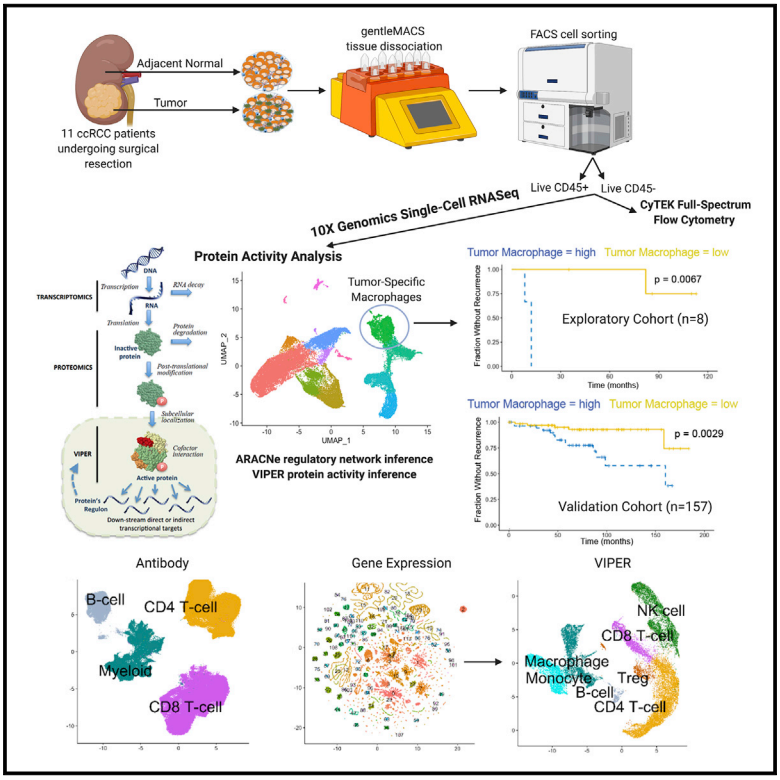
Cell

Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages

Graphical abstract



Highlights

- We present a robust pipeline for scRNA-seq analysis by protein activity inference
- Our pipeline identifies a C1Q^{+}TREM2^{+}APOE^{+} macrophage population enriched in ccRCC
- Protein activity signature of these cells is prognostic of post-surgical recurrence
- C1Q^{+} Macrophage association with recurrence is validated by immunohistochemistry

Authors

Aleksandar Obradovic, Nivedita Chowdhury, Scott M. Haake, ..., James McKiernan, Andrea Califano, Charles G. Drake

Correspondence

ac2248@cumc.columbia.edu (A.C.), cgd2139@columbia.edu (C.G.D.)

In brief

Analysis of the tumor microenvironment using tumor and tumor-adjacent tissue of treatment-naive clear cell renal carcinoma resections from patients by combining single-cell sequencing and single-cell protein activity uncovers a tumor-specific infiltrating macrophage subpopulation associated with disease recurrence.

Obradovic et al., 2021, Cell *184*, 1–18
May 27, 2021
https://doi.org/10.1016/j.cell.2021.04.038

Cell

CellPress

Article

Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages

Aleksandar Obradovic,[1,3,12] Nivedita Chowdhury,[1,4,12] Scott M. Haake,[11,12] Casey Ager,[1] Vinson Wang,[2] Lukas Vlahos,[3] Xinzheng V. Guo,[1] David H. Aggen,[1,13] W. Kimryn Rathmell,[10] Eric Jonasch,[10] Joyce E. Johnson,[10] Marc Roth,[10] Kathryn E. Beckermann,[10] Brian I. Rini,[10] James McKiernan,[2,5] Andrea Califano,[3,4,5,6,7,8,9,*] and Charles G. Drake[1,2,5,14,*]

[1]Columbia Center for Translational Immunology (CCTI), Columbia University Irving Medical Center (CUMC), New York, NY 10032, USA
[2]Department of Urology, Herbert Irving Comprehensive Cancer Center (HICC), New York, NY 10032, USA
[3]Department of Systems Biology, HICC, New York, NY 10032, USA
[4]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[5]HICC, Columbia University, New York, NY, USA
[6]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY
[7]Department of Biomedical Informatics, Columbia University, New York, NY, USA
[8]Department of Medicine, Columbia University College of Physicians and Surgeons, New York, NY, USA
[9]J.P. Sulzberger Columbia Genome Center, New York, NY, USA
[10]Vanderbilt University Medical Center, Nashville, TN, USA
[11]MD Anderson Cancer Center, Houston, TX, USA
[12]These authors contributed equally
[13]Present address: Division of Solid Tumor Oncology, MSKCC, New York, NY 10065, USA
[14]Lead contact
*Correspondence: ac2248@cumc.columbia.edu (A.C.), cgd2139@columbia.edu (C.G.D.)
https://doi.org/10.1016/j.cell.2021.04.038

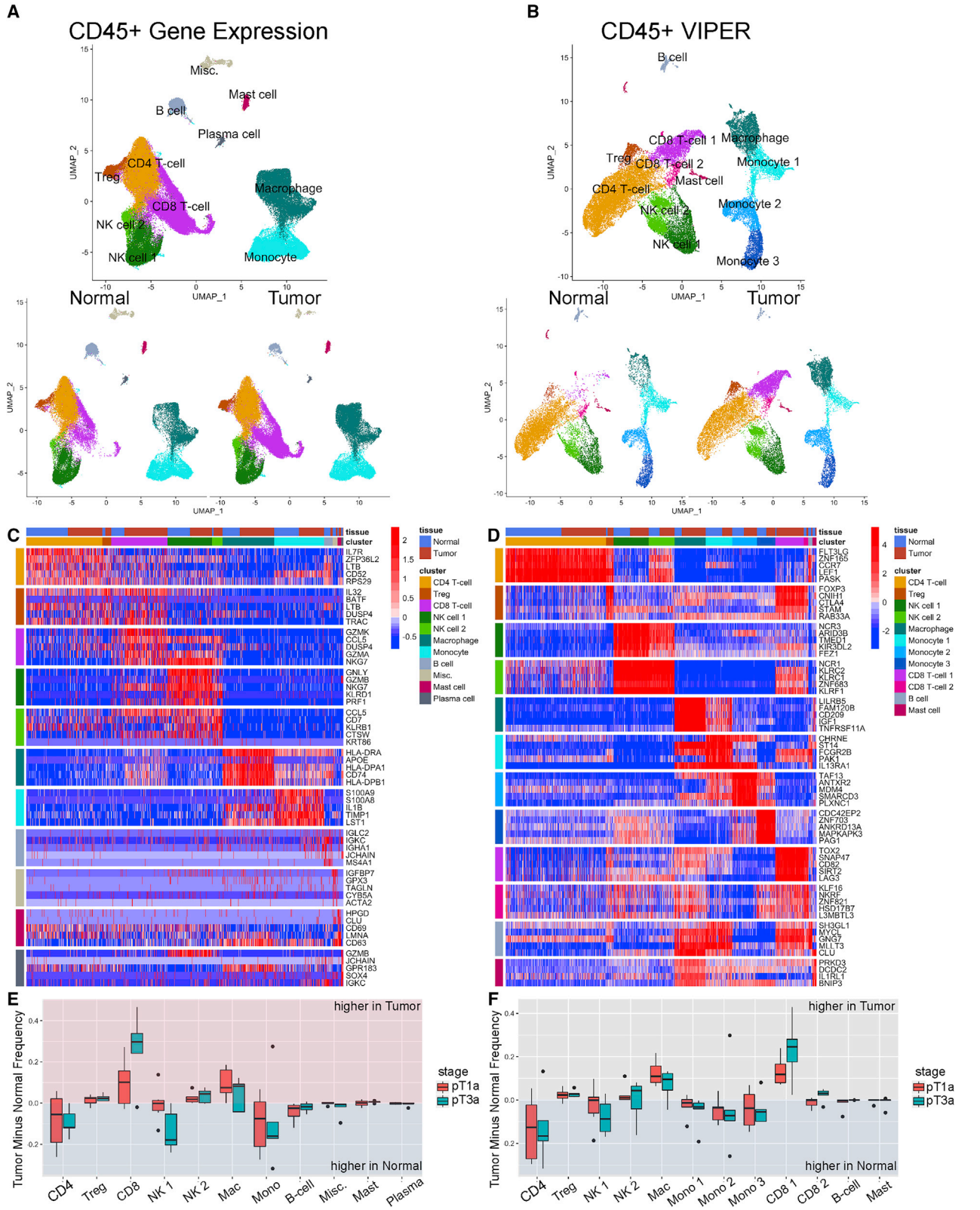
SUMMARY

Clear cell renal carcinoma (ccRCC) is a heterogeneous disease with a variable post-surgical course. To assemble a comprehensive ccRCC tumor microenvironment (TME) atlas, we performed single-cell RNA sequencing (scRNA-seq) of hematopoietic and non-hematopoietic subpopulations from tumor and tumor-adjacent tissue of treatment-naive ccRCC resections. We leveraged the VIPER algorithm to quantitate single-cell protein activity and validated this approach by comparison to flow cytometry. The analysis identified key TME subpopulations, as well as their master regulators and candidate cell-cell interactions, revealing clinically relevant populations, undetectable by gene-expression analysis. Specifically, we uncovered a tumor-specific macrophage subpopulation characterized by upregulation of TREM2/APOE/C1Q, validated by spatially resolved, quantitative multispectral immunofluorescence. In a large clinical validation cohort, these markers were significantly enriched in tumors from patients who recurred following surgery. The study thus identifies TREM2/APOE/C1Q-positive macrophage infiltration as a potential prognostic biomarker for ccRCC recurrence, as well as a candidate therapeutic target.

INTRODUCTION

Clear cell renal carcinoma (ccRCC) is the most common histological subtype of renal carcinoma. Although primary disease is treated surgically, approximately 40% of resected ccRCC patients will relapse and develop metastases (Koul et al., 2011). With a 5-year survival of 10% (Sánchez-Gastaldo et al., 2017), metastatic ccRCC is a lethal disease, underscoring the need to understand the cellular and molecular mechanisms in primary lesions that are prognostic for recurrence, both as biomarkers and as potential targets for intervention. Although ccRCC is an immunogenic tumor, the tumor-immune cell dynamics that regulate effective anti-tumor responses remain incompletely characterized. Consistent with other immunogenic tumors, overall immune infiltration and tumor mutation burden are partially predictive of response to therapy; yet, the value of these biomarkers in clinical decision making remains elusive. Indeed, the complete picture of anti-tumor immune response drivers is complex (Davoli et al., 2017; Şenbabaoğlu et al., 2016; Turajlic et al., 2017). Predictors of post-surgical disease recurrence are also limited, with previous gene-expression studies suggesting CD44 as a marker of recurrence (Li et al., 2015).

To date, the most comprehensive studies of the primary ccRCC tumor micro-environment (TME) used cytometry by time of flight (CyTOF) to interrogate markers of innate and adaptive immunity (Chevrier et al., 2017). These studies showed that expression of T cell exhaustion markers and $CD38^{+}$ myeloid cell infiltration was associated with worse overall outcome. High-throughput droplet-based single-cell RNA sequencing (scRNA-seq) has recently emerged as a valuable tool to catalog the diverse cellular subpopulations that comprise the TME, with the ability to identify representative gene-expression signatures



(legend on next page)

from thousands of individual cells in a single sample (Zheng et al., 2017; Finak et al., 2015). In contrast to bulk RNA-seq, scRNA-seq can characterize the transcriptional state of individual cell types, highlighting the role of rare populations whose gene-expression signature would be diluted below the limits of detection in bulk samples (Stuart et al., 2019). In contrast to flow cytometry or CyTOF, scRNA-seq generates a genome-wide profile of each individual cell's transcriptome, without requiring selection of predefined markers. The value of scRNA-seq has been demonstrated in recent studies of melanoma (Sade-Feldman et al., 2018; Jerby-Arnon et al., 2018) and breast cancer (Chung et al., 2017). However, no systematic, single-cell studies have been performed to study tumor tissue versus adjacent normal in ccRCC.

A key technical limitation of scRNA-seq is that gene-expression profiles are extremely sparse, with ~80%–90% of genes undetected in every cell, a phenomenon known as "gene dropout." While such data are effective in characterizing more molecularly distinct cellular subpopulations, they are not well suited to study specific genes and may also fail to detect more subtle differences, for instance, due to activation of a few critical lineage markers (Elyada et al., 2019). Although dimensionality reduction tools, such as the Seurat pipeline (Butler et al., 2018), are successful in identifying individual subpopulations, the sparse and noisy nature of the data often prevents elucidation of finer-grain biological mechanisms.

To address this issue, we have developed the metaVIPER algorithm (Ding et al., 2018), which leverages highly multiplexed, tissue-specific gene-reporter assays to accurately measure the activity of up to ~6,500 regulatory proteins on a single-cell basis, including transcription factors (TFs), co-factors (co-TFs), signaling proteins (SPs), and surface markers (SMs), based on the expression of their downstream regulatory targets (*regulon*). MetaVIPER extends the VIPER algorithm (Alvarez et al., 2016) to single cells, independent of lineage. For simplicity, here we will use the term VIPER to refer to its single-cell implementation. Single-cell, tissue-specific regulons are inferred using ARACNe, an information theoretic algorithm that has been experimentally validated in multiple tissue contexts, with a >70% accuracy in target identification (Basso et al., 2005).

To comprehensively characterize the interaction of immune- and non-immune cells in the ccRCC TME, we generated scRNA-seq data from fluorescence-activated cell sorting (FACS)-purified hematopoietic and non-hematopoietic cells (Figure S1) dissociated from tumor and adjacent non-tumor tissue of 11 treatment-naive primary ccRCC patients. To analyze these data, we developed a VIPER-based scRNA-seq analysis pipeline to assess single-cell protein activity from single-cell ARACNe networks followed by an optimized single-cell clustering approach. These studies revealed a population of tumor-specific C1Q$^+$TREM2$^+$APOE$^+$ macrophages associated with early post-surgical disease recurrence, as well as a potential target for therapeutic intervention. To validate VIPER predictions, we generated spectral flow cytometry and scRNA-seq from matched patient samples, as well as quantitative, multispectral immunofluorescence (qmIF) data for a set of proteins significantly activated in a macrophage subpopulation prognostic for post-surgical disease recurrence. Taken together, these data provide a comprehensive atlas of primary ccRCC TME subpopulations—including the master regulator (MR) proteins that control their transcriptional state, lineage markers, and predicted cell-cell interactions.

RESULTS

Protein activity analysis of CD45$^+$ TME cells reveals tumor-specific immune subpopulations

To study hematopoietic and non-hematopoietic populations in the primary ccRCC TME at single-cell resolution, we isolated live cells from 11 treatment-naive resected tumors, along with adjacent normal tissue. Expression-based clustering of scRNA-seq profiles revealed populations broadly consistent across patients (Figure S2). We initially focused on the hematopoietic compartment (CD45$^+$), which was visualized following UMAP dimensionality reduction and clustered using the Seurat Louvain algorithm (Stuart et al., 2019). To optimize often arbitrary clustering while retaining scalability to hundreds of thousands of cells, we performed Louvain clustering across a range of 100 resolution values and selected optimal clustering resolution by optimizing a bootstrapped mean silhouette score (see STAR Methods). This clustering approach resolved CD4 and CD8 lymphocytes, regulatory T cells (Tregs), natural killer (NK) cells (two populations), macrophages, monocytes, and small populations of B cells, mast cells, and plasma cells (Figure 1A), which were represented in all patient samples. For visualization purposes, we show a heatmap for the top five transcripts most uniquely upregulated in each cluster (Figure 1C). These data confirmed SingleR-inferred cellular identify of each cluster (Figure S3)—including expression of IL7R in CD4 T cells, CD3 and granzyme in CD8 T cells, and S100A8/S100A9 in monocytes.

However, expression-based clustering missed multiple established markers of these populations. For example, Tregs did not show differential expression of the canonical FOXP3 transcription factor; rather, the most overexpressed gene was

Figure 1. Deep profiling of CD45$^+$ microenvironment by gene expression and protein activity reveals tumor-specific immune populations

(A) UMAP plots for single-cell gene expression pooled across CD45$^+$ samples, clusters visualized and labeled by cell type. Bottom plot is split by tumor versus adjacent normal label.

(B) UMAP plots for VIPER-Inferred protein activity pooled across CD45$^+$ samples. Bottom plot is split by tumor versus adjacent normal label.

(C) Heatmap of top5 upregulated genes for each cluster by expression; each row represents a gene and each column represents a cell. Legend shows cluster identity with cell type inferred by SingleR and tumor (red) or adjacent normal (blue) tissue source.

(D) Heatmap of top5 differentially upregulated proteins for each cluster by VIPER-inferred activity. Legend is as in (C).

(E) Bar plots of patient-by-patient cluster frequency in tumor minus frequency in adjacent normal for each gene-expression cluster, grouped by stage; values <0 (blue) indicate higher frequency in adjacent normal and values >0 (red) indicate higher frequency in tumor.

(F) Bar plots of patient-by-patient cluster frequency in tumor minus frequency in adjacent normal for each VIPER cluster, grouped by stage, as in (E).

See also Figures S2 and S3.

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

interleukin-32 (IL-32). While other genes in this cluster, e.g., BATF, TIGIT, and TNFRSF18, are expressed in Tregs (Chao and Savage, 2018), none is considered a canonical marker. Further, expression-based clustering failed to recapitulate the heterogeneity of these subpopulations; for example, it distinguished only two populations of myeloid cells (Figure 1). Finally, considering intra-cluster statistics, differentially expressed genes had poor reproducibility. For instance, average within-cluster standard deviation of classic markers such as IL7R, KLRD1, and CD8B was quite high, $\sigma_{IL7R} = 3.19$, $\sigma_{KLRD1} = 3.79$, and $\sigma_{CD8B} = 3.01$, respectively, exceeding the mean expression values of the gene: $\mu_{IL7R} = 2.52$, $\mu_{KLRD1} = 3.17$, and $\mu_{CD8B} = 1.41$.

We next proceeded to assess whether protein activity-based clustering, using VIPER, would yield additional robustness and biological insight. Since we previously showed that regulatory networks of lineage-related cells have >95% overlap (Mani et al., 2008), we generated (n = *69*) ARACNe-inferred regulatory models—one for each gene-expression cluster in each patient—as we expected that finer differences would likely be present within primary lineages. For each single cell, we then used the cluster-specific regulatory networks to perform VIPER-based protein activity inference. While the resulting clusters were generally consistent with those derived by gene expression, protein activity-based clusters showed multiple critical differences (Figures 1B and 1D). For example, the most differentially active protein in the Treg cluster was FOXP3, consistent with well-established Treg biology (Chao and Savage, 2018) (Figure 1D), and Cytotoxic T Lymphocyte Antigen-4 (CTLA-4), which is upregulated on the surface of tumor-infiltrating Tregs (Arce Vargas et al., 2017), was inferred as differentially active by VIPER. VIPER identified a distinct CD8 T cell population with markers consistent with exhaustion, including differential activation of LAG-3, TOX2, and PD1, which had been missed by expression-based analysis (Figures 1D and 2A). Additionally, myeloid cells were further stratified by VIPER into macrophages and three distinct monocyte subpopulations.

VIPER analysis identified several populations that were differentially represented as a function of tumor stage and localization in tumor versus adjacent non-tumor; these were undetectable by gene-expression analysis (Figures 1E and 1F). Specifically, Treg, CD8 T cell, and macrophage normalized counts were higher in the tumor compared to normal adjacent tissue ($p = 0.012$, $p = 0.006$, $p = 0.013$, respectively). In contrast, monocyte, B cell, and CD4 T cell counts were higher in adjacent normal ($p = 0.097$, $p = 0.017$, $p = 0.018$, respectively). Two NK cell clusters were identified, one with higher counts in the adjacent normal (NK cell 1) and a second with higher counts in tumor (NK cell 2) ($p = 0.09$, $p = 0.008$, respectively). Consistent with prior data (Becht et al., 2015), activity-based but not expression-based clustering identified higher counts of exhausted CD8 T cells in tumor versus adjacent normal ($p = 0.0005$), and also in stage pT3a versus pT1a tumors ($p = 0.015$) (Figure 1F). Further, the tumor-specific macrophage population identified by inferred protein activity was more significantly enriched in tumor as compared to adjacent normal than the coarse macrophage population identified by gene expression, i.e., $p = 0.0006$ versus $p = 0.013$.

Reproducibility of individual markers was also significantly improved by VIPER (Figure 1D), compared to gene expression (Figure 1C). For example, the standard deviation of the classic markers (IL7R, KLRD1, and CD8B) was much lower, $\sigma_{IL7R} = 0.75$, $\sigma_{KLRD1} = 0.48$, and $\sigma_{CD8B} = 0.49$, respectively, whereas their mean activity value was substantially larger $\mu_{IL7R} = 5.28$, $\mu_{KLRD1} = 6.54$, and $\mu_{CD8B} = 5.88$, as further confirmed by significant improvement in silhouette scores, SC = 0.7 by VIPER as compared to a SC = 0.35 by gene-expression-based clustering (Figure S4). Thus, based on the coefficient of variation (i.e., σ/μ), reproducibility was increased between 8.9-fold (IL7R) and 25.6-fold (CD8B) by VIPER-based clustering.

Relative specificity of tumor-infiltrating immune cell populations

To visualize protein activity differences in key cell-type markers over-represented in tumor versus adjacent normal, including macrophages, Tregs, and exhausted CD8 T cells, we generated violin plots (Figure 2A). As shown, LILRB5 was identified as the most differentially active among VIPER-inferred markers of tumor-specific macrophages. These macrophages, as well as the three monocyte clusters, showed high APOE activity. Among T cells, FOXP3 was identified as the most activated protein in TIL Treg; relative activation was also noted in tumor-associated CD8 T cells. CTLA-4 followed a similar pattern, consistent with previous data from bulk TIL Treg studies (Arce Vargas et al., 2017). LAG-3 and PD-1 (PDCD1) showed similar inferred protein activity distributions, with higher activity in CD8 T cells and, in particular, in the CD8 TIL cluster (CD8 T cell 1) versus the (CD8 T cell 2) cluster that was also represented in adjacent normal. PD-1 also showed significant expression in TIL Treg. Interestingly, we found significant activation of TOX2—a protein related to TOX1, which plays a critical role in epigenetic reprogramming of exhausted CD8 T cells (Khan et al., 2019; Scott et al., 2019)—in the CD8 TIL cluster, consistent with an exhausted phenotype. We complemented these differential activity analyses by examining genes that were differentially expressed in the tumor-specific macrophage population, which were only be identified by VIPER-based protein activity clustering (Figure 1C). Once a finer-grain cluster structure was revealed, several genes were found to be overexpressed in specific clusters, including APOE, C1QA-C, and TREM2, demonstrating the ability to integrate both differential protein activity and differential gene expression in the analysis (Figure 2B). Notably, differential expression of these genes would have gone undetected if the cluster structure produced by gene-expression-based clustering had been used. As discussed below, we subsequently validated the tumor-specific macrophage marker proteins identified by these analyses by immunofluorescence and correlated those data with clinical outcome.

Master regulators of subpopulation transcriptional state

In addition to recapitulating differential activity of established subpopulation markers, VIPER analysis identified novel proteins that were differentially active in both common and rare subpopulations. For TFs and co-TFs, these master regulator (MR) proteins represent novel mechanistic drivers of the transcriptional state of these cells, while signaling and surface-marker proteins may represent novel lineage markers for FACS-based isolation, as further confirmed by the protein-based validation assays

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

CellPress

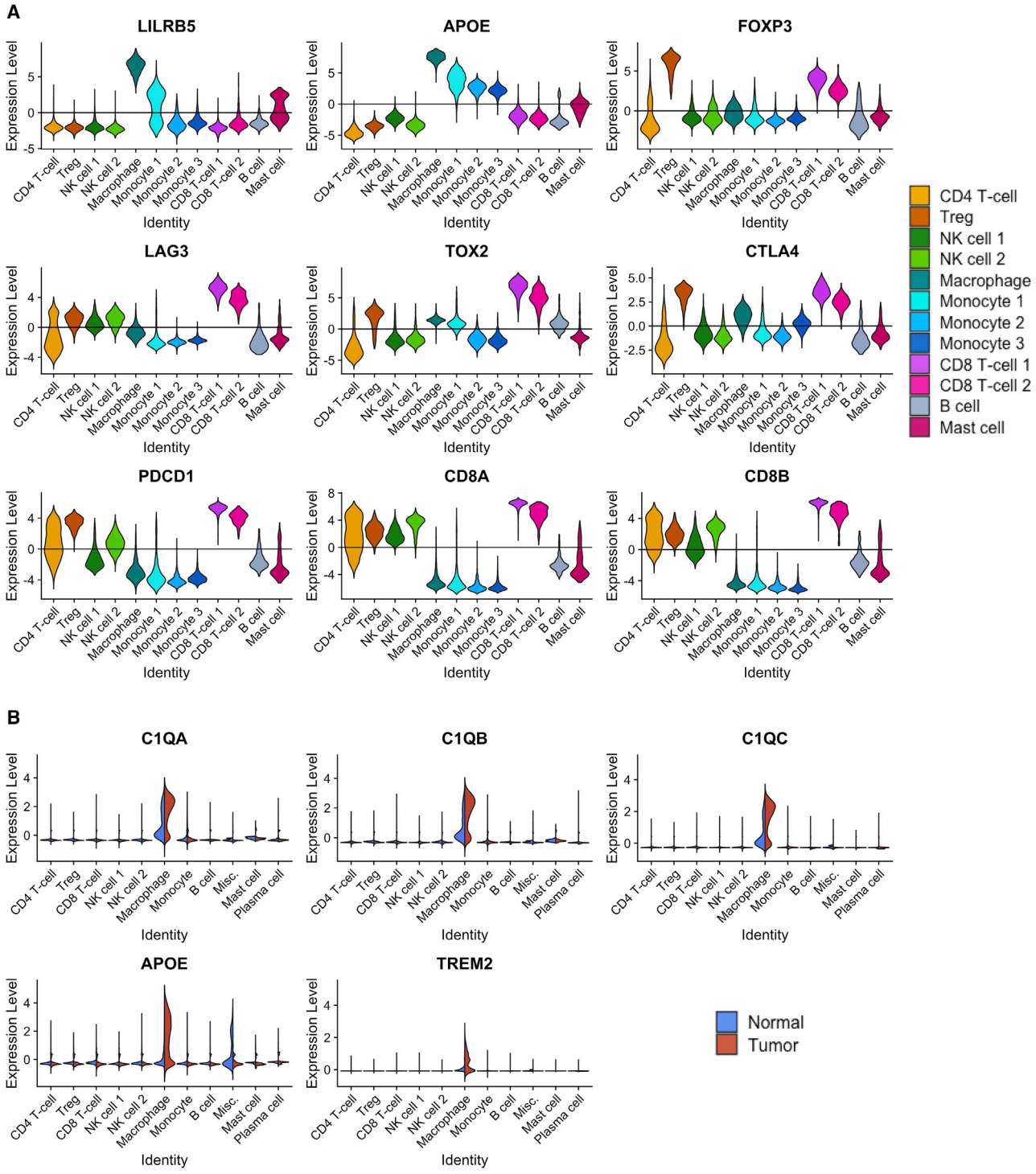
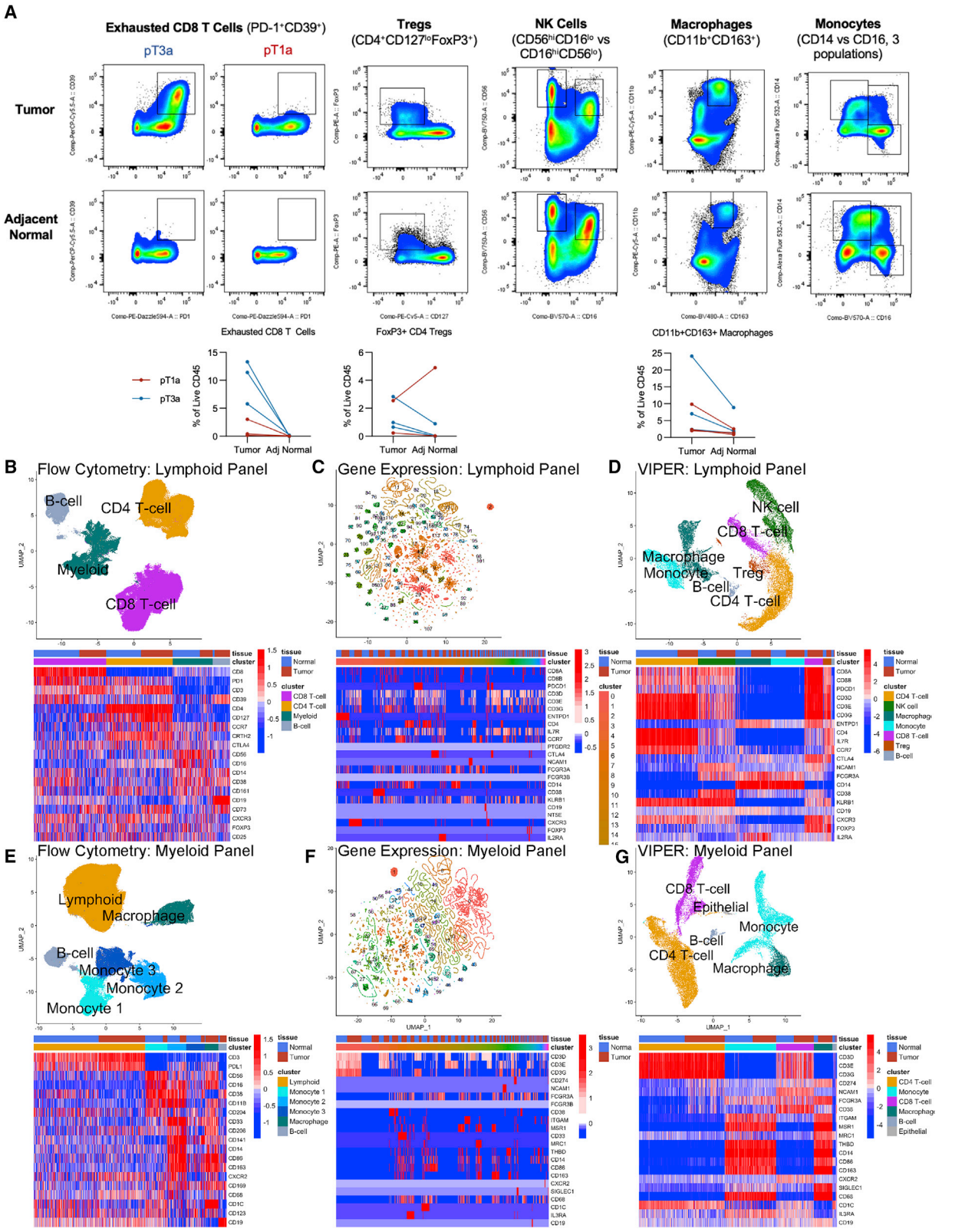


Figure 2. Known and novel tumor-infiltrating immune population markers discovered from single-cell transcriptomic and inferred proteomic data

(A) Violin plots of VIPER-inferred proteins upregulated in CD45^{+} cell subsets corresponding to Tregs (FOXP3, CTLA4), exhausted CD8s (TOX2, LAG3, PD1, CTLA4), and tumor-specific macrophages (LILRB5, APOE).

(B) Violin plots of top transcriptional markers (C1Q, APOE, TREM2) specifically upregulated in tumor-infiltrating macrophages as compared to other cell populations as well as non-tumor macrophages.

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038



(legend on next page)

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

discussed below. Of note, we previously showed that a majority (>70%) of VIPER-inferred proteins control the transcriptional state of a cell by direct, physical regulation of their transcriptional targets and can be used to efficiently reprogram cell state by ectopic expression/co-expression (Carro et al., 2010; Dutta et al., 2016; Talos et al., 2017; Arumugam et al., 2020). As a result, these proteins may also represent attractive drug targets for modulation of specific subpopulations. Candidate MR proteins of each VIPER-inferred cluster are reported in Table S2; the top MRs are also shown in Figure 1D. In particular, MRs of tumor-enriched cell populations (i.e., Tregs, CD8 T 1 cells, and macrophages) included both established drug targets for Tregs or exhausted CD8 T cells, such as CTLA-4 and PD-1, as well as less well-characterized markers of Tregs (CNIH1, STAM, RAB33A, etc.), exhausted CD8 T cells (TOX2, SNAP47, CD82, SIRT2, LAG-3, etc.), and tumor-infiltrating macrophages (LILRB5, FAM120B, CD209, IGF1, TNFRSF11A, etc.). As a result, these data provide a valuable resource of proteomic regulators for the full complement of cell phenotypes in the ccRCC TME.

VIPER-inferred protein activity recapitulates flow cytometry, overcoming scRNA-seq-related gene dropout

To more fully characterize the proteomic profile of the ccRCC TME and to benchmark VIPER results, we analyzed a subset of samples for which scRNA-seq data were available using high-dimensional flow cytometry with a 19-marker lymphoid panel and 19-marker myeloid panel. Manual gating of specific marker pairs broadly recapitulated the populations identified by VIPER-based cluster analysis. For example, flow cytometry identified a population of $CD8^+/PD1^+/CD39^+$ CD8 T cells, with numerically higher normalized counts in tumor versus adjacent non-tumor samples ($p = 0.057$), consistent with the cluster of tumor-enriched exhausted CD8 T cells identified by VIPER (Figure 3A). Cytometry also identified a population of $CD4^+/CD127^{low}/FOXP3^+$ Tregs, with higher representation in the majority of tumor versus adjacent normal samples ($p = 0.072$). Similarly, these flow studies confirmed the existence of two distinct NK cell subpopulations (i.e., $CD56^{high}/CD16^{low}$ versus $CD56^{low}/CD16^{high}$), a $CD11C^+/CD163^+$ macrophage population with higher representation in tumor versus adjacent normal ($p = 0.076$), and three distinct monocyte subpopulations ($CD14^+/CD16^+$, versus $CD14^+/CD16^-$, versus $CD14^-/CD16^+$). While these populations were validated by manual gating of specific proteins, our analysis shows that they could not have been inferred directly from the high-dimensional flow cytometry data (Figures 3B–3G). To test this, we performed unsupervised clustering of the flow cytometry dataset, using the Resolution-Optimized Louvain cluster analysis algorithm by which we infer expression and activity-based clusters. Based on the lymphoid panel, the analysis identified four distinct clusters: CD4 and CD8 T cells, myeloid cells, and B cells (Figure 3B) while the myeloid panel yielded 6 clusters: lymphoid cells, B cells, 3 monocyte cell types characterized as $CD14^+/CD16^+$, $CD14^+/CD16^-$, and $CD14^-/CD16^+$, respectively, and a macrophage cluster only represented in the tumor compartment, with relative overexpression of CD86, CD1D, CD16, CD163, CD169, CD56, CXCR2, CD14, and CD33 proteins (Figure 3E).

To assess whether scRNA-seq data could recapitulate these findings, we restricted unsupervised cluster analysis to genes encoding for proteins represented in the flow cytometry panels (Figures 3C and 3F). This analysis failed to reveal biologically relevant clusters due to high gene dropout rates, even on markers that should be highly expressed. Indeed, expression of the genes encoding for the 19 lymphoid and 19 myeloid proteins was too noisy and sparse to support cluster inference consistent with established cell types. This result emphasizes the extremely noisy nature of scRNA-seq measurements when restricted to specific genes of interest.

We next tested whether analysis of VIPER-inferred activity for lymphoid and myeloid markers could recapitulate meaningful cell types. As shown in Figure 3D, the vast majority of proteins in the lymphoid flow cytometry panel were well resolved by VIPER, with the exceptions of PTGDR2, FCGR3B, and NT5E. Similarly, the majority of the proteins in the myeloid flow panel were also well resolved by VIPER, with the exception of FCGR3B and CD33 (Figure 3G). Taken together, 34 of 39 proteins (77%) were well represented by VIPER, consistent with the 70%–80% previously reported recovery in protein activity measurements (Alvarez et al., 2016). This was even more remarkable because the panel analyzed here included mostly surface markers not directly involved in transcriptional regulation, with a few exceptions (e.g., FOXP3).

Protein activity analysis restricted to the lymphoid panel was effective in recovering lymphoid cell diversity and was able to distinguish monocytes from macrophages despite limited profiling of macrophage lineage markers by the lymphoid panel (Figure 3D), thus comparing favorably with analysis of flow cytometry data, which only identified 4 of these 7 cell types. For instance, protein activity analysis was effective in identifying Tregs, which were missed by flow-based clustering due to low intensity of FOXP3 staining. Comparing protein abundance and activity in matched flow-cytometry and VIPER clusters

Figure 3. Flow cytometry is better recapitulated by protein activity than by gene expression

(A) Representative flow cytometry gating in tumor and adjacent normal and frequency plots in tumor and adjacent normal for all manually gated populations. Populations of $PD1^+CD39^+$ exhausted CD8 cells, Tregs, and $CD11B^+CD163^+$ macrophages are of higher frequency in tumor than adjacent normal. Representative plots showing two distinct NK cell phenotypes and three monocytic sub-phenotypes, consistent with Figure 1B.

(B) UMAP projection, clustering, and heatmap by flow cytometry proteins profiled in CyTEK lymphoid panel.

(C) UMAP and clustering by scRNA-seq gene-expression subset to the proteins profiled in (B), showing noise-induced decrease in clustering resolution.

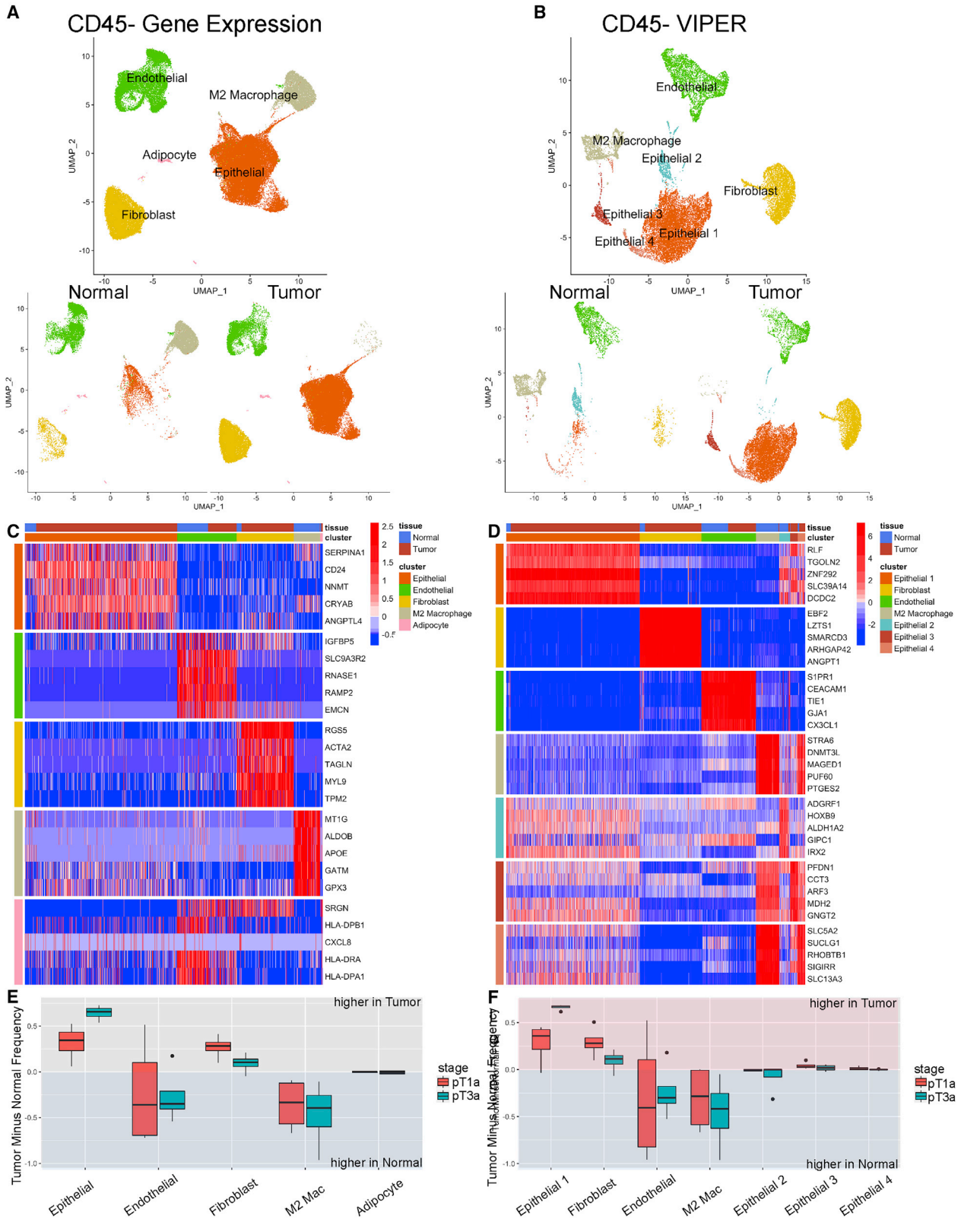
(D) UMAP and clustering by scRNA-seq VIPER inference subset to the proteins profiled in (B).

(E) UMAP and clustering by flow cytometry proteins profiled in CyTEK myeloid panel.

(F) UMAP and clustering by scRNA-seq gene expression, subset to the proteins profiled in (E).

(G) UMAP and clustering by scRNA-seq VIPER inference, subset to the proteins profiled in (E).

See also Figure S5.



(legend on next page)

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

(e.g., CD4 T cells), the reproducibility of activity data was approximately 2-fold higher, on average, based on coefficient of variation (CV) analysis, defined as the ratio of the standard deviation over the mean. When averaged over the top proteins differentially represented in the CD4 and CD8 T cell clusters, flow-based analysis produced CV_{CD4} = 0.206 and CV_{CD8} = 0.209, while activity-based analysis yielded CV_{CD4} = 0.151 and CV_{CD8} = 0.124, reflecting higher noise in antibody-based measurements.

These findings were validated in a public CITE sequencing (CITE-seq) dataset profiling antibody staining and gene expression simultaneously in single cells derived from cord blood (Stoeckius et al., 2017). Here, clustering by antibody profile was lost by single-cell gene expression but was completely recovered by VIPER analysis (Figure S5A). Moreover, the cell-matched coefficient of variation within each cluster was significantly lower for VIPER-inferred protein activity as compared to gene expression (p = 0.0004) and compared favorably with antibody-based measurements (p = 0.0083) (Figure S5B). As above, VIPER-measured protein activity correlated more closely with antibody staining than did gene expression (Figures S5C–S5E). Taken together, these results show that protein activity-based analyses can recapitulate cell-type identification based on protein-level data derived by flow cytometry even from a relatively restricted set of marker proteins, suggesting that such analyses effectively mitigate the gene dropout inherent in scRNA-seq.

Protein activity in CD45$^-$ cells distinguishes tumor cells from adjacent normal epithelium

We next compared expression and activity-based clustering of non-hematopoietic (CD45$^-$) ccRCC TME cells. Expression-based clustering of CD45$^-$ cells from all patients (including tumor and adjacent normal) revealed four predominant cell types: epithelial cells, endothelial cells, fibroblasts, and M2 macrophages (Figure 4A). These populations showed differential representation in tumor versus adjacent normal, with M2 macrophages predominant in adjacent normal (p_{M2} = 0.007), and fibroblasts and epithelial cells over-represented in tumor tissue (p_{FB} = 0.009, p_{Epi} = 0.0005). The epithelial cluster, mostly composed of tumor cells, showed tumor compartment specificity as compared to adjacent normal. As expected, it was more highly represented in patients with pT3a compared to pT1a disease ($p_{Epi|pT3\text{-}pT1}$ = 0.011) (Figure 4E). The 5 most upregulated genes for each cluster are shown in Figure 4C. Of note, epithelial cells overexpressed SERPINA1, a protease inhibitor upregulated in multiple cancer types, as well as CD24, recently described as a macrophage immune checkpoint protein (Barkal et al., 2019), whose expression is associated with worse prognosis in ccRCC (Arik et al., 2017). Additionally, the epithelial cluster showed upregulation of keratin 16 (*KRT16*) and 8 (*KRT8*), which have also been associated with poor prognosis in ccRCC (Tan et al., 2017).

VIPER was equally successful in identifying fibroblast, endothelial, and M2 macrophage clusters, while revealing a deeper level of heterogeneity. Specifically, epithelial cells stratified into four distinct clusters (Figure 4B). While clusters E1, E3, and E4 were more represented in the tumor as compared to adjacent non-tumor (p_{E1} = 0.001, p_{E3} = 0.056, p_{E4} = 0.028), cluster E2 was more represented in adjacent normal (p_{E2} = 0.312) (Figure 4E). This population represents normal epithelial cells, whose gene expression was not sufficiently distinct from tumor cells to be effectively stratified without VIPER. For visualization purposes, we show the 5 most differentially active proteins for each cluster (Figure 4D).

Further analyses showed that cluster E1—the most prevalent among the four epithelial clusters—was significantly over-represented in stage 3 tumors as compared to stage 1 tumors ($p_{E1|S3:S1}$ = 0.018), while lower-frequency populations E2, E3, and E4 were represented in both stage 1 and stage 3 patients. This analysis suggests that differential frequency of a dominant epithelial cell population whose transcriptional state is virtually identical across patients effectively stratifies stage 1 versus stage 3 tumors (Figures 4E and 4F). These data are relevant because transcriptionally distinct tumor cell subpopulations may have differential drug sensitivity and because protein activity-based analysis but not gene expression allowed distinct identification of normal versus tumor-related cells.

Inferred tumor-cell copy-number alterations are characteristic of ccRCC

To further understand the epithelial cell clusters, we assessed VIPER-inferred activity of PAX8, PAX2, and CAIX, proteins expressed in renal epithelium; these are upregulated in malignancy and commonly used as markers for ccRCC (Farber et al., 2017). This analysis confirmed increased activity of these markers in epithelial clusters E1, E3, and E4 (Figure 5A). To more precisely determine which epithelial clusters represent tumor cells, we performed copy-number alteration (CNA) inference clustered by expression-based (Figure 5B) or activity-based (Figure 5C) analysis (Tickle et al., 2019). We inferred CNAs for each CD45 negative cell, using CD45 positive cells as normal ploidy controls. The results (Figure 5C) showed that aberrant CNA regions are present in epithelial clusters E1, E3, and E4 but not E2, including recurrent 3p chromosomal deletions not detected in any other cell type. Of note, chromosome 3p deletions occur in >96% of all ccRCC patients, as that region contains the VHL

Figure 4. Deep profiling of CD45$^-$ cells by gene expression and protein activity distinguishes tumor cells from normal epithelium

(A) UMAP of single-cell gene expression pooled across all CD45$^-$ samples, clusters labeled by cell type. Bottom plot is split by tumor versus adjacent normal label.

(B) UMAP of VIPER-inferred protein activity pooled across all CD45$^-$ samples, clusters labeled by cell type. Bottom plot is split by tumor versus adjacent normal.

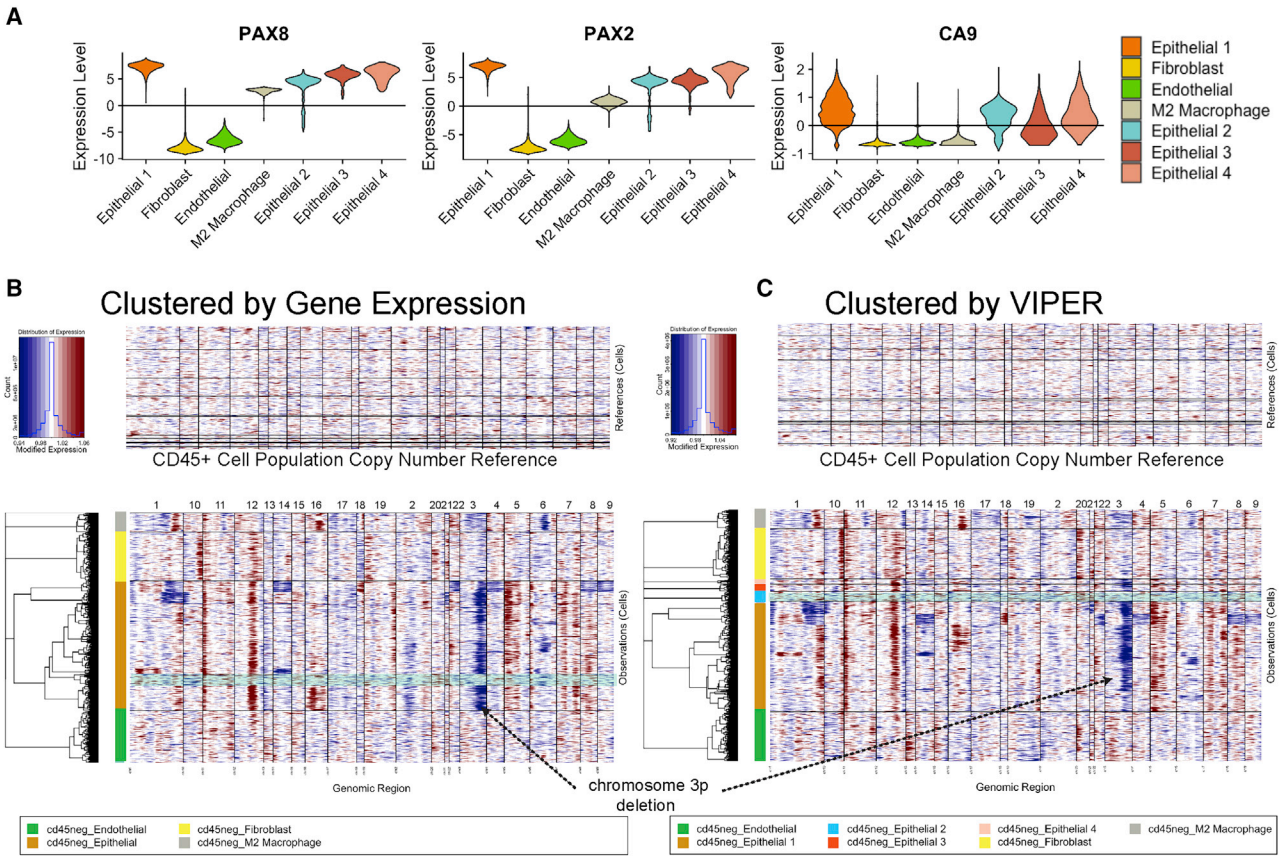
(C) Heatmap of top5 differentially upregulated genes for each cluster by expression; each row represents a gene and each column represents a cell. The legend shows cluster identity with cell type inferred by SingleR and tumor (red) or adjacent normal (blue).

(D) Heatmap of top5 differentially upregulated proteins for each cluster by VIPER-inferred activity. Legend is as in (C).

(E) Bar plots of patient-by-patient cluster frequency in tumor minus frequency in adjacent normal for each gene-expression cluster, grouped by stage; values <0 (blue) indicate higher frequency in adjacent normal, values greater < 0 (red) indicate higher frequency in tumor.

(F) Bar plots of patient-by-patient cluster frequency in tumor minus frequency in adjacent normal for each VIPER cluster, grouped by stage, as in (E).

See also Figures S2 and S3.



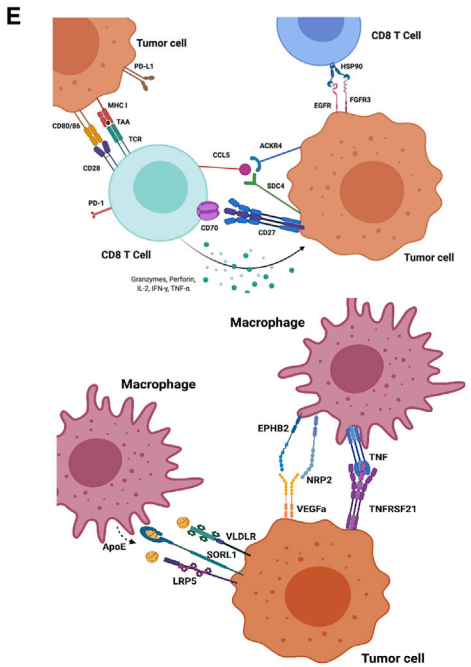
D

Tumor :: T cell Interactions

Ligand	Receptor	Ligand Cluster	Receptor Cluster
CCL5	ACKR4	CD8 T-cell 1	Epithelial 3
CCL5	SDC4	CD8 T-cell 1	Epithelial 1
CD70	CD27	Epithelial 1	CD8 T-cell 1
HSP90AA1	EGFR	CD8 T-cell 2	Epithelial 1
HSP90AA1	FGFR3	CD8 T-cell 2	Epithelial 4

Tumor :: Macrophage Interactions

Ligand	Receptor	Ligand Cluster	Receptor Cluster
APOE	LRP5	Macrophage	Epithelial 4
APOE	SORL1	Macrophage	Epithelial 1
APOE	VLDLR	Macrophage	Epithelial 1
CYR61	ITGB5	Epithelial 1	Macrophage
ICAM1	EGFR	Macrophage	Epithelial 1
MDK	LRP1	Epithelial 1	Macrophage
MDK	SDC3	Epithelial 1	Macrophage
MDK	TSPAN1	Epithelial 1	Macrophage
RARRES2	CMKLR1	Epithelial 1	Macrophage
SERPINA1	LRP1	Epithelial 3	Macrophage
TNF	TNFRSF21	Macrophage	Epithelial 4
VEGFA	EPHB2	Epithelial 1	Macrophage
VEGFA	NRP2	Epithelial 1	Macrophage



(legend on next page)

tumor suppressor locus (Hsieh et al., 2018). The epithelial cluster inferred by expression-based analysis included cells lacking gross copy-number alterations (highlighted in Figure 5B), corresponding exactly to activity-based cluster E2 (Figure 5C).

Dissecting receptor-ligand interactions in ccRCC

A critical challenge that may benefit from VIPER-based protein activity measurements is the elucidation of cross-compartment interactions that may modulate tumor homeostasis. The extensive scRNA-seq dataset generated by these studies, which included data from both hematopoietic and non-hematopoietic cells, supported *in silico* interrogation of putative receptor/ligand interactions between cell types. To that end, we identified overexpressed genes encoding for secreted ligands and differential VIPER activity of their cognate binding receptors between all possible subpopulation pairs and across patients. Referencing a curated public database of 2,557 known receptor-ligand interaction pairs (Lizio et al., 2019), we identified interactions supported by significant overexpression of the ligand in any cell population and concomitant activation of the cognate receptor by VIPER in any patient-matched subpopulation. A total of 276 candidate receptor-ligand pairs were identified (Table S4). Of these, several had been previously established in ccRCC. For example, we identified receptor/ligand pair KDR/VEGFA in tumor cells and endothelial cells, respectively. Figure 5D shows a curated subset of predicted receptor/ligand interactions between tumor cells and tumor-enriched immune populations (T cells and macrophages). Of these, the potential interaction between CD70 and its cognate receptor CD27, in tumor and CD8 T cells, respectively, is of interest, given ongoing investigation of CD70 as a therapeutic target (Jacobs et al., 2015). The majority of interactions thus identified were not previously reported, providing a systematic resource for future studies.

A tumor-specific macrophage signature is associated with disease recurrence

As above, protein activity analysis identified a novel, tumor-specific subpopulation of macrophages and their top differentially active proteins (i.e., LILRB5, APOE, and TREM2) and differentially expressed genes (i.e., C1QA-C, APOE, and TREM2) (Table S2). To assess the clinical significance of this population, we leveraged single-cell ARACNe networks to transform bulk RNA-seq data from two independent cohorts (n = 8 and n = 157) using VIPER. Here, gene expression of each cohort was scaled by the mean and standard deviation of each gene and VIPER was applied. We defined a set of statistically upregulated proteins ($p < 0.05$) in the tumor-specific macrophage population, and computed normalized enrichment score (NES) of this gene set in the ranked differential protein activity signature of patients with post-surgical disease recurrence compared to those without recurrence. This analysis was first performed in a small (n = 8), well clinically annotated cohort of bulk RNA-seq samples from untreated ccRCC surgical resections; here, we found a significant enrichment of tumor macrophage signature in 4 patients with recurrence compared to 4 age- and stage-matched controls (Figure 6A) (normalized enrichment score [NES] = 4.08, $p = 4.5 \times 10^{-5}$). We found that the leading-edge proteins included marker proteins APOE and TREM2, as well as other macrophage-associated proteins of potential clinical interest, such as LILRB5, MERTK, and IGF1 (Figure 6C). Sample-by-sample NES of the tumor macrophage gene set was computed directly on the ranked VIPER activity of proteins in each bulk RNA-seq sample and was consistent with the group-wise analysis in Figure 6A, such that all non-recurrent patients had significant depletion of tumor macrophage markers and recurrent patients had strong enrichment, with the sole exception of a single patient who recurred late (82 months post initial surgery) (Figure 6D). To further explore this clinical association, we performed univariate Cox regression of NES versus time to recurrence (TTR) on a patient-by-patient basis ($p = 0.057$). Binary log-rank test of macrophage enrichment, with NES > 0 = "high" and NES $\leq$ 0 = "low," showed a strong statistically significant association between signature enrichment and shorter time-to-recurrence ($p = 6.7 \times 10^{-3}$) despite a relatively small sample size, suggesting a strong effect (Figure 6B). To validate the association of markers representative of this rare population with recurrence, we next measured their enrichment in a larger cohort of bulk RNA-seq samples from 157 treatment-naive ccRCC surgical resections, annotated with time to post-surgical recurrence. This validation cohort showed a consistent upregulation of tumor macrophage markers in patients with post-surgical recurrence (Figure 6E) (NES = 4.33, $p = 1.5 \times 10^{-5}$), with a significant cox regression p value of 0.012 and binarized log-rank p value of 0.0029 (Figure 6F).

C1Q/TREM2-expressing macrophages are tumor restricted and associated with post-surgical recurrence

We next queried whether markers of the macrophage population associated with poor outcome were co-expressed in cells by Immuno-fluorescence staining and analyzed their spatial localization to determine whether these markers were tumor, T cell, or

Figure 5. Tumor cell labeling is validated by copy-number inference and tumor marker expression

(A) Violin plots of VIPER-inferred activity for ccRCC tumor markers PAX2, PAX8, and CA9. Plots grouped by $CD45^-$ cluster label revealing increased expression in epithelial cells.

(B) CNA inference for all $CD45^-$ populations, using $CD45^+$ cells as reference. Columns represent chromosomal regions and rows represent cells, grouped by gene-expression cluster, with a subset of copy-number-normal epithelial cells highlighted in green.

(C) CNA inference re-grouped by VIPER cluster. Epithelial cell clusters 1, 3, and 4 contain consistent chromosome 3p deletions characteristic of ccRCC, while epithelial cluster 2, highlighted in green, is grossly copy-number normal.

(D) Table of known receptor-ligand interaction pairs in which ligand is significantly upregulated by gene expression in one cluster and receptor is significantly upregulated by VIPER in another. Subset to interactions inferred between tumor cells and T cells, or between $APOE^+/TREM2^+/C1Q^+$ tumor macrophages and tumor cells.

(E) Visualization of receptor-ligand interaction pairs shown in (D).

See also Figure S7.

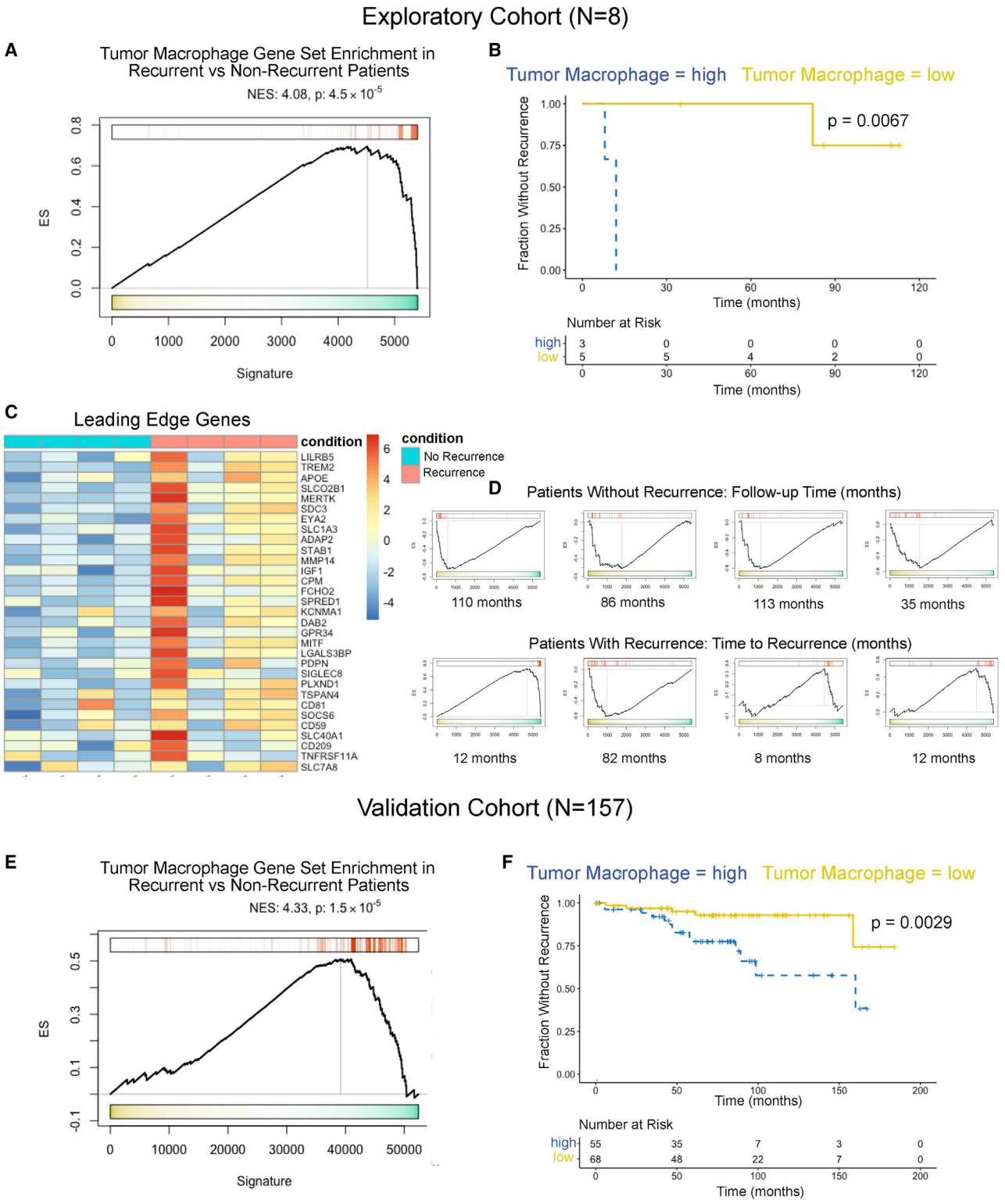


Figure 6. Enrichment of tumor-specific macrophage markers defined from single-cell RNA-seq in bulk RNA-seq data is associated with shorter time to recurrence

(A) Gene set enrichment analysis (GSEA) of tumor-specific macrophage marker proteins in VIPER-transformed bulkRNA-seq data from 4 patients with post-surgical recurrence versus 4 without. Proteins were ranked by the fold change in recurrence versus no recurrence; the p value was computed by GSEA versus gene shuffling the null model with 1,000 permutations. Note enrichment in patients with recurrence (NES = 4.08, p = $4.5*10^{-5}$).

(legend continued on next page)

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

macrophage related. We specifically interrogated C1Q, APOE, and TREM2, as the latter two were identified as protein activity markers strongly associated with clinical outcome (Figure 6C) and C1Q was highly overexpressed but could only be identified following VIPER clustering of the single-cell data. For these studies, we developed a fluorescence-based panel that included the three markers, as well as subpopulation-specific markers CA9 (tumor cells), CD3 (T cells), and CD69/CD163 (pan-macrophage) (Figure 7A). Using this panel, we stained each of the 11 samples interrogated by scRNA-seq and quantified expression in multiple segments of tumor and adjacent non-tumor tissue. Both C1Q and TREM2 were strongly enriched in macrophages across all samples (Figure 7B).

To determine which macrophage populations were comparatively tumor restricted, we tested the representation of cells co-staining for C1Q, TREM2, and CD68/CD163 within tumor stroma versus adjacent normal. In contrast to total CD68/CD163 macrophages, and as predicted by VIPER analysis, $C1Q^+$ and $TREM2^+$ macrophages were significantly tumor restricted, while the double-positive population ($C1Q^+$, $TREM2^+$) appeared to be almost completely exclusive to tumor tissue (Figure 7C). Furthermore, these $C1Q^+/TREM2^+/APOE^+$ macrophages were localized more closely to tumor cells than control macrophages ($C1Q^-TREM2^-APOE^-$) with a relative distance to the nearest $CA9^+$ cell of 15.25 um versus 23.28 um, $p = 1.7 \times 10^{-14}$, respectively. The strong tumor restriction of this population did not appear to correlate with tumor stage (Figure 7D). To further assess for association with disease recurrence, we stained the samples from the same dataset we had used for initial bulk RNA-seq studies (Figures 6A–6D). These samples showed that both TREM2 and C1Q were significantly enriched in the tumor stroma of patients with disease recurrence as compared to patients without recurrence (Figure 7E) ($p_{C1Q} = 0.047$, $p_{TREM2} = 0.038$, $p_{C1Q/TREM2} = 0.009$). $C1Q^+$ macrophages, in particular, were significantly associated with disease recurrence ($p = 0.028$). These data suggest that assessment of intra-tumoral $C1Q^+$ macrophage density by IF (Figure S6) may provide a useful prognostic biomarker for recurrence. We explored this hypothesis by first calculating a cutoff for C1Q macrophage frequency that maximized the log-rank statistic, and next performing log-rank regression (Figure 7F). A $C1Q^+$ macrophage frequency threshold of 0.01 significantly separated patients with post-surgical recurrence from those without recurrence, with a log-rank p value of 6.7×10^{-3} and area under the curve (AUC) of 0.9375. These data recapitulated the disease recurrence Kaplan-Meier curve defined by gene set enrichment analysis (GSEA) analysis (Figure 6B) and independently support the association of tumor-infiltrating macrophage density with post-surgical recurrence, highlighting a strong consistency between IF staining and scRNA-seq analysis.

DISCUSSION

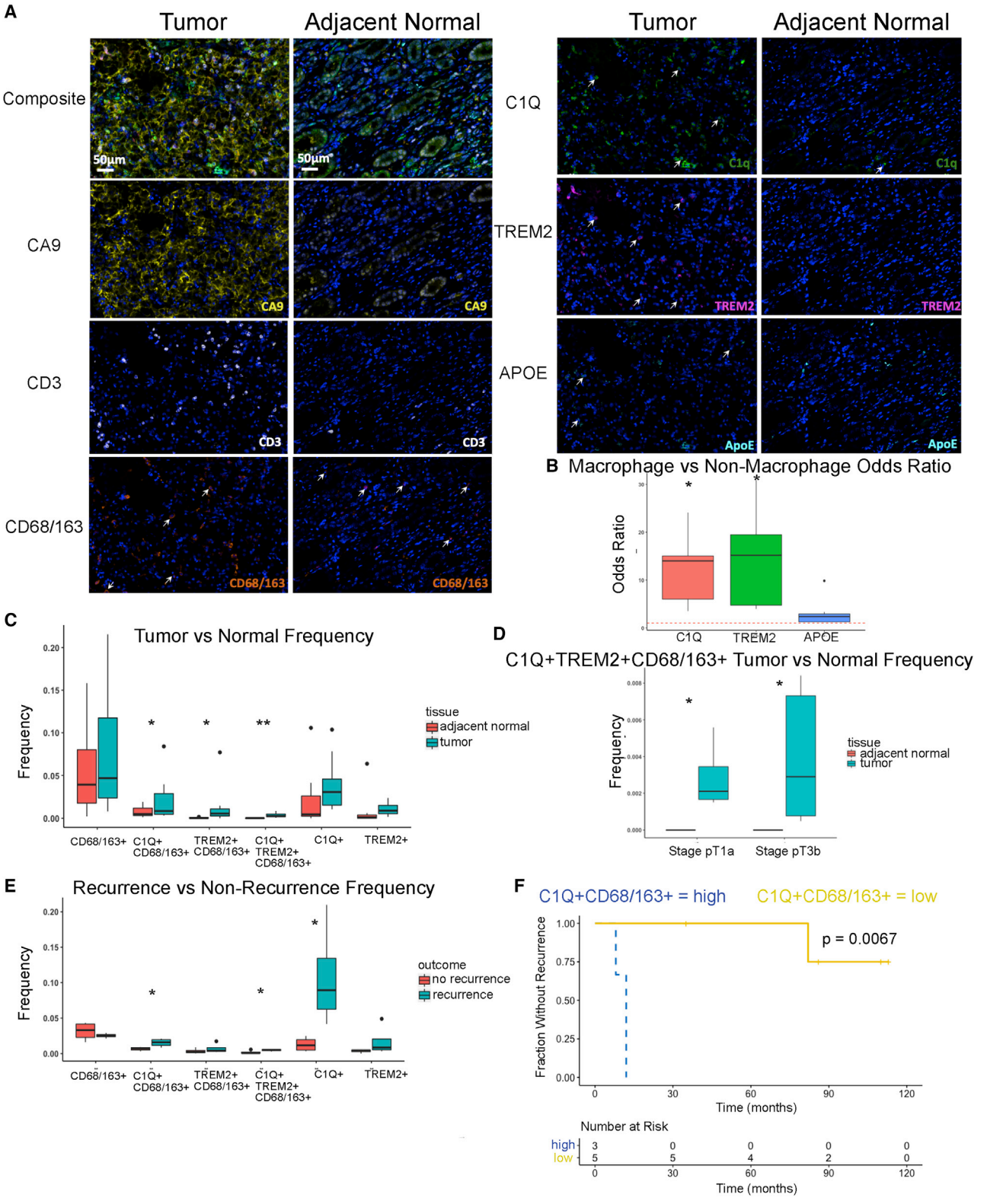
We report a systematic single-cell analysis of the cell populations that comprise the immune and non-immune compartments of clear cell renal cell carcinoma (ccRCC), from >200,000 cells representing tumor and adjacent normal tissue from 11 patients, with either stage 1 or stage 3 disease. By incorporating both transcriptomic and VIPER-based proteomic data, our analysis characterizes subpopulations, key regulatory proteins, and candidate ligand/receptor-mediated interactions, providing a previously unavailable window into the microenvironment of ccRCC.

These studies provide insight that could only be gleaned using our comprehensive VIPER-based scRNA-seq protein activity analysis pipeline. In particular, key tumor-specific populations, comprising both immune and non-immune cells, and their established lineage markers were missed by expression-based cluster analysis and by flow cytometry, due to significant gene dropout effects, a limited set of antibodies, and measurement reproducibility. By contrast, activity-based analyses provided high-resolution sub-structure and revealed a novel tumor-specific macrophage population prognostic for recurrence.

To confirm that activity-based analysis tracked protein expression (quantified using flow cytometry and IF), we performed a comprehensive validation of VIPER results using high-parameter spectral flow cytometry. Our results show that VIPER-based analyses may potentially outperform antibody-based measurements in terms of both detection and reproducibility, while providing quantitative activity assessment for >6,000 proteins in a single experiment. By contrast, gene-expression-based analyses of scRNA-seq data could not recapitulate flow cytometry results, due to significant gene dropout effects. Thus, a key novel finding of this study is feasibility, accuracy, and reproducibility of network-based protein activity inference from single-cell gene-expression profile data.

Importantly, activity-based analyses identified several known immune checkpoint and master regulatory proteins missed by gene-expression analysis alone. In exhausted CD8 T cells, for instance, these included LAG-3, PD-1, and CTLA-4, while, in Tregs, they included FOXP3 and CTLA-4. Thus, the full set of differentially active regulatory proteins reported in Table S2 represents a previously unavailable resource for the study of these cell types. Analysis of both hematopoietic and non-hematopoietic cells allowed us to study the interaction between tumor-related cells and immune subpopulations in the TME, especially with respect to tumor-infiltrating macrophages. Inference of interactions among established receptor-ligand pairs (Lizio et al., 2019) identified >200 ligand/receptor-mediated cell-cell interactions, which were consistently yet independently detected across all patients (Table S4). Among multiple potentially

(B) Kaplan-Meier curve of sample-by-sample tumor-specific macrophage GSEA associated with time to recurrence, yellow line indicates patients with low enrichment, blue line indicates patients with high enrichment. Log-rank p value = 0.0067.
(C) Heatmap of leading-edge protein set from (A).
(D) Sample-by-sample tumor macrophage GSEA, annotated with each sample's recurrence status and time to recurrence or total observation time. Proteins were ranked by inferred activity.
(E) Macrophage signature GSEA in recurrence versus no recurrence in validation cohort (N = 157).
(F) Kaplan-Meier curve of sample-by-sample GSEA in association with time to recurrence in the validation cohort, log-rank p value = 0.0029.



(legend on next page)

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

significant interactions identified (Figure S7), we highlight CD70, localized to the surface of tumor cells, interacting with its cognate ligand, CD27, on tumor-infiltrating T cells. CD70 is expressed in many solid tumors, including RCC (Jilaveanu et al., 2012), and may facilitate tumor cell escape by inducing tumor cell proliferation and survival (Jacobs et al., 2015); these data establish it as a potential therapeutic target in ccRCC.

In terms of potential clinical relevance, activity-based analysis identified a tumor-specific macrophage subpopulation characterized by upregulation of C1Q, APOE, and TREM2 and high activity of the LILRB5 protein (Figures 1 and 2). This subpopulation was consistently detected in all tumors, and GSEA analysis of its single-cell RNA-seq protein signature in independent bulk RNA-seq profiles revealed its significant association with shorter time to post-surgical recurrence. These findings were confirmed in a validation cohort of 157 patients (Figure 6). Of note, the VIPER-based test to measure activity of these proteins in patients, based on their tumor's mRNA profile (OncoTarget) (Zeleke et al., 2020), recently received CLIA certification by the New York and California departments of health (Alvarez et al., 2018; Alvarez and Califano, 2018).

Protein-level qmIF confirmed the clinical significance of $C1Q^+$ tumor-specific macrophages and recapitulated association with shorter time to post-surgical recurrence identified at the transcriptional level (Figure 7F). Though the functional role of these tumor-specific macrophages is currently unknown, a recent study in primary renal tumors also found that high density of C1Q-expressing cells correlates with poor prognosis (Roumenina et al., 2019) although that study did not characterize these findings at the single-cell level or report their tumor specificity and interaction with tumor cells. Functionally, complement-deficient mice developed high densities of C1Q expressing macrophages with concurrent upregulation of immune checkpoints PD-1, LAG-3, and PD-L1. These orthogonal data are consistent with our observations in human single-cell data, as we documented a C1Q-expressing tumor-specific macrophage population and a high frequency of likely exhausted $LAG\text{-}3^+$ $PD\text{-}1^+$ T cells in the ccRCC TME.

Our studies also highlighted APOE and TREM2—a member of the immunoglobulin superfamily that plays an important immunomodulatory role in the regulation of inflammatory processes (Ramirez et al., 2015; Roussos et al., 2015) and enhances tumor proliferation (Wang et al., 2016; Yao et al., 2016). The role of the TREM2-ApoE pathway in RCC tumor biology has not been fully explored. A recent study profiling a murine ccRCC model using scRNA-seq in conjunction with intracellular proteomic staining identified a population of $TREM2^+$ tumor-infiltrating macrophages, which appears to be phenotypically similar to the population we discovered in patients (Katzenelenbogen et al., 2020). Functional studies showed that co-culture of CD8 T cells with these macrophages significantly impeded T cell proliferation, and that TREM2 knockdown led to favorable pre-clinical outcomes.

In conclusion, we report the development and application of a novel and broadly generalizable scRNA-seq analytic pipeline, which complements gene expression with inferred protein activity to comprehensively dissect the repertoire of subpopulations in the TME. While our analysis focused on treatment-naive clear cell renal carcinoma, our validation with proteins concurrently profiled by flow cytometry suggests that this approach could be effectively applied to any tumor of interest and potentially to other tissue-based studies. Our scRNA-seq data are limited by the relatively small number of cases, but it should be noted that we cumulatively profiled > 200,000 cells with high data quality (Table S1) and that the populations identified were remarkably consistent across patients (Figure S2), suggesting that additional patients would not dramatically affect the conclusions of the study. Enrichment of single-cell signatures in bulk data and qmIF studies showed strong and statistically significant association between tumor infiltration by a C1Q-expressing macrophage subpopulation and disease recurrence. One implication of these findings is that ccRCC patients with an increased density of C1Q-expressing macrophages in the tumor stroma at baseline might be at increased risk of post-surgical disease recurrence and thus may be suitable candidates for adjuvant therapy or more aggressive neoadjuvant approaches in the context of clinical trials. A more intriguing possibility is that these cells could be causal of (rather than associated with) recurrence; hence, targeting their top master regulators and/or proteomic markers could be of clinical value in ccRCC.

Limitations of the study

Consistent with prior results, we found that recovery rates for protein activity inference using this analysis pipeline were in the 70%–80% range, i.e., 20%–30% of differentially active proteins may be missed. Although this compares favorably with

Figure 7. A novel population of $C1Q/TREM2^+$ macrophages are tumor specific and associated with shorter time to recurrence by immunohistochemistry

(A) Representative IHC images for each marker in tumor stroma versus adjacent normal. Note high C1Q/TREM2/APOE staining within $CA9^+$ tumor as compared to tumor-adjacent ($CA9^-$) regions.

(B) Odds ratios (OR) across samples of tumor-specific macrophage markers C1Q, TREM2, and APOE co-staining with $CD68/CD163^+$ macrophage cells versus $CD68/CD163^-$ non-macrophage cells; note association of C1Q and TREM2 with macrophage markers. Dotted red line represents OR = 1. Individual OR for C1Q and TREM2 co-staining with CD68/CD163 is statistically significant by Fisher's exact test ($p < 0.01$).

(C) Frequency by IHC of $C1Q^+$ or $TREM2^+$ macrophages in tumor stroma versus adjacent normal across the 11 patient samples profiled by scRNA-seq. Enrichment in tumor compared to adjacent normal assessed by paired Wilcox test, $^*p < 0.05$, $^{**}p < 0.01$.

(D) Frequency of $C1Q^+TREM2^+CD68/CD163^+$ macrophages in tumor versus adjacent normal, plotted by stage (pT1a versus pT3b). No $C1Q^+TREM2^+CD68/CD163^+$ cells were present in adjacent normal.

(E) Frequency of $C1Q^+$ or $TREM2^+$ macrophages in tumor stroma of patients with or without post-surgical recurrence, from the cohort profiled by bulkRNA-seq in Figures 6A–6D. Higher frequency in patients with recurrence assessed by unpaired Wilcox test, $^*p < 0.05$.

(F) Kaplan-Meier plot of $C1Q^+CD68/CD163^+$ frequency in association with time to recurrence. Log-rank p value = 0.0067, with sample-by-sample frequency binarized by log-rank maximization to >0.01 = "high" and <0.01 = "low."

See also Figure S6.

gene expression, where >80%–90% of genes may be undetected, we expect that future studies aimed at improving the population-specific reporter assays used to infer protein activity by VIPER will address these limitations.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Tissue dissociation
  - Single-cell RNA sequencing
  - Single-cell RNA-seq gene expression processing
  - Resolution-optimized Louvain clustering algorithm
  - Semi-supervised cell type calling
  - Regulatory network inference
  - Protein activity inference
  - Copy number inference
  - Flow cytometry data acquisition
  - Multi-dimensional analysis of flow cytometry data
  - Receptor-ligand interaction inference
  - Staining for multiplex immunohistochemistry
  - Multispectral imaging
  - Image analysis
  - Association of tumor macrophage signature with clinical recurrence
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.04.038.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute's Office of Cancer Genomics Cancer Target Discovery and Development (CTD2) initiative. The results published here are based in whole or in part upon data generated by Cancer Target Discovery and Development (CTD2) Network (https://ocg.cancer.gov/programs/ctd2/data-portal) established by the National Cancer Institute's Office of Cancer Genomics. This research was also supported by National Institutes of Health (NIH) grants R01 CA127153, 1P50CA58236-15, and P30CA006973 and CUMC institutional funds to C.G.D.; by NIH grants R35CA197745, U01DA217858, S10 OD012351, and S10 OD021764 to A.C. and by NIH grant R38 grant CA231577 to D.H.A. The independent validation RNA-seq dataset was supported by DOD/CDMRP award W81XWH-17-1-0309 under proposal CA160728. The P30CA013696 Cancer Center grant supports of the Flow Cytometry Core at HICCC Siu-Hong-Ho and Luke Caisheng assisted with FACS sorting.

AUTHOR CONTRIBUTIONS

A.O., N.C., C.G.D., and A.C. wrote and edited the manuscript. A.O. generated figures and performed bioinformatic/statistical analyses. A.O., N.C., A.C., and C.G.D. conceived of and designed the study. J.M. provided clinical samples. N.C., V.W., D.A., and A.O. processed tissue for scRNA-seq. X.G. performed scRNA-seq. C.A. and D.A. performed flow cytometry. A.O. and L.V. coded the single-cell analysis pipeline. N.C. generated multispectral image data. S.H. coordinated collection of tumor samples and RNA sequencing. J.J. reviewed samples for inclusion. M.R., W.K.R., and E.J. performed clinical annotation. K.B. and B.R. coordinated data transfer. C.G.D. oversaw flow, immunostaining, and overall experimental design. A.C. advised on data processing/visualization and algorithms.

DECLARATION OF INTERESTS

C.G.D. is a co-inventor on patents licensed from JHU to BMS and Janssen; has served as a paid consultant to AZ Medimmune, BMS, Pfizer, Roche, Sanofi Aventis, Genentech, Merck, and Janssen; and has received sponsored research funding to his institution from BMS IIoN and Janssen. A.C. is founder, equity holder, consultant, and director of DarwinHealth Inc., which has licensed IP related to these algorithms from Columbia University. Columbia University is an equity holder in DarwinHealth Inc. B.I.R. has served as a paid consultant to BMS, Pfizer, GNE/Roche, Aveo, Synthorx, Compugen, Merck, Corvus, Surface Oncology, 3DMedicines, Arravive, Alkermes, Arrowhead, GSK, and Shionogi and holds stock in PTC therapeutics.

Received: June 2, 2020
Revised: February 10, 2021
Accepted: April 23, 2021
Published: May 20, 2021

REFERENCES

Alvarez, M.J., and Califano, A. (2018). Darwin OncoTarget/OncoTreat: NY CLIA certified tests to identify effective drugs on an individual cancer patient basis from RNASeq profiles (Dept of Pathology and Cell Biology Web Site, Columbia University).

Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat. Genet. *48*, 838–847.

Alvarez, M.J., Subramaniam, P.S., Tang, L.H., Grunn, A., Aburi, M., Rieckhof, G., Komissarova, E.V., Hagan, E.A., Bodei, L., Clemons, P.A., et al. (2018). A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. Nat. Genet. *50*, 979–989.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172.

Arce Vargas, F., Furness, A.J.S., Solomon, I., Joshi, K., Mekkaoui, L., Lesko, M.H., Miranda Rota, E., Dahan, R., Georgiou, A., Sledzinska, A., et al.; Melanoma TRACERx Consortium; Renal TRACERx Consortium; Lung TRACERx Consortium (2017). Fc-Optimized Anti-CD25 Depletes Tumor-Infiltrating Regulatory T Cells and Synergizes with PD-1 Blockade to Eradicate Established Tumors. Immunity *46*, 577–586.

Arik, D., Can, C., Dündar, E., Kabukçuoğlu, S., and Paşaoğlu, Ö. (2017). Prognostic Significance of CD24 in Clear Cell Renal Cell Carcinoma. Pathol. Oncol. Res. *23*, 409–416.

Arumugam, K., Shin, W., Schiavone, V., Vlahos, L., Tu, X., Carnevali, D., Kesner, J., Paull, E.O., Romo, N., Subramaniam, P., et al. (2020). The Master Regulator Protein BAZ2B Can Reprogram Human Hematopoietic Lineage-Committed Progenitors into a Multipotent State. Cell Rep. *33*, 108474.

Barkal, A.A., Brewer, R.E., Markovic, M., Kowarsky, M., Barkal, S.A., Zaro, B.W., Krishnan, V., Hatakeyama, J., Dorigo, O., Barkal, L.J., et al. (2019). CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature *572*, 392–396.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat. Genet. *37*, 382–390.

Becht, E., Giraldo, N.A., Beuselinck, B., Job, S., Marisa, L., Vano, Y., Oudard, S., Zucman-Rossi, J., Laurent-Puig, P., Sautès-Fridman, C., et al. (2015). Prognostic and theranostic impact of molecular subtypes and immune classifications in renal cell cancer (RCC) and colorectal cancer (CRC). OncoImmunology *4*, e1049804.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318–325.

Chao, J.L., and Savage, P.A. (2018). Unlocking the Complexities of Tumor-Associated Regulatory T Cells. J. Immunol. *200*, 415–421.

Chevrier, S., Levine, J.H., Zanotelli, V.R.T., Silina, K., Schulz, D., Bacac, M., Ries, C.H., Ailles, L., Jewett, M.A.S., Moch, H., et al. (2017). An Immune Atlas of Clear Cell Renal Cell Carcinoma. Cell *169*, 736–749.

Chung, W., Eum, H.H., Lee, H.O., Lee, K.M., Lee, H.B., Kim, K.T., Ryu, H.S., Kim, S., Lee, J.E., Park, Y.H., et al. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat. Commun. *8*, 15081.

Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*, eaaf8399.

Ding, H., Douglass, E.F., Jr., Sonabend, A.M., Mela, A., Bose, S., Gonzalez, C., Canoll, P.D., Sims, P.A., Alvarez, M.J., and Califano, A. (2018). Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. Nat. Commun. *9*, 1471.

Dutta, A., Le Magnen, C., Mitrofanova, A., Ouyang, X., Califano, A., and Abate-Shen, C. (2016). Identification of an NKX3.1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. Science *352*, 1576–1580.

Elyada, E., Bolisetty, M., Laise, P., Flynn, W.F., Courtois, E.T., Burkhart, R.A., Teinor, J.A., Belleau, P., Biffi, G., Lucito, M.S., et al. (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov. *9*, 1102–1123.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

Farber, N.J., Kim, C.J., Modi, P.K., Hon, J.D., Sadimin, E.T., and Singer, E.A. (2017). Renal cell carcinoma: the search for a reliable biomarker. Trans. Cancer Res. *6*, 620–632. https://doi.org/10.21037/tcr.2017.05.19.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278.

Hsieh, J.J., Le, V.H., Oyama, T., Ricketts, C.J., Ho, T.H., and Cheng, E.H. (2018). Chromosome 3p Loss-Orchestrated VHL, HIF, and Epigenetic Deregulation in Clear Cell Renal Cell Carcinoma. J. Clin. Oncol. *36*, JCO2018792549.

Jacobs, J., Deschoolmeester, V., Zwaenepoel, K., Rolfo, C., Silence, K., Rottey, S., Lardon, F., Smits, E., and Pauwels, P. (2015). CD70: An emerging target in cancer immunotherapy. Pharmacol. Ther. *155*, 1–10.

Jerby-Arnon, L., Shah, P., Cuoco, M.S., Rodman, C., Su, M.J., Melms, J.C., Leeson, R., Kanodia, A., Mei, S., Lin, J.R., et al. (2018). A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell *175*, 984–997.

Jilaveanu, L.B., Sznol, J., Aziz, S.A., Duchen, D., Kluger, H.M., and Camp, R.L. (2012). CD70 expression patterns in renal cell carcinoma. Hum. Pathol. *43*, 1394–1399.

Katzenelenbogen, Y., Sheban, F., Yalin, A., Yofe, I., Svetlichnyy, D., Jaitin, D.A., Bornstein, C., Moshe, A., Keren-Shaul, H., Cohen, M., et al. (2020). Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer. Cell *182*, 872–885.

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

Koul, H., Huh, J.S., Rove, K.O., Crompton, L., Koul, S., Meacham, R.B., and Kim, F.J. (2011). Molecular aspects of renal cell carcinoma: a review. Am. J. Cancer Res. *1*, 240–254.

Lachmann, A., Giorgi, F.M., Lopez, G., and Califano, A. (2016). ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics *32*, 2233–2235.

Li, X., Ma, X., Chen, L., Gu, L., Zhang, Y., Zhang, F., Ouyang, Y., Gao, Y., Huang, Q., and Zhang, X. (2015). Prognostic value of CD44 expression in renal cell carcinoma: a systematic review and meta-analysis. Sci. Rep. *5*, 13157.

Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C.C., de Hoon, M., Severin, J., Oki, S., Hayashizaki, Y., et al. (2019). Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. Nucleic Acids Res. *47* (D1), D752–D758.

Mani, K.M., Lefebvre, C., Wang, K., Lim, W.K., Basso, K., Dalla-Favera, R., and Califano, A. (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. Mol. Syst. Biol. *4*, 169.

Martens, J.H., and Stunnenberg, H.G. (2013). BLUEPRINT: mapping human blood cell epigenomes. Haematologica *98*, 1487–1489. https://doi.org/10.3324/haematol.2013.094243.

Ramirez, A., Karsak, M., Thelen, M., Glebov, K., Nagler, M., Rading, S., Thiele, H., Lernnarz, M., Jessen, F., and Maier, W. (2015). Functional characterization of a novel TREM2 coding variant linked to familial Alzheimer's disease. Alzheimers Dement. *11*, 500.

Roumenina, L.T., Daugan, M.V., Noé, R., Petitprez, F., Vano, Y.A., Sanchez-Salas, R., Becht, E., Meilleroux, J., Clec'h, B.L., Giraldo, N.A., et al. (2019). Tumor Cells Hijack Macrophage-Produced Complement C1q to Promote Tumor Growth. Cancer Immunol. Res. *7*, 1091–1105.

Roussos, P., Katsel, P., Fam, P., Tan, W., Purohit, D.P., and Haroutunian, V. (2015). The triggering receptor expressed on myeloid cells 2 (TREM2) is associated with enhanced inflammation, neuropathological lesions and increased risk for Alzheimer's dementia. Alzheimers Dement. *11*, 1163–1170.

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.

Sánchez-Gastaldo, A., Kempf, E., González Del Alba, A., and Duran, I. (2017). Systemic treatment of renal cell cancer: A comprehensive review. Cancer Treat. Rev. *60*, 77–89.

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

Şenbabaoğlu, Y., Gejman, R.S., Winer, A.G., Liu, M., Van Allen, E.M., de Velasco, G., Miao, D., Ostrovnaya, I., Drill, E., Luna, A., et al. (2016). Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol. *17*, 231.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., et al. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 7.

Tan, H.-S., Jiang, W.-H., He, Y., Wang, D.-S., Wu, Z.-J., Wu, D.-S., Gao, L., Bao, Y., Shi, J.-Z., Liu, B., et al. (2017). KRT8 upregulation promotes tumor metastasis and is predictive of a poor prognosis in clear cell renal cell carcinoma. Oncotarget *8*, 76189–76203.

Talos, F., Mitrofanova, A., Bergren, S.K., Califano, A., and Shen, M.M. (2017). A computational systems approach identifies synergistic specification genes that facilitate lineage conversion to prostate tissue. Nat. Commun. *8*, 14662.

Tickle, T., Tirosh, I., Georgescu, C., Brown, M., and Haas, B. (2019). inferCNV of the Trinity CTAT Project. (Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA).

Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J.L., Wong, Y.N.S., Rowan, A., Kanu, N., Al Bakir, M., et al. (2017). Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. Lancet Oncol. *18*, 1009–1021.

Wang, X.Q., Tao, B.B., Li, B., Wang, X.H., Zhang, W.C., Wan, L., Hua, X.M., and Li, S.T. (2016). Overexpression of TREM2 enhances glioma cell proliferation and invasion: a therapeutic target in human glioma. Oncotarget *7*, 2354–2366.

Yao, Y., Li, H., Chen, J., Xu, W., Yang, G., Bao, Z., Xia, D., Lu, G., Hu, S., and Zhou, J. (2016). TREM-2 serves as a negative immune regulator through Syk pathway in an IL-10 dependent manner in lung cancer. Oncotarget *7*, 29620–29634.

Zeleke, T., Pan, Q., Chiuzan, C., Onishi, M., Alvarez, M.J., Honan, E., Yang, M., Chia, P.L., Mukhopadhyay, P., Kelly, S., et al. (2020). Network-based assessment of HDAC6 activity is highly predictive of pre-clinical and clinical responses to the HDAC6 inhibitor ricolinostat. medRxiv. https://doi.org/10.1101/2020.04.23.20066928.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-Human CD45 Qdot 800, clone HI30	Thermo Fisher Scientific	Cat#Q10156; RRID:AB_1500477
Mouse anti-Human CD19 BB515, clone HIB19	BD Biosciences	Cat#564456
Anti-Human CD45RA FITC, clone HI100	BioLegend	Cat#304148; RRID:AB_2564157
Anti-Human FoxP3 PE, clone 236A/E7	Thermo Fisher Scientific	Cat#12-4777-42
Mouse anti-Human PD-1 PE-Dazzle/594, clone EH12.2H7	BioLegend	Cat#329939
Mouse anti-Human CD127 PE-Cy5, clone A019D5	BioLegend	Cat#351525
Mouse anti-Human CD38 PerCP, clone HIT2	BioLegend	Cat#303519; RRID:AB_893315
Mouse anti-Human CD39 PerCP-Cy5.5, clone A1	BioLegend	Cat#328218
Anti-Human CD25 Pe-Cy7, clone BC96	BioLegend	Cat#302611
Mouse anti-Human CD14 APC, clone 63D3	BioLegend	Cat#367118; RRID:AB_2566792
Anti-Human CTLA-4 Alexa 647, clone L3D10	BioLegend	Cat#349920; RRID:AB_2566185
Mouse anti-CD3 Alexa 700, clone HIT3a	BioLegend	Cat#300324; RRID:AB_493739
Anti-Human CD73 APC-Cy7, clone AD2	BioLegend	Cat#344021
Anti-Human CXCR3 BV421, clone G025H7	BioLegend	Cat#353715; RRID:AB_11124720
Anti-Human CD8 Pacific Blue, clone SK1	BioLegend	Cat#344718
Mouse anti-Human CD4 BV480, clone SK3	BD Biosciences	Cat#566104
Anti-Human CD16 BV570, clone 3G8	BioLegend	Cat#302035; RRID:AB_10915988
Anti-Human CRTH2 BV605, clone BM16	BioLegend	Cat#350121; RRID:AB_2566759
Mouse anti-Human HLA-DR BV650, clone L243	BioLegend	Cat#307649; RRID:AB_2562544
Mouse anti-Human CD161 BV711, clone DX12	BD Biosciences	Cat#563865
Mouse anti-Human CD56 BV750, clone 5.1H11	BioLegend	Cat#362555; RRID:AB_2734396
Anti-Human CCR7 BV785, clone G043H7	BioLegend	Cat#353230; RRID:AB_2563630
Mouse anti-Human CD86 FITC, clone BU63	BioLegend	Cat#374203; RRID:AB_2721573
Anti-Human CD14 Alexa 532, clone 61D3	Thermo Fisher	Cat#58-0149-41
Anti-Human CD141 PE, clone M80	BioLegend	Cat#344103; RRID:AB_1877220
Mouse anti-Human PD-L1 PE-Dazzle/594, clone 29E.2A3	BioLegend	Cat#329731
Anti-Human CD11b PE-Cy5, clone ICRF44	BioLegend	Cat#301307
Mouse anti-Human CXCR2 PerCP-Cy5.5, clone 538/CXCR2	BioLegend	Cat#320717
Anti-Human CD204 PE-Cy7, clone PSL204	Thermo Fisher Scientific	Cat#25-2045-42
Mouse anti-Human CD68 APC, clone Y1/82A	BioLegend	Cat#333809; RRID:AB_10567107

(*Continued on next page*)

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Anti-Human CD206 APC/Fire 750, clone 15-2	BioLegend	Cat#321133
Mouse anti-Human CD1c BV421, clone L161	BioLegend	Cat#331525; RRID:AB_10933249
Anti-Human CD123 Pacific Blue, clone 6H6	BioLegend	Cat#306043
Mouse anti-Human CD163 BV480, clone GHI/61	BD Biosciences	Cat#746549
Mouse anti-Human CD169 BV605, clone 7-239	BioLegend	Cat#346009; RRID:AB_2721538
Anti-Human CD33 BV711, clone P67.6	BioLegend	Cat#366623; RRID:AB_2721556
Mouse anti-Human HLA-ABC BV786, clone G46-2.6	BD Biosciences	Cat#740982
Mouse anti-Human CD81 PE-Cy7, clone 5A6	BioLegend	Cat#349511
Mouse anti-Human CD74 PE, clone LN2	BioLegend	Cat# 326807; RRID:AB_2229059
Mouse anti-Human HLA-DR Alexa 700, clone L243	BioLegend	Cat# 307625; RRID:AB_493770
Mouse anti-Human CD169 APC, clone 7-239	BioLegend	Cat# 346007; RRID:AB_11150773
Mouse anti-Human CD68 PE, clone Y1/82A	BioLegend	Cat# 333807
Mouse anti-Human CD206 BV785, clone 15-2	BioLegend	Cat# 321141; RRID:AB_2734301
Mouse anti-Human CD45 BV605, clone HI30	BioLegend	Cat# 304041; RRID:AB_2562105
Rabbit anti-human TREM2, clone D8I4C	Cell Signaling Technologies	Cat# 91068S
Mouse anti-human C1q, clone C1QA/2956	AbCam	Cat# ab268120
Mouse anti-human CD3, clone LN10	Leica Biosystems	Cat# NCL-L-CD3-565
Rabbit anti-human ApoE, clone D17N	Cell Signaling Technologies	Cat# 13366S
Rabbit anti-human CA-9, polyclonal	AbCam	Cat# ab15086
Mouse anti-human CD68, clone KP1	BioGeneX	Cat# AM416-5M
Mouse anti-human CD163, clone 10D6	AbCam	Cat# ab74604
Human TruStain FcX	BioLegend	Cat#422302
eBioscience FoxP3/Transcription Factor Staining Buffer Set	Thermo Fisher Scientific	Cat# 00-5523-00
GIBCO Fetal Bovine Serum, Certified, Heat Inactivated	Thermo Fisher Scientific	Cat#10-082-147
Bovine Serum Albumin	Sigma	Cat#A9647-100 g
GIBCO DPBS (1x)	Thermo Fisher Scientific	Cat#14190-144
Biological samples		
Untreated ccRCC Tumor and Adjacent Normal Tissue: Fresh Samples	Columbia University Irving Medical Center	N/A
Untreated ccRCC Tumor FFPE tissue	Columbia University Irving Medical Center	N/A
Untreated ccRCC Tumor FFPE tissue	Vanderbilt University Medical Center	N/A
Critical commercial assays		
Miltenyi Mouse Tumor Dissociation Kit	Miltenyi Biotec	Cat#130-096-730; RRID:SCR_020285
10x Genomics Chromium Single Cell 3′ Reagent Kit	10X Genomics	N/A
Opal-7-color multiplex IHC kit	Akoya Biosciences	SKU NEL811001KT

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and Analyzed data used to generate analyses shown in this Manuscript: scRNASeq, bulkRNASeq, qmIF, cyTEK spectral flow cytometry	This Manuscript	Available on Mendeley at https://dx.doi.org/10.17632/nc9bc8dn4m.1 And Github at https://github.com/Aleksobrad/single-cell-rcc-pipeline
Software and algorithms		
FlowJo v10.6.2	BD Biosciences	https://www.flowjo.com/
SpectroFlo v2.0	Cytek	https://cytekbio.com/
GraphPad Prism v8.4	GraphPad	https://www.graphpad.com/scientific-software/prism/
InForm Advanced Image Analysis Software v 2.4.6	Akoya Biosciences	https://www.akoyabio.com/phenoptics/software/inform-tissue-finder/
PhenoChart v 1.0.12	Akoya Biosciences	https://akoyabio.helpdocs.com/phenoptics-software-updates/phenochart-version-1011
Seurat v3	(Stuart et al., 2019)	https://cran.r-project.org/web/packages/Seurat/index.html
SingleR	Aran et al., 2019	https://bioconductor.org/packages/release/bioc/html/SingleR.html
InferCNV	(Tickle et al., 2019)	https://github.com/broadinstitute/infercnv
ARACNe	Lachmann et al., 2016	https://github.com/califano-lab/ARACNe-AP
VIPER	Alvarez et al., 2016	http://bioconductor.org/packages/release/bioc/html/viper.html
Resolution-Optimized Louvain Clustering	This manuscript	https://github.com/Aleksobrad/single-cell-rcc-pipeline
R Code Used to Generate Figures and perform analyses shown in this manuscript	This manuscript	https://github.com/Aleksobrad/single-cell-rcc-pipeline
Single-Cell ARACNe & VIPER analysis pipeline in active development	This manuscript	https://github.com/califano-lab/PISCES

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Dr. Charles G. Drake (cgd2139@columbia.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data files and specific code used to perform all analyses in this manuscript are available at https://github.com/Aleksobrad/single-cell-rcc-pipeline. General pipeline for VIPER analysis of scRNASeq data is available as an actively maintained and updated R package at https://github.com/califano-lab/PISCES. Source data for all analysis in this study will also be publicly hosted on Mendeley Data: https://doi.org/10.17632/nc9bc8dn4m.1

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fresh Surgical Tumor Tissue and Matched Adjacent Normal Tissue were obtained from 11 adult patients with clear cell renal carcinoma undergoing nephrectomy to treat primary, non-metastatic disease, ranging from grade 1 to grade 4 and including 6 patients with stage pT1a disease and 5 patients with pT3a disease (Table S1). These were dissociated immediately for Single-cell RNASeq and flow Cytometry analysis. Formalin-Fixed Paraffin-Embedded (FFPE) Pathology samples of the same patient tumors were obtained for follow-up Immunohistochemistry analysis. A separate cohort of FFPE tissue from primary tumor of 11 adult patients also treated for clear cell renal carcinoma by surgical nephrectomy was curated retrospectively to identify patients who recurred after

surgery and match them by age and tumor stage to those who didn't. This cohort was tracked over a period of 5-113 months, during which time 6 patients experienced disease recurrence (between 5-82 months after surgery) and 5 patients had no recurrence (between 35-113 months after surgery). We used this cohort for validation of Immunohistochemical analysis, as well as for exploratory profiling by bulk RNA sequencing and association of markers identified from single-cell profiling with time-to-recurrence. A larger validation cohort was curated from the Vanderbilt tissue bank, identifying 157 treatment-naive patients with varying follow-up time and annotation of post-surgical disease recurrence and profling them by bulk RNA sequencing. The studies were conducted in accordance with the guidelines approved by the Institutional Review Board (IRB) protocols, AAAO5706 and AAAA9967, respectively.

METHOD DETAILS

Tissue dissociation

Fresh Tumor or Adjacent Normal tissue were minced to 2-4 mm sized pieces in separate 6-cm dishes and digested to single cell suspension using Multi Tissue Human Tumor Dissociation Kit 1 (Miltenyi Biotec) and a gentleMACS OctoDissociator (Miltenyi Biotec) according to the manufacturer's instructions. Dissociated cells from both Tumor and Adjacent Normal tissue were aliquoted for Flow Cytometry Analysis and single-cell sequencing, with 2-3x10^6 cells allocated for flow cytometry and the remainder used for single-cell sequencing. Cells aliquoted for sequencing were stained for Live/Dead (eBioscience, cat#50-112-9035) and CD45 (BioLegend, cat#368524) and then fluorescence-activated cell sorted (FACS) using BD Influx^{TM} cell sorter into a Live CD45 positive population and a Live CD45 negative population, each of which were separately loaded for single-cell RNA sequencing. Boundaries between positive and negative cell fractions were drawn based on single-color stain. An example gating strategy is shown in Figure S1. For the first set of three patients processed (Patients A-C), only the sorted CD45-positive population was further processed for single-cell RNASequencing, and for a second set of eight patients (patients 1-8), both CD45-positive and CD45-negative cells were processed for single-cell RNASequencing.

Single-cell RNA sequencing

Sorted CD45-positive and CD45-negative samples were processed for single-cell gene expression capture (scRNASeq) using the 10X Chromium 3′ Library and Gel Bead Kit (10x Genomics), following the manufacturer's user guide at the Columbia University Human Immune Monitoring Core (HIMC). After GelBead in-Emulsion reverse transcription (GEM-RT) reaction, 12-15 cycles of polymerase chain reaction (PCR) amplification were performed to obtain cDNAs used for RNaseq library generation. Libraries were prepared following the manufacturer's user guide and sequenced on Illumina NovaSeq 6000 Sequencing System. Single-cell RNASeq data were processed with Cell Ranger software at the Columbia University Single Cell Analysis Core. Illumina base call files were converted to FASTQ files with the command ''cellranger mkfastq.'' Expression data were processed with ''cellranger count'' on the pre-built human reference set of 30,727 genes. Cell Ranger performed default filtering for quality control, and produced for each sample a barcodes.tsv, genes.tsv, and matrix.mts file containing counts of transcripts for each sample, such that expression of each gene is in terms of the number of unique molecular identifiers (UMIs) tagged to cDNA molecules corresponding to that gene. These data were loaded into the R version 3.6.1 programming environment, where the publicly available Seurat package was used to further quality-control filter cells to those with fewer than 10% mitochondrial RNA content, more than 1,500 unique UMI counts, and fewer than 15,000 unique UMI counts. Pooled distribution across all samples of UMI counts, unique gene counts, and percentage of mitochondrial DNA after QC-filtering is shown in Figure S1, with total post-filtering cell counts and median UMIs/cell shown for each individual sample in Table S1.

Single-cell RNA-seq gene expression processing

Gene Expression UMI count matrices for each sample were processed in R using the Seurat SCTransform command to perform a regularized negative binomial regression based on the 3000 most variable genes. Each sample was then individually clustered by the Resolution-Optimized Louvain Clustering Algorithm described below, and within each cluster metaCells were computed for downstream regulatory network inference by summing SCTransform-corrected template counts for the 10 nearest neighbors of each cell by Pearson correlation distance. Normalized datasets for both Tumor and Adjacent Normal tissue across all patients were combined separately for CD45-positive and CD45-negative samples using the FindIntegrationAnchors and IntegrateData functions in Seurat, with the default parameters. The resulting datasets of 102,509 CD45-positive cells from 11 patients and 61,423 CD45-negative cells from 8 patients were projected into their first 50 principal components using the RunPCA function in Seurat, and further reduced into a 2-dimensional visualization space using the RunUMAP function with method umap-learn and Pearson correlation as the distance metric between cells. Differential Gene Expression between clusters was computed by the MAST hurdle model for single-cell gene expression modeling, as implemented in the Seurat FindAllMarkers command, with log fold change threshold of 0.5 and minimum fractional expression threshold of 0.25, indicating that the resulting gene markers for each cluster are restricted to those with log fold change greater than 0 and non-zero expression in at least 25% of the cells in the cluster.

Resolution-optimized Louvain clustering algorithm

For each clustering step in the analysis, clustering was performed in two steps. The Louvain algorithm as implemented in Seurat uses the FindNeighbors and FindClusters functions, such that the FindClusters function includes a resolution parameter that allows

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

selection of a progressively higher number of clusters as the parameter is increased, which does not control for over-clustering or allow for objective evaluation of cluster purity. Therefore, clustering was performed with resolution values ranging from 0.01 to 1.0 at intervals of 0.01, and cluster quality was evaluated at each resolution value to select an optimum in this range. For each resolution value, the clustered cells were subsampled to 1000, and silhouette score was computed for these 1000 cells and their cluster labels. For gene expression data, Pearson correlation was used as the distance metric in computation of silhouette score, and for VIPER-inferred protein activity data ViperSimilarity as implemented in the VIPER package was used as the distance metric. This procedure was repeated for 100 random samples of 1000 cells to compute a mean and standard deviation of average silhouette score at each resolution value. Examples can be seen plotted in Figure S4. The highest resolution value that maximizes mean silhouette score was selected as the optimal resolution at which to cluster the data without over-clustering.

Semi-supervised cell type calling

For each single cell gene expression sample, cell-by-cell identification of cell types was performed using the SingleR package and the preloaded Blueprint-ENCODE reference, which includes normalized expression values for 259 bulk RNASeq samples generated by Blueprint and ENCODE from 43 distinct cell types representing pure populations of stroma and immune cells (Martens and Stunnenberg, 2013; ENCODE Project Consortium, 2012). The SingleR algorithm computer correlation between each individual cell and each of the 259 reference samples, and then assigns both a label of the cell type with highest average correlation to the individual cell and a p value computed by wilcox test of correlation to that cell type compared to all other cell types. Projection of cell-by-cell SingleR labels with $p < 0.05$ onto the Gene Expression UMAP space is shown in Figure S3, such that localization of SingleR labels is highly concordant with the unsupervised clustering. Unsupervised Clusters determined by the resolution-optimized Louvain algorithm are labeled as a particular cell type based on the most-represented SingleR cell type label within that cluster.

Regulatory network inference

From each sample, metaCells were computed within each gene expression cluster by summing SCTransform-corrected template counts for the 10 nearest neighbors of each cell by Pearson correlation distance. 200 metaCells per cluster were sampled to compute a regulatory network from each cluster in each patient. All regulatory networks were reverse engineered by the ARACNe algorithm. ARACNe was run with 100 bootstrap iterations using 1785 transcription factors (genes annotated in gene ontology molecular gunction database as GO:0003700, "transcription factor activity," or as GO:0003677, "DNA binding" and GO:0030528, "transcription regulator activity," or as GO:0003677 and GO:0045449, "regulation of transcription"), 668 transcriptional cofactors (a manually curated list, not overlapping with the transcription factor list, built upon genes annotated as GO:0003712, "transcription cofactor activity," or GO:0030528 or GO:0045449), 3455 signaling pathway related genes (annotated in GO biological process database as GO:0007165, "signal transduction" and in GO cellular component database as GO:0005622, "intracellular" or GO:0005886, "plasma membrane"), and 3620 surface markers (annotated as GO:0005886 or as GO:0009986, "cell surface"). ARACNe is only run on these gene sets so as to limit protein activity inference to proteins with biologically meaningful downstream regulatory targets, and we do not apply ARACNe to infer regulatory networks for proteins with no known signaling or transcriptional activity for which protein activity may be difficult to biologically interpret. Parameters were set to zero DPI (Data Processing Inequality) tolerance and MI (Mutual Information) p value threshold of 10^{-8}, computed by permuting the original dataset as a null model. Each gene list used to run ARACNe is available on github, along with the generated patient-by-patient ARACNe tables organized into CD45-positive and CD45-negative clusters.

Protein activity inference

Protein activity was inferred for CD45-positive cells from each patient by running the metaVIPER algorithm with all CD45-positive ARACNe networks across all patients on the SCTransform-scaled and Anchor-Integrated gene expression signature of single cells from each patient. Because the SCTransform-scaled gene expression signature is already normalized, VIPER normalization parameter was set to "none." The resulting patient-by-patient VIPER matrices were combined by sub-setting to the VIPER proteins for which activity was inferred in each patient sample, resulting in 2,562 proteins with successfully inferred activity across all CD45-positive patient samples. For CD45-negative single cells, protein activity was inferred by running the metaVIPER algorithm with all CD45-negative ARACNe networks across all patients in the same way, and then taking the intersection of 2,667 proteins with successfully inferred activity across all CD45-negative patient samples. VIPER-Inferred Protein Activity matrices were loaded into a Seurat Object with CreateSeuratObject, then projected into their first 50 principal components using the RunPCA function in Seurat, and further reduced into a 2-dimensional visualization space using the RunUMAP function with method umap-learn and Pearson correlation as the distance metric between cells. Differential Gene Expression between clusters identified by resolution-optimized Louvain was computed using bootstrapped t test, run with 100 bootstraps, and top proteins for each cluster were ranked by p value.

Copy number inference

Copy Number Alteration (CNA) across CD45-negative cells was inferred from gene expression counts at the single cell level using the InferCNA package. Cells were clustered according to their unsupervised clustering label by either gene expression or VIPER. At the first iteration of CNA inference, the entire set of CD45-positive cells was taken as a reference set to infer CNAs shown for each of the CD45-negative populations (Figure 3).

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

Flow cytometry data acquisition

From each of the 8 patient samples profiled by single-cell sequencing of both the CD45-positive and CD45-negative cells, an aliquot of roughly $2x10^6$ cells was taken for staining and high-throughput flow cytometry on CyTEK Aurora flow cytometer. Cells were stained for 10 minutes with Zombie NIR dye (1:1000 concentration), then stained with surface antibodies for 30 minutes on ice protected from light. After washing, cells stained with myeloid panel antibodies were run fresh on the cytometer. Cells stained with the lymphoid panel were fixed with the FoxP3 Fix/Perm kit (ThermoFisher) for at least 30 minutes, then stained with intracellular markers for 30 minutes on ice protected from light. All antibodies used can be found in Table S10. For both panels, single stain reference controls were created using UltraComp eBeads (ThermoFisher). Due to poor staining quality in Patient8, samples from this patient were excluded in downstream analysis of flow cytometry data. Data was evaluated by multi-dimensional analysis in R, and follow-up manual gating was performed as shown in Figure 2 using FlowJo v10.5.3.

Multi-dimensional analysis of flow cytometry data

Flow cytometry samples from all samples were combined and gated on Live CD45-positive, then the gated .fcs files for both lymphoid and myeloid antibody panel were separately exported from FlowJo software and analyzed in R with the flowCore and ggcyto packages. For each panel, raw fluorescence data were normalized with the estimateLogicle and transform functions, and in order to reduce computational burden of downstream analysis a sampled set of normalized fluorescence data from 250,000 cells were then loaded into a Seurat object with CreateSeuratObject. Two-dimensional representation of these data was computed by RunUMAP and resolution-optimized Louvain clustering was performed. Fluorescence of all markers was visualized in a heatmap, with cells grouped by cluster. For single-cell sequencing data, the CD45-positive gene expression matrix and VIPER-inferred protein activity matrix were each subset to genes corresponding to the proteins profiled by flow cytometry, and re-clustered by the resolution-optimized Louvain algorithm. Side by side comparison of the clustering and heatmaps for flow cytometry protein expression, gene expression, and inferred protein activity are shown in Figure 3.

Receptor-ligand interaction inference

A curated database of 2,557 known receptor-ligand interaction pairs was downloaded from the RIKEN FANTOM5 database. This list of receptor-ligand pairs was subset to pairs for which the ligand is significantly upregulated by gene expression in at least one VIPER cluster across patients and the receptor is significantly upregulated by protein activity in at least one VIPER cluster across patients. This reduced the total set of receptor-ligand pairs detected in our dataset to 276. For each pair we annotate a ligand cell type with highest median gene expression and a receptor cell type with highest median protein activity. Filtering to interactions involving the Tumor macrophage cluster and any T cell cluster returns 5 interaction pairs, filtering to interactions involving any Tumor cell clusters and any T cell cluster returns 5 interaction pairs, and filtering to interactions involving any Tumor cell and the Tumor macrophage cluster returns 13 interaction pairs. These are shown in Figure 5D.

Staining for multiplex immunohistochemistry

After consulting with a pathologist, patient FFPE tissue blocks with at least 50% tumor were chosen for sectioning on to Superfrost™ slides. Representative full section 4 μm slides of tissue specimens were stained for H&E and viewed by the pathologist to determine areas of tumor, stroma, regression, and immune infiltrates. Each patient's tissue specimen was then stained using Opal™ 7-color multiplex IHC kit, according to the manufacturer's protocol (Akoya Biosciences) with minor modifications. Briefly, the slides were baked at 60°C for approximately 2 hr before de-paraffinization and retrieval of antigen at pH 9. The slides were then blocked using 3% hydrogen peroxide (in 1X Tris Buffer with 0.05%Tween20), followed by an additional block using the antibody diluent, before staining with the primary antibodies, which include (in the order of staining) TREM2 (clone-D8I4C, Cell Signaling, cat# 91068S, 1:400, AR9), C1q (clone-C1QA/2956, AbCam, cat#ab268120, 1:100, AR6), CD3 (clone-LN10, Leica, cat#NCL-L-CD3-565,1:100, AR6), ApoE (clone – D17N, Cell Signaling, cat#13366S, 1:300, AR6), CA9 (polyclonal, AbCam, cat#ab15086, 1:1000, AR9) and CD68 (clone – KP1, BioGeneX, cat#AM416-5M, RTU, AR6) along with CD163(clone – 10D6, AbCam, cat#ab74604, ready-to-use (RTU), AR6). For each staining cycle, the slides were first incubated with primary antibody, followed by the secondary HRP-polymerization, and signal amplification using Tyramide conjugated to an Opal fluorophore and microwave treated in the AR6 or AR9 buffer as required by the next round of primary antibody staining. Single color controls for each fluorochrome and an unstained slide were processed in the same batch and used to create the library for spectral unmixing.

Multispectral imaging

For each patient specimen, slides were scanned using Vectra 3 (PerkinElmer), with nine representative areas chosen for multispectral imaging – (i) areas with 50% tumor and 50% stroma, (ii) areas with > 90% tumor, and (iii) one area with > 90% adjacent normal, wherever possible. These images were factored equally for each patient during analysis using InForm™ software (PerkinElmer). Single stained slides and unstained slides were used for building the spectral library and for unmixing, taking autofluorescence spectrum of patient tissue into account.

Please cite this article in press as: Obradovic et al., Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.038

Image analysis

Using the spectral library, the nine representative regions for each patient were spectrally unmixed before manual tissue segmentation using inForm software (Version 2.6, PerkinElmer). Tissue segmentation included highlighting examples of CA9+ renal tumor tissue, classifying the CA9- highly cellular regions as stroma; and the spatially distant CA9- tubular regions as adjacent normal (Figure S7). This trained the InForm algorithm to characterize each of the three tissue types and segment all the corresponding regions for each patient sample. Cellular components were then identified for each cell using the DAPI nuclear counter stain to define the nucleus; and CD3 and C1q stains to detect the associated membrane and cytoplasm, respectively. Using DAPI, we adjusted the nuclear splitting intensity to prevent incorrect identification and quantification of cells due to clumping. Individual cells were then phenotyped manually on the basis of their staining as Tumor cells (CA9+, yellow), macrophages (CD68+ or CD163+, orange), T cells (CD3+, white), ApoE+ (aqua), TREM2+ (magenta) and C1q+ (green). A training set was defined for the InForm algorithm of around 30 cells for each phenotype, from which we were able to distinguish between the cell densities across all tissue types within the nine fields for each patient. The cells were then scored for the staining intensities of each individual marker, including co-expression across the three tissue types – tumor, stroma and adjacent normal, and threshold fluorescence value of positive staining versus background was computed for each marker by the InForm software. The data from each field was compiled to summarize the position, phenotype and density of cells for each patient. Data were further analyzed in R version 3.6.1 using the phenoptr package, such that all fields for each patient sample were combined into a single data frame with cell-by-cell annotation of classified tissue context (tumor, tumor stroma, or adjacent normal), and fluorescence intensity of all markers.

Co-staining of C1Q, TREM2, and APOE with known macrophage markers was determined by generating contingency tables of C1Q, TREM2, or APOE positive cells with CD68/CD163 positive cells, and testing for statistical over-representation of C1Q/TREM2/APOE on macrophages by Fisher's Exact Test. Odds ratios of co-staining with CD68/CD163+ versus CD68/CD163- cells were computed across all 11 patients in the cohort profiled by single-cell RNASeq, shown as a boxplot in Figure 7B. Cell counts were computed and normalized in the tumor stromal and adjacent normal tissue contexts for each combination of C1Q+/TREM2+/APOE+ Macrophages, defined by positive staining for DAPI and CD68/CD163 and negative staining for CD3 and CA9. Frequencies of each cell population in tumor stroma versus adjacent normal tissue were compared by paired Wilcox test for the 7 patients in which regions of both tumor stroma and adjacent normal tissue were identified on the same stained tissue slices (Figures 7C and 7D).

In the separate validation cohort of 8 patients for which 4 experienced early post-surgical recurrence and 4 did not, frequency of each combination of C1Q+/TREM2+/APOE+ Macrophage cells was computed in the same way as described above, and frequencies of each population in tumor stroma of recurrent versus non-recurrent patients were compared by unpaired Wilcox test (Figure 7E). Since C1Q+ cells and C1Q+CD68/CD163+ cells were significantly enriched in tumor stroma of patients with early recurrence, fraction of cells staining for these markers was tested for association with time-to-recurrence. Threshold for defining high versus low fraction of cells positive for these markers was determined by maximization of the log-rank statistic, such that frequency of C1Q+ cells > 0.02 was determined to be high C1Q+ and frequency of C1Q+CD68/CD163+ cells > 0.01 was determined to be high C1Q+CD68/CD163+. Kaplan-Meier curve was plotted for each population, with statistical significance assessed by log-rank test.

Association of tumor macrophage signature with clinical recurrence

A protein signature for the Tumor-Specific Macrophage cluster was defined based on proteins differentially upregulated in the VIPER macrophage cluster (see Table S2 for gene and protein marker lists defining each VIPER cluster). In the dataset of FFPE samples profiled by bulkRNASeq that had been followed for time-to-recurrence after nephrectomy, outlier samples with low total read-counts were filtered out, and signature of remaining patients with recurrence (time-to-recurrence 8 months, 12 months, 12 months, and 82 months) versus patients without recurrence (observation period 35 months, 86 months, 110 months, and 113 months), was computed by z-score scaling of log10(TPM) normalized counts. Protein activity was computed from gene signature by VIPER using the CD45+ ARACNe networks inferred from single-cell data. Enrichment of the Tumor-Specific Macrophage protein marker set in the VIPER-transformed signature of recurrence versus no recurrence from bulkRNASeq was computed by Gene Set Enrichment Analysis (GSEA), with normalized enrichment score and p value determined by 1000 random permutations of gene labels. Activity of proteins in the leading edge of the enrichment was plotted sample-by-sample in a gene expression heatmap. Sample-by-Sample Normalized Enrichment Scores were also computed by ranking proteins in each sample according to decreasing activity. Cox regression of the raw normalized enrichment scores against time to disease recurrence was performed. Normalized enrichment scores for each sample were then binarized to less than zero (low) or greater than zero (high), and Kaplan-Meier curve showing association with time to recurrence was plotted along with the binarized log-rank p value.

Results were further validated by repeating the sample-by-sample gene set enrichment of VIPER macrophage markers in a larger cohort of 157 patients profiled by bulk-RNASeq, where enrichment of macrophage signature was associated with shorter time to post-surgical disease recurrence with log-rank p value of 0.0029. This analysis was performed using the ggsurvplot and survminer packages in R, and is shown in Figure 6. Validation of tumor-specific macrophage association with time-to-recurrence was also performed by immunohistochemical staining of FFPE tissue from the same 8 patients analyzed by RNA sequencing. Immunohistochemical staining and fluorescence thresholding was performed as described above, and proportion of C1Q, TREM2, APOE, and CD68/CD163 positive cells in the tumor stroma was compared in recurrent versus non-recurrent patient samples, along with the proportions of cell co-staining for every combination of those markers. Significance of the difference in frequency between recurrence and

non-recurrence samples was assessed by unpaired Wilcox test. Cell populations with significant difference in staining between the two groups were further assessed by log-rank regression against time-to-recurrence. Frequency threshold for high versus low level of staining was determined by maximizing the log-rank statistic, and Kaplan-Meier curve associating IHC staining with time-to-recurrence was generated, shown in Figure 7F.

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantitative and statistical analyses were performed using the R computational environment and packages described above. Differential gene expression was assessed at the single-cell level by the MAST single-cell statistical framework as implemented in Seurat v3 (Finak et al., 2015), and differential VIPER activity was assessed by t test, each with Benjamini-Hochberg multiple-testing correction. Comparisons of cell frequencies were performed by non-parametric Wilcox rank-sum test, and survival analyses were performed by log-rank test. In all cases, statistical significance was defined as an adjusted p value less than 0.05. Details of all statistical tests used can be found in the corresponding figure legends.

Supplemental figures

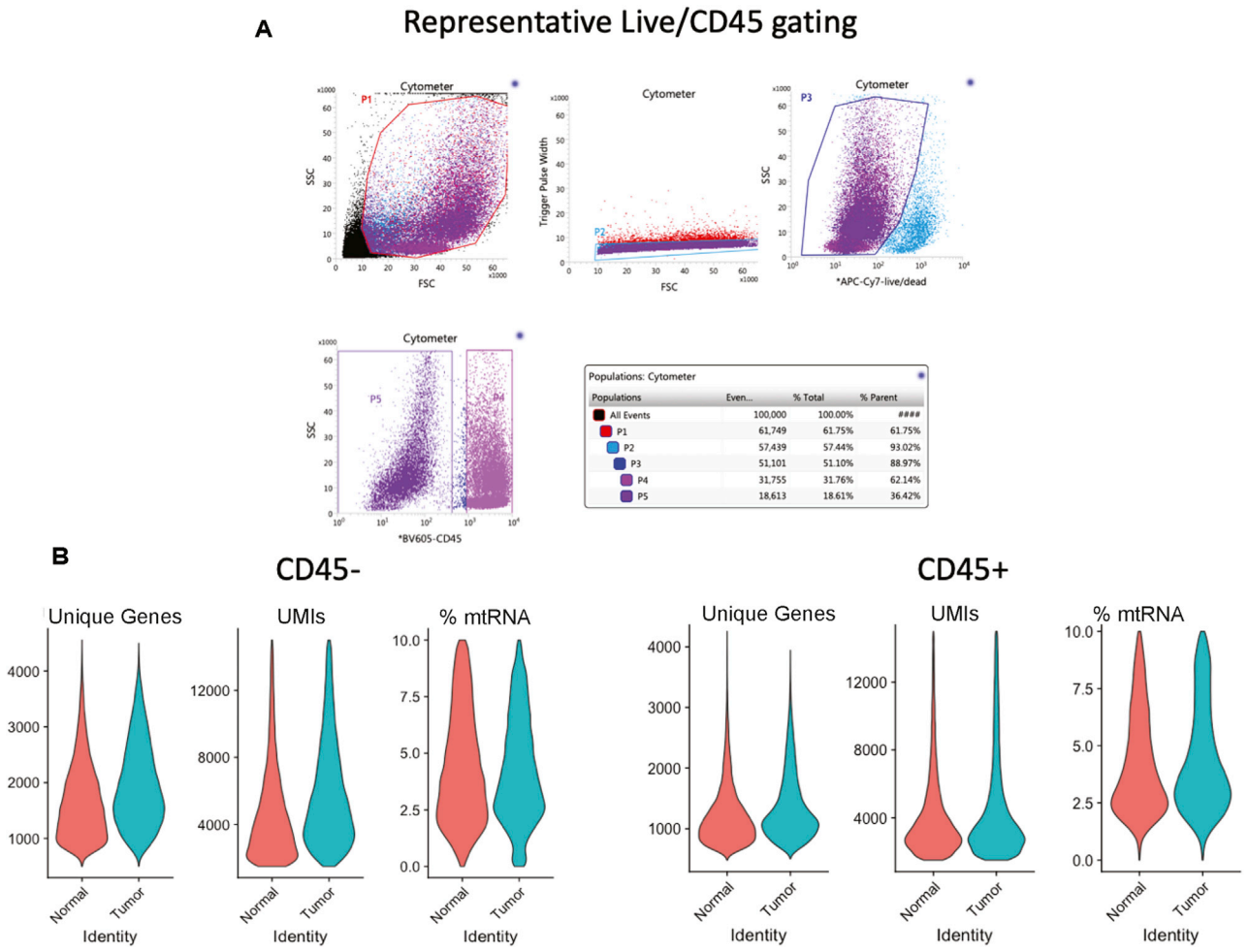


Figure S1. Gating and quality control, related to STAR Methods

A) Representative Live CD45+/− gating. B) distribution of combined % mitochondrial genes, number of UMIs/cell, and number of genes/cell for CD45+ and CD45- Tumor and Adjacent Normal samples, following quality-control filtering.

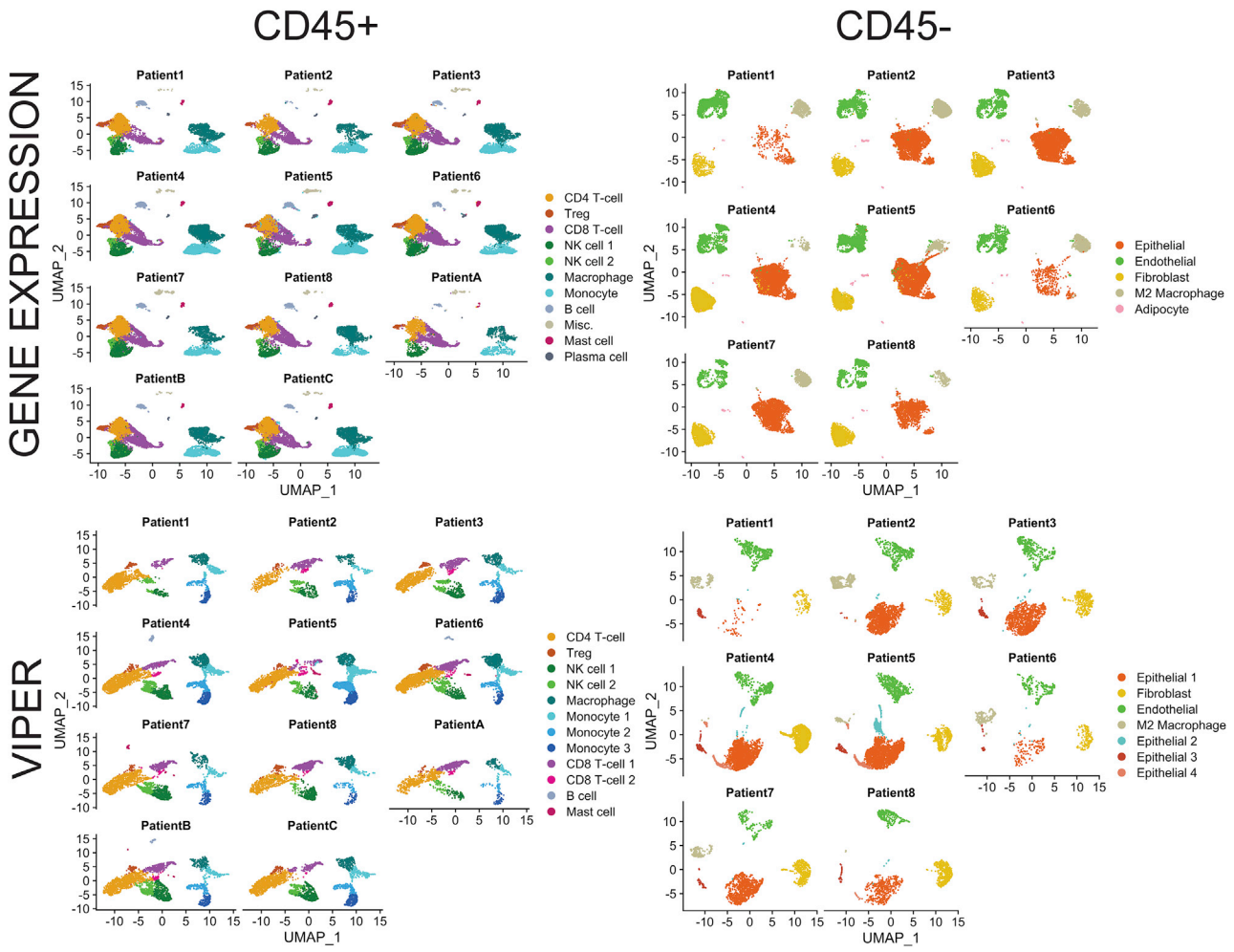


Figure S2. Consistency of cell-type clustering across patients, related to Figures 1 and 4

UMAP plots of global gene expression clustering and global VIPER clustering for both CD45+ and CD45- cells, split by individual patient identity, such that the overall cell types identified are consistent across patients, with minimal batch effect.

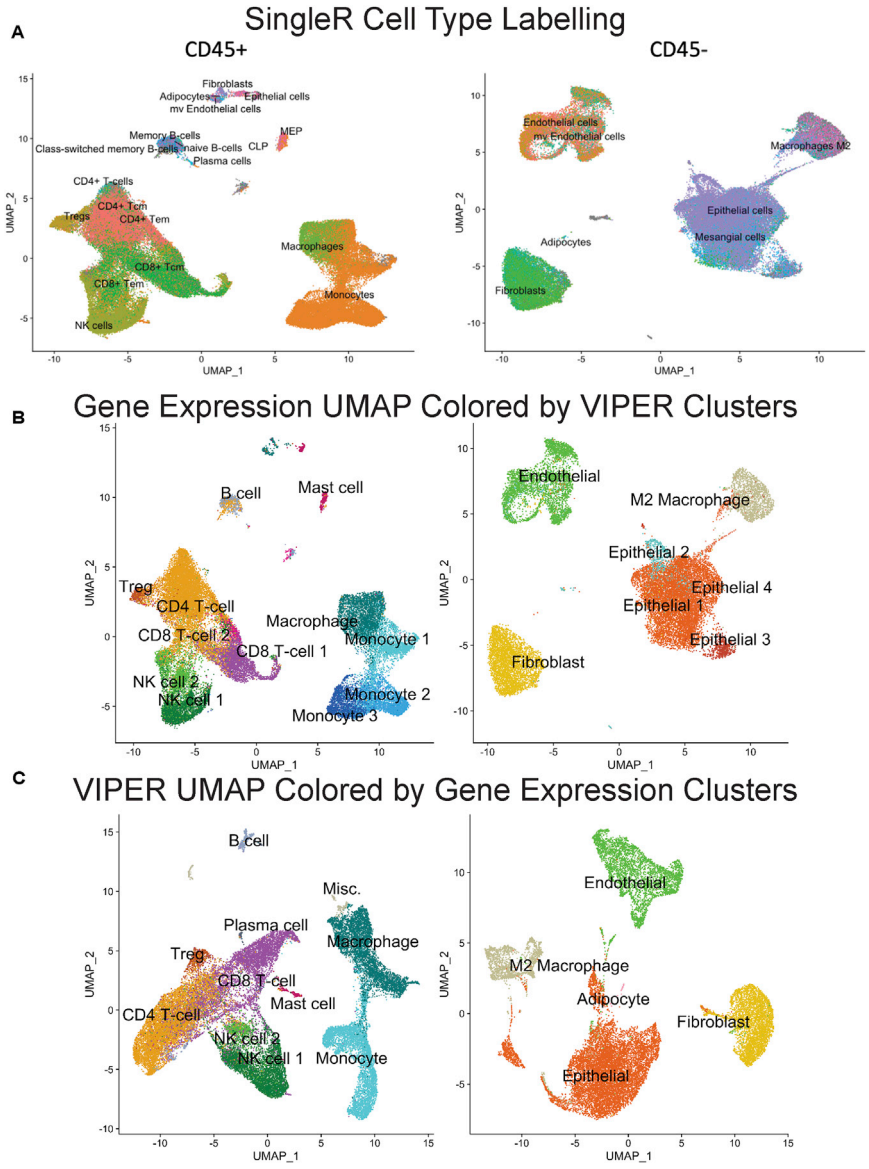


Figure S3. Visualization of cell-type labels, related to Figures 1 and 4

A) UMAP plots of global gene expression clustering for both CD45+ and CD45- cells overlaid with cell-by-cell SingleR cell type labels, excluding labels with p > 0.05. B) Gene Expression UMAP plot re-colored by VIPER clusters C) VIPER UMAP plot re-colored by Gene Expression clusters.

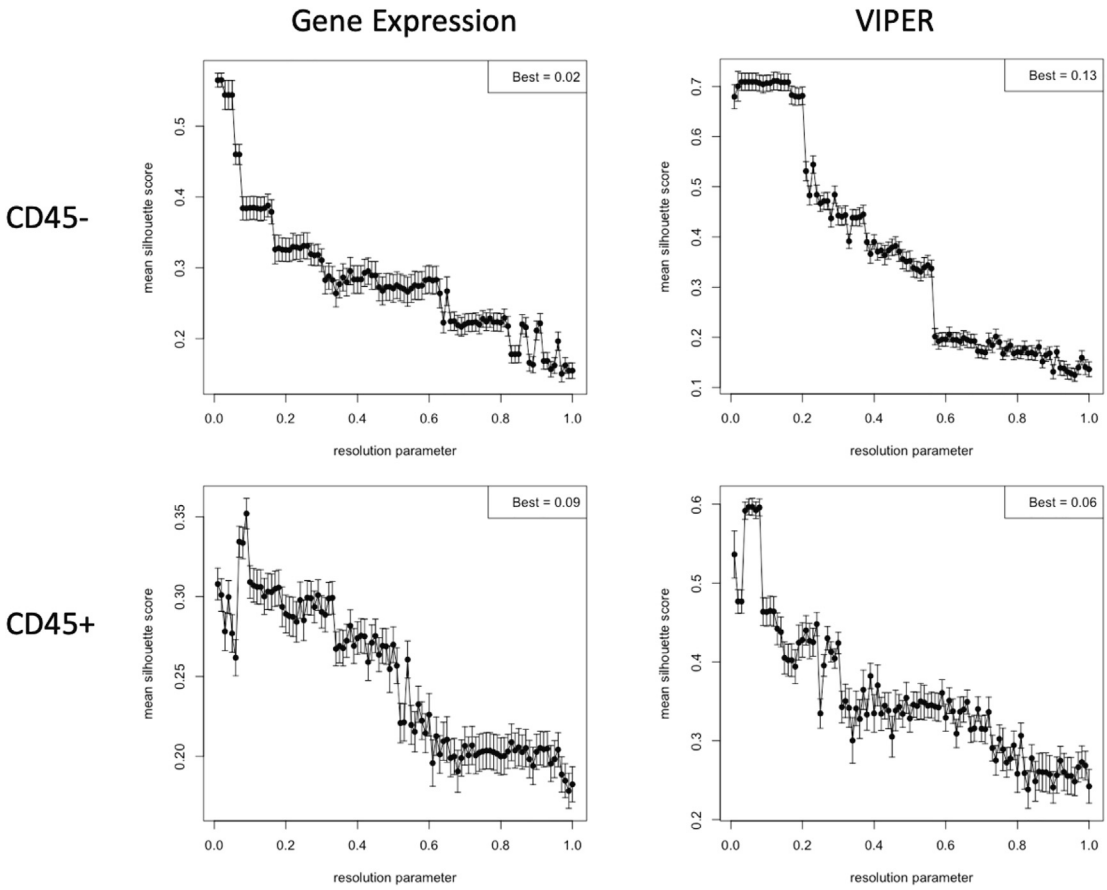
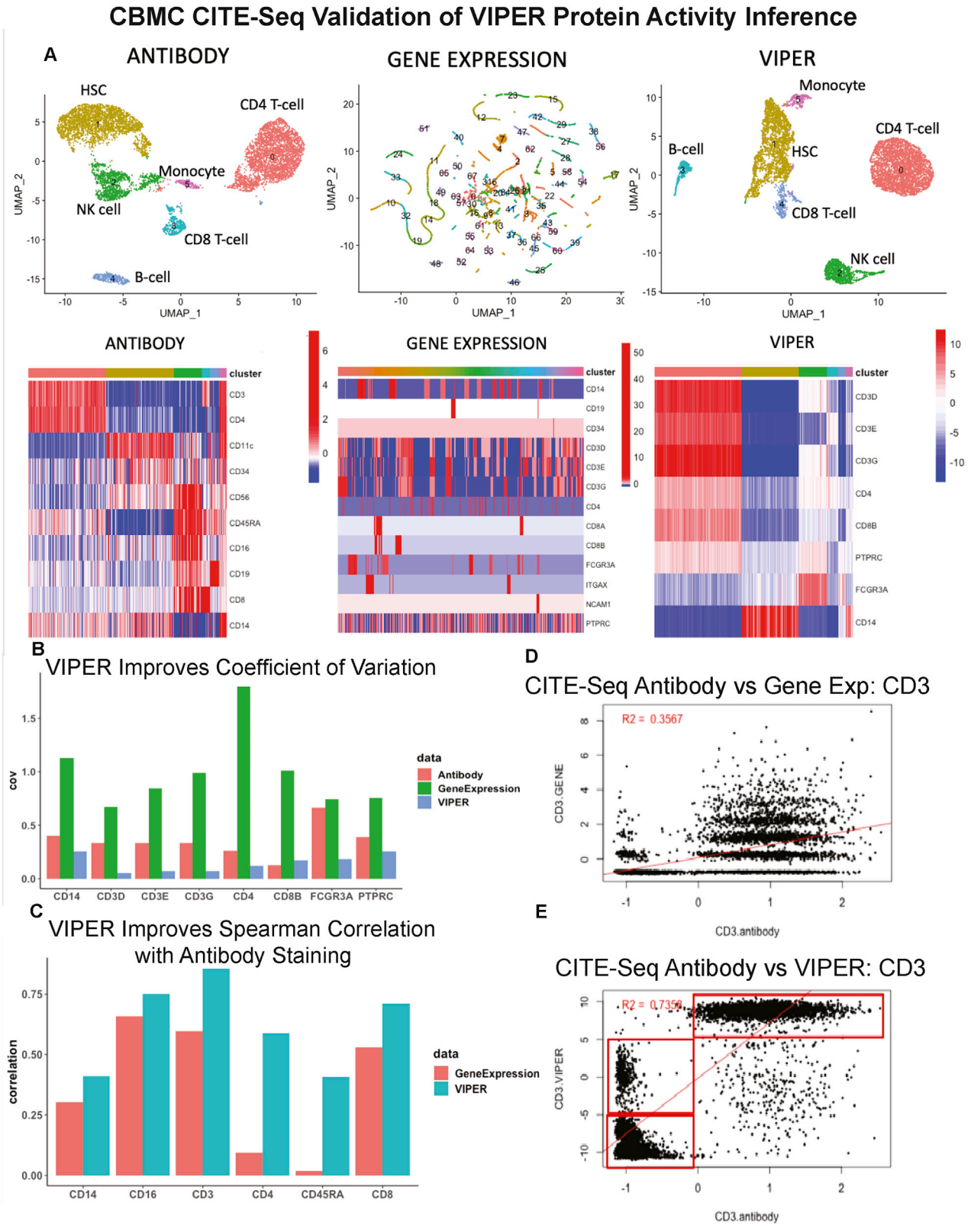


Figure S4. Resolution-optimized Louvain clustering silhouette scores, related to STAR Methods

Mean and Standard Deviation of Silhouette Score by Resolution-Optimized Louvain algorithm for each resolution value ranging along the x axis from 0 to 1.0 at intervals of 0.01, showing "best" resolution in the top-right as the resolution that maximizes mean silhouette score. Includes CD45+ Gene Expression clustering, CD45- Gene Expression clustering, CD45+ VIPER clustering, and CD45- VIPER clustering.



(legend on next page)

Figure S5. Validation of VIPER inference robustness on paired antibody and scRNA-seq data by CITE-seq, related to Figure 3

A) Additional validation of VIPER inference robustness on public CITE-Seq dataset concurrently profiling antibody staining on small set of proteins and gene expression in the same single-cells from cord blood mononuclear cells (CBMC). Analysis performed as in Figure 3, demonstrating recovery of cell types identified with antibody staining by VIPER despite data dropout by gene expression. B) Coefficient of Variation for each gene by antibody staining, gene expression, and VIPER, showing dramatic reduction by VIPER relative to gene expression for all markers, improving even on variability of antibody staining. C) pairwise Spearman correlation of each gene to antibody staining by gene expression or by VIPER, showing stronger correlation by VIPER to antibody staining for all genes. D) Representative pairwise scatterplot of gene expression and antibody staining intensity in the same cells, showing CD3E gene against CD3 protein stain, with weak positive Pearson correlation. E) Pairwise scatterplot of the same cells shown in D by VIPER activity versus antibody staining intensity, with significantly stronger Pearson correlation and visual segregation of distinct cell lineages outlined by red boxes.

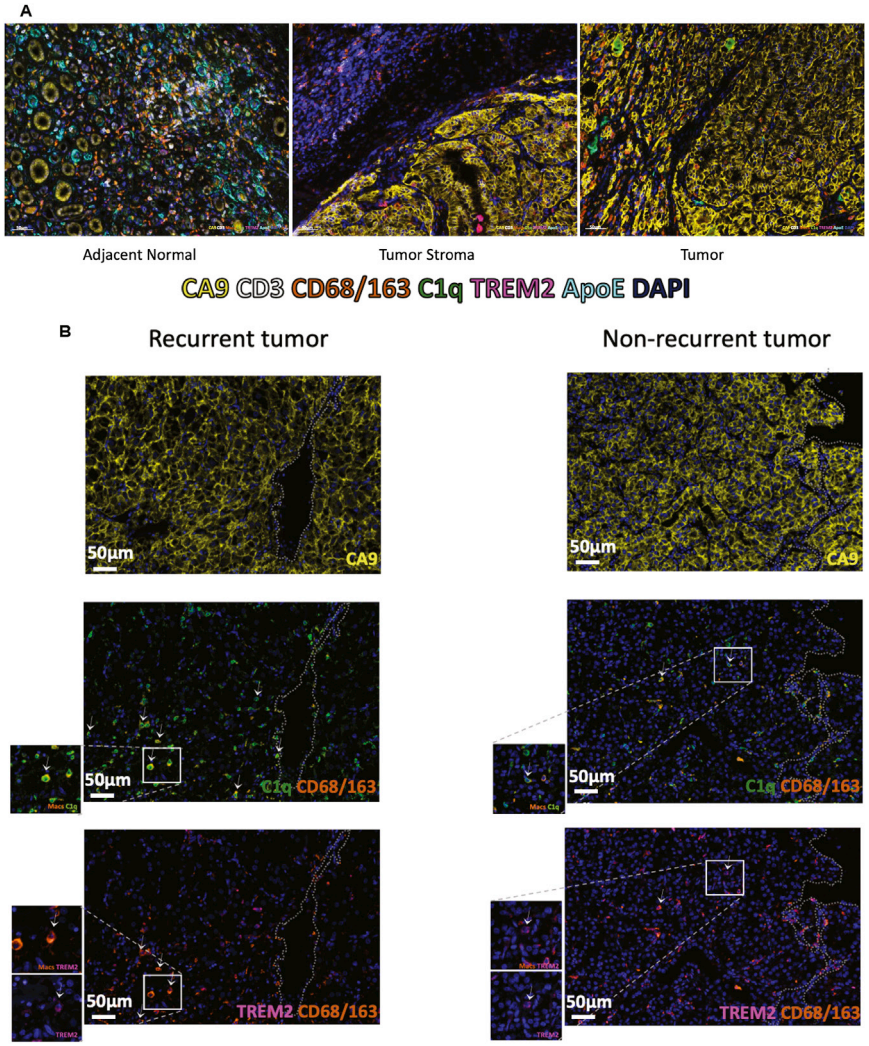


Figure S6. Visualization of C1Q^{+}/TREM2^{+} cells by immunohistochemistry, related to Figure 7

A) IHC plots showing typical morphology of tumor and adjacent non-tumor tissue B) C1Q/TREM2 co-staining with CD68/CD163 showing higher density in a representative recurrent patient (left) than a representative non-recurrent patient (right).

A

Tumor :: T cell Interactions

Ligand	Receptor	ligand_cluster	receptor_cluster
CCL5	ACKR4	CD8 T-cell 1	Epithelial 3
CCL5	SDC4	CD8 T-cell 1	Epithelial 1
CD70	CD27	Epithelial 1	CD8 T-cell 1
HSP90AA1	EGFR	CD8 T-cell 2	Epithelial 1
HSP90AA1	FGFR3	CD8 T-cell 2	Epithelial 4

Macrophage :: T cell Interactions

Ligand	Receptor	ligand_cluster	receptor_cluster
CCL5	CCR5	CD8 T-cell 1	Macrophage
CXCL3	CXCR1	Macrophage	CD8 T-cell 2
ICAM1	IL2RG	Macrophage	CD8 T-cell 1
LTB	CD40	Treg	Macrophage
LTB	LTBR	Treg	Macrophage

Macrophage :: Tumor Interactions

Ligand	Receptor	ligand_cluster	receptor_cluster
APOE	LRP5	Macrophage	Epithelial 4
APOE	SORL1	Macrophage	Epithelial 1
APOE	VLDLR	Macrophage	Epithelial 1
CYR61	ITGB5	Epithelial 1	Macrophage
ICAM1	EGFR	Macrophage	Epithelial 1
MDK	LRP1	Epithelial 1	Macrophage
MDK	SDC3	Epithelial 1	Macrophage
MDK	TSPAN1	Epithelial 1	Macrophage
RARRES2	CMKLR1	Epithelial 1	Macrophage
SERPINA1	LRP1	Epithelial 3	Macrophage
TNF	TNFRSF21	Macrophage	Epithelial 4
VEGFA	EPHB2	Epithelial 1	Macrophage
VEGFA	NRP2	Epithelial 1	Macrophage

B

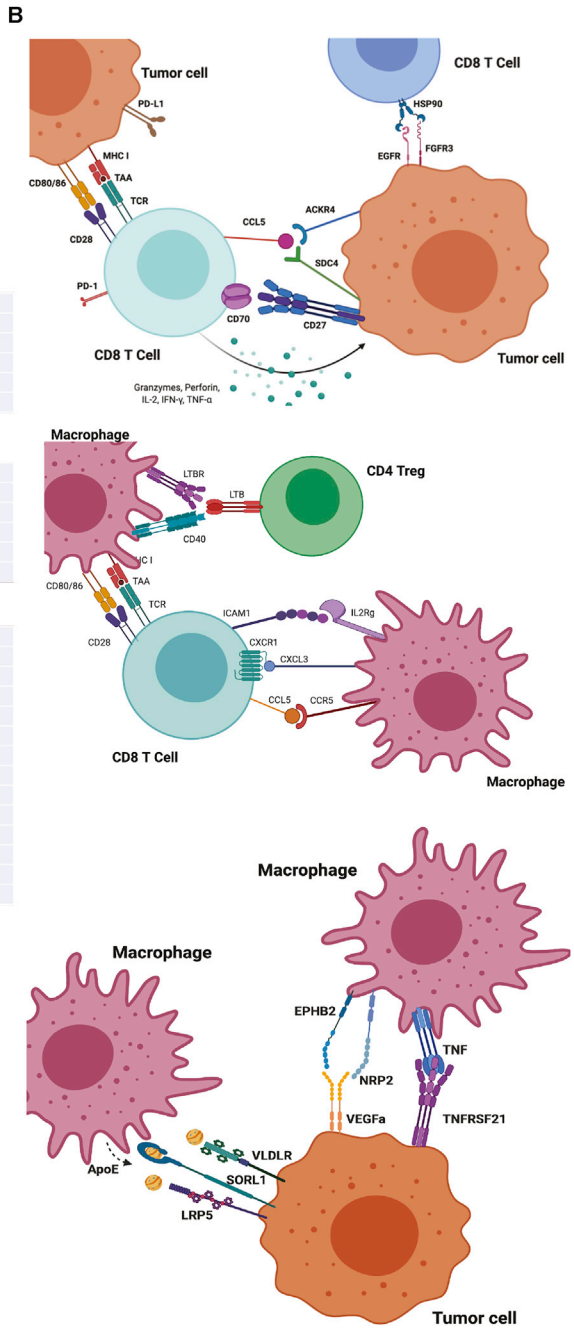


Figure S7. Receptor-ligand interactions are inferred between tumor cells, tumor macrophages, and T cells, related to Figure 5

A) Table of known receptor-ligand interaction pairs for which ligand is significantly upregulated by Gene Expression in one cluster and receptor is significantly upregulated by VIPER in another. Subset to interactions inferred to occur between Tumor Macrophages and T cells, between Tumor cells and T cells, or between Tumor Macrophages and Tumor cells. B) Visualization of the receptor-ligand interaction pairs shown in A.