$ compartments (**Figure S4L**). Using this T_{ex} gene signature score we observed enrichment of exhaustion in $CD8$ biased archetypes including the immune desert (ID) archetype (**Figure S4M-N**) and with low MHC I expression by the tumor cells in these archetypes (**Figure S4O**).

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 phagocytic cell (MPC) subsets including monocytes (Mo), macrophages (Mp), classical dendritic cells (cDC2, cDC1) and plasmacytoid dendritic cells (pDCs) for each archetype using flow cytometry (**Figure S4P**). This revealed that, despite, slight enrichment between archetypes e.g. of cDC2 over cDC1 in IR $CD4$ Bias, the frequencies of cDC1/cDC2, monocytes, and macrophages showed variability within archetypes, indicating that MPC frequencies are segregating the 6-features archetypes. (**Figure 4G** and **Figure S4Q**).

Combes et al.

10-Feature Archetype Definition

The complexity of the MPC 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 across all samples, we generated a scRNAseq sub-study of sorted tumor associated MPC to identify specific gene signatures for the principal MPC subsets (**Figure S5A**). Following removal of non-MPC 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 to their respective MPC 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). When validated against cellular abundance by flow cytometry, the correlation of each MPC subset feature score across all indications 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 S5D** and extended method). Again, when we analyzed the distributions of these five MPC subset features scores by indication, a high heterogeneity by indication was apparent (**Figure 5A** and **Figure S5E**).

Thus, we repeated the unsupervised clustering after adding four MPC subset features including macrophages, monocytes and the two types of classical dendritic cells to the previous six features (**Supplementary Table 3**). The 10-feature clustering produced 12

Combes et al. Figure 5

Combes et al.

unique tumor immune archetypes after DBI minimization, with no bias toward specific archetypes found between scores calculated from flow and scores directly from RNAseq (**Figure 5A and Figure S5F-H**). As previously observed when adding T cell subsets, the addition of the four MPC measures subdivided preexisting immune archetypes but did not increase those by 16-fold as could occur if these were randomly assorted (**Figure 5B-C**). MPC inclusion also resulted in some specimens shifting between MC and ID archetypes, driven by strong monocytes or macrophages enrichment.

We found that generally both IR and ID sub-archetypes were distinguished by specific pairings of T cell and myeloid subsets but that the specific pairing 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 depending on archetypes (**Figure 5D-N**). In the IS archetypes (**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

Combes et al.

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 The 10-Feature Archetypes

Focusing on 10-feature archetypes, we next surveyed the association of the archetypes that emerged, with cell fractions 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) (**Figure 6A**). Plasma cells, defined by enrichments of IgG genes, were notably also enriched in T cell Centric Macrophage bias tumors whereas B cell genes (CD20, CD79) were more prevalent in T cell Centric DC rich (Chen et al., 2020; van Galen et al., 2019) (**Figure 6A**). 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 S5H**). An archetype-specific immune composition pattern again correlated with specific enrichment of groups of chemokine transcripts. (Nagarsheth et al., 2017; Tokunaga et al., 2018) (**Figure 6B**). Notably, chemokines with specificity for families of chemokine receptors now frequently co-clustered within 10-feature archetypes.

Combes et al.

We next used the myeloid RNAseq compartment to explore the level of expression of genes previously associated to tumor-infiltrating MPC, 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) across all 12 archetypes. Despite the heterogeneity in individual gene expression associated to 'M1' and 'M2' macrophage phenotype among the different 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 with the exception of LAMP3 which was significantly enriched in the IR CD4 archetype (**Figure 6C**). LAMP3 expression has been recently associated to a distinct tumor-associated cDCs subset characterized by high expression of regulatory molecules such as PD-L1 and correlating with Treg abundance in tumor (Cheng et al., 2021; Maier et al., 2020; Zhang et al., 2019).

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

Combes et al.

enriched in genes associated with the type 1 response (IFNG, TNF, IL1B) both in T_{conv} and MPC 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 represent a collection of distinct immune phenotypes made up of a unique combination of cell composition and transcriptomic profiles.

However, certain immune populations remained variable across the archetypes. For instance, neutrophil infiltration, presents only a slight and statistically insignificant rise in archetypes with low immune infiltration (**Figure S6A-B**). Conversely, a gene signature previously associated to stimulatory dendritic cells abundance and better outcome in Melanoma patient are slightly enriched in archetypes with high total immune infiltration (Barry et al., 2018) (**Figure S6A and S6C**). Altogether, those data suggest that while the 12 immune archetypes identified may represent the dominant combination of cell types and transcriptomic profiles present in tumors, they may independently be populated with other cell types or signatures.

Combes et al.

10-Features Tumor Archetype Ties Closely to Tumor Biology And Disease Outcome

Finally, we sought to examine the relationship of these immune archetypes to the phenotype of the tumor cells themselves. As previously shown for the coarse archetypes, the 10-feature archetypes are also diverse in their tissue of origin; however, some tumor indications remain highly represented within particular archetypes (e.g., Kidney: Immune Rich CD8, Melanoma: T cell centric archetypes) (**Figure 7A and Figure S7**). Regardless of the frequent intra-indication 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 ISG, Senescence or EMT (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 differently 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

Combes et al. Figure 7

Combes et al. Figure S7

Combes et al.

from the IR CD8 Macrophage archetype, concordant with increased type 1 response in immune cells from the same archetype.

In order 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 a set of 12 unique gene signatures from which we observed both immune (e.g. LAYN, CTLA4, CSF1, CCL19) and non-immune (CRHBP, PTGFR, HSPA2, MTCL1) related genes. Using these minimal 12 signatures on the TCGA dataset, we demonstrated that the tumor archetypes identified in our IPI dataset were retrieved with a similar relative composition of archetypes by indications 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 S7**). This suggests the predictive potential of these signatures for rapid classification of patients into archetypes.

Finally, survival analysis of the different tumor archetypes identified in the TCGA dataset for each cancer type separately showed that better outcome may be indication-specific and each indication may have a different type of TIME promoting the best immune response (**Figure S7**). However, when survival analysis was performed across indications, we detected significant outcome differences between archetypes that have similar T cell subset enrichment (**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:2993 days) and more so when compared to ID and IS CD8 biased

Combes et al.

archetypes (p-value $4.82E-12$) (**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 (22217 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 $6.89E-2$) (**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 cancer and to understand how to best modulate these archetypes depending on specific immune context.

Combes et al.

DISCUSSION

In this study, we present a holistic survey of dominant immune archetypes across 12 cancer indications using fresh surgical tumor specimen from 364 patients. Empowered by complementary profiling assays that assembled corresponding information from matched compositional, transcriptomic and imaging data on a variety of tumor-infiltrating cells (**Figure 1**, <http://immunoprofiler.org/>), 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. 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 TIME of solid tumors. This is partly in line with previous work which identified 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 immune gene expression signatures (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 MPC subsets. However, our analysis leveraged our linked compositional data via flow cytometry and RNAseq data from sorted cell populations to identify further the origin of those different immune networks which may partly explain our higher number of archetypes. Future work using single-cell omics across tumor types will be critical to determine if the 12 immune

Combes et al.

archetypes identified in our study can be further subdivided or if cell subsets identified using these technologies generally align with our current landscape (Cheng et al., 2021; Slyper et al., 2020; Zhang et al., 2020).

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 (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 and 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, a pattern previously identified as supporting plasma cell residency in bone marrow (Glatman Zaretsky et al., 2017). It is tempting to propose that conserved relationships of cell densities found in such a non-cancerous setting might be because some tumors hijack a particular immune archetype that may normally be used in a completely different setting. This idea is perhaps reinforced by the distinct chemokine expression profile identified for each tumor immune archetype. While IR CD8 archetypes are defined by a CXCL9, CXCL10, CXCL11/CXCR3 axis already described in the TME, ID CD8

Combes et al.

Macrophage bias archetype is defined by a strong expression of chemokines binding CX3CR1, an axis essential in immune surveillance and homeostasis (Gerlach et al., 2016).

The association between chemokine expression programs and the tumor immune archetypes also suggests an association with a specific spatial organization. While we confirmed the overall correspondence between flow cytometry, transcriptomic data and immunohistochemistry assays, the spatial landscape of the tumor immune archetypes would need further investigation. In future, the 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 enable this analysis, using the knowledge that only some archetypes have the relationship to begin with. In this way, archetype-discovery plays an important role in selecting tissues containing similar biology for concerted study.

Tumor-specific genetics and mutational burden have been proposed to be key for anti-tumor immunity in multiple indications (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 (Spranger et al., 2016). Conversely, we demonstrated that immune tumor archetypes are associated to tumor proliferation, diverse tumor transcriptomic programs and overall survival. Studying the synergy between immune archetypes and tumor mutational profile,

Combes et al.

may elucidate further the importance of tumor mutational burden in anti-tumor immune response (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 model 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; Sade-Feldman et al., 2018). 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.

Combes et al.

ACKNOWLEDGEMENTS

We thank all members of the Krummel Lab, UCSF ImmunoX, UCSF CoLabs and the UCSF Immunoprofiler Consortium for discussion and guidance while developing this study. We would like to thank Dr Nicholas Kuhn for scientific discussion. We would like to thank Dr Kenneth Hu for editing the manuscript. We would like to particularly thank Isabelle Tingin, Garry Shumakher, Elizabeth Edmiston and Meghan Zubradt for their constant support and help during this study. We would like to thank Dr Ana Catharina Silva for her help in designing the graphical abstract. Acquisition and analysis of certain human samples described in this study was partially funded by contributions from AbbVie, Amgen, Bristol-Myers Squibb, and Pfizer as part of the UCSF Immunoprofiler Initiative. Further support came from the NIH (R01 U01) for MFK and P30 : P30CA082103 for ABO. Finally, we thank all patients and their families, for placing their trust in us.

AUTHOR CONTRIBUTIONS

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 of 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, AAR, DA, 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.

REFERENCES

- Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson, R.M., Lowe, J.B., Liu, S.D., Zhao, S., et al. (2014). Multiplexed ion beam imaging of human breast tumors. *Nat Med* 20, 436–442.
- Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* 18, 220.
- Arce Vargas, F., Furness, A.J.S., Litchfield, K., Joshi, K., Rosenthal, R., Ghorani, E., Solomon, I., Lesko, M.H., Ruef, N., Roddie, C., et al. (2018). Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. *Cancer Cell* 33, 649-663.e4.
- Asp, M., Giacomello, S., Larsson, L., Wu, C., Fürth, D., Qian, X., Wärdell, E., Custodio, J., Reimegård, J., Salmén, F., et al. (2019). A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* 179, 1647-1660.e19.
- Azizi, E., Carr, A.J., Plitas, G., Cornish, A.E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kisieliovas, V., Setty, M., et al. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* 174, 1293-1308.e36.
- Barry, K.C., Hsu, J., Broz, M.L., Cueto, F.J., Binnewies, M., Combes, A.J., Nelson, A.E., Loo, K., Kumar, R., Rosenblum, M.D., et al. (2018). A natural killer-dendritic cell axis defines checkpoint therapy-responsive tumor microenvironments. *Nat Med* 24, 1178–1191.
- Beltra, J.-C., Manne, S., Abdel-Hakeem, M.S., Kurachi, M., Giles, J.R., Chen, Z., Casella, V., Ngiew, S.F., Khan, O., Huang, Y.J., et al. (2020). Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms. *Immunity* 52, 825-841.e8.
- Bi, K., He, M.X., Bakouny, Z., Kanodia, A., Napolitano, S., Wu, J., Grimaldi, G., Braun, D.A., Cuoco, M.S., Mayorga, A., et al. (2021). Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. *Cancer Cell*.
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795.
- Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med* 24, 541–550.
- Binnewies, M., Mujal, A.M., Pollack, J.L., Combes, A.J., Hardison, E.A., Barry, K.C., Tsui, J., Ruhland, M.K., Kersten, K., Abushawish, M.A., et al. (2019). Unleashing Type-2 Dendritic Cells to Drive Protective Antitumor CD4+ T Cell Immunity. *Cell* 177, 556-571.e16.

Combes et al.

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

Blank, C.U., Haining, W.N., Held, W., Hogan, P.G., Kallies, A., Lugli, E., Lynn, R.C., Philip, M., Rao, A., Restifo, N.P., et al. (2019). Defining “T cell exhaustion.” *Nat Rev Immunol* 19, 665–674.

Blondel, V.D., Guillaume, J.-L., Hendrickx, J.M., de Kerchove, C., and Lambiotte, R. (2008). Local leaders in random networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 77, 036114.

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

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

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

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

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

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

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

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

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

Combes et al.

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.

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

Davies, D.L., and Bouldin, D.W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1*, 224–227.

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

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

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

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

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

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

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

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

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

Combes et al.

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

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

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

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

Gotwals, P., Cameron, S., Cipolletta, D., Cremasco, V., Crystal, A., Hewes, B., Mueller, B., Quarantino, 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.

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

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

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

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

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

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

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

Combes et al.

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

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

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

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

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

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

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

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

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

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

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

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

Combes et al.

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

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

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

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

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

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

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

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

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

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

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

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

Combes et al.

Sade-Feldman, M., Yizhak, K., Bjorgaard, S.L., Ray, J.P., de Boer, C.G., Jenkins, R.W., Lieb, D.J., Chen, J.H., Frederick, D.T., Barzily-Rokni, M., et al. (2018). Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell* 175, 998-1013.e20.

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

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

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

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

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

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

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

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

Tokunaga, R., Zhang, W., Naseem, M., Puccini, A., Berger, M.D., Soni, S., McSkane, M., Baba, H., and Lenz, H.-J. (2018). CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy. *Cancer Treat Rev* 63, 40–47.

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

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

Combes et al.

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

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

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

Zhu, X., Zhang, J., Xu, Y., Wang, J., Peng, X., and Li, H.-D. (2020). Single-Cell Clustering Based on Shared Nearest Neighbor and Graph Partitioning. *Interdiscip Sci* *12*, 117–130.

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

FIGURE LEGENDS

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), Monophagocytes (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. 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 extended method section). E. Cross-whisker plots comparing median Tcell, Myeloid and CD90+ CD44+ Stroma gene signature scores by tumor indication to the median flow score with the interquartile range on both axes. F. Individual and combined spider plots of median Tcell, Myeloid, CD90+ CD44+ Stroma scores by tumor indication.

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 (a graph-based community detection algorithm) 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 summarized by the 3 features. 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 abundance 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.

Figure 3: Coarse immune archetypes are independent of tissue origin and associated to overall survival. A. UMAP display and graph-based clustering of tumor immune archetypes using Tcell, Myeloid and CD90+ CD44+ Stroma features color-coded by tumor indication. 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 indication 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 Neutrophils frequency in tumor measured by flow cytometry for each cluster/archetype identified. E. Box and whisker plot of CD4⁺ regulatory T cells frequency in tumor measured by flow cytometry for each cluster/archetype identified. F. Box and whisker plot of Ln CD4⁺ to CD8⁺ conventional T cell frequency ratio in tumor measured by flow cytometry for each cluster/archetype identified using three features.

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 summarized by the 6 features. (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 Ln CD4⁺ to CD8⁺ conventional T cell feature gene signature score ratio in tumor measured each of the cluster/archetype identified with 6 features 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 IPI cohort using six features. G. Box and whisker plot of Ln cDC2 to cDC1 ratio (top) and Mono to Macs ratio (bottom) measured by flow cytometry for each cluster/archetype identified using 6 features.

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 feature scores, calculated in the myeloid compartment, in the IPI cohort. (Right) UMAP display and graph-based clustering of tumor immune archetypes using Tcell, Myeloid, CD90⁺ CD44⁺ Stroma, CD4, CD8, Macrophages, Monocytes, cDC1 and cDC2 feature scores 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 using 10 features. O, R. UMAP overlay of Macrophages (Macs) and Monocytes (Mono) and classical dendritic cell type 1 (cDC1) and 2 (cDC2) feature scores (R) (See Fig S5H) for each cluster/archetype identified using 10

Combes et al.

features. 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 using 10 features.

Figure 6: immune gene expression pattern specific to each tumor archetype

A. Bubble plot of NK cells (natural killer cells), B cells and plasma cells associated gene expression in the live compartment grouped by cluster/archetypes in the IPI cohort using 10 features. B. Heatmap and hierarchical clustering of the median chemokine gene expression of all chemokines in the Chemokine phenotype signature, grouped by cluster/archetype in IPI cohort using 10 features. 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 using 10 features. 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 RNAseq compartment) grouped by cluster/archetype in IPI cohort using 10 features.

Figure 7: Immune archetypes tie closely to tumor biology and disease outcome. A. UMAP display and graph-based clustering of tumor immune archetypes using Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8, Macrophages, Monocytes, cDC1 and cDC2 feature scores color-coded by tumor indication. Each dot represents a single patient summarized by the 10 features. 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 using 10 features. 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. Kaplan-Meier overall survival curve from the TCGA cohort for each immune archetype using gene signatures in (E), split by T conventional subset enrichment). Median survival (MS) and p-value associated to each survival curve are noted.

Supplementary Materials For:

A Pan-Cancer Census of Dominant Tumor Immune Archetypes

Alexis J. Combes^{1,2,3,4}, Bushra Samad^{1,2,3,4}, Jessica Tui^{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. Barczac⁴, Michael Adkisson⁴, Austin Edwards^{1,2,3,4}, Mohammad Naser^{1,2,3,4}, Kevin C. Barry⁵, Tristan Courau^{1,2,3,4}, Taymour Hammoudi^{1,2,3,4}, Rafael J Arguëllo⁶, 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. Kelley⁷, Jocelyn S. Chapman^{3,9}, Chloe E. Attreya^{3,7,10}, Amar Patel^{3,10}, Adil I. Daud¹¹, Patrick Ha¹², Aaron A. Diaz¹³, 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 Boissonnas¹⁷, Saurabh Asthana^{3,8}, Vincent Chan^{2,3,18}, & Matthew F. Krummel^{1,2,3,†}

This PDF file includes:

Supplementary Figure Legends

Supplementary Tables 1 to 5

Extended STAR Methods and Associated References

SUPPLEMENTARY FIGURE LEGENDS:

Figure S1: Generation and validation steps of T cells, Myeloid cells and Stromal cells features from solid tumors using flow cytometry and bulk RNA-sequencing, related to Figure 1. A. Volcano plots, Venn diagrams and gene names showing the method of feature gene signature discovery for the Tcell, Myeloid and CD90+ cd 44+ Stroma features using differential gene expression in tumor associated sorted compartments (see extended method section). B. Correlation plots of Tcell and Myeloid gene signature score against their corresponding flow population fraction (top) and CIBERSORT derived fraction (bottom) color-coded by tumor indication. C. Box and whisker plots of Tcell, Myeloid and CD90+ CD44 Stroma features gene scores in the IPI cohort. D. Cross-whisker plots comparing median Tcell, Myeloid and CD90+ CD44+ Stroma gene signature scores by tumor indication between the IPI and TCGA cohorts with interquartile range on both axes.

Figure S2: Identification of coarse immune archetypes in solid tumors using Louvain clustering on two independent datasets, related to Figure 2. A, B. Scatter plot of the Davies-Bouldin index and cluster size over multiple iterations of Louvain clustering and varying parameters using 3 features in the IPI (A) or TCGA (B) cohort. C. 3D plot of Tcell, Myeloid and CD90+ CD44+ Stroma scores color-coded by their cluster assignment from Louvain clustering these features in the IPI (left) and TCGA (right) cohorts. D. Violin plots of the Tcell, Myeloid and CD90+ CD44+ Stroma feature score for each cluster/archetype in TCGA cohort. E. Box and whisker plot of a pan chemokine gene score by cluster/archetype identified in TCGA cohort. F. Scatter plot of immune cell population fraction using only viable cells or total cells as denominator. G, H, I, J, K, L. Representative H&E (top) and immunofluorescence (bottom) images of tumor biopsies using CD45 (red) and DAPI (blue) staining for each cluster/archetype identified in IPI cohort.

Figure S3: Coarse immune archetypes are independent of tissue origin and associated to overall survival., related to Figure 3. A. UMAP display and graph-based clustering of immune archetypes using Tcell, Myeloid and CD90+ CD44+ Stroma features to cluster patient with cohorts color coded by tumor type. B, C, D. (Left) Pie charts representing distribution of each archetype by indication from IPI (top) and TCGA (bottom) cohorts using 3 features. (Right) Kaplan-Meier overall survival curve for each immune archetype identified on TCGA dataset for Bladder urothelial carcinoma (B, BLCA), Gynecologic tumors (C, UCS + UCEC + OV,) and Head and Neck squamous cell carcinoma (D, HNSC).

Figure S4: Generation and validation of CD4 CD8 and Treg signatures and further exploration of 6 feature archetypes, related to Figure 4. A. Venn diagram and gene names of the Treg (CD4+ regulatory T cell) feature gene signature score (see extended method section). B. Volcano plot visualizing differential gene expression between CD4+ or CD8+ T cells in the tcell compartment (see extended method section). Feature gene signatures are listed in red squares. C. Correlation plots of CD4, CD8 and Treg feature score versus their corresponding flow population fractions (top) and CIBERSORT derived fractions (bottom) color-coded by tumor indication. D. Violin plots of the 6 features distributions in each cluster/archetype in the IPI cohort. E. Scatter plot of the Davies-Bouldin index and cluster size over multiple iterations of Louvain clustering and varying parameters

using 6 features in the IPI cohort. F UMAP of immune archetypes using Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8 and Treg features in the IPI cohort using KNN and Louvain clustering. (Right) UMAP overlay of samples that needed to calculate their CD4 and CD8 feature score flow cytometry population fractions (see extended method section). G, H. Box and whisker plot of total immune cell population fraction (G) and Neutrophils (H) using flow cytometry. I. Box and whisker plot pan chemokine gene score by cluster/archetypes in the IPI cohort using 6 features. J. Heatmap of the Spearman correlation coefficient of gene expression between CTLA4, CD38, PDCD1, HAVCR2 and LAG3 and our Exhaustion phenotype gene signature with additional genes associated with exhaustion used as a control. K. Gating strategy of tumor associated conventional CD4+ and CD8+ T cells and the markers used to define exhaustion. L. Correlation plots of individual and combined CD4+ PD1+ CTLA4+ and CD8+ PD1+ CTLA4+ conventional T cells population fractions against the Exhaustion phenotype gene signature score color-coded by tumor indication. M. (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. (Right) UMAP overlay of the Exhaustion phenotype score (see J and extended method) calculated in the tcell compartment. N, O. Box and whisker plots of Exhaustion phenotype score (N) and MHC Class I phenotype score calculated in the tumor compartment (O) for each cluster/archetype identified in IPI cohort using 6 features. P. UMAP overlay of classical dendritic cell type 1 (cDC1) and 2 (cDC2), plasmacytoid dendritic cells (pDCs), monocytes (Mono) and macrophages (Mac) frequencies measured by flow cytometry.

Figure S5 Single-cell RNA sequencing-derived myeloid gene signatures refine immune archetypes, related to Figure 5. A. Details of the processing pipeline for digesting fresh tumor biopsies into single cell suspension, submitting to multi-parametric flow cell sorting of tumor associated myeloid population and encapsulation for single-cell RNA sequencing. B. t- display and graph-based clustering of tumor associated myeloid cells from Melanoma processed for single-cell RNA sequencing. Each dot represents a cell. C. Dot plot of the top differentially-expressed-genes (DEG) between clusters identified in tumor associated mononuclear phagocytic cell (MPC) subsets. D. Gating strategy of tumor associated myeloid cells. E. Correlation plots of Macrophages, Monocytes, cDC1, cDC2 and pDC gene score against their corresponding flow population fractions in the myeloid compartment color-coded by tumor indication. F. Box and whisker plots gene score for pDCs (out of myeloid compartment) in IPI cohort. G. Scatter plot of the Davies-Bouldin index and cluster size over multiple iterations of Louvain clustering and varying parameters using 10 features in the IPI cohort. H. (Right) UMAP of tumor immune archetypes using Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8, Macrophages, Monocytes, cDC1 and cDC2 feature scores to cluster patients in the IPI cohort, using KNN and Louvain clustering. (Center-Left) UMAP overlay of samples that needed to calculate their CD4 and CD8 feature score and their Macrophages, Monocytes, cDC1 and cDC2 feature scores using flow cytometry population fractions (see extended method section). I. Violin plots of the 10 features distributions in each cluster/archetype in the IPI cohort.

Figure S6: immune cell frequency and transcriptomic profile pattern associated to each tumor archetype, related to Figure 6. A. (Left) UMAP display and graph-based clustering of tumor immune archetypes using Tcell, Myeloid, CD90+ CD44+ Stroma, CD4, CD8, Macrophages, Monocytes, cDC1 and cDC2 feature scores to cluster patients in the IPI cohort. Each dot represents a single

patient summarized by the 10 features. (Right) UMAP overlay of CD45+ cell frequency measured by flow cytometry. B. UMAP overlay of neutrophils frequency out the total immune cells measured by flow cytometry. C. UMAP overlay of a Stimulatory DC gene signature score measure on 'all viable' RNAseq compartment

Figure S7: Immune archetypes tie closely to tumor biology and disease outcome, Related to Figure 7

A to J. (Left) Pie charts representing the distribution of each archetype by indication in IPI 10-feature clustering (top) and TCGA (bottom) cohorts. (Right) Kaplan-Meier overall survival curve for the most abundant immune tumor archetype by tumor indication in TCGA cohort (see Extended method).

[illegible]

Supplementary Table 1: List of UCSF Immunoprofiler Initiative samples and associated clinical data. List of each tumor biopsy included in the first Louvain clustering Figure 2A and associated demographic and Clinical data.

		Bulk RNAseq	Multi-parametric flow cytometry		
Sample	3 Feature Archetype	Live	Tcell Panel	Myeloid panel	Sort Panel
			CD45+,CD3+, HLA-DR+/-	CD45+,CD3- CD19-,CD20- CD56- HLA-DR+	CD45-,EPCAM- CD90+,CD44+
IPIBLAD001.T1		—	X	X	—
IPIBLAD002.T1		—	X	X	—
IPIBLAD003.T1		—	X	—	—
IPIBLAD004.T1		—	X	—	—
IPIBLAD005.T1	Immune Rich	X	—	—	—
IPIBLAD006.T1	Myeloid Centric	X	X	X	—
IPIBLAD007.T1		—	X	X	X
IPIBLAD009.T1		—	X	X	X
IPIBLAD010.T1		—	X	X	X
IPIBLAD011.T1		—	X	—	—
IPIBLAD012.T1		—	X	—	—
IPIBLAD013.T1	Immune Desert	X	—	—	—
IPIBLAD014.T1	Immune Stromal Rich	X	—	—	—
IPIBLAD015.T1		—	X	—	—
IPIBLAD019.T1	Immune Stromal Rich	X	—	—	—
IPIBLAD020.T1	Myeloid Centric	X	—	—	—
IPIBLAD024.T1	Tcell Centric	X	—	—	—
IPIBLAD026.T1	Immune Rich	X	—	—	—
IPIBLAD028.T1	Immune Desert	X	—	—	—
IPIBLAD029.T1	Immune Rich	X	—	—	—
IPIBLAD030.T1	Immune Stromal Rich	X	—	—	—
IPIBLAD032.T1	Myeloid Centric	X	X	X	X
IPIBLAD033.T1	Immune Stromal Desert	X	X	X	X
IPIBLAD034.T1	Immune Desert	X	X	X	X
IPIBLAD036.T1	Immune Stromal Rich	X	X	X	X
IPIBLAD040.T1		—	X	—	—
IPIRC010.T1		—	X	—	X
IPIRC011.T1		—	X	—	X
IPIRC012.T1		—	X	—	X
IPIRC030.T1		—	X	X	X
IPIRC031.T1		—	X	X	—

IPIRC032.T1	Immune Rich	X	X	X	X
IPIRC034.T1		—	X	X	—
IPIRC035.T1	Immune Stromal Desert	X	—	—	—
IPIRC039.T1	Tcell Centric	X	—	—	—
IPIRC040.T2	Immune Stromal Rich	X	—	—	—
IPIRC048.T1	Tcell Centric	X	—	—	—
IPIRC054.T1	Immune Stromal Rich	X	—	—	—
IPIRC055.T1	Immune Stromal Rich	X	—	—	—
IPIRC057.T1	Immune Rich	X	X	X	X
IPIRC058.T1	Immune Rich	X	—	—	—
IPIRC060.T1	Immune Stromal Desert	X	X	X	X
IPIRC062.T1	Tcell Centric	X	—	—	—
IPIRC063.T1	Immune Stromal Rich	X	—	—	—
IPIRC064.T1	Tcell Centric	X	—	—	—
IPIRC067.T1	Immune Desert	X	—	—	—
IPIRC070.T1	Tcell Centric	X	—	—	—
IPIRC071.T2	Immune Desert	X	—	—	—
IPIRC072.T1	Immune Rich	X	X	X	—
IPIRC073.T1	Immune Desert	X	X	X	X
IPIRC074.T1	Immune Desert	X	—	—	X
IPIRC075.T1	Immune Rich	X	—	—	X
IPIRC076.T1	Immune Desert	X	—	—	X
IPIRC077.T1	Tcell Centric	X	X	X	—
IPIRC078.T2	Immune Stromal Desert	X	—	—	—
IPIRC079.T1	Myeloid Centric	X	X	X	—
IPIRC080.T1	Immune Stromal Desert	X	X	X	X
IPIRC087.T1	Immune Stromal Rich	X	—	—	—
IPIRC088.T1	Tcell Centric	X	X	X	—
IPIRC089.T1	Immune Rich	X	X	X	—
IPIRC090.T1	Immune Rich	X	X	X	X
IPIRC093.T1	Immune Desert	X	X	X	X
IPIGBM001.T1		—	X	X	X
IPIGBM002.T1		—	X	X	X
IPIGBM003.T1		—	X	X	X
IPIGBM004.T1		—	X	X	X

IPIGBM005.T1		—	X	X	X
IPIGBM006.T1	Immune Desert	X	X	X	X
IPIGBM007.T1		—	X	X	X
IPIGBM008.T1		—	X	X	—
IPIGBM011.T1		—	X	X	—
IPIGBM012.T1		—	X	—	
IPIGYN003.T1	Immune Stromal Rich	X	X	X	X
IPIGYN007.T1	Immune Desert	X	—	—	—
IPIGYN011.T1	Immune Stromal Desert	X	—	—	—
IPIGYN011.T2	Immune Stromal Desert	X	—	—	—
IPIGYN013.T1	Immune Rich	X	—	—	—
IPIGYN015.T1	Immune Desert	X	—	—	—
IPIGYN018.T2	Tcell Centric	X	X	X	X
IPIGYN019.T1	Immune Stromal Desert	X	—	—	—
IPIGYN021.T1	Immune Stromal Desert	X	X	X	X
IPIGYN026.T1	immune Stromal Desert	X	—	—	—
IPIGYN027.T1	Tcell Centric	X	—	—	—
IPIGYN040.T1	Immune Desert	X	—	—	—
IPIGYN043.T1	Myeloid Centric	X	—	—	—
IPIGYN045.T1	Myeloid Centric	X	X	X	X
IPIGYN048.T1	Tcell Centric	X	—	—	—
IPIGYN059.T1	Immune Desert	X	—	—	—
IPIGYN060.T1	Tcell Centric	X	—	—	—
IPIGYN064.T1	Tcell Centric	X	—	—	—
IPIGYN065.T1	Myeloid Centric	X	X	X	X
IPIGYN069.T1	Immune Stromal Rich	X	—	—	—
IPIGYN070.T1		—	X	X	X
IPIGYN073.T1		—	X	X	X
IPIGYN074.T1		—	X	X	X
IPIGYN078.T1	Immune Desert	X	—	—	—
IPIGYN082.T1	Immune Stromal Rich	X	—	—	—
IPIGYN083.T1	Immune Stromal Rich	X	—	—	—
IPIGYN084.T1	Immune Rich	X	X	X	X
IPIGYN085.T1	Immune Stromal Rich	X	—	—	—
IPIGYN086.T1	Immune Desert	X	X	X	X

IPIGYN087.T1	immune Stromal Desert	X	X	X	X
IPIGYN088.T1	Immune Desert	X	X	X	—
IPIGYN089.T1	Myeloid Centric	X	—	—	—
IPIGYN091.T1	immune Stromal Desert	X	—	—	—
IPIGYN093.T1	Immune Desert	X	—	—	—
IPIGYN095.T1	Myeloid Centric	X	X	X	X
IPIGYN096.T1	Immune Desert	X	X	X	X
IPIGYN098.T1	Myeloid Centric	X	—	—	—
IPIGYN099.T1	immune Stromal Desert	X	—	—	—
IPIGYN100.T1	Tcell Centric	X	X	X	X
IPIGYN103.T1	Immune Desert	X	—	—	—
IPIGYN104.T1	Tcell Centric	X	—	—	—
IPIGYN105.T1	Immune Desert	X	—	—	—
IPIGYN106.T1	Immune Desert	X	—	—	—
IPIGYN107.T1	Myeloid Centric	X	—	—	—
IPIGYN108.T1	Immune Desert	X	—	—	—
IPIGYN109.T1	Immune Desert	X	—	—	—
IPIGYN110.T1	Immune Stromal Rich	X	—	—	—
IPIGYN116.T1	Immune Desert	X	X	X	X
IPIGYN124.T1	Immune Stromal Rich	X	X	X	—
IPIGYN125.T1	Immune Desert	X	X	X	X
IPIGYN130.T1		—	X	X	X
IPIGYN134.T1		—	X	X	X
IPIHEP001.T1	immune Stromal Desert	X	X	X	X
IPIHEP003.T1	Tcell Centric	X	—	—	—
IPIHEP012.T1	Immune Stromal Rich	X	—	—	X
IPIHEP013.T1	Tcell Centric	X	X	X	X
IPIHEP015.T1	Myeloid Centric	X	X	X	X
IPIHEP016.T1	Immune Stromal Rich	X	—	—	—
IPIHEP017.T1	Immune Desert	X	—	X	X
IPIHEP021.T1		—	X	X	X
IPIHEP028.T1		—	X	X	X
IPIHEP029.T1		—	X	X	—
IPIHEP030.T1		—	X	X	X
IPIHEP031.T1		—	X	X	—

IPIHEP032.T1		—	X	X	X
IPIHEP033.T1		—	X	—	—
IPIHNSC021.T1	Immune Desert	X	X	X	—
IPIHNSC022.T1		—	X	X	X
IPIHNSC023.T1		—	X	X	X
IPIHNSC024.T1	Immune Desert	X	X	X	—
IPIHNSC025.T1		—	X	—	—
IPIHNSC026.T1	Tcell Centric	X	X	—	—
IPIHNSC029.T1		—	X	X	—
IPIHNSC030.T1	Tcell Centric	X	—	—	—
IPIHNSC031.T1	Tcell Centric	X	X	X	X
IPIHNSC050.T1	Myeloid Centric	X	X	X	X
IPIHNSC054.T1	Immune Stromal Rich	X	—	—	—
IPIHNSC056.T1	Immune Stromal Rich	X	X	—	—
IPIHNSC057.T1	Myeloid Centric	X	X	—	—
IPIHNSC059.T1	Tcell Centric	X	X	X	X
IPIHNSC060.T1	Immune Stromal Rich	X	X	—	—
IPIHNSC061.T1	Immune Desert	X	X	—	—
IPIHNSC062.T1	Immune Stromal Rich	X	X	—	—
IPIHNSC063.T1	Immune Stromal Rich	X	X	—	—
IPIHNSC065.T1	Immune Desert	X	X	X	X
IPIHNSC069.T1	Immune Desert	X	X	—	X
IPIHNSC070.T1	Immune Stromal Rich	X	X	X	—
IPIHNSC073.T1	Immune Stromal Desert	X	—	—	—
IPIHNSC074.T1	Immune Stromal Rich	X	—	—	*
IPIHNSC075.T1	Immune Stromal Rich	X	X	X	—
IPIHNSC076.T1	Immune Desert	X	X	X	X
IPIHNSC077.T1	Immune Stromal Rich	X	X	X	X
IPIHNSC078.T1	Immune Desert	X	X	X	X
IPIHNSC079.T1	Immune Stromal Rich	X	X	—	—
IPIHNSC083.T1		—	—	X	—
IPIKID022.T1		—	X	X	—
IPIKID023.T1		—	X	X	—
IPIKID027.T1	Immune Stromal Rich	X	X	X	X

IPIKID030.T1	Immune Rich	X	—	—	—
IPIKID031.T1	Immune Stromal Rich	X	X	X	X
IPIKID031.T2	Myeloid Centric	X	X	X	X
IPIKID033.T2	Tcell Centric	X	X	X	—
IPIKID043.T1	Immune Rich	X	X	X	X
IPIKID047.T1	Immune Rich	X	—	—	—
IPIKID048.T1	Tcell Centric	X	—	—	—
IPIKID049.T1	Tcell Centric	X	—	—	—
IPIKID050.T1	Tcell Centric	X	—	—	—
IPIKID051.T1	Immune Rich	X	—	—	—
IPIKID052.T1	Tcell Centric	X	—	—	—
IPIKID057.T1	Immune Rich	X	—	—	—
IPIKID060.T1	Immune Rich	X	—	—	—
IPIKID061.T1	Myeloid Centric	X	—	—	—
IPIKID062.T1	Immune Rich	X	—	—	—
IPIKID066.T1	Immune Rich	X	—	—	—
IPIKID067.T1	Immune Rich	X	—	—	—
IPIKID069.T1	Tcell Centric	X	—	—	—
IPIKID070.T1	Immune Stromal Rich	X	X	—	—
IPIKID071.T1	Immune Rich	X	—	—	—
IPIKID073.T1	Immune Desert	X	X	X	X
IPIKID074.T1	Immune Rich	X	—	—	—
IPIKID076.T1	Immune Rich	X	—	—	—
IPIKID077.T1	Immune Rich	X	—	—	—
IPIKID078.T1	Immune Rich	X	—	—	—
IPIKID079.T1	Immune Rich	X	—	—	—
IPIKID083.T1	Immune Rich	X	X	X	X
IPIKID084.T1	Immune Rich	X	X	X	X
IPIKID085.T1	Immune Rich	X	X	—	X
IPIKID086.T1	Immune Rich	X	X	—	X
IPIKID086.T2	Immune Rich	X	X	X	X
IPIKID088.T1	Myeloid Centric	X	X	X	X
IPIKID089.T1	Immune Rich	X	X	—	X
IPIKID090.T1	Immune Rich	X	X	X	X
IPIKID091.T1	Immune Rich	X	X	—	—

IPIKID092.T1	Immune Stromal Rich	X	X	X	X
IPIKID093.T1	Immune Rich	X	X	—	X
IPIKID093.T2	Myeloid Centric	X	X	—	X
IPIKID094.T1	Immune Stromal Rich	X	X	—	X
IPIKID094.T2	Myeloid Centric	X	X	—	X
IPIKID095.T1	Immune Stromal Rich	X	X	X	—
IPIKID097.T1	Immune Desert	X	X	X	—
IPIKID098.T1	Myeloid Centric	X	—	—	—
IPIKID099.T1	Immune Rich	X	—	—	—
IPIKID100.T1	Immune Rich	X	X	X	X
IPIKID102.T1	Immune Rich	X	X	X	X
IPIKID103.T1	Immune Rich	X	X	X	X
IPIKID104.T1	Immune Rich	X	X	X	X
IPIKID106.T1	Tcell Centric	X	—	—	—
IPIKID108.T1	Immune Rich	X	X	—	X
IPIKID109.T1	Immune Stromal Rich	X	X	X	X
IPIKID110.T1	Immune Rich	X	X	X	X
IPIILUNG001.T1		—	X	X	—
IPIILUNG002.T1		—	X	X	—
IPIILUNG003.T1	Immune Stromal Rich	X	X	X	—
IPIILUNG004.T1		—	X	X	—
IPIILUNG006.T1	Immune Rich	X	X	X	—
IPIILUNG008.T1	Immune Rich	X	—	—	—
IPIILUNG009.T1	Immune Desert	X	—	—	—
IPIILUNG011.T1	Immune Rich	X	—	—	—
IPIILUNG013.T1	Tcell Centric	X	—	—	—
IPIILUNG014.T1	Immune Rich	X	X	X	X
IPIILUNG015.T1	Immune Rich	X	X	X	X
IPIILUNG021.T1	Immune Desert	X	—	—	—
IPIILUNG024.T1		—	X	X	—
IPIILUNG025.T1	immune Stromal Desert	X	X	X	—
IPIILUNG030.T1	Tcell Centric	X	—	—	—
IPIILUNG030.T2	Tcell Centric	X	—	—	—
IPIILUNG031.T1	Tcell Centric	X	X	X	X
IPIILUNG037.T1	Tcell Centric	X	X	X	—

IPI LUNG037.T2	immune Stromal Desert	X	—	—	
IPI LUNG041.T1	Tcell Centric	X	X	X	X
IPI LUNG049.T1	Tcell Centric	X	—	—	—
IPI LUNG051.T1	Tcell Centric	X	—	—	—
IPI LUNG052.T1	Tcell Centric	X	—	—	—
IPI LUNG054.T1	Immune Rich	X	—	—	—
IPI LUNG055.T1	immune Stromal Desert	X	—	—	—
IPI LUNG056.T1	Tcell Centric	X	X	X	X
IPI LUNG058.T1	Immune Rich	X	—	—	—
IPI LUNG062.T1	Immune Rich	X	—	—	—
IPI LUNG064.T1	Tcell Centric	X	—	—	—
IPI LUNG066.T1	Immune Stromal Rich	X	—	—	—
IPI LUNG069.T1	Immune Rich	X	—	—	—
IPI LUNG070.T1	Myeloid Centric	X	—	—	—
IPI LUNG074.T1	Immune Rich	X	—	—	—
IPI LUNG076.T1	Immune Rich	X	—	—	—
IPI LUNG077.T1	Immune Stromal Rich	X	—	—	—
IPI LUNG080.T1	Immune Stromal Rich	X	X	X	X
IPI LUNG081.T1	Myeloid Centric	X	—	—	—
IPI LUNG085.T1	Immune Rich	X	—	—	—
IPI LUNG088.T1	Immune Stromal Rich	X	X	X	X
IPI MEL025.T1		—	X	X	—
IPI MEL027.T1		—	X	X	—
IPI MEL031.T1		—	X	X	—
IPI MEL056.T1	Tcell Centric	X	X	X	X
IPI MEL069.T1	Tcell Centric	X	—	X	—
IPI MEL105.T2	Tcell Centric	X	X	X	—
IPI MEL106.T1		—	X	X	X
IPI MEL108.T1	immune Stromal Desert	X	X	X	—
IPI MEL113.T1	Tcell Centric	X	X	X	—
IPI MEL125.T1		—	X	X	—
IPI MEL127.T1		—	X	X	—
IPI MEL131.T1		—	X	X	X
IPI MEL141.T2	immune Stromal Desert	X	—	—	—

IPIMEL146.T1	Tcell Centric	X	X	X	X
IPIMEL175.T1	Immune Rich	X	—	—	—
IPIMEL181.T1		—	X	—	
IPIMEL191.T1		—	X	—	
IPIMEL208.T1	Tcell Centric	X	X	X	—
IPIMEL212.T1	Tcell Centric	X	—	—	—
IPIMEL221.T1	Tcell Centric	X	—	—	—
IPIMEL226.T1		—	X	—	
IPIMEL252.T1	Tcell Centric	X	—	—	—
IPIMEL253.T1	Immune Rich	X	X	X	X
IPIMEL255.T1	Tcell Centric	X	—	—	—
IPIMEL271.T1	Myeloid Centric	X	—	—	—
IPIMEL273.T1	Tcell Centric	X	—	—	—
IPIMEL274.T1	Tcell Centric	X	X	X	X
IPIMEL283.T1	Myeloid Centric	X	—	—	—
IPIMEL288.T1	Tcell Centric	X	X	X	X
IPIMEL296.T1	Tcell Centric	X	X	X	X
IPIMEL298.T1	Tcell Centric	X	—	—	—
IPIMEL301.T1	Tcell Centric	X	—	—	—
IPIMEL321.T1	Immune Desert	X	—	—	—
IPIMEL325.T1	Immune Desert	X	—	—	—
IPIMEL343.T1	Myeloid Centric	X	—	—	—
IPIMEL348.T1	Immune Desert	X	—	—	—
IPIMEL353.T1	Tcell Centric	X	—	—	—
IPIMELB002.T1	Tcell Centric	X	—	—	—
IPIMELB003.T1	Immune Stromal Desert	X	—	—	—
IPIMELB003.T2	Immune Desert	X	—	—	—
IPIMELB003.T3	Myeloid Centric	X	—	—	—
IPIPDAC001.T1		—	X	—	—
IPIPDAC002.T1		—	X	X	—
IPIPDAC003.T1		—	X	X	X
IPIPDAC004.T1		—	X	X	X
IPIPDAC005.T1		—	X	X	X
IPIPDAC013.T1		—	X	X	X
IPIPDAC021.T1		—	X	X	X
IPIPDAC029.T1		—	X	X	X
IPIPDAC032.T1		—	X	—	—

IIPDAC034.T1		—	X	—	—
IIPDAC037.T1	Myeloid Centric	X	X	X	X
IIPDAC041.T1		—	X	—	X
IIPDAC047.T1		—	X	X	X
IIPDAC050.T1		—	X	X	X
IIPDAC051.T1		—	X	X	X
IIPDAC052.T1		—	X	—	
IIPDAC053.T1		—	X	X	X
IIPDAC056.T1		—	X	X	—
IIPDAC058.T1		—	X	X	—
IIPNET005.T1	immune Stromal Desert	X	X	X	—
IIPNET007.T1	Myeloid Centric	X	X	X	X
IIPNET009.T1	Immune Desert	X	—	—	—
IIPNET010.T1	immune Stromal Desert	X	X	X	X
IIPNET014.T1	Immune Desert	X	—	—	—
IIPNET016.T1	Myeloid Centric	X	X	X	X
IISRC001.T1	Myeloid Centric	X	X	X	X
IISRC003.T1	Myeloid Centric	X	X	X	X
IISRC004.T1	Immune Stromal Rich	X	X	X	X
IISRC009.T1	Immune Stromal Rich	X	X	X	X
IISRC011.T1	Immune Stromal Rich	X	—	—	—
IISRC012.T1	Myeloid Centric	X	—	—	—
IISRC013.T1	Myeloid Centric	X	—	—	—
IISRC014.T1	Immune Stromal Rich	X	—	—	—
IISRC015.T1	Immune Desert	X	X	X	X
IISRC019.T1	Myeloid Centric	X	—	—	—
IISRC021.T1		—	X	X	—
IISRC022.T1		—	X	—	—
IISRC024.T1		—	X	X	—
IISRC025.T1		—	X	—	—
IISRC026.T1		—	X	X	X
IISRC027.T1		—	X	X	—
IISRC035.T1		—	X	X	—

Supplementary Table 2: List of UCSF Immunoprofiler Initiative samples and associated flow cytometry data and bulk RNA seq data used for 3 features clustering. List of each tumor biopsy included in Figure 1A and 1C with their associated different flow cytometry data or 'all live' bulk RNA sequencing compartment. A cross mean that the given sample have data associated and it has been used in this manuscript. Cluster assignment for the 3 feature clustering is provided.

		Bulk RNAseq		Multi-parametric flow cytometry								
Sample		Live	Tcell	Myeloid	Tcell Panel	Tcell Panel	Tcell Panel	Myeloid Panel	Myeloid Panel	Myeloid panel	Myeloid panel	Myeloid panel
	10 Feature Archetypes				Treg:CD45+,CD3+, CD4+,CD8- ,CD25+,FOXP3+	CD4+ Tconv: CD45+,CD3+,CD4+,CD 8-,FOXP3-,CD25+/-	CD8+ Tconv: CD45+,CD3+,CD4- ,CD8+	Monocytes:CD45+,CD 3-,CD19-,CD20-CD56- HLA-DR+, CD14+,CD16- ,CD11c+,BDCA3- ,BDCA1-,CD64lo	Macrophages:CD45+, CD3-,CD19-,CD20- CD56- HLA-DR+, CD14+,CD16+,CD11c+ ,BDCA3-,BDCA1- ,CD64hi	cDC2:CD45+,CD3- ,CD19-,CD20-CD56- HLA-DR+, CD14+/-,CD16- ,CD11c+,BDCA3- ,BDCA1+	cDC1:CD45+,CD3- ,CD19-,CD20-CD56- HLA-DR+, CD14-,CD16- ,CD11c+,BDCA3+,BDC A1-	pDCs:CD45+,CD3- ,CD19-,CD20-CD56- HLA-DR+, CD14-,CD16- ,CD11c-,BDCA3-,BDCA1- ,BDCA4+
IPIBLAD005.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	—	X	X	X	—	—
IPIBLAD006.T1	Immune Stromal CD8 Bias	X	X	X	—	—	—	X	X	X	X	—
IPIBLAD013.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD014.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD019.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD020.T1	Myeloid Centric DC2 Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD024.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD026.T1	Immune Rich CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD028.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD029.T1	Tcell Centric Macro Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD030.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
IPIBLAD032.T1		X	—	—	X	X	X	—	—	—	—	—
IPIBLAD033.T1	Immune Desert CD4 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
IPIBLAD034.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	X
IPIBLAD036.T1	Tcell Centric Macro Bias	X	X	X	X	X	X	X	X	X	X	X
IPICRC032.T1	Myeloid Centric DC2 Bias	X	—	—	X	X	X	X	X	X	X	—
IPICRC035.T1	Immune Desert Mono Bias	X	X	—	X	X	X	X	X	X	X	—
IPICRC039.T1		X	X	X	—	—	—	—	—	—	—	—
IPICRC040.T2	Myeloid Centric DC1 Bias	X	X	—	—	—	—	X	X	X	X	—
IPICRC048.T1	Tcell Centric Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPICRC054.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	—	X	X	X	X	—
IPICRC055.T1	Immune Stromal CD4 Bias	X	X	X	—	—	—	X	X	X	X	—
IPICRC057.T1	Immune Desert Mono Bias	X	X	—	X	X	X	X	X	X	X	—
IPICRC058.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPICRC060.T1		X	—	X	X	X	X	—	—	—	—	—
IPICRC062.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPICRC063.T1	Immune Stromal CD8 Bias	X	X	X	—	—	—	X	X	X	X	—
IPICRC064.T1	Immune Desert Mono Bias	X	X	X	—	—	—	X	X	X	X	—
IPICRC067.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	X
IPICRC070.T1	Tcell Centric Macro Bias	X	X	X	—	—	—	X	X	X	X	X
IPICRC071.T2	Immune Desert CD4 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
IPICRC072.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	X
IPICRC073.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
IPICRC074.T1	Immune Desert Mono Bias	X	X	X	—	—	X	X	X	X	X	X
IPICRC075.T1	Myeloid Centric DC2 Bias	X	X	X	—	—	X	X	X	X	X	X
IPICRC076.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	X	X	X	X	X	X
IPICRC077.T1	Immune Desert Mono Bias	X	X	X	X	X	—	X	X	X	X	X
IPICRC078.T2	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	X
IPICRC079.T1	Myeloid Centric DC1 Bias	X	X	X	X	X	X	X	X	X	X	X
IPICRC080.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	X
IPICRC087.T1	Immune Rich CD4 Macro Bias	X	X	X	X	X	X	—	—	—	—	—
IPICRC088.T1	Immune Rich CD4 Macro Bias	X	—	—	X	X	X	X	X	X	X	X
IPICRC089.T1	Immune Rich CD8 Macro Bias	X	X	X	X	X	X	—	—	—	—	—
IPICRC090.T1	Immune Rich CD4 Macro Bias	X	—	X	X	X	X	X	X	X	X	X
IPICRC093.T1	Myeloid Centric DC1 Bias	X	X	X	X	X	X	X	X	X	X	X
IPIGBM006.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
IPIGYN003.T1	Myeloid Centric DC1 Bias	X	—	X	X	X	X	X	X	X	X	—
IPIGYN007.T1		X	—	X	—	—	—	—	—	—	—	—
IPIGYN011.T1	Immune Desert Mono Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN011.T2	Immune Desert Mono Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN013.T1	Immune Rich CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN015.T1	Immune Desert CD4 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
IPIGYN018.T2	Immune Desert Mono Bias	X	—	—	X	X	X	X	X	X	X	—
IPIGYN019.T1	Immune Desert CD4 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
IPIGYN021.T1	Immune Desert CD4 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
IPIGYN026.T1		X	X	X	—	—	—	—	—	—	—	—
IPIGYN027.T1	Tcell Centric Macro Bias	X	X	—	—	—	—	X	X	X	X	—
IPIGYN040.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN043.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN045.T1	Immune Stromal CD4 Bias	X	—	X	X	X	X	X	X	X	X	—
IPIGYN048.T1		X	X	X	—	—	—	—	—	—	—	—
IPIGYN059.T1	Immune Desert CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN060.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN064.T1		X	X	X	—	—	—	—	—	—	—	—
IPIGYN065.T1	Immune Stromal CD8 Bias	X	—	X	X	X	X	X	X	X	X	—
IPIGYN069.T1	Immune Stromal CD4 Bias	X	X	X	—	—	—	X	X	X	X	—
IPIGYN078.T1	Myeloid Centric DC1 Bias	X	X	X	—	X	—	X	X	X	X	—
IPIGYN082.T1	Immune Stromal CD4 Bias	X	X	X	—	X	—	X	X	X	X	—

PIGYN083.T1	Immune Desert CD4 Macro Bias	X	X	X	—	X	—	X	X	X	X	—
PIGYN084.T1	Immune Rich CD8 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PIGYN085.T1	Immune Stromal CD8 Bias	X	X	X	—	X	—	X	X	X	X	—
PIGYN086.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
PIGYN087.T1	Myeloid Centric DC1 Bias	X	—	X	X	X	X	X	X	X	X	—
PIGYN088.T1	Immune Stromal CD8 Bias	X	X	—	X	X	—	X	X	X	X	—
PIGYN089.T1	Immune Stromal CD4 Bias	X	X	X	X	X	—	X	X	X	X	—
PIGYN091.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PIGYN093.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PIGYN095.T1	Immune Desert CD4 Macro Bias	X	—	—	X	X	X	X	X	X	X	X
PIGYN096.T1	Immune Desert CD8 Macro Bias	X	—	X	X	X	X	X	X	X	X	X
PIGYN098.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	X
PIGYN099.T1	Immune Desert CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	X
PIGYN100.T1	Immune Rich CD8 Macro Bias	X	—	X	X	X	X	X	X	X	X	X
PIGYN103.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN104.T1	Tcell Centric Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN105.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN106.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN107.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN108.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN109.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN110.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN116.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIGYN124.T1	Myeloid Centric DC1 Bias	X	X	X	X	X	X	—	—	—	—	—
PIGYN125.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIHEP001.T1	Immune Desert Mono Bias	X	—	—	X	X	X	X	X	X	X	—
PIHEP003.T1	Tcell Centric Macro Bias	X	X	X	X	X	X	X	X	X	X	—
PIHEP012.T1		X	X	X	X	X	X	—	—	—	—	—
PIHEP013.T1	Tcell Centric Macro Bias	X	X	X	X	X	X	X	X	X	X	—
PIHEP015.T1	Myeloid Centric DC2 Bias	X	—	—	X	X	X	X	X	X	X	—
PIHEP016.T1	Tcell Centric DC Rich	X	X	—	X	X	X	X	X	X	X	—
PIHEP017.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	—
PIHNSC021.T1	Immune Desert CD8 Macro Bias	X	—	X	—	X	X	X	X	X	X	—
PIHNSC024.T1	Immune Desert CD8 Macro Bias	X	—	—	—	X	X	X	X	X	X	—
PIHNSC026.T1		X	—	X	—	X	X	—	—	—	—	—
PIHNSC030.T1	Tcell Centric Macro Bias	X	X	—	—	—	—	X	X	X	X	—
PIHNSC031.T1		X	—	X	X	X	X	—	—	—	—	—
PIHNSC050.T1	Immune Stromal CD4 Bias	X	X	X	X	X	X	X	X	X	X	—
PIHNSC054.T1		X	X	X	—	—	—	—	—	—	—	—
PIHNSC056.T1	Tcell Centric Macro Bias	X	—	—	—	—	—	X	X	X	X	—
PIHNSC057.T1	Immune Stromal CD4 Bias	X	X	X	X	—	X	X	X	X	X	—
PIHNSC059.T1	Tcell Centric Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PIHNSC060.T1	Tcell Centric Macro Bias	X	X	X	X	—	—	X	X	X	X	—
PIHNSC061.T1	Immune Desert Mono Bias	X	X	X	X	—	—	X	X	X	X	—
PIHNSC062.T1	Immune Stromal CD8 Bias	X	X	X	X	X	—	X	X	X	X	—
PIHNSC063.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
PIHNSC065.T1	Immune Stromal CD8 Bias	X	—	—	X	X	X	X	X	X	X	—
PIHNSC069.T1	Immune Desert CD4 Macro Bias	X	X	X	X	X	X	X	X	X	X	—
PIHNSC070.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
PIHNSC073.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	—
PIHNSC074.T1	Tcell Centric Macro Bias	X	X	X	X	X	X	X	X	X	X	X
PIHNSC075.T1	Immune Stromal CD4 Bias	X	X	—	X	X	—	X	X	X	X	—
PIHNSC076.T1	Immune Stromal CD8 Bias	X	—	X	X	X	X	X	X	X	X	X
PIHNSC077.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	X
PIHNSC078.T1	Myeloid Centric DC1 Bias	X	X	X	X	X	X	X	X	X	X	X
PIHNSC079.T1	Tcell Centric Macro Bias	X	X	X	X	—	—	X	X	X	X	X
PIKID027.T1	Immune Rich CD8 Macro Bias	X	—	X	X	X	X	X	X	X	X	—
PIKID030.T1	Immune Rich CD8 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
PIKID031.T1		X	—	—	X	X	X	—	—	—	—	—
PIKID031.T2		X	—	X	X	X	X	—	—	—	—	—
PIKID033.T2		X	—	—	—	X	X	—	—	—	—	—
PIKID043.T1	Immune Rich CD8 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PIKID047.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PIKID048.T1	Immune Rich CD8 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
PIKID049.T1	Tcell Centric Macro Bias	X	X	—	—	—	—	X	X	X	X	—
PIKID050.T1	Immune Rich CD8 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
PIKID051.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PIKID052.T1		X	X	X	—	—	—	—	—	—	—	—
PIKID057.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	—	X	X	X	X	—
PIKID060.T1	Immune Desert Mono Bias	X	X	X	—	—	—	X	X	X	X	—
PIKID061.T1	Immune Stromal CD4 Bias	X	X	X	—	—	—	X	X	X	X	—
PIKID062.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	—	—	X	X	—
PIKID066.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PIKID067.T1	Immune Desert CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—

PIKID069.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	—	X	X	X	X	—
PIKID070.T1	Immune Stromal CD8 Bias	X	X	X	X	X	—	X	X	X	X	—
PIKID071.T1	Immune Rich CD8 Mono Bias	X	X	X	—	X	—	X	X	X	X	—
PIKID073.T1	Immune Desert CD4 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PIKID074.T1	Immune Rich CD8 Macro Bias	X	X	X	—	X	—	X	X	X	X	—
PIKID076.T1	Immune Rich CD8 Macro Bias	X	X	X	—	X	—	X	X	X	X	—
PIKID077.T1	Immune Rich CD8 Mono Bias	X	X	X	—	X	—	X	X	X	X	—
PIKID078.T1	Immune Rich CD8 Macro Bias	X	X	X	—	X	—	X	X	X	X	—
PIKID079.T1	Immune Rich CD8 Macro Bias	X	X	X	—	X	—	X	X	X	X	—
PIKID083.T1	Immune Rich CD8 Macro Bias	X	X	X	X	X	X	X	X	X	X	—
PIKID084.T1	Immune Rich CD8 Macro Bias	X	X	X	X	X	X	X	X	X	X	—
PIKID085.T1	Tcell Centric Macro Bias	X	X	X	X	X	X	X	X	X	X	—
PIKID086.T1	Immune Rich CD8 Macro Bias	X	X	X	X	—	—	X	X	X	X	—
PIKID086.T2	Tcell Centric Macro Bias	X	—	X	X	X	X	X	X	X	X	—
PIKID088.T1	Immune Rich CD8 Mono Bias	X	—	X	X	X	X	—	—	—	—	—
PIKID089.T1	Immune Rich CD8 Mono Bias	X	X	X	X	—	—	X	X	X	X	—
PIKID090.T1	Immune Rich CD8 Mono Bias	X	—	—	X	X	X	—	—	—	—	—
PIKID091.T1	Immune Rich CD8 Macro Bias	X	X	X	X	—	—	X	X	X	X	—
PIKID092.T1	Immune Rich CD8 Macro Bias	X	X	X	X	—	—	X	X	X	X	X
PIKID093.T1	Immune Rich CD8 Macro Bias	X	X	X	—	X	X	X	X	X	X	X
PIKID093.T2	Immune Stromal CD8 Bias	X	X	X	—	—	—	—	—	—	—	—
PIKID094.T1	Immune Stromal CD8 Bias	X	X	X	—	—	—	X	X	X	X	X
PIKID094.T2	Myeloid Centric DC2 Bias	X	X	X	—	—	—	X	X	X	X	X
PIKID095.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	—	X	X	X	X	X
PIKID097.T1	Myeloid Centric DC1 Bias	X	X	X	—	—	—	X	X	X	X	X
PIKID098.T1	Immune Stromal CD4 Bias	X	X	X	—	—	—	X	X	X	X	X
PIKID099.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	X
PIKID100.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	X	X	X	X	X	X
PIKID102.T1	Immune Rich CD8 Macro Bias	X	X	X	—	—	X	X	X	X	X	X
PIKID103.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	X	—	—	—	—	—
PIKID104.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	X	—	—	—	—	—
PIKID106.T1	Immune Desert Mono Bias	X	X	—	—	—	—	X	X	X	X	X
PIKID108.T1	Immune Rich CD4 Macro Bias	X	X	X	—	—	X	—	—	—	—	—
PIKID109.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	X	—	—	—	—	—
PIKID110.T1	Immune Desert CD4 Macro Bias	X	—	X	—	—	X	—	—	—	—	—
PILUNG003.T1	Immune Rich CD8 Mono Bias	X	—	X	X	X	X	—	—	—	—	—
PILUNG006.T1	Immune Rich CD8 Macro Bias	X	X	X	X	X	X	X	X	X	X	—
PILUNG008.T1	Immune Rich CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PILUNG009.T1	Immune Desert Mono Bias	X	X	X	—	—	—	X	X	X	X	—
PILUNG011.T1	Myeloid Centric DC2 Bias	X	X	—	—	—	—	X	X	X	X	—
PILUNG013.T1	Immune Rich CD4 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PILUNG014.T1	Immune Rich CD4 Macro Bias	X	X	X	X	—	X	X	X	X	X	—
PILUNG015.T1	Immune Rich CD4 Macro Bias	X	X	X	X	—	X	X	X	X	X	—
PILUNG021.T1	Myeloid Centric DC1 Bias	X	X	X	—	—	X	X	X	X	X	—
PILUNG025.T1	Immune Rich CD4 Macro Bias	X	—	—	X	X	X	—	—	—	—	—
PILUNG030.T1	Immune Rich CD4 Macro Bias	X	X	X	X	—	X	X	X	X	X	—
PILUNG030.T2	Tcell Centric DC Rich	X	X	X	—	—	—	X	X	X	X	—
PILUNG031.T1	Immune Rich CD8 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PILUNG037.T1	Tcell Centric Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PILUNG037.T2	Immune Rich CD4 Macro Bias	X	—	—	—	—	—	—	—	—	—	—
PILUNG041.T1	Immune Rich CD4 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PILUNG049.T1	Immune Desert CD4 Macro Bias	X	X	—	—	X	—	X	X	X	X	—
PILUNG051.T1	Immune Desert Mono Bias	X	X	X	—	X	X	X	X	X	X	—
PILUNG052.T1	Tcell Centric Macro Bias	X	X	X	*	—	—	X	X	X	X	—
PILUNG054.T1	Immune Rich CD4 Macro Bias	X	X	X	—	X	—	X	X	X	X	—
PILUNG055.T1	Immune Desert CD4 Macro Bias	X	X	—	—	X	—	X	X	X	X	—
PILUNG056.T1	Tcell Centric Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PILUNG058.T1	Immune Rich CD4 Macro Bias	X	X	—	—	X	—	X	X	X	X	—
PILUNG062.T1	Immune Stromal CD4 Bias	X	X	—	—	X	—	X	X	X	X	—
PILUNG064.T1	Immune Rich CD4 Macro Bias	X	—	—	—	—	—	—	—	—	—	—
PILUNG066.T1	Immune Rich CD4 Macro Bias	X	X	—	—	X	—	—	—	—	—	—
PILUNG069.T1	Immune Rich CD4 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
PILUNG070.T1	Immune Rich CD4 Macro Bias	X	X	—	—	—	—	—	—	—	—	—
PILUNG074.T1	Immune Stromal CD4 Bias	X	X	—	X	—	X	X	X	X	X	—
PILUNG076.T1	Immune Rich CD4 Macro Bias	X	X	—	—	—	—	X	X	X	X	—
PILUNG077.T1	Immune Rich CD4 Macro Bias	X	X	—	—	—	—	—	—	—	—	—
PILUNG080.T1	Immune Rich CD4 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PILUNG081.T1	Immune Stromal CD8 Bias	X	X	—	—	—	—	X	X	X	X	—
PILUNG085.T1	Immune Rich CD4 Macro Bias	X	X	—	—	—	—	—	—	—	—	—
PILUNG088.T1	Tcell Centric Macro Bias	X	—	X	X	X	X	X	X	X	X	X
PIMEL056.T1	Immune Desert Mono Bias	X	—	—	X	X	X	X	X	X	X	—
PIMEL069.T1	Tcell Centric DC Rich	X	X	X	—	—	—	X	X	X	X	—
PIMEL105.T2	Immune Desert Mono Bias	X	—	—	X	X	X	X	X	X	X	—
PIMEL108.T1	Immune Rich CD4 Macro Bias	X	X	X	X	—	X	—	—	—	—	—
PIMEL113.T1	Myeloid Centric DC1 Bias	X	—	—	X	X	X	X	X	X	X	—
PIMEL141.T2	Immune Rich CD4 Macro Bias	X	X	X	—	—	—	—	—	—	—	—
PIMEL146.T1	Immune Rich CD4 Macro Bias	X	—	—	X	X	X	—	—	—	—	—
PIMEL175.T1	Immune Rich CD4 Macro Bias	X	—	—	—	—	—	—	—	—	—	—
PIMEL208.T1	Immune Desert Mono Bias	X	—	—	X	X	X	X	X	X	X	—
PIMEL212.T1	Tcell Centric Macro Bias	X	X	—	X	X	—	X	X	X	X	—

PI MEL221.T1	Immune Desert Mono Bias	X	X	X	X	X	X	X	X	X	X	—
PI MEL252.T1	Immune Rich CD4 Macro Bias	X	X	—	—	X	X	X	X	X	X	—
PI MEL253.T1	Immune Rich CD4 Macro Bias	X	X	—	X	X	X	X	X	X	X	—
PI MEL255.T1	Immune Rich CD4 Macro Bias	X	X	X	—	X	—	X	X	—	X	—
PI MEL271.T1	Myeloid Centric DC1 Bias	X	X	X	X	X	—	X	X	X	X	—
PI MEL273.T1	Tcell Centric DC Rich	X	X	X	—	X	—	X	X	X	X	—
PI MEL274.T1	Tcell Centric Macro Bias	X	X	X	—	X	X	X	X	X	X	—
PI MEL283.T1	Immune Stromal CD8 Bias	X	X	X	—	X	—	X	X	X	X	—
PI MEL288.T1	Tcell Centric DC Rich	X	X	X	X	X	X	X	X	X	X	—
PI MEL296.T1		X	*	—	X	—	—	—	—	—	—	—
PI MEL298.T1	Tcell Centric DC Rich	X	X	X	—	—	X	X	X	X	X	—
PI MEL301.T1		X	X	X	—	—	—	—	—	—	—	—
PI MEL321.T1	Immune Stromal CD8 Bias	X	X	X	—	—	X	X	X	X	X	—
PI MEL325.T1		X	X	X	—	—	X	X	X	X	X	—
PI MEL343.T1	Immune Stromal CD8 Bias	X	X	X	—	—	—	X	X	X	X	X
PI MEL348.T1	Immune Desert Mono Bias	X	X	X	—	—	—	—	—	—	—	—
PI MEL353.T1	Immune Desert Mono Bias	X	X	X	—	—	—	X	X	X	X	X
PI MELB002.T1	Immune Rich CD8 Macro Bias	X	X	X	—	X	X	X	X	X	X	—
PI MELB003.T1	Immune Desert CD8 Macro Bias	X	X	X	—	—	—	X	X	X	X	—
PI MELB003.T2		X	X	X	—	—	—	—	—	—	—	—
PI MELB003.T3		X	X	X	—	—	—	—	—	—	—	—
PI PDAC037.T1	Myeloid Centric DC1 Bias	X	X	X	X	X	X	X	X	X	X	—
PI PNET005.T1	Immune Desert Mono Bias	X	X	X	X	X	—	X	X	X	X	—
PI PNET007.T1		X	—	—	X	X	X	—	—	—	—	—
PI PNET009.T1	Immune Desert Mono Bias	X	X	X	—	X	X	X	X	X	X	—
PI PNET010.T1	Immune Desert Mono Bias	X	—	—	X	X	X	X	X	X	X	—
PI PNET014.T1	Myeloid Centric DC1 Bias	X	X	X	—	X	X	X	X	X	X	X
PI PNET016.T1	Myeloid Centric DC1 Bias	X	—	—	X	X	X	X	X	X	X	—
PI SRC001.T1	Immune Stromal CD8 Bias	X	X	—	X	X	X	X	X	X	X	—
PI SRC003.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
PI SRC004.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
PI SRC009.T1	Immune Stromal CD8 Bias	X	X	X	X	X	X	X	X	X	X	—
PI SRC011.T1	Immune Rich CD8 Mono Bias	X	X	X	—	—	—	X	X	X	X	—
PI SRC012.T1	Immune Stromal CD8 Bias	X	X	X	—	—	—	X	X	X	X	—
PI SRC013.T1	Immune Stromal CD8 Bias	X	X	X	—	—	—	X	X	X	X	X
PI SRC014.T1	Immune Stromal CD8 Bias	X	X	X	—	—	—	X	X	X	X	X
PI SRC015.T1	Immune Desert CD8 Macro Bias	X	—	—	X	X	X	X	X	X	X	—
PI SRC019.T1		X	—	—	—	—	—	—	—	—	—	—

Supplementary Table 3: List of UCSF Immunoprofiler Initiative samples and associated flow cytometry data and bulk RNA seq data used for 6 and 10 feature clustering. List of each tumor biopsy included in the 6 and 10 feature clustering and their associated different flow cytometry data and or different bulk RNA sequencing compartment. A cross mean that the given sample have data associated and it has been used in this manuscript. Cluster assignment for the 10-feature clustering is also provided

[illegible]

Supplementary Table 5: List of genes used for the phenotype gene signatures. List of each gene included in the different phenotype gene signatures used to perform scoring and hierarchical clustering of the different tumor immune archetypes.

Material and Methods

KEY RESOURCES 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

Chemicals, Peptides, and Recombinant Proteins		
N/A		
Critical Commercial Assays		
N/A		
Deposited Data		
GEO		
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

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		

METHOD 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.

Human tissue digestion and Multiparametric Flow Cytometry staining and sorting

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

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

Cell sorting for bulk RNA sequencing

Bulk RNAseq compartment	Gating Strategy	Potential Caveats
All viable cells 'Live' compartment	All cells Viability dye negative LIVE/DEAD™Fixable Aqua Dead Cell Stain Cat:L34957	This was performed after tissue enzymatic digestion and allow us to sequence only viable cells after isolation. Therefore it is slightly different than bulk RNAseq from whole tissue prior digestion as it was done in the TCGA dataset.
Conventional CD4+ and CD8+ T cells compartment	CD4+ and CD8+ Tconv: CD45+, CD3+ , CD4+, CD8-, CD25- CD19/20-, CD56-	Using this gating strategy we aimed to enrich for both Conventional CD4+ and CD8+ T cells but we don't separate CD4+ and CD8+
Regulatory CD4+ T cells compartment	Treg: CD45+, CD3+, CD4+, CD8- , CD25+, CD19/20-, CD56-	Using this gating strategy we aimed to enrich for Tregulatory T cells using 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+, Cd11b+/-	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 permissive and this compartment may contain other tumor associated non immune cell types

For cell sorting single cell suspension was stained with cell surface antibodies mix diluted in BV stain buffer (BD Biosciences) following manufacturer instruction for 30 minutes on ice in the dark and subsequently washed three time and resuspended in of sort buffer (PBS/2% FCS/2mM EDTA) after filtering in 100um filter (thermo). When possible, a maximum of 6 different tumor associated cell subsets were sorted with a BD FACSAria Fusion into 1.5mL Eppendorf tubes containing 150ul of lysis buffer used for mRNA extraction (Invitrogen). The 6 different tumor associated cell types between

5000 and 50000 cells were sorted based their protein expression profile as describe here: 'live cells' all cells negative for the viability dye ; Tconv: negative for viability dye CD45+ CD3+ CD19- CD56- CD25- ; Treg: negative for viability dye CD45+ CD3+ CD19- CD56- CD4+ CD25+ ; Myeloid: negative for viability dye CD45+ CD3- CD19- CD56- HLADR+ CD11b+/- ; CD90+ CD44+ Stroma: negative for viability dye CD45- CD44+ CD90+ ; Tumor: negative for viability dye CD45- CD44- CD90- EPCAM+/- . 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 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. After sorting, cells were pelleted and resuspended at 1x10³ cells/ml 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).

Imaging Staining and Acquisition

H&E

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

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.

Imaging Slides

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 DinaBead 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 MiSeq/MiniSeq to check if more than 10 percent of the reads aligned to coding regions and contained more than 1,000 unique reads in total. The libraries were then re-pooled based on the percentage of reads in a coding region. 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 (Li, 2013) 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).

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 were subjected to a non-linear dimensionality reduction using tSNE 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.

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. Heterotypic doublets were removed from the major analysis (see platelet-first analysis).

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.

Assembling Cohorts

Immunoprofiler (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. 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 indications (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., 2019). 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, in order to parallel the IPI cohort sample types as accurately as possible. This reduced the patient sample set to 4341 tumor samples.

QUANTIFICATION & STATISTICAL ANALYSIS

Feature Gene Signatures

Tcell, Myeloid, CD90+ CD44+ Stroma and Treg

Gene signatures for each of the features (Supplementary Table4) 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 Treg feature gene signature follows 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. The process of feature gene signature selection for each of the four 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.

CD4 and CD8

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.

Macrophages, Monocytes, cDC1 and cCD2

Gene signatures were generated from single cell RNAseq. Differential gene expression (DGE) between each clusters identified were performed on log-normalized gene counts using the Poisson test provided by the FindMarkers/FindAllMarkers functions in Seurat. Genes with > 0.4 log-fold changes, an adjusted p value of 0.05 (based on Bonferroni correction), at least 35% expressed in tested groups, were regarded as significantly differentially expressed genes (DEGs). Cluster marker genes were identified by applying the DE tests for upregulated genes between cells in one cluster to all other clusters in the dataset. Genes were curated based on log-fold changes and a minimum expression by the other cluster/cell types. The different subsets of mononuclear phagocytic cells were identified by comparing cluster marker genes with public sources referenced in the text. (Supplementary Table 4)

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 Chemokine gene signature was a selection of 39 chemokines of interest. 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.

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 et al., 2019; Wiley et al., 2019). 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) and the Plasma cells gene signature was from (Chen et al., 2020; van Galen et al., 2019). 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 (B et al., 2018; Kumar et al., 2017; Savas et al., 2018; Zhou et al., 2009). A bubble plot was made to visualize the median TPM expression per archetype in the tcell compartment. The bubble size and color represented values of the median TPM in each archetype, transformed to a z-score. A hierarchical clustered heatmap was made to visualize the Th1, Th2, Th17, Trm

and Tex gene signature genes in the samples of the Immune Rich CD8 Bias archetype. The TPM of the gene signature genes for the archetype samples were clustered using a Euclidian distance metric and the Ward linkage method. The clustered heatmap was visualized using Seaborn with rug for the indications in the archetype samples

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 Signature

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 $\log_{10}(\text{TPM}+0.001)$ 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 Signature

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 $\log_{10}(\text{TPM}+0.001)$ 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.

Scores

Gene Signature Score

The feature gene signature scores were calculated using an $m \times n$ matrix where m represented the TMM (Robinson and Oshlack, 2010) normalized logCPM ($\log_2$ counts per million) expression of the feature signature genes and n represented the selected sample set.

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: $E = m \times n$ matrix of gene expression

$P = m \times n$ 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 indication 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	$5.00e-02 < p \leq 1.00e+00$
*	$1.00e-02 < p \leq 5.00e-02$
**	$1.00e-03 < p \leq 1.00e-02$
***	$1.00e-04 < p \leq 1.00e-03$
****	$p \leq 1.00e-04$

For the IPI cohort the gene signature scores for the Tcell, Myeloid, CD90+ CD44+ Stroma, Treg, Chemokine features were calculated using the live compartment. The CD4, CD8 and T cell Exhaustion were calculated using the tcell compartment. The Macrophage, Classical Monocyte, CDC1, cDC2 were calculated using the myeloid compartment. The MHC Class I was calculated using the tumor compartment.

For the TCGA cohort the gene signature scores for the Tcell, Myeloid and CD90+ CD44+ Stroma features were calculated for all indications. 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, indications were used for further analysis as their, median scores per indication appeared to correlate well with the IPI patient samples. The rationale behind including only indications that correlated well with IPI was to ensure that further analysis in TCGA would have extrapolated relevance in the IPI patient samples. The excluded indications' 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 indication, 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) indications as there were not many representative samples for these indications 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.

Flow Score

The flow scores for the flow cytometry population ratios were calculated by converting the ratios to percentiles and then further transforming the percentile to percentiles to increase the spread in the scores. Boxplots were generated to visualize the scores by cancer indication using Seaborn.

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 and 189 overlapping patient samples with the myeloid compartment sequenced. As some feature gene signature scores, noted above, were evaluated in the tcell and myeloid 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 and of the 71 missing myeloid compartment samples 15 did not have flow data.

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. Survival analysis was performed per indication and across all indications using lifelines (Davidson-Pilon et al., 2020), 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 indication between IPI and TCGA.

Feature Clustering

The feature gene signature scores in the IPI cohort were clustered using SCANPY (Wolf et al., 2018) a Python-based toolkit. The initial clustering was done using three features (IPI-3f), Tcell, Myeloid and CD90+ CD44+ Stroma; subsequent clustering was done with additional features: six features (IPI-6f), which was IPI-3f + Treg, CD4 and CD8 and ten features (IPI-10f), which was IPI-6f + Macrophage, Classical Monocytes,

cDC1 and cDC2. 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 = [0.3,0.4,0.5...,2] (resolution)

In the TCGA cohort only the 3f gene feature signature score clustering (TCGA-3f) was done as above with the exception of 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 for the IPI-3f, IPI-6f, IPI-10f and TCGA-3f 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.

The IPI-3f and TCGA-3f final clustering selection had six clusters and hence six archetypes. Survival analysis was performed on TCGA-3f per indication and across all indications using lifelines. Pie charts were made to compare the archetype abundance per indication in IPI-3f and TCGA-3f. Additionally, 3D plots of IPI-3f score and TCGA-3f score visualized the distribution of the unclustered scores.

The IPI-6f and IPI-10f final clustering selection had eight and twelve clusters respectively resulting in eight archetypes for IPI-6f and twelve archetypes for IPI-10f. All the final clustering selections (IPI-3f, IPI-6f, IPI-10f, TCGA-3f) were visualized as UMAPs (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. For all the UMAPs we used a value of 0.25 for min-dist. Population flow proportions and phenotype gene signatures were overlayed on the UMAP clustering. Alluvial plots were made using RAWgraphs (Mauri et al., 2017) to visualize the stability of clusters as the clustering progressed from IPI-3f to IPI-6f to IPI-10f.

REFERENCES

- Arce Vargas, F., Furness, A.J.S., Litchfield, K., Joshi, K., Rosenthal, R., Ghorani, E., Solomon, I., Lesko, M.H., Ruef, N., Roddie, C., et al. (2018). Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. *Cancer Cell* 33, 649-663.e4.
- 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).
- Barry, K.C., Hsu, J., Broz, M.L., Cueto, F.J., Binnewies, M., Combes, A.J., Nelson, A.E., Loo, K., Kumar, R., Rosenblum, M.D., et al. (2018). A natural killer–dendritic cell axis defines checkpoint therapy–responsive tumor microenvironments. *Nat. Med.* 24, 1178–1191.
- Biswas, S.K., Allavena, P., and Mantovani, A. (2013). Tumor-associated macrophages: functional diversity, clinical significance, and open questions. *Semin. Immunopathol.* 35, 585–600.
- Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks.
- Cassetta, L., Fragkogianni, S., Sims, A.H., Swierczak, A., Forrester, L.M., Zhang, H., Soong, D.Y.H., Cotechini, T., Anur, P., Lin, E.Y., et al. (2019). Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-Specific Reprogramming, Biomarkers, and Therapeutic Targets. *Cancer Cell* 35, 588-602.e10.
- Chen, 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-immunoglobulin analysis reveals the diversity of B cells in non-small cell lung cancer. *Genome Biol.* 21, 152.
- Combes, A.J., Courau, T., Kuhn, N.F., Hu, K.H., Ray, A., Chen, W.S., Chew, N.W., Cleary, S.J., Kushnoor, D., Reeder, G.C., et al. (2021). Global absence and targeting of protective immune states in severe COVID-19. *Nature*.
- Davidson-Pilon, C., Kalderstam, J., Jacobson, N., sean-reed, Kuhn, B., Zivich, P., Williamson, M., AbdealiJK, Datta, D., Fiore-Gartland, A., et al. (2020). CamDavidsonPilon/lifelines: v0.25.6.
- Davies, D.L., and Bouldin, D.W. (1979). A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-1, 224–227.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dominguez, D., Tsai, Y.-H., Gomez, N., Jha, D.K., Davis, I., and Wang, Z. (2016). A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. *Cell Res.* 26, 946–962.
- Ephrem, A., Epstein, A.L., Stephens, G.L., Thornton, A.M., Glass, D., and Shevach, E.M. (2013). Modulation of Treg cells/T effector function by GITR signaling is context-dependent. *Eur. J. Immunol.* 43, 2421–2429.

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

Goldman, M., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2019). The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *BioRxiv* 326470.

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.

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95.

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

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

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

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio*.

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

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

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

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

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

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf.* 56–61.

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

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

R Development Core Team (2010). a language and environment for statistical computing: reference index (Vienna: R Foundation for Statistical Computing).

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

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

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

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

Rossum, G. (1995). Python reference manual (NLD: CWI (Centre for Mathematics and Computer Science)).

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

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

Virtanen, P., Gommers, R., Oliphant, T.E., 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.

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

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY : large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 1–5.

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

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

Home | HUGO Gene Nomenclature Committee.