JCI The Journal of Clinical Investigation

Failed immune responses across multiple pathologies share pan-tumor and circulating lymphocytic targets

Anne Monette, … , Igor Jurisica, Rejean Lapointe

J Clin Invest. 2019. https://doi.org/10.1172/JCI125301.

Research In-Press Preview **Immunology** **Oncology**

Rationale Tumor infiltrating lymphocytes are widely associated with positive outcomes, yet carry key indicators of a systemic failed immune response against unresolved cancer. Cancer immunotherapies can reverse their tolerance phenotypes, while preserving tumor-reactivity and neoantigen-specificity shared with circulating immune cells. Objectives We performed comprehensive transcriptomic analyses to identify gene signatures common to circulating and tumor infiltrating lymphocytes in the context of clear cell renal cell carcinoma. Modulated genes also associated with disease outcome were validated in other cancer types. Findings Using bioinformatics, we identified practical diagnostic markers and actionable targets of the failed immune response. On circulating lymphocytes, three genes, *LEF1*, *FASLG*, and *MMP9*, could efficiently stratify patients from healthy control donors. From their associations with resistance to cancer immunotherapies and microbial infections, we uncovered not only pan-cancer, but pan-pathology failed immune response profiles. A prominent lymphocytic matrix metallopeptidase cell migration pathway, is central to a panoply of diseases and tumor immunogenicity, correlates with multi-cancer recurrence, and identifies a feasible, non-invasive approach to pan-pathology diagnoses. Conclusions The non-invasive differently expressed genes we have identified warrant future investigation towards the development of their potential in precision diagnostics and precision pan-disease immunotherapeutics.

Failed immune responses across multiple pathologies share pan-tumor and circulating lymphocytic targets

Anne Monette[1,2,3,4], Antigoni Morou[1,3], Nadia A. Al-Banna[1,2,5,6], Louise Rousseau[1], Jean-Baptiste Lattouf[1,2,7], Sara Rahmati[8], Tomas Tokar[8], Jean-Pierre Routy[9], Jean-François Cailhier[1,2,3,10], Daniel E. Kaufmann[1,3], Igor Jurisica[8,11,12,13], Réjean Lapointe[1,2,3]

[1] University of Montreal Hospital Research Centre, Montreal, Quebec, Canada.
[2] Montreal Cancer Institute, Montreal, Quebec, Canada.
[3] Department of Medicine, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.
[4] Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada.
[5] Faculty of Medicine, McGill University, Montreal, Quebec, Canada.
[6] Department of Basic Medical Sciences, College of Medicine, QU Health Cluster, Qatar University, Doha, Qatar.
[7] Department of Surgery, University of Montreal, Montreal, Quebec, Canada.
[8] Krembil Research Institute, Toronto Western Hospital, Toronto, Ontario, Canada.
[9] Chronic Viral Illnesses Service and Division of Hematology, McGill University Health Centre, Montreal, Quebec, Canada.
[10] Nephrology Division, Department of Medicine, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada.
[11] Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada.
[12] Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.
[13] Institute of Neuroimmunology, Slovak Academy of Sciences, Slovak Republic.

Address correspondence to: Réjean Lapointe, Centre de recherche du CHUM (CRCHUM), 900 rue St-Denis, Tour Viger, Montreal, Quebec, Canada, H2X 0A9. Phone: 011.514.890.8000; Email: rejean.lapointe@umontreal.ca. Or to: Anne Monette, Lady Davis Institute for Medical Research, 3755 ch. Côte Ste-Catherine Road, Montreal, Quebec, Canada, H3T 1E2. Phone: 011.514.340-8222. Email: anne.monette@mail.mcgill.ca.

Abstract

Tumor infiltrating lymphocytes are widely associated with positive outcomes, yet carry key indicators of a systemic failed immune response against unresolved cancer. Cancer immunotherapies can reverse their tolerance phenotypes, while preserving tumor-reactivity and neoantigen-specificity shared with circulating immune cells. We performed comprehensive transcriptomic analyses to identify gene signatures common to circulating and tumor infiltrating lymphocytes in the context of clear cell renal cell carcinoma. Modulated genes also associated with disease outcome were validated in other cancer types. Through comprehensive bioinformatics analyses, we identified practical diagnostic markers and actionable targets of the failed immune response. On circulating lymphocytes, three genes, *LEF1*, *FASLG*, and *MMP9*, could efficiently stratify patients from healthy control donors. From their associations with resistance to cancer immunotherapies and microbial infections, we uncovered not only pan-cancer, but pan-pathology failed immune response profiles. A prominent lymphocytic matrix metallopeptidase cell migration pathway, is central to a panoply of diseases and tumor immunogenicity, correlates with multi-cancer recurrence, and identifies a feasible, non-invasive approach to pan-pathology diagnoses. The differently expressed genes we have identified warrant future investigation towards the development of their potential in non-invasive precision diagnostics and precision pan-disease immunotherapeutics.

Introduction

Pan-cancer studies have demonstrated that tumor infiltrating lymphocytes (TILs) are prognostic determinants of intratumoral heterogeneity (1). The failure of TILs at discriminating and eliminating neoantigen content from general tissue-specific, tumor-associated, or tumor-selective antigens, should reclassify the emergence of cancer as facet of the failed immune response (FIR). The one-size-fits-all potential for pan-cancer treatment by immune checkpoint blockade (ICB) is being investigated, but a significant fraction of patients do not respond to any single currently available therapies, while others progress or develop resistance. To identify TIL phenotypes contributing to responses, pan-cancer studies have correlated epithelial-to-mesenchymal transition and immune activation (2), where immune activation and $CD8^+$ TIL landscapes associate with favorable prognostic genes (3). Pre-treatment pan-cancer surveillance of interferon signaling and antigen presentation factors (4), or pan-cancer “immunophenoscore” (5), may become adopted by mainstream oncology. Indeed, comprehensive multi-cancer databases are vastly extending our knowledge of tumorigenesis by providing avenues for deciphering diagnostic pan-cancer signatures distinguishing tumor types, and having prognostic, predictive, and therapeutic potential (6, 7). Integrative pan-cancer analyses elucidate tumor lineage unique signatures (8), and trace metastatic lesions to tissues of origin (9). However, the use of whole tumor datasets as precise scoring determinants of immune inference requires complex deconvolution (10), complicated by influence of tumor expression programs on TILs (11).

Baseline-tumor specimens provide a wealth of prospective information (12). However, their use for routine prediction to response is challenged by specimen sizes,

and limits imposed by cancer heterogeneity, and invasiveness and delays from surgical procurement (13). Liquid biopsies have the advantage of being easily accessible (14), where circulating tumor cells and DNA (15, 16), and pan-cancer platelets (17) have been investigated for diagnostics. Yet, their non-specificities and inabilities of pinpointing the nature of primary tumors have delayed adoption of these methods (18). The traceability of circulating immune cells, that are themselves targets for immunotherapies, should make their shifting phenotypes superior predictive biomarkers. Pan-cancer diagnostic and predictive biomarkers from circulating effector peripheral blood lymphocytes (PBLs) could mirror neoantigen-specific FIR TIL phenotypes, providing a "peripheral immunoscore" correlating with tumor burden or response to therapies (19, 20).

We have overcome the limitations of tumor heterogeneity and biospecimen and biomarker accessibility, by identifying cell surface FIR pan-cancer diagnostic and actionable targets circulating PBLs. Clear cell renal cell carcinoma (ccRCC) was used as the primary model for biomarker discovery for several reasons: it features highest numbers of differentially expressed genes (DEGs) (21), observed PBL profiles correlate with TIL profiles (22), and its strong yet dysfunctional immunogenicity represents an enigma (23), despite RCC trials demonstrating better responses to ICB than other targeting therapies (24).

We performed comprehensive microarray analyses on paired patient (pt) $CD8^+$ T cells and $CD19^+$ B cells isolated from ccRCC tumors (TIL, TIL-B), normal tumor-adjacent tissue infiltrating immune cells (TIIC), peripheral blood mononuclear cells (PBMC) from patients (ptPBL), and from age-matched healthy control donors (cdPBL). DEGs from TILs and ptPBLs were compared to The Cancer Genome Atlas (TCGA) Kidney Renal Clear

Cell Carcinoma (KIRC) cohort, identifying DEGs associated with patient prognosis. Both cell surface DEGs, and DEGs having pre-existing targeting compounds, more amenable to antagonistic or agonistic design or drug repurposing, were retained. ccRCC prognostic, pan-cancer DEGs were identified in lung, breast, ovarian and gastric cancers. A scoring system was implemented to retain DEGs whose expression was restricted to lymphocytes, were supported by immunology and oncology literature, and had significant protein-protein interactions (PPI) and gene expression correlations. Top scoring DEGs were validated on a new independent RCC cohort, where a minimum set of three DEGs could stratify patients. A central DEG, matrix metallopeptidase 9 (MMP-9), could also stratify patients having pan-cancer recurrence. Effects of pan-cancer genes were investigated for splicing defects and used in pathway discovery. Our 'blinded' gene-discovery pipeline design is supported by identification of DEGs previously reported as biomarkers conferring resistance, and others slated as novel immunotherapeutic targets. The demonstration that pan-cancer FIR-DEGs are essential for controlling human immunodeficiency virus 1 (HIV-1) and other microbial infections, implicates these as pan-pathology immune biomarkers with diagnostic and therapeutic potential.

Results

Distinct cell surface coding DEG profiles from ccRCC CD8$^+$ and CD19$^+$ PBLs and TILs

To investigate pan-cancer immunity, we performed a comprehensive microarray analyses on matched case-control pairs of CD8$^+$ TILs and CD19$^+$ TIL-Bs from ccRCC tumors, CD8$^+$ and CD19$^+$ TIICs from normal tumor-adjacent tissues, and CD8$^+$ and CD19$^+$ PBLs from ccRCC patients; along with CD8$^+$ and CD19$^+$ PBLs from matched healthy control donors (**Supplemental Figure S1**). Study patient clinicopathologic characteristics are presented in **Supplemental Table S1**. Quality control experiments for yield and quality of various rapidly isolated immune cell subsets from tumors were performed (**Supplemental Figure S2A-D**), in addition to stringent bulk total RNA quality testing prior to its amplification and application to comprehensive microarrays interrogating >67,000 transcripts (**Supplemental Figure S2E**). The Affymetrix Transcriptome Analysis Console (TAC) was used to observe prominent DEGs in TILs and ptPBLs relative to TIICs and cdPBLs (**Supplemental Figure S2F**), totaling 7,300 (i.e., CD8$^+$ and CD19$^+$ TIL-B/TIIC and ptPBL/cdPBL; 1.5 fold-change; $P < 0.05$) (**Supplemental Figure S1**). Principal component analyses (PCA) were generated using the Partek Genomics Suite for all paired CD8$^+$ or CD19$^+$ biospecimens and PBL controls (**Figure 1A, B**). Venn diagrams were generated to demonstrate overlaps in DEGs represented by CD8$^+$ and CD19$^+$ ptPBLs (20.4%) and TILs (37.8%) (**Figure 1C**), and to demonstrate overlaps of possible splice junctions generating spliceoforms common to CD8$^+$ ptPBLs and TILs, made possible by using comprehensive HTA 2.0 microarrays (**Figure 1D**), and suggesting that patient-inherent post-translational modification

106 programs generating distinct RNA isoforms may also influence the behavior of TILs. To
107 assess feasibility of pursuing DEGs more easily amenable to therapeutic interventions
108 such as ICB (i.e., actionable targets), we used unsupervised clustering and PCA to
109 examine DEGs coding for molecules expressed on the plasma membranes (PM); these
110 efficiently separating immune isolates – with largest differences maintained between TILs
111 and TIICs (**Figure 1E**) – and also permitting efficient separation of ptPBLs and cdPBLs
112 (**Figure 1E, F**).
113

114 ***Prognostic ccRCC DEGs have pan-cancer prognostic potential***

115 To identify prognostically important ccRCC DEGs, we generated Kaplan-Meier
116 plots and P values for the 7,300 significant DEGs using TCGA KIRC RNA-seq and
117 associated clinical datasets (n = 534 tumor, n = 72 healthy control donors). This step
118 resulted in detecting 2,257 prognostic DEGs (**Supplemental Figure S1**). To further refine
119 prognostic DEGs and find the most feasible actionable targets, we focused on PM-
120 associated proteins, or those having known targeting compounds. Partek and PANTHER
121 Gene Ontology were both used to identify PM proteins, ensuring most PM-associated
122 DEGs would be retained. ChEMBL target searches were used to identify proteins with
123 known targeting compounds. Together, these two approaches reduced target DEGs to
124 779, which were then investigated for their pan-cancer potential using >11,500 lung,
125 breast, gastric, and ovarian cancer patients from Kaplan Meier-plotter, generating 467
126 (i.e., 62%) target DEGs with pan-cancer potential. This refined list represented pan-
127 cancer FIR biomarkers, grouped as either: 1) agonistic targets decreased in tumors
128 relative to normal tissues and having a positive prognosis, or 2) antagonistic targets

increased in tumors relative to normal tissues and having negative prognosis (**Supplemental Figure S1**). PCA transposition permitted visualization of how these pan-cancer FIR-DEGs identified from ptPBLs (**Figure 2A, Supplemental Figure S3A**) or TILs (**Figure 2B, Supplemental Figure S3B**) were distributed across the five cancers, and how they correlated with each other; where numerous were common to both $CD8^+$ TILs and $CD19^+$ TIL-Bs relative to their TIIC counterparts (**Figure 2C, full gene list in Supplemental Figure S3C**). A subset of pan-cancer FIR-DEGs were also found to be common between TILs and ptPBLs (**Supplemental Figure S3D**).

Correlograms reflected increased correlations between the five cancers used to refine for pan-cancer target FIR-DEGs (**compare Supplemental Figure S3E and Figure 2D**). Because the selected 467 pan-can FIR-DEGs were discovered using whole tumor TCGA datasets, we compared percentages of correlations between five cancers to that of their immune infiltrates (n > 11,500), providing similar trends, suggesting strong likelihood that global FIR-DEG signatures were immune-based (**Figure 2E**). Of these 467 pan-cancer FIR-DEGs, proportions of agonistic and agonistic targets derived from ptPBLs were equal, whereas those derived from TILs had increased agonistic target representation (**Figure 2F**).

To further refine pan-cancer FIR-DEG targets, nominal derivatives (binomial values) were generated to integrate quantitative and non-quantitative, and thus non-harmonizable datasets and analyses, and were used get an overall score representing their 1) coupled expression and effect on overall survival (n = 5 cancers, Kaplan Meier-plotter), 2) coupled RNA and protein expression in myeloid and lymphoid cells relative to 12 other cancers, 3) modified in expression levels in cancers relative to normal tissues (n

= 17 cancers; The Human Protein Atlas), and 4) direct published literary evidence of DEG expression in the immune subtypes from which they were identified (**Supplemental Table S2**). The top 200 scoring pan-cancer ptPBL and TIL FIR-DEGs were subjected to PPI analyses using rudimentary search engine (STRING) (**Supplemental Figure S3F**) (25), providing a PPI enrichment value (*P* = 1.85e-10) warranting further investigation. For more comprehensive PPI analyses, we used IID, pathDIP and NAViGaTOR, providing new evidence of interactions (**Figure 3**), with most highly associated pathways to antagonistic targets including, immune system, TNF signaling, APC complex inactivation, NFkB, and agonistic pathways WNT signaling, chemokine signaling, proteoglycans, and GPCRs; (*P* < 1e-10) (**Supplemental File S1A**). Finally, the top scoring 200 ptPBL and TIL pan-cancer FIR-DEGs were further refined by retaining those that were the most correlated in differential pan-cancer gene expression, towards discovery of novel mechanistic pathways not deciphered from above analyses (**Figure 4**). The combination of these scoring methods was used to select pan-cancer FIR-DEG for validation on a new RCC patient cohort (**Supplemental Figure S1**).

Pan-cancer and polarizing DEGs stratify CD8$^+$, CD19$^+$, PBLs, TILs and TIICs

Twenty eight pan-cancer FIR-DEGs, and 62 commonly used T cell polarizing genes defining known T cell subsets, were selected for validation on a new, independent 74 patient RCC cohort, using TaqMan Gene Expression Assays on 96.96 microfluidic BioMark HD Real-Time PCR system dynamic arrays (Fluidigm), providing the advantage of DEG co-expression analysis. Total CD8$^+$ ptPBL RNA from 41 ccRCC, eight RCC, and six pRCC patients, and CD8$^+$ cdPBL RNA from control donors were analyzed, with three

ccRCC patient duplicates added as inter-assay RNA extraction controls. We also included five total ptPBMC and five total ndPBMC RNA preparations. Finally, to maximize use of the microfluidics chip, and to determine if these could provide a baseline for DEG expression, we also added pooled total RNA samples from: $CD8^+$ (*n* = 50 patients) and $CD19^+$ (*n* = 50 patients) ptPBLs, $CD8^+$ (*n* = 15 patients) and $CD19^+$ (*n* = 15 patients) cdPBLs, ccRCC PBMCs (*n* = 10 patients), pRCC PBMCs (*n* = 10 patients), ndPBMCs (*n* = 10 control donors), and paired ccRCC TILs (*n* = 8) and TIICs (*n* = 8). BioMark HD-generated heatmaps, housekeeping genes, and loading controls are shown in **Supplemental Figure S4A-D**.

Following normalization, correlograms were used to visualize co-expression dynamics between all DEGs (**Supplemental Figure S4E**). Unsupervised clustering demonstrated that pooled RNA fractions were stratified as expected: with $CD8^+$ and $CD19^+$ isolates stratifying furthest apart, and total PBMC isolates stratifying independently, but remaining closer to $CD8^+$, as a function of T cells (7-24%) representing a larger frequency of total PBMC than B cells (1-7%) (**Figure 5A**). Also expected, TILs stratified closest to total PBMC, yet remained close to TIICs – reflecting tissue infiltrating immune profiles. Finally, ccRCC ptPBLs and cdPBLs from either $CD19^+$ or $CD8^+$ isolates clustered closely, at opposite ends of the heat map. Unsupervised clustering was also used to observe that individual ccRCC ptPBMCs were efficiently stratified from ndPBMCs (**Figure 5B**). PCA was used to visualize coupling of pooled RNA fractions and DEG co-expression, here demonstrating that patient TILs, PBMCs and $CD8^+$ ptPBLs were distantly stratified from both TIICs and $CD19^+$ ptPBLs (**Figure 5C**). This 3-dimensional

197 view also provided evidence of co-expressing groups of pan-cancer FIR-DEGs and
198 polarizing genes.
199
200 ***Pan-cancer DEGs stratify RCC patients from control donors***
201 Differential expression and correlation analyses were coupled to identify pan-
202 cancer FIR-DEG combinations most efficiently stratifying patients. Several pan-cancer
203 FIR-DEGs (*ICOS*, *PF4V1*, *IFNG*, *LAG3*, *TIGIT*, *CDA*, *PDK4*, *KLF4*, *PIM2*, *TIMP1*,
204 *IGF2BP3*, *IL23A*, *LEF1*, *TCF7*), in combination with other T cell genes, efficiently stratified
205 patients from control donors to an accuracy of 90.1% (**Figure 6A, D**). Absence of novel
206 discovered pan-cancer FIR-DEGs uncommon to T cell polarization caused loss of patient
207 stratification (**Supplemental Figure S5A**); however, control donors still stratified with a
208 *LEF1* and *NT5E* expressing population, which included other biomarkers of activation,
209 and immune checkpoint *BTLA*, which we and others believe marks T cells having
210 enhanced survival properties (26).
211 Combination testing revealed that a smaller set of these patient stratifying pan-
212 cancer genes (*IFNG, CDA, PDK4, KLF4, IGF2BP3, LEF1),* could also stratify patients to
213 an accuracy of 89.1% (**Figure 6B, D**), which could not be met in their absence
214 (**Supplemental Figure S5B-C**). Additional combination testing identified a minimal set of
215 three DEGs (*MMP9, FASLG, LEF1*) stratifying patients with a 79.3% (**Figure 6C, D**).
216 Interestingly, aside from stratifying patients from control donors, pan-cancer FIR-DEG
217 PCAs revealed two dominant $CD8^+$ ptPBL populations containing either *FASLG* or *LEF1*
218 together responsible for triggering cell death or cell activation. In addition, the three
219 internal patient duplicates remained closely clustered throughout PCAs, whereas pooled

RNA factions were centralized among their counterparts (**Figure 6A-C**). Further correlation analyses performed on RCC patients populating yellow PCA quadrants occupied by control donors demonstrated these to have increased *CXCR3* (*P*=0.0021; *r*=0.4898; CI=0.1874-0.7074) and *CXCR5* (*P*=0.0029; *r*=0.4764; CI=0.1705-0.6988) (Spearman method), suggesting that these may be two key RCC fitness genes, also recently linked to increased abilities of broadly neutralizing antibody production by HIV-1 elite controllers (27).

Pan-cancer FIR-DEGs common to RCC and HIV

This link between RCC ptPBL DEGs and HIV-1 controllers prompted our examination of other pan-cancer FIR-DEGs commonly expressed by HIV-1 controllers. Intriguingly, the majority of our validated pan-cancer FIR-DEGs were represented in an HIV-1 elite DEG screen (28). As such, we searched the literature to elucidate which of these DEGs were useful to both cancer and HIV-1 when expressed by PBLs, demonstrating that 60% of these similarly polarized T cells towards permissiveness to cancer development and HIV-1 infection (**Supplemental Table S5**). In comparing HIV-1 controller DEGs similarities to ccRCC DEGs, we observe that the pan-cancer DEG prioritization pipeline increased identity to HIV-1 controller DEGs from 17% (467 pan-can DEGs) to 50% (top 100 pan-can DEGs) (**Figure 6E**). This led us to consider whether the pan-cancer FIR-DEG pipeline was actually identifying pan-pathology genes. We thus compared our pan-cancer DEGs to datasets from another study aimed at identifying frontline biomarkers common to numerous pathologies (29). Strikingly, 82.1% of our ptPBL-based, and 42.8% of our TIL-based top 200 pan-can FIR-DEGs were confirmed

by their findings, with 51% of 467 pan-can DEGs, and 59% of top 100 pan-can DEGs present (**Figure 6F**), and where a total of 71.1% of DEGs were commonly reflected by bacterial infection datasets. Potentially revealing pan-pathology T cell biomarkers, we then compared our lists to cancer patient datasets of response to anti-PD-1 immunotherapy (30, 31), highlighting a few of our pan-cancer FIR-DEGs (**Supplemental Table S6**), notably including our *MMP9, FASLG,* and *LEF-1* minimal triad stratifying ccRCC patients from control donors (**Figure 6C**; and see **Supplemental Table S7** for summary of validated DEGs common to other datasets).

Pan-cancer DEGs are associated with pan-cancer recurrence and T cell activation

Within the validation cohort, 10 of 28 ccRCC patients (35.7%) were recorded as having been previously treated for other malignancies including: kidney, bladder, blood, breast, colon, liver, melanoma, ovary, prostate, rectal, and uterine cancers, where a few had suffered from three different malignancies with no recorded metastases. We used this opportunity to compare validated DEGs across control donors, RCC patients, and those positive or negative for pan-cancer recurrence. Strikingly, *MMP9* expression best stratified pan-cancer recurring patients ($P<0.0001$, t = 1.779, df=26 T-test; $P=0.007$, MDiff=1.716, 95% CI of diff=0.3989 to 3.034, df=2366, 2-way ANOVA, Tukey's post-test) (**Figure 7A**). Indeed, all *MMP9*^{hi} patients had previously suffered from RCC, along with blood, breast, colon, melanoma, ovarian, prostate, or uterine cancers; with a high proportion of adenocarcinomas (80.0%). A disproportionate number of *MMP9*^{low} patients had previously suffered from a bladder or prostate cancer (66.6%). Other DEGs stratifying

pan-cancer recurring patients were *KLF4*, *RORC*, *PDK4*, and *CCR4*; yet these genes were decreased in these recurring patients.

From PCA analyses demonstrating that pan-cancer DEGs stratified two $CD8^+$ T cell pools in addition to individual patients (**Figure 6**), along with observations of their ability to stratify patients according to pan-cancer recurrence (**Figure 7A**), we applied correlograms to observe whether their combined expression could tip the balance between tolerant/anergic and activated/effector T cell profiles. The merging of correlograms providing a split in DEGs populations, expression levels of DEGs between isolates, and balance of DEGs formerly documented in the literature as being associated to activation or tolerance phenotypes, suggested existence of a dominant activated effector $CD8^+$ ptPBL population (**Figure 7B**). The majority of these effector DEGs (69.2%) providing an activation phenotype were downregulated in patients with pan-cancer recurrence, suggesting these patients may lack the ability to mount an immune response.

Pan-cancer DEGs synergize towards prognosis and are subject to splicing defects

The TCGA KIRC-probing prognostic algorithm was modified to test all combinations of additive effects of pan-cancer FIR-DEGs on patient prognosis. The only DEGs with marked additive effects on patient survival were *MMP9*, *LEF1*, *PF4V1*, and *TIMP-1* – this observation gaining additional support from correlograms providing evidence of their co-expression (**Figure 7C**). Additionally, relative to the three-DEG signature stratifying patients (**Figure 6C**), while *FASLG* did not itself associate with prognosis (P = 0.401), $MMP9^{hi}LEF1^{hi}FASLG^{hi}$ KIRC patients had reduced survival (HR =

289 0.0988-0.6324; *P* = 3.71e-05). While the TCGA KIRC dataset represents whole tumor
290 RNA expression, The Human Protein Atlas showed that unlike the others having additive
291 effects on prognosis, expression of *MMP9* RNA and protein are strictly associated with
292 lymphoid and myeloid systems; thus, possibly enhancing prognostic effects by identifying
293 immune relevant signature populations from whole tumor datasets. Inverse correlation
294 observed between oncostatic melatonin receptor 1A (MTNR1A), extensively expressed
295 by lymphocytes, and MMP-9 expression in RCC as a plausible mechanism for our
296 findings (32), prompted us to re-examine microarray datasets to see that *MTNR1A* was
297 indeed reduced in ccRCC $CD8^+$ TILs (*P* = 8.3 e-04) and $CD19^+$ TIL-Bs (*P* = 1.4 e-04).
298 To gain insight on other possible mechanisms behind effects of pan-cancer FIR-
299 DEGs on patients, and because we used the Human Transcriptome Array (HTA) 2.0
300 microarray able to distinguish between differential gene expression and transcript isoform
301 modulation, we used the microarray dataset to observe whether these differed at the
302 isoform level. In paired patient $CD8^+$ ptPBLs and TILs, with exception of *HIST1H2BG*,
303 *ICOS* and *IFNG*, all other validated pan-cancer FIR-DEGs had modified isoforms, where
304 47.36% of these were found to be mirrored between $CD8^+$ ptPBLs and TILs relative to
305 TIICs (**Figure 7D**) (**Supplemental Table S8**). Additionally, as determined by Affymetrix
306 TAC software, there were many more distinct transcript isoforms present and heightened
307 splicing indices for ptPBLs than for TILs, relative to TIICs (i.e., ptPBLs vs. TIICs, 71.97%,
308 avg. splicing index = 18.432, avg. splicing event score = 0.224; TILs vs. TIICs, 28.57%,
309 avg. splicing index = 1.727, avg. splicing event score = 0.306). Thus, the transcript isoform
310 repertoire of $CD8^+$ ptPBLs is much larger than that of $CD8^+$ TIL, likely due to similarities
311 for tissue-infiltrates; but with a few notable differences including higher isoform numbers

for immune checkpoints *TIGIT* and *LAG3*. Both *MMP9* and *TCF7* common isoforms were further increased in ptPBLs relative to TILs, and both *CD69* and *IQGAP1* common isoforms were modified in ptPBLs relative to cdPBLs. For *MMP9, TIMP1, IQGAP1, MPHOSPH8, CD69, TCF7, LAG3*, and *TIGIT*, the same isoforms are repeatedly represented among isolate types (i.e., $CD8^+$ ptPBLs and TILs, relative to $CD8^+$ TIICs and cdPBLs) (**Supplemental Table S8**). Together, these results suggest that prognostic effects of pan-cancer FIR-DEGs may also be the result of deficiencies in transcript isoforms required for optimal T cell fitness.

Enrichment of pan-cancer disrupted MMP9-pathways in ccRCC ptPBLs

Our initial strategy to use PPI analyses for refinement of ccRCC DEGs for validation was only partially useful. Now armed with validation experiments and strength in statistics for individual DEGs, by repeating PPI analysis using rudimentary search engine (String), the importance of *MMP9* having highest combined interaction annotation score (14.91), and its positioning as a central interacting node of pan-cancer FIR-DEGs (*TIMP1, PDK4, LEF1, CDA, KLF4, PF4V1, SELF, PIM2, ICOS, IFNG, IL23A, IL6ST, TCF7, SELL, SERPINE1, OSM, CXCL5, HBA1, COLA1, MAB2, LIFR, IQGAP1, MAPK8, PIK3CA, BLC2, LAG3,* and *TIGIT*), with associated cytokine production, and immune cell migration and adhesion cellular processes, held more weight (**Supplemental Figure S6**).

MMP9 in $CD8^+$ PBLs was one of three DEGs able to stratify RCC patients from control donors, and *MMP9* was increased in pan-cancer recurring RCC patients. We thus used advanced PPI and pathway analyses (IID and pathDIP, using NAViGaTOR) to re-examine the microarray datasets with aim to decipher significant role of MMP-9 in

signaling cascades at play in ccRCC patients. A comprehensive pathway enrichment analysis using all 1,036 non-redundant ptPBL DEGs, identified pathways including amyloid fiber formation, platelet activation, SIRT, HDACs, leukocyte transendothelial migration, alcoholism, SUMOylation, androgen receptor, and TNFalpha (P < e-10) all had links to *MMP9* regulation (**Supplemental File S1B**). To identify the most relevant MMP-9 pathways of the 235 revealed by pathDIP in ccRCC ptPBLs, we performed correlation analyses revealing that 216 of the 1,036 DEGs were significantly correlated with MMP-9-positive pathways (**Supplemental Table S9**). Generating physical PPI networks using NAViGaTOR demonstrated that all but 6 of these 216 DEGs (97.71%) do interact (**Figure 8A, see Supplemental Figure S7 for full PPI**). From pathway enrichment analysis using pathDIP, many disease, cancer, and immunity pathways could be repeatedly observed in MMP-9 significant DEG-associated pathways (**Figure 8B**, see **Supplemental Figure S8** for full analyses). Tissue-specific disrupted PPI networks among MMP-9 interactors in 13 cancers were examined, whereby the majority of identified genes represented in cell/leukocyte migration and adhesion processes, and extracellular matrix disassembly and collagen metabolism (**Supplemental Figure S8**), as recently reported as representing pre-treatment serum biomarkers in response to ICB (33), and genes common to ccRCC ptPBLs are involved in immune response and activation, regulation of apoptosis, and migration in response to bacteria (34). Interestingly, cancers having the highest MMP-9 gained and lost PPIs were colon, mouth, and lung (**Figure 8C**). Finally, an independent differential correlation analysis and organization of MMP-9 pathways and their significantly associated DEGs, was used to validate that though extracted from ccRCC ptPBLs expression signatures, the majority of MMP-9 pathways filtered on ccRCC

DEGs were most linked to a variety of renal diseases, numerous viral, bacterial and parasitic infections, numerous cancers, and immunity and antigen recognition and activation; differentiation, and cellular survival pathways (**Figure 9**, see **Supplemental Figure S9** for expanded pathway DEG names).

Discussion

The importance of combining TIL and ptPBL profiles, large patient datasets, and bioinformatics to resolve singular predictive biomarkers representing pan-pathology personalized immunotherapeutic targets cannot be understated. Here, we set out to elucidate novel pan-cancer targetable immune biomarkers using a non-biased approach, where we discovered numerous pan-cancer FIR-DEGs correlating with patient survival, several of which have also been identified as immunotherapy resistance, HIV-controller, and bacterial infection biomarkers. In developing an assay for RCC patient stratification, we found that this was possible with as little as three $CD8^+$ ptPBL DEGs (i.e., *MMP9*, *LEF1* and *IFNG*). We discovered that *MMP9* was increased among pan-cancer recurring patients, with some suffering from as many as three malignancies without recorded metastases. PPI networks placed *MMP9* as a central node of interaction matrices amid other pan-cancer FIR-DEGs identifying cell migration and cytokine pathways.

Matrix metalloproteinase 9 (MMP-9; leukocyte Gelatinase B), is a type IV collagenase (35) belonging to the >25 MMP family of secreted and transmembrane-bound, zinc-dependent endopeptidases controlling tissue remodeling by degrading components of the extracellular matrix, proteinases, adhesion molecules, chemokines and cytokines (36). Following its activation by cleavage, secreted pro-MMP-9 is regulated

by tissue inhibitor of metalloproteinases 1 (TIMP-1), also identified as a pan-cancer FIR-DEG. MMP-9 is a biomarker of pathogenesis and progression of various diseases, including heart disease, atherosclerosis, hypertension, chronic obstructive pulmonary disease, blood–brain barrier disorders, endometriosis, Down syndrome, and streptococcal pneumonia (37, 38). MMP-9 is important for productive infection by numerous viruses, including hepatitis viruses, vesicular stomatitis virus, respiratory syncytial virus, and HIV-1 (39).

While many MMPs associate with poor prognoses, MMP-9 is the most extensively studied, and associates with aggressive phenotypes and poor prognoses in several solid malignancies (40). In cancer, MMP-9 is associated with genetic instability, tissue remodeling, tumor cell proliferation, invasion and motility, progression, extravasation, metastasis, epithelial-mesenchymal transition, angiogenesis, apoptosis, inflammation, and immunosurveillance (35). MMP-9 in PBLs, serum, and tumors, predicts prognosis, invasiveness, grade and differentiation, recurrence, metastasis, and treatment-resistance, for bladder, lung, blood, colorectal, prostate, and liver cancers (41-48). In RCC, MMP-9 is increased in tumors and plasma (49), and correlates with histological grade (50), poor prognosis and lowered survival (51), metastasis, decreased time-to-progression, and poor response to sunitinib (52).

Though many studies have examined MMPs facilitating epithelial cancer invasion via migration towards chemokines (53), others indicate MMPs shape the aggressive stromal compartment of the tumor microenvironment (54). MMPs are expressed by various immune subsets within stromal compartment across cancers, with TILs as the highest producers of MMP-9 (55). MMP-9 expression disproportionately correlates with

immune response genes rather than extracellular matrix genes in lung adenocarcinomas (56). MMP-9 is constitutively expressed at higher levels in splenic T cells and TILs than tumor cells (55). Immunosurveillance requires an invasive phenotype, supporting a role for MMP-9, upregulated in response to cytokines, and facilitating migration of PBLs from the vasculature into sites of inflammation. Induced by adhesion-molecule interactions between APCs and lymphocytes, MMP-9 is considered a pure effector molecule, amplifying immune function by catalysing innate and acquired immunity (57), and is critical for antigen-specific activation-induced proliferation of T cells (58). We observed increased *MMP9* expression in both $CD8^+$ and $CD19^+$ ccRCC ptPBLs, and other genes we find correlating with *MMP9* expression are implicated in $CD8^+$ cross-priming and antigen-mediated activation and proliferation (*CD69, STAT4, NFIL3, IL10, JAK1*) (39, 58, 59). In patients with recurring pan-cancers, *MMP9* was inversely correlated with its regulator STAT3, which restricts tumor-penetration of anti-tumor $CD8^+$ T cells (60, 61). In pooled RNA PCA analyses, *MMP9* expression correlated with TILs and *CXCL13,* which induces MMP-9 expression towards leukocyte migration in perivascular spaces (62, 63). *BATF* was the only pan-cancer FIR-DEG negatively correlating with *MMP9* in RCC, ccRCC, and in pan-cancer recurring patients, and is known to suppress MMP-9 and effector molecules for $CD8^+$ T cell differentiation and survival (64). Other DEGs in pan-cancer recurring patients (*KLF4*, *RORC*, *CCR4*, *PDK4*) are implicated in MMP-9 regulation for cell migration (65-68). Some DEGs stratifying RCC patients and correlating with *MMP9*, have been associated with response to immunotherapies (*MMP9, IL10, NFIL3, LEF1, FASLG, MKI67, STAT4, CD244, JAK1)*. In a recent study examining correlates with ccRCC response to ICB, despite an underpowered discovery cohort,

increased *MMP9* correlated with progressive disease relative to low *MMP9* in partial responders or patients with stable disease (69). Another study demonstrated that reduction of MMP-9 on monocytes from combination immunotherapies led to conversion of tumor microenvironments from “cold” to “inflamed” states eliciting protective T cell responses (70). Finally, *MMP9* and numerous other extracellular matrix pan-cancer genes we have identified, have also just recently been shown to be differentially expressed between ICB responders and non-responders (71), and where, notably, we identify almost half (44%) of their defined immunotherapy failure signature as gained PPIs of MMP-9 during cancer.

Possible mechanisms of MMP-9 regulation in cancer are supported by our observations of alternative *MMP9* RNA isoforms in RCC $CD8^+$ ptPBLs and TILs, and by previous association of MMP9 polymorphisms with disease (35). Another possible mechanism stems from reports of pharmacologic concentrations of melatonin inhibiting MMP-9 through melatonin receptor 1A (MTNR1A), as an oncostatic agent inhibiting tumor growth and invasiveness of renal, breast, gastric, hematological, prostate, bone, skin, liver, and brain tumors (32). MTNR1A is extensively expressed by splenic, thymic and circulating $CD4^+$, $CD8^+$ and B^+ lymphocytes (72), and was found to be significantly reduced in ccRCC TILs and TIL-Bs, and thus less likely to respond to melatonin or regulate MMP-9. Thus, representing an attractive target for the treatment of numerous diseases, a specific MMP-9 inhibitor, JNJ0966, has been developed for clinical utility in immune disorders (73), and could be repurposed for the control of numerous cancers and pathologies. Finally, another possible MMP-9-cancer onset mechanism stems from our observations of its gained and disrupted PPIs were most significant in colon, mouth, and

lung cancers. Thus, deregulation of MMP-9 roles in maintaining colonic microbiota (74), may have profound effects on cancers in these systems.

We have focused on MMP-9 because this molecule efficiently stratifies RCC patients via both $CD8^+$ and $CD19^+$ immune cells, yet extends itself beyond pan-cancer: as an attractive pan-pathology immune biomarker and target for the treatment of numerous diseases, and is a central interactor for numerous other pan-cancer biomarkers we have identified herein. MMP-9 is an attractive pan-pathology biomarker because its elevated expression can be detected in tumors, but also by less invasive methods examining blood, and by completely non-invasive methods examining urine, feacal samples, tears and exhaled breath condensate (75-77). Since MMP-9 expression is controlled by circadian rhythm, diurnal variations may be responsible for its initial immune deregulation in patients. With the possibility of its expression causing varied responses among patients, MMP-9 monitoring may be valuable over treatment course, be it radio-, chemo-, or immuno-therapeutic. Further work determining the precise polarity of T cell subsets expressing MMP-9, and its function permitting their infiltration into tumors, may resolve imminent questions of the field concerning tumor immunogenicity. Aside from *MMP9,* however, numerous other identified pan-cancer FIR-DEGs warrant further investigation.

In the discovery of how pan-cancer FIR-DEGs may have pan-pathology effects, we observed that blinded refining of target list using informatics without previous knowledge of their onco-static or -promoting abilities, increased identity of pan-cancer DEGs also important for HIV-1 controllers, as the first data-driven suggestion these may actually represent pan-pathology markers. To be certain of this link to HIV-1 fitness

correlated by our pan-cancer FIR-DEGs, we performed extensive literature reviews indicating that the majority shared similar effects promotion or inhibition of HIV-1 infection. Other PBL DEGs shared by many infectious and autoimmune diseases are common to those discovered here (e.g., *MMP9*, *IGF2BP3, TIMP1, CDA, IFNG, PFAV1, LAG3, PIM2, ICOS, TIGIT, IL23A*).

To our knowledge, we are the first to present a pan-pathology biomarker pipeline starting from $CD8^+$ and $CD19^+$ DEGs from paired cancer patient PBLs, TILs and TIICs. Prognostic DEGs were validated across five cancers and as little as three could stratified patients and pan-cancer recurrence. We focused on DEGs that could be easily detected as biomarkers or targeted by therapies. The resulting pan-cancer genes are of additional interest as they may reflect a pan-pathology state of the FIR, with numerous DEGs similarly modulated and important for resistance to immunotherapy, and unresolved viral and bacterial infections. The non-invasive FIR-DEGs we have identified warrant future investigation towards the development of their potential in precision diagnostics and precision pan-disease immunotherapeutics.

Methods

Complete detailed methods are in the supplemental materials.

Statistics

For the training ccRCC cohort, a sample size of $n = 5$ paired patient TILs, TIICs, and PBLs was determined as having above 0.9 power according to the GeneChip Human Transcriptome Array 2.0 manufacturer guidelines (Affymetrix, Thermo Fisher Scientific).

Training set microarrays power calculations by the manufacturer used an inference of means calculation from https://www.stat.ubc.ca/~rollin/stats/ssize/n2. Multiple hypothesis test correction was performed using the FDR Benjamini–Hochberg step-up procedure. For the RCC validation cohort (n = 74), power analysis determined that a minimal sample size of n = 62 to reach a power of 0.80 at α = 0.05 (two-tailed) (G*Power ver. 3.1.9.2; Universitat Düsseldorf, Germany). For algorithms used, statistical methodology for algorithms is described within scripts and https://www.biostars.org/p/153013/. Limma and survival packages for R are used for single and synergistic ccRCC prognostic algorithms. Dendrogram, heatmap and PCA unsupervised algorithms used the Euclidean distance metric and complete linkage clustering method. Correlogram algorithm uses the R corrplot library, and was created from http://www.sthda.com/, using r-project corrplot and vignette packages. Binomial correlations for testing of validated DEGs against clinical patient parameters used two-tailed nonparametric Spearman correlation with 95% CI (Prism V6.01, GraphPad). An unpaired 2-tailed student's t test with FDR of 1% was used to compare two groups, and two-way ANOVA (with Sidak's multiple-comparisons test) and 95% CI was used for multiple comparisons. Pathway enrichment analysis results were adjusted for multiple testing by applying FDR and Bonferroni methods. *P*-values of less than 0.05 were considered to indicate a statistically significant difference.

Study approval

Renal cancer patients underwent resection for stage I-IV RCC between 2013 and 2017 at the CHUM (Montreal, Canada). Study approval, and written and informed consent procedures approval was granted by the CHUM Research Ethics Board. Written informed

consent was received from all study participants prior to inclusion in the study. All methods were performed in accordance with the relevant guidelines and regulations. Clinical participant data was randomly numbered for complete anonymity.

Data availability

TCGA KIRC RNA-seq datasets and associated clinical datasets are available at the cBioPortal for Cancer Genomics at http://gdac.broadinstitute.org/. Pan-cancer testing patient cohort GEO, the EGA and TCGA datasets are available at http://kmplot.com/. DEG protein profiles in cells and across 17 cancers are available from https://www.proteinatlas.org/. Transcriptomic datasets from melanoma and NSCLC patients treated with anti-PD-1 therapy are available from Hugo et al. (30), and Rizvi et al. (31). The HIV-1 elite controllers dataset is available from Zhang et al. (28), and the bacterial datasets are listed in Song et al. (29). Comprehensive pathway enrichment analysis and PPI analyses are available as Supplemental Data. The microarray data is published at the National Center for Biotechnology Information Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under GEO accession number GSE117230.

Author contributions

A.Monette, I.J., and R.L. designed experiments; A.Monette, N.A.A-B., and L.R. selected and consented patients, and transported biospecimens; J.-B.L. performed surgeries; J.-B.L. and L.R. contributed to clinical data; A.Monette and N.A.A-B. performed immune cell isolation and biobanking; N.A.A-B. assisted in flow-cytometry analyses; A.Monette performed RNA extractions, organized collaboration and outsourcing; A.Monette and

A.Morou created algorithms; A.Morou selected T cell polarizing validation genes, performed microfluidics qRT-PCR validation experiments, and normalized and analyzed the data; I.J. performed integrative analysis, interaction network, and pathway enrichment analysis, and suggested experiments and analyses; S.R. performed pan-cancer analyses; T.T. performed pathway enrichment analysis; A.Monette and R.L. organized, harmonized, analyzed and interpreted the data, created figures and wrote the manuscript; A.Monette, I.J., D.E.K. and R.L. supervised the study; I.J., A.Morou, D.E.K., N.A.A-B., J.-P.R., and R.L. critically reviewed the manuscript. All authors revised the final manuscript.

Supplemental material

Supplemental Figures and Tables

Supplemental Tables

Supplemental Methods

Supplemental File S1

Acknowledgements

The authors wish to acknowledge all study participants and patients; The Cancer Genome Atlas; Mathieu Latour and Roula Albadine and supporting staff of the CHUM pathology department; Manon de Ladurantaye and Anne-Marie Mes-Masson from the CRCHUM for RNA quality profiling, Geneviève Cormier and Fred Saad from the CRCHUM for blood drawing of control donors; Gilles Corbeil of the CRCHUM genomics department for RNA quality testing and microarray profiling; Francois Harvey of the CRCHUM bioinformatics department; Peter Graf and Patrick Sabourin from Affymetrix for providing reagents and

561 technical assistance; Zeeshan Farroq and Ofir Goldberger from Fluidigm, Erika Diaz from
562 StemCell, Andrew Mouland from McGill University, Simon Turcotte from University of
563 Montreal, and Sascha Ring and Tris from Biostars for their advice. This work was partially
564 performed at the Institut du Cancer de Montréal, CRCHUM and University of Montreal, in
565 Montreal, Quebec, Canada. This work was supported by a Canadian Cancer Society
566 Research Institute grant (CCSRI #702036 to R.L. and I.J.), and a Biomedical Research
567 Grant from the Kidney Foundation of Canada (KFOC130019 to R.L.). R.L. is supported
568 by the Quebec Cell, Tissue and Gene Therapy Network –ThéCell (a thematic network
569 supported by the Fonds de recherche du Québec–Santé (FRQS)), the FRQS, and the
570 Immunotherapy network (iTNT) from the Terry Fox Research Institute (TFRI), A.Monette
571 is supported by Mitacs, Merck, l'Institut du cancer de Montreal (ICM), the Society for
572 Immunotherapy of Cancer, and the Lady Davis Institute for Medical Research; N.A-B. is
573 supported by FRQS post-doctoral award and Qatar University. J.-B.L. is supported by
574 Institut du Cancer de Montréal. J.-P.R. holds the Louis Lowenstein Chair, McGill
575 University. D.E.K. is supported by an FRQS Research Scholar Award (#31035), CIHR
576 #377124, NHLBI RO1-HL-092565 and the Canada Foundation for Innovation (CFI
577 #31756). I.J. and computational analysis was supported by the Canada Research Chair
578 Program (CRC #225404), Ontario Research Fund (#34876), Natural Sciences Research
579 Council (NSERC #203475), Canada Foundation for Innovation (CFI #203373, #30865),
580 Krembil Foundation and IBM.

581 # Footnotes

582 ## Conflict of interest:

583 The authors have declared that no conflict of interest exists.

584

References

1. Morris LG, Riaz N, Desrichard A, Senbabaoglu Y, Hakimi AA, Makarov V, et al. Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget.* 2016;7(9):10051-63.
2. Mak MP, Tong P, Diao L, Cardnell RJ, Gibbons DL, William WN, et al. A Patient-Derived, Pan-Cancer EMT Signature Identifies Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition. *Clin Cancer Res.* 2016;22(3):609-20.
3. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med.* 2015;21(8):938-45.
4. Budczies J, Bockmayr M, Klauschen F, Endris V, Frohling S, Schirmacher P, et al. Mutation patterns in genes encoding interferon signaling and antigen presentation: A pan-cancer survey with implications for the use of immune checkpoint inhibitors. *Genes Chromosomes Cancer.* 2017;56(8):651-9.
5. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* 2017;18(1):248-62.
6. Li Y, Kang K, Krahn JM, Croutwater N, Lee K, Umbach DM, et al. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC Genomics.* 2017;18(1):508.
7. Kaczkowski B, Tanaka Y, Kawaji H, Sandelin A, Andersson R, Itoh M, et al. Transcriptome Analysis of Recurrently Deregulated Genes across Multiple Cancers Identifies New Pan-Cancer Biomarkers. *Cancer Res.* 2016;76(2):216-26.
8. Akbani R, Ng PK, Werner HM, Shahmoradgoli M, Zhang F, Ju Z, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nature communications.* 2014;5:3887.
9. Xu Q, Chen J, Ni S, Tan C, Xu M, Dong L, et al. Pan-cancer transcriptome analysis reveals a gene expression signature for the identification of tumor tissue origin. *Mod Pathol.* 2016;29(6):546-56.
10. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 2016;17(1):174.
11. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science.* 2016;352(6282):189-96.
12. Gnjatic S, Bronte V, Brunet LR, Butler MO, Disis ML, Galon J, et al. Identifying baseline immune-related biomarkers to predict clinical outcome of immunotherapy. *J Immunother Cancer.* 2017;5:44.
13. Lim C, Tsao MS, Le LW, Shepherd FA, Feld R, Burkes RL, et al. Biomarker testing and time to treatment decision in patients with advanced nonsmall-cell lung cancer. *Ann Oncol.* 2015;26(7):1415-21.
14. Crowley E, Di Nicolantonio F, Loupakis F, and Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol.* 2013;10(8):472-84.
15. Haber DA, and Velculescu VE. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. *Cancer Discov.* 2014;4(6):650-61.
16. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature.* 2018;563(7732):579-83.

17. Best MG, Sol N, Kooi I, Tannous J, Westerman BA, Rustenburg F, et al. RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell.* 2015;28(5):666-76.
18. Alix-Panabieres C, and Pantel K. Challenges in circulating tumour cell research. *Nat Rev Cancer.* 2014;14(9):623-31.
19. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat Med.* 2016;22(4):433-8.
20. Farsaci B, Donahue RN, Grenga I, Lepone LM, Kim PS, Dempsey B, et al. Analyses of Pretherapy Peripheral Immunoscore and Response to Vaccine Therapy. *Cancer Immunol Res.* 2016;4(9):755-65.
21. Mishra S, Kaddi CD, and Wang MD. Pan-cancer analysis for studying cancer stage using protein and gene expression data. *Conf Proc IEEE Eng Med Biol Soc.* 2016;2016:2440-3.
22. MacFarlane AWt, Jillab M, Plimack ER, Hudes GR, Uzzo RG, Litwin S, et al. PD-1 expression on peripheral blood cells increases with stage in renal cell carcinoma patients and is rapidly reduced after surgical tumor resection. *Cancer Immunol Res.* 2014;2(4):320-31.
23. Fridman WH, Pages F, Sautes-Fridman C, and Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer.* 2012;12(4):298-306.
24. Motzer RJ, Tannir NM, McDermott DF, Aren Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *The New England journal of medicine.* 2018;378(14):1277-90.
25. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research.* 2015;43(Database issue):D447-52.
26. Haymaker CL, Wu RC, Ritthipichai K, Bernatchez C, Forget MA, Chen JQ, et al. BTLA marks a less-differentiated tumor-infiltrating lymphocyte subset in melanoma with enhanced survival properties. *Oncoimmunology.* 2015;4(8):e1014246.
27. Martin-Gayo E, Cronin J, Hickman T, Ouyang Z, Lindqvist M, Kolb KE, et al. Circulating CXCR5(+)CXCR3(+)PD-1(lo) Tfh-like cells in HIV-1 controllers with neutralizing antibody breadth. *JCI Insight.* 2017;2(2):e89574.
28. Zhang LL, Zhang ZN, Wu X, Jiang YJ, Fu YJ, and Shang H. Transcriptomic meta-analysis identifies gene expression characteristics in various samples of HIV-infected patients with nonprogressive disease. *J Transl Med.* 2017;15(1):191.
29. Song F, Qian Y, Peng X, Li X, Xing P, Ye D, et al. The frontline of immune response in peripheral blood. *PLoS One.* 2017;12(8):e0182294.
30. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell.* 2016;165(1):35-44.
31. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* 2015;348(6230):124-8.
32. Lin YW, Lee LM, Lee WJ, Chu CY, Tan P, Yang YC, et al. Melatonin inhibits MMP-9 transactivation and renal cell carcinoma metastasis by suppressing Akt-MAPKs pathway and NF-kappaB DNA-binding activity. *J Pineal Res.* 2016;60(3):277-90.
33. Jensen C, Madsen DH, Hansen M, Schmidt H, Svane IM, Karsdal MA, et al. Non-invasive biomarkers derived from the extracellular matrix associate with response to immune checkpoint blockade (anti-CTLA-4) in metastatic melanoma patients. *J Immunother Cancer.* 2018;6(1):152.

34. Rahmati S, Abovsky M, Pastrello C, and Jurisica I. pathDIP: an annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. *Nucleic acids research.* 2017;45(D1):D419-D26.
35. Farina AR, and Mackay AR. Gelatinase B/MMP-9 in Tumour Pathogenesis and Progression. *Cancers (Basel).* 2014;6(1):240-96.
36. Goetzl EJ, Banda MJ, and Leppert D. Matrix metalloproteinases in immunity. *J Immunol.* 1996;156(1):1-4.
37. Liu H, Wang J, Wang H, Tang N, Li Y, Zhang Y, et al. Correlation between matrix metalloproteinase-9 and endometriosis. *Int J Clin Exp Pathol.* 2015;8(10):13399-404.
38. Vissers M, Hartman Y, Groh L, de Jong DJ, de Jonge MI, and Ferwerda G. Recognition of Streptococcus pneumoniae and muramyl dipeptide by NOD2 results in potent induction of MMP-9, which can be controlled by lipopolysaccharide stimulation. *Infect Immun.* 2014;82(12):4952-8.
39. Chen J, Xu W, Chen Y, Xie X, Zhang Y, Ma C, et al. Matrix Metalloproteinase 9 Facilitates Hepatitis B Virus Replication through Binding with Type I Interferon (IFN) Receptor 1 To Repress IFN/JAK/STAT Signaling. *J Virol.* 2017;91(8).
40. Kunz P, Sahr H, Lehner B, Fischer C, Seebach E, and Fellenberg J. Elevated ratio of MMP2/MMP9 activity is associated with poor response to chemotherapy in osteosarcoma. *BMC Cancer.* 2016;16:223.
41. Wieczorek E, Jablonowski Z, Tomasik B, Konecki T, Jablonska E, Gromadzinska J, et al. MMP, VEGF and TIMP as prognostic factors in recurring bladder cancer. *Clin Biochem.* 2015;48(18):1235-40.
42. Sun Y, Chen Y, Li S, Lei Y, Xu D, Jiang N, et al. NanoVelcro-captured CTC number concomitant with enhanced serum levels of MMP7 and MMP9 enables accurate prediction of metastasis and poor prognosis in patients with lung adenocarcinoma. *Int J Nanomedicine.* 2017;12:6399-412.
43. Ramon de Fata F, Ferruelo A, Andres G, Gimbernat H, Sanchez-Chapado M, and Angulo JC. The role of matrix metalloproteinase MMP-9 and TIMP-2 tissue inhibitor of metalloproteinases as serum markers of bladder cancer. *Actas Urol Esp.* 2013;37(8):480-8.
44. Gusella M, Bolzonella C, Paolini R, Rodella E, Bertolaso L, Scipioni C, et al. Plasma matrix metalloprotease 9 correlates with blood lymphocytosis, leukemic cell invasiveness, and prognosis in B-cell chronic lymphocytic leukemia. *Tumour Biol.* 2017;39(2):1010428317694325.
45. Alonso S, Mayol X, Nonell L, Salvans S, Pascual M, Pera M, et al. Peripheral blood leucocytes show differential expression of tumour progression-related genes in colorectal cancer patients who have a postoperative intra-abdominal infection: a prospective matched cohort study. *Colorectal Dis.* 2017;19(5):O115-O25.
46. Marin-Aguilera M, Reig O, Lozano JJ, Jimenez N, Garcia-Recio S, Erill N, et al. Molecular profiling of peripheral blood is associated with circulating tumor cells content and poor survival in metastatic castration-resistant prostate cancer. *Oncotarget.* 2015;6(12):10604-16.
47. Shah K, Patel S, Mirza S, and Rawal RM. Unravelling the link between embryogenesis and cancer metastasis. *Gene.* 2018;642:447-52.
48. Tang D, Zhang J, Yuan Z, Zhang H, Chong Y, Huang Y, et al. PSC-derived Galectin-1 inducing epithelial-mesenchymal transition of pancreatic ductal adenocarcinoma cells by activating the NF-kappaB pathway. *Oncotarget.* 2017;8(49):86488-502.
49. Lein M, Jung K, Laube C, Hubner T, Winkelmann B, Stephan C, et al. Matrix-metalloproteinases and their inhibitors in plasma and tumor tissue of patients with renal cell carcinoma. *Int J Cancer.* 2000;85(6):801-4.
50. Kawata N, Nagane Y, Hirakata H, Ichinose T, Okada Y, Yamaguchi K, et al. Significant relationship of matrix metalloproteinase 9 with nuclear grade and prognostic impact of tissue inhibitor of metalloproteinase 2 for incidental clear cell renal cell carcinoma. *Urology.* 2007;69(6):1049-53.

51. Kallakury BV, Karikehalli S, Haholu A, Sheehan CE, Azumi N, and Ross JS. Increased expression of matrix metalloproteinases 2 and 9 and tissue inhibitors of metalloproteinases 1 and 2 correlate with poor prognostic variables in renal cell carcinoma. *Clin Cancer Res.* 2001;7(10):3113-9.
52. Perez-Gracia JL, Prior C, Guillen-Grima F, Segura V, Gonzalez A, Panizo A, et al. Identification of TNF-alpha and MMP-9 as potential baseline predictive serum markers of sunitinib activity in patients with renal cell carcinoma using a human cytokine array. *Br J Cancer.* 2009;101(11):1876-83.
53. Chang C, and Werb Z. The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. *Trends Cell Biol.* 2001;11(11):S37-43.
54. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, and Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Res.* 2016;18(1):84.
55. Owen JL, Iragavarapu-Charyulu V, Gunja-Smith Z, Herbert LM, Grosso JF, and Lopez DM. Up-regulation of matrix metalloproteinase-9 in T lymphocytes of mammary tumor bearers: role of vascular endothelial growth factor. *J Immunol.* 2003;171(8):4340-51.
56. Creighton C, and Hanash S. Expression of matrix metalloproteinase 9 (MMP-9/gelatinase B) in adenocarcinomas strongly correlated with expression of immune response genes. *In Silico Biol.* 2003;3(3):301-11.
57. Opdenakker G, Van den Steen PE, and Van Damme J. Gelatinase B: a tuner and amplifier of immune functions. *Trends Immunol.* 2001;22(10):571-9.
58. Benson HL, Mobashery S, Chang M, Kheradmand F, Hong JS, Smith GN, et al. Endogenous matrix metalloproteinases 2 and 9 regulate activation of CD4+ and CD8+ T cells. *Am J Respir Cell Mol Biol.* 2011;44(5):700-8.
59. Kothari P, Pestana R, Mesraoua R, Elchaki R, Khan KM, Dannenberg AJ, et al. IL-6-mediated induction of matrix metalloproteinase-9 is modulated by JAK-dependent IL-10 expression in macrophages. *J Immunol.* 2014;192(1):349-57.
60. Liu F, Zhang T, Zou S, Jiang B, and Hua D. B7H3 promotes cell migration and invasion through the Jak2/Stat3/MMP9 signaling pathway in colorectal cancer. *Mol Med Rep.* 2015;12(4):5455-60.
61. Yue C, Shen S, Deng J, Priceman SJ, Li W, Huang A, et al. STAT3 in CD8+ T Cells Inhibits Their Tumor Accumulation by Downregulating CXCR3/CXCL10 Axis. *Cancer Immunol Res.* 2015;3(8):864-70.
62. Singh S, Singh R, Singh UP, Rai SN, Novakovic KR, Chung LW, et al. Clinical and biological significance of CXCR5 expressed by prostate cancer specimens and cell lines. *Int J Cancer.* 2009;125(10):2288-95.
63. Modvig S, Degn M, Horwitz H, Cramer SP, Larsson HB, Wanscher B, et al. Relationship between cerebrospinal fluid biomarkers for inflammation, demyelination and neurodegeneration in acute optic neuritis. *PLoS One.* 2013;8(10):e77163.
64. Hasegawa H, Senga T, Ito S, Iwamoto T, and Hamaguchi M. A role for AP-1 in matrix metalloproteinase production and invadopodia formation of v-Crk-transformed cells. *Exp Cell Res.* 2009;315(8):1384-92.
65. Li W, Liu M, Su Y, Zhou X, Liu Y, and Zhang X. The Janus-faced roles of Kruppel-like factor 4 in oral squamous cell carcinoma cells. *Oncotarget.* 2015;6(42):44480-94.
66. Sato W, Tomita A, Ichikawa D, Lin Y, Kishida H, Miyake S, et al. CCR2(+)CCR5(+) T cells produce matrix metalloproteinase-9 and osteopontin in the pathogenesis of multiple sclerosis. *J Immunol.* 2012;189(10):5057-65.
67. Yang Y, Du L, Yang X, Qu A, Zhang X, Zhou C, et al. Aberrant CCR4 expression is involved in tumor invasion of lymph node-negative human gastric cancer. *PLoS One.* 2015;10(3):e0120059.
68. Cheng CY, Hsieh HL, Hsiao LD, and Yang CM. PI3-K/Akt/JNK/NF-kappaB is essential for MMP-9 expression and outgrowth in human limbal epithelial cells on intact amniotic membrane. *Stem Cell Res.* 2012;9(1):9-23.

69. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ, et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science.* 2018;359(6377):801-6.
70. Perry CJ, Munoz-Rojas AR, Meeth KM, Kellman LN, Amezquita RA, Thakral D, et al. Myeloid-targeted immunotherapies act in synergy to induce inflammation and antitumor immunity. *The Journal of experimental medicine.* 2018;215(3):877-93.
71. Chakravarthy A, Khan L, Bensler NP, Bose P, and De Carvalho DD. TGF-beta-associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. *Nature communications.* 2018;9(1):4692.
72. Pozo D, Delgado M, Fernandez-Santos JM, Calvo JR, Gomariz RP, Martin-Lacave I, et al. Expression of the Mel1a-melatonin receptor mRNA in T and B subsets of lymphocytes from rat thymus and spleen. *FASEB J.* 1997;11(6):466-73.
73. Scannevin RH, Alexander R, Haarlander TM, Burke SL, Singer M, Huo C, et al. Discovery of a highly selective chemical inhibitor of matrix metalloproteinase-9 (MMP-9) that allosterically inhibits zymogen activation. *J Biol Chem.* 2017;292(43):17963-74.
74. Pujada A, Walter L, Patel A, Bui TA, Zhang Z, Zhang Y, et al. Matrix metalloproteinase MMP9 maintains epithelial barrier function and preserves mucosal lining in colitis associated cancer. *Oncotarget.* 2017;8(55):94650-65.
75. Moses MA, Wiederschain D, Loughlin KR, Zurakowski D, Lamb CC, and Freeman MR. Increased incidence of matrix metalloproteinases in urine of cancer patients. *Cancer Res.* 1998;58(7):1395-9.
76. Annahazi A, Abraham S, Farkas K, Rosztoczy A, Inczefi O, Foldesi I, et al. A pilot study on faecal MMP-9: a new noninvasive diagnostic marker of colorectal cancer. *Br J Cancer.* 2016;114(7):787-92.
77. Markoulli M, Papas E, Cole N, and Holden BA. The diurnal variation of matrix metalloproteinase-9 and its associated factors in human tears. *Investigative ophthalmology & visual science.* 2012;53(3):1479-84.

Figures

Figure 1

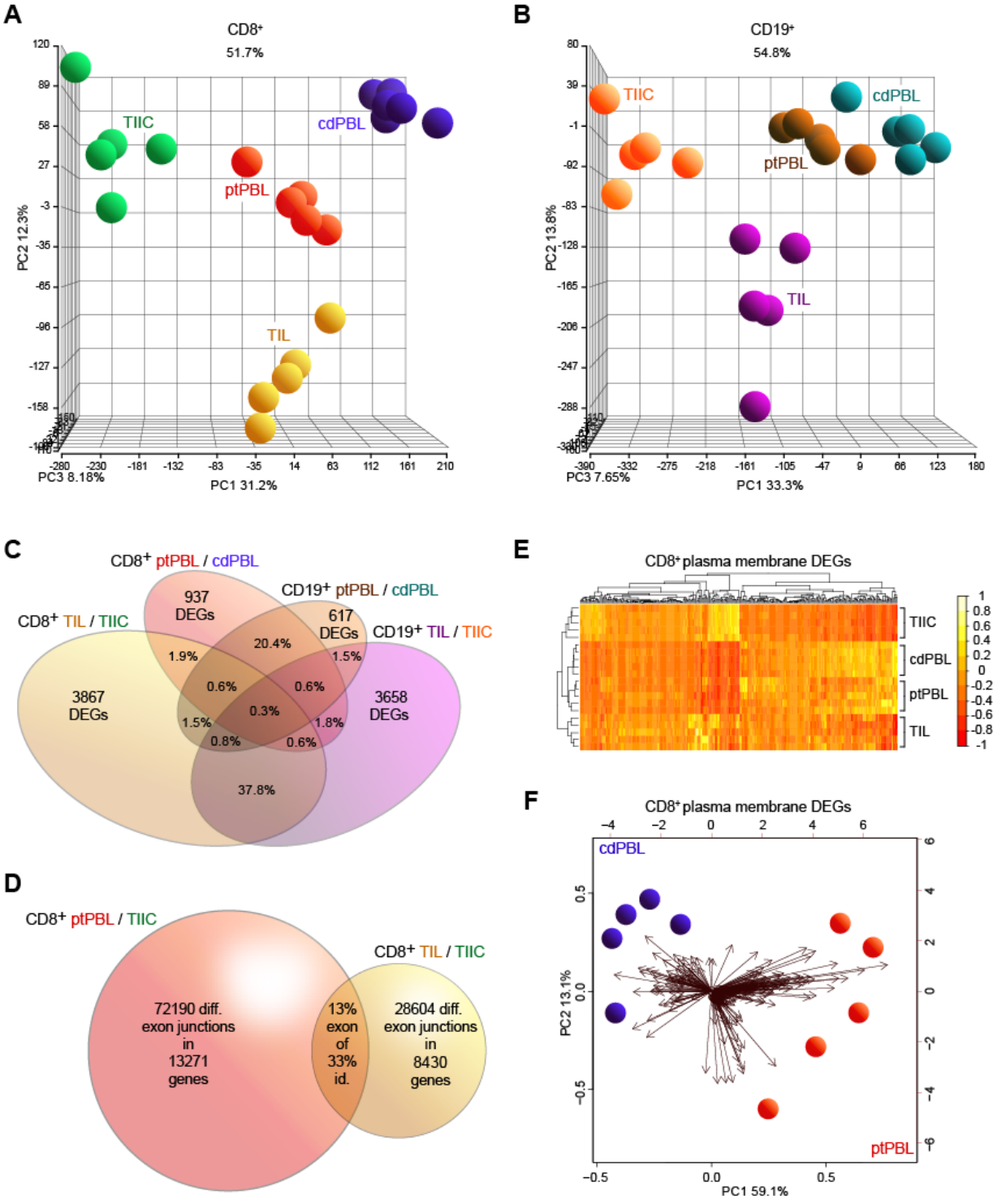

Figure 1. Distinct comprehensive transcriptomics from paired $CD8^+$ and $CD19^+$ profiles from ccRCC blood, tumors and tissues, and control donor blood isolates. (**A-B)** PCA demonstrating distinct DEG profiles from comprehensive HTA 2.0 microarray analyses of (**A**) $CD8^+$ (*n* = 15) and (**B**) $CD19^+$ (*n* = 15) immune cell subsets from paired patient tumors (TIL; TIL-B), normal adjacent tissues (TIIC), and circulating lymphocytes (ptPBL), and control donor matched circulating lymphocytes (cdPBL) (*n* = 10). (**C**) 4-way Venn diagram demonstrating percent overlaps of DEGs identified by microarrays across different source biospecimens analyzed. (**D**) Venn diagram demonstrating that ptPBLs have greater numbers of differentially represented exon-exon probe selected regions (PSR) junctions compared to TILs; relative to TIICs from paired $CD8^+$ samples ($P<0.05$; ANOVA, Transcriptome Analysis Console v.3, Affymetrix). Thirteen percent of shared PSR junctions exist between ptPBLs and TILs, representing 33% of total genes common to ptPBLs and TILs having shared isoform identity. (**E**) Gene Ontology plasma membrane (PM) proteins identified by Partek, and unsupervised hierarchical clustering algorithm generated heatmaps demonstrating that the four different $CD8^+$ isolates are stratified according to PM, using log2 expression values applying the Euclidean distance metric and complete linkage clustering method (R programming language; R-studio). Heatmaps demonstrate the unsupervised clustering of PBL isolates as most closely related, with TILs and TIICs at their boundaries, suggesting that their profiles may be influenced by the cancer microenvironment. (**F**) Feasibility of utilizing PM associated proteins towards identifying pan-cancer DEGs that can stratify patients is demonstrated by PCA biplots of PM DEGs from $CD8^+$ cdPBL and ptPBL isolates created on log2 values using the biplot function (R; R-studio). TIL, tumor infiltrating lymphocytes; TIIC, adjacent normal tissue infiltrating immune cells; ptPBL, patient peripheral blood lymphocyte; cdPBL, control donor peripheral blood lymphocyte; diff., differential; id., identity.

Figure 2

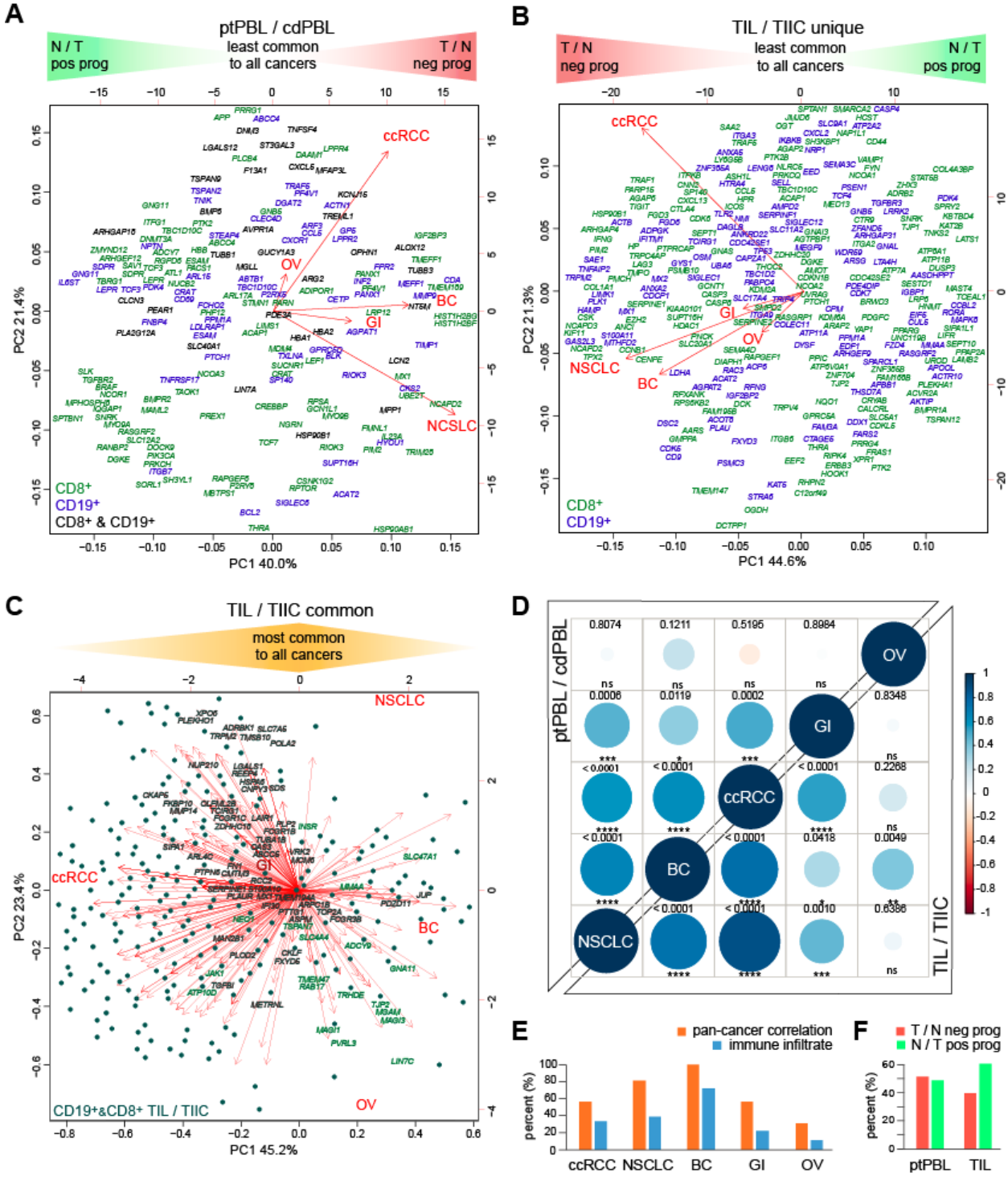
A
ptPBL / cdPBL
least common to all cancers
N / T pos prog
T / N neg prog
ccRCC
OV
BC
GI
NCSLC
CD8+
CD19+
CD8+ & CD19+
PC1 40.0%
PC2 21.4%
B
TIL / TIIC unique
least common to all cancers
T / N neg prog
N / T pos prog
ccRCC
GI
OV
NSCLC
BC
CD8+
CD19+
PC1 44.6%
PC2 21.3%
C
TIL / TIIC common
most common to all cancers
NSCLC
ccRCC
GI
BC
OV
CD19+&CD8+ TIL / TIIC
PC1 45.2%
PC2 23.4%
D
ptPBL / cdPBL
TIL / TIIC
OV
GI
ccRCC
BC
NSCLC
E
pan-cancer correlation
immune infiltrate
percent (%)
ccRCC NSCLC BC GI OV
F
T / N neg prog
N / T pos prog
percent (%)
ptPBL TIL

Figure 2. A subset of prognostic ccRCC DEGs have pan-cancer prognostic potential. (A-B) PCAs nominal derivatives of combined modulation of expression and effects on prognosis to visualize $CD8^+$ and $CD19^+$ DEGs from (**A**) ptPBLs and (**B**) TILs with significant gene modulation and effect on prognosis across the five cancers tested. Genes on far left are more highly expressed in normal tissues than tumors and have positive prognostic effects (N/T pos prog), representing agonistic targets. Genes on far right are more highly expressed in tumors than normal tissues and have negative effects on prognosis (T/N neg prog), representing antagonistic targets. PCAs also illustrate linkage between gene coexpression and cancer types, where BC ptPBLs and NSCLC TILs are most related to other cancers. In (**A**), all ptPBL DEGs are shown. In (**B**), DEGs unique to $CD8^+$ TILs or $CD19^+$ TIL-Bs are shown. (**C**) DEGs common to $CD8^+$ TILs and $CD19^+$ TIL-Bs are shown, where dark highlighted gene names represent best antagonistic targets, and green highlighted gene names represent best agonistic targets. (**D**) Correlograms representing linkage between the five cancers from nominal derivatives, demonstrating that NSCLC and BC are most related to ccRCC, independently of patient sample number (Spearman method, coexpression coefficient ladder on right). (**E**) Graph demonstrating similar expression patterns of pan-cancer DEGs and genes representing infiltrating immune cell subsets used: CD45, CD3, CD4, CD8, CD20, CD56, and CD68 across pan-cancers (*n* = 11,577). (**F**) Graph demonstrating distributions of relative ratios of 483 agonistic vs. agonistic pan-cancer genes, where TILs have higher percentages of genes that are lower in tumors and have positive prognostic value. TIL, tumor infiltrating lymphocytes; TIIC, normal adjacent tissue infiltrating immune cells; ptPBL, patient peripheral blood lymphocyte; cdPBL, control donor peripheral blood lymphocyte, BC, breast cancer; NSCLC, non-small cell lung cancer; GI, gastro-intestinal, OV, ovarian.

Figure 3

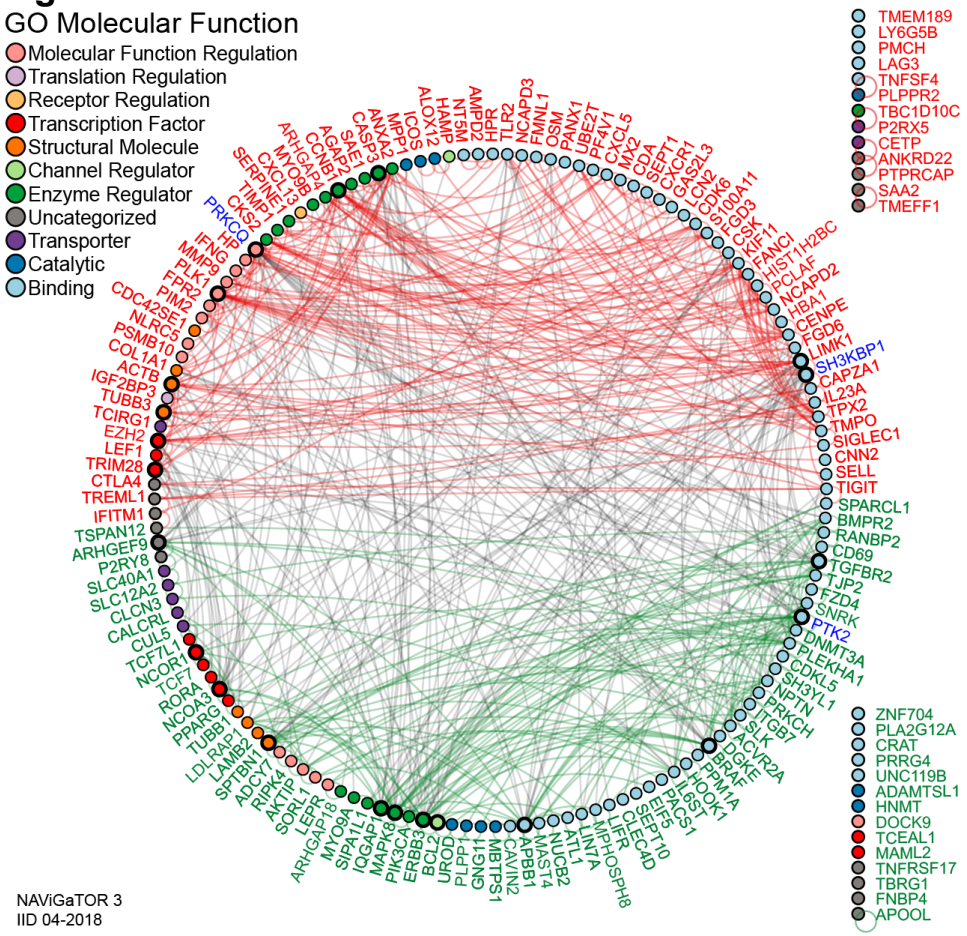

Figure 3. Pan-cancer DEGs have extensive protein-protein interactions (PPI). PPI networks of top 200 DEGs. A high PPI enrichment value (*P* = 1.85e-10) indicating interactions among these DEGs is very significant relative to proteins drawn at random, indicating biological connection as groups in defined pathways. Pan-cancer agonistic (red) and antagonistic (green) DEGs (nodes/circles) and their interactions (edges/lines), demonstrate groupings of these two pan-cancer DEG subclasses, and grey lines highlight interactions between them. Non-interacting DEGs are on right (NAViGaTOR v3 and IID v04-2018). DEG nodes are colored according to GO Molecular Functions listed in the top left legend. Larger node circles represent highest degree DEGs interactors within network, and blue DEG names represent centrality of interactors (as determined by the all pairs shortest path algorithm in NAViGaTOR). GO, Gene Ontology.

Figure 4

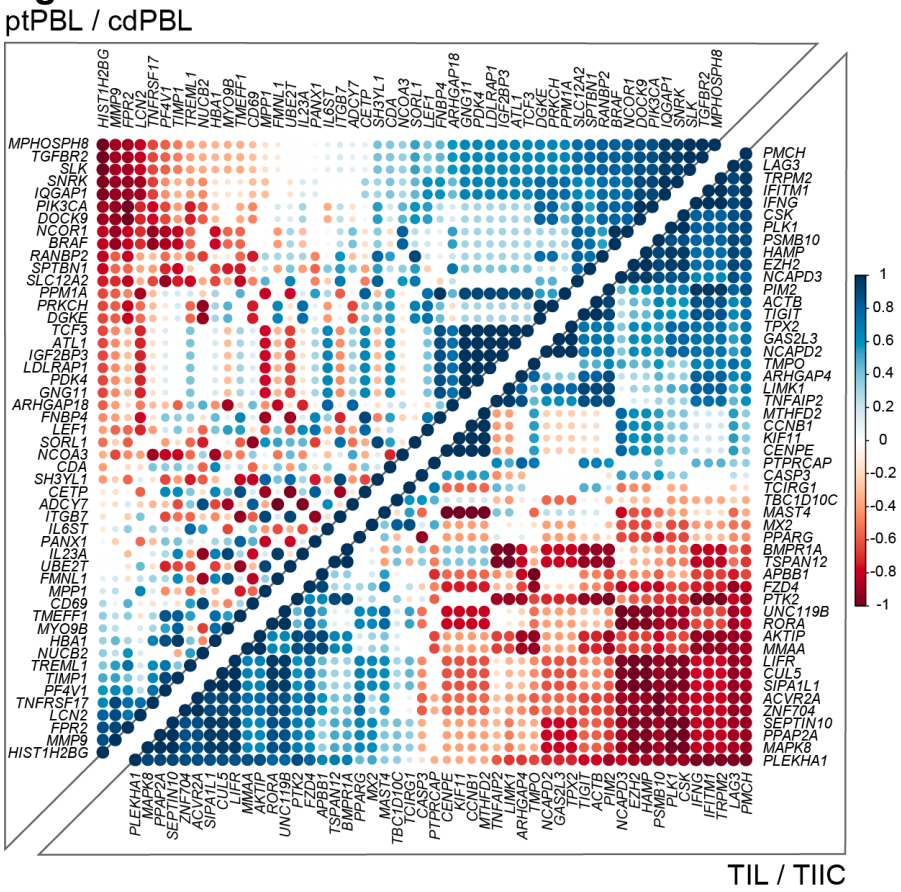

Figure 4. Pan-cancer DEGs have extensive coexpression dynamics. Correlograms of top 200 selected prognostic pan-cancer DEGs demonstrate extensive co-expression dynamics in CD8^{+} ccRCC isolates (Spearman method, expression ladder on right) (*n* = 20). Predominant pathways of the four most highly correlating pan-cancer gene groups included Gene Ontology (GO) biological processes: cellular responses to stimulus, receptor signaling, and regulation of metabolic processes; and Kegg pathways: adherens junctions and colorectal, endometrial, blood and pancreatic cancers for the top right correlating ptPBL gene group, whereas bottom left ptPBL gene groups had extracellular matrix disassembly. For TILs, the bottom left gene group was stronger for GO biological processes: receptor signaling, developmental processes, cell communication, and signal transduction, while the top right TIL gene group was dominated by cell cycle regulation processes (*P* = 4.98e-05), but also: regulation of T cell activation and cytokine production. TIL, tumor infiltrating lymphocytes; TIIC, normal adjacent tissue infiltrating immune cells; ptPBL, patient peripheral blood lymphocyte; cdPBL, control donor peripheral blood lymphocyte.

Figure 5

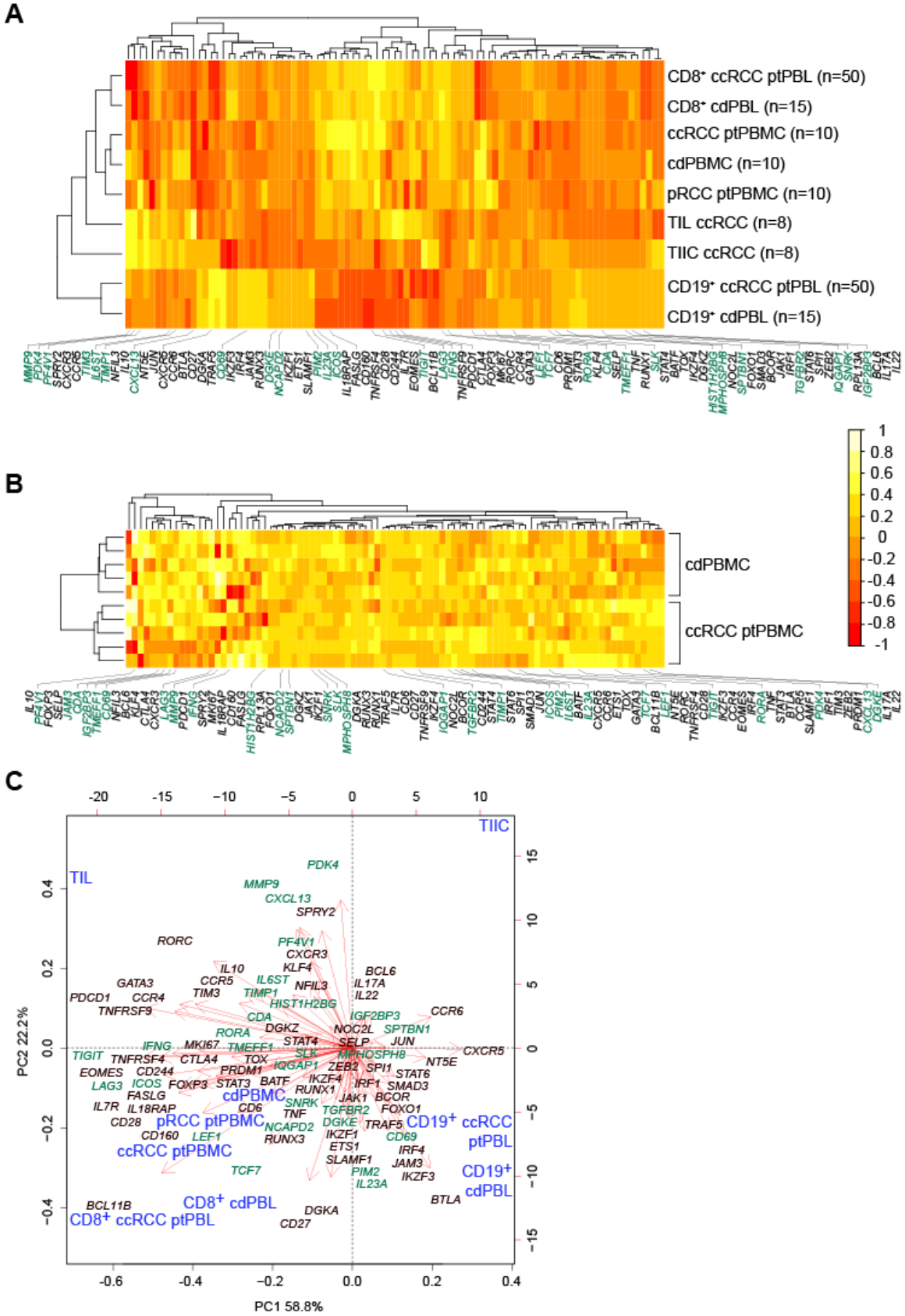
A
CD8+ ccRCC ptPBL (n=50)
CD8+ cdPBL (n=15)
ccRCC ptPBMC (n=10)
cdPBMC (n=10)
pRCC ptPBMC (n=10)
TIL ccRCC (n=8)
TIIC ccRCC (n=8)
CD19+ ccRCC ptPBL (n=50)
CD19+ cdPBL (n=15)
B
cdPBMC
ccRCC ptPBMC
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
C
TIIC
TIL
pRCC ptPBMC
ccRCC ptPBMC
cdPBMC
CD8+ cdPBL
CD8+ ccRCC ptPBL
CD19+ ccRCC ptPBL
CD19+ cdPBL
PC1 58.8%
PC2 22.2%

Figure 5. Coexpressing pan-cancer and polarizing DEGs stratify $CD8^+$, $CD19^+$, PBLs, TILs and TIICs. Using all genes from qRT-PCR validation, (**A-B**) unsupervised hierarchical clustering algorithms, using -ΔCT normalized qRT-PCR expression values, applying the Euclidean distance metric and complete linkage clustering method, were used to generate heatmap clustering and associated dendrograms (R programming language; R-studio). Heatmaps demonstrated that (**A**) pooled fraction used for validation experiments can efficiently stratify all isolates as expected from their genetic linkages and immune cell subset ratios of PBL populations (*n* = 74 patients, *n* = 176 samples, *n* = 9 sample pools), and that (**B**) total individual ptPBLs and cdPBLs cluster separately (*n* = 10). Pooled fractions are also used for principal component analysis (PCA) in (**C**), using -ΔCT normalized qRT-PCR expression values, applying the Euclidean distance metric and complete linkage clustering method (R programming language; R-studio) (*n* = 176 samples), demonstrating that PBLs are most closely linked to circulating $CD8^+$ T cells, and are different in DEG composition relative to circulating $CD19^+$ B cells, TIL-Bs, and TIICs. PCA presented also demonstrates the common and differing coexpression of certain T cell polarizing and pan-cancer DEGs in TIL and $CD8^+$ ptPBL isolates. Pan-cancer genes are highlighted in green throughout. ccRCC, clear cell renal cell carcinoma; pRCC, papillary renal cell carcinoma; pt, patient; nd, normal donor; PBL, peripheral blood lymphocyte; PBMC, peripheral blood mononuclear cells; tot, total; TIL, tumor infiltrating lymphocytes; TIIC, normal adjacent tissue infiltrating immune cells; N, normal tissues DEG; T tumor tissues DEG; pos, positive; neg, negative; prog, prognosis n, number of patients in pool.

Figure 6

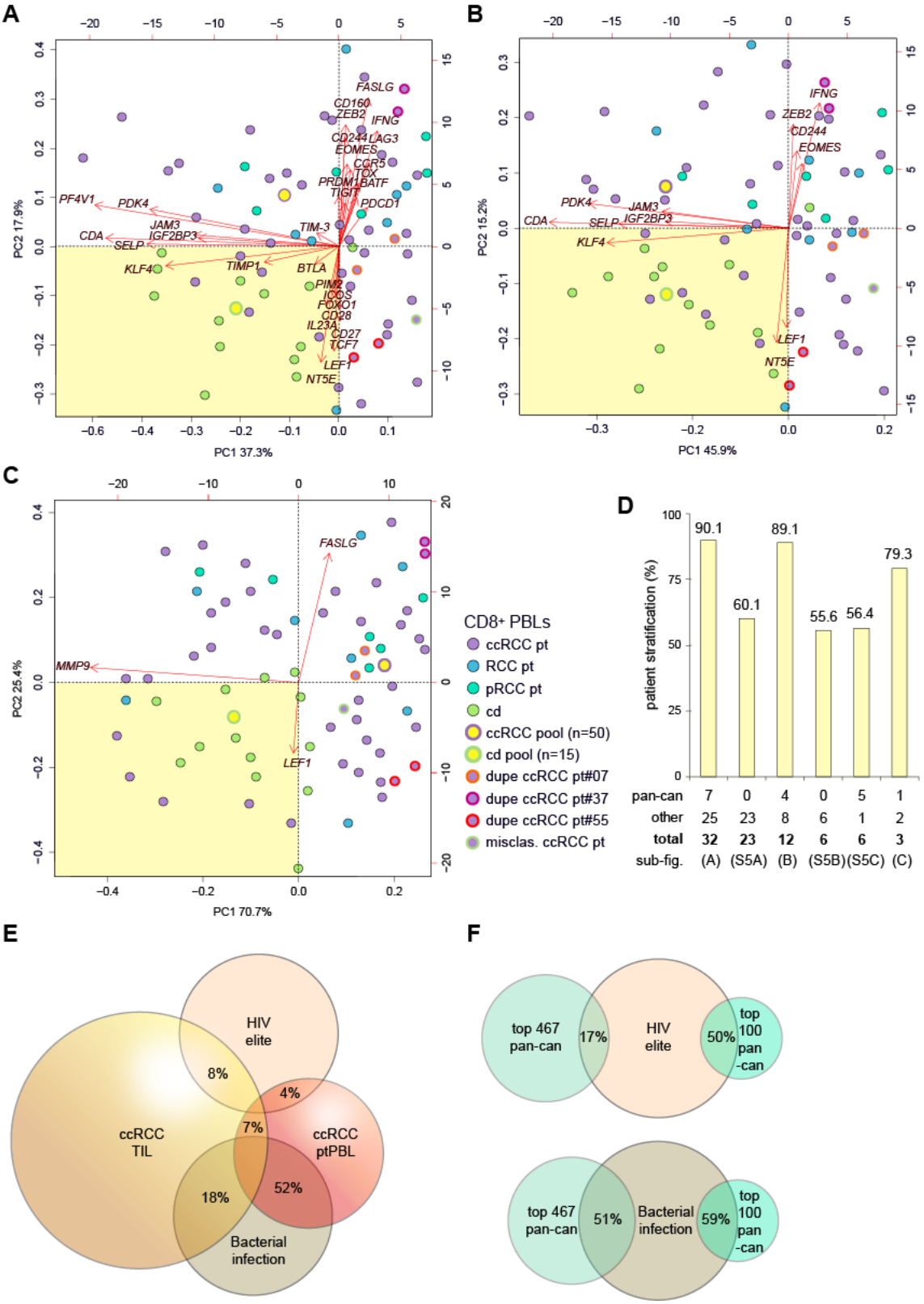
A
PC1 37.3%
PC2 17.9%
FASLG
CD160
ZEB2
IFNG
CD244
LAG3
EOMES
CCR5
TOX
PRDM1
BATF
TIGIT
PDCD1
PF4V1
PDK4
JAM3
IGF2BP3
TIM-3
CDA
SELP
KLF4
TIMP1
BTLA
PIM2
ICOS
FOXO1
CD28
IL23A
CD27
TCF7
LEF1
NT5E
B
PC1 45.9%
PC2 15.2%
IFNG
ZEB2
CD244
EOMES
PDK4
JAM3
IGF2BP3
CDA
SELP
KLF4
LEF1
NT5E
C
PC1 70.7%
PC2 25.4%
FASLG
MMP9
LEF1
CD8+ PBLs
ccRCC pt
RCC pt
pRCC pt
cd
ccRCC pool (n=50)
cd pool (n=15)
dupe ccRCC pt#07
dupe ccRCC pt#37
dupe ccRCC pt#55
misclas. ccRCC pt
D
patient stratification (%)
90.1
60.1
89.1
55.6
56.4
79.3
pan-can 7 0 4 0 5 1
other 25 23 8 6 1 2
total 32 23 12 6 6 3
sub-fig. (A) (S5A) (B) (S5B) (S5C) (C)
E
HIV elite
ccRCC TIL
ccRCC ptPBL
Bacterial infection
8%
4%
7%
18%
52%
F
top 467 pan-can
17%
HIV elite
50%
top 100 pan -can
top 467 pan-can
51%
Bacterial infection
59%
top 100 pan -can

Figure 6. Iterative DEG combination testing defining minimal gene sets required for stratifying patients from control donors according to circulating CD8$^+$ T cells. Normalized -ΔCT qRT-PCR DEG expression values from individual and pooled CD8$^+$ ptPBLs and cdPBLs were used for PCA using applying the Euclidean distance metric and complete linkage clustering method (R programming language; R-studio) (*n* = 69). (**A**) Patients are stratified using 32 DEGs including pan-cancer (*ICOS*, *PF4V1*, *IFNG*, *LAG3*, *TIGIT*, *CDA*, *PDK4*, *KLF4*, *PIM2*, *TIMP1*, *IGF2BP3*, *IL23A*, *LEF1*, *TCF7*), T cell polarizing (*FASLG*, *ZEB2*, *EOMES*, *CCR5*, *TOX*, *PRDM1*, *BATF*, *FOXO1*, *CD28*, *CD27*), adhesion (*JAM3*, *SELP*), and immune checkpoint DEGs (*CD160*, *CD244*, *PDCD1*, *TIM-3*, *BTLA*, *NT5E*). (**B**) Patients are stratified using 12 DEGs including pan-cancer (*CDA*, *PDK4*, *KLF4*, *IGF2BP3*) and adhesion (*JAM3*, *SELP*) DEGs. (**C**) Patients are stratified using three DEGs (pan-cancer, *MMP9* and *LEF1*; T cell polarizing, *FASLG*). Pale yellow background boxes highlight PCA stratified control donors used to calculate percent patient stratification. (**D**) Graph representing percent patient stratification from DEG groups in (**A-C**), and in **Supplemental Figure S5**, with representative numbers of pan-cancer genes among groups at bottom (*n* = 66, non-duplicate samples). (**E-F**) Venn diagrams demonstrating overlaps between (**E**) CD8$^+$ ccRCC ptPBL DEGs, CD8$^+$ ccRCC TIL DEGs, CD8$^+$ HIV elite controllers, and PBMC from bacteria infected patients, and (**F**) effect of pan-cancer pipeline on enhancing CD8$^+$ DEG identity. PBL, peripheral blood lymphocyte; pt, patient; nd, normal donor; TIL, tumor infiltrating lymphocytes; dupe, duplicate sample; misclas., misclassified benign kidney lesion; ccRCC, clear cell renal cell carcinoma; pRCC, papillary renal cell carcinoma; RCC, renal cell carcinoma; n, number of pooled samples; pan-can, pan-cancer; other, other DEGs; sub-fig., associated sub-figure.

Figure 7

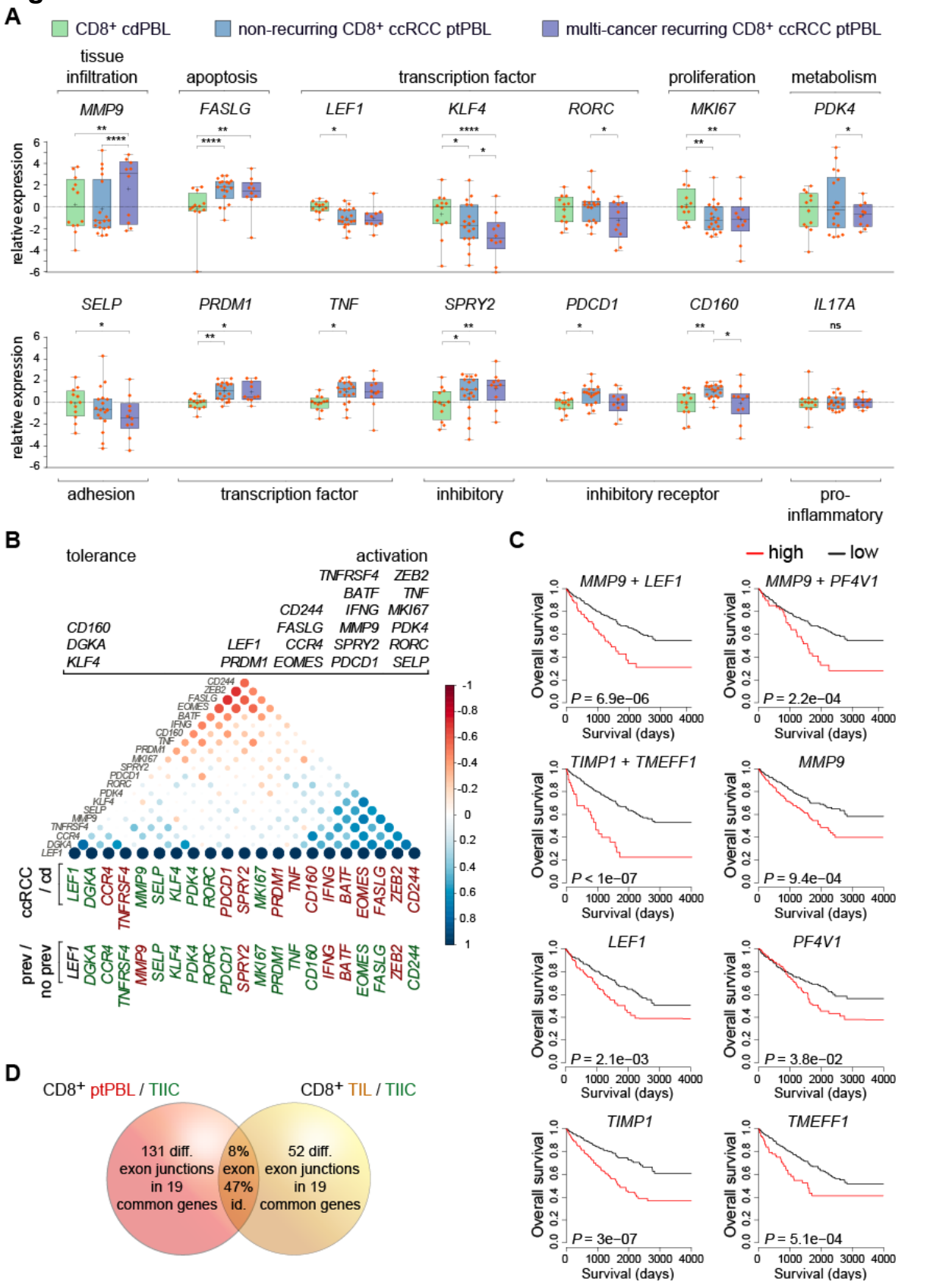
A
CD8+ cdPBL
non-recurring CD8+ ccRCC ptPBL
multi-cancer recurring CD8+ ccRCC ptPBL
tissue infiltration
apoptosis
transcription factor
proliferation
metabolism
MMP9
FASLG
LEF1
KLF4
RORC
MKI67
PDK4
relative expression
SELP
PRDM1
TNF
SPRY2
PDCD1
CD160
IL17A
ns
adhesion
transcription factor
inhibitory
inhibitory receptor
pro-inflammatory
B
tolerance
activation
CD160
DGKA
KLF4
LEF1
PRDM1
CCR4
EOMES
PDCD1
CD244
FASLG
MMP9
SPRY2
RORC
SELP
TNFRSF4
BATF
IFNG
ZEB2
TNF
MKI67
PDK4
ccRCC / cd
prev / no prev
C
high
low
MMP9 + LEF1
P = 6.9e−06
MMP9 + PF4V1
P = 2.2e−04
TIMP1 + TMEFF1
P < 1e−07
MMP9
P = 9.4e−04
LEF1
P = 2.1e−03
PF4V1
P = 3.8e−02
TIMP1
P = 3e−07
TMEFF1
P = 5.1e−04
Overall survival
Survival (days)
D
CD8+ ptPBL / TIIC
CD8+ TIL / TIIC
131 diff. exon junctions in 19 common genes
8% exon 47% id.
52 diff. exon junctions in 19 common genes

Figure 7. Additive prognostic pan-cancer DEGs stratify multi-cancer recurring ccRCC patients having activated CD8$^+$ T cell profiles. (**A**) DEGs from validation cohort were compared among cdPBL (*n* = 12), and ptPBL with (*n* = 10) or without (*n* = 18) recurring multi-cancers. *P*, two-way ANOVA, Tukey post-test; *, $P < 0.05$; **, $P < 0.01$; ****, $P < 0.0001$; boxes, upper and lower quartiles; whiskers, all points maxima to minima; +, mean; line, median. Functional classifications of DEG groups are listed above and below, and the literature was used to (**B**) segregate DEGs according to tolerance or activation phenotypes. Correlograms (Spearman method) using normalized -ΔCT qRT-PCR expression values for visualization of two groups of pan-cancer and T cell polarizing DEGs, with differences observed between all ccRCC patients vs. control donors, and patients with vs. without recurring cancers (Student T test, $P < 0.05$) (red, increased expression; green, decreased expression). Only *MMP9* is significantly increased in multi-cancer patients relative to all others. (**C**) Pan-cancer DEG combinations tested for additive prognostic effects using TCGA KIRK dataset. Only *MMP9*, *LEF1*, *PF4V1*, *TIMP1*, and *TMEFF1* demonstrate additive prognostic effects, and these cluster in correlograms (as above) enquiring pan-cancer DEGs with combinatorial effects on prognosis. Kaplan-Meier plots *P*, log-rank. (**D**) Venn diagram illustrating that ptPBLs have more differentially represented exon-exon probe selected regions (PSR) junctions relative to TILs; both relative to TIICs ($P < 0.05$; ANOVA, Transcriptome Analysis Console v.3, Affymetrix), with 8% overlap of total PSR junctions between ptPBLs and TILs, and 47% of all pan-cancer DEGs having shared ptPBL and TIL PSR junction identity (see **Supplemental Table S8**). tot, total; ccRCC, clear cell renal cell carcinoma; ptPBL, patient peripheral blood lymphocyte; cdPBL, control donor peripheral blood lymphocyte; TIL, tumor infiltrating lymphocytes; TIIC, normal adjacent tissue infiltrating immune cells; diff., differential; id., identity.

Figure 8

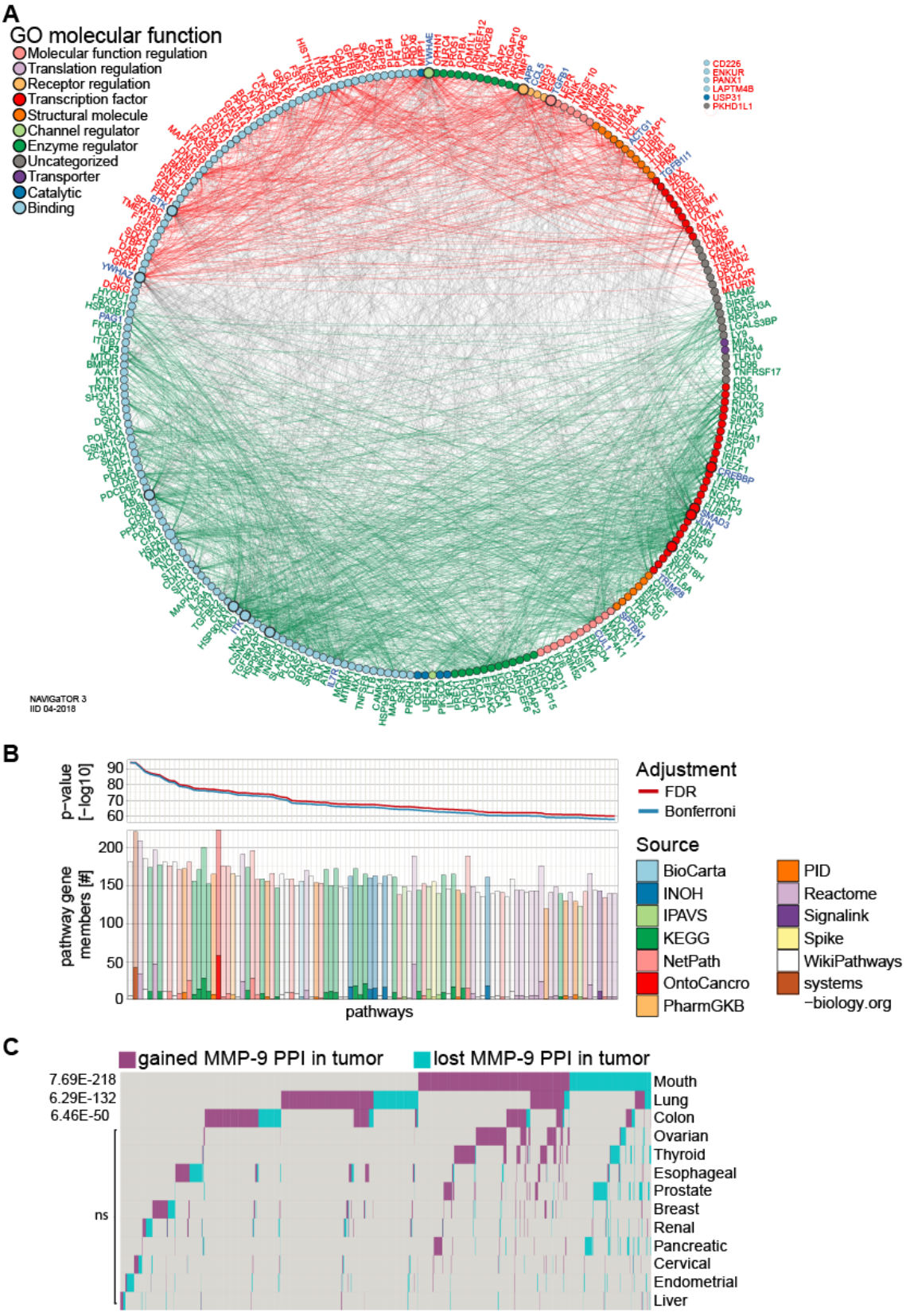
A
GO molecular function
Molecular function regulation
Translation regulation
Receptor regulation
Transcription factor
Structural molecule
Channel regulator
Enzyme regulator
Uncategorized
Transporter
Catalytic
Binding
CD226
ENKUR
PANX1
LAPTM4B
USP31
PKHD1L1
NAViGaTOR 3
IID 04-2018
B
p-value [−log10]
90
80
70
60
pathway gene members [#]
200
150
100
50
0
pathways
Adjustment
FDR
Bonferroni
Source
BioCarta
INOH
IPAVS
KEGG
NetPath
OntoCancro
PharmGKB
PID
Reactome
Signalink
Spike
WikiPathways
systems −biology.org
C
gained MMP-9 PPI in tumor
lost MMP-9 PPI in tumor
7.69E-218
6.29E-132
6.46E-50
ns
Mouth
Lung
Colon
Ovarian
Thyroid
Esophageal
Prostate
Breast
Renal
Pancreatic
Cervical
Endometrial
Liver

Figure 8. Enrichment of disrupted MMP9 pathways in ccRCC patient circulating cells and various cancers. (**A**) Protein-protein interaction (PPI) network linking pan-cancer proteins from significant MMP-9 pathways-associated ccRCC ptPBL DEGs. DEGs (nodes/circles), and their interactions (edges/lines) are colored red (highly expressed) and green (lowly expressed), and grey edges highlight interactions between them (NAViGaTOR v3 and IID v04-2018). Non-interacting proteins are listed on top right. DEG nodes are colored according to Gene Ontology (GO) Molecular Functions listed in legend. Larger node circles, represent high degree of interactions with all other DEGs, and blue DEG names represent centrality of interactors. (**B**) Pathway enrichment analysis graphs depicting results of pathDIP analysis for MMP-9 pathway interactors from correlation analyses. Upper panel shows significance of enrichment obtained for individual pathways (p-value, -log10) adjusted for multiple testing using FDR and Bonferroni methods. Lower barplot shows overlap between query genes and members of individual pathways. Respective numbers of known and predicted pathway members are distinguished by opacity, and fill color indicates source of given pathway. Plots are restricted to top 100 most significant (full pathways in **Supplemental Figure 7A**). (**C**) Tissue-specific disrupted PPI networks among MMP-9 interactors in cancer. Gained and lost MMP-9 PPIs in thirteen epithelial tissues-cancers highlighting its tissue-specific role in cancer, where 106 disrupted MMP-9 PPIs were identified (n = 2,801) (see **Supplemental Figure S7**). 1,814 disrupted PPIs were found, 81% of which are disrupted in only one or two tissues and only less than 5% are present in more than three tissues.

Figure 9

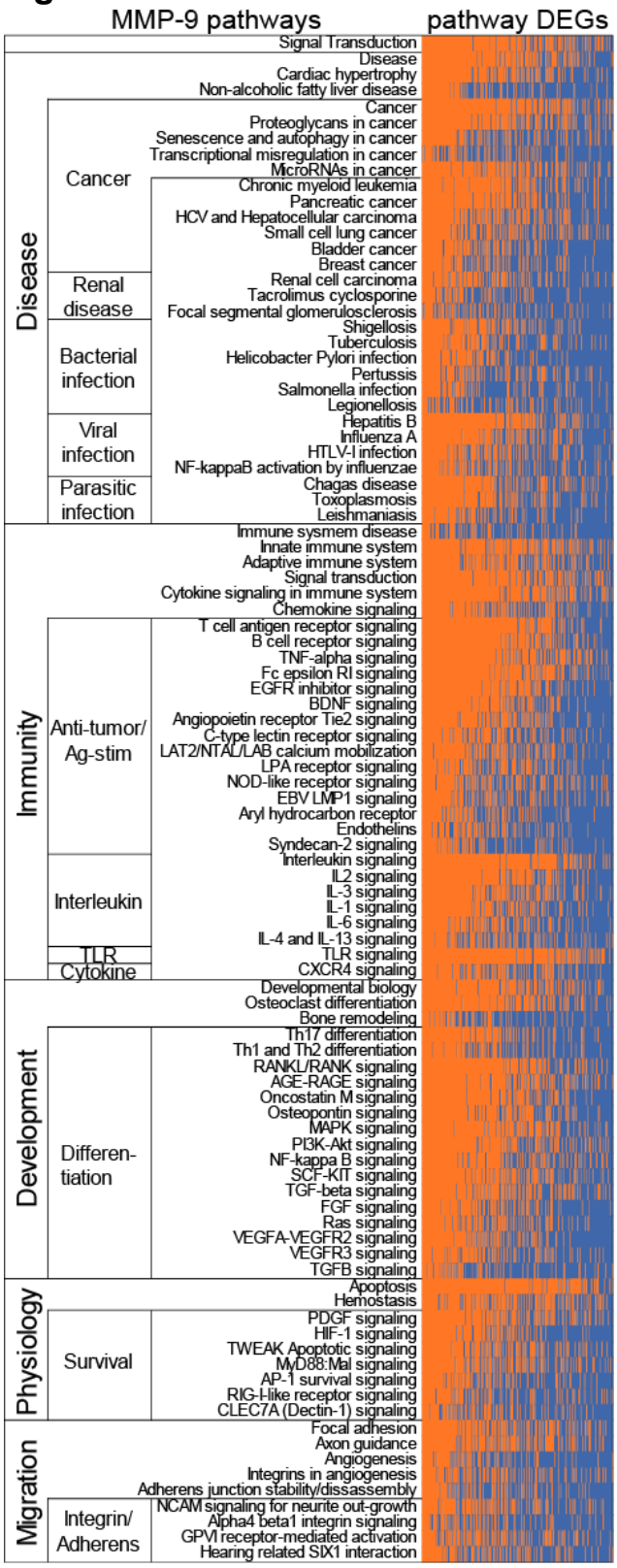
MMP-9 pathways
pathway DEGs
Signal Transduction
Disease
Cardiac hypertrophy
Non-alcoholic fatty liver disease
Cancer
Proteoglycans in cancer
Senescence and autophagy in cancer
Transcriptional misregulation in cancer
MicroRNAs in cancer
Chronic myeloid leukemia
Pancreatic cancer
HCV and Hepatocellular carcinoma
Small cell lung cancer
Bladder cancer
Breast cancer
Renal cell carcinoma
Tacrolimus cyclosporine
Focal segmental glomerulosclerosis
Shigellosis
Tuberculosis
Helicobacter Pylori infection
Pertussis
Salmonella infection
Legionellosis
Hepatitis B
Influenza A
HTLV-I infection
NF-kappaB activation by influenzae
Chagas disease
Toxoplasmosis
Leishmaniasis
Immune sysmem disease
Innate immune system
Adaptive immune system
Signal transduction
Cytokine signaling in immune system
Chemokine signaling
T cell antigen receptor signaling
B cell receptor signaling
TNF-alpha signaling
Fc epsilon RI signaling
EGFR inhibitor signaling
BDNF signaling
Angiopoietin receptor Tie2 signaling
C-type lectin receptor signaling
LAT2/NTAL/LAB calcium mobilization
LPA receptor signaling
NOD-like receptor signaling
EBV LMP1 signaling
Aryl hydrocarbon receptor
Endothelins
Syndecan-2 signaling
Interleukin signaling
IL2 signaling
IL-3 signaling
IL-1 signaling
IL-6 signaling
IL-4 and IL-13 signaling
TLR signaling
CXCR4 signaling
Developmental biology
Osteoclast differentiation
Bone remodeling
Th17 differentiation
Th1 and Th2 differentiation
RANKL/RANK signaling
AGE-RAGE signaling
Oncostatin M signaling
Osteopontin signaling
MAPK signaling
PI3K-Akt signaling
NF-kappa B signaling
SCF-KIT signaling
TGF-beta signaling
FGF signaling
Ras signaling
VEGFA-VEGFR2 signaling
VEGFR3 signaling
TGFB signaling
Apoptosis
Hemostasis
PDGF signaling
HIF-1 signaling
TWEAK Apoptotic signaling
MyD88:Mal signaling
AP-1 survival signaling
RIG-I-like receptor signaling
CLEC7A (Dectin-1) signaling
Focal adhesion
Axon guidance
Angiogenesis
Integrins in angiogenesis
Adherens junction stability/dissassembly
NCAM signaling for neurite out-growth
Alpha4 beta1 integrin signaling
GPVI receptor-mediated activation
Hearing related SIX1 interaction
Disease
Cancer
Renal disease
Bacterial infection
Viral infection
Parasitic infection
Immunity
Anti-tumor/ Ag-stim
Interleukin
TLR
Cytokine
Development
Differen-tiation
Physiology
Survival
Migration
Integrin/ Adherens

Figure 9. MMP-9 pathway DEGs are linked to a variety of renal diseases, cancers, microbial infections and immune activation. MMP-9 pathways associated ptPBL DEGs from pathDIP matrix were correlated, and DEGs significantly associated with MMP-9 pathways ($P < 0.05$) were used to generate supervised heatmap of most highly correlating MMP-9-pathway genes and associated pathways. Majority of FIR DEG-associated MMP-9 pathways are enriched for those of renal diseases, numerous viral, bacterial and parasitic infections, numerous cancers, and immunity and antigen recognition and activation; differentiation, and cellular survival pathways. Orange, DEG represented in pathway; blue DEG not represented in pathway (full pathway DEGs list in **Supplemental Figure S9**).