Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer

Anusha-Preethi Ganesan[1,2], James Clarke[3], Oliver Wood[3], Eva M Garrido-Martin[4], Serena Chee[3], Toby Mellows[4], Daniela Samaniego-Castruita[1], Divya Singh[1], Grégory Seumois[1], Aiman Alzetani[5], Edwin Woo[5], Peter S Friedmann[4], Emma V King[3], Gareth J Thomas[3], Tilman Sanchez-Elsner[4], Pandurangan Vijayanand[1,4,6] & Christian H Ottensmeier[3,5,6].

[1]La Jolla Institute for Allergy & Immunology, La Jolla, CA.

[2]Division of Pediatric Hematology Oncology, Rady Children's Hospital, University of California San Diego, San Diego, CA.

[3]Cancer Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

[4]Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, Faculty of Medicine University of Southampton, Southampton, United Kingdom.

[5]Southampton University Hospitals NHS foundation Trust, Southampton, United Kingdom.

[6]These authors jointly directed this work.

Correspondence should be addressed to P. V. (vijay@lji.org)

ABSTRACT

Therapies that boost anti-tumor responses of cytotoxic T lymphocytes (CTLs) have shown promise, however clinical responses to currently available immunotherapeutic agents vary considerably, for which the molecular basis is unclear. We performed transcriptomic profiling of tumor-infiltrating CTLs from treatment-naïve patients with lung cancer to define the molecular features associated with robustness of anti-tumor immune responses. We observed marked heterogeneity in the expression of molecules associated with T cell antigen receptor (TCR) activation and immune checkpoints such as 4-1BB, PD-1, TIM-3. Transcripts linked to tissue-resident memory cells (T_{RM}), such as *CD103*, were enriched in tumors containing high density of CTLs, and CTLs from $CD103^{hi}$ tumors displayed features of enhanced cytotoxicity. Higher density of T_{RM} cells in tumors predicted better survival outcomes in lung cancer, and this effect was over and above that conferred by CTL density. Hereby, we define the molecular fingerprint of tumor-infiltrating CTLs and identify potentially new immunotherapy targets.

INTRODUCTION

Immunotherapy is rapidly gaining its place as a standard treatment for solid tumors[1,2], including lung cancer[3]. Nonetheless, only ~30% of patients benefit from this approach[4]. Much remains to be learned about how immunotherapies work and how to choose the right treatment or combination for a particular patient. Understanding the mechanisms and molecular basis of effective anti-tumor immune responses will be essential to develop novel immunotherapeutic agents for those patients who do not respond to currently available immunotherapies.

Immunotherapies are thought to enhance the antitumor responses of cytotoxic T lymphocytes (CTLs) *i.e.*, $CD8^+$ T cells that infiltrate into the tumor[5]. Indeed, a high density of tumor-infiltrating lymphocytes (TILs) predicts good prognosis in a wide range of cancers[6,7]. However, it remains unclear why the degree of infiltration by TILs varies significantly even between individuals with the same cancer. It is also unknown whether there are merely quantitative differences in the number of TILs or whether qualitative differences also exist in TILs from tumors with high TIL density that may contribute to the superior outcome seen in these patients. An understanding of the TIL transcriptome and the molecular basis of TIL heterogeneity could lead not only to novel biomarkers for

patient stratification for therapy, but also identify novel immune pathways to be targeted by future immunotherapeutic strategies.

To date, transcriptional studies of $CD8^+$ T cells from cancer patients have analyzed cells in peripheral blood or metastatic sites[8-11]. The precise state of $CD8^+$ T cell activation, differentiation and function within primary tumors, is poorly understood; however, this must be a key reference point from which to begin unraveling the biology of immune attack at the time of diagnosis, tumor progression and after intervention with immunotherapies. In order to fully characterize the molecular nature of immune responses at the tumor site, we have taken an unbiased approach to define the global transcriptional profile of purified $CD8^+$ TILs from well-characterized cohorts of patients with two epithelial cancers, non-small cell lung cancer (NSCLC) and head and neck squamous cell cancer (HNSCC).

RESULTS

Major transcriptional changes characterize $CD8^+$ TILs

To identify the core transcriptional signature of $CD8^+$ TILs, we performed RNA sequencing of purified populations of $CD8^+$ T cells present in tumor samples ($CD8^+$ TILs) from 36 patients with treatment-naïve early stage non-small cell lung cancer (NSCLC) (**Supplementary Fig. 1a, Supplementary Table 1** and **2**). We also generated matched transcriptional profiles of $CD8^+$ T cells isolated from the adjacent non-tumor lung tissue ($CD8^+$ N-TILs) to discriminate features linked to lung tissue residence from those related to tumor infiltration. To assess the conservation of the transcriptional program of $CD8^+$ TILs in a related solid tumor of epithelial-origin, we utilized a similar data set generated in 41 patients with head and neck squamous cell carcinoma (HNSCC) from both human

papilloma virus (HPV)-positive (virally-driven) and HPV-negative subtypes (**Supplementary Fig. 1a**, **Supplementary Table 1** and **2**).

We identified a large number of transcripts (*n* = 1403) that were differentially expressed by $CD8^+$ TILs when compared to $CD8^+$ N-TILs (**Fig. 1a** and **Supplementary Table 3**), suggesting major changes in the transcriptional landscape of $CD8^+$ TILs in lung tumor tissue. The expression of lung cancer '$CD8^+$ TIL-associated transcripts' did not differ according to histological subtype (**Supplementary Fig. 1b**). Principal-component analysis (PCA) and hierarchical clustering also showed that $CD8^+$ TILs from both subtypes of lung cancer mostly clustered together, distinct from the $CD8^+$ N-TILs (**Fig. 1b** and **Supplementary Fig. 1c,d**). Interestingly, this set of lung cancer '$CD8^+$TIL-associated transcripts' were similarly expressed in $CD8^+$ TILs in both subtypes of HNSCC (**Fig. 1a** and **Supplementary 1b**), which also clustered together with $CD8^+$ TILs from lung cancer (**Fig. 1b** and **Supplementary Fig. 1c,d**), indicating a conserved TIL transcriptome for these two tumor types.

Features associated with inhibited T cell function, anergy and senescence have been described in TILs[12-14]. Gene set enrichment analysis (GSEA) revealed significant enrichment of genes linked to the so-called exhaustion stage, such as *PDCD1* (which encodes for PD-1), *CTLA4*, *HAVCR2* (which encodes for TIM-3) and *KLRG1*, while genes associated with T cell anergy and senescence were not enriched (**Fig. 1c,d**). T cell-associated genes derived from The Cancer Genome Atlas (TCGA) of lung cancer[15] were also enriched in our data set (**Fig. 1d**). Together these findings suggest that our strategy for micro-scaled RNA-Seq analysis of freshly purified *ex vivo* $CD8^+$ TILs and $CD8^+$ N-TILs reliably identifies transcripts previously linked to TILs.

Cell cycle and TCR activation pathways in $CD8^+$ TILs

To gain broad insight into the functional relevance of the $CD8^+$ TIL transcriptional program, we performed gene pathway analysis. Interestingly, in TILs, we observed significant enrichment of transcripts encoding overlapping sets of genes involved in cell cycle control, mitosis, DNA replication and signaling *via* the tumor suppressor p53, ataxia telangiectasia mutated (ATM) and polo-like kinase (PLK) pathways (**Fig. 2a-c** and **Supplementary Table 4**), indicating that proliferating $CD8^+$ T cells are enriched in TILs (tumors). Furthermore, we observed enrichment of canonical pathways involved in antigen-specific T cell activation, especially the 4-1BB (tumor necrosis factor receptor superfamily member 9, TNFRSF9)-mediated and CD27 co-stimulatory pathways that are activated following TCR engagement and co-stimulation by antigen-presenting cells (APCs), respectively[16,17] (**Fig. 2a**,**d**). The increased expression of 4-1BB in $CD8^+$ TILs was confirmed at the protein level by flow cytometry (**Fig. 2e**). Together these data suggest that TCR engagement and co-stimulation, presumably provided by APCs expressing tumor-associated antigens (TAA), are likely to be involved in antigen-specific activation and proliferation of $CD8^+$ TILs, implying that the tumor milieu sustains clonal expansion of presumed TAA-specific $CD8^+$ T cells. This suggestion was further supported by analysis of the TCR repertoire, which indicated significantly greater clonal expansion of $CD8^+$ TILs compared to N-TILs (**Fig. 2f** and **Supplementary Table 5**).

Heterogeneity in immunotherapy target molecules

The remarkable success of immune checkpoint blockers such as anti-PD-1 and anti-CTLA-4 agents in humans and in model organisms[4,18] suggests that $CD8^+$ TILs with features of TCR engagement and strong co-stimulation are likely to mount robust anti-tumor immune responses. However, the response to such treatments is highly variable and limited to a minority of patients. Such inter-individual variability in response may be dictated by the underlying molecular profile of $CD8^+$ TILs, which may also reveal

alternative immune evasion mechanisms besides PD-1 and CTLA-4-based pathways. Therefore, we examined the expression of a spectrum of potential immunotherapy target molecules to uncover the extent of molecular heterogeneity in $CD8^+$ TILs. We observed marked variability in the expression of transcripts encoding *PDCD1* and other potential immunotherapy target molecules across different lung cancer (**Fig. 3a**,**b**) and HNSCC patients (**Supplementary Fig. 2**). We confirmed PD-1 expression at the protein level and showed that the abundance of *PDCD1* transcripts correlates with the average number of PD-1 expressing cells in the tumors (**Supplementary Fig. 3a,b**). We also found varying combinations of the expression of co-inhibitory molecules; for example, $CD8^+$ TILs from some lung cancer patients displayed upregulation of transcripts encoding 4 immunotherapy target molecules, namely, PD-1, TIM-3, LAG-3 and CTLA-4, while others showed different patterns of three, two or even single molecule expression (**Fig. 3a**,**b**). The high molecular resolution and breadth of our data suggests that baseline transcriptional profiling of tumor-infiltrating $CD8^+$ T cells may guide selection of appropriate immunotherapies for each patient and the development of biomarkers that predict clinical response to checkpoint blockade with mono- or combination therapies.

PDCD1 expression correlates with TIL density

The marked heterogeneity observed in *PDCD1* transcript abundance led us to investigate factors linked to *PDCD1* expression in $CD8^+$ TILs. Despite the perceived negative regulatory role of PD-1 as an immune checkpoint, it serves as a marker for clonally expanded, antigen-specific T cells capable of lysing autologous tumor cells[19,20]. Furthermore, we found a strong positive correlation between the expression of *PDCD1* and *4-1BB* (**Fig. 3c**), a molecule expressed following TCR engagement and thus a marker of antigen-specific T cells[16,17,21]. The heterogeneity in the expression of these surrogate markers for antigen specificity suggests that not all tumors contain similar

numbers of tumor-reactive $CD8^+$ TILs. Hence, we asked what factors might influence the enrichment of *PDCD1*- and *4-1BB*-expressing $CD8^+$ TILs, *i.e.* cells we presume to be TAA-specific. We found no correlation of *PDCD1* or *4-1BB* transcript abundance with clinical or pathological characteristics such as patient age, gender, histological subtype, stage of disease, performance status or smoking status (**Supplementary Fig. 3c**). However, there was a positive correlation between the abundance of both these transcripts and the average number of $CD8^+$ TILs infiltrating each tumor sample (**Fig. 3c**). A similar correlation was also observed with the abundance of *CD8A* transcripts in lung tumor samples from TCGA RNA-Seq data set (http://cancergenome.nih.gov) (**Fig. 3d**).

Besides *PDCD1* and *4-1BB*, tumors with high TIL density (classified based on the average number of $CD8\alpha^+$ T cells infiltrating tumors into TIL^{hi}, TIL^{int} and TIL^{lo} tumors, **Supplementary Fig. 4)** relative to TIL^{lo} tumors also had higher transcript expression for several other immunotherapy target molecules such as TIM-3, LAG-3 and TIGIT (**Fig. 3e**). Previous studies have linked PD-1 and 4-1BB to both exhaustion[22] and antigen-specific TCR activation[19,20], but the positive correlation of their expression with TIL density implies that their higher expression likely reflects enrichment of activated TAA-specific $CD8^+$ T cells.

$CD8^+$ T_{RM} cells are enriched in TIL^{hi} tumors

Tumors with high TIL density have better outcomes when compared to those with low TIL density[6]. Besides the numerical changes in T cells, it is not known if there are qualitative differences in tumor-infiltrating $CD8^+$ T cells between these groups. Defining such features may provide insights into the mechanisms that govern the magnitude and specificity of anti-tumor $CD8^+$ T cells responses.

We found 109 transcripts for which expression differed significantly between TIL^{hi} *versus* TIL^{lo} tumors (**Fig. 4a** and **Supplementary Table 6**). As expected, transcripts involved in TCR activation (*4-1BB, PDCD1*) were upregulated in TIL^{hi} tumors (**Fig. 4a**), consistent with the enrichment of presumed TAA-specific $CD8^+$ T cells, although the specificity requires further experimental confirmation. Several other transcripts associated with tissue retention of lymphocytes and tissue-resident memory T cells (T_{RM}) were differentially expressed in TIL^{hi} tumors as compared to TIL^{lo} tumors (**Supplementary Table 6** and **Supplementary Fig. 5a**). For example, *ITGAE* (*CD103*) encodes the α_E subunit of the integrin molecule $\alpha_E\beta_7$ (human mucosal lymphocyte-1 antigen), which binds the adhesion molecule E-cadherin expressed by epithelial cells in barrier tissues[23,24]. Expression of this marker of T_{RM} cells was enriched in TIL^{hi} tumors (**Fig. 4a**,**b**) and positively correlated with the average number of $CD8^+$ cells within tumors (**Fig. 4c**). This finding was also validated in the TCGA lung cancer data set (**Fig. 4c**). We confirmed CD103 expression in $CD8^+$ TILs at the protein level by immunohistochemistry and flow cytometry (**Fig. 4d**,**e**)**.** Surface molecules linked to T_{RM} cells[25,26] such as CD69 and CD49a (ITGA1) were highly co-expressed with CD103, and those linked to effector memory cells (KLRG1) and central memory cells (CCR7 and CD62L) had reduced expression in $CD103^+CD8^+$ TILs compared to $CD103^-CD8^+$ TILs, suggesting that the former population represents T_{RM} cells (**Fig. 4f** and **Supplementary Fig. 5b**). We also observed co-expression of PD-1 and 4-1BB in 6% and 4% respectively of $CD103^+CD8^+$ TILs as demonstrated in a representative patient sample (**Fig. 4e**)**.**

Another transcript enriched in TIL^{hi} tumors was *CXCR6* (**Fig. 4a**,**b**), whose expression is not only linked to T_{RM} cells[25], but is also important for the localization and function of tissue-residing T cells[27,28]. *S1PR1* and *KLF2* transcripts, known to be downregulated in T_{RM} cells[23], were also diminished in TIL^{hi} tumors (**Fig. 4b**).

Downregulation of *S1PR1*, which encodes sphingosine 1-phosphate receptor 1 (S1P1), is necessary for the egress of T cells from the lymph nodes and subsequent retention in tissues, as T cells expressing high amounts of S1P1 are retained in the lymph nodes and also easily exit from tissues due to the higher concentrations of its ligand, sphingosine-1 phosphate (S1P) in the lymph nodes and blood. *S1PR1* is a target gene of KLF2, a transcription factor; its downregulation has been shown to result in reduced *S1PR1* expression, and both of these genes together play an important role in the establishment and retention of T_{RM} cells in tissues[29]. GSEA also revealed that TIL^{hi} tumors exhibit low expression of genes that are typically downregulated in a core set of T_{RM} signature genes, such as *SIPR5*, *STK38*, *FAM65B*[23,25] (**Fig. 4g**). Pathway analysis of the genes enriched in TIL^{hi} tumors revealed a significant overrepresentation of genes involved in the canonical interferon (IFN) pathway (**Supplementary Fig. 5c**), which was also predicted to be an upstream regulator by ingenuity pathway analysis (**Fig. 4h**). Because IFN-γ produced by T_{RM} cells has been shown to recruit circulating T cells to potentiate robust immune responses in tissues[30,31], we infer that the IFN response signature seen in TIL^{hi} tumors may be the result of T_{RM} activation by TAA (tumor-specific T_{RM} activity). Overall, these results demonstrate that T_{RM} cells are enriched in TIL^{hi} tumors.

CD103 density predicts survival in lung cancer

We next examined $CD8^+$ TILs from tumors enriched for T_{RM} cells ($CD103^{hi}$) for features that would support a robust anti-tumor immune response. Ingenuity pathway analysis of the genes differentially expressed in $CD103^{hi}$ *versus* $CD103^{lo}$ TILs (classified based on the expression of *ITGAE* (*CD103*) transcripts in $CD8^+$ TILs, **Supplementary Fig. 6a,b** and **Supplementary Table 7**) pointed to cell proliferation and cytotoxicity as the key activated functions (**Supplementary Table 8**). Consistent with this analysis,

several transcripts linked to cell cycle and proliferation[32] were markedly upregulated in $CD103^{hi}$ $CD8^+$ TILs (**Fig. 5a**,**b**). We confirmed by flow cytometry that $CD103^+CD8^+$ TILs express the cell proliferation marker Ki67 (**Fig. 5c**). Several transcripts linked to cytotoxic function of $CD8^+$ T cells (*IFNG, GZMA, GZMB, SEMA7A, KLRB1, CCL3, STAT1, RAB27A, IL21R, FKBP1A*[32]) were also significantly upregulated in $CD103^{hi}$ tumors (**Fig. 5d,e** and **Supplementary Table 7**). We confirmed at the protein level that $CD103^+CD8^+$ TILs express molecules linked to cytotoxicity such as granzyme B, granzyme A, perforin, CD107a and produce interferon-γ (**Fig. 5f** and **Supplementary Fig. 7a,b**), and demonstrated that $CD103^+CD8^+$ TILs were the major producers of both granzyme A and granzyme B (**Supplementary Fig. 7c**). However, we found no significant difference between matched $CD103^+CD8^+$ TILs and $CD103^-CD8^+$ TILs when the proportion of cells expressing granzyme A, granzyme B, IFN-γ and CD107a was compared, with the exception of perforin that showed reduced expression in $CD103^+CD8^+$ TILs (**Supplementary Fig. 7d**). To address the question of whether $CD8^+$ TILs from $CD103^{hi}$ tumors have greater effector potential, we compared the mean fluorescence intensity (MFI) and percentage of cells expressing these molecules in $CD103^{hi}$ tumors *versus* $CD103^{lo}$ tumors (**Fig. 5f** and **Supplementary Fig. 7e**). Notably, we found that granzyme B was expressed at significantly higher amounts in $CD8^+$ TILs from $CD103^{hi}$ tumors compared to $CD103^{lo}$ tumors (**Fig. 5f**). These results suggest that tumors rich in T_{RM} cells ($CD103^{hi}$) harbor $CD8^+$ T cells that actively proliferate in the tumor milieu and display enhanced cytotoxic molecule production, all hallmarks of robust anti-tumor immunity.

Based on this finding, we hypothesized that a high density of CD103 in tumors (T_{RM}-enriched tumors) may also confer a survival advantage beyond that previously found to be associated with $CD8^+$ TIL density[6,7]. In an independent, large cohort of predominantly early stage lung cancer patients (*n* = 689; 83% Stage I to IIIA,

Supplementary Table 9) followed up from 2007 to 2016, we assessed retrospectively the survival outcome for patients whose tumors were classified based on the density of cells expressing CD8α or CD103 (**Supplementary Table 9**). A higher density of $CD8^+$ TILs was associated with a 28% reduction in mortality (Cox proportional hazards model, P = 0.077, **Fig. 5g**). Importantly, lung cancer patients with $CD103^{hi}$ tumors had significantly reduced mortality (34% reduced risk of mortality, Cox proportional hazards model, P = 0.045) compared to those with $CD103^{lo}$ tumors (**Fig. 5h**). This finding was also observed in the TCGA data set for lung cancer (**Supplementary Fig. 7f**). To better understand the dependence of CD103 and CD8 density in tumors, we determined the status of CD103 density ($CD103^{hi}$, $CD103^{int}$, $CD103^{lo}$) in tumors pre-classified based on CD8 density. As expected, the proportion of $CD103^{hi}$ tumors was higher in $CD8^{hi}$ compared to $CD8^{lo}$ tumors; however, there was some discordance as tumors with $CD103^{lo}$ or $CD103^{int}$ status were also observed in $CD8^{hi}$ tumors (**Fig. 5i**). Notably, even in the subgroup of lung cancer patients with high $CD8^+$ TIL density ($CD8^{hi}$ tumors), patients with higher CD103 density had significantly reduced mortality (60% reduced risk of mortality, Cox proportional hazards model, P = 0.043) and survived significantly longer compared to patients with $CD103^{lo}$ tumors (**Fig. 5i**). Our results suggest that patients with a robust intra-tumoral T_{RM} response have better long-term survival outcomes, and this effect is over and above that conferred by density of $CD8^+$ TILs.

New molecules linked to tumor immune response

Transcripts for molecules that have been shown to be effective immunotherapy targets, such as *PDCD1*, *TIM3* and *LAG3*, were among the most enriched in tumors with $CD8^{hi}$ and $CD103^{hi}$ TIL status, which were both independently linked to better anti-tumor immunity and survival outcome. Therefore, we reasoned that other molecules in the list of genes upregulated in tumors with $CD8^{hi}$ and $CD103^{hi}$ TIL status may also play an

important functional role in modulating the magnitude and specificity of anti-tumor immune responses (**Fig. 6a** and **Supplementary Table 7**). Some examples include CD39, encoded by ectonucleoside triphosphate diphosphohydrolase 1 (*ENTPD1*), a cell-surface ectonucleotidase that dephosphorylates ATP to AMP (**Fig. 6b**,**c**). We found that CD39 protein expression was strikingly increased in $CD103^{+}CD8^{+}$ TILs when compared to $CD103^{-}CD8^{+}$ TILs (**Fig. 6d**). High concentrations of ATP observed in the tumor microenvironment can be toxic to cells via purinergic receptor (P2RX7) signaling[33,34]. Given that $CD8^{+}$ TILs (from both $CD103^{hi}$ and $CD103^{lo}$) exhibited similar expression of P2RX7 receptor transcripts (**Fig. 6c**), we speculate that the higher abundance of CD39 are likely to preferentially protect T_{RM} cells ($CD103^{+}CD8^{+}$ TILs) from ATP-induced cell death. Note, however, that adenosine produced by CD39 may also be suppressive for NKT cells, NK cells and probably $CD8^{+}$ T cell function[35,36]. CD38 is another ectonucleotidase and type II trans-membrane glycoprotein with various functions including regulation of adenosine signaling, adhesion, and transduction of activation and proliferation signals[37,38]. CD38 protein expression was also increased in $CD103^{+}CD8^{+}$ TILs when compared to $CD103^{-}CD8^{+}$ TILs (**Fig. 6d**). Given that purinergic receptors can be therapeutically targeted, it will be pertinent to test how CD39 and CD38 modulate ATP and purinergic signaling pathways to influence the development and function of anti-tumor T_{RM} cells ($CD103^{+}CD8^{+}$ TILs).

Among transcription factors, several transcripts in the Notch signaling pathway (*NOTCH*, *RBPJ*, *DTX2*, *UBC*, *UBB*) were enriched in $CD103^{hi}$ TILs, suggesting an important role for this pathway in boosting T_{RM} responses in lung cancer, a speculation supported by a recent report showing that the Notch pathway supports T_{RM} cell development in the lungs[39] (**Fig. 6b**,**c**). Transcripts of two transcription factors (BATF and NAB1) that are potentially linked to $CD4^{+}$ T cell–mediated help of $CD8^{+}$ T cells were enriched in $CD103^{hi}$ $CD8^{+}$ TILs (**Fig. 6b**,**c**).

Other examples of transcripts upregulated in CD103^{hi} CD8^{+} TILs include *KIR2DL4*, which encodes a killer cell immunoglobulin-like receptor KIR2DL4 with activating and inhibitory functions[40]; its protein expression was confirmed in CD103^{+}CD8^{+} TILs (**Fig. 6d**). HLA-G, a non-classical MHC class I molecule, has been shown to engage KIR2DL4 and increase cytokine and chemokine production by NK cells[41]. Though the expression of HLA-G is highly restricted, several reports have shown its increased expression in tumor tissue, especially in lung cancer[42], so we speculate that HLA-G expressed in tumors may also convey activation signals *via* the KIR2DL4 receptor to CTLs. *SIRPG* encodes for SIRPG, a member of the immunoglobulin superfamily of signal-regulatory proteins (SIRPs) that interact with the ubiquitously expressed CD47 molecule[43]. Interestingly, SIRPG is the only member of the SIPR family that is expressed on T cells, and its interaction with CD47 expressed on APCs was shown to enhance T cell proliferation and IFN-γ production[44]. Based on the increased expression of *SIRPG* transcripts in CD103^{hi}CD8^{+} TILs (**Fig. 6b**,**c**), we speculate that SIRPG may also serve as an important co-stimulatory molecule and its function could be exploited to enhance the anti-tumor function of CTLs. Several candidate molecules described here have not been fully assessed for their potential as immunotherapeutic targets in cancer; the importance of these molecules in boosting anti-tumor immune responses should be verified in further functional studies.

DISCUSSION

We have taken an unbiased discovery-based approach to identify transcripts that are enriched in CD8^{+} TILs and those that are linked to robust anti-tumor immune responses and good outcomes. Prior transcriptional studies of anti-tumor CD8^{+} T cells from patients with cancer have been largely restricted to analysis of whole tumor tissue or CD8^{+} T cells in peripheral blood or metastatic sites[8-11]. Our study is the first and largest survey of over 100 transcriptomes from purified CD8^{+} TILs and N-TILs, derived

from tumor tissue and the best available control tissue, adjacent uninvolved lung. Bioinformatic analysis of these data sets revealed a core $CD8^+$ TIL transcriptional profile comprising of ~1400 genes that is shared across different tumor subtypes and is distinct from N-TILs. This profile suggests extensive molecular reprogramming within the tumor microenvironment and the enrichment of presumably TAA-specific cells that are actively proliferating following TCR engagement and co-stimulation, all hallmarks of effective anti-tumor immunity.

Despite the use of purified $CD8^+$ TIL populations for the analyses, we observed significant heterogeneity in the expression of cell cycle, TCR activation, co-stimulation and inhibitory genes across patients. This underlying molecular heterogeneity in anti-tumor CTL response may partly explain the variability in clinical responses to currently available immune checkpoint blockers. We propose that baseline transcriptional profiling of purified tumor-infiltrating CTLs may enable rationale selection of immunotherapies. Our strategy of purifying relevant immune cell populations from relatively small tumor samples and performing micro-scaled RNA-Seq assays to generate high-resolution genome-wide data can be readily applied to any accessible tumor type. This approach can thus be used to develop biomarkers of response to immunotherapies and to discover novel targets for immunotherapy.

Another unique aspect of our study is the evaluation of $CD8^+$ TIL transcriptomes in relation to TIL density (a feature linked to outcome). This analysis revealed a number of features linked to robust anti-tumor immune responses, *i.e.* TIL density; the most striking of these was tissue residence. $CD8^+$ TILs enriched for T_{RM} cells ($CD103^{hi}$) had features of enhanced cytotoxicity and proliferation, suggesting that patients whose tumors have a high density of T_{RM} cell markers, such as CD103, have a more robust anti-tumor immune response, and that this feature in the tumor may independently influence clinical outcomes. In a large independent cohort of lung cancer patients, we

showed that a higher density of CD103 predicts for better survival outcomes. Most importantly, we confirmed that this effect is over and above that conferred by density of $CD8^+$ TILs; this novel finding is biologically relevant and not addressed by prior work[45-47]. Thus, our study not only reveals the close link between TIL density, T_{RM} features and enhanced survival, but also sheds light on the global molecular features that endow $CD8^+$ TILs from T_{RM}-rich tumors with robust anti-tumor properties. We speculate that the generation of a robust anti-tumor T_{RM} response should be an important goal of vaccination approaches targeting neo-antigens or shared tumor antigens.

Since lung cancer patients with a high density of $CD8^+$ or $CD103^+$ TILs had better survival outcomes, our comparisons of the transcriptional profile of $CD8^+$ TILs from tumors with either a high or a low density of CD8 or CD103 are likely to highlight features linked to the *de novo* generation of robust anti-tumor immunity. The list of differentially expressed transcripts included molecules such as PD-1, TIM-3, CTLA-4, and LAG-3, CD27, CD8 and OX40 that are effective cancer immunotherapy targets in humans or in model organisms. Other molecules in this gene list might also have an important role in modulating the magnitude and specificity of anti-tumor immune response. For example, several promising molecules that we identified such as CD38, CD39, BATF, NAB1, KIR2DL4, SIPRG and Notch signaling players deserve further investigation as immunotherapeutic targets in cancer. BATF has recently been shown to regulate $CD8^+$ T cell metabolism and survival, and to reduce the inhibited phenotype of $CD8^+$ T cells[48,49]. In a model of LCMV infection, the expression of BATF in $CD8^+$ T cells, induced by CD4-derived IL-21, was essential for the maintenance of CTL effector response, and overexpression of BATF rescued the effector function of unhelped $CD8^+$ T cells[49]. NAB1, encoded by *NAB1*, is a transcription factor whose mouse homologue NAB2 is induced in $CD8^+$ T cells that have received $CD4^+$ T cell help, and is required to prevent activation-induced cell death of those "helped" $CD8^+$ T cells[50]. Thus we

hypothesize that NAB1, which has high sequence homology to NAB2, may also play a similar role in preventing apoptosis of tumor-infiltrating CTLs, and that its increased expression may identify tumors where $CD8^+$ TILs have received $CD4^+$ T cell help. Further functional experiments will be required to verify the role of these novel molecules.

Our study reveals the transcriptional program of $CD8^+$ TILs at the tumor site and uncovers the inter-patient heterogeneity that presumably underlies the variability in clinical response to checkpoint blockade. It provides key insights into the molecular mechanisms governing robust anti-tumor CTL responses and lends support to the notion that tumor vaccines should be designed to enable $CD8^+$ T_{RM} generation for durable immunity. The ability to perform micro-scaled RNA-seq analyses on purified $CD8^+$ TILs from patients' tumors allowed us to identify gene expression programs that may inform personalized immunotherapeutic treatment strategies, thereby providing a useful tool for translational application.

ACCESSION CODES

Sequencing data is deposited into the Gene Expression Omnibus (GSE90730).

ACKNOWLEDGMENTS

We thank M. Chamberlain, K. Amer, B. Johnson for assistance with recruitment of study subjects and processing of samples, and A. Easton for help with TIL scoring. Funding support for sample collection was provided by the Wessex Clinical Research Network and the National Institute of Health Research, UK. We thank Z. Fu and J. Greenbaum at the LJI bioinformatics core for help with processing and analysis of sequencing data. We also thank H Cheroutre, M Kronenberg and J Moore for reviewing the manuscript and providing valuable insights. This work was supported by William K. Bowes Jr Foundation

(P.V.), Faculty of Medicine, University of Southampton (P.V., T.S.-E, C.H.O) and Cancer Research UK (C.H.O).

AUTHOR CONTRIBUTIONS

A.-P.G., P.S.F., T.S.-E., C.H.O. and P.V. conceived the work, designed, performed and analyzed experiments; O.W., J.C., E.M.G.-M., T.M. performed the cell isolations and immunohistochemistry data analysis under the supervision of G.J.T. and C.H.O.; S.C., E.W., A.A. and E.V.K. assisted in patient recruitment, consenting and sample collection; A.-P.G., D.S., D.S-C. performed micro-scaled RNA-Seq experiments and analysis under the supervision of G.S., C.H.O. and P.V.; A.-P.G., performed data analysis and wrote the manuscript under the supervision of C.H.O. and P.V.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Figure Legends

Figure 1. Core CD8$^+$ TIL transcriptional profile. (**a**) RNA-Seq analysis showing row-wise *z*-scores of normalized read counts for each differentially expressed gene (rows) obtained by pairwise comparison of lung CD8$^+$ N-TILs (n = 32) *versus* NSCLC CD8$^+$ TILs (n = 36) (DESeq2 analysis, Benjamini-Hochberg adjusted $P < 0.05$ and 1.5 fold change) for various cancer subtypes, highlighted at the top: NSCLC adenocarcinoma (red) and squamous carcinoma (pink), HPV-negative (light blue) and HPV-positive (dark

blue) HNSCC; *n* = 41 for HNSCC $CD8^+$ TILs. RNA-Seq data from each independent sample is shown in columns. (**b**) Principal component analysis of each $CD8^+$ T cell core transcriptome (symbol) from N-TILs and TILs of indicated cancer subtypes. Percentage of variance in each principal component (PC) is shown in parentheses next to the PC. (**c**) RNA-Seq analysis of the expression of exhaustion-associated genes, represented as University of California Santa Cruz (UCSC) genome browser tracks (RPKM, reads per kilobase per million mapped) (top) and dot plots (log_2 normalized counts; error bars are mean ± SEM) (below); each dot represents data from an independent sample. (**d**) GSEA enrichment plots for the indicated gene sets in the transcriptome of NSCLC $CD8^+$ TILs *versus* N-TILs (Online Methods). The top portion of the plot shows the running enrichment score (RES) for the gene set as the analysis walks down the ranked list of genes and reflects the degree to which the gene set is overrepresented at the top or bottom of the ranked list of genes. The middle portion of the plot shows where the members of the gene set (indicated as blue lines) appear in the ranked list of genes. The bottom portion of the plot shows the value of the ranking metric.

Figure 2. Pathways enriched in $CD8^+$ TILs. (**a**) Ingenuity Pathway Analysis (IPA) of canonical pathways enriched in $CD8^+$ TILs, *P* values calculated by Fisher Exact Test (Online Methods); HBCS, hereditary breast cancer signaling. (**b**) Overlap of several cell cycle and proliferation pathways. Number in parentheses indicates the number of genes in the pathway. Number on the connecting lines indicates number of shared genes between pathways. (**c**) RNA-Seq analysis of the expression of the indicated transcripts, represented as University of California Santa Cruz (UCSC) genome browser tracks (RPKM, reads per kilobase per million mapped) (top) and bar graphs (log_2 normalized counts; error bars are mean ± SEM) (below); each dot represents data from an

independent sample. (**d**) IPA shows the differentially expressed genes in $CD8^+$ TILs (highlighted in yellow) present in the canonical 4-1BB and CD27 signaling pathways in lymphocytes. (**e**) Contour plots show surface expression of 4-1BB and CD8 in live and singlet-gated $CD45^+CD3^+$ T cells obtained from matched PBMC, lung N-TILs and NSCLC TILs from the same patient. (**f**) Graph shows the average numbers of frequently occurring clonotypes in $CD8^+$ N-TILs *versus* NSCLC $CD8^+$ TILs, derived from analysis of TCR beta chain sequences from RNA-Seq data set (error bars are mean ± SEM); * $P < 0.05$ (unpaired Student's two-tailed *t*-test).

Figure 3. Heterogeneity in immunotherapy target molecules. (**a**) RNA-Seq analysis of NSCLC $CD8^+$ TILs showing the row-wise normalized expression of indicated transcripts (rows) for each patient (column). Panel above heat map shows $CD8^+$ TIL density (top) and tumor stage (bottom) of each patient. (**b**) PCA (center plot) of $CD8^+$ TILs from NSCLC patients categorized based on $CD8^+$ TIL density in tumors as TIL^{lo} (blue), TIL^{int} (green) and TIL^{hi} (magenta) with surrounding plots showing expression of the indicated transcripts as in (**a**); each dot represents data from a patient. (**c**) Correlation between *PDCD1* and *4-1BB* transcript expression (log_2 normalized counts) (left) in NSCLC $CD8^+$ TIL; correlation of *PDCD1* (center) and *4-1BB* (right) transcript expression with the number of tumor-infiltrating $CD8^+$ cells quantified by immunohistochemistry. (**d**) Correlation between *PDCD1* and *4-1BB* transcript expression (left); correlation of *PDCD1* (center) and *4-1BB* (right) transcript expression with *CD8A* transcripts in the TCGA lung cancer RNA-Seq data set ($n = 1013$), extreme outliers are not shown. *r* value indicates the Spearman correlation coefficient. (**e**) RNA-Seq analysis of the expression of the indicated transcripts, represented as University of California Santa Cruz (UCSC) genome browser tracks (RPKM, reads per kilobase per million

mapped) (top) and bar graphs (log_2 normalized counts; error bars are mean ± SEM) (below); each dot represents data from an independent sample.

Figure 4. Tissue residency features in TIL^{hi} tumors. (**a**) RNA-Seq analysis of NSCLC CD8$^+$ TILs showing row-wise *z*-scores of normalized read counts for each differentially expressed gene (rows) obtained by comparison of TIL^{lo} *versus* TIL^{hi} tumors (DESeq2 analysis, Benjamini-Hochberg adjusted $P < 0.05$). RNA-Seq data from each independent sample is shown in columns. (**b**) RNA-Seq analysis of the expression of the indicated transcripts, represented as University of California Santa Cruz (UCSC) genome browser tracks (RPKM, reads per kilobase per million mapped) (top) and bar graphs (log_2 normalized counts; error bars are mean ± SEM) (below); each dot represents data from an independent sample. (**c**) Correlation of *ITGAE* (*CD103*) transcript expression (log_2 normalized counts) in NSCLC CD8$^+$ TILs with the number of tumor-infiltrating CD8$^+$ cells quantified by immunohistochemistry (left); correlation of *ITGAE* (*CD103*) transcript expression with *CD8A* transcripts in the TCGA lung cancer RNA-Seq data set ($n = 1013$) (right), extreme outliers are not shown. *r* value indicates the Spearman correlation coefficient. (**d**) Immunohistochemistry staining for CD8α (left), PD-1 (center) and CD103 (right) in TIL^{lo} (top panel) and TIL^{hi} NSCLC tumors (bottom panel); scale bar is 100 µm. (**e**) Contour plots show surface expression of CD103 and CD8 in live and singlet-gated CD45$^+$CD3$^+$ T cells obtained from matched PBMC, lung N-TILs and NSCLC TILs from the same patient (top); plots below show co-expression of PD-1 with CD103 and 4-1BB with CD103 in live and singlet-gated CD45$^+$CD3$^+$CD8$^+$ T cells. (**f**) Top panel contour plots show co-expression of CD69 and CD49a with CD103 in live and singlet-gated CD45$^+$CD3$^+$CD8$^+$ T cells, and overlay of CD103$^+$CD8$^+$ TILs (red color) with CD103$^-$CD8$^+$ TILs (blue color); bottom panel shows expression of KLRG1, CD62L and CCR7 in live and singlet-gated CD45$^+$CD3$^+$CD8$^+$ T cells. (**g**) GSEA plots for T_{RM} signature genes (left,

upregulated genes; right, downregulated genes) in the transcriptome of $CD8^+$ TILs from NSCLC TIL^{hi} tumors. (**h**) IPA analysis shows transcripts upregulated in NSCLC TIL^{hi} tumors that are regulated by interferon-γ.

Figure 5. CD103 density predicts survival in lung cancer. (**a**) RNA-Seq analysis of NSCLC $CD8^+$ TILs showing row-wise *z*-scores of normalized read counts for cell cycle and proliferation-related gene transcripts (rows) (Online Methods), expressed by $CD8^+$ TILs from NSCLC *ITGAE*^{lo} *versus ITGAE*^{hi} (referred as $CD103^{lo}$ *versus* $CD103^{hi}$) tumors. RNA-Seq data from each independent sample is shown in columns. (**b**) RNA-Seq analysis of the expression of the indicated transcripts, represented as University of California Santa Cruz (UCSC) genome browser tracks (RPKM, reads per kilobase per million mapped) (top) and bar graphs (log_2 normalized counts; error bars are mean ± SEM) (below); each dot represents data from an independent sample. (**c**) Contour plots show co-expression of Ki67 with CD103 in live and singlet-gated $CD45^+CD3^+CD8^+$ T cells obtained from matched PBMC, lung N-TILs and NSCLC TILs from the same patient. (**d**) PCA (center plot) of $CD8^+$ TILs from NSCLC patients categorized based on *ITGAE* transcript abundance as $CD103^{lo}$ (blue), $CD103^{int}$ (green) and $CD103^{hi}$ (magenta) (center plot) with surrounding plots showing expression of the indicated cytotoxicity-related gene transcripts. (**e**) Expression of the *GZMB*, *GZMA and IFNG* transcripts is represented as bar graphs (log_2 normalized counts; error bars are mean ± SEM); each dot represents data from an independent sample. (**f**) Bar graph shows mean expression (error bars are mean ± SEM of geometric MFI values) of granzyme B in $CD8^+$ TILs in $CD103^{lo}$ (n = 5) *versus* $CD103^{hi}$ (n = 7) tumors; * P = 0.0025 (Mann-Whitney test); each dot represents data from a patient. Contour plots show expression of granzyme B, granzyme A, perforin, CD107a (LAMP-1) and interferon-γ in live and singlet-gated

$CD45^+CD3^+CD8^+$ T cells obtained from NSCLC TILs. (**g**,**h**) Kaplan–Meier curves for lung cancer mortality according to density of $CD8^+$ and $CD103^+$ cells in tumors (log-rank test, $P = 0.086$ and $P = 0.043$, respectively; $n = 689$) (**i**) CD103 density ($CD103^{hi}$, $CD103^{int}$, $CD103^{lo}$) in tumors pre-classified based on CD8 density (left); Kaplan–Meier curves for lung cancer mortality in $CD8^{hi}$ tumors sub-classified according to density of CD103 (right) (log-rank test, $P = 0.036$).

Figure 6. New molecules linked to tumor immune response. (**a**) RNA-Seq analysis of NSCLC $CD8^+$ TILs showing row-wise *z*-scores of normalized read counts for each differentially expressed gene (rows) obtained by comparison of $CD103^{lo}$ *versus* $CD103^{hi}$ tumors (DESeq2 analysis, Benjamini-Hochberg adjusted $P < 0.05$). RNA-Seq data from each independent sample is shown in columns. (**b**) RNA-Seq analysis of the expression of the indicated transcripts (axes) (*z*-scores of normalized counts) in NSCLC $CD8^+$ TILs, presented as a scatterplots. Each dot shown in magenta and blue is data from a patient with $CD103^{hi}$ and $CD103^{lo}$ tumor, respectively. (**c**) Expression of the same transcripts as in **b.** is represented as bar graphs ($\log_2$ normalized counts; error bars are mean ± SEM); each dot represents data from a patient. (**d**) Contour plots show expression of KIR2DL4, CD38 and CD39 in live and singlet-gated $CD45^+CD3^+CD8^+$ T cells obtained from NSCLC TILs. Graphs show percentage of CD38 and CD39 in the indicated populations ($n = 11$); * $P = 0.0006$ and $P < 0.0001$, respectively (paired Student's two-tailed *t*-test); each dot represents data from a patient.

ONLINE METHODS

Patient characteristics and sample processing. The Southampton and South West Hampshire Research Ethics Committee and the Ethics Committee of La Jolla Institute approved the study, and written informed consent was obtained from all subjects. Newly diagnosed, untreated patients with NSCLC and HNSCC (**Supplementary Table 1**) referred to Southampton University Hospitals NHS Foundation Trust and Poole Hospital NHS Foundation trust, UK between 2014 and 2016 were prospectively recruited. Freshly resected tumor tissue and matched adjacent non-tumor lung tissue (in the case of

patients with NSCLC) was obtained following surgical resection. T cells were isolated from tumor (TILs) or adjacent uninvolved lung (N-TILs) using a combination of mechanical and enzymatic dissociation. Briefly, tumor or lung tissue was cut into small fragments and incubated at 37 °C for 15 min in an orbital shaker with 2 ml RPMI 1640 medium (Fisher Scientific) containing 20 units/ml Liberase DL (Roche) and 800 units/ml DNase I (Sigma-Aldrich). Dispersed cells were then passed through a 70 μm filter, centrifuged and re-suspended in MACS buffer (phosphate-buffered saline containing 2 mM EDTA and 0.5% bovine serum albumin) for flow sorting or flow cytometric analysis. For isolating and phenotyping $CD8^+$ T cells from tumor or lung tissue, dispersed cells were first incubated with FcR block (Miltenyi Biotec), then stained with a mixture of fluorescently conjugated antibodies: anti-CD45-FITC (HI30; BioLegend), anti-CD4-PE (RPA-T4; BD Biosciences), anti-CD3-PE-Cy7 (SK7; BioLegend), anti-CD8α-PerCP-Cy5.5 (cSK1; BD Biosciences), anti-HLA-DR-APC (L243; BD Biosciences), anti-CD14-APC-H7 (MφP9; BD Biosciences), anti-CD19-PerCP-Cy5.5 (clone HIB19; BioLegend) and anti-CD20-PerCP-Cy5.5 (clone 2H7; BioLegend). Stained samples were analyzed using BD FACSAria™ (BD Biosciences) and FlowJo software (Treestar), and $CD8^+$ T cells were sorted into ice-cold TRIzol LS reagent (Ambion)[51,52]**.** Phenotypic analysis of $CD8^+$ TILs for T_{RM} markers was performed by staining with anti-CD69-BV605 (FN50; BioLegend), anti-CD49a-PE (TS2/7; BioLegend), anti-KLRG1-APC (SA231A2; BioLegend), anti-CD62L-BV510 (DREG-56; BioLegend), anti-CCR7-AF700 (TS2/7; BioLegend). Flow cytometric analysis of $CD8^+CD103^+$ T cells and intra-cellular assessment of Ki67 were carried out with the following antibodies: anti-CD45-FITC (HI30; BioLegend), anti-Ki67-PE (Ki67; BioLegend), anti-CD3-APC-Cy7 (SK7; BioLegend), anti-CD8α-PerCP-Cy5.5 (SK1; BD Biosciences), anti-CD103-APC (Ber-ACT8; BioLegend), anti-PD-1-PE-Cy7 (eBioJ105; eBioscience), anti-4-1BB-Pacific blue

(4B4-1; BioLegend). The True-Nuclear™ Transcription Factor Buffer set (BioLegend) was used for the intracellular staining of Ki67. Flow cytometric analysis of novel molecules and intracellular assessment of cytotoxic molecules were performed using the following antibodies: anti-Granzyme A-APC (CB9; BioLegend), anti-Granzyme B-PE (REA226; Miltenyi Biotec), anti-Perforin-PE or –BV421 (B-D48; BioLegend), anti-KIR2DL4-PE (mAb33; BD BioLegend), anti-CD38-APC-Cy7 (HB-7; BioLegend), anti-CD39-PE (A1; BioLegend). For cytokine and CD107a assay, $CD8^+$ TILs were stimulated *ex vivo* with 20 nM PMA and 1 μM ionomycin for 4 h and 5 μg/ml brefeldin was added during the final 2 h of stimulation. Anti-CD107a-PE (H4A3; BioLegend) was added to the PMA and ionomycin stimulation mix for the final 2 h. Intra-cellular assessment of interferon-γ was performed using anti-IFNG-BV-421 (4S.B3; BioLegend) at the end of stimulation. Assays were performed in at least 6 patients and representative plots shown. Stained samples were analyzed using the BD FACSCanto II (BD Biosciences). Dead cells were excluded using the LIVE/DEAD® Fixable Aqua dead cell stain kit (Life Technologies) or DAPI.

Histology and immunohistochemistry. Immunohistochemistry (IHC) was performed on FFPE tumor sections against cytokeratin (AE1/AE3; Dako), CD8α (C8/144B; Dako), CD103 (EPR4166(2); Abcam), PD-1 (NAT105; Abcam) and granzyme B (GrB-7; Dako). Assays were performed in at least 10 patients and representative staining shown. TILs were quantified using a Zeiss AxioCam MRc5 microscope (Zeiss) and Zeiss Axiovision software (version 4.8.1.0; Zeiss). An average of 10 high-power (×400) fields across representative areas of each tumor was counted to account for intratumoral heterogeneity; these were averaged to generate an intratumoral TIL score. Tumors with an average CD8 count in the top 1/3 or bottom 1/3 percentile were classified as TIL^{hi} or

TIL^{lo}, respectively; the lowest CD8 count in the TIL^{hi} tumors was at least 2 fold greater than the highest CD8 count in the TIL^{lo} tumors (**Supplementary Fig. 4**). For overall survival analyses (**Fig. 5e**-**g**), tumor tissue microarrays from NSCLC patients were stained with anti-CD8α (C8/144B; Dako) and anti-CD103 (EPR4166(2); Abcam) antibodies and viewed under low-power magnification (×2.5 objective) to determine CD8 and CD103 density, as described previously[53].

Survival data and analysis. In an independent large cohort of predominantly early stage NSCLC patients (*n* = 689, **Supplementary Table 9**) followed up from 01/2007 to 06/2016 (minimum follow up 3.4 years) we retrospectively analyzed survival according to CD8 and CD103 TIL density. The primary endpoint was overall survival, and survival time was measured from the date of diagnosis until date of death or date last seen alive. Kaplan–Meier plots (with log-rank tests to determine significance of overall survival, *P* values shown in **Fig. 5g-i**) and unadjusted Cox proportional hazards model (to determine relative risk of death) were used to analyze the survival data, as described previously[54]. Patients were excluded from analysis if survival was < 30 days to exclude possibility of surgery-related mortality**.** Survival analysis based on the expression of *ITGAE* (*CD103*) transcripts in tumor samples from lung adenocarcinoma patients in the TCGA was derived from http://www.oncolnc.org.

RNA sequencing. Total RNA was purified using a miRNAeasy micro kit (Qiagen) and quantified as described previously[55] (on average, ~4230 $CD8^+$ T cells per sample were processed for RNA-Seq analysis). Purified total RNA was amplified following the smart-seq2 protocol[55]. cDNA was purified using AMPure XP beads (1:1.1 ratio, Beckman Coulter). From this step, 1 ng of cDNA was used to prepare a standard Nextera XT

sequencing library (Nextera XT DNA sample preparation kit and index kit, Illumina). Samples were sequenced using HiSeq2500 (Illumina) to obtain 50-bp single-end reads. Quality control steps were included to determine total RNA quality and quantity, optimal number of PCR pre-amplification cycles, and cDNA fragment size[56]. Samples that failed quality control were eliminated from further downstream steps.

RNA-Seq analysis. RNA-Seq data was mapped against the hg19 reference using TopHat[57] (v1.4.1., --library-type fr-secondstrand -C) and the RefSeq gene annotation downloaded from the UCSC Genome Bioinformatics site. Sequencing read coverage per gene was counted using HTSeq-count (-m union -s yes -t exon -i gene_id, http://www-huber.embl.de/users/anders/HTSeq/). To identify genes differentially expressed between patient groups, we performed negative binomial tests for paired and unpaired comparisons by employing the Bioconductor package DESeq2 disabling the default options for independent filtering and Cooks cutoff[58]. We considered genes differentially expressed between any comparison when the DESeq2 analysis resulted in a Benjamini-Hochberg–adjusted *P* value < 0.05. The Qlucore Omics Explorer 3.2 software package was used for visualization and representation (heat maps, principal component analysis) of RNA-Seq data[52]. Unsupervised hierarchical clustering of samples based on the expression of genes (*n* = 1000) with the highest variance, which accounted for 20% of the total variance, was performed using DESeq package functions and custom scripts on R (**Supplementary Fig. 1c**). TCR sequences were retrieved from $CD8^+$ T cell RNA-Seq data sets and the frequencies of TCRβ chain clonotypes were determined using default parameters of the MiXCR package[59] (**Supplementary Table 5**). The CD103 status of TILs was determined based on the transcript abundance of *ITGAE* (*CD103*) in $CD8^+$ TILs. Tumors with $CD8^+$ TILs expression of *ITGAE* transcripts in the top 1/3 or bottom

1/3 percentile were classified as $CD103^{hi}$ or $CD103^{lo}$, respectively (**Supplementary Fig. 6**).

Knowledge-based network generation and pathway analysis. The biological relevance of differentially expressed genes identified by DESeq2 analysis was further investigated using the Ingenuity Pathways Analysis platform. The enrichment of canonical pathways (pre-defined, well-described metabolic and signaling pathways curated from literature reviews) amongst differentially expressed genes was assessed, with significance determined by right-tailed Fisher's exact test, $P < 0.05$. For network analysis, differentially expressed genes were progressively linked together based on a measure of their interconnection, which is derived from previously characterized functional interactions.

Gene Set Enrichment Analysis (GSEA). The Qlucore Omics Explorer 3.2 software package was used for GSEA analysis. GSEA was used to further assess whether specific biological pathways or signatures were significantly enriched between two groups. GSEA determines whether an *a priori* defined 'set' of genes (such as a signature) show statistically significant cumulative changes in gene expression between phenotypic subgroups[60]. In brief, all genes are ranked based on their differential expression between two groups. Next, a running enrichment score (RES) is calculated for a given gene set based on how often its members appear at the top or bottom of the ranked differential list. 1000 random permutations of the phenotypic subgroups are used to establish a null distribution of RES against which a normalized running enrichment score (NES) and FDR-corrected q values are calculated using Kolmogorov-Smirnov statistic. GSEA was run with a focused group of gene signatures, namely exhaustion[22], lung cancer associated T cell signature[15], anergy[61], senescence[62], tissue residency[25].

These gene signatures (**Fig. 1d**, **Fig. 4g** and **Supplementary Table 10**) were selected to test the null hypothesis that different $CD8^+$ T cell phenotypes were not significantly enriched in $CD8^+$ T cell groups.

Statistical analysis. Comparison between two groups was assessed with two-tailed unpaired or paired Student's *t*-test (**Fig. 2f, Fig. 6d, Supplementary Fig. 5b and Supplementary Fig. 7c,d**) or Mann-Whitney test (**Fig. 5f** and **Supplementary Fig. 7e**) or Kolmogorov-Smirnov test (**Supplementary Fig. 3c**) using GraphPad Prism 6. Spearman correlation coefficient (*r* value) was calculated to assess significance of correlations between the expression of any two transcripts of interest.

Data availability. Sequencing data is deposited into the Gene Expression Omnibus (GSE90730). Source data are attached as supplementary files.

References

1. Pardoll, D.M. The blockade of immune checkpoints in cancer immunotherapy. *Nature reviews. Cancer* **12**, 252-264 (2012).

2. Drake, C.G., Lipson, E.J. & Brahmer, J.R. Breathing new life into immunotherapy: review of melanoma, lung and kidney cancer. *Nature reviews. Clinical oncology* **11**, 24-37 (2014).

3. Garon, E.B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *The New England journal of medicine* **372**, 2018-2028 (2015).

4. Sharma, P. & Allison, J.P. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* **161**, 205-214 (2015).

5. Tumeh, P.C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568-571 (2014).

6. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960-1964 (2006).

7. Yang, Y. Cancer immunotherapy: harnessing the immune system to battle cancer. *The Journal of clinical investigation* **125**, 3335-3337 (2015).

8. Baitsch, L. *et al.* Exhaustion of tumor-specific CD8(+) T cells in metastases from melanoma patients. *The Journal of clinical investigation* **121**, 2350-2360 (2011).

9. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782-795 (2013).

10. Collins, N. *et al.* Transcriptional Hallmarks Of Tumor Infiltrating Lymphocyte Responses To Melanoma. *Blood* **122** (2013).

11. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).

12. Pauken, K.E. & Wherry, E.J. Overcoming T cell exhaustion in infection and cancer. *Trends in immunology* **36**, 265-276 (2015).

13. Brown, I.E., Blank, C., Kline, J., Kacha, A.K. & Gajewski, T.F. Homeostatic proliferation as an isolated variable reverses CD8+ T cell anergy and promotes tumor rejection. *Journal of immunology* **177**, 4521-4529 (2006).

14. Montes, C.L. *et al.* Tumor-induced senescent T cells with suppressor function: a potential form of tumor immune evasion. *Cancer research* **68**, 870-879 (2008).

15. Johnston, R.J. *et al.* The immunoreceptor TIGIT regulates antitumor and antiviral CD8(+) T cell effector function. *Cancer cell* **26**, 923-937 (2014).

16. Curran, M.A. *et al.* Systemic 4-1BB activation induces a novel T cell phenotype driven by high expression of Eomesodermin. *The Journal of experimental medicine* **210**, 743-755 (2013).

17. Willoughby, J.E. *et al.* Differential impact of CD27 and 4-1BB costimulation on effector and memory CD8 T cell generation following peptide immunization. *Journal of immunology* **193**, 244-251 (2014).

18. Topalian, S.L., Drake, C.G. & Pardoll, D.M. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer cell* **27**, 450-461 (2015).

19. Gros, A. *et al.* PD-1 identifies the patient-specific CD8(+) tumor-reactive repertoire infiltrating human tumors. *The Journal of clinical investigation* **124**, 2246-2259 (2014).

20. Fourcade, J. *et al.* Upregulation of Tim-3 and PD-1 expression is associated with tumor antigen-specific CD8+ T cell dysfunction in melanoma patients. *The Journal of experimental medicine* **207**, 2175-2186 (2010).

21. Wolfl, M. *et al.* Activation-induced expression of CD137 permits detection, isolation, and expansion of the full repertoire of CD8+ T cells responding to antigen without requiring knowledge of epitope specificities. *Blood* **110**, 201-210 (2007).

22. Wherry, E.J. *et al.* Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* **27**, 670-684 (2007).

23. Mackay, L.K. *et al.* The developmental pathway for CD103(+)CD8+ tissue-resident memory T cells of skin. *Nature immunology* **14**, 1294-1301 (2013).

24. Mueller, S.N., Gebhardt, T., Carbone, F.R. & Heath, W.R. Memory T cell subsets, migration patterns, and tissue residence. *Annual review of immunology* **31**, 137-161 (2013).

25. Mackay, L.K. *et al.* Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459-463 (2016).

26. Sathaliyawala, T. *et al.* Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity* **38**, 187-197 (2013).

27. Lee, L.N. *et al.* CXCR6 is a marker for protective antigen-specific cells in the lungs after intranasal immunization against Mycobacterium tuberculosis. *Infection and immunity* **79**, 3328-3337 (2011).

28. Tse, S.W., Radtke, A.J., Espinosa, D.A., Cockburn, I.A. & Zavala, F. The chemokine receptor CXCR6 is required for the maintenance of liver memory CD8(+) T cells specific for infectious pathogens. *The Journal of infectious diseases* **210**, 1508-1516 (2014).

29. Skon, C.N. *et al.* Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8+ T cells. *Nature immunology* **14**, 1285-1293 (2013).

30. Ariotti, S. *et al.* T cell memory. Skin-resident memory CD8(+) T cells trigger a state of tissue-wide pathogen alert. *Science* **346**, 101-105 (2014).

31. Schenkel, J.M. *et al.* T cell memory. Resident memory CD8 T cells trigger protective innate and adaptive immune responses. *Science* **346**, 98-101 (2014).

32. Best, J.A. *et al.* Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. *Nature immunology* **14**, 404-412 (2013).

33. Bono, M.R., Fernandez, D., Flores-Santibanez, F., Rosemblatt, M. & Sauma, D. CD73 and CD39 ectonucleotidases in T cell differentiation: Beyond immunosuppression. *FEBS letters* **589**, 3454-3460 (2015).

34. Kuhny, M., Hochdorfer, T., Ayata, C.K., Idzko, M. & Huber, M. CD39 is a negative regulator of P2X7-mediated inflammatory cell death in mast cells. *Cell communication and signaling : CCS* **12**, 40 (2014).

35. Ohta, A. A Metabolic Immune Checkpoint: Adenosine in Tumor Microenvironment. *Frontiers in immunology* **7**, 109 (2016).

36. Gupta, P.K. *et al.* CD39 Expression Identifies Terminally Exhausted CD8+ T Cells. *PLoS pathogens* **11**, e1005177 (2015).

37. Sandoval-Montes, C. & Santos-Argumedo, L. CD38 is expressed selectively during the activation of a subset of mature T cells with reduced proliferation but improved potential to produce cytokines. *Journal of leukocyte biology* **77**, 513-521 (2005).

38. Munoz, P. *et al.* Antigen-induced clustering of surface CD38 and recruitment of intracellular CD38 to the immunologic synapse. *Blood* **111**, 3653-3664 (2008).

39. Hombrink, P. *et al.* Programs for the persistence, vigilance and control of human CD8+ lung-resident memory T cells. *Nature immunology* **17**, 1467-1478 (2016).

40. Faure, M. & Long, E.O. KIR2DL4 (CD158d), an NK cell-activating receptor with inhibitory potential. *Journal of immunology* **168**, 6208-6214 (2002).

41. Rajagopalan, S. & Long, E.O. KIR2DL4 (CD158d): An activation receptor for HLA-G. *Frontiers in immunology* **3**, 258 (2012).

42. Wisniewski, A. *et al.* Genetic polymorphisms and expression of HLA-G and its receptors, KIR2DL4 and LILRB1, in non-small cell lung cancer. *Tissue antigens* **85**, 466-475 (2015).

43. Brooke, G., Holbrook, J.D., Brown, M.H. & Barclay, A.N. Human lymphocytes interact directly with CD47 through a novel member of the signal regulatory protein (SIRP) family. *Journal of immunology* **173**, 2562-2570 (2004).

44. Piccio, L. *et al.* Adhesion of human T cells to antigen-presenting cells through SIRPbeta2-CD47 interaction costimulates T-cell proliferation. *Blood* **105**, 2421-2427 (2005).

45. Wang, Z.Q. *et al.* CD103 and Intratumoral Immune Response in Breast Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* (2016).

46. Webb, J.R., Milne, K. & Nelson, B.H. Location, location, location: CD103 demarcates intraepithelial, prognostically favorable CD8+ tumor-infiltrating lymphocytes in ovarian cancer. *Oncoimmunology* **3**, e27668 (2014).

47. Djenidi, F. *et al.* CD8+CD103+ tumor-infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. *Journal of immunology* **194**, 3475-3486 (2015).

48. Kurachi, M. *et al.* The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8+ T cells. *Nature immunology* **15**, 373-383 (2014).

49. Xin, G. *et al.* A Critical Role of IL-21-Induced BATF in Sustaining CD8-T-Cell-Mediated Chronic Viral Control. *Cell reports* **13**, 1118-1124 (2015).

50. Wolkers, M.C. *et al.* Nab2 regulates secondary CD8+ T-cell responses through control of TRAIL expression. *Blood* **119**, 798-804 (2012).

51. Vijayanand, P. *et al.* Invariant natural killer T cells in asthma and chronic obstructive pulmonary disease. *The New England journal of medicine* **356**, 1410-1422 (2007).

52. Engel, I. *et al.* Innate-like functions of natural killer T cell subsets result from highly divergent gene programs. *Nature immunology* **17**, 728-739 (2016).

53. Marsh, D. *et al.* Stromal features are predictive of disease mortality in oral cancer patients. *The Journal of pathology* **223**, 470-481 (2011).

54. Ward, M.J. *et al.* Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. *British journal of cancer* **110**, 489-500 (2014).

55. Seumois, G. *et al.* Transcriptional Profiling of Th2 Cells Identifies Pathogenic Features Associated with Asthma. *Journal of immunology* **197**, 655-664 (2016).

56. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* **9**, 171-181 (2014).

57. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).

58. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550 (2014).

59. Bolotin, D.A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nature methods* **12**, 380-381 (2015).

60. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550 (2005).

61. Safford, M. *et al.* Egr-2 and Egr-3 are negative regulators of T cell activation. *Nature immunology* **6**, 472-480 (2005).

62. Fridman, A.L. & Tainsky, M.A. Critical pathways in cellular senescence and immortalization revealed by gene expression profiling. *Oncogene* **27**, 5975-5987 (2008).