The Ratio of Exhausted to Resident Infiltrating Lymphocytes Is Prognostic for Colorectal Cancer Patient Outcome

Momeneh Foroutan1, Ramyar Molania2,3, Aline Pfefferle1,4, Corina Behenbruch5, Sebastian Scheer1, Axel Kallies6, Terence P. Speed2,7, Joseph Cursons1,4, and Nicholas D. Huntington1,4

ABSTRACT

Immunotherapy success in colorectal cancer is mainly limited to patients whose tumors exhibit high microsatellite instability (MSI). However, there is variability in treatment outcomes within this group, which is in part driven by the frequency and characteristics of tumor-infiltrating immune cells. Indeed, the presence of specific infiltrating immune-cell subsets has been shown to correlate with immunotherapy response and is in many cases prognostic of treatment outcome. Tumor-infiltrating lymphocytes (TIL) can undergo distinct differentiation programs, acquiring features of tissue-residency or exhaustion, a process during which T cells upregulate inhibitory receptors, such as PD-1, and lose functionality. Although residency and exhaustion programs of CD8+ T cells are relatively well studied, these programs have only recently been appreciated in CD4+ T cells and remain largely unknown in tumor-infiltrating natural killer (NK) cells. In this study, we used single-cell RNA sequencing (RNA-seq) data to identify signatures of residency and exhaustion in colorectal cancer–infiltrating lymphocytes, including CD8+, CD4+, and NK cells. We then tested these signatures in independent single-cell data from tumor and normal tissue–infiltrating immune cells. Furthermore, we used versions of these signatures designed for bulk RNA-seq data to explore tumor-intrinsic mutations associated with residency and exhaustion from TCGA data. Finally, using two independent transcriptomic datasets from patients with colon adenocarcinoma, we showed that combinations of these signatures, in particular combinations of NK-cell activity signatures, together with tumor-associated signatures, such as TGFβ signaling, were associated with distinct survival outcomes in patients with colon adenocarcinoma.

Introduction

Immune checkpoint blockade (ICB) in colorectal cancer has shown clinical benefit in a subset of patients with colorectal cancer whose tumors have high microsatellite instability (MSI) or are mismatch repair deficient (dMMR), and in 10% of patients whose tumors are microsatellite stable (MSS) or MMR proficient (pMMR; ref. 1). However, due to the complex interplay between tumor cells and immune cells in the context of cancer immunotherapy, MSI status alone is often not enough to precisely predict response to ICB (2). The main factors impacting the outcome of immunotherapy, and, in general, survival outcome in patients, are the frequency and the characteristics of the immune cells in the tumor microenvironment (TME; refs. 3–5).

Single-cell analysis studies in several cancer types indicate a high level of heterogeneity in tumor-infiltrating lymphocytes (TIL; refs. 6–8). One subset of CD8+ T cells is characterized by “exhaustion,” a process that occurs in activated T cells in response to persistent antigen exposure in the TME, resulting in a loss of functionality and expression of coinhibitory receptors such as PD-1 and CTLA-4 (9). Heterogeneity within the exhausted T-cell population ranges from progenitor or precursor cells, which exhibit self-renewal capacity, to terminally differentiated cells (10–14), and this heterogeneity affects the success of immunotherapy (15–17).

T-cell exhaustion has largely been studied in the context of melanoma and non–small cell lung cancer (NSCLC; refs. 7, 18, 19), although it is beginning to be assessed in colorectal cancer (8, 20). However, studies looking at T-cell differentiation states and their tumor reactivity have yielded conflicting outcomes. In one study, it was reported that exhausted T cells have high reactivity toward tumor cells (measured by IFNγ and TNFα production ex vivo) and show high proliferation at earlier stages of the exhaustion program (19). Huang and colleagues presented evidence that a subset of circulating CD8+ T cells expressing markers of exhausted cells (e.g., PDCD1, CTLA4, and HAVCR2) are the most proliferative T cells following ICB therapy, and that a high ratio of exhausted CD8+ T-cell reinvigoration to tumor burden is prognostic in patients with head and neck cancer (21). Others have reported that a low ratio of exhausted T cells to early activated cells is associated with better response to ICBS (16). In general, the presence of tissue-resident memory T cells (Trem) is associated with beneficial survival outcome and better response to immunotherapy in several cancer types (22–25). Although several studies have explored residency programs (or molecular signatures associated with residency) in T cells, these processes remain poorly understood in NK cells. Tissue-resident NK cells in steady-state are...
cleared from conventional NK cells and are referred to as type 1 innate lymphoid cells (ILC1; ref. 26). We have previously observed ILC1s in murine tumor models and suggested these cells are derived from infiltrating NK cells that differentiated into ILC1s under conditions of high TGFβ signaling in the TME, permitting tumor cell escape from immunosurveillance (27).

TGFβ signaling not only plays critical roles in differentiation, development, and maintenance of various resident immune cells (28–31), but also is one of the key drivers of epithelial-to-mesenchymal transition (EMT) in cancer cells (32), which could impact the interplay between tumor cells and infiltrating immune cells. Studies have shown that TGFβ signaling is associated with T-cell exclusion and tumor escape from immunosurveillance (33) and that there is a better IC8 response after blockade of TGFβ signaling, driven by increasing T-cell infiltration of the tumors (34). Because of the crucial roles of TGFβ signaling in differentiation of resident cells, T-cell exclusion in tumors, and its associations with survival outcome and response to IC8s, it is of utmost importance to better understand the relationship and interplay between TGFβ and immune or tumor cells.

In this study, we dissected transcriptional programs underpinning lymphocyte residency and exhaustion of tumor-infiltrating T and NK cells using integrated analyses of single-cell data from colorectal cancer and normal gut, as well as bulk transcriptomic datasets. We identified numerous tumor-intrinsic mutations associated with these programs and used our residency and exhaustion signatures in combination with other tumor-associated programs (e.g., TGFβ signaling) to identify patients who have better survival outcome or may benefit from IC8 immunotherapies. We identified improved survival outcome for patients with high exhaustion and low residency programs in NK cells, and patients with high exhaustion programs in both CD8+ and NK cells and low TGFβ signaling. This work has implications for cancer immunotherapy as it suggests that strategies to prevent tumor residency may improve NK-cell and CD8+ T-cell antitumor immunity and patient outcomes.

Materials and Methods

All analyses for this study were performed using R version ≥ 4.0.2, with Bioconductor version ≥ 3.11 (35). A list of all the packages and tools used in this study with their versions and citations are given in Supplementary Table S1. General data wrangling and visualizations were performed using the following R packages: tidyverse (36), ComplexHeatmap (37), ggplot2 (38), RColorBrewer (https://cran.r-project.org/web/packages/RColorBrewer), and gridExtra (https://cran.r-project.org/web/packages/gridExtra). The ggseignR package (https://cran.r-project.org/web/packages/ggseignR) was also used to add P values to the boxplots.

Deriving the initial exhaustion and residency signatures

We used high-quality single-cell data (Smart-seq2) from Zhang and colleagues (20) to obtain gene expression signatures. Transcript per million (TPM)–like data were obtained from GSE146771 on June 9, 2020, with data extracted for CD8+, CD4+, and NK cells using predefined annotations from the metadata. Each cell type was analyzed separately.

In each case, we started with a selection of canonical markers associated with residency (Res) and exhaustion (Exh) in each of the cell types. The Res canonical markers consisted of CD69, ITGAE (CD103), ITGAI, and RGS1 for CD8+, CD4+, and NK cells (referred to as CD8_Res and CD4_Res, respectively), and we also additionally included ZNF683 (Hobit) for NK cells (referred to as NK_Res). The five Exh markers we used were: HAVCR2 (Tim-3), PDCD1 (PD-1), CTLA4, LAYN, CXCL13 for CD8+ cells (referred to as CD8_Exh); the same genes plus LAG3 for CD4+ cells (referred to as CD4_Exh), and; HAVCR2, PDCD1, CTLA4, TIGIT, and LAG3 for NK cells (referred to as NK_Exh).

Next, we curated an extensive list of genes reported in the literature to be associated with Res or Exh across different species (human and mouse), as well as different conditions (infectious disease, cancer and normal tissue; refs. 4, 7, 18, 23, 27, 29, 39–48; Supplementary Table S2). For each cell type, we then calculated the Spearman correlation (ρ) between candidate Exh and Res markers and the list of canonical markers for that cell type. For CD8+ and NK cells, genes with ρ > 0.35 with any of the canonical Res signatures were considered Res genes, and those with ρ > 0.25 with canonical Exh genes were counted as Exh markers. These thresholds were slightly different for CD4+ cells due to differences in gene expression distributions (ρ > 0.3 for CD4_Res and CD4_Exh). After removing genes that were overlapping between the Res and Exh gene list, we performed a sparse canonical correlation analysis (CCA; using PMA R package; https://cran.r-project.org/web/packages/PMA) and removed genes with a coefficient < −0.1 in the first component in order to reduce the correlation between the Res and Exh sets. We refer to the resultant gene sets as the “initial” Res and Exh genes; these were different for each cell subset, although some genes were overlapping across subsets.

Deriving the final exhaustion and residency signatures

Using the initial Res and Exh signatures and the singcore R package (49), we scored the Zhang Smart-seq2 NK, CD4+ and CD8+ cells (20). We then stratified cells based on their Res and Exh scores and performed differential expression (DE) analyses between cells with high Res/low Exh scores and those with high Exh/low Res scores. We used the Seurat R package (50) to perform DE analyses with two different approaches: (i) Wilcoxon rank sum test, and; (ii) the hurdle model from the MAST package (51). We retained genes that were considered as differentially expressed using either approach (P_adj < 0.05 and logFC > 0.3), and further refined this list by comparing transcript abundance percentiles for Res versus Exh versus circulating blood cells. Specifically, we retained Exh genes whose 65th percentile abundance in Exh cells was higher than the 90th percentile expression in both Res and blood cells. Similar comparisons were performed to refine Res genes, and a 75th percentile threshold was used when defining the NK_Exh signature.

Analysis of single-cell datasets

For each of the single-cell datasets used in this study, we considered each cell subset independently, and therefore, the data filtration, clustering, and dimensionality reductions were also performed using slightly different parameters in each subset of a given dataset. We refer readers to the code available on Github for details on parameter settings (see “Code availability”). For all raw single-cell data, we performed the SCTransform method (52) from the Seurat package (50) using the top 5,000 variable genes, regressing on the percent of mitochondrial genes. We also used the Seurat package (50) to perform clustering, dimensionality reduction, and DE analysis. The slingshot R package (53) was used to perform the trajectory analysis on the SingleCellExperiment object generated by SingleCellExperiment R package (54). For all single-cell datasets, we used the singscore package (49) to score single cells. Blend plots were generated using the Seurat visualization functions with single-cell scores obtained from singscore.
Processed Zhang and colleagues’ single-cell RNA sequencing (RNA-seq) and metadata for both Smart-seq2 ($N_{patients} = 12$) and 10X ($N_{patients} = 18$) data were obtained from GSE146771 on the June 9, 2020 (20). No further normalization or processing was performed. The de Vries and colleagues’ single-cell data RNA-seq data as well as cell annotation information were received from the authors on July 23, 2020 (55); we included two cell groups annotated as proliferative cells and unclassified cells when analyzing the data for each of the NK, CD4+, and CD8+ cells separately. The raw single-cell RNA-seq data from the human gut atlas by James and colleagues were downloaded on June 9, 2020 from the Gut Cell Atlas (https://www.gutcellatlas.org/; ref. 56). We performed batch correction to remove donor differences using the integration approach provided in the Seurat package (50).

Raw single-cell RNA-seq data from de Andrade and colleagues were downloaded for 1 patient with melanoma (CY129; with a relatively large number of infiltrating NK cells) on June 15, 2020 from GSE139249 (57). We analyzed these data in two steps: first, we ran the Seurat pipeline (50) and normalized the data to identify potential contaminating cells based on the expression of marker genes for cells other than NK cells. To be consistent with the original manuscript (57), we normalized to remove donor differences using the integration approach (https://www.gutcellatlas.org/; ref. 56). We performed batch correction to remove donor differences using the integration approach provided in the Seurat package (50).

We then filtered the pan-cancer data to only retain genes with cpm $> 20\%$ mitochondrial genes, number of RNA counts $\geq 25,000$, or number of features $\geq 500$, which resulted in 4,195 cells. Next, we used the Seurat package (50) to analyze these data using similar settings provided above; we performed SCTransform (52) for normalization, used the first 20 principal components (PC) for identifying the k-nearest neighbors of each cell, and performed dimensional reduction. For running Uniform Manifold Approximation and Projection (UMAP), we also specified n.neighbors = 50, min.dist = 0.4.

Refinement of the signatures based on bulk cell line and laser capture microdissection data

The RNA-seq count data and meta-data for all cell lines from the Cancer Cell Line Encyclopedia (CCLE) were downloaded from the Broad Institute Data Portal (https://portals.broadinstitute.org/ccle/; ref. 58) on the Feb 7, 2020. We downloaded gene annotation information (e.g., gene lengths) from GenCode (v19) and used the edgeR Bioconductor package (59) to generate a DGEList object in R. We then filtered the pan-cancer data to only retain genes with cpm $> 1$ in at least five cell lines, performed Trimmed Mean of M-values (TMM) normalization, and calculated the Reads Per Kilobase Million (RPKM) values. The data were then subset to only visualize the expression of the marker genes in the colorectal cancer cell lines. Comparing against these data, we considered Exh and Res markers to pass the “bulk tumor threshold” if they had median logRPKM expression $\leq 0$.

We also downloaded processed microarray data (log2, RMA normalized) from high-purity colorectal cancer samples collected with laser capture microdissection (LCM) by Tsukamoto and colleagues (ref. 60; GSE21510) using the GEOquery R package (61) in September 2020. Probes that mapped to more than one Entrez ID were excluded, and for probes mapping to the same gene, we considered the average expression value. In these data, we considered Exh and Res markers to pass the “bulk tumor threshold” if they had median logRPKM expression $\leq$ median expression of all genes in the data. Finally, if a given gene passed the “bulk tumor threshold” in CCLE or LCM, we considered that marker to be suitable for use in bulk tumor samples. The scores from these markers have “Bulk” prefix in their names. We note that abundance data for MARCH3 did not exist within the CCLE data, and 14 genes (KRT6, APOBEC3H, TTC2A, MIR4435–2HG, HRNRPA1L2, HLA-DRB3, HLA-DRB1, KLRC2, ADGRE5, HSPA1A, HSPA1B, XCL2, HSPA8, NOTO) were not present within the LCM microarray data, and, therefore, could not be validated against these datasets.

Refinement of the signatures based on their expression in stroma cells

We further refined the Exh_Bulk and Res_Bulk signatures to remove genes that showed high expression in stroma. To assess the expression of these genes in the stromal cells, we used the single-cell data (Smart-seq2) for cancer-associated fibroblasts (CAF) and myofibroblasts from Zhang and colleagues (20) and considered them as stromal cells (referred to as Fibroblasts in the figures). We also scored each of the NK, CD8+, and CD4+ cells in the same data set using their corresponding Res and Exh signatures (i.e., all NK cells were scored against NK_Res and NK_Exh signatures) to define subsets of resident and exhausted cells in each of the cell types. We then compared the median expression of our signature genes in fibroblasts versus each corresponding cell type; for example, we compared the median expression of NK_Res_Bulk signature genes in NK_Res cells with those in fibroblasts. Doing this for all the signature genes, we identified 6 genes that showed higher median expression in stroma cells compared with each of the cell types, and therefore they were removed from the signatures (CSF1 from CD8_Exh_Bulk, PPP1R15A, NROA3 and ITGAI from CD4_Res_Bulk; FNDC3B and IL1R1 from CD4_Exh_Bulk).

Analysis of the transcriptomics data from bulk tumor samples

Raw counts for TCGA colon adenocarcinoma (COAD; $N_{samples} = 479$) harmonized RNA-seq data (62) were downloaded using the TCGAbiolinks R package (63). We only considered protein-coding genes and filtered to remove genes expressed at low level (genes with count $< 15$ across 90% of the samples, applied in cancer and normal samples separately) and data were normalized using RUV-III (ref. 64; https://cran.r-project.org/web/packages/ruv).

Processed data from Marisa and colleagues ($N_{samples} = 585$; ref. 65) were downloaded from GSE39582 on March 18, 2019 using the GEOquery R package (61), and a SummarizedExperiment object was generated using the SummarizedExperiment Bioconductor package (66). Probe IDs were mapped to gene symbols using the hgu133-plus2.db R package (DOI: 10.18129/B9.bioc.hgu133plus2cdf). Consensus molecular subtyping in both datasets was performed using the CMScaler R package (67).

Single-cell and single-sample scoring and molecular signatures

We used the singscore Bioconductor package (49) to score both single cells and single samples where relevant in this study. We only used signatures with known or expected upregulated genes, and did not center scores, such that scores range between 0 and 1, where 0.25 corresponds to a mean signature gene rank in the 25th percentile by abundance, and 0.75 corresponds to a mean signature gene rank in the 75th percentile by abundance. A list of signatures that were used to score tumor data and perform survival analysis as well as signatures used to score single-cell data is given in Supplementary Table S3. These included several signatures from MSigDB (68) obtained through the msigdf R package (https://github.com/stephenturner/msigdf), and several tumor- and immune-related signatures (31, 32, 69–73).
Transcription factors and G-protein-coupled receptors

A list of 1,549 transcription factors (TF) was generated by searching for "transcription factor" in gene description or protein class fields from the Human Protein Atlas (HPA). We generated a list of genes that were G-protein-coupled receptors (GPCR) or were associated with GPCR signaling pathways from several sources, including the HPA (searching for "coupled" in gene description and protein class fields), GO terms (obtained from QuickGO), and uniprot (searching for "gpcr" under the Type field), resulting in 2,456 genes (Supplementary Table S4).

Identification of genomic mutations associated with residency and exhaustion

The TCGA mutation data [maf format mapped to hg38, from the Genomic Data Commons (GDC; https://gdc.cancer.gov/)] were downloaded using the TCGAbiolinks package (63) on October 15, 2018. We filtered the mutation data for those mutations with low or modifier impacts, and retained mutations with moderate or high impacts. Several mutations in the same gene for the same sample were annotated as multi-hit (MHT), and duplicated records were removed to retain one record of mutation in a given gene within individual samples. Using these data, we defined the mutation load as the total number of mutated genes in a given sample.

We then subset the mutation data to keep genes mutated within at least 10 patient samples. For each of the Exh and Res scores (from the Bulk signatures) for different cell types, we used the tidyverse (36) and broom (74) R packages to generate several linear models (obtained from the literature, which were functional programs (see “Materials and Methods”); Supplementary Fig. S1 summarizes the workflow for this study). We then used these initial signatures (Supplementary Table S5) with singcore (49) to quantify the relative concordance of single cells against these signatures, where a high score represents higher relative expression of genes within a given signature (Fig. 1A). Although residency scores for each cell type clearly separated immune cells in peripheral blood from those infiltrating tissue (either tumor or adjacent normal samples; across the y-axis in Fig. 1A and Supplementary Fig. S2), the exhaustion scores further separated cells with differences in their relative expression of exhaustion genes (across the y-axis in Fig. 1A and Supplementary Fig. S2). We then obtained genes that were differentially expressed between exhausted cells (with high exhaustion and low residency scores) and resident cells (with high residency and low exhaustion scores; thresholds shown in Fig. 1A),
and further filtered the resulting genes by group comparisons with percentile thresholds (see "Materials and Methods").

The resultant gene sets were residency (Res) and exhaustion (Exh) signatures for the CD8$^+$, CD4$^+$, and NK cells that can be used for infiltrating immune cells (called CD8_Res ($N_{\text{genes}} =$ 43), CD8_Exh ($N_{\text{genes}} =$ 27), CD4_Res ($N_{\text{genes}} =$ 31), CD4_Exh ($N_{\text{genes}} =$ 83), NK_Res ($N_{\text{genes}} =$ 61), and NK_Exh ($N_{\text{genes}} =$ 19). Supplementary Table S6). We further merged all the Res signatures together to define a general Res signature (All_Res; $N_{\text{genes}} =$ 100) and all the Exh signatures together to define a general Exh signature (All_Exh; $N_{\text{genes}} =$ 118). The single-cell scores from these signatures could identify a subset of cells with higher evidence of Exh or Res programs in the Zhang Smart-seq2 data (Supplementary Fig. S3), with exhausted NK cells largely restricted to tumors with MSI status. In general, our results showed that Exh programs were similar between CD8$^+$ T cells and NK cells; however, Res signatures were more similar between CD8$^+$ and CD4$^+$ cells. Zhang and colleagues reported more similarity between Th1-like clusters of CD4$^+$ cells with exhausted CD8$^+$ cells (8), but in our analysis, we observed higher exhaustion scores in a large number of tumor-specific CD4$^+$ Tregs, which is more similar to the CD4$^+$ dysfunctional Treg signature in melanoma samples reported by Li and colleagues (19).

HAVCR2 (encoding TIM3) was the only gene shared across the Exh signatures for all three cell types, but we identified ten overlapping genes across all the Res signatures (Supplementary Table S6). A small number of genes also overlapped between each pair of signatures.
Some of the TF-encoding genes that we identified in the All_Res signature included genes associated with activating protein 1 (AP-1) TFs, which are downstream of T-cell receptor (TCR) activation, NF-kB signaling, and early growth response downstream of AP-1 and NF-kB signaling (Supplementary Table S6). Although it was identified as one of the CD8+ Res markers, GPR15 is one of the less well-studied GPCRs and it also showed restricted expression in CD4+ and NK Res cells (Supplementary Fig. S4). From the full list of genes in the All_Exh (N_genes = 118) signature, we identified three genes associated with GPCR signaling (e.g., CCR8) and ten TFs that were not associated with GPCR signaling (Supplementary Table S6).

Consistent with the idea that Exh signatures show a strong degree of overlap with activation signatures, a small subset of the Exh cells from all the three cell types exhibited a high level of proliferation. Using a proliferation signature and the Exh signatures obtained here, we further identified signatures for proliferative Exh cells (see Supplementary Fig. S5; Supplementary Table S7 for the list of DEGs).

After clustering and dimensionality reduction on our cell type–specific marker genes, we performed unsupervised trajectory analyses, which not only separated blood from tissue-infiltrating immune cells (Supplementary Fig. S2), but also showed resident cells as an intermediate step between naive cells (mostly in blood circulation) and exhausted cells (Fig. 1B; Supplementary Fig. S2). Although the Exh subset of CD4+ and NK cells are a downstream lineage of Res cells in tissues, the Exh subset in CD8+ T cells seem to derive from both tissue-infiltrating Res cells as well as circulating CD8+ T cells (Supplementary Fig. S2).

To further examine the development of NK cells in tumor tissue, we assessed the expression of several genes defining NK-cell maturation and tissue residency (46) in the NK cells extracted from tissue and blood (Supplementary Fig. S6). We observed a population of tissue-infiltrating NK cells (annotated as the GZMK population by Zhang and colleagues; ref. 20), with expression of naive markers (CCR7, SELL, and GPR138) as well as inhibitory markers (at least one of PDCD1, TIGIT, HAVCR2, or CTLA-4; Supplementary Fig. 57). Overlaid expression of these genes on the trajectory plots identified these cells as belonging to an intermediate stage of the trajectory (Supplementary Fig. S7).

To examine the association between these programs and TGFβ signaling, we also scored single cells against a TGFβ signature for Tim3 cells reported by Nath and colleagues (31). Our results showed higher correlation between TGFβ signature scores and Res scores in CD8+ and NK cells compared to Exh scores, whereas we observed higher correlation between the CD4_Exh and TGFβ program (Supplementary Fig. S8). Most of the CD4_ Exh cells that have high TGFβ scores are tumor-associated Tregs, and TGFβ is important for Treg infiltration, retention, and suppressive function (79).

Transcriptional signatures identify residency and exhaustion in single-cell data

Using the signature genes obtained above, we generated UMAP plots in three independent single-cell data sets: 10X colorectal cancer data from Zhang and colleagues (20), colorectal cancer data from de Vries and colleagues (55), and normal human gut data from James and colleagues (ref. 56; see “Materials and Methods”. Fig. 2A). Similar to the Zhang and colleagues’ Smart-seq UMAPs, the first dimension of the UMAP plots in the Zhang 10X data generated using our marker genes separated the immune cells in peripheral blood from those in the tissue. We showed that whereas a subset of cells only exhibited evidence for Exh or Res, some cells displayed a “hybrid” program. In normal human gut, we observed relatively more Res cells, which confirmed previous observations that showed more Res cells in normal samples (8). We also observed a small subset of exhausted cells in normal gut tissue, perhaps due to the persistent presence of microbiota-derived antigens in the gut.

There are differences in cellular context and immune-cell infiltration across different tissue types. However, due to the low frequency of NK cells in the colorectal cancer test datasets as well as the current lack of further single-cell datasets on human colorectal cancer that include NK cells, we further examined our NK markers in approximately 4,000 NK cells extracted from a patient with melanoma (Fig. 2B; ref. 57). Using this dataset, we found that NK cells infiltrating in a distinct tumor nodule (in the cortex of the tumor) had the highest NK_Res signature, followed by NK cells within the cortex (margin) and center (core) of the tumor. NK cells from peripheral blood exhibited the lowest NK_Res score (Fig. 2B; Supplementary Fig. S9). Although NK_Exh scores were also relatively higher in the NK cells extracted from the nodule compared with the NK cells from blood, the differences between NK_Exh scores from different sources were not as high as those in NK_Res scores. This is consistent with the observed evidence of NK_Exh across NK cells extracted from different sources in the colorectal cancer single-cell test datasets (Fig. 2A and B), and suggests that NK cells from different sources (blood, cortex, center, and nodule) in patients with cancer may exhibit features of exhaustion, similar to what has been reported in patients with chronic hepatitis B (80).

Exhaustion and residency programs are enriched in MSI/dMMR and CMS1/CMS4 samples

As tumor samples from patients include both cancer cells and noncancer cells from the tumor microenvironment (such as infiltrating immune cells and stromal cells), we further refined our signatures to retain only genes with low expression in tumor cells and stromal cells. This increases our confidence when scoring bulk tumor samples, ensuring that scores are primarily defined by infiltrating immune cells and not tumor or stromal cells. To this end, we examined the expression of all our marker genes in the colorectal cancer cell lines from the CCLE and colorectal cancer LCM cells from Tsukamoto and colleagues (GSE21510; ref. 60), and retained genes that showed low expression in either of the two data sets (see “Materials and Methods”). The new signatures are named with “Bulk” suffix (e.g., CD8_Res_Bulk; Fig. 3; Supplementary Table S6). Next, for the list of genes in each of the Res_Bulk and Exh_Bulk signatures, we further examined their expression in stromal cells and removed genes that showed relatively higher expression in stromal cells (see “Materials and Methods”; Supplementary Figs. S10–S12; Supplementary Table S6); these included CSF1 from CD8_Exh_Bulk; PPAR15A, NRA43, and ITGA1 from CD4_Res_Bulk; FNDCB and ILLIR1 from CD4_Exh_Bulk. The final signatures after removal of these genes can be used to identify Res and Exh programs in bulk RNA-seq from tumor samples across larger patient cohorts.

We then used these signatures in colon cancer patient samples from two independent datasets: RNA-seq from TCGA (N_samples = 479, including 40 matched normal samples), and microarray from Marisa and colleagues (ref. 65; N_samples = 585, including 19 matched normal samples). Stratifying samples based on scores and MSI/dMMR status showed relatively higher Exh_Bulk and Res_Bulk scores in MSI-high (MSI-H) and dMMR samples, although the difference between MSI-H/dMMR and MSS/pMMR was larger for Exh_Bulk scores compared with Res_Bulk scores (Fig. 4A). We then divided samples based on the four colorectal cancer CMS, and in both tumor datasets we observed relatively higher Exh_Bulk and Res_Bulk scores in CMS1 and CMS4 subtypes (Fig. 4B), which is consistent with the characteristics of these subtypes (higher number of MSI samples and higher immune...
activation in CMS1, CSM4 and CMS2/3, respectively; refs. 81, 82). As the CMS1 subtype has higher occurrence in proximal colorectal cancer samples (right-side), we further confirmed higher scores, specifically for NK_Exh_Bulk signature, for proximal (right) compared with distal (left) tumors (Supplementary Fig. S13).

Consistent with higher Res scores in single-cell data from normal gut samples (Fig. 2), we also observed relatively higher Res_Bulk scores in normal bulk samples (scores were equivalent or higher than the scores in CMS1 and CMS4), with relatively lower Exh_Bulk scores, specifically for CD8⁺ and NK cells (Fig. 4B). These results not only are consistent with previous research showing higher immune infiltration in MSI-H and CMS1 samples (81, 82), but also suggest that the presence of exhausted cells is more cancer specific, with relatively less enrichment in normal tissues.

To examine the associations between our Exh_Bulk and Res_Bulk scores and signals from other immune cells, we scored TCGA and Marisa data against different immune-cell signatures from CIBERSORT as well as a general immune and stroma signatures.

Figure 2.
Test signatures in independent single-cell datasets of cancer and normal samples. A, UMAP blend plots of CD8⁺, CD4⁺, and NK cells generated using signature genes in two colorectal cancer (CRC) single-cell data sets (top and middle row) and normal gut cells (bottom row). Scores were scaled for better visualization of both scores on the same plot. B, Boxplots and UMAP blend plot for the NK cells extracted from a patient with melanoma, colored by the scaled NK_Exh and NK_Res scores. The middle lines in the boxplots represent mean scores, whereas the lower lines and upper lines represent the 25th percentile and 75th percentile of scores, respectively. See Supplementary Fig. S9 for the tissue source distribution of cells in the UMAP plot.
Our results showed high correlations between Exh_Bulk scores, specifically NK_Exh_Bulk scores, with the CIBERSORT scores for activated NK cells, activated dendritic cells, T cells, CD8+ T cells, neutrophils, and M1 macrophages in both datasets (Supplementary Fig. S14). All these signature scores showed relatively lower correlations with Res_Bulk scores.

To identify tumor intrinsic mutations associated with residency and exhaustion programs, we used somatic mutation data from the TCGA COAD cohort for two analyses. First, we examined the associations between gene mutations and the different signature scores.

Figure 3. Exhaustion and residency markers and their refinement in two cancer data sets CCLE and LCM. A, Hierarchical clustering of marker genes from different cell types; genes are annotated based on whether or not they pass the “bulk tumor threshold” in CCLE and LCM expression data. B, Expression of Res and Exh markers (colored orange and blue) in the CCLE and LCM data. N represents the number of cell lines or samples in the data sets. Genes annotated in black are those that pass the bulk tumor thresholds in at least one of the two datasets. We removed six other genes with relatively high expression in stromal cells: CSF1 from CD8_Exh_Bulk; PPP1R15A, NR4A3, and ITGA1 from CD4_Res_Bulk; and FNDC3B and IL1R1 from CD4_Exh_Bulk. CRC, colorectal cancer.
Exh_Bulk and Res_Bulk scores. In general, we found more gene mutations associated with exhaustion signatures than with residency programs (Supplementary Table S8), with the highest number related to NK_Exh_Bulk score \((N_{\text{Genes}} = 3,371)\), followed by CD8_Exh_Bulk \((N_{\text{Genes}} = 2,150)\). Among the residency signatures, CD8_Res_Bulk scores were associated with the highest number of gene mutations \((N_{\text{Genes}} = 928)\), followed by NK_Res_Bulk score \((N_{\text{Genes}} = 832)\).

Among the list of significant genes, we found several well-known colorectal cancer–associated gene mutations. Mutant BRAF showed significantly higher scores for all Exh_Bulk and Res_Bulk signatures, except for CD4_Res_Bulk. APC mutations, on the other hand, were associated with lower Exh_Bulk and Res_Bulk scores, with the difference being more profound for Exh_Bulk scores compared to Res_Bulk scores (Fig. 5). These results are consistent with previous research demonstrating more MSI samples with BRAF mutations and more APC mutations in MSS samples (82, 83). We also identified mutations within several genes that impact the tumor antigen presentation pathway and JAK/STAT signaling. Mutations in JAK1 were associated with both higher Exh_Bulk and Res_Bulk scores, whereas STAT1 mutations had significantly higher Exh_Bulk scores but not Res_Bulk scores. Finally, our results identified genes that are not traditionally associated with these programs, for example, SPEN (a transcriptional repressor, which synergizes with RUNX2) and several dynein
axonemal heavy chain (DNAH) genes (components of microtubules, involved in ATP-dependent cell motility; Supplementary Table S8). Focusing on the top gene mutations (\(P_{adj} < 0.0001\)) associated with the NK_Exh_Bulk program (\(N_{Genes} = 1,027\)), we identified several GO terms, such as actin filament–based processes, calcium ion transmembrane transport, regulation of GTPase activity, response to oxygen levels, response to TGF\(\beta\), and stem-cell proliferation (Supplementary Table S9; Supplementary Fig. S15).

In the second analysis, we focused on NK-cell signatures as they had a greater number of gene mutations associated with exhaustion. We performed elastic net regression by including all significant genes obtained in the first analysis (for each of the NK_Exh_Bulk and NK_Res_Bulk), as well as clinical factors (see “Materials and Methods”). This analysis highlighted 44 genes predictive of the NK_Exh_Bulk score; some examples of these predictive genes have links to antigen presentations (HLA-B, JAK1, TRRAP) and while

Figure S5.
Exh_Bulk and Res_Bulk scores for each cell type for TCGA samples with mutant (Mut) and WT forms of selected genes. The middle lines in the boxplots represent mean scores, whereas the lower lines and upper lines represent the 25th percentile and 75th percentile of scores, respectively. The \(P\) values for each gene represent the results of Wilcoxon tests between samples with WT and Mut for that gene. ***, \(P < 0.001\); **, \(P < 0.01\); and *, \(P < 0.05\); NS, not significant.
others (POLD1 and TGM6) have functional links that are less apparent. Mutations in all genes, other than APC, were associated with higher NK_Exh_Bulk scores. In addition, CMS3 showed a significantly negative coefficient for NK_Exh_Bulk, which likely reflects previous observations that this is an “immune-excluded” subtype (82). CMS4 was one of the predictive factors for both NK_Exh_Bulk and NK_Res_Bulk signatures. Indeed, CMS4 was the most predictive factor for NK_Res_Bulk scores (compared with all the genes and clinical factors included as predictors), and this is of interest as CMS4 is associated with EMT programs and TGFβ signaling, the latter playing an important role in differentiation of tumor-infiltrating resident cells (28, 29). We found 75 genes predictive of NK_Res_Bulk (mutations in 55 genes were associated with higher scores), of which 12 genes overlapped between the two signatures, including BRAF, APC, and HTR5A (Supplementary Table S10; Supplementary Figs. S16–S18). One of the predictive genes for NK_Res_Bulk was PAK4, which was recently identified as a target in cancer immune evasion as its expression was inversely correlated with CD8+ T-cell infiltration and ICB response in melanoma, and PAK4 overexpression correlates with an activated Wnt-β-catenin pathway (84, 85).

**Combined Exh and Res scores identify patients with distinct survival outcome**

As heterogeneity among infiltrating immune cells is associated with patient responses to different therapies and survival outcomes, we next examined whether the Res_Bulk and Exh_Bulk signatures were prognostic for survival. We used the OS data available for both the TCGA and Marisa datasets, as well as PFI in TCGA and RFS from the Marisa data (65).

Although each individual Res_Bulk and Exh_Bulk score did not consistently show significant associations with survival (with clinical factors also included as covariates for Cox models) in both datasets (Supplementary Fig. S19, Method section), stratifying samples based on several score pairs (e.g., NK_Exh_Bulk and TGFβ signaling in Fig. 6A) resulted in significant associations with survival outcome, which were consistent across these independent data (see “Materials and Methods”; Supplementary Fig. S20). Specifically, we observed poorer OS for patients with high NK_Exh_Bulk and high TGFβ signaling when compared with patients with low NK_Exh_Bulk and low TGFβ signaling within both datasets (Fig. 6B). Similar results were observed when comparing samples stratified by CD8NK_Exh_Bulk and TGFβ signaling (Supplementary Figs. S20 and S21). We observed a better OS for patients with high NK_Exh_Bulk and low NK_Res_Bulk compared with those with low NK_Exh_Bulk and high NK_Res_Bulk; this trend was also significant when performing a multivariate Cox analysis using PFI and RFS data, and was consistent in both datasets (Fig. 7). Similar trends were observed in PFI and RFS data when samples were stratified on the basis of CD8NK_Exh_Bulk and NK_Res_Bulk score pairs, or based on CD8NK_Res_Bulk and NK_Exh_Bulk score pairs. In this case, high Exh_Bulk and low Res_Bulk scores were associated with better outcomes compared with low Exh_Bulk and high Res_Bulk scores (Supplementary Fig. S20; Supplementary Figs. S22 and S23). Samples with the opposite characteristics included different CMS subtypes (Fig. 7A) and different MSI status (Supplementary Fig. S24). This suggests that our scoring approach adds a new layer of information for sample stratification beyond known clinical variables, with Exh

---

**Figure 6.**

A. Relative NK_Exh_Bulk and TGFβ signaling scores for samples in TCGA and Marisa data, annotated (marker color) with CMS subtype. Pearson correlation coefficients between the two signature scores are shown in parentheses in the title of each scatter-plot. B, Kaplan-Meier curves demonstrating OS for patients stratified on the basis of scores shown in A. The numbers in parentheses represent the number of samples in each group. The P values from log-rank test as well as Cox proportional hazard test are shown in the title of each plot. NA, unclassified samples.
and Res scores predictive of survival outcome in colorectal adenocarcinoma.

Although Res_Bulk and Exh_Bulk scores mainly showed positive correlations with TGFβ signaling, mesenchymal programs (driven by TGFβ or hypoxia and EGF) and KRAS signaling within both datasets, negative correlations were observed between Res_Bulk and Exh_Bulk scores and DNA damage repair (DDR), oxidative phosphorylation (OXPHOS), peroxisome, and epithelial scores (Supplementary Fig. S20). However, when focusing on each of the cancer-related signature scores, there were differences in their correlation with Res_Bulk and Exh_Bulk scores. For example, consistent with our findings in single-cell data (Supplementary Fig. S8), we observed higher correlation between TGFβ signaling scores and Res_Bulk scores compared with those from TGFβ signaling scores and Exh_Bulk scores (Supplementary Fig. S25).

Discussion

Several genes in our Exh and Res signatures have been previously reported as part of the exhausted/dysfunctional or resident T-cell signatures in different cancer types (7, 8, 18, 19), although some have received less attention in this context. We identified members of the APOBEC family of single-stranded DNA cytosine deaminases in our CD8_Exh signatures (APOBEC3H and APOBEC3C). APOBEC genes...
are thought to have similar functions in the context of HIV and cancer (86). Indeed, some APOBEC genes are expressed in tumor-infiltrating T cells and are associated with better clinical outcome (87, 88). GAPDH, one of our identified Exh markers, induces the translation of IFNG and IL2 in highly glycolytic T cells and suppresses translation of key genes in CD4+ T cells following inhibition of glycolysis (89). In addition, GAPDH is upregulated in rapidly cycling and activated NK cells compared with slowly cycling NK cells (90). One of the CD4_Exe genes associated with GPCR was CCR8, and combining anti-CCR8 with other immunotherapy agents in colorectal cancer has been shown to prevent the suppressive effect of CD4+ Tregs and improve patient outcomes (91). GPR15, one of our CD8_Res markers, which also has restricted expression in resident CD4+ and NK cells, encodes a GPCR involved in T-cell trafficking in colon, and its expression in Tregs is promoted by TGFβ, mediating the suppression of inflammation in the colon (79). CXCRI4, identified as one of our CD4_Res genes involved in GPCR signaling, was just recently suggested to mediate immune suppression in pancreatic and colorectal cancer samples, and the use of anti-CXCRI4 in MSS samples increased the immune response (92).

There appears to be heterogeneity in tissue residency programs across different cancer types. For example, our analyses indicate that ZNF668 (HOBIT), a central mediator of lymphocyte tissue-residency in mice (93), may act more as an Exh marker in colorectal cancer, with higher relative abundances within exhausted CD8+ and NK cells compared with their respective resident cells. This is in contrast to NSCLC, where relatively high expression is seen in resident T cells (7, 93). These data are consistent with other studies in which ZNF668 expression is shown to be higher in CD8+ effector T cells in NSCLC compared with hepatocellular carcinoma and colorectal cancer (6–8). Although we identified the NRE4A family genes as Res signature genes, the NRE4A family has been previously linked to CD8+ T-cell exhaustion (94) and shown to limit the function of chimeric antigen receptor (CAR) T cells in solid tumors (95). Marquardt and colleagues reported the upregulation of ZNF331 within CD69+ CD49a+ CD103+ resident NK cells from normal lung tissue compared with CD69- cells (96), and we identified this gene as one of our CD8_Res genes. They also observed higher expression of IRF4 in the resident NK cells, whereas we removed this gene from our list due to a particularly high correlation between ZNF668 in CD69+ CD49a+ CD103+ cells and ZNF331 in hepatocellular carcinoma and colorectal cancer (6–8). Although we identified the NRE4A family genes as Res signature genes, the NRE4A family has been previously linked to CD8+ T-cell exhaustion (94) and shown to limit the function of chimeric antigen receptor (CAR) T cells in solid tumors (95). Marquardt and colleagues reported the upregulation of ZNF331 within CD69+ CD49a+ CD103+ resident NK cells from normal lung tissue compared with CD69- cells (96), and we identified this gene as one of our CD8_Res genes. They also observed higher expression of IRF4 in the resident NK cells, whereas we removed this gene from our list due to a particularly high correlation with both Res and Exh programs in NK cells. Notably, we here identified IRF8 as a TF involved in Res programs.

The unsupervised trajectory analysis suggested that exhausted CD8+ T cells derive from both resident cells and circulating CD8+ T cells. This is in line with the reported diversity in the origins of exhausted CD8+ T cells (5). Zhang and colleagues observed a developmental link between effector memory T (Tmem) and exhausted T cells in patients with colorectal cancer (8), whereas in NSCLC, Guo and colleagues showed a direct link between Tmem and exhausted cells (7). Li and colleagues showed that the dysfunctional (or exhausted) cells in patients with melanoma derive from transitional (early effector GZMK+) T cells through a local differentiation process, without any evidence of exhaustion in peripheral blood (19), whereas Huang and colleagues analyzed exhausted CD8+ T cells in peripheral blood of patients with melanoma that were treated by anti-PD-1 therapy, and had T-cell clones in common with the tumor (21). The obtained trajectory of the NK and CD8+ Res cells in our results passed through an intermediate subpopulation with coexpression of GZMK, naïve cell markers, and some inhibitory receptors. Similarly, Galletti and colleagues have reported two subsets of CCR7+CD8+ T cells, one of which expresses GZMK along with PDCD1 and TIGIT (termed T precursor exhausted-like cells), which commit to a terminally dysfunctional state (97). Therefore, this intermediate subpopulation of NK cells may represent the NK precursor—exhausted cells.

On the basis of our mutation analysis, the NK_Exh_Bulk signature was associated with the highest number of gene mutations. This suggests that tumor mutation burden and related neoantigen expression may drive tumor inflammation and consequently impact NK-cell activation and exhaustion. Another interpretation is that in the presence of highly activated NK cells, tumor cells undergo positive selection for cells with mutations resulting in decreased immunogenicity to these cytotoxic lymphocytes as an immune surveillance escape strategy (98). This second possibility is also supported by our findings of gene mutations that were predictive of NK_Exh_Bulk (e.g., genes associated with antigen presentations). Another predictive gene of NK_Exh_Bulk was POLD1, mutations of which have received attention as biomarkers for cancer immunotherapy due its crucial role in DNA replication and repair (99, 100). We also found several less appreciated genes in this context, such as TGM6, a transglutaminase (TGM), which similar to other TGMs, such as TGM2, performs posttranslational modification in a Ca2+-dependent manner, and may play a role in cancer (101). We also identified calcium ion transmembrane transport as one of the significant GO terms associated with NK_Exh_Bulk genes; intracellular Ca2+ is required for cancer cell proliferation and apoptosis and for the efficacy of cytotoxic T-cell and NK-cell function (102).

GO enrichment of the mutations associated with NK_Exh_Bulk also suggested the importance of several actin filament–based and cytoskeletal processes, which are much less appreciated in the context of tumor immune infiltration. Solid tumors undergo massive remodeling after attack by cytotoxic T and NK cells; this remodeling not only impacts tumor-cell recognition by lymphocytes, but also influences immune-cell activation (103). We also noted that some of our Exh signature genes (such as MYO7A and AFAP1L2, shared by CD8+ and NK cells) were those associated with actin and myosin-based processes. Remodeling of actin filaments and myosin contractibility are required for the formation of immunologic synapse as well as transportation and secretion of cytotoxic granules in T and NK cells (104), and increased expression of some actin modulating genes is reported in rapidly cycling NK cells that exhibited reduced cytotoxicity (90).

In both single-cell and bulk tumor data, we observed positive correlations between TGFβ-related signature scores and Res/Exh scores, and this correlation was higher for Res scores compared with Exh scores. In this study, we took this a step further and demonstrated how combining TGFβ signaling scores with NK_Exh_Bulk scores could predict survival outcome. We also showed the importance of the combined use of NK_Res_Bulk and NK_Exh_Bulk signatures for predicting survival outcome, independent of known colorectal cancer clinical factors (e.g., age, stage, MSI and CMS). This could be clinically important given that MSI status alone is not sufficient in predicting response to immunotherapy or survival outcome (2, 82), approximately 13% of patients with colorectal cancer cannot be assigned to any of the CMS subtypes and almost all CMS subtypes include a mixture of samples with both poor and good prognosis, the exception being CMS4 and relapsed CMS1, which are associated with poorer survival outcomes (105). By taking into account that: (i) samples with lower TGFβ signaling and lower NK_Res_Bulk are associated with better survival, whereas samples with lower NK_Exh_Bulk and higher TGFβ signaling show poorer outcome. (ii) TGFβ is associated with T-cell exclusion in several cancer types and targeting TGFβ is beneficial for response to ICIs (33), and (iii) we observed better survival outcome for...
patients with high NK_Exh_Bulk and low NK_Res_Bulk programs (compared with those with low NK_Exh_Bulk and high NK_Res_Bulk), it is tempting to speculate that part of the suppressive biology associated with TGFβ signaling is mediated by the cytokine promoting NK-cell residency.

This study suggested a better survival outcome for patients with higher Exh to Res programs (specifically for NK and combined NK and CD8 signatures) by defining these programs via nonoverlapping gene signatures, whereas several previous studies reported associations between CD8 Res cells and better response to ICBs and improved survival outcome (23–25). In some cases, these differences could be due to cancer types and heterogeneity in Res cell populations, but it is also at least partially due to the use of a nonspecific signature or a mixed Exh and Res gene sets in these studies, and as noted above, there is a strong degree of overlap between T-cell activation and Exh markers, which are difficult to distinguish in the absence of time course data. For example, many studies have defined Res cells using ITGAE (CD103) (24, 25). We removed this gene based on the results of the canonical correlation analysis on the initial Res and Exh signatures in NK and CD69 cells, which means that the presence of this gene was associated with both Res and Exh programs. Duhem and colleagues identified a subset of CD8+ TILs with Tm phenotype with coexpression of ITGAE and ENTPD1 (CD39), which were also enriched for Exh markers (PDCD1, CTLA4, and HAVCR2) and had high expression of CD69; they showed that these cells are associated with high tumor reactivity and better prognosis in patients with head and neck cancer (106). Therefore, high expression of ITGAE and better survival outcome may not be specifically attributed to Res cells. Another example is the Tm signature reported by Byrne and colleagues (23), which includes a mix of Res (e.g., CD69) and even more Exh markers (HAVCR2, PDCD1, TIGIT, LAYN, etc). Our current study not only introduces novel distinct Exh and Res programs in colorectal cancer T cells and in particular in NK cells, but also shows the importance of such signatures and their combinations with known genomics and molecular information in stratification of samples to predict survival outcome and potentially their response to immunotherapy.

Authors’ Disclosures

M. Foroutan reports other support from Prostate Cancer Foundation during the conduct of the study. A. Pfeifferle reports personal fees from oNKo-innate Pty Ltd. outside the submitted work. J. Cursons reports other support from oNKo-innate Pty Ltd during the conduct of the study, as well as other support from oNKo-innate Pty Ltd. outside the submitted work. N.D. Huntington reports grants from National Health and Medical Research Council and Prostate Cancer Foundation during the conduct of the study, as well as personal fees from oNKo-innate Pty Ltd. outside the submitted work. No disclosures were reported by the other authors.

Authors’ Contributions

M. Foroutan: Conceptualization, resources, data curation, software, formal analysis, validation, investigation, visualization, methodology, writing—original draft, writing—review and editing. R. Molania: Resources, software, formal analysis. A. Pfeifferle: Formal analysis, writing—review and editing. C. Behrenbruch: Resources, visualization, methodology. S. Scheer: Investigation. A. Kallies: Resources, writing—review and editing. T.P. Speed: Formal analysis, supervision, methodology. J. Cursons: Conceptualization, resources, data curation, software, formal analysis, supervision, methodology, writing—review and editing. N.D. Huntington: Conceptualization, resources, supervision, funding acquisition, writing—original draft, writing—review and editing.

Acknowledgments

This work was supported by project grants from the National Health and Medical Research Council (NHMRC) of Australia (1124907, 1124784, 1049407, 1066770, 1057852, and 1027472) and a fellowship (1124788, to N.D. Huntington). N.D. Huntington is a recipient of Research Grants from the Harry J. Lloyd Charitable Trust (USA), Melanoma Research Alliance (USA), Tour de Cure (AUS), the Ian Potter Foundation (AUS), Cancer Council of Victoria (1145730), a CLIP grant from Cancer Research Institute (USA), and a Prostate Cancer Foundation Young Investigator (PCF-YI) Award (USA). The results shown in this study are in part based upon data generated by the TCGA Research Network (https://www.cancer.gov/tcga) and CCLE data generated by the Broad Institute (https://portals.broadinstitute.org/ccle). The authors thank Dr. Miranda and Dr. de Vries from Leiden University Medical Center for providing single-cell data (de Vries and colleagues, Gut 2020) used in this study. The authors also extend their gratitude to the patient donors who contributed samples that were analyzed in this work.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 19, 2021; revised June 14, 2021; accepted August 17, 2021; published first August 19, 2021.

References

1. Llosa NJ, Luber B, Tam AI, Smith KN, Siegel N, Awan AH, et al. Intratumoral adaptive immnosuppression and type 17 immunity in mismatch repair proficient colorectal tumors. Clin Cancer Res 2019;25:5256–9.
2. Baretz M, Le DT. DNA mismatch repair in cancer. Pharmacol Ther 2018;189:45–62.
3. Huntington ND, Cursons J, Rautela J. The cancer–natural killer cell immunity cycle. Nat Rev Cancer 2020;20:437–54.
4. Cursons J, Souza Fonseca Guimaraes F, Foroutan M, Anderson A, Hollande F, Hedjyd-Jedh S, et al. A gene signature predicting natural killer cell infiltration and improved survival in melanoma patients. Cancer Immunol Res 2019;7:1162–74.
5. Yu X, Zhang L, Chaudhry A, Rapaport AS, Ouyang W. Unravelling the heterogeneity and dynamic relationships of tumor infiltrating T cells by single-cell RNA sequencing analysis. J Leukoc Biol 2020;107:917–32.
6. Zheng C, Zheng L, Yoo J, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell 2017;169:1342–1356.e16.
7. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med 2018;24:978–85.
8. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature 2018;564:268–72.
9. Blank CU, Haining WN, Held W, Hogan PG, Kallies A, Lugli E, et al. Defining ‘t cell exhaustion’. Nat Rev Immunol 2019;19:665–74.
10. Kallies A, Zehn D, Utschneider DT. Precursor exhausted T cells: key to successful immunotherapy? Nat Rev Immunol 2020;20:128–36.
11. Utschneider DT, Charmoy M, Chennupati V, Pouille L, Ferreira DP, Calderon-Copete S, et al. T cell factor 1-expressing memory-like CD8(+) T cells sustain the immune response to chronic viral infections. Immunity 2016;45:415–27.
12. Im SJ, Hashimoto M, Gerner MY, Lee J, Kissick HT, Burger MC, et al. Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. Nature 2016;537:417–21.
13. Leong YA, Chen Y, Ong HS, Wu D, Man K, Delage C, et al. CXCGR5+ follicular cytotoxic T cells control viral infection in B cell follicles. Nat Immunol 2016;17:1187–96.

14. Wherry EJ, Karachali M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol 2015;15:486–99.

15. Siddiqui J, Chaeubple K, Chennupati V, Fuertes Marraco SA, Calderon-Copete S, Pais Ferreira D, et al. Intratumoral Tcf11+(+)-PD-1(-) CD8(+)+ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. Immunity 2019;50:195–211 e10.

16. Sade-Feldman M, Yi H, Shek K, Reidy JA, Ray JP, de Boer CG, Jenkins RW, et al. Defining T cell states associated with response to checkpoint immunotherapy. Nat Immunol 2019;20:326–36.

17. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A cancer-intrinsic transcriptional program promotes T cell exclusion and resistance to checkpoint blockade. Nat Immunol 2019;20:1676:775.

18. Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O18. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A cancer-intrinsic transcriptional program promotes T cell exclusion and resistance to checkpoint blockade. Nat Immunol 2019;20:1676:775.

19. Zhang Y, Souza-Fonseca-Guimaraes F, Bald T, Ng SS, Young A, Ngiow SF, et al. Antigen-driven tumor-specific T cell responses differentially mediate tumor control and respond to checkpoint blockade. Immunity 2016;50:195–206.

20. Byrne A, Savas P, Sant L, Li R, Virassamy B, Luen SJ, et al. Tissue-resident memory T cells: new insights into their development and functions. Nat Immunol 2013;14:1294–301.

21. Dumesnil N, Labano S, Romero P. Tumor resident memory T cells: new insights into their development and functions. Nat Immunol 2013;14:1294–301.

22. Dumauthioz N, Labiano S, Romero P. Tumor resident memory T cells: new insights into their development and functions. Nat Immunol 2013;14:1294–301.

23. Casey KA, Fraser KA, Schenkel JM, Moran A, Abt MC, Beura LK, et al. Antigen-specific regulatory programs conferring phenotypic identities to human NK cells. Cell 2019;178:341–54.e20.

24. Edwards J, Wilmott JS, Madore J, Gide TN, Queck C, Tasker A, et al. CD103(+) tumor-infiltrating lymphocytes: a tumor-specific, tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. J Immunol 2015;194:3475–86.

25. Sellier C, Belz GT, Huntingdon ND. Development, homeostasis, and heterogeneity of NK cells and ILC1. Curr Top Microbiol Immunol 2016;395:37–48.

26. Gao Y, Souza-Fonseca-Guimaraes F, Bald T, Ng SS, Young A, Ngiow SF, et al. Antigen-driven tumor-specific T cell responses differentially mediate tumor control and respond to checkpoint blockade. Immunity 2016;50:195–206.

27. Utzschneider DT, Gabriel SS, D103(+) tumor-resident CD8(+) T cells with enhanced protective function. Immunity 2014;40:961-973.

28. Foroutan M, Cursons J, Hediyeh-Zadeh S, Thompson EW, Davis MJ. A developmental pathway for CD103(+) tumor-resident CD8(+) T cells. Cell 2020;181:442–56.e12.

29. The Immunological Genome Project. ImmGen at 15. Nat Immunol 2020;21:700–3.

30. Corridoni D, Antanaviciute A, Gupta T, Fawker-Corbett D, Aulicino A, Jagielowicz M, et al. Single-cell atlas of colonic CD8(+) T cells in ulcerative colitis. Nat Med 2020;26:1480–90.

31. Nath AP, Braun A, Ritchie SC, Carbonne FR, Mackay LK, Gebhardt T, et al. Comparative analysis reveals a role for TGF-β in shaping the residency-related transcriptional signature in tissue-resident memory CD8(+) T cells. PloS One 2019;14:e0201045.

32. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.

33. Wickham H, Averick M, Bryan J, Chang W, McGowan L, Francois R, et al. Welcome to the tidyverse. J Open Source Softw 2019;4:1868.

34. Gu Z, Eds R, Schlener M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–9.

35. Wickham H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York; 2016.

36. Jansen CS, Prokhnova N, Master VA, Sanda MG, Carlisle JW, Bilen MA, et al. An intra-tumoral niche maintains and differentiates stem-like CD8 T cells. Nature 2019;576:465–70.

37. Comprehensive integration of single-cell data. Cell 2019;177:1888–1902.

38. Jansen CS, Prokhnova N, Master VA, Sanda MG, Carlisle JW, Bilen MA, et al. An intra-tumoral niche maintains and differentiates stem-like CD8 T cells. Nature 2019;576:465–70.

39. McDavid AFG, Yajima M. MAST: model-based analysis of single cell transcriptomics data. Cell 2019;177:1888–1902.

40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for the analysis of gene expression data from RNA-seq counts. Nat Biotechnol 2010;28:577–85.

41. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.

42. Wickham H, Averick M, Bryan J, Chang W, McGowan L, Francois R, et al. Welcome to the tidyverse. J Open Source Softw 2019;4:1868.

43. Gu Z, Eds R, Schlener M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–9.

44. Wickham H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York; 2016.

45. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.

46. Wickham H, Averick M, Bryan J, Chang W, McGowan L, Francois R, et al. Welcome to the tidyverse. J Open Source Softw 2019;4:1868.

47. Gu Z, Eds R, Schlener M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–9.

48. Wickham H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York; 2016.

49. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.

50. Wickham H, Averick M, Bryan J, Chang W, McGowan L, Francois R, et al. Welcome to the tidyverse. J Open Source Softw 2019;4:1868.

51. McDavid AFG, Yajima M. MAST: model-based analysis of single cell transcriptomics data. Cell 2019;177:1888–1902.

52. Combinatorial analysis reveals a role for TGF-β in shaping the residency-related transcriptional signature in tissue-resident memory CD8(+) T cells. PloS One 2019;14:e0201045.

53. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
