







ORIGINAL RESEARCH

 OPEN ACCESS 

Intratumoral T-cell receptor repertoire composition predicts overall survival in patients with pancreatic ductal adenocarcinoma

Vikram S. Pothuri ^{a*}, Graham D. Hogg ^{b*}, Leah Conant^a, Nicholas Borcharding^b, C. Alston James^a, Jacqueline Mudd^a, Greg Williams^a, Yongwoo David Seo^{c,d}, William G. Hawkins^{a,e}, Venu G. Pillarisetty^{c,f}, David G. DeNardo ^{b,e*}, and Ryan C. Fields ^{a,e*}

^aDepartment of Surgery, Washington University School of Medicine, St. Louis, MO, USA; ^bDepartment of Medicine, Washington University School of Medicine, St. Louis, MO, USA; ^cDepartment of Surgery, University of Washington School of Medicine, Seattle, WA, USA; ^dDepartment of Surgical Oncology, MD Anderson Cancer Center, Houston, TX, USA; ^eSiteman Cancer Center, Washington University School of Medicine, St. Louis, MO USA; ^fFred Hutchinson Cancer Center, Seattle, WA USA

ABSTRACT

Pancreatic ductal adenocarcinoma (PDAC) is a lethal malignancy that is refractory to immune checkpoint inhibitor therapy. However, intratumoral T-cell infiltration correlates with improved overall survival (OS). Herein, we characterized the diversity and antigen specificity of the PDAC T-cell receptor (TCR) repertoire to identify novel immune-relevant biomarkers. Demographic, clinical, and TCR-beta sequencing data were collated from 353 patients across three cohorts that underwent surgical resection for PDAC. TCR diversity was calculated using Shannon Wiener index, Inverse Simpson index, and "True entropy." Patients were clustered by shared repertoire specificity. TCRs predictive of OS were identified and their associated transcriptional states were characterized by single-cell RNAseq. In multivariate Cox regression models controlling for relevant covariates, high intratumoral TCR diversity predicted OS across multiple cohorts. Conversely, in peripheral blood, high abundance of T-cells, but not high diversity, predicted OS. Clustering patients based on TCR specificity revealed a subset of TCRs that predicts OS. Interestingly, these TCR sequences were more likely to encode CD8⁺ effector memory and CD4⁺ T-regulatory (Tregs) T-cells, all with the capacity to recognize beta islet-derived autoantigens. As opposed to T-cell abundance, intratumoral TCR diversity was predictive of OS in multiple PDAC cohorts, and a subset of TCRs enriched in high-diversity patients independently correlated with OS. These findings emphasize the importance of evaluating peripheral and intratumoral TCR repertoires as distinct and relevant biomarkers in PDAC.

ARTICLE HISTORY

Received 7 November 2023
Revised 13 February 2024
Accepted 14 February 2024

KEYWORDS

Pancreatic cancer; T-cell receptor repertoire; cancer immunology; tumor microenvironment; immunotherapy; biomarkers

Introduction


Pancreatic Ductal Adenocarcinoma (PDAC) is on track to become the second leading cause of cancer-related death by 2030 and remains refractory to checkpoint therapies.^{1,2} Despite poor efficacy of T-cell-directed immunotherapies in PDAC, T-cell infiltration and neo-antigenicity appear to be predictors of outcomes.^{3,4} This suggests that the composition of the T-cell receptor (TCR) repertoire could also predict outcomes and perhaps be used to identify patients who could benefit from checkpoint inhibitors. Multiple studies have analyzed intratumoral TCR repertoires using TCR clonality or its inverse, diversity.⁵ High diversity indicates a high number of unique TCRs that are evenly distributed, while high clonality indicates a disproportionate expansion of a subset of TCRs.^{5,6} Across multiple cancer types, including melanoma, lung, and breast, elevated TCR diversity is associated with increased overall survival (OS).⁷ However, in melanoma, patients with low TCR diversity (high clonality) prior to treatment demonstrate an improved response to PD-1 blockade and longer progression-free survival.⁶⁻⁸ These studies suggest that the ideal TCR

repertoire may vary by cancer type and treatment modality. In 60 PDAC patients, long-term survivors had a more diverse intratumoral TCR repertoire than short-term survivors.³ However, this was demonstrated in the absence of controlling for T-cell abundance or clinical characteristics.³ We therefore sequenced the complementarity determining region 3 (CDR3) of peripheral blood and tumor-infiltrating TCRs to better understand whether repertoire diversity predicts OS in PDAC. Our analysis revealed that across three cohorts of patients with PDAC, intratumoral TCR diversity was a stronger predictor of OS than T-cell abundance, suggesting that a diverse repertoire with broad specificity may play a more significant role in determining patient outcomes than sheer intratumoral T-cell abundance.

The CDR3 region of the TCR-beta chain is a unique barcode for the specificity of a particular T-cell. Repertoire diversity provides insights into the global adaptive response; however it is difficult to draw conclusions about specificity to conserved tumor antigens. Novel tools have emerged to better predict both the antigen specificity of a particular CDR3 sequence

CONTACT David G DeNardo  ddenardo@wustl.edu; Ryan C Fields  rcfields@wustl.edu  Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, 63110, USA

*Co-first Authors.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/2162402X.2024.2320411>.

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

(*NetTCR*), and to group CDR3 sequences by structural motifs (*GLIPH2*, *TCRdist3*, *TESSA*).^{9–15} Structurally-related CDR3 sequences may bind the same peptide-major histocompatibility complex (p-MHC). The TCR repertoire is vast, with $\sim 10^8$ distinct rearrangements in any individual, making it rare to observe shared clonotypes.^{16,17} Motif-based algorithms such as *GLIPH2* and *TCRdist3* can identify highly similar but non-identical “metaclusters” of TCR-beta sequences with shared specificity.^{18,19} To investigate the relationship between TCR diversity and repertoire composition, we employed a novel technique of unsupervised nearest-neighbors analysis to cluster patients based on shared TCR specificity.¹⁰ Our analysis identified a cluster of patients with high intratumoral diversity and improved OS. The TCR repertoires of these patients were enriched for a set of TCRs with shared antigen specificity and amino acid properties that independently predicted OS in two additional datasets. Interestingly, these CDR3-beta rearrangements were more likely to encode effector memory CD8⁺ T-cells and T-regulatory (Treg) CD4⁺ T-cells, with the potential to bind pancreas-specific autoantigens.

We present evidence that despite the relatively low mutational burden of PDAC, T-cell mediated anti-tumor immunity may still delay disease progression.²⁰ By identifying TCR repertoire characteristics associated with survival, we hope to help identify patients that may benefit from T-cell-directed therapies.

Methods

Study population

Washington University in St. Louis (WashU): 162 patients with PDAC, treated at Siteman Cancer Center (2011–2019) were included. Tumor and peripheral blood mononuclear (PBMC) samples were taken for 122 patients, and for 40 additional patients, only PBMC samples were taken. Ten patients were excluded due to missing data, yielding 152 patients with PBMC and tissue and 116 patients with tissue only. **University of Washington (UW):** Previously published clinical and immunosequencing data was accessed via collaboration.²¹ The cohort included 54 patients with resected PDAC and three were excluded due to missing data, yielding 51 with tissue data. **The Cancer Genome Atlas (TCGA):** RNA-seq data was available for 171 patients from the PAAD cohort. Clinical data were downloaded from the TCGA database. Nine were excluded due to missing data, yielding 162 patients with RNA-seq data.²² **MD Anderson (MDA):** 5'-scRNA-seq and clinical data was available for seven patients with PDAC.²³

TCR repertoire data processing and analysis

UW and WashU: UW and WashU cohorts were sent to Adaptive Biotechnologies, genomic DNA was extracted, and immunosequencing of the CDR3 regions of TCR-beta chains was performed using the immunoSEQ Assay. Data is available on the Adaptive Biotechnologies Portal. WashU ImmunoSEQ data was converted to the VDJTools clonotype table,²⁴ filtered to remove nonfunctional clonotypes, and analyzed with VDJTools diversity functions. PBMC and tissue samples were

downsampled (10,000 and 5,000 reads/clonotype respectively) to account for the differential depth of sequencing. Nine tissue samples were excluded due to low clonotype counts. To verify the VDJTools analysis, we utilized a model created by Bortone et al. that produces corrected diversity estimates.⁶ The *predict_true_entropy_from_counts* routine was used to calculate True entropy.

TCGA: TCR CDR3 sequences were extracted from NCI GDC Data Portal PDAC bulk RNAseq data with the *MiTCR* package.²² Twelve samples were excluded due to low CDR3-beta clonotype counts, yielding 150 samples for analysis.

TCR metacluster analysis and unsupervised clustering

TCR distances were then computed between expanded intratumoral clones (productive frequency > 0.1%) representing the ‘centroids’ and all other intratumoral clones with the python *tcrcdist3* package (v0.2.0).¹⁰ A *TCRdist* radius of 15 was used as a cutoff. TCR distances are calculated based on a weighted *BLOSUM62* amino acid substitution matrix weighted 3× toward the CDR3 junction residues. CDR1, CDR2, and CDR2.5 residues are also incorporated into the model and are inferred from the VDJ usage.¹⁰ Resulting metaclusters were used as the features of a sparse matrix, containing the sum of the productive frequency of rearrangements within the *TCRdist* radius by patient. The sparse matrix was processed as a single-cell-experiment.²⁵ Data were log normalized by multiplying by 10^6 and taking the natural log of the value + 1. Variable features were selected (p value < 0.25), scaled, and used for principal components analysis (scatter package: v1.1.8).²⁶ The first 40 principal components were used to generate a K-nearest-neighbors map, ($k = 50$). Clustering was performed with the *igraph* (v1.2.6) implementation of the Louvain algorithm. Dimensional reduction was performed with the R implementation of the UMAP algorithm using the first 6 principal components as determined by elbow plot.²⁷

TCR dimensional reduction with TESSA autoencoder

Amino acid sequences for CDR3-beta sequences were input into the Briseis Autoencoder to generate TCR embeddings.¹¹ In brief, a matrix of five amino acid properties or Atchley factors multiplied by the length of the CDR3 rearrangement was dimensionally reduced with a neural network to a single vector of length 30, or the TCR embedding. This 30-variable embedding was used to plot TCR relationships with the python implementation of the UMAP algorithm with 100 neighbors (*umap-learn*: v0.5.2). *TESSA* subsets were identified using the python density-based clustering algorithm DBSCAN (*hdbscan*: v0.8.28).

Single cell RNAseq analysis

Single Cell 5' FASTQ files were aligned with *cellranger* v6.1.2 to map whole transcriptome and VDJ sequences using the GRCh38–2020 and GRCh38-alts-ensembl-7.1.0 references provided by 10× Genomics. Filtered feature barcode matrices were processed with *Seurat* v4.2.0. Barcodes with fewer than 200 features or more than 10% mitochondrial genes were

removed. Counts were normalized with the *sctransform* function (v0.3.5).²⁸ T-cell states were defined with the *ProjectTILs* package (v3.0.1).²⁹

TCR – p-MHC specificity prediction

CDR3-beta rearrangements were queried against the IEDB database of known TCR-p-MHC interactions (accessed 2/10/23). CDR3-beta similarity was calculated with the *TCRmatch* algorithm, which utilizes a k-mer based comparison of two sequences, scoring each position with a BLOSUM62 amino acid substitution matrix.¹³ A significance cutoff of 0.97 was used.

Statistics

Cohorts were compared in Table 1 with chi-squared and ANOVA. The tissue samples from the WashU cohort were combined with the UW cohort. Each cohort was divided into “Upper Quartile” and “Lower 3 Quartiles” based on True entropy. Fraction T-cells of nucleated cells is a metric provided by Adaptive Biotechnologies that quantifies the relative frequency of T-cells among nucleated cells. Each cohort was divided into “Above Median” and “Below Median” based on the fraction of T-cells. Cox proportional hazards models were used to compute hazard ratios (HR) and 95% confidence intervals (CI). Overall survival (OS) was defined as time between surgery and death or last follow-up if alive. The multivariable regressions were adjusted for factors that affect both TCR diversity and survival, including age, nodal status, neoadjuvant therapy, and fraction T-cells (when available). The proportional hazards assumption was satisfied for all models. The Kaplan – Meier method was used to generate survival curves with the log-rank test for statistical significance. Two-sided *p* values < 0.05 were considered statistically significant and false discovery rate correction was used for multiple comparisons. A beta-binomial model was used for differential analysis of metacluster abundance between patient clusters (FDR < 0.25) due to the inflation of zero values in TCR sequencing data (corncob: v0.3.1). In brief, metacluster cumulative

productive frequency was multiplied by 1×10^6 and treated as ‘Taxa’ in a differential model controlling for the effects of cluster identity on dispersion. Statistical analysis was performed with R v4.2.2 or Python v3.8.5.

Study approval

Tumor and blood samples were obtained from patients treated at Washington University School of Medicine in St. Louis, after patients signed written informed consent. Electronic medical records were accessed under the Institutional Review Board (IRB) approved protocol (201108117). This study adheres to the ethical standards established by the Declaration of Helsinki.

Results

TCR diversity is increased in peripheral blood compared to tumor

Few studies have comprehensively analyzed the composition of the TCR repertoire within human PDAC.^{3,21,30,31} To address this gap, we sequenced the CDR3-beta region in tumor and/or peripheral blood samples from 152 PDAC patients (WashU). Of the 152 patients, all included PBMC data and 116 included matched tumor-PBMC data. These pairs enabled a comparison of peripheral and intratumoral repertoires. Diversity accounts for two qualities: ‘richness’ (number of unique clonotypes) and ‘evenness’ (distribution of clonotypes).^{5,6} We initially used the Shannon-Weiner index, which reflects richness and evenness, and the Inverse Simpson index, which reflects evenness.^{5,6} Importantly, the principle of evenness provides insight into clonal expansion of TCRs, with higher levels of evenness corresponding to reduced clonal expansion. To accurately compare diversity metrics among patients with variable sequencing depth, we used down-sampling. Eight tissue samples fell below the down-sample threshold and were removed from this analysis. Additionally, we employed “True entropy” which is adjusted for disparities in read count, obviates the need for down-sampling, and enables diversity comparisons across cohorts.⁶ As expected, TCR diversity in PBMC was

Table 1. Clinical and pathologic characteristics of pancreatic ductal adenocarcinoma cases according to patient cohort.

Characteristics ^a	TCGA (n = 150)	Patient Cohort UW (n = 51)	WashU (n = 152)	<i>p</i> ^b
Sex				0.652
Female	67 (45%)	19 (37%)	65 (43%)	
Male	83 (55%)	32 (63%)	87 (57%)	
Mean age ± SD (years)	65.5 ± 11.1	63 ± 9.9	66.5 ± 10.3	0.116
Nodal Status				0.002
Negative	40 (27%)	26 (51%)	41 (27%)	
Positive	110 (73%)	25 (49%)	111 (73%)	
Mean Survival ± SD (days)	521.4 ± 424.4	1421.1 ± 1057.3	852.5 ± 699.4	<0.001
Survival Group				<0.001
PO	4 (3%)	0 (0%)	5 (3%)	
ST	29 (19%)	1 (2%)	24 (16%)	
MT	39 (26%)	25 (49%)	47 (31%)	
LT	13 (9%)	25 (49%)	37 (24%)	
s	65 (43%)	0 (0%)	39 (26%)	

^aPercentage indicates the proportion of patients with a specific clinical or pathologic characteristic among patients in each cohort. ^bChi-squared test used for categorical variables (with continuity correction) and regular ANOVA used for continuous variables. PO perioperative (≤3 months), ST short-term (>3 months and <1 year), MT medium-term (≥1 year and ≤3 years), LT long-term, S surviving (surviving, but <3 years). SD Standard Deviation.

significantly greater than within tumor, likely reflecting the greater number of naïve T-cells in circulation and the greater total number of T-cells (Figure 1(a–c)).³¹ There was concordance among the three metrics, suggesting that True entropy eliminates the need for down-sampling and would allow comparisons across cohorts and the formation of composite cohorts (Figure 1(a–c)).

Long-term survivors have increased TCR repertoire diversity within the tumor but not in PBMC

To investigate a relationship with OS, we assigned patients to survival groups based on previously published cutoffs: Short Term (ST), Medium Term (MT), Long Term (LT), Perioperative (PO), Survivors (S).³ We hypothesized that LT survivors would have decreased diversity, indicative of clonal proliferation of tumor-reactive T-cells.^{32,33} We compared the same diversity metrics across survival groups in both PBMC and tumor samples. Overall ANOVA testing revealed significant differences in intratumoral TCR diversity in tissue with all three indices (Figure 1(d–f)). Unexpectedly, pairwise t-tests with false discovery rate (FDR) correction showed significantly greater True entropy in LT survivors compared to MT survivors and significantly greater inverse Simpson index (more evenly distributed TCRs) in LT survivors compared to ST survivors (Figure 1d–f). There were also trends toward increased True entropy and Shannon Wiener index in LT survivors compared to ST survivors. Furthermore, the one hundred TCRs with the highest cumulative productive frequency made up less of the overall repertoire in LT survivors compared to ST and MT survivors (Figure S1). This corroborates the difference in inverse Simpson index in LT compared to ST survivors and suggests that intratumoral repertoire diversity is associated with OS. Interestingly, we observed no difference in the three diversity metrics in PBMC samples Figure 1(g–i). Additionally, we observed no significant variation in overlap between peripheral and intratumoral repertoires based on survival groups (Figure S2). This suggests that the composition of the intratumoral TCR repertoire may be a better predictor of outcomes than the peripheral repertoire and that increased intratumoral TCR diversity may be associated with more favorable outcomes. However, it is crucial to determine whether TCR diversity is an independent prognostic factor rather than a reflection of other patient characteristics.

TCR repertoire diversity decreases with age but is not affected by neoadjuvant therapy, stage, or grade

Increased age has been associated with both poorer outcomes and reduced TCR repertoire diversity, likely secondary to thymus involution and accumulation of clonally expanded memory T-cell subsets.¹⁷ In PBMC and tissue, there was a significant negative correlation between age and TCR diversity (Figure S3a–b).

We also hypothesized that tumor stage, grade, and neoadjuvant therapy may shape the TCR repertoire. In PBMC and tissue, there was no significant difference in either estimated T-cell fraction or TCR diversity when patients were divided by stage (Figure S4a–b) and only a difference in TCR diversity in

tissue when divided by grade (Fig. S5a–b). A sizable fraction of patients received neoadjuvant therapy (68 (44.7%) PBMC, 54 (46.5%) tissue). Patients were either assigned to neoadjuvant chemotherapy (Chemo), chemo-radiation (Chemo+Rad), or no neoadjuvant therapy (None). In PBMC, we observed a significant difference in TCR diversity across these three groups, but pairwise comparisons were nonsignificant (Figure S6b). In tissue, there was no significant difference in TCR diversity across these three groups (Figure S6b). There were also no overall differences in T-cell abundance in either PBMC or tissue with treatment (Figure S6a). Due in part to the heterogeneity of neoadjuvant regimens, further research is needed to assess this relationship.

High intratumoral TCR repertoire diversity predicts OS

In PBMC and tissue cohorts, patients were divided into diversity quartile groups. We again found a positive association between intratumoral diversity and OS, which was absent in the periphery (Figure S7a–b). Additionally, we observed that survival curves for the lower quartile and middle quartile groups largely overlapped, while the upper quartile diversity group had a significant survival advantage (Figure S7a). This suggests there may be a threshold of intratumoral repertoire diversity that is a strong predictor of outcomes. Thus, for the remainder of our analysis, we compared upper quartile diversity to all other quartiles.

For confirmation, we incorporated two additional datasets of intratumoral CDR3-beta sequences: a cohort with immunosequencing data from the University of Washington (UW) and a cohort with bulk-RNaseq extracted CDR3-beta sequences from TCGA. Due to the relatively small sample size of the UW cohort, we combined the UW and WashU cohorts for survival analysis. In this combined cohort, we observed a similar effect of upper quartile diversity on OS (Figure S7c).

To exclude the possibility that age, nodal status, neoadjuvant status or T-cell fraction were confounding the association between intratumoral repertoire diversity and OS, we used multivariate Cox regression models. We defined three cohorts: UW and WashU tissue, TCGA tissue, and WashU PBMC. In UW and WashU tissue, upper quartile intratumoral diversity was a strong and significant predictor of OS (HR (95% CI): 0.55 (0.34–0.89); Figure 2(a–f)). Interestingly, within the tumor, estimated T-cell fraction did not predict OS (Figure 2(b–f)). This suggests that the composition of the intratumoral TCR repertoire, rather than the sheer number of T-cells within the tumor, may be more important for anti-tumor immunity. The TCGA tissue data also demonstrated that upper quartile TCR diversity is a significant predictor of OS (HR (95% CI): 0.52 (0.28–0.99); Figure 2(e–f)). In multiple cohorts and when controlling for covariates, increased intratumoral TCR diversity is associated with improved OS.

In contrast to the tissue datasets, in the WashU PBMC data, upper quartile TCR repertoire diversity did not significantly predict OS (Figure 2(c–f)). However, the estimated T-cell fraction was a significant predictor of OS (HR (95% CI): 0.66 (0.43–0.998); Figure 2(d–f)). This suggests that in peripheral blood, the estimated T-cell fraction could either

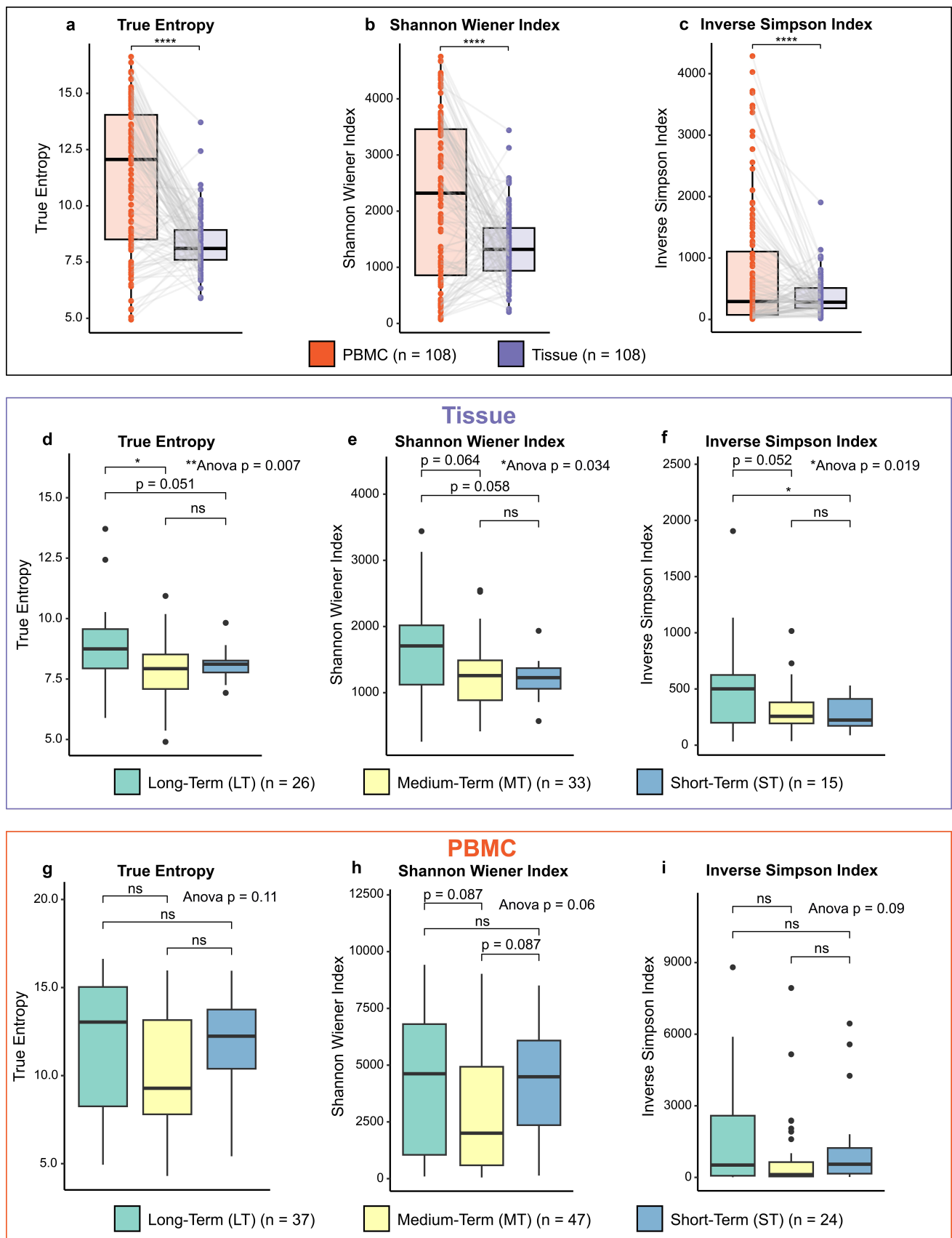


Figure 1. TCR Repertoire Diversity Metrics in the WashU Cohort. A-C: comparison of matched tissue and PBMC samples ($n = 108$) by (a) True entropy, (b) Shannon Wiener index, (c) Inverse Simpson index. D-F: comparison of survival groups in tissue samples by (d) True entropy, (e) Shannon Wiener index, (f) Inverse Simpson index (D: LT $n = 27$, MT $n = 39$, ST $n = 15$; E/F: LT $n = 26$, MT $n = 33$, ST $n = 15$). G-I: comparison of survival groups in PBMC samples by (g) True entropy, (h) Shannon Wiener index, (i) Inverse Simpson index (LT $n = 37$, MT $n = 47$, ST $n = 24$). *7 patients had clonotype counts below downsampling threshold and were excluded from E/F (A-C) **** $p < 0.0001$ as calculated by paired t-test. (D-I) * $p < 0.05$, ** $p < 0.01$ as calculated by ANOVA for overall differences and t-test with false discovery rate correction for multiple pairwise comparisons. ns = not significant

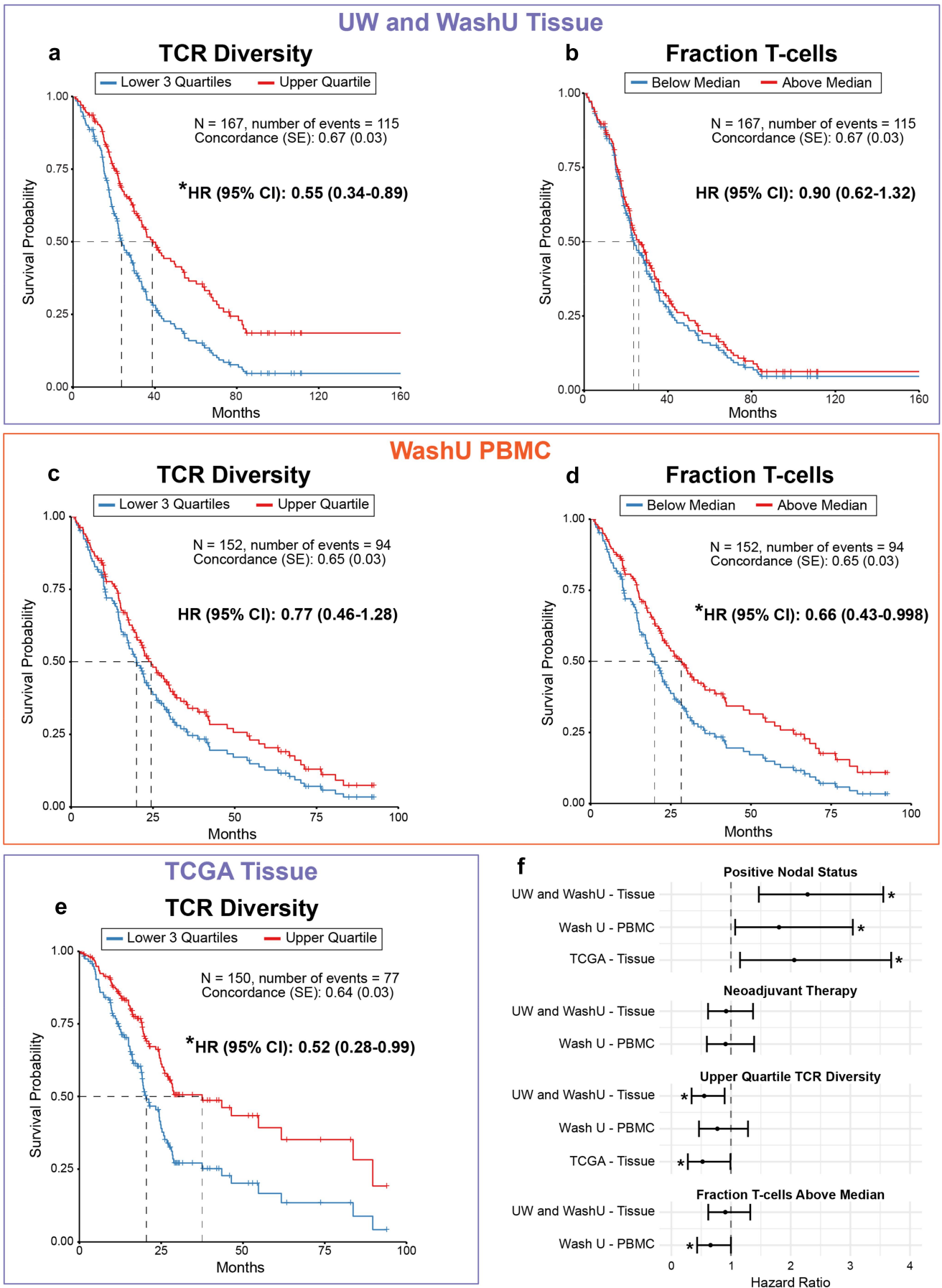


Figure 2. Multivariate cox proportional-hazards models for the UW and WashU – tissue, WashU – PBMC and TCGA – tissue cohorts. (a,b) Multivariate cox proportional-hazards model for UW and WashU – tissue cohort (N = 167), using the variables: age, nodal status, neoadjuvant therapy, TCR diversity (true entropy) quartile group and

indicate the magnitude of the T-cell response against the tumor or perhaps is more reflective of global immunocompetence.

Unsupervised analysis defines a repertoire cohort of LT survivors with high diversity

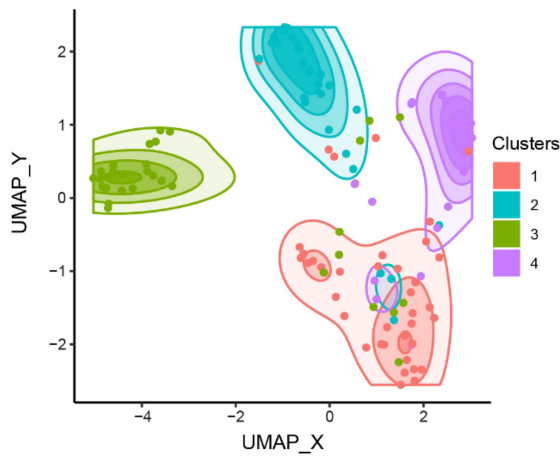
As shown, intratumoral T-cell repertoire diversity is predictive of OS; however, how that diversity relates to antigen specificities of the T-cell pool remains unknown. To address this, we sought to group patients by their shared TCR specificities. Tools like *TCRdist3* have been developed to infer shared TCR specificity from highly diverse CDR3 rearrangements with little overlap between individuals.³⁴ In tumor, where the resolution of CDR3 sequences is lower than blood, it is rare to observe identical rearrangements between individuals. Thus, grouping highly similar TCRs enables a robust comparison of TCR specificity groups or metaclusters between patients. We define metaclusters in this context as groups of multiple TCR CDR3 rearrangements that have a high probability of reacting with the same p-MHC.¹⁹ We utilized *TCRdist3* to identify intratumorally expanded metaclusters.¹⁰ Expanded clones were defined as having a productive frequency greater than 0.1%.¹⁷ In the WashU tissue cohort, we calculated the TCR distances (a measure of similarity) between tissue-expanded clones and all other clones. Rearrangements found to have a low TCR distance were grouped in metaclusters. These metaclusters were then used to construct a sparse matrix (Fig. S8a). The productive frequency of these metaclusters, each of which consists of TCRs with similar predicted specificity, was used for unsupervised clustering, resulting in four clusters of patients (Figure 3a). These patient clusters are defined as groups of patients with shared metaclusters, where each metacluster represents multiple TCRs with common specificity. Interestingly, Cluster 3 has significantly greater OS than all other clusters (Figure 3b-c). No significant differences were found in estimated T-cell fraction between clusters, indicating again that intratumoral T-cell infiltration alone is not a significant predictor of survival (Figure 3e). On the other hand, cluster 3, with the highest OS, had significantly increased intratumoral repertoire diversity compared to cluster 1, which had the lowest OS (Figure 3d). There was no significant difference in stage or age between patient clusters, and there was a nonsignificant trend toward a higher proportion of LT survivors in patient cluster 3 (Fig. S8c-e). This suggests that TCRs in patients from cluster 3 may represent a link between a highly diverse repertoire and distinct antigen specificities that confer a survival advantage.

Patient cluster 3 contains a subset of metaclusters that significantly predict OS in multiple datasets

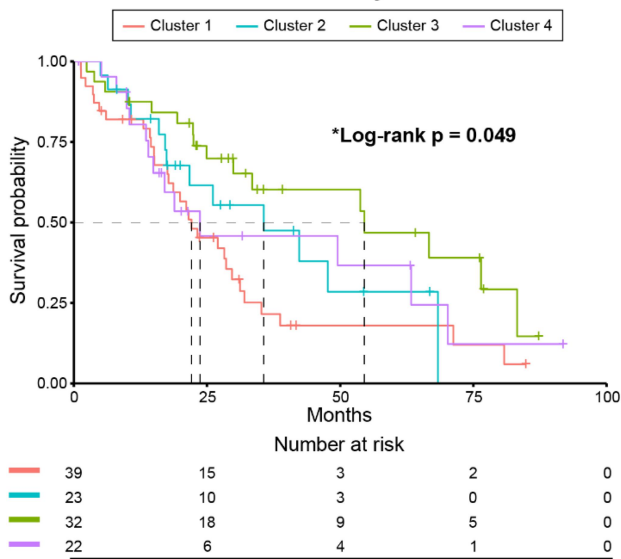
We next sought to understand which metaclusters within patient cluster 3 were driving the survival benefit. To perform differential testing on metacluster relative abundance, we employed a beta-binomial model (Figure 3f & S8F).³⁵ 571 metaclusters were enriched within patient cluster 3 with a lenient FDR cutoff of <0.25 . We next sought to identify a smaller subset of these 571 enriched metaclusters with shared structural relationships. TCRs belonging to the same metacluster are predicted to react to the same p-MHC; however, that does not preclude TCRs in separate metaclusters from having highly similar CDR3 sequences due to overlapping VDJ usage. To define a subset of metaclusters with shared structural motifs, we used the *TESSA* algorithm to group TCRs based on their amino acid properties.¹¹ The *TESSA* algorithm is a neural-network-based autoencoder that reduces the dimensionality of a matrix of amino acid properties for a single TCR to generate a latent dimensional representation of the CDR3 sequence.¹¹ We ran the *TESSA* autoencoder on all CDR3-beta sequences from the WashU tissue cohort and visualized the resulting matrix with UMAP dimensional reduction, where *TESSA* subsets were defined by the density-based *DBSCAN* method (Fig. S9a). With this method, *TESSA* subsets are defined as groups of TCRs that have shared amino acid properties. These *TESSA* subsets contain numerous metaclusters as defined by *TCRdist3*; however, some metaclusters may be distributed across a small number of *TESSA* subsets. Highly uncommon TCRs failed to cluster with any *TESSA* subset and were removed (Figure 4a). CDR3-beta sequences belonging to metaclusters enriched in patient cluster 3 were then highlighted (Figure 4b). These highlighted TCRs were then merged based on metacluster identity and colored by their *TESSA* subset assignment (vertex size corresponds to TCR counts) (Figure 4c). Metaclusters spanning multiple *TESSA* subsets were linked by gray edges (Figure 4c). There was significant overlap between metaclusters and *TESSA* subsets with 32% of metaclusters contained within one *TESSA* subset and roughly 75% of metaclusters contained in fewer than three *TESSA* subsets. No metacluster was distributed across more than five *TESSA* subsets and all but one was contained within four *TESSA* subsets (Fig. S9b-d). Additionally, on average, 82% of the TCRs in a given metacluster are found in a single *TESSA* subset (Fig. S9c). The patient cluster 3 enriched metaclusters (highlighted in Figure 4b) were separated based on their assigned *TESSA* subset. We hypothesized that by taking a group of TCRs with shared specificity that were associated with a survival advantage (metaclusters enriched in patient cluster 3) and separating them by amino acid properties (*TESSA*), this could reveal a smaller, core set of TCRs that confer a survival benefit. Thus, each subset was tested as

fraction T-cell group. Impact of upper quartile compared to lower 3 quartile TCR diversity (a) and impact of above median compared to below median fraction T-cells (b) on overall survival. (c,d) Multivariate cox proportional-hazards model for Wash U-PBMC cohort ($N = 152$), using the variables: age, nodal status, neoadjuvant therapy, TCR diversity (true entropy) quartile group and fraction T-cell group. Impact of upper quartile compared to lower 3 quartile TCR diversity (c) and impact of above median compared to below median fraction T-cells (d) on overall survival. (e) Multivariate cox proportional-hazards model for TCGA-Tissue cohort ($N = 150$), using the variables: age, nodal status, and TCR diversity (true entropy) quartile group. Impact of upper quartile compared to lower 3 quartile TCR diversity (e) on overall survival. F: hazard ratios for variables in the three models with 95% confidence intervals calculated with multivariate cox regression. (a-f) * $p < 0.05$ as calculated by multivariate Cox proportional-hazards model. Likelihood ratio test p value < 0.05 for all three models.

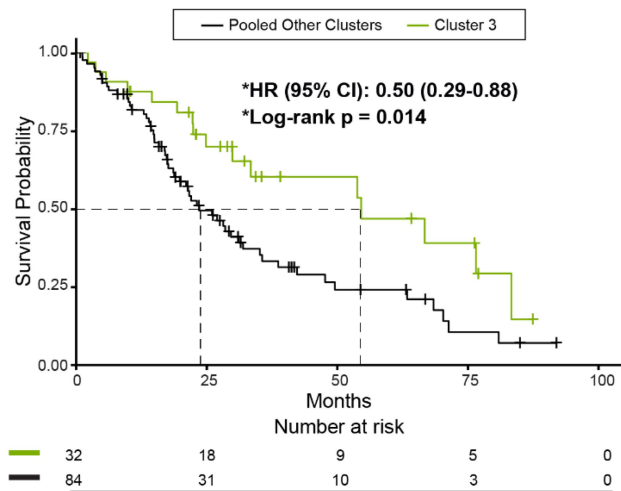
a Unsupervised Clustering of T-cell Repertoire



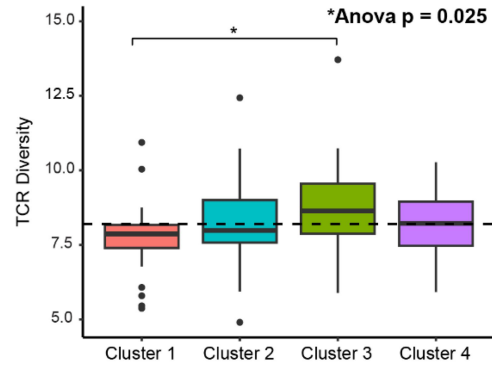
b Overall Survival by Clusters



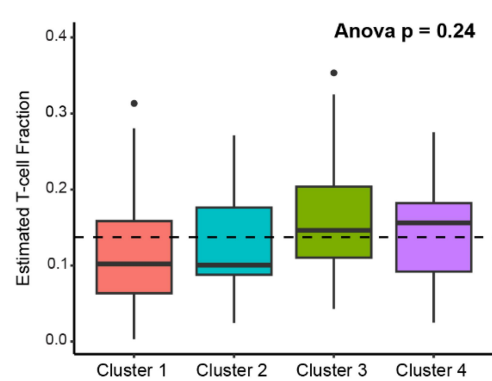
c Overall Survival by Pooled Clusters



d TCR Diversity



e Estimated T-cell Fraction



f Metacluster Differential Abundance

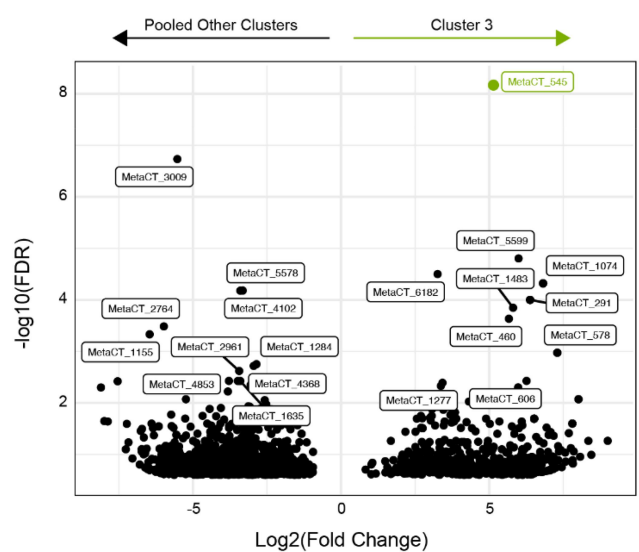


Figure 3. Clustering of WashU tissue patients by shared TCR specificity. (a) Unsupervised clustering of WashU tissue patients into four distinct patient clusters. (b) Kaplan-Meier curve of overall survival by individual cluster. (c) Kaplan-Meier curve of overall survival comparing cluster 3 to pooled clusters 1,2 and 4. D-E: comparison of patient clusters by (d) TCR diversity and (e) estimated T-cell fraction. F: relative abundance of TCR Metaclusters in patient cluster 3 compared to pooled clusters 1,2 and 4. (b) $p < 0.05$ as calculated by log-rank test. (c) $p < 0.05$ as calculated by log-rank test. HR (95% CI) calculated by univariate cox proportional hazards model. (d-e) $p < 0.05$ as calculated by ANOVA for overall differences and t-test with false discovery rate correction for multiple pairwise comparisons. (f) Calculated by beta-binomial regression model with significance cutoff of FDR < 0.25.

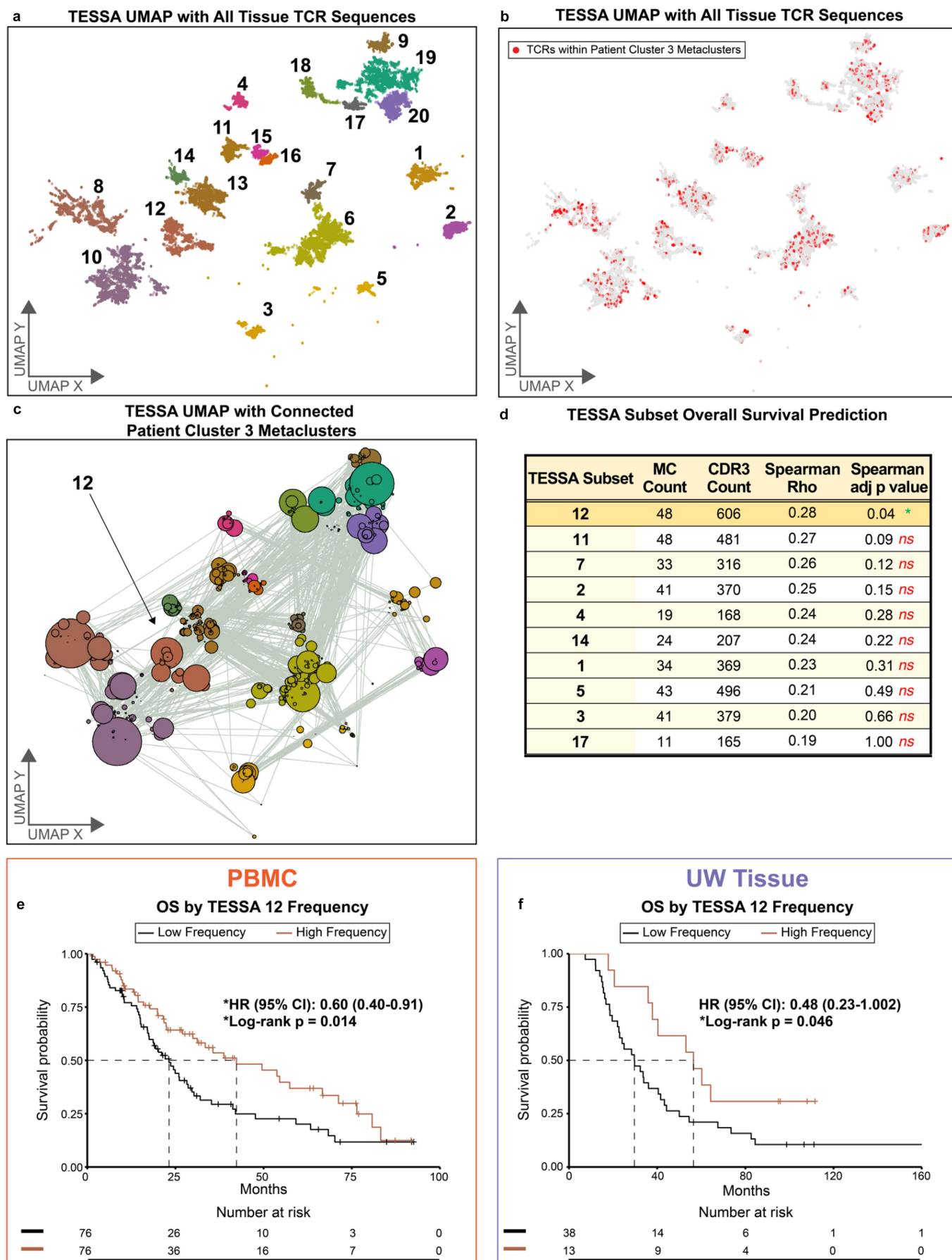


Figure 4. Clustering of WashU tissue TCRs based on amino acid properties (TESSA). A-B: UMAP dimensional reduction of TESSA embedding on all tissue CDR3b sequences from the WashU tissue data set with TCRs unassigned to a cluster removed (a) and with CDR3b sequences belonging to cohort 3 enriched metaclusters

a univariate predictor of OS in the WashU tissue cohort. Productive frequency of metaclusters within each *TESSA* subset was analyzed as a continuous variable. Interestingly, *TESSA* subset 12 had the highest correlation coefficient, and was the only subset to meet a significance threshold after adjusting for multiple comparisons (Figure 4d).

Having identified metaclusters within *TESSA* subset 12 as a predictor of survival in the WashU tissue dataset, we explored whether these metaclusters could be used as a generalizable biomarker of OS. We first tested whether the *TESSA* 12 metaclusters predicted survival in the WashU PBMC dataset, which includes an additional 36 patients that lacked tissue data. When patients were split by the median cumulative productive frequency of *TESSA* 12 metaclusters, high frequency significantly predicted OS (HR (95% CI): 0.60 (0.40–0.91) log-rank $p = 0.014$, Fig. 4e). The distribution of *TESSA* 12 cumulative productive frequency was similar in PBMC and tissue, but the mean productive frequency is higher in tissue compared to PBMC (0.34% vs. 0.16%, $p = 0.01$, Fig. S10a-c). Taken together this indicates that the *TESSA* 12 metaclusters could represent a clinically relevant tumor enriched subset of TCRs.

We next used the UW tissue dataset to corroborate the prognostic value of *TESSA* 12 metacluster frequency. Although this is a small dataset ($n = 51$), making survival analysis difficult, when we identified matched *TESSA* 12 metaclusters in the UW tissue dataset (*TCRdist* radius < 16) and stratified patients based on productive frequency, we again found that *TESSA* 12 metaclusters were predictive of OS ($p = 0.046$, Figure 4f). Both the hazard ratio and Pearson correlation with OS trended toward significance (HR (95%CI): 0.48 (0.23–1.002) $R = 0.24$, $p = 0.086$) (Figure 4f and S11a). These results were not attributable to differential intratumoral T-cell abundance as the high-frequency *TESSA* 12 group had a significantly reduced estimated T-cell fraction (Fig. S11b). Here we have shown that a group of TCRs with shared specificity and amino acid properties that are found within patients with diverse TCR repertoires can be used as a biomarker in both blood and tissue to predict OS in PDAC.

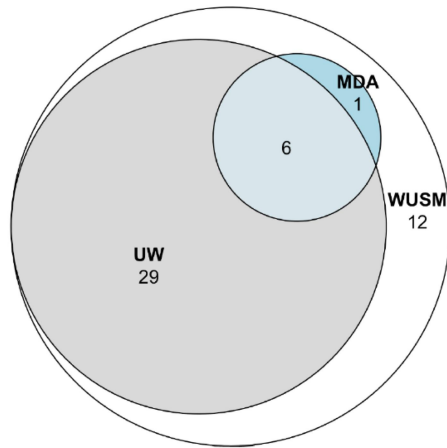
TESSA 12 metaclusters contain a higher frequency of autoreactive effector CD8⁺ T-cells

Having identified a set of metaclusters of interest, we next investigated their antigen specificity and transcriptional identity. We employed a newly published dataset from MD Anderson (MDA) of 5-prime single-cell sequenced sorted CD3⁺ tumor-infiltrating T-cells from patients with resectable PDAC.²³ 5-prime single-cell RNA sequencing includes the CDR3-beta region, permitting the union of TCR specificity data with transcriptomic profiles. We used *TCRdist3* to identify CDR3-beta rearrangements predicted to have the same

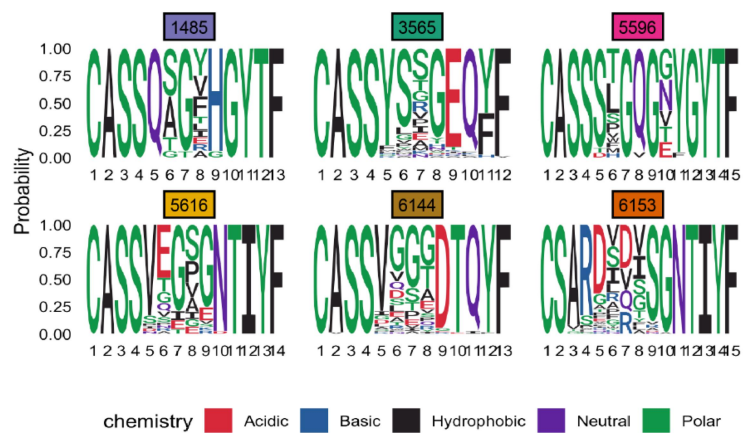
specificity as *TESSA* 12 metaclusters. The utility of TCR distances, rather than exact rearrangements to determine shared specificity was evident, as we were able to identify six metaclusters shared among all three PDAC tissue datasets (WashU, UW and MDA) (Figure 5a). In contrast, no identical clones were shared among all three datasets (Fig. S11c). Of the six shared metaclusters, five had a similar start sequence of CASS (Figure 5b). Additionally, 80% of the sequences contained the TRBJ1 segment, with the remaining sequences containing TRBJ2, both of which end in polar Threonine or Tyrosine residues (Figure 5b). We next inferred the T-cell identities in the MDA dataset using the *ProjectTILs* package to map the transcriptomes onto a reference dataset (Figure 5c).²⁹ *TESSA* 12 metacluster matched T-cells were then highlighted on the UMAP (Figure 5d). Interestingly, when comparing the identities of all MDA sequenced T-cells with the subset of *TESSA* 12 metacluster T-cells, there was a significant enrichment of CD8⁺ effector memory T-cells ($p = 0.046$) and a reduction in CD8⁺ naïve-like T-cells ($p = 0.021$ Figure 5e). This suggests that *TESSA* 12 CD8⁺ T-cells tend to be less naïve and more activated, characteristic of a tumor-reactive T-cell population. Among CD4⁺ T-cells, there was an expansion of Tregs ($p = .009$ Figure 5e). Tregs can be a negative prognostic factor, as they function to enforce self-tolerance by dampening potentially harmful adaptive immune responses to autoantigens or commensal pathogens.³⁶ However, there are conflicting reports as to whether Tregs are negatively or positively associated with OS in PDAC patients.^{37,38} The expansion of Tregs in *TESSA* 12 metaclusters, suggests that these TCRs may have the capacity to react with self. We therefore assessed the antigen specificity of the six *TESSA* 12 metaclusters shared across the three cohorts (Figure 5a). We utilized the *TCRmatch* algorithm through the *IEDB* database, which queries CDR3-beta sequences against a database of known TCR – p-MHC interactions.³⁹ Sequences from all six shared metaclusters were submitted, and a confidence cutoff of 0.97 was used as suggested by the initial publication.¹³ In two of the six metaclusters, we identified TCRs with known autoreactivity to beta islet-derived proteins. Specifically, metacluster #5616 contained a TCR with reactivity to an insulin peptide, while metacluster #6144 contained a TCR with reactivity to Zinc Transporter 8a (SLC30A8), both of which are expressed in the endocrine pancreas and associated with Type 1 Diabetes (Figure 5f).^{40,41} The CDR3 sequence specific to the insulin peptide was predicted to react with HLA-DQB1 *03:02, a heterodimer of MHC class II recognized by CD4⁺ T-cells. As predicted, one of the T-cells within metacluster #5616 was assigned as a Type 1 CD4⁺ T helper cell (Th1), suggesting that a T-cell specific to an insulin peptide could bind antigen-presenting cells, proliferate, and produce inflammatory IFN γ . While negative in most circumstances, an autoimmune

highlighted (b). (c) Highlighted TCRs in (b) belonging to a given metacluster within the same *TESSA* subset were collapsed into vertices and those belonging to a matched metacluster in different *TESSA* subsets were connected by gray lines (vertex size corresponds to TCR count). TCRs not present in metaclusters enriched in patient cluster 3 (highlighted in b) are excluded from this panel and subsequent analysis. (d) Metaclusters from (b) and (c) were separated based on *TESSA* subset and tested as a correlate of overall survival in the WashU tissue cohort. (e) Kaplan-Meier curve of overall survival comparing patients with high frequency of *TESSA* 12 metaclusters (above median) vs. low frequency (below median) in the WashU PBMC cohort. (f) Kaplan-Meier curve of overall survival comparing patients with high frequency of *TESSA* 12 metaclusters (upper quartile) vs. patients with low frequency (lower 3 quartiles) in the UW tissue cohort as determined by TCR distances. (d) * $p < 0.05$ as calculated by log rank or Spearman correlation with Bonferroni correction for multiple comparisons. ns = not significant (e,f) * $p < 0.05$ as calculated by log-rank test. HR (95% CI) calculated by univariate Cox proportional hazards model.

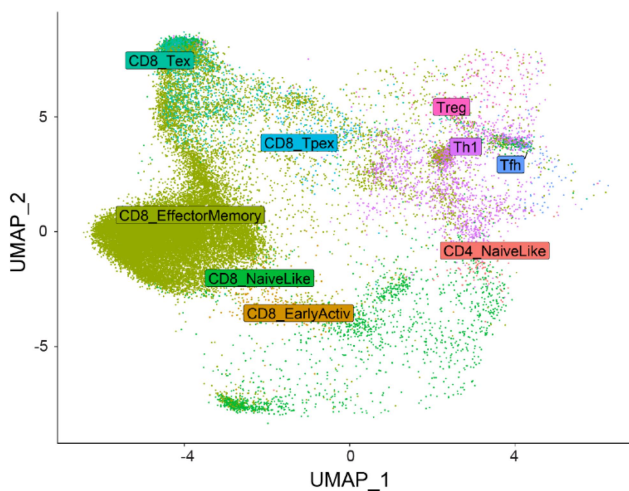
a Overlap of Metaclusters in TESSA 12 by Dataset



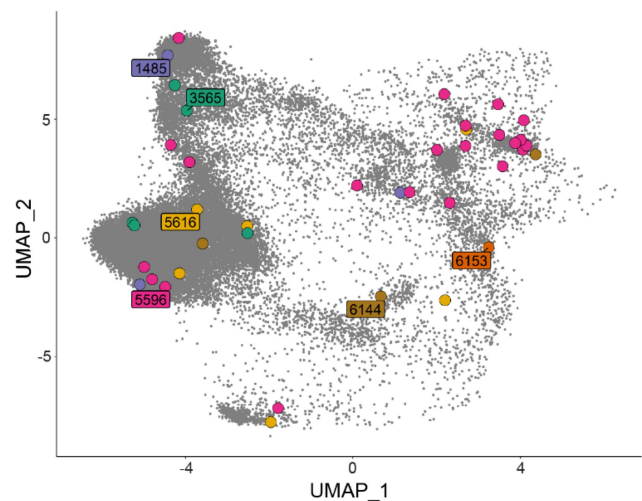
b Consensus Sequences of Shared Metaclusters



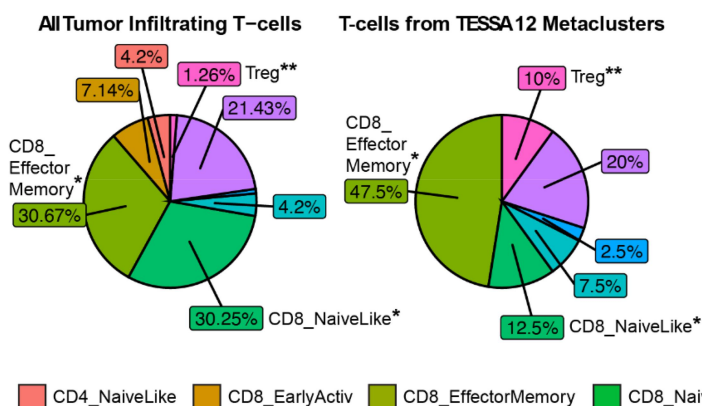
c Tumor Infiltrating T-cell Subsets



d TESSA12 Shared Metaclusters



e T-cell Subsets



f Predicted Specificity of Metaclusters in TESSA 12

MCID	5616	6144
Input TCR	ASSVGNTIY	ASSVGVDTQY
Matched TCR	ASSAGNTIY	ASSSVGVDTQY
IEDB TCRmatch Score	0.9707	0.9715
HLA	HLA-DQB1*03:02	HLA-A*02:01
VDJ	TRBV5-1 TRBJ1-3	TRBV6-1 TRBD1 TRBJ2-3
Epitope	VEALYLVAEE	VAANIVLTV
Autoantigen	insulin B (11-23)	zinc transporter 8a [SLC30A8] (186-194)
Autoantigen Expression	Endocrine Pancreas	Endocrine Pancreas

Figure 5. T-cell identity and TCR specificity in metaclusters of interest. (a) Euler plot depicting overlap of *TESSA 12* metaclusters between patient cohorts. (b) consensus amino acid sequence of metaclusters shared by WashU tissue, UW tissue and MDA. (c,d) Inference of T-cell identities in MDA dataset based on mapping of scRNA-seq transcriptomes to a reference dataset (c) with T-cells bearing TCRs from *TESSA 12* metaclusters highlighted (d). (e) Comparison of T-cell subset frequencies in all tumor-infiltrating T-cells (left, $n = 7$ patients) and T-cells from *TESSA 12* metaclusters (right). (f) predicted specificity of two of the six shared metaclusters from (a) determined with TCRmatch and the IEDB database with a significance cutoff of 0.97. (e) * $p < 0.05$, ** $p < 0.01$ as calculated by Fisher's Exact Test

signature may be of benefit in the context of malignancy. The majority of Tregs were found within metacluster #5596, which unfortunately we were unable to match with a cognate p-MHC. However, the sheer abundance of Tregs emphasizes the

autoimmune signature within *TESSA 12* metaclusters. Most T-cells within metacluster #6144 were assigned as CD8⁺ effector memory T-cells, which are capable of binding with the predicted HLA-A *02:01. These cytotoxic CD8⁺ T-cells could

directly engage and kill pancreatic islet cells expressing SLC30A8. Perhaps, these effector memory CD8⁺ T-cells are resident within the pancreas and are activated as bystanders during carcinogenesis. Here we describe a subset of intratumoral T-cells enriched in multiple PDAC datasets that are a positive prognostic factor for OS and are predicted to react against pancreas-specific autoantigens.

Discussion

Herein we describe the relationship between intratumoral and peripheral TCR repertoires with survival in patients with pancreatic cancer. We rigorously analyzed TCR sequencing data from three independent cohorts of patients. Intratumoral T-cell diversity was consistently predictive of OS when controlling for T-cell abundance and clinical characteristics. While studies have suggested this relationship without controlling for additional variables, to our knowledge this is the first report linking intratumoral TCR diversity to OS independent of T-cell abundance and in multiple cohorts.³ This adds to a body of evidence that PDAC may not be a ‘cold’ tumor and that despite the lack of success of checkpoint therapies, there may be a prominent role for T-cells in limiting cancer progression.^{3,4,42} Additionally, we find that increased abundance of T-cells within the tumor does not independently predict patient outcome, which emphasizes that not all T-cells contribute to anti-tumor immunity.⁴³ This is a differing result than has been previously reported, which could be attributable to the method of measurement (immunohistochemistry vs. estimated fraction of Total Reads).⁴⁴ However, even previous studies have only shown a weak correlation with total T-cell count, and a very strong correlation with Granzyme⁺ CD8⁺ T-cells, highlighting that the composition of the intratumoral T-cell compartment has a significant impact on patient outcome.^{3,44}

T-cell abundance in peripheral blood was a positive prognostic factor, while repertoire diversity failed to correlate with OS. This inverse finding may suggest that the composition of the peripheral blood T-cell compartment is skewed by a multitude of previous antigenic exposures and is more reflective of an individual’s antigenic history rather than a distinct anti-tumor repertoire. The observation that the abundance of T-cells within peripheral blood correlated with OS may be more indicative of an immunocompetent individual as opposed to a patient with globally depressed lymphocyte counts.

A limitation of many TCR studies that employ diversity statistics is the difficulty in linking a summary metric to distinct CDR3 identities and specificities. We utilized a novel method to bridge this divide. Patients were clustered based on shared TCR specificity, thereby aggregating patients by their ability to target similar antigens. This identified a patient cluster with both significantly greater OS, and the highest mean diversity score. Within this cluster we were able to distinguish a subset of metaclusters that predicted OS both in blood and in tissue across datasets. Further research is needed to determine whether these metaclusters could be used as a biomarker to identify checkpoint or chemotherapy-

responsive patients. It is also important to note that predicted antigen specificity based on TCR-beta sequencing is limited to well-described antigens at this time, and in the absence of careful HLA-haplotyping, is rife with false-positive p-MHC discoveries. Despite this limitation, there is still value to cognate antigen analysis as the likelihood of true positives also increases when analyzing metaclusters shared across a large population of patients. Our initial characterization of these metaclusters demonstrates that within the CD8⁺ T-cell compartment, metaclusters tend to preferentially encode effector memory cells as opposed to naïve cells, and within the CD4⁺ T-cell compartment, they tend to encode Tregs. The increased abundance of Tregs suggested that these cells might react with self-antigens, and indeed when we inferred p-MHC specificity, we found that two of the clusters were capable of engaging endocrine pancreas-specific autoantigens associated with Type I Diabetes. Although diabetes is a risk factor for PDAC, it is possible that the presence of self-reactive T-cells within the pancreas could be of benefit as they may be rapidly and nonspecifically activated in the inflammatory TME.^{45,46}

Many groups have examined the effects of checkpoint therapies on T-cell repertoire diversity.^{47–49} They may both reinvigorate exhausted T-cells and promote the emergence of new T-cell clones, thus reshaping repertoire composition.^{50,51} Our results suggest that a diverse intratumoral T-cell repertoire is associated with improved OS, particularly a repertoire that contains pancreas autoreactive T-cells. Further research is needed to determine how checkpoint therapies might affect these autoreactive populations, potentially reinvigorating a tissue resident bystander population to mediate anti-tumor immunity. It is our hope that the characterization of TCR repertoire characteristics associated with improved OS could identify patients that might benefit from T-cell directed immunotherapies that have so dramatically altered the prognosis of other lethal malignancies.

Disclosure statement

N.B. is a consultant for Santa Ana Bio and Omniscope. No potential conflicts of interest were reported by the other authors.

Funding

The work was supported by the National Institutes of Health [P30CA091842, F30CA254087], Washington University School of Medicine Dean’s Fellowship, The David Riebel Cancer Research Fund, Washington University SPORE in Pancreas Cancer [P50CA196510-05], The DeNardo Scholar Fellowship, Washington University School of Medicine Dean’s Medical Student Research Fellowship for the MPHS Yearlong Research Program. This work was also supported by Adaptive Biotechnologies.

ORCID

Vikram S. Pothuri  <http://orcid.org/0000-0003-4725-8706>
 Graham D. Hogg  <http://orcid.org/0000-0002-7143-0695>
 David G. DeNardo  <http://orcid.org/0000-0002-3655-5783>
 Ryan C. Fields  <http://orcid.org/0000-0001-6176-8943>

Author contributions

RF, DD, GH, and VP conceptualized and designed the study. RF, DD, GH, VP, and NB developed the methodology. GH, VP, LC, AJ, GW, JM and YDS acquired data. RF, DD, GH, VP, GW, and NB analyzed and interpreted data. GH, VP, LC, AJ, GW, and JM provided administrative, technical, or material support. RF, DD, GH, VP, LC, AJ, GW, YDS, JM, and VP wrote, reviewed, and/or revised the manuscript. RF and DD supervised the study. Order of co-first authorship was determined by coin-toss.

Data availability statement

All software packages in this study are publicly available. All code used to generate figures is available at this location (<https://github.com/WashU-TCR-PDAC>), and de-identified patient data can be made available upon request. TCRb CDR3 sequences will be made publicly available on Adaptive Technologies Immunoseq Analyzer Portal.

References

- O'Reilly EM, Oh D-Y, Dhani N, Renouf DJ, Lee MA, Sun W, Fisher G, Hezel A, Chang S-C, Vlahovic G. et al. Durvalumab with or without tremelimumab for patients with metastatic pancreatic ductal adenocarcinoma: a phase 2 randomized clinical trial. *JAMA Oncol.* 2019;5(10):1431–1438. doi:10.1001/jamaoncol.2019.1588.
- Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* 2014;74(11):2913–2921. doi:10.1158/0008-5472.CAN-14-0155.
- Balachandran VP, Luksza M, Zhao JN, Makarov V, Moral JA, Remark R, Herbst B, Askan G, Bhanot U, Senbabaoglu Y. et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature.* 2017;551(7681):512–516. doi:10.1038/nature24462.
- Carstens JL, Correa de Sampaio P, Yang D, Barua S, Wang H, Rao A, Allison JP, LeBleu VS, Kalluri R. Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat Commun.* 2017;8(1):15095. doi:10.1038/ncomms15095.
- Aversa I, Malanga D, Fiume G, Palmieri C. Molecular T-Cell repertoire analysis as source of prognostic and predictive biomarkers for checkpoint blockade immunotherapy. *Int J Mol Sci.* 2020;21(7):2378. doi:10.3390/ijms21072378.
- Bortone DS, Woodcock MG, Parker JS, Vincent BG. Improved T-cell receptor diversity estimates associate with survival and response to anti-PD-1 therapy. *Cancer Immunol Res.* 2021;9(1):103–112. doi:10.1158/2326-6066.CIR-20-0398.
- Valpione S, Mundra PA, Galvani E, Campana LG, Lorigan P, De Rosa F, Gupta A, Weightman J, Mills S, Dhomen N. et al. The T cell receptor repertoire of tumor infiltrating T cells is predictive and prognostic for cancer survival. *Nat Commun.* 2021;12(1):4098. doi:10.1038/s41467-021-24343-x.
- Hogan SA, Courtier A, Cheng PF, Jaberg-Bentele NF, Goldinger SM, Manuel M, Perez S, Plantier N, Mouret J-F, Nguyen-Kim TDL. et al. Peripheral blood TCR repertoire profiling may facilitate patient stratification for immunotherapy against melanoma. *Cancer Immunol Res.* 2019;7(1):77–85. doi:10.1158/2326-6066.CIR-18-0136.
- Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM. Analyzing the mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol.* 2020;38(10):1194–1202. doi:10.1038/s41587-020-0505-4.
- Mayer-Blackwell K, Schattgen S, Cohen-Lavi L, Crawford, JC, Souquette A, Gaevart, JA, Hertz T, Thomas, PG, Bradley P, Fiore-Gartland A. TCR meta-clonotypes for biomarker discovery with tcrdist3: identification of public, HLA-restricted SARS-CoV-2 associated TCR features. *bioRxiv.* 2021;10: e68605. doi:10.7554/eLife.68605.
- Zhang Z, Xiong D, Wang X, Liu H, Wang T. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat Methods.* 2021;18(1):92–99. doi:10.1038/s41592-020-01020-3.
- Sidhom J-W, Larman HB, Pardoll DM, Baras AS. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat Commun.* 2021;12(1):1605. doi:10.1038/s41467-021-21879-w.
- Chronister WD, Crinklaw A, Mahajan S, Vita R, Koşaloğlu-Yalçın Z, Yan Z, Greenbaum JA, Jessen LE, Nielsen M, Christley S. et al. Tcrmatch: predicting T-Cell receptor specificity based on sequence similarity to previously characterized receptors. *Front Immunol.* 2021;12. doi:10.3389/fimmu.2021.640725.
- Valkiers S, Van Houcke M, Laukens K, Meysman P, Boeva DV. ClusTCR: a python interface for rapid clustering of large sets of CDR3 sequences with unknown antigen specificity. *Bioinformatics.* 2021;37(24):4865–4867. doi:10.1093/bioinformatics/btab446.
- Montemurro A, Schuster V, Povlsen HR, Bentzen AK, Jurtz V, Chronister WD, Crinklaw A, Hadrup SR, Winther O, Peters B. et al. NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCR α and β sequence data. *Commun Biol.* 2021;4(1):1060. doi:10.1038/s42003-021-02610-3.
- Holt RA. Interpreting the T-cell receptor repertoire. *Nat Biotechnol.* 2017;35(9):829–830. doi:10.1038/nbt.3957.
- Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee J-Y, Olshen RA, Weyand CM, Boyd SD, Goronzy JJ. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci USA.* 2014;111(36):13139–13144. doi:10.1073/pnas.1409155111.
- Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, Crawford JC, Clemens EB, Nguyen THO, Kedzierska K. et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature.* 2017;547(7661):89–93. doi:10.1038/nature22383.
- Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, Ji X, Han A, Krams SM, Pettus C. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature.* 2017;547(7661):94–98. doi:10.1038/nature22976.
- Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med.* 2017;377(25):2500–2501. doi:10.1056/NEJMc1713444.
- Seo YD, Jiang X, Sullivan KM, Jalikis FG, Smythe KS, Abbasi A, Vignali M, Park JO, Daniel SK, Pollack SM. et al. Mobilization of CD8+ T cells via CXCR4 blockade facilitates PD-1 checkpoint therapy in human pancreatic cancer. *Clin Cancer Res.* 2019;25(13):3934–3945. doi:10.1158/1078-0432.CCR-19-0081.
- Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA. et al. The immune landscape of cancer. *Immunity.* 2018;48(4):812–830. e814. doi:10.1016/j.immuni.2018.03.023.
- Schalck A, Sakellariou-Thompson D, Forget M-A, Sei E, Hughes TG, Reuben A, Bai S, Hu M, Kumar T, Hurd MW. et al. Single-cell sequencing reveals trajectory of tumor-infiltrating lymphocyte states in pancreatic cancer. *Cancer Discov.* 2022;12(10):2330–2349. doi:10.1158/2159-8290.CD-21-1248.
- Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI. et al. Vdjtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput Biol.* 2015;11(11): e1004503. doi:10.1371/journal.pcbi.1004503.
- Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, Marini F, Rue-Albrecht K, Risso D, Sonesson C. et al. Orchestrating single-cell analysis with bioconductor. *Nat Methods.* 2020;17(2):137–145. doi:10.1038/s41592-019-0654-x.
- McCarthy DJ, Campbell KR, Lun ATL, Wills QF, Hofacker I. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics.* 2017;33(8):1179–1186. doi:10.1093/bioinformatics/btw777.

27. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* 2019;37(1):38–44. doi:10.1038/nbt.4314.
28. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019;20(1):296. doi:10.1186/s13059-019-1874-1.
29. Andreatta M, Corria-Osorio J, Müller S, Cubas R, Coukos G, Carmona SJ. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat Commun.* 2021;12(1):2965. doi:10.1038/s41467-021-23324-4.
30. Stromnes IM, Hulbert A, Pierce RH, Greenberg PD, Hingorani SR. T-cell localization, activation, and clonal expansion in human pancreatic ductal adenocarcinoma. *Cancer Immunol Res.* 2017;5(11):978–991. doi:10.1158/2326-6066.CIR-16-0322.
31. Poschke I, Faryna M, Bergmann F, Flossdorf M, Lauenstein C, Hermes J, Hinz U, Hank T, Ehrenberg R, Volkmar M. et al. Identification of a tumor-reactive T-cell repertoire in the immune infiltrate of patients with resectable pancreatic ductal adenocarcinoma. *Oncoimmunology.* 2016;5(12):e1240859–e1240859. doi:10.1080/2162402X.2016.1240859.
32. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, Jamal-Hanjani M, Wilson GA, Birkbak NJ, Hiley CT. et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science.* 2016;351(6280):1463–1469. doi:10.1126/science.aaf1490.
33. Campana LG, Mansoor W, Hill J, Macutkiewicz C, Curran F, Donnelly D, Hornung B, Charleston P, Bristow R, Lord GM. et al. T-Cell infiltration and clonality may identify distinct survival groups in colorectal cancer: development and validation of a prognostic model based on the cancer genome atlas (TCGA) and Clinical Proteomic Tumor Analysis Consortium (CPTAC). *Cancers Basel.* 2022;14(23):5883. doi:10.3390/cancers14235883.
34. Soto C, Bombardi RG, Kozhevnikov M, Sinkovits RS, Chen EC, Branchizio A, Kose N, Day SB, Pilkinton M, Gujral M. et al. High frequency of shared clonotypes in human T cell receptor repertoires. *Cell Rep.* 2020;32(2):107882. doi:10.1016/j.celrep.2020.107882.
35. Martin BD, Witten D, Willis AD. Modeling microbial abundances and dysbiosis with beta-binomial regression. *Ann Appl Stat.* 2020;14(1):94–115. doi:10.1214/19-AOAS1283.
36. Shang B, Liu Y, Jiang SJ, Liu Y. Prognostic value of tumor-infiltrating FoxP3+ regulatory T cells in cancers: a systematic review and meta-analysis. *Sci Rep.* 2015;5(1):15179. doi:10.1038/srep15179.
37. Kiryu S, Ito Z, Suka M, Bito T, Kan S, Uchiyama K, Saruta M, Hata T, Takano Y, Fujioka S. et al. Prognostic value of immune factors in the tumor microenvironment of patients with pancreatic ductal adenocarcinoma. *BMC Cancer.* 2021;21(1):1197. doi:10.1186/s12885-021-08911-4.
38. Brouwer T, Ijsselstein J, Oosting J, Ruano D, van der Ploeg M, Dijk F, Bonsing B, Fariña A, Morreau H, Vahrmeijer A. et al. A paradoxical role for regulatory T cells in the tumor microenvironment of pancreatic cancer. *Cancers Basel.* 2022;14(16):3862. doi:10.3390/cancers14163862.
39. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 2018;47(D1):D339–D343. doi:10.1093/nar/gky1006.
40. Wang Y, Sosinowski T, Novikov A, Crawford F, Neau DB, Yang J, Kwok WW, Marrack P, Kappler JW, Dai S. et al. C-terminal modification of the insulin B: 11–23 peptide creates superagonists in mouse and human type 1 diabetes. *Proc Natl Acad Sci U S A.* 2018;115(1):162–167. doi:10.1073/pnas.1716527115.
41. Culina S, Lalanne AI, Afonso G, Cerosaletti K, Pinto S, Sebastiani G, Kuranda K, Nigi L, Eugster A, Østerbye T. et al. Islet-reactive CD8 + T cell frequencies in the pancreas, but not in blood, distinguish type 1 diabetic patients from healthy donors. *Sci Immunol.* 2018;3(20). doi:10.1126/sciimmunol.aao4013.
42. Shibuya KC, Goel VK, Xiong W, Sham JG, Pollack SM, Leahy AM, Whiting SH, Yeh MM, Yee C, Riddell SR. et al. Pancreatic ductal adenocarcinoma contains an effector and regulatory immune cell infiltrate that is altered by multimodal neoadjuvant treatment. *PLoS One.* 2014;9(5):e96565. doi:10.1371/journal.pone.0096565.
43. Jang J-E, Hajdu CH, Liot C, Miller G, Dustin ML, Bar-Sagi D. Crosstalk between regulatory T cells and tumor-associated dendritic cells negates anti-tumor immunity in pancreatic cancer. *Cell Rep.* 2017;20(3):558–571. doi:10.1016/j.celrep.2017.06.062.
44. Muller M, Haghnejad V, Schaefer M, Gauchotte G, Caron B, Peyrin-Biroulet L, Bronowicki J-P, Neuzillet C, Lopez A. The immune landscape of human pancreatic ductal carcinoma: key players, clinical implications, and challenges. *Cancers.* 2022;14(4):995. doi:10.3390/cancers14040995.
45. Nelson CE, Thompson EA, Quarnstrom CF, Fraser KA, Seelig DM, Bhela S, Burbach BJ, Masopust D, Vezys V. Robust iterative stimulation with self-antigens overcomes CD8(+) T cell tolerance to self- and tumor antigens. *Cell Rep.* 2019;28(12):3092–3104.e5. doi:10.1016/j.celrep.2019.08.038.
46. Leem G, Jeon M, Kim KW, Jeong S, Choi SJ, Lee YJ, Kim E-S, Lee J-I, Ha SY, Park S-H. et al. Tumour-infiltrating bystander CD8 + T cells activated by IL-15 contribute to tumour control in non-small cell lung cancer. *Thorax.* 2022;77(8):769–780. doi:10.1136/thoraxjnl-2021-217001.
47. Hopkins AC, Yarchoan M, Durham JN, Yusko EC, Rytlewski JA, Robins HS, Laheru DA, Le DT, Lutz ER, Jaffee EM. et al. T cell receptor repertoire features associated with survival in immunotherapy-treated pancreatic ductal adenocarcinoma. *JCI Insight.* 2018;3(13). doi:10.1172/jci.insight.122092.
48. Hosoi A, Takeda K, Nagaoka K, Iino T, Matsushita H, Ueha S, Aoki S, Matsushima K, Kubo M, Morikawa T. et al. Increased diversity with reduced “diversity evenness” of tumor infiltrating T-cells for the successful cancer immunotherapy. *Sci Rep.* 2018;8(1):1058. doi:10.1038/s41598-018-19548-y.
49. Robert L, Tsoi J, Wang X, Emerson R, Homet B, Chodon T, Mok S, Huang RR, Cochran AJ, Comin-Anduix B. et al. CTLA4 blockade broadens the peripheral T-Cell receptor repertoire. *Clin Cancer Res.* 2014;20(9):2424–2432. doi:10.1158/1078-0432.CCR-13-2648.
50. Wu TD, Madireddi S, de Almeida PE, Banchereau R, Chen YJJ, Chitre AS, Chiang EY, Iftikhar H, O’Gorman WE, Au-Yeung A. et al. Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature.* 2020;579(7798):274–278. doi:10.1038/s41586-020-2056-8.
51. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, McNamara KL, Granja JM, Sarin KY, Brown RA. et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med.* 2019;25(8):1251–1259. doi:10.1038/s41591-019-0522-3.