



Published in final edited form as:

Cancer Discov. 2018 September ; 8(9): 1142–1155. doi:10.1158/2159-8290.CD-17-1246.

An Empirical Approach Leveraging Tumorgrafts to Dissect the Tumor Microenvironment in Renal Cell Carcinoma Identifies Missing Link to Prognostic Inflammatory Factors

Tao Wang, Ph.D.^{1,2,3,*}, Rong Lu, Ph.D.^{#1,4}, Payal Kapur, M.D.^{#2,5}, Bijay S. Jaiswal, Ph.D.^{#6}, Raquibul Hannan, M.D., Ph.D.^{2,7}, Ze Zhang, B.S.¹, Ivan Pedrosa, M.D.^{2,8}, Jason J. Luke, M.D., FACP⁹, He Zhang, Ph.D.⁴, Leonard D Goldstein, Ph.D.⁶, Qurratulain Yousuf, B.S.^{2,10}, Gu Yi-Feng, Ph.D.^{2,10}, Tiffani McKenzie, B.S.², Allison Joyce, B.S.^{2,10}, Min S Kim, Ph.D.^{2,4}, Xinlei Wang, Ph.D.¹¹, Danni Luo, M.S.⁴, Oreoluwa Onabolu, B.S.^{2,10}, Christina Stevens, B.S.^{2,10}, Zhiqun Xie, Ph.D.¹, Mingyi Chen, M.D., Ph.D.⁵, Alexander Filatenkov, M.D.⁵, Jose Torrealba, M.D.⁵, Xin Luo, Ph.D.⁴, Wenbin Guo¹, Jingxuan He, B.S.¹, Eric Stawiski, Ph.D.^{6,12}, Zora Modrusan, Ph.D.⁶, Steffen Durinck, Ph.D.^{6,12}, Somasekar Seshagiri, Ph.D.^{6,*}, and James Brugarolas, M.D., Ph.D.^{2,10,*}

¹Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.

²Kidney Cancer Program, Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.

³Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.

⁴Bioinformatics Core Facility, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.

⁵Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.

⁶Molecular Biology Department, Genentech, Inc., South San Francisco, CA, USA, 94080.

⁷Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.

⁸University of Texas Southwestern Medical Center, Dallas, TX, USA, 75390.

***Corresponding authors:** Tao Wang, Ph.D., Quantitative Biomedical Research Center, Department of Clinical Sciences, Center for the Genetics of Host Defense, Harold C. Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, TX, 75390, USA; Phone: 214-648-4082; Tao.Wang@UTSouthwestern.edu, James Brugarolas, M.D., Ph.D., Department of Internal Medicine & Kidney Cancer Program, Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas, 75390, USA; Phone: 214-648-4059; James.Brugarolas@UTSouthwestern.edu, Somasekar Seshagiri, Ph.D., Molecular Biology Department, Genentech, Inc., South San Francisco, CA, USA, 94080; Phone: 650-225-3351; seshagiri.somasekar@gene.com.

Conflict of interest: The authors declare no conflict of interest

This study was supported by the National Institutes of Health (NIH) [R03 ES026397-01/TW, R01 GM110050-01A1/MC, R01 CA175754/JB, SPORE 1P50CA19651601/PK, JB, TW, IP, R01 CA154475/IP, CCSG 5P30CA142543/TW], Center for Translational Medicine of UT Southwestern [SPG2016-018/TW], Cancer Prevention and Research Institute of Texas [CPRIT RP150596/MK, HZ, RL, DL], and American Cancer Society [RSG-16-001-01-CCE/RH]. This work was also partially supported by fundraising efforts orchestrated by the KCP Patient Council and the Kidney Cancer Coalition.

⁹Department of Medicine, University of Chicago, Chicago, IL, USA, 60637.

¹⁰Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA

¹¹Department of Statistical Science, Southern Methodist University, Dallas, Texas, USA

¹²Bioinformatics and Computational Biology Department, Genentech, Inc., South San Francisco, CA, USA, 94080.

These authors contributed equally to this work.

Abstract

By leveraging tumorgraft (PDX) RNA-Seq data, we developed an empirical approach, DisHet, to dissect the tumor microenvironment (eTME). We found that 65% of previously defined immune signature genes are not abundantly expressed in Renal Cell Carcinoma (RCC), and identified 610 novel immune/stromal transcripts. Using eTME, genomics, pathology and medical record data involving >1,000 patients, we established an inflamed pan-RCC subtype (IS) enriched for T_{regs}, NK cells, T_{H1} cells, neutrophils, macrophages, B cells, and CD8⁺ T cells. IS is enriched for aggressive RCCs, including *BAP1*-deficient clear-cell and type 2 papillary tumors. The IS subtype correlated with systemic manifestations of inflammation such as thrombocytosis and anemia, which are enigmatic predictors of poor prognosis. Furthermore, IS was a strong predictor of poor survival. Our analyses suggest that tumor cells drive the stromal immune response. These data provide a missing link between tumor cells, the TME, and systemic factors.

Statement of significance: We undertook a novel empirical approach to dissect the renal cell carcinoma tumor microenvironment by leveraging tumorgrafts. The dissection and downstream analyses uncovered missing links between tumor cells, the tumor microenvironment, systemic manifestations of inflammation, and poor prognosis.

Keywords

tumor heterogeneity; dissection; renal cell carcinoma; inflammation; immunogenicity

INTRODUCTION

The interplay between a tumor and its microenvironment influences a variety of processes, including invasion, metastasis and response to therapy. However, much still remains unknown about this interplay, and the triggered inflammatory response, which could involve immune cells from both the innate and adaptive immune systems. To understand the intricate interactions between tumor and the microenvironment, it is helpful to identify and characterize each cell population present in bulk tumor. One school of *in silico* approaches relies on signature genes and a backend deconvolution method, such as single sample Gene Set Enrichment Analysis (ssGSEA), an algorithm that correlates particular cell types with a set of signature genes (1). Indeed, several sets of gene signatures have been produced (2–4). However, these signatures suffer from poor definition of immune-specific genes, which can be affected by the type of cancer these cells infiltrate or, when derived from laser microdissection approaches, by technically challenging sample processing (5). Newer

approaches involve single cell sequencing, but they are still quite expensive and also require extensive sample handling, which can affect gene expression (6).

Therefore, improved approaches are essential for more accurate dissection of the heterogeneous cell populations in tumors. To generate such gene signatures with empirical evidence, it is helpful to obtain the *in vivo* expression of immune cells infiltrating bulk tumors by dissection of bulk tumor expression profiles. However, there are two major problems associated with current tumor dissection studies. First, in addition to the immune cell and the tumor cell components, normal parenchyma may contaminate the bulk tumor specimen. Thus, earlier two-component dissection models (7,8) may be insufficient. Second, tumor cell line cultures, which have been used in some studies as surrogates of the tumoral component (9,10), are subject to significant phenotypic drift from the original tumors (11,12) and are not optimal.

Here, we undertook a novel approach to dissect the tumor cells and the tumor microenvironment in renal cancer by leveraging tumorgrafts. Tumorgrafts, or patient derived xenografts (PDX), represent original tumor cell populations transplanted directly from a patient into a host, typically a mouse. In comparison with cell line-based xenografts, tumorgrafts more faithfully reproduce the histology, gene expression, mutations, and treatment responsiveness of individual human tumors (13–15). Despite recent reports questioning the validity of tumorgrafts (16), the reported limitations only affect certain tumor types subject to genome duplication over long passages. Importantly, in the tumorgrafts, only the human tumor cells expand, and the tumor stroma is soon replaced by the host (13,15,17). Thus, inasmuch as human transcripts can be distinguished from the mouse transcripts (through alignment to both reference genomes), tumorgrafts can be used to specifically define the tumor cell transcriptome. By subtracting this transcriptome and the normal tissue transcriptome from the corresponding patient tumor through a three-component dissection analysis, a gene expression profile may be generated corresponding to the immune/stroma component, namely the tumor microenvironment (TME) component.

We applied this approach to Renal Cell Carcinoma (RCC). RCC include clear cell RCC (ccRCC), the most common subtype accounting for 70% of all RCCs; papillary RCC (pRCC); and chromophobe RCC (chRCC) (18). RCC is a paradigmatic tumor type characterized by a rich inflammatory stroma (19,20). Interestingly, several extensively validated prognostic factors in RCCs (21,22) are also indicators of systemic inflammation, such as thrombocytosis, anemia, and neutrophilia. However, a detailed description of how these extensively validated RCC prognostic factors are linked to the biology of the tumor cells and the tumor microenvironment is lacking. RCC is also one of two tumors traditionally regarded as immune-responsive (23–25). Two other tumors considered as immunogenic include melanoma and, recently, lung cancer. Whereas melanoma and lung cancer stand out by their high mutation burdens, RCCs exhibit a modest overall mutational load (26). However, a recent study reported that RCCs have the highest number of indel mutations on a pan-cancer basis (19), which may yield a disproportionately higher amount of neoantigens and elicit robust adaptive immune responses.

Here, we set out to evaluate the tumor microenvironment in RCC using DisHet, a three-component dissection algorithm we developed. Based on DisHet analyses, we obtained an empirically-defined tumor microenvironment gene expression signature (eTME). Using eTME, we refined previously published signatures and assigned other eTME genes to specific immune cell types using the BLUEPRINT project data. Based on expression of eTME genes, we defined an inflamed subtype (eTME-IS), which is correlated with systemic manifestations of inflammation such as thrombocytosis and anemia. eTME-IS patients have abundant tumor infiltration of T_{regs}, NK cells, T_{h1} cells, neutrophils, macrophages, B cells, and CD8⁺ T cells, and are enriched for clear-cell tumors with *BAP1* mutations and papillary tumors of type 2 histology. Our analyses also show that tumor cells could actively modulate the immune response in the tumor microenvironment.

RESULTS

Leveraging tumorgrafts to dissect bulk tumor RNA-Seq data into different cellular components

We explored whether tumorgrafts could be used to dissect the immune/stroma component's gene expression through subtraction from the corresponding patient tumor. First, we evaluated whether tumorgrafts are purely comprised of tumor cells. We leveraged the UT Southwestern (UTSW) Kidney Cancer Program (KCP) platform, and identified 35 patients (29 ccRCC and 6 non-ccRCC) for whom we have carried out RNA-Seq for one bulk tumor, one tumorgraft, and one normal tissue sample (**Sup. Table 1**). We profiled the expression of genes from the ESTIMATE signature (2), a signature made up of stromal and immune genes. Overall, we found that ESTIMATE gene expression was much lower in the human cell component of tumorgrafts when compared with the bulk tumor (**Fig. 1a**; note log scale). As a control, we evaluated three previously identified RCC tumor-specific genes, *FABP7* (27), *PTHLH* (28), and *CD70* (29). As shown in **Fig. 1a**, all three transcripts were expressed at high levels in both the bulk tumor and the human component of the tumorgrafts. Next, we identified 38 RCC patients for whom we have carried out exome-seq for at least one tumor, one tumorgraft and one normal sample (27 overlapping with the 35 previously mentioned patients). We calculated the average somatic variant allele frequencies (VAFs) in both tumors and tumorgraft samples and found that, as we previously showed (15), VAFs generally increased in the tumorgrafts (**Fig. 1b**). As only the human tumor cells carry these somatic mutations, these data are consistent with the notion that only the tumor cell population expands in mice and that the stroma is replaced by the host.

One premise of our dissection analysis is that the gene expression of tumorgrafts does not shift significantly from those of the original human tumors. We carried out RNA-Seq for one patient (XP389) for whom we have obtained primary tumors, and have established one tumorgraft and one tumor cell line. We calculated the proportions of genes whose expression deviation in tumorgrafts or cell culture from primary tumors are larger than a series of cutoffs. We calculated this proportion with respect to all genes, or to genes whose expression in tumorgraft and cell culture are both larger than that in the primary tumors (such genes are more likely to be the tumor-specific genes). **Fig. 1c** shows that although there are still some differences between tumorgrafts and the tumor cells in primary tumors, the phenotypic drift

in tumorgrafts is smaller than that in cell culture. These data are consistent with our previous report that the majority of tumorgrafts cluster with the corresponding patient tumor in unsupervised tumor analyses (13). Ben-David *et al* (16) recently reported significant clonal evolution in tumorgrafts over passages through analyses of copy number variation (CNV) data in multiple tumor types. However, it is worth noting that the drift applied almost exclusively to tumors having undergone genome duplication, which is a rare event in RCC. We also show the whole genome CNVs in the primary tumor and tumorgraft of XP166 (11th passage) and XP243 (1st passage). In most parts of the genomes of both early and late passages of tumorgrafts, the CNVs are in a direction (amplification or loss) consistent with that of the primary tumor (**Fig. 1d**). This suggests that the genomic instability in RCC tumorgrafts is relatively small, even in late passages. The CNVs in the tumorgraft are generally larger in amplitude than in primary tumors, which is expected, as tumorgrafts have a higher representation of human tumor cells and the CNVs are not masked by non-tumor cells. In addition, one other advantage of tumorgrafts is that they allow rapid sample processing and minimize ischemia time. On the other hand, approaches like Laser Capture Microdissection (LCM) and single cell analyses would allow direct measurement of cells in the stroma and tumor tissues, but is limited by preanalytical variables such as ischemia time.

Building on the previous analysis, we subtracted from the bulk tumor gene expression, tumor cell and normal parenchyma expression using tumorgraft and adjacent normal tissue RNA-Seq data. To meet the analytic needs, we developed DisHet, for Dissecting Heterogeneous bulk tumors, an algorithm based on a Bayesian Hierarchical model. The dissection task can be expressed as solving for $Ax=b$, where A is a matrix of component-specific expression, x is a vector of mixture proportions and b is a vector of bulk tissue expressions. The three components include tumor cells, immune/stroma, and normal parenchyma. Importantly, the immune/stroma component further contains various types of immune cells, activated fibroblasts, and other cell types such as endothelial cells and pericytes. The bulk tumor expression (b) and the normal and tumor components of the A matrix are observed from the bulk tumor, normal tissue, and tumorgraft RNA-Seq data, respectively, while the goal is to estimate the third unknown component in A . This concept is shown in **Fig. 1e**. The bulk sample expression is determined by both the amount and expression of each cellular component, both of which can be estimated by DisHet. The **Supplementary File** explains the full technical details of DisHet.

DisHet accurately dissects component proportions and immune/stroma component expression of bulk tumors

To evaluate DisHet, we initially performed simulation analyses. As shown in **Sup. Fig. 1a**, the component proportions estimated by DisHet are almost the same as the simulated true proportions. **Sup. Fig. 1b** shows that DisHet can accurately estimate the average expression of each gene across all patients. Finally, DisHet can accurately estimate the individual variation of each gene by calculating a Spearman correlation between the simulated expression for each gene across all patients and the estimated expression by DisHet in that gene across all individuals. We show that most of these correlations are very close to 1 (**Sup. Fig. 1c**), confirming the accuracy of DisHet in capturing the individual expression variation.

We applied DisHet on the real data from the 35 RCC patients for which matched bulk tumor, tumorgraft and normal tissue RNA-Seq were conducted. As a validation, pathologic analyses of sections flanking the analyzed tumor sample for tumor proportions correlate well (Spearman correlation=0.65, $p=0.0004$) with estimations of DisHet (**Fig. 2a**). Secondly, there is a strong positive correlation (Spearman correlation=0.72, $p=2.5\times 10^{-5}$) between the predicted tumor proportions with average VAFs of somatic mutations (**Fig. 2b**). Finally, ssGSEA was applied to the combined ESTIMATE stromal and immune cell genes. We observed a positive correlation (Correlation=0.88, $P<1\times 10^{-10}$) between the ESTIMATE enrichment score and the immune/stroma component proportions predicted by DisHet (**Fig. 2c**).

In addition to component proportions, we also validated the dissected immune/stroma component expression. We applied ssGSEA on the dissected immune/stroma component expression with B and T cell signatures from the Immunome immune cell signatures (4). We found that patients with a higher level of lymphocyte infiltration, according to H&E slides reviewed by our pathologist, did in fact have higher ssGSEA scores for T cell and B cell genes (**Fig. 2d**). We also applied ssGSEA using the 16 Winslow signatures defined in breast cancer (3) with genes in the same signature being highly correlated. We showed that genes in the same signature are indeed highly correlated in the expression of the immune/stroma component (**Sup. Fig. 2a**). Lastly, we performed DisHet analyses of 9 RNA-Seq datasets from 4 unique patients, each with more than one set of matched bulk tumor, tumorgraft, and normal tissue samples available. We showed that these 9 samples largely clustered by each patient according to their immune/stroma component expression (**Sup. Fig. 2b**). Overall, these extensive analyses attest to the accuracy of the dissected gene expression across individuals.

We compared the performance of DisHet with another popular algorithm, DeMix (8), for gene expression deconvolution in a mixture of cell types. However, DeMix considers only two components: tumor and stroma, with the stroma component being the sum of the normal parenchyma and immune/stroma components defined by DisHet. DisHet shows much better dissection accuracy than DeMix (compare **Sup. Fig. 3ab** with **Fig. 2ac**). DisHet is also much more computationally efficient than DeMix (**Sup. Fig. 3c**).

Identifying novel immune/stroma-specific genes with empirical evidence and redefining immune cell-specific signatures

The matched trio of bulk tumor, tumorgraft, and normal tissue RNA-Seq data and DisHet analyses enabled us to define immune/stroma-specific genes with empirical evidence. We created one immune/stroma-specific gene list with 2,080 genes that are expressed on average at least 3 times higher in the immune/stroma component than in the tumor, and another with 904 genes expressed at least 20 times higher (**Sup. Table 2**). The false discovery rates (FDRs) are estimated to be 4.7% for the first set and 0.0044% for the second set based on 500 permutations. We termed these immune/stroma-specific signature genes eTME, for empirically-defined gene signature of the tumor microenvironment (immune/stroma). We investigated the relationship of eTME genes with previously published signatures. To our surprise, 65% of the Immunome signature genes (4) are not abundantly expressed in the

RCC immune/stroma component even according to the less stringent cutoff of $3\times$ (**Fig. 3a**). In addition, 22% of the ESTIMATE genes and 9% of the Winslow genes are also not abundantly expressed in the RCC immune/stroma component. Although we applied these gene signatures in previous sections for validating the gene expression dissected by DisHet, they are not optimal, and further refinement filtering out non-specific genes would be fitting. Conversely, among the 904 eTME genes expressed at $>20\times$ higher levels, we identified 778 genes without a clear interpretation in the Immunome signatures. Furthermore, even after excluding genes in ESTIMATE and Winslow signatures, there were 610 immune/stroma-specific genes that were novel.

The eTME genes may hold clues to the kidney cancer inflammatory environment and may help explain why renal cancer is highly immunogenic. To investigate the identity of these genes, we carried out gene ontology analyses for the 904 eTME genes. As expected, there was a significant enrichment for immune pathways in these genes (**Sup. Table 3**). Dominant enrichment of immune-related pathways are even seen in the subset of 778 and 610 eTME genes that are not captured by the Immunome alone, or Immunome+ESTIMATE+Winslow signatures (**Sup. Table 3**). eTME also contain a small fraction of genes related to non-immune pathways/cell types. For example, fibroblast-related genes such as FAP (30) and POSTN (31) are found in the set of 904 eTME genes. This is also reflected by the extracellular matrix-related GO terms that are weakly enriched in eTME (**Sup. Table 3**). To derive a more specific interpretation for each gene, we evaluated the gene expression data of hematopoietic lineages (BLUEPRINT, Expression Atlas; <http://www.ebi.ac.uk/gxa/home>) and called signature genes by requiring their expression to be at least 2 times larger than any of the other 33 lineages, and conducting a Kruskal-Wallis test on signature gene sets requiring P value <0.001 . For these analyses we used the eTME genes with 3 fold cutoff. In this way, we assigned cell type identities to 293 genes, including 231 genes missed by the previous signatures (**Sup Table 2**). Despite the use of multiple signatures, more than 50% of the eTME genes remain unassigned. While these genes may be expressed in more than one lineage, they may also carry information about novel cell types or differentiation states.

To validate the eTME gene signatures, we carried out single cell RNA-Seq for five RCC patients without treatments prior to tumor resection. The lymphoid and myeloid cells were separately enriched from these five patients by FACS (using CD45 positivity and separated based on scatter properties), generating two pools of cells ($n=482$ for lymphoid and $n=487$ for myeloid, all patients combined), which were sequenced separately. We first aggregated single cell gene expression data and normalized each gene by expression of the same gene in the tumor component, to derive a ratio similar to that in **Fig. 3a**. We plotted these ratios (logged) for “Non-eTME” genes - genes not belonging to the eTME gene signature ($3\times$ cutoff), and eTME genes defined with $3\times$ and $20\times$ fold cutoffs. In **Fig. 3b**, we show that these single cell/tumor expression ratios for the eTME genes are mostly positive, proving the overall concordance between single cell data and eTME. The P values of two-way T tests comparing non-eTME vs. eTME ($3\times$) genes, and non-eTME vs. e-TME ($20\times$) genes are smaller than 1×10^{-15} . We also asked a complementary question, by examining genes that are most highly expressed in the single cells. By limiting to genes whose $\log(\text{single cell/tumor})$

ratios are larger than 3.5, 4 and 4.5, 71%, 86%, and 98% of the genes can be recapitulated by eTME (3x).

Next, we determined the immune cell type of each sequenced cell according to expression of eTME signature genes. **Fig. 3c** shows the number of cells assigned to each cell type and the proportions of these cells in the myeloid pool and lymphoid pool. Reassuringly, most designated neutrophils, macrophages, and eosinophils were found in the myeloid pool and most designated T cells, NK cells, and B cells were found in the lymphoid pool. Moreover, most lymphoid cells were designated as T cells, which is consistent with previous reports of counts of lymphoid cells in RCC (32). Admittedly, technical issues may exist that led to some apparent artefacts, such as assignment of some B cells in the myeloid pool. To further validate eTME, we profiled the infiltration levels of several immune cell types in a small set of patient RNA-Seq data using ssGSEA on the refined gene signatures that passed the eTME criterion of 3x. For the same patients, we carried out immunohistochemistry (IHC) for the cell type surface markers. On average, the correlation between predicted infiltration abundances and IHC was 0.69 (**Fig. 3d**). On balance, these results confirm the accuracy of the eTME genes signatures that we defined with empirical evidence. We have also incorporated the eTME signatures in our DisHet package and have provided them publicly.

Discovery of a highly-inflamed pan-RCC subtype based on eTME expression profiles

The identification of a highly specific RCC eTME gene signature enabled us to evaluate the functional importance of the tumor immune infiltrates. We deployed the eTME genes to assess the largest publicly available cohort of RCC samples with RNA-Seq data, 884 RCCs from TCGA. This cohort includes tumors of different histologies. We conducted Principal Component Analysis (PCA) on these tumors based on eTME gene expression with the more stringent cutoff of 20 $\times$ and used tumor specific genes as a control. Notably, the eTME gene expression signature was able to cluster RCCs of the same histology with good accuracy (**Sup. Fig. 4**). We contrasted the eTME-based clustering with clustering based on tumor-specific genes and observed a similar clustering structure based on tumor specific gene expression. However, the clustering was less robust in separating out the pRCC and chRCC histologies, than eTME genes. These results show that the tumor microenvironment carries important discriminating features corresponding to the different histologies.

Building on the previous observations, we took a more focused approach to elucidate the heterogeneity of eTME expression profiles in each patient. We focused on ccRCC first and clustered all TCGA ccRCC patients (n=529) based on expression of the eTME genes (20 $\times$). We observed two clusters of patients, one accounting for about 42% of patients and the other accounting for about 58% of patients (**Fig. 4a**). To characterize each cluster, we evaluated the infiltrating levels of various types of immune cells in each group of patients, by applying the ssGSEA method on cell type-specific signature genes that passed the more relaxed eTME criterion of 3x. We found that the first cluster of patients was enriched for T_{reg} cells, NK cells, T_H cells, neutrophils, macrophages, eosinophils, B cells, CD8⁺ T cells, and aDC cells. This cluster was also characterized by high expression levels of the complement system proteins (represented by the C1q signature from the Winslow signatures). The second cluster was, in contrast, enriched for angiogenesis, pDC, and mast cells. Thus, the eTME

signature separated tumors into an inflamed subtype (IS) that included an extensive immune infiltrate and a second subtype that was not inflamed (NIS). We found that IS was enriched for mutations in *BAP1* ($p=7.7\times10^{-5}$), a tumor suppressor associated with high grade cancers (15,33). In contrast, *PBRM1* mutations were more homogeneously distributed across both subtypes, though enriched in NIS.

We extended our observations to two other, albeit significantly smaller, ccRCC cohorts. We trained a support vector machine (SVM) classifier based on the unsupervised clusters and predicted the subtype identities of ccRCC and non-clear cell RCC (nccRCC) patients from the UTSW KCP database (181 patients in total, 39% ccRCCs, 24% pRCCs, 15% chRCCs, 22% other) with bulk tumor RNA-seq data available (**Fig. 4b**). As in the TCGA ccRCC cohort, in the KCP cohort, eTME-IS ccRCC patients were associated with enriched *BAP1* mutations, although statistical significance was not achieved ($p=0.21$, $N=70$). Next, we analyzed the exome-Seq and RNA-Seq data from another ccRCC cohort (34). In this cohort, we found 9 patients with *BAP1* mutations out of all 95 available patients. The SVM model predicted 41 eTME-IS patients, which include 7 *BAP1*-mutated tumors ($p=0.032$). Given the intriguing correlation of *BAP1* and eTME-IS, we carried out CD4 and CD8 IHC staining in our recently established RCC mouse models, *Six2-Cre;Vhl^{F/F};Pbrm1^{F/F}* and *Six2-Cre;Vhl^{F/F};Bap1^{F/+}* (35). We observed more CD4 and CD8 T cell infiltration in *Bap1*-mutated mice than in *Pbrm1*-mutated mice, suggesting a causal relationship may exist between *BAP1* mutations and the eTME-IS phenotype (**Sup. Fig. 5**).

Next we extended our analyses to nccRCC. We evaluated TCGA pRCC and chRCC patients (**Fig. 4c**). The eTME-IS patients accounted for a minority of nccRCC tumors. Interestingly, the pRCC tumors in the eTME-IS group were predominantly type 2 ($p=7.2\times10^{-5}$), which have a worse prognosis than type 1 (36). We also carried out ssGSEA analysis using the refined eTME signatures in both KCP and TCGA nccRCC cohorts (**Fig. 4bc**). Across all cohorts examined, there was enrichment of T_{reg} cells, NK cells, T_{H1} cells, neutrophils, macrophages, B cells, and $CD8^{+}$ T cells, and the complement system in eTME-IS patients. Therefore, the grouping of RCC patients based on eTME expression profiles revealed a pan-RCC subtype with shared similarities across different RCC histologies.

Recently, Ayers *et al* (37) reported an 18-gene IFN- γ gene signature whose activation predicts clinical response to PD-1 blockade. We analyzed this gene signature in all three patient cohorts (**Fig. 4**), and we discovered a very strong correlation of the IFN- γ activation signature with the eTME-IS patient subtype. In synergy with this, the expression of *PD1* (*PDCDI*) is higher in eTME-IS patients than eTME-NIS patients (**Fig. 4a**). These results reveal the potential of the eTME analysis and eTME-IS/NIS patient subtyping to be translated to clinical practice.

Next we asked whether the differences in inflammatory phenotype correlated with tumor characteristics on imaging. We focused on the 35 RCC patients analyzed by DisHet. The SVM classifier predicted that 21 of these patients belonged to eTME-IS. An experienced radiologist (I.P.) examined all MRI/CT images for these patients (32 as available) blindly, and observed (**Fig. 4d**) a significant correlation ($p=0.034$) between eTME-IS and infiltrative tumor margins (including partial infiltration). In addition, we also reviewed H&E staining

images (P.K.), which showed that eTME-IS samples are associated with more necrosis than eTME-NIS samples (**Fig. 4e**), which also indicates poor prognosis (38).

Tumors induce stromal responses when engrafted in mice

To determine whether the inflamed tumor microenvironment is driven by the tumor cells, we asked whether human tumors, when transplanted in mice, stimulated an inflammatory response similar to that in the human stroma as determined by tumor infiltration by inflammatory murine cells. We first focused on the 35 RCC patients. Because the tumorgraft hosts, NOD/SCID mice, are inherently deficient in B and T cells, we focused on neutrophils and NK cells. By quantifying the human and mouse immune cell populations in the original human tumors and the mouse-specific gene expression from tumorgrafts, respectively, we found that the mouse neutrophil and NK cell infiltrates were induced to an extent comparable in humans (**Fig. 5a**). Comparing eTME-IS and eTME-NIS tumors, we showed that eTME-IS tumors have a higher level of neutrophils than eTME-NIS tumors ($p=0.021$), and they also elicit higher levels of neutrophils in the mouse ($p=0.031$). A similar trend was also observed for NK cells (**Fig. 5a**, $p=0.036$ in human and $p=0.23$ in mice).

Although the above analyses showed a trend of concordancy, the characteristics of the immune infiltrates could be significantly different between human disease and murine tumorgraft environment relating to factors such as species and microbiota. Therefore, we employed another analysis that more broadly investigates the eTME genes as a whole. We hypothesized that expression of mouse immune/stroma-specific genes in different tumorgrafts established from the same patients should be similar to each other, if tumors indeed drive stromal responses. We analyzed RNA-Seq data for a set of 37 tumorgrafts from a total of 14 unique patients from our database, for whom at least two tumorgrafts had been established and analyzed by RNA-Seq for each patient (**Sup. Table 1**). We carried out Principal Component Analysis of mouse gene expression quantified from the tumorgraft RNA-Seq data, for genes that belong to the innate immunity system and satisfied the eTME cutoff of 3x. As can be seen in **Fig. 5b**, tumorgrafts from the same patients tend to cluster into the same entities. As a negative control, we carried out this analysis for all non-eTME ($3\times$ cutoff) mouse transcripts, and observed almost no separation between tumorgraft samples of the same patients (**Fig. 5b**). To further illustrate this analysis from another perspective, we calculated the Euclidean distances between tumorgraft pairs from the same patients as one group, and pairs from different patients as another group, determined by murine expression of eTME genes in innate immunity system, all eTME genes, and all other genes. The density plot in **Fig. 5c** shows that the distances between tumorgrafts from the same patients are much smaller than those from different patients, in terms of expression of eTME genes in the innate immune system (KS test, $p=7.7\times 10^{-15}$). In contrast, inter-patient and intra-patient distances are almost indistinguishable according to expression of other non-eTME genes ($p=0.98$, **Fig. 5c**). Interestingly, echoing this active role of tumor cells on stromal immune responses, our previous work showed that RCC tumors that induced hypercalcemia in humans also induced hypercalcemia in tumorgraft mice (13).

Inflammatory tumor microenvironment correlates with systemic inflammatory manifestations and worse prognosis

We next asked whether there was a correlation between the inflammatory subtype (eTME-IS) we identified from expression profiles with systemic manifestations of inflammation and other systemic factors. We extracted clinical lab records from UTSW KCP patients from our database. We found that patients with eTME-IS tumors had higher platelet counts ($p=9.4\times 10^{-4}$), and lower hemoglobin levels ($p=8.1\times 10^{-10}$) (**Fig. 6a**), which are established prognostic markers in RCC and may signal systemic inflammation (22). In contrast, neutrophil counts were not significantly different between these two groups of patients ($p=0.71$). As a control, creatinine, a marker of renal function (39), showed no significant difference between patients with eTME-IS and eTME-NIS tumors ($p=0.20$) (**Sup. Fig. 6**). While not all of these parameters are available for the TCGA cohort, we asked whether a correlation existed between the inflamed tumor microenvironment with thrombocytosis and anemia. In all the TCGA RCC patients combined, higher proportions of eTME-IS patients had elevated platelet counts ($p=0.039$) and low hemoglobin levels ($p=0.019$) compared with eTME-NIS patients (**Fig. 6b**). Overall, these data show that tumors with eTME-IS are associated with thrombocytosis and anemia, which may represent systemic manifestations of tumor-induced inflammation.

Terminal cancer, poorly controlled inflammatory states and wasting syndromes are associated with malnourishment and lower protein (albumin) levels in the blood. We asked whether a correlation existed between the inflammatory microenvironment and low albumin. Interestingly, eTME-IS patients had lower albumin levels than eTME-NIS patients ($p=2.8\times 10^{-7}$) (**Sup. Fig. 6**). Another indicator of aggressive cancer (40), whose pathogenesis is not always understood, is low circulating sodium. Our results show that eTME-IS patients seem to have lower sodium levels as well ($p=0.014$) (**Sup. Fig. 6**).

Given our findings that the inflamed tumor microenvironment correlates with systemic markers of inflammation, which are validated prognostic factors in RCC (21,22), we asked whether the inflamed tumor microenvironment correlated with patient survival. Our Kaplan-Meier analyses showed that eTME-IS patients had significantly worse overall survival across all three patient cohorts analyzed (**Fig. 6c**) (HR=1.8 (95% CI=1.3–2.5) and $p=1.4\times 10^{-4}$ in the TCGA ccRCC cohort, HR=11.9 (4.1–34.5) and $p=4.6\times 10^{-6}$ in the UTSW cohort, and HR=2.3 (1.2–4.5) and $p=0.014$ in the TCGA nccRCC cohort).

DISCUSSION

There are several unique advantages of our methodologies over previous efforts to dissect the tumor microenvironment. In our study, we took the novel approach of including matched tumors, tumorgrafts, and normal tissue samples from the same patients to dissect out the different components of the tumor microenvironment. The tumorgrafts provided more accurate transcriptomic profiles of tumor cells grown *in vivo* than cultured tumor cells, and allow the accurate subtraction of the tumor cell component from the tumor bulk. Second, this analysis involving matched trio samples is empowered by DisHet, a novel and efficient Bayesian algorithm. DisHet properly considers the possible existence of all three cellular components in a bulk tumor.

With the DisHet analyses of the matched trio RNA-Seq data, we refined, with accurate empirical evidence, previous immune/stroma-specific genes such as Immunome, ESTIMATE, and Winslow and in addition annotated many other immune/stroma-specific genes not previously reported. Our analyses enabled the generation of an empirical tumor microenvironment gene signature, which we refer to as eTME. Importantly, immune cell quantification methods like ssGSEA rely on accurate gene signatures. The eTME genes defined by the DisHet analysis could be used to refine these signatures, as is done in our study, and to make the immune cell quantification much more relevant. Moreover, the tumor microenvironment of each cancer type could be very different. Our eTME gene signatures are defined with RCC data and should have very high accuracy in RCC, while ESTIMATE identified leukocytes in breast, colorectal and ovarian cancer. Ideally, such tumor-specific gene signatures should be established for each type of cancer. The NCI Patient-Derived Models Repository (PDMR) repository was recently released and has provided RNA-Seq data for more than 600 tumorgraft models for 24 cancer types. Valuable resources such as PDMR can be utilized for achieving this goal.

The advantages of our approach complement insights that can be provided through other techniques such as single cell sequencing and Laser Capture Microdissection (LCM). It will be more powerful if these approaches and tumorgraft models can be utilized in a synergistic manner to tackle tumor heterogeneity and tumor immunity, as is also exemplified by some of the analyses carried out in this study. Over time, the cost and complexity of single cell analysis and LCM are expected to decrease, which should make these approaches more affordable and feasible.

Recently, Ben-David *et al* (16) reported that tumorgrafts do not represent an exact replica of the original tumors in multiple tumor types. Notably, their study shows that RCC are the second most stable tumor next to brain tumors. Indeed, our own work (13,15) has shown that passage has insignificant impact on copy number variation and gene expression, even in late passage. While in this work all tumorgrafts except for one were from early passage (≤ 3), one was from passage 11, and we did not find substantial differences compared to earlier passages. In addition, more generally, the genomic instability reported by Ben-David *et al* is inferred from gene expression data, rather than directly measured at the DNA level, and focuses on arm-level copy number variation. Furthermore, as reported by the authors, their findings are largely relegated to tumors that experience genome duplication, which is rare in RCC.

The eTME signature we identified was applicable across RCC histologies and we show that it offers greater discriminatory power of histological subtypes than a signature focusing on tumor-specific genes. These data suggest the existence of unique TME features that accompany the different histological subtypes. Furthermore, our work shows that eTME gene expression can separate tumors into a highly inflammatory subtype (eTME-IS) and a non-inflammatory subtype (eTME-NIS) with a characteristic immune cell infiltration repertoire and stromal pathway activation. This clustering of eTME-IS and eTME-NIS patients and their associated immune infiltration patterns seems to be corroborated by flow cytometric findings by Giraldo *et al* (41). In their study, the authors identified a ccRCC “immune-regulated” subtype of patients, characterized by inflammation, CD4⁺ICOS⁺ cells

with a T_{reg} phenotype, and high risk of disease progression, which likely corresponds to the eTME-IS phenotype. On the other hand, we found higher numbers of mast cells in eTME-NIS tumors, which are associated with a better outcome in keeping with a recent report showing that tumor infiltrating mast cells are associated with a survival advantage in localized ccRCC (42).

The tumor microenvironment may have either a pro-tumor or anti-tumor effect. Interestingly, eTME-IS tumors may be more susceptible to immunotherapy. eTME-IS tumors are characterized by an IFN- γ signature, which has been associated with response to PD-1 blockade (37). It is also in alignment with the fact that patients with worse prognosis may have a higher chance of response to Nivolumab+Ipilimumab, and perhaps even nivolumab alone (43,44).

Unsupervised clustering analyses revealed that *BAP1*-mutant ccRCC tumors are correlated with an inflammatory tumor microenvironment (eTME-IS). Interestingly, *BAP1* deficiency has been associated with a peritoneal inflammatory response upon exposure to asbestos fibers in mice (45), and with an inflammatory phenotype in primary uveal melanoma (46). Our CD4 and CD8 IHC experiments in the GEMMs suggest a causal relationship between *BAP1* loss and immune infiltrations in RCC, though more extensive experiments are needed for confirmation.

Our data provides insight into a missing link between tumor subtypes and important systemic prognostic factors, which may be indicators of inflammation. Specifically, anemia, thrombocytosis and neutrophilia correlate with inflammatory conditions beyond cancer. These markers are associated, for instance, with autoimmune conditions such as rheumatoid arthritis, Crohn's disease, and ulcerative colitis, as well as some infections such as granulomatous infections. However, their molecular pathogenesis, and specifically how these manifestations are linked to RCC, is poorly understood. Our study shows that the eTME-IS phenotype is associated with anemia and thrombocytosis. Inasmuch as the inflammatory tumor microenvironment may account for thrombocytosis or anemia, this study sets the foundation for identifying mediators of this systemic responses, and as such, begins to shed light into features in the tumor/TME that may affect prognosis.

Overall, our work, which was enabled by important methodological advances, provides novel biological insights of the tumor microenvironment based on empirical analyses of renal tumors, which could eventually lead to novel therapeutic interventions. Especially in the modern era of immunotherapy, a better characterization of the tumor-microenvironment interactions may hold clues to more informed treatment decisions involving immunotherapy. The mechanism of tumor immunogenicity could be vastly different in different tumor types. For example, in some other tumor types (*e.g.* melanoma or head & neck), high lymphocyte infiltration is associated with better prognosis. However, our methodology is general and should have wide applicability for understanding the roles of immunogenicity and inflammation in development, progression, metastasis, and treatment of a large variety of cancers.

MATERIALS AND METHODS

UTSW patient samples

The study was performed following a protocol approved by the Institutional Review Board (IRB) of the University of Texas Southwestern Medical Center that involves analyses of kidney cancer (or control) samples donated by patients. The studies were conducted in accordance with the Belmont Report and U.S. Common Rule guidelines. The investigators obtained informed written consent from the subjects regarding whether their tissue samples can be used in this study. We carried out RNA-Seq for 403 samples and exome-Seq for 343 samples from 195 patients of the UTSW KCP cohort (**Sup. Table 1**). RNA-Seq and Exome-seq analyses were performed from the same samples using protocols that allow simultaneous DNA and RNA extraction (47). Small samples to be extracted (~50 mg) were flanked by sections used for pathological imaging analyses. In addition to genomics data, we also queried clinical lab results and follow up data for the UTSW patients. H&E images were reviewed by a pathologist (P.K.) blindly to report percentages of tumor nuclei. Among the 35 tumorgraft models that were used in the primary analysis (**Sup. Table 1**), 25 are first passage, 7 are second passage, 2 are third passage, and one is the 11th passage. The cell line derived from the tumor of patient XP389 (**Fig. 1c**) is established in 2012 in Dr. James Brugarolas lab.

Immunohistochemistry

Four-micron-thick sections of formalin-fixed, paraffin-embedded tissue were subjected to heat-induced epitope retrieval using CC1 (Ventana Medical Systems Inc., Tucson, Arizona, USA), a Tris-based buffer at pH 8–8.5, and immunohistochemistry staining. We used rabbit monoclonal CD3 (Ventana Medical Systems, Inc.), mouse monoclonal CD4 (Leica Biosystems Ltd, Newcastle, UK), rabbit monoclonal CD8 (Ventana Medical Systems, Inc.), mouse monoclonal CD20 (Ventana Medical Systems, Inc.), mouse monoclonal CD68 (Dako, Carpinteria, California, USA), mouse monoclonal CD56 (Dako), and mouse-monoclonal CD117 (Dako). The slides were scanned for each case using the Aperio ScanScope (Aperio Technologies, Vista, CA, USA). After scanning the slides, relevant areas of tumor were selected for analysis. Aperio membrane IHC algorithm was used for each marker, which is based on the spectral differentiation between brown (positive) and blue (counter) staining. Results of Aperio IHC quantification were also confirmed manually. We calculated the proportion of positive cells out of all cells in the sampled areas. The final results were validated by another pathologist, M.C.

The DisHet model

The DisHet model dissects the bulk tumor RNA-Seq data such that patient-to-patient differences in the immune/stroma gene expression profiles are minimized, based on the assumption that the immune/stroma components of different patients share similarities in gene expression. At the same time, the “remaining” patient-to-patient differences can reflect the true individual variation in gene expression of the immune/stroma component. The DisHet algorithm is based on a Bayesian Hierarchical Model and details of model setup and estimation procedures can be found in the **Supplementary File**.

High-throughput sequencing data analysis

RNA-Seq reads that passed quality control were aligned to mouse genome (mm10) and human genome (hg38) using HISAT2. Reads that were primarily aligned to mouse (mm10) were removed. By aligning a human tumor RNA-Seq dataset competitively to the combined genome of human and mouse, we estimated the possibility of misalignment of a sequencing read to the wrong genome to be only 0.047%. The remaining reads were then re-aligned to the human reference genome (hg38). Exon level read counting was performed using featureCounts. Multi mapping reads were included in the final count for each exon. For each sample, a constant value of 1 was added to the raw read counts per gene. Expression matrix was quantile normalized. For exome-Seq data, BWA was used to map Fastq files to the combined reference genome of human (hg19) and mouse (mm10). Reads that had primary alignments in mouse genome were dropped. Then remaining reads were mapped again to hg19 using BWA. Duplicates were removed by Picard tools and sequencing data of the same sample across lanes were merged. Realignment was carried out around indels using RealignerTargetCreator from GATK toolkit, and BaseRecalibrator from GATK toolkit was used to carry out base recalibration. Somatic SNP and indel mutations were called using MuTect2, VarScan, Strelka, and SpeedSeq and combined. Copy number variation was detected from exome-seq data using in-house software (publication in preparation).

Single cell sequencing data generation and analysis

Resected tumors from five RCC patients were rinsed with PBS, dissociated mechanically into tiny pieces, and transferred into a 50 ml conical tube containing 10 ml pre-warmed Digest media (10ml Processing media, 40ul Dispase stock, 40ul collagenase stock, and 20ul DNase stock, pre-prepared). Tumor tissues were incubated in digestion media for 10 minutes at 37°C, then vortexed for 10 seconds and pipetted up and down for 1 minute using pipettes of descending sizes (25 ml, 10 ml and 5 ml). The cell suspension was filtered using a 70µm nylon mesh to remove residual cell clumps and centrifuged at 580g at 4°C for 6 minutes. CD45 positive cells of different granularity that were viable (Calcein blue high) were sorted into each well of a 96-well plate (1 plate myeloid, 1 plate lymphoid) in the UTSW Flow Cytometry Core. Reverse transcription and whole transcriptome amplification of single cells were performed with SMART-Seq v4 Ultra Low Input RNA Kit (Takara) following the manufacturer's protocol. Amplified cDNA products were then purified with 0.9X Agencourt AmPure XP-beads (Beckman). Following purification, the PCR product of each single cell were analyzed for fragment size distribution with Bioanalyzer (Agilent Technologies) using High Sensitivity DNA reagent (Agilent Technologies) and quantified with Quant-IT picogreen dsDNA assay kit (Thermo Fischer scientific). Single cell RNAseq libraries were then constructed using Nextera XT DNA sample preparation kit (Illumina) and Nextera XT index kit (Illumina). Libraries were pooled together, QC and quantified using Bioanalyzer (Agilent Technologies) and Qubit HsDNA kit (Invitrogen). Constructed libraries were then sequenced by an Illumina Hiseq2500 or Hiseq4000 sequencer with 1X50 bp reads.

Single cell RNA-Seq data were processed by standard bioinformatics pipelines and assembled into a gene expression matrix, E_{ij} ($i = 1 \dots I, j = 1 \dots J$), of I genes x J cells with log transformation and quantile normalization. We used the following approach to assign cell

type to each sequenced cell. First, we centered each gene's expression by $E_{ij}^c = E_{ij} - \text{median}(E_i)$. We denote the set of signature genes for each cell type, $k = 1 \dots K$, as S_k . The signature enrichments for this cell type across all cells, $i = 1 \dots J$, are calculated as $e_j^k = \text{median}(E_{ij}^c \mid i \in S_k)$. This set of enrichment scores for each cell type is normalized by median and standard deviation: $e_j^{n,k} = (e_j^k - \text{median}(e^k)) / \text{sd}(e^k)$. A sequenced cell j is assigned cell type k if and only if $e_j^{n,k} = \max(e_j^{n,\cdot})$ and $e_j^{n,k} > \text{median}(e^{n,k})$. We successfully assigned cell type identities to 663 cells (68.5% of all cells).

Statistical analysis

All computations and statistical analyses were carried out in the R computing environment. All hypothesis tests performed in this study are two-sided, unless otherwise specified. For comparison with DeMix, we run the DeMix Linux version with the following parameters: Cit=64, Matched=1, Nstart=2, Nref=0, and Nhuman=1 (8). ssGSEA analysis was performed using the R GSVA package by calling the gsva function with parameter method="ssgsea" and mseq=T (48). In this study, the ESTIMATE stromal and immune signatures are combined as one unique signature. Gene ontology analysis is carried out on the GOrilla server (49). Hierarchical clustering of TCGA ccRCC patients was based on Euclidean distance and the "complete" agglomeration method. We used a support vector machine to train a classifier to learn the two subtypes generated by unsupervised clustering of TCGA ccRCC patients. We used the svm function in the e1071 R package with sigmoid kernel. For survival analyses, we employed the log-rank test based on dichotomized patient sets by the eTME-IS/NIS subtypes. Patients' overall survival data were visualized by Kaplan-Meier estimator. For all boxplots appearing in this study, box boundaries represent interquartile ranges, whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range, and the line in the middle of the box represents the median. We used Spearman correlation for all correlations calculated for validation purpose in **Fig. 2** and **Sup. Fig. 3**.

Data availability

Data are provided for UTSW patients specifically consenting to placement of their raw genomic data in a protected publicly accessible database. Sequencing data from consented patients are deposited in EGA, with accession numbers EGAS00001002786 and EGAS00001000926 (also shown in **Sup. Table 1**).

Code availability

The DisHet algorithm is freely available at <https://cran.r-project.org/web/packages/DisHet> and <http://lce.biohpc.swmed.edu/dishet>. The eTME gene signatures for immune cells we defined in this study are also incorporated into both the R package and the webserver.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We acknowledge the UT Southwestern Tissue Shared Resource for collecting patient tissue samples. We acknowledge the UTSW Flow Cytometry Core for providing help with the single cell sequencing experiments. We acknowledge the UTSW Kidney Spore Histology core for providing help with all IHC experiments. We acknowledge Eva Moresco and Jessie Norris for their helpful comments on the writing of this paper.

Financial support: This study was supported by the National Institutes of Health (NIH) [R03 ES026397–01/TW, R01 GM110050–01A1/MC, R01 CA175754/JB, SPORE 1P50CA19651601/PK, JB, TW, IP, R01 CA154475/IP, CCSG 5P30CA142543/TW], Center for Translational Medicine of UT Southwestern [SPG2016–018/TW], Cancer Prevention and Research Institute of Texas [CPRIT RP150596/MK, HZ, RL, DL], and American Cancer Society [RSG-16–001-01-CCE/RH]. This work was also partially supported by fundraising efforts orchestrated by the KCP Patient Council and the Kidney Cancer Coalition.

Bibliography

1. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462:108–12. [PubMed: 19847166]
2. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. [PubMed: 24113773]
3. Winslow S, Leandersson K, Edsjö A, Larsson C. Prognostic stromal gene signatures in breast cancer. *Breast Cancer Res*. 2015;17:23. [PubMed: 25848820]
4. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39:782–95. [PubMed: 24138885]
5. Chung SH, Shen W. Laser capture microdissection: from its principle to applications in research on neurodegeneration. *Neural Regen Res*. 2015;10:897–8. [PubMed: 26199603]
6. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet*. 2016;17:175–88. [PubMed: 26806412]
7. Yadav VK, De S. An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinformatics*. 2015;16:232–41. [PubMed: 24562872]
8. Ahn J, Yuan Y, Parmigiani G, Suraokar MB, Diao L, Wistuba II, et al. DeMix: deconvolution for mixed cancer transcriptomes using raw measured data. *Bioinformatics*. 2013;29:1865–71. [PubMed: 23712657]
9. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet*. 2015;47:1168–78. [PubMed: 26343385]
10. Hackl H, Charoentong P, Finotello F, Trajanoski Z. Computational genomics tools for dissecting tumour-immune cell interactions. *Nat Rev Genet*. 2016;17:441–58. [PubMed: 27376489]
10. Williams SA, Anderson WC, Santaguida MT, Dylla SJ. Patient-derived xenografts, the cancer stem cell paradigm, and cancer pathobiology in the 21st century. *Lab Invest*. 2013;93:970–82. [PubMed: 23917877]
11. Malaney P, Nicosia SV, Davé V. One mouse, one patient paradigm: New avatars of personalized cancer therapy. *Cancer Lett*. 2014;344:1–12. [PubMed: 24157811]
12. Sivanand S, Peña-Llopis S, Zhao H, Kucejova B, Spence P, Pavia-Jimenez A, et al. A validated tumorgraft model reveals activity of dovitinib against renal cell carcinoma. *Sci Transl Med*. 2012;4:137ra75.
13. Braekeveldt N, Wigerup C, Gisselsson D, Mohlin S, Merselius M, Beckman S, et al. Neuroblastoma patient-derived orthotopic xenografts retain metastatic patterns and geno- and phenotypes of patient tumours. *Int J Cancer*. 2015;136:E252–61. [PubMed: 25220031]
14. Peña-Llopis S, Vega-Rubín-de-Celis S, Liao A, Leng N, Pavia-Jiménez A, Wang S, et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet*. 2012;44:751–9. [PubMed: 22683710]

15. Ben-David U, Ha G, Tseng Y-Y, Greenwald NF, Oh C, Shih J, et al. Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat Genet.* 2017;49:1567–75. [PubMed: 28991255]
16. Bradford JR, Wappett M, Beran G, Logie A, Delpuech O, Brown H, et al. Whole transcriptome profiling of patient-derived xenograft models as a tool to identify both tumor and stromal specific biomarkers. *Oncotarget.* 2016;7:20773–87. [PubMed: 26980748]
17. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. *BMJ.* 2014;349:g4797. [PubMed: 25385470]
18. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* 2017;18:1009–21. [PubMed: 28694034]
19. Spranger S, Luke JJ, Bao R, Zha Y, Hernandez KM, Li Y, et al. Density of immunogenic antigens does not explain the presence or absence of the T-cell-inflamed tumor microenvironment in melanoma. *Proc Natl Acad Sci USA.* 2016;113:E7759–68. [PubMed: 27837020]
20. Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. *J Clin Oncol.* 2002;20:289–96. [PubMed: 11773181]
21. Heng DY, Xie W, Regan MM, Warren MA, Golshayan AR, Sahi C, et al. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. *J Clin Oncol.* 2009;27:5794–9. [PubMed: 19826129]
22. Itsumi M, Tatsugami K. Immunotherapy for renal cell carcinoma. *Clin Dev Immunol.* 2010;2010:284581. [PubMed: 21253521]
23. Haanen JBAG. Immunotherapy of melanoma. *EJC Suppl.* 2013;11:97–105. [PubMed: 26217118]
24. Somasundaram A, Burns TF. The next generation of immunotherapy: keeping lung cancer in check. *J Hematol Oncol.* 2017;10:87. [PubMed: 28434399]
25. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* 2016;17:231. [PubMed: 27855702]
26. Domoto T, Miyama Y, Suzuki H, Teratani T, Arai K, Sugiyama T, et al. Evaluation of S100A10, annexin II and B-FABP expression as markers for renal cell carcinoma. *Cancer Sci.* 2007;98:77–82. [PubMed: 17083565]
27. Yao M, Murakami T, Shioi K, Mizuno N, Ito H, Kondo K, et al. Tumor signatures of PTHLH overexpression, high serum calcium, and poor prognosis were observed exclusively in clear cell but not non clear cell renal carcinomas. *Cancer Med.* 2014;3:845–54. [PubMed: 24861371]
28. Jilaveanu LB, Sznol J, Aziz SA, Duchon D, Kluger HM, Camp RL. CD70 expression patterns in renal cell carcinoma. *Hum Pathol.* 2012;43:1394–9. [PubMed: 22401771]
29. Chung K-M, Hsu S-C, Chu Y-R, Lin M-Y, Jiaang W-T, Chen R-H, et al. Fibroblast activation protein (FAP) is essential for the migration of bone marrow mesenchymal stem cells through RhoA activation. *PLoS ONE.* 2014;9:e88772. [PubMed: 24551161]
30. Ratajczak-Wielgomas K, Grzegorzolka J, Piotrowska A, Gomulkiewicz A, Witkiewicz W, Dziegiel P. Periostin expression in cancer-associated fibroblasts of invasive ductal breast carcinoma. *Oncol Rep.* 2016;36:2745–54. [PubMed: 27633896]
31. Moeckel J, Staiger N, Mackensen A, Meidenbauer N, Ullrich E. Sunitinib does not impair natural killer cell function in patients with renal cell carcinoma. *Oncol Lett.* 2017;14:1089–96. [PubMed: 28693278]
32. Kapur P, Peña-Llopis S, Christie A, Zhrebker L, Pavía-Jiménez A, Rathmell WK, et al. Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *Lancet Oncol.* 2013;14:159–67. [PubMed: 23333114]
33. Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet.* 2013;45:860–7. [PubMed: 23797736]

34. Gu Y-F, Cohn S, Christie A, McKenzie T, Wolff N, Do QN, et al. Modeling renal cell carcinoma in mice: bap1 and pbrm1 inactivation drive tumor grade. *Cancer Discov.* 2017;7:900–17. [PubMed: 28473526]
35. Mejean A, Hopirtean V, Bazin JP, Larousserie F, Benoit H, Chrétien Y, et al. Prognostic factors for the survival of patients with papillary renal cell carcinoma: meaning of histological typing and multifocality. *J Urol.* 2003;170:764–7. [PubMed: 12913693]
36. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest.* 2017;127:2930–40. [PubMed: 28650338]
37. Pichler M, Hutterer GC, Chromecki TF, Jesche J, Kappel-Kettner K, Rehak P, et al. Histologic tumor necrosis is an independent prognostic indicator for clear cell and papillary renal cell carcinoma. *Am J Clin Pathol.* 2012;137:283–9. [PubMed: 22261455]
38. Perrone RD, Madias NE, Levey AS. Serum creatinine as an index of renal function: new insights into old concepts. *Clin Chem.* 1992;38:1933–53. [PubMed: 1394976]
39. Jeppesen AN, Jensen HK, Donskov F, Marcussen N, von der Maase H. Hyponatremia as a prognostic and predictive factor in metastatic renal cell carcinoma. *Br J Cancer.* 2010;102:867–72. [PubMed: 20145619]
40. Giraldo NA, Becht E, Vano Y, Petitprez F, Lacroix L, Validire P, et al. Tumor-Infiltrating and Peripheral Blood T-cell Immunophenotypes Predict Early Relapse in Localized Clear Cell Renal Cell Carcinoma. *Clin Cancer Res.* 2017;23:4416–28. [PubMed: 28213366]
41. Fu H, Zhu Y, Wang Y, Liu Z, Zhang J, Wang Z, et al. Tumor Infiltrating Mast Cells (TIMs) Confers a Marked Survival Advantage in Nonmetastatic Clear-Cell Renal Cell Carcinoma. *Ann Surg Oncol.* 2017;24:1435–42. [PubMed: 27896514]
42. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med.* 2015;373:1803–13. [PubMed: 26406148]
43. Escudier B, Tannir NM, McDermott DF, Frontera OA, Melichar B, Plimack ER, et al. LBA5CheckMate 214: Efficacy and safety of nivolumab + ipilimumab (N+I) v sunitinib (S) for treatment-naïve advanced or metastatic renal cell carcinoma (mRCC), including IMDC risk and PD-L1 expression subgroups. *Ann Oncol.* 2017;28.
44. Napolitano A, Pellegrini L, Dey A, Larson D, Tanji M, Flores EG, et al. Minimal asbestos exposure in germline BAP1 heterozygous mice is associated with deregulated inflammatory response and increased risk of mesothelioma. *Oncogene.* 2016;35:1996–2002. [PubMed: 26119930]
45. Gezgin G, Dogrusöz M, van Essen TH, Kroes WGM, Luyten GPM, van der Velden PA, et al. Genetic evolution of uveal melanoma guides the development of an inflammatory microenvironment. *Cancer Immunol Immunother.* 2017;66:1–10. [PubMed: 27714433]
46. Peña-Llopis S, Brugarolas J. Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat Protoc.* 2013;8:2240–55.
47. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7. [PubMed: 23323831]
48. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 2009;10:48. [PubMed: 19192299]

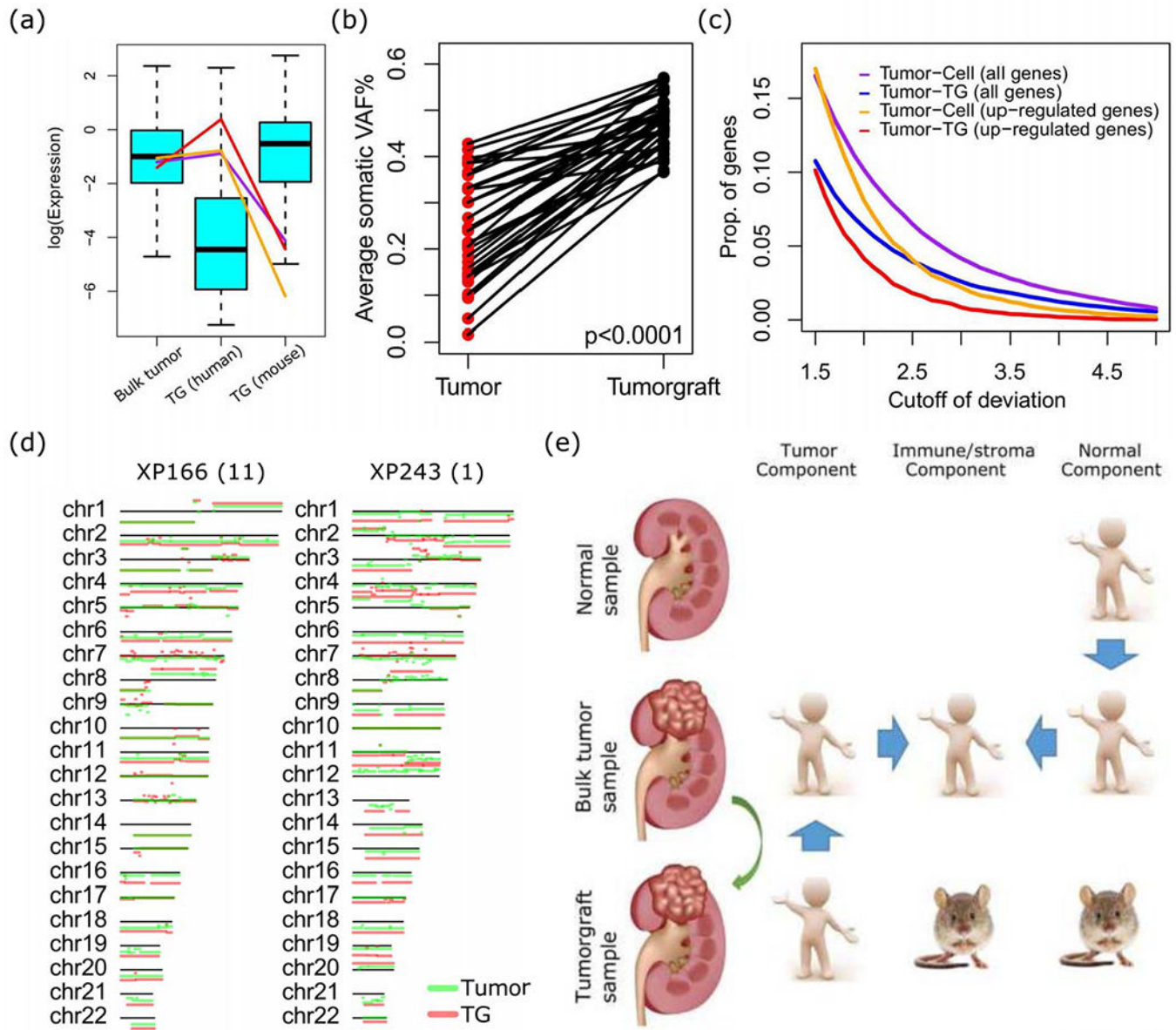
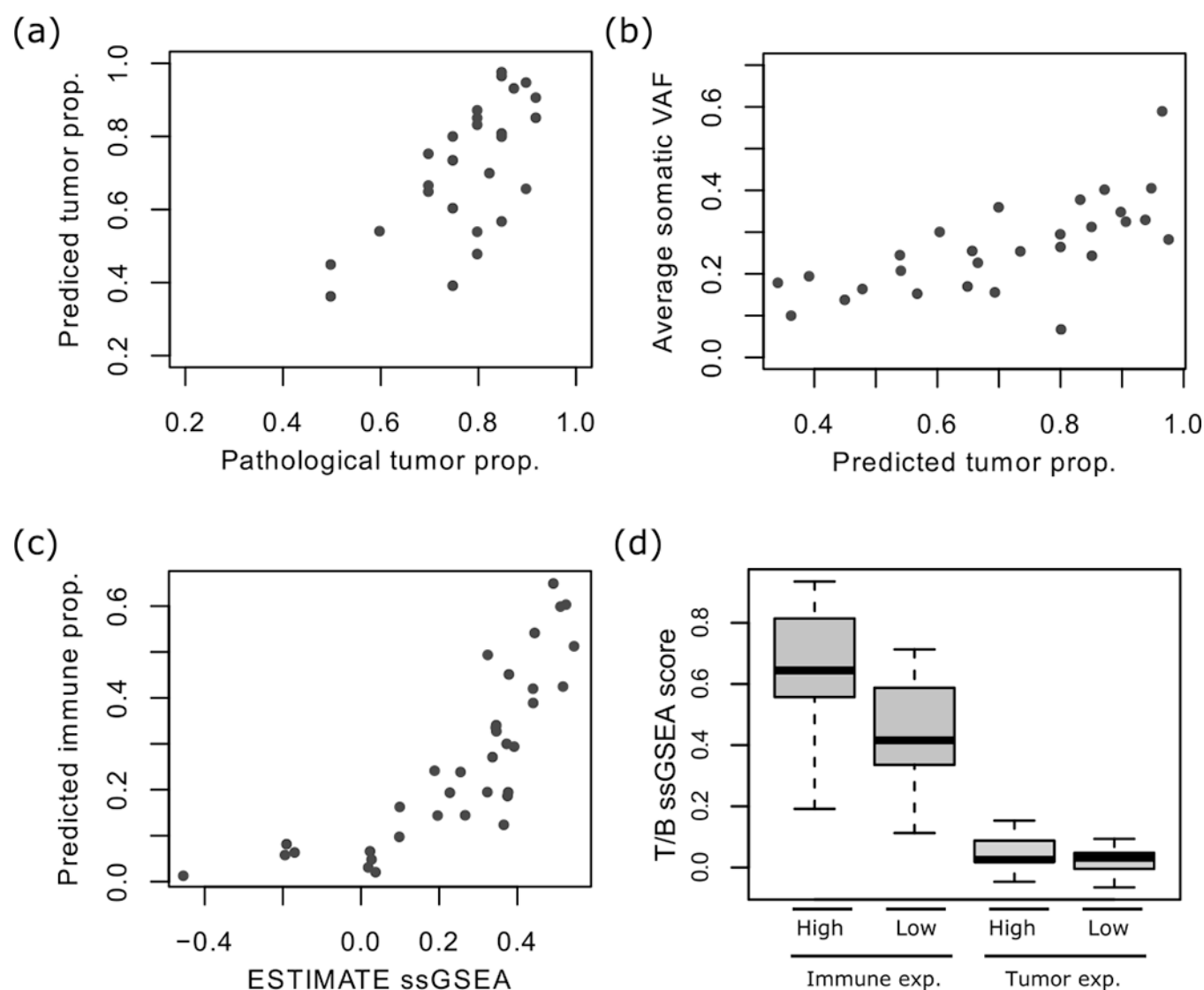


Fig. 1.

DisHet builds on the observation that only tumor cells are expanding in the mouse tumorgraft. (a) Average logged expression of ESTIMATE genes in the bulk tumor, human cell component of tumorgraft, and mouse cell component of tumorgraft. As a control, the expression of three known RCC-specific genes is shown: FABP7 (purple), PTHLH (red), and CD70 (orange). N=35. Box boundaries represent interquartile ranges, whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range, and the line in the middle of the box represents median. (b) Average variant allele frequencies (VAFs) of somatic mutations in the human cell component of the tumorgraft samples compared with bulk tumor samples. N=38. (c) Proportion of genes with expression deviation larger than a series of cutoffs comparing primary tumor vs. tumorgraft or tumor cell culture. The calculation of proportions is either limited to all genes or genes whose expression in

tumorgraft and tumor cell culture is larger than that in the primary tumor. (d) Whole genome copy number variation in the primary tumor and tumorgraft of XP166 (11th passage) and XP243 (1st passage). Regions where copy number variations are detected are drawn for primary tumor (green) and tumorgrafts (red). Black line refers to copy number of 2. (e) Rationale of DisHet for dissecting bulk tumor RNA-Seq samples. The normal tissue RNA-Seq dataset provides expression of the normal parenchyma. The tumorgraft RNA-Seq dataset provides expression of the tumor cells, after the human cell and mouse cell transcription profiles have been separated by a competitive alignment process. The bulk tumor gene expression contains expression of the immune/stroma component, tumor cell component and the normal parenchyma component. The latter two components and the whole tumor mixture are observed from the RNA-Seq data, which allows DisHet to dissect the unknown immune/stroma component.

**Fig. 2.**

Validation of the dissection results of DisHet. (a) Correlation of predicted tumor proportions and proportions assessed by pathological reviews. (b) Correlation of predicted tumor proportions and average VAF of somatic mutations. (c) Correlation of ssGSEA scores of the ESTIMATE signature in the bulk tumor and predicted immune/stroma component proportions. (d) ssGSEA signature scores (Y axis) for combined signature of T cells and B cells in patient samples whose flanking H&E staining slides were called by pathologist (P.K.) to have high and low lymphocytic infiltration levels (X axis). The same analysis is done in tumor cells as control.

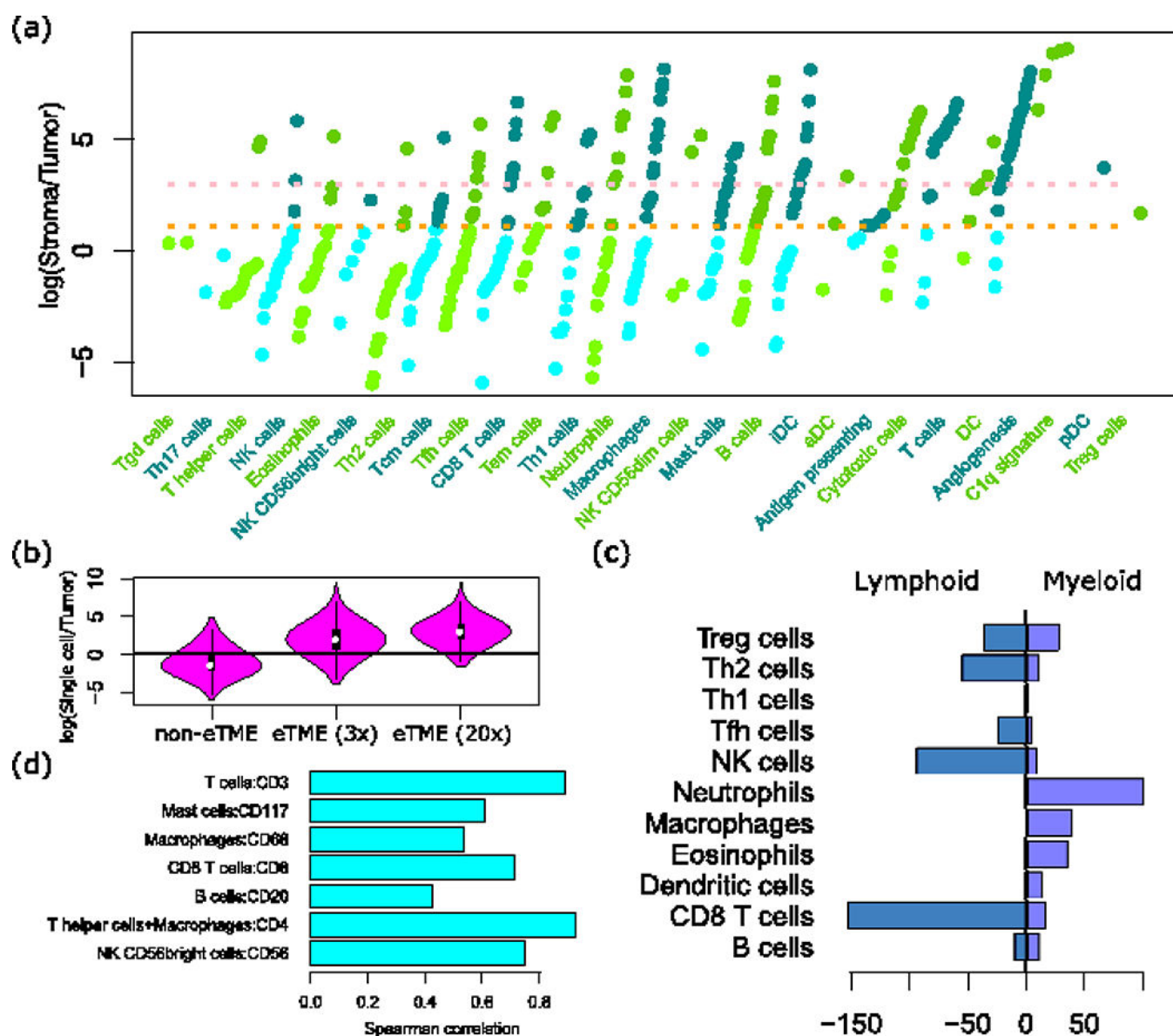


Fig. 3.

Defining the eTME genes. (a) Shows the relative expression level of each gene of each previously published signature (Immunome and the complement signature from Winslow) in the immune/stroma component over tumor component. Dotted lines show 3× and 20× cutoffs. (b) Aggregated single cell gene expression normalized by expression of the same gene in the tumor component. “Non-eTME” genes refer to genes not belonging to the eTME genes (3x), “eTME (3x)” and “eTME (20x)” refer to the eTME genes defined with 3× and 20× fold cutoffs. (c) eTME gene signatures assigned cell types to 969 cells according to their single cell RNA-Seq data. (d) Validation of refined gene signatures using Immunohistochemistry (IHC). RNA-Seq data from 7 patients were analyzed to predict the immune cell infiltration levels using gene signatures refined by the eTME criteria for a series of cell types by ssGSEA. IHC was performed on FFPE samples from the same patients for validation. Spearman correlations between them were shown in histograms.

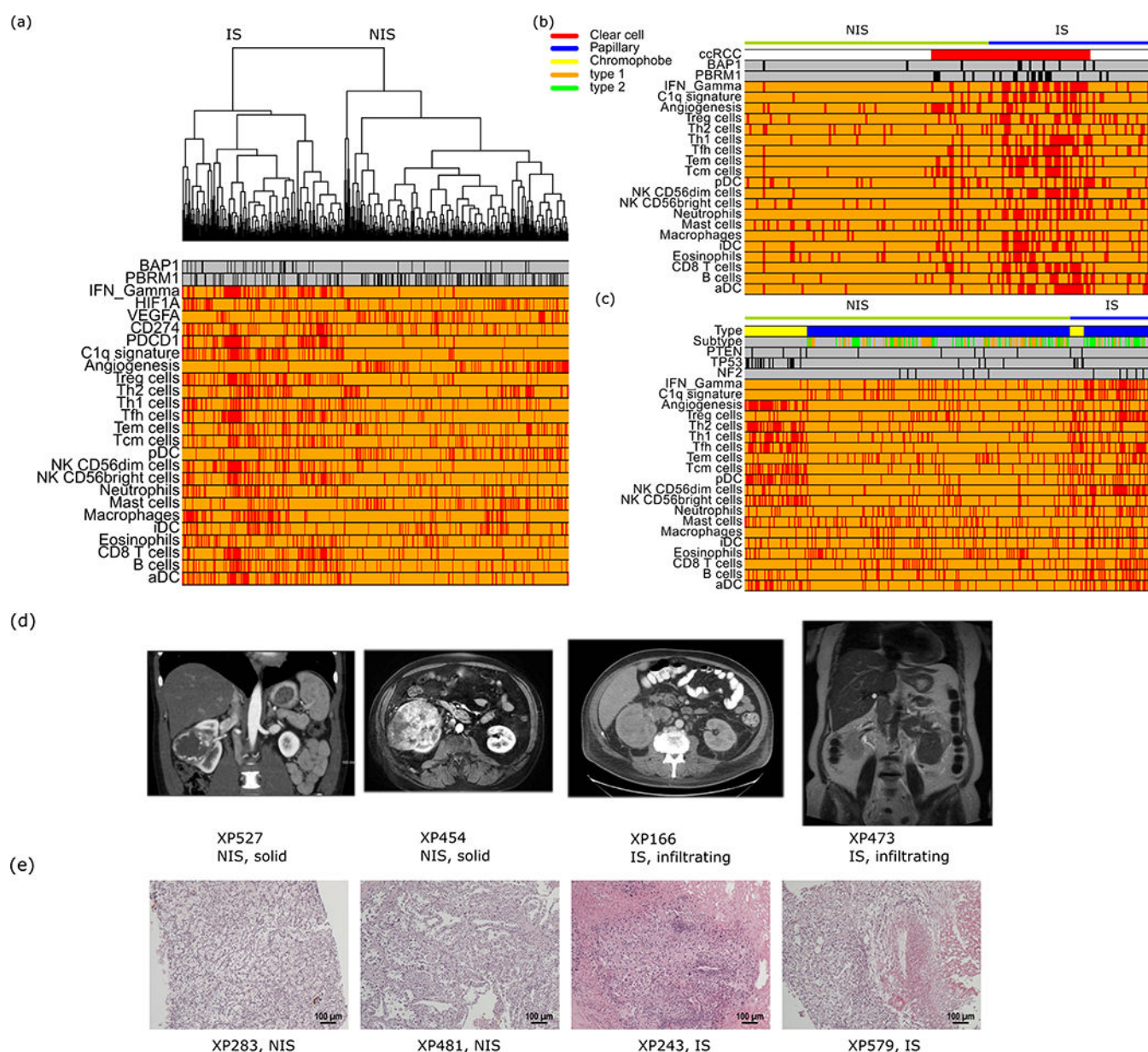


Fig. 4. Discovery of a highly-inflamed pan-RCC subtype based on eTME expression profiles. (a) Hierarchical clustering of TCGA ccRCC tumors based on eTME genes that passed the cutoff of 20x. *BAP1* and *PBRM1* mutations are shown as black bars. ssGSEA activation scores for immune cells and stromal pathways are shown, with red meaning higher activation scores. Validation of the eTME-based subtyping in the KCP cohort (b) and TCGA nccRCC cohort (c). The eTME-based subtyping is predicted by a SVM model. In (b), red bars denote ccRCC (N=70) and white bars denote all other RCCs (N=111). In (c), yellow bar refers to chRCC and blue bars refer to pRCCs. Green bars refer to type 2 pRCCs and orange bars refer to type 1 pRCCs. (d) Radiological assessment of eTME-IS/NIS tumor types shows correlation of eTME-IS with infiltrative margins. Representative MRI/CT images of eTME-NIS and eTME-IS patients were shown. (e) Pathological assessment of

eTME-IS/NIS tumor types shows correlation of eTME-IS with necrosis. Representative H&E staining images of eTME-NIS and eTME-IS patients were shown.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

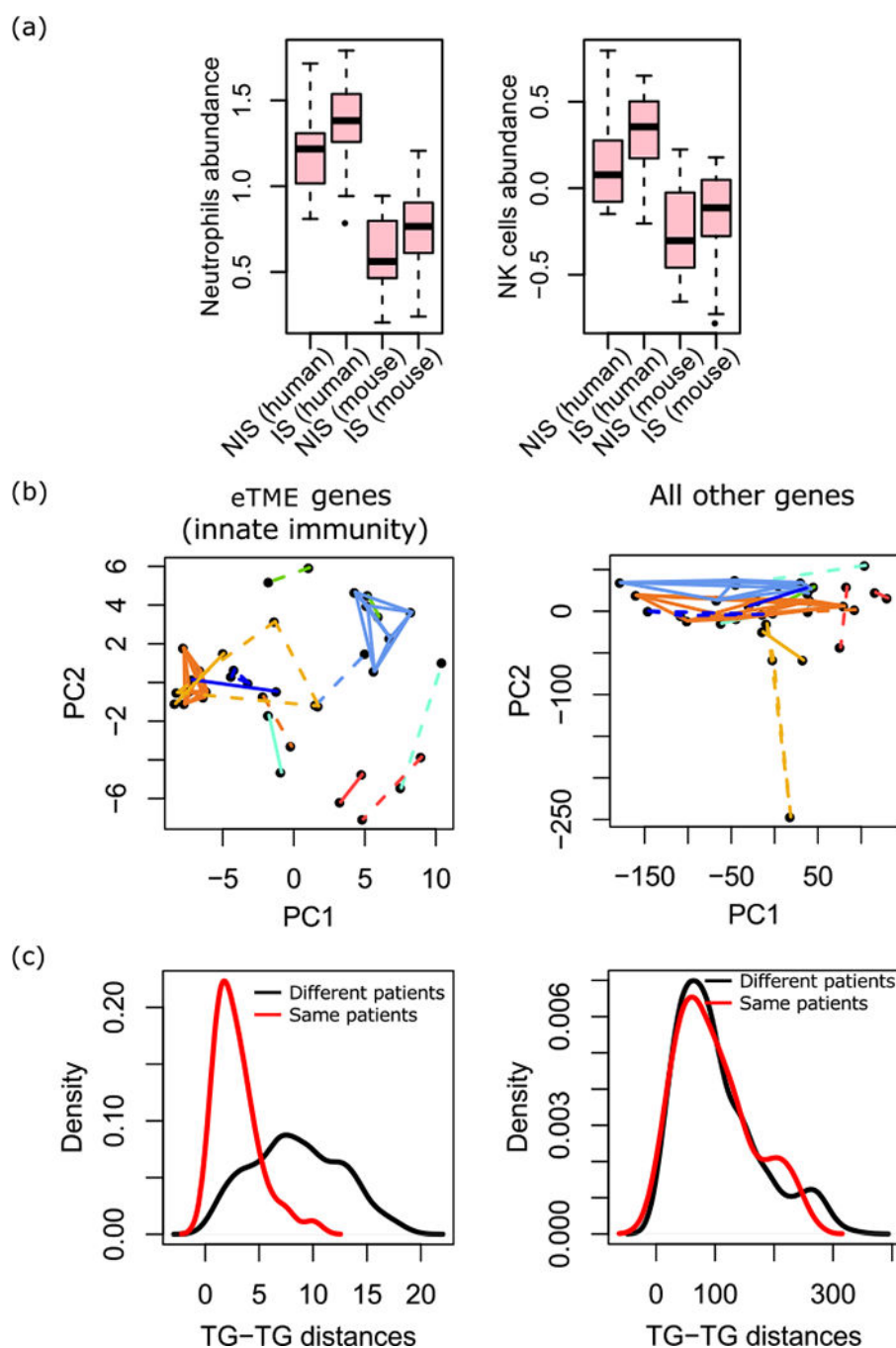


Fig. 5. Tumors drive stromal responses when engrafted in mice. (a) Infiltration levels of neutrophils, and NK cells in the human and mouse stroma of eTME-NIS (N=14) and eTME-IS (N=20) patients. (b) PCA plot of 37 tumorgraft samples corresponding to 14 unique patients based on expression of eTME genes (cutoff of 3x) that belong to the innate immune system (left panel) and non-eTME genes (3 fold cutoff, right panel). Tumorgraft samples from the same patients are connected by lines of different combinations of colors and styles. (c) Euclidean distances between tumorgrafts from the same patients (red) or different patients (black),

defined by expression of eTME genes that belong to the innate immunity system (left panel) and non-eTME genes (right panel).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

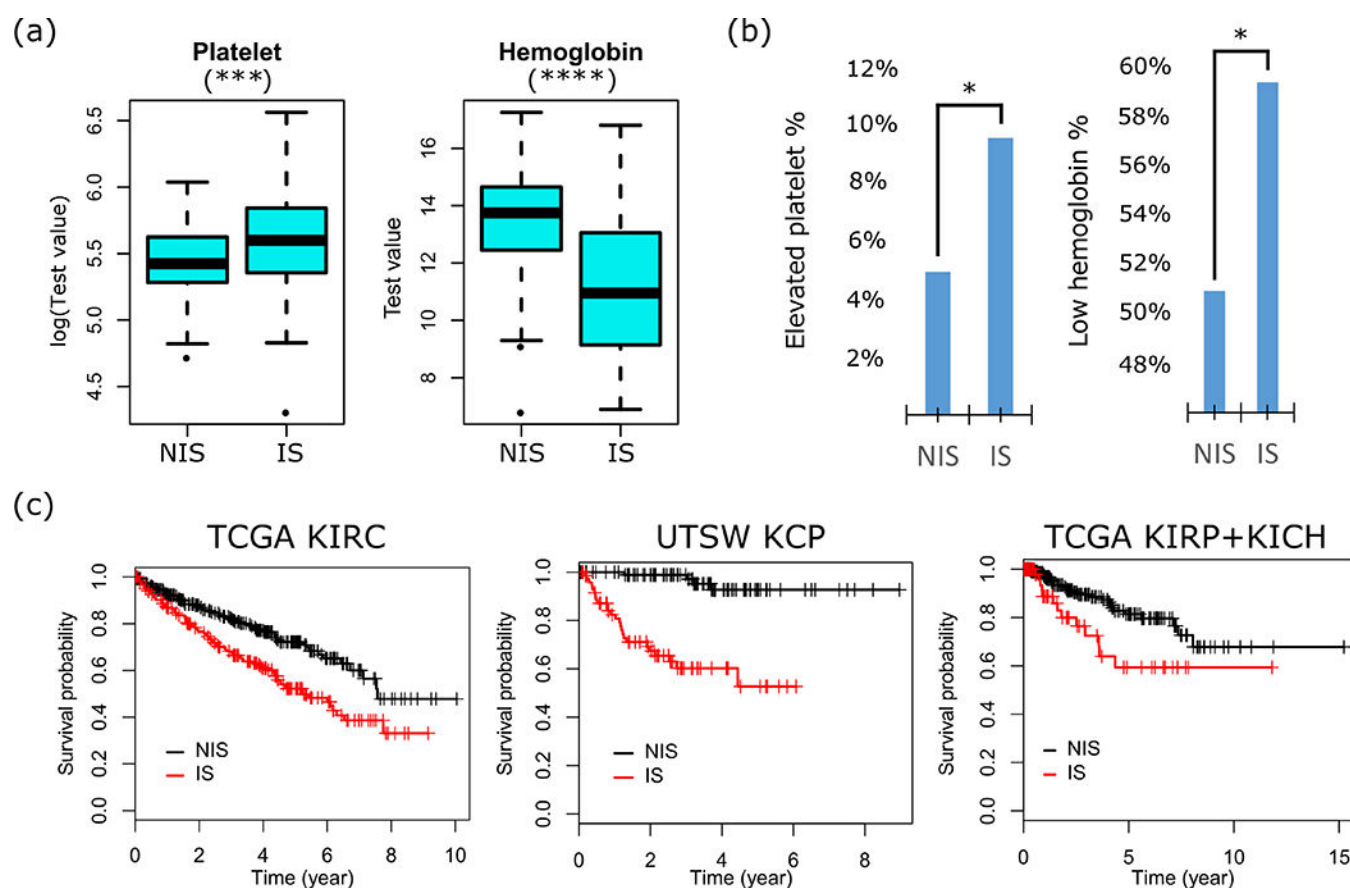


Fig. 6. Inflammation in the tumor microenvironment correlates with systemic manifestations of inflammation and worse prognosis. (a) Platelet (N=168) and hemoglobin (N=169) levels for eTME-NIS and eTME-IS patients in the UTSW KCP cohort. Units: platelet, log(counts/uL); hemoglobin, g/dL. (b) Percentages of eTME-NIS and eTME-IS patients with elevated platelet (N=393) and low hemoglobin (N=444) levels in the TCGA pan-RCC cohort. (c) Kaplan-Meier estimators for TCGA ccRCC patients (N=529), UTSW KCP RCC patients (N=181), and TCGA pRCC/chRCC patients (N=351), respectively. Patients are dichotomized by the eTME-NIS and eTME-IS designations.