





Integration of tumor extrinsic and intrinsic features associates with immunotherapy response in non-small cell lung cancer

Denise Lau ¹✉, Sonal Khare¹, Michelle M. Stein¹, Prerna Jain¹, Yinjie Gao¹, Aicha BenTaieb ¹, Tim A. Rand¹, Ameen A. Salahudeen ¹ & Aly A. Khan ^{1,2}✉

The efficacy of immune checkpoint blockade (ICB) varies greatly among metastatic non-small cell lung cancer (NSCLC) patients. Loss of heterozygosity at the HLA-I locus (HLA-LOH) has been identified as an important immune escape mechanism. However, despite HLA-I disruptions in their tumor, many patients have durable ICB responses. Here we seek to identify HLA-I-independent features associated with ICB response in NSCLC. We use single-cell profiling to identify tumor-infiltrating, clonally expanded CD4⁺ T cells that express a canonical cytotoxic gene program and NSCLC cells with elevated HLA-II expression. We postulate cytotoxic CD4⁺ T cells mediate anti-tumor activity via HLA-II on tumor cells and augment HLA-I-dependent cytotoxic CD8⁺ T cell interactions to drive ICB response in NSCLC. We show that integrating tumor extrinsic cytotoxic gene expression with tumor mutational burden is associated with longer time to progression in a real-world cohort of 123 NSCLC patients treated with ICB regimens, including those with HLA-LOH.

¹Tempus Labs, Inc., Chicago, IL 60654, USA. ²Department of Pathology, University of Chicago, Chicago, IL 60637, USA. ✉email: publications@tempus.com; aakhan@uchicago.edu

PD-1 and PD-L1 immune checkpoint blockade (ICB) has produced clinical responses in metastatic non-small cell lung cancer (NSCLC)^{1,2}. However, clinical response rates and duration of response are highly variable, and the mechanisms of action of ICBs in patients remain poorly understood. Identifying patients who are likely to respond to a particular therapy is a cornerstone of personalized cancer medicine. Currently, tumor intrinsic features, such as tumor PD-L1 immunohistochemistry (IHC) and tumor mutational burden (TMB), are the most studied biomarkers for the prediction of response to ICB in NSCLC. While PD-L1 IHC staining has been clinically approved as a diagnostic for use in the front line to guide ICB therapy, it is complicated by the variety of assays and scoring criteria for measuring it. More notably, PD-L1 expression alone is insufficient to identify all potential responders to ICB in NSCLC^{3,4}. Similarly, TMB lacks definitive clinical utility as large, randomized clinical trials have generated contradictory results⁵⁻⁷. Tumor extrinsic features such as the presence of tumor-infiltrating T cells have demonstrated some predictive value for ICB response^{8,9} but have yet to be independently assessed in NSCLC. Thus, there is an unmet clinical need for biomarkers that complement PD-L1 staining to better predict individualized outcomes for ICB in NSCLC.

With rapid advancements in clinical DNA and RNA sequencing and improved access to electronic health records, there is an increasing volume of real-world data that can be used to supplement clinical trial studies. These linked molecular and clinical real-world datasets can be used to evaluate transcriptional signatures or other genomic markers associated with cancer treatment outcomes in routine oncology practice. In this study, we evaluated the association between tumor intrinsic features and ICB outcomes by analyzing real-world molecular profiling data and outcomes from 123 NSCLC patients. In addition, we examined whether identifying and integrating tumor extrinsic features could more accurately predict the real-world clinical responses of NSCLC patients. A lack of NSCLC tumor biopsies with paired DNA and RNA sequencing data, as well as real-world outcome data, has hampered previous research into these questions.

An essential component in the model for ICB function is the direct killing of tumor cells by cytotoxic CD8⁺ T cells in response to tumor antigen presentation on HLA class I (HLA-I) molecules. Thus, disruption of HLA-I antigen presentation in tumors via somatic alterations, such as mutations in its co-receptor beta 2 microglobulin (*B2M*) or HLA-I loss of heterozygosity (HLA-LOH), has been proposed as an important mechanism of immune escape and resistance to ICB¹⁰⁻¹². Accordingly, the rate of HLA-LOH has been reported to be as high as 40% in NSCLC patients¹³ and has been linked to worse survival on ICB^{14,15}. However, evidence has emerged that patients with disrupted tumor HLA-I presentation can still have durable responses to ICB^{9,16-19}. The mechanistic basis for ICB response in HLA-I disrupted tumors remains incompletely understood.

One potential mechanism for an HLA-I-independent immune response is through CD4⁺ T cells. Classically, effector CD4⁺ T cells help CD8⁺ T cells by licensing dendritic cells and secreting pro-inflammatory cytokines^{20,21}. However, earlier studies in animal models have shown that CD4⁺ T cells, which recognize antigen via HLA class II (HLA-II) molecules rather than HLA-I, can also directly kill tumor cells^{22,23}. Recently, studies using single-cell RNA sequencing (scRNAseq) have characterized antigen-specific cytotoxic CD4⁺ T cells in various cancers, including melanoma, breast, and colon cancer²⁴⁻²⁶. Furthermore, a population of cytotoxic CD4⁺ T cells was characterized in bladder cancer that directly killed autologous tumor cells in an HLA-II-dependent manner and was associated with improved response to ICB²⁷. However, the presence, functions, and clinical

implications of cytotoxic CD4⁺ T cells in NSCLC have not yet been characterized. Separately, multiple studies have found that HLA-II, which is typically expressed only on immune cells, can also be expressed on NSCLC tumor cells²⁸⁻³⁰, and tumor HLA-II expression has been linked to increased survival in NSCLC³¹. We postulate that cytotoxic CD4⁺ T cells exist in NSCLC and may function via HLA-II expression on tumor cells^{32,33}, which, when viewed collectively, could be a mechanistic basis for anti-tumor immune responses observed in HLA-I-disrupted tumors.

In this work, we investigate the impact of HLA-LOH on ICB efficacy in a real-world patient cohort and identify HLA-I-independent features associated with ICB response in NSCLC. We use single-cell multi-omic profiling (a combination of scRNAseq, T cell receptor [TCR] sequencing, and surface protein profiling) to characterize the tumor and T cell compartments in NSCLC tumors from 10 patients. We identify a robust population of tumor-infiltrating, clonally expanded CD4⁺ T cells expressing a canonical cytotoxic gene program. Concordantly, we find tumor cells with elevated HLA-II expression in NSCLC patients. Following this discovery, we develop an integrative model by combining tumor extrinsic cytotoxic gene expression with TMB to predict ICB outcomes in a real-world cohort of 123 NSCLC patients, including those with HLA-I-disrupted tumors. Overall, this study recommends integrating tumor extrinsic and intrinsic features to more accurately model real-world clinical responses to immunotherapy in NSCLC.

Results

HLA class I-disrupted NSCLC patients can have durable responses to checkpoint inhibitors. To determine whether patients with HLA-I-disrupted NSCLC tumors can respond to ICB, we assembled a real-world cohort of 123 patients with metastatic, non-squamous NSCLC who were molecularly profiled prior to initiating standard ICB treatment. Patients with actionable *EGFR* or *ALK* alterations were excluded from the cohort. Patient tumor samples were profiled using targeted DNA sequencing^{34,35} or whole-exome sequencing, along with whole-transcriptome RNA sequencing^{36,37}. Additionally, DNA sequencing was performed on matched normal samples obtained from blood or saliva when available (102/123). Response to therapy was evaluated using the time to progression (TTP)³⁸ with a median cohort TTP of 210 days. In this cohort, 52% of patients received ICB as a monotherapy while the rest were treated in combination with chemotherapy, and 50% of patients were treated with ICB in the first line. Other key clinical metrics are detailed in Supplementary Data 1.

The most frequently mutated driver genes were *TP53* (62%), followed by *KRAS* (50%) and *STK11* (19%) (Fig. 1a). Patients with *TP53*-mutated tumors had significantly longer TTP (HR = 0.60, $p = 0.028$, log-rank). We also assessed a number of previously described tumor intrinsic immunotherapy biomarkers, such as PD-L1 IHC, TMB, and HLA-LOH. Patients with PD-L1-high tumors ($\geq 50\%$ tumor cell staining) in this real-world cohort did not have improved TTP compared to PD-L1-negative and -low patients (HR = 0.86, $p = 0.60$, log-rank) (Fig. 1b). Tumors in this cohort had a median TMB of 5.71 mutations per megabase (mut/Mb). Patients with a high TMB (≥ 10 mut/Mb) had longer TTP than those with a low TMB (HR = 0.60, $p = 0.077$, log-rank) (Fig. 1c).

Similar to previous studies^{13,14}, we found that a significant proportion of this real-world cohort had defects in class I antigen presentation, with 26% having LOH in at least one HLA-I gene and 2% harboring a *B2M* mutation (Fig. 1a, center). Furthermore, 24% of patients were homozygous for at least one HLA class I gene. Interestingly, patients in this cohort with HLA-LOH trended toward improved TTP compared to those with intact

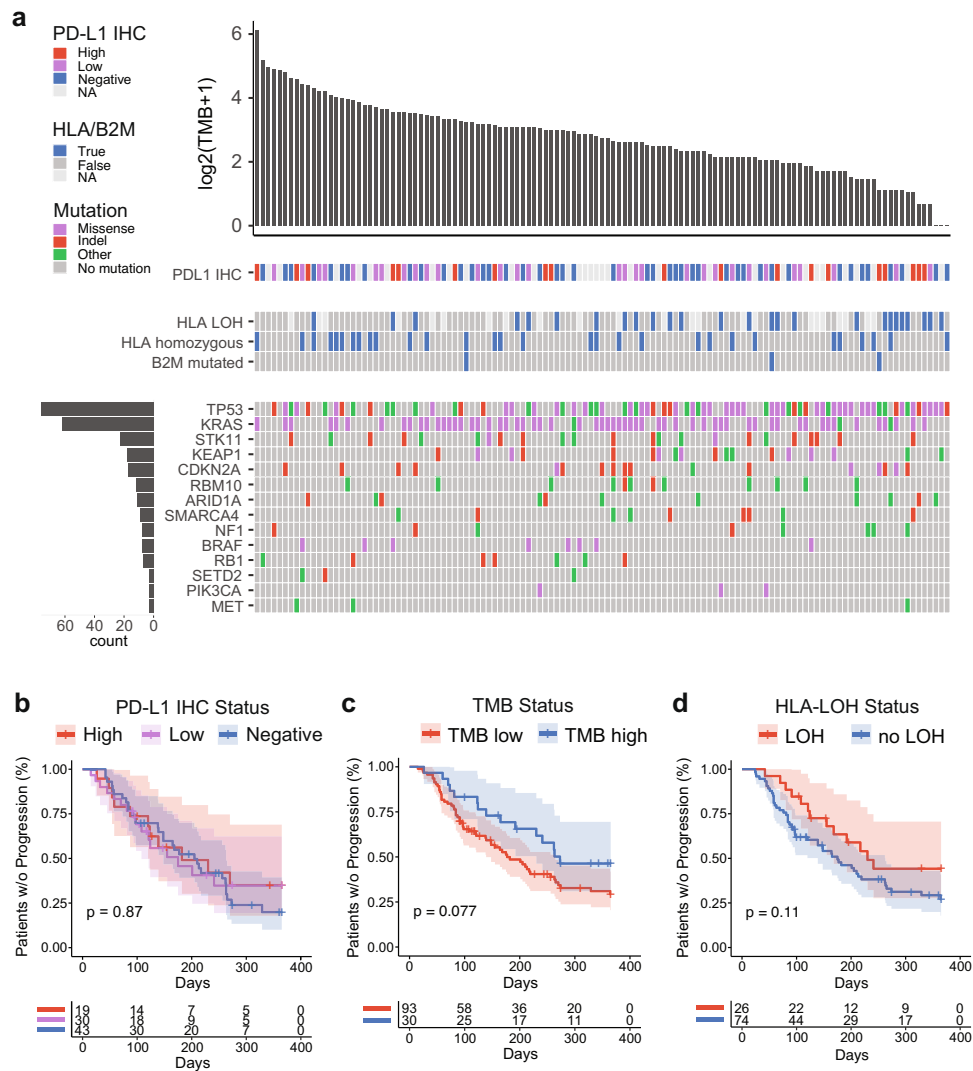


Fig. 1 Cohort clinical and genomic characteristics. **a** Mutation plot showing the distribution of TMB and frequency of driver mutations across the cohort ($n = 123$). Each column represents a patient and columns are ordered by TMB. Immunotherapy biomarker status and mutation type are denoted by color. Genes are sorted by frequency of mutation. **b** Kaplan-Meier plots showing time to progression on ICB therapy, stratified by PD-L1 IHC status ($n = 92$, $p = 0.87$, log-rank), **c** TMB status ($n = 123$, HR = 0.60, $p = 0.083$, log-rank), and **d** HLA-LOH ($n = 100$, HR = 1.64, $p = 0.11$, log-rank). Source data are provided as a Source Data file.

HLA-I (HR = 0.61, $p = 0.11$, log-rank). There was no significant association between *B2M* mutation or HLA homozygosity and TTP (*B2M*: HR = 1.24, $p = 0.77$, HLA homozygosity: HR = 0.88, $p = 0.64$) (Supplementary Fig. 1). These findings show that in a real-world setting, some patients with limitations in class I antigen presentation can still have durable clinical responses to immunotherapy.

Identification of multiple cytotoxic T cell populations in NSCLC. To investigate the existence of HLA-I-independent mechanisms of ICB response, we sought to characterize the landscape of infiltrating T cells in the tumor-immune micro-environment of NSCLC. Single-cell profiling has become an important technique in investigating the heterogeneity of the human immune system and identifying clinically important cell types that would not be possible with bulk RNA expression profiling technology^{39,40}. We performed single-cell profiling on 10 dissociated tumor samples obtained from patients with NSCLC who had never received treatment (Supplementary Data 2). Flow cytometry was used to separate samples into

CD45⁺ and CD45⁻ fractions, which were then subjected to gene expression profiling via scRNAseq on the 10X Genomics Chromium platform. In addition, single-cell TCR and cell surface protein profiling using DNA-barcoded antibodies were performed on the CD45⁺ fraction (Fig. 2a). We profiled a total of 2806 CD45⁻ cells and 62,723 CD45⁺ cells. For our protein expression analysis, we computationally isolated the CD4⁺ T cell compartment (16,008 cells) and the CD8⁺ T cell compartment (13,935 cells) (Fig. 2b, Supplementary Fig. 2). After filtering on highly variable genes and normalizing the expression data to unit variance, we used Leiden clustering to identify subpopulations within the CD4⁺ and CD8⁺ T cell compartments that have distinct transcriptional programs (Fig. 2c).

Within the CD8⁺ T cell compartment, we identified six subpopulations. Notably, five of the six clusters, CD8_{GZMK}, CD8_{GZMB-1}, CD8_{GZMB-2}, and CD8_{prolif} exhibited elevated expression of canonical cytotoxic genes, including granzymes and perforins, as well as associated transcription factors like *EOMES* (Fig. 2d). Cytotoxic T cells are critical effectors of immune response in solid tumors⁴¹. We found that

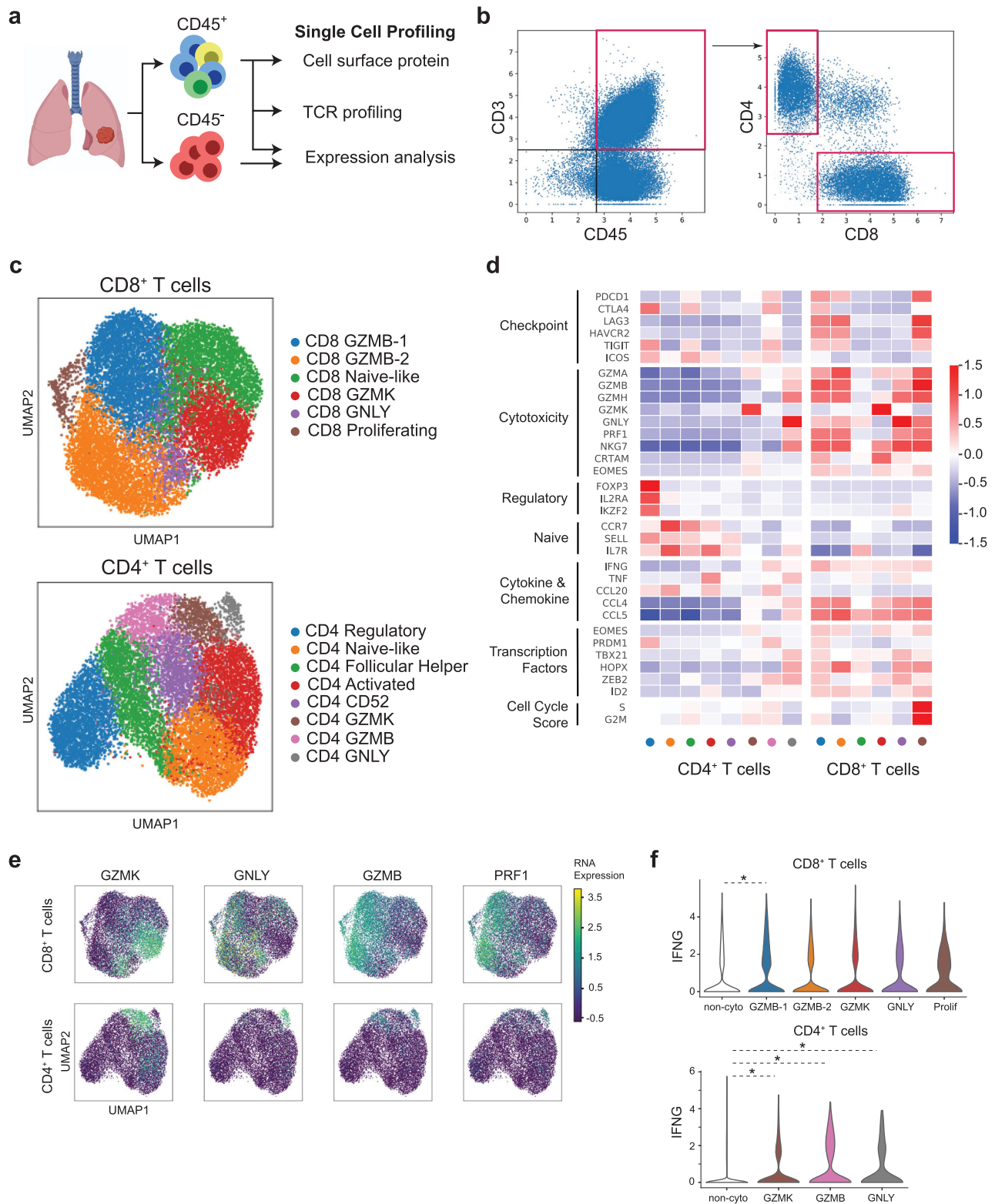


Fig. 2 Single-cell characterization of tumor-infiltrating T cells. **a** Schematic overview of the experimental design ($n = 10$ tumor samples). **b** Gating strategy for the computational isolation of the CD4⁺ (12,711 cells) and CD8⁺ T cell compartment (13,935 cells). Expression is shown in centered log-ratio (CLR)-transformed counts. **c** UMAP projections show a subpopulation of cells in the T cell compartment based on Leiden clustering. Cells were labeled based on the expression of key immune function genes. **d** Heatmap showing the scaled expression of key immune function genes in the T cell compartment. **e** UMAP projections show the distinct patterns of expression for the cytotoxic genes *PRF1*, *GZMB*, *GZMK*, and *GNLY*. **f** Violin plots show the log-transformed RNA expression of *IFNG* in the Leiden clusters. *Represents clusters with significantly different expressions ($p < 0.05$) compared to the other clusters after Bonferroni correction (two-sided Wilcoxon Rank-Sum). For top panel: $p_{GZMB-1} = 1.04 \times 10^{-23}$, $n = 13,935$ cells; bottom panel: $p_{GZMK} = 4.32 \times 10^{-12}$, $p_{GZMB} = 5.92 \times 10^{-44}$, $p_{GNLY} = 1.40 \times 10^{-5}$, $n = 12,711$ cells). Source data are provided as a Source Data file.

the pattern of expression for specific cytotoxic genes varied across the CD8⁺ T cell compartment (Fig. 2e). *GZMA*, *GZMH*, and *NKG7* were expressed broadly across the cytotoxic CD8⁺ population, however, elevated expression of *GZMB* and *GZMK* was observed in distinct clusters (Fig. 2e, Supplementary Fig. 3). A prior study using peripheral blood showed that *GZMK*⁺ CD8⁺ T cells are at an earlier stage of maturation than *GZMB*⁺ CD8⁺ T cells⁴². Consistent with that finding, *CRTAM*, an immunoglobulin superfamily protein induced after TCR activation^{43,44}, was primarily expressed in the CD8_{GZMK} population, indicating that these cells are in an earlier stage of effector cell differentiation than the other cytotoxic CD8⁺ populations. Interestingly, the CD8_{GZMB} cluster, which had the highest expression of the pore-forming peptide granulysin (*GZMB*), expressed both *GZMB* and *GZMK*, suggesting an intermediate state between the CD8_{GZMK} and CD8_{GZMB} clusters.

The two *GZMB*-high clusters identified, CD8_{GZMB-1} and CD8_{GZMB-2}, shared a number of characteristics. They both exhibited elevated *CD103* and *CD39* expression, a phenotype that has previously been associated with an enrichment of tumor-reactive T cells^{45,46}. In addition, they expressed high levels of immune checkpoint genes, such as *PDCD1*, *LAG3*, and *TIGIT*, suggesting that they are exhausted and could be targeted by ICB regimens. They also expressed *CXCL13*, which has been associated with better response and survival in NSCLC patients treated with PD-1 blockade due to enhanced immune cell recruitment to tertiary lymphoid structures⁴⁷. Notably, the CD8_{GZMB-1} population was enriched for *FABP5* and *41BB* expressing cells, which is indicative of a heightened oxidative metabolic rate. A comparable population has been observed in hepatocellular carcinoma, and it is hypothesized that the metabolic adaptation of these cells confers a survival advantage in the tumor microenvironment⁴⁸.

The final cytotoxic cluster, CD8_{prolif}, appeared to be actively proliferating as evidenced by high S and G2M cell cycle scores relative to other clusters. These cells most likely represent clones that were recently TCR stimulated and are now undergoing active clonal expansion (Fig. 2d). The only non-cytotoxic cluster, CD8_{naive-like}, expressed markers such as *CCR7* and *IL7R*, and most likely represent bystander T cells. All six populations, including the CD8_{naive-like} cells, expressed *IFNG* (Fig. 2d, f). The IFN- γ -regulating transcription factors *EOMES* and *TBX21* were expressed in multiple populations, suggesting redundancy and overlapping transcriptional control of *IFNG*.

We next sought to characterize the relationship between the CD8⁺ clusters using partition-based graph abstraction (PAGA)⁴⁹ (Supplementary Fig. 4a). Similar to our expectations based on the patterns of RNA expression, the CD8_{GZMK} and CD8_{GZMB} clusters had much greater connectivity to the CD8_{naive-like} than the two CD8_{GZMB} clusters. Interestingly, the CD8_{prolif} cluster was connected to the CD8_{GZMB} clusters but not the other cytotoxic clusters, suggesting that *GZMB* expressing cells are the primary CD8⁺ T cell population undergoing proliferation and clonal expansion in the tumor microenvironment.

We also identified populations of T cells with cytotoxic gene expression within the CD4⁺ T cell compartment, as previously noted in NSCLC⁵⁰ and other cancers^{24,27,51}. The pattern of cytotoxic gene expression was heterogeneous in these CD4⁺ T cells, but broadly paralleled that observed in the CD8⁺ T cell compartment (Fig. 2e). The expression of *GZMB* and *GZMK* was again elevated in distinct clusters (Fig. 2e, Supplementary Fig. 3). The *GZMB*-high cluster, CD4_{GZMB}, co-expressed *GZMB* but not *GZMK*. Similarly, *CRTAM* and *EOMES*, which have been implicated in the induction of the cytotoxic program in CD4⁺ T cells^{52,53}, were most highly expressed in the CD4_{GZMK} population. Interestingly, the CD4_{GZMB} cluster, but not the

CD4_{GZMK} or CD4_{GZMB} clusters, expressed high levels of immune checkpoints such as *PDCD1* and *CTLA4* (Fig. 2d, Supplementary Fig. 3), and are therefore the cells likely to be responsive to ICB therapy⁵⁴.

Notably, cytotoxic CD4⁺ T cells, particularly the CD4_{GZMB} and CD4_{GZMK} cells, expressed *IFNG* at significantly higher levels than non-cytotoxic CD4⁺ T cells ($p < 0.0001$ for both, Mann-Whitney *U*) (Fig. 2f). *IFNG* has been shown to directly increase HLA-II expression in tumor cells^{23,30,31,55}. Thus, cytotoxic CD4⁺ T cells may serve as a reservoir for the paracrine induction of HLA-II antigen presentation machinery in tumor cells.

Consistent with other known functions of CD4⁺ T cells, we identified a T follicular helper cluster, CD4_{follicular}, with high *IL7R* and *CD200* expression. Single-cell analyses in an NSCLC model have indicated that T follicular helper cells promote cytotoxic CD8⁺ T cell proliferation and tissue residence in the tumor microenvironment⁵⁶. In addition, we also identified a regulatory T cell cluster, CD4_{regulatory}, with high *FOXP3* expression and an activated T cell cluster, CD4_{activated}, characterized by *CD69* expression (Fig. 2d).

We also sought to characterize the relationship between clusters in the CD4⁺ compartment using a PAGA graph (Supplementary Fig. 4b). Similar to the findings in CD8⁺ T cells, the CD4_{GZMK} and CD4_{GZMB} clusters had close connections to the CD4_{naive-like} cluster, but the CD4_{GZMB} cluster did not. Taken together, the comparison of CD8⁺ and CD4⁺ T cells reveals the distinct development of cytotoxic populations that share a number of transcriptional features, including increased levels of granzymes, perforin, and immune checkpoint genes.

Cytotoxic T-cell populations are clonally expanded. Clonal expansion is a key trait of antigen-experienced T cells. To determine whether the cytotoxic T cell populations exhibited evidence of clonal expansion within the tumor microenvironment, we analyzed the TCR repertoires derived from the scRNAseq data. We define an expanded TCR clone as a population that contains more than one cell with identical TCR alpha and TCR beta CDR3 sequences and found evidence of extensive clonal expansion within all cytotoxic T cell populations.

Almost all CD8⁺ cytotoxic cells (83%) were members of an expanded T cell clone, and the Shannon entropy of the TCR repertoire from cytotoxic CD8⁺ cells was significantly lower than that of non-cytotoxic CD8⁺ cells ($p < 0.001$, Hutcheson's *t*-test) (Supplementary Fig. 5a). As expected, the CD8_{prolif} population was the most clonally expanded, followed by the CD8_{GZMB-1} and CD8_{GZMB-2} clusters (Fig. 3a, Supplementary Fig. 5b). Notably, the CD8_{prolif} cluster shared a high proportion of TCR clones with the two CD8_{GZMB} populations, indicating that these effector cells are being antigen-stimulated and actively undergoing clonal expansion (Fig. 3b). In contrast, the CD8_{prolif} cluster contained a lower proportion of CD8_{GZMK} and very few CD8_{GZMB} clones, implying that these early-stage cytotoxic CD8⁺ T cells proliferate less actively than CD8_{GZMB} populations.

There were far fewer clonally expanded CD4⁺ T cells than CD8⁺ T cells (Fig. 3c, Supplementary Fig. 5b). However, large clonal expansions were detected specifically within the cytotoxic CD4⁺ populations, with 54% of all cytotoxic CD4⁺ cells belonging to an expanded clone. Additionally, the Shannon entropy of the TCR repertoire of cytotoxic CD4⁺ cells was significantly lower than that of non-cytotoxic CD4⁺ cells ($p < 0.001$, Hutcheson's *t*-test). This suggests that similar to cytotoxic CD8⁺ T cells, these cells are responding to tumor antigens.

When we looked at the prevalence of shared clones across CD4⁺ T cell subpopulations, we discovered that 26% of

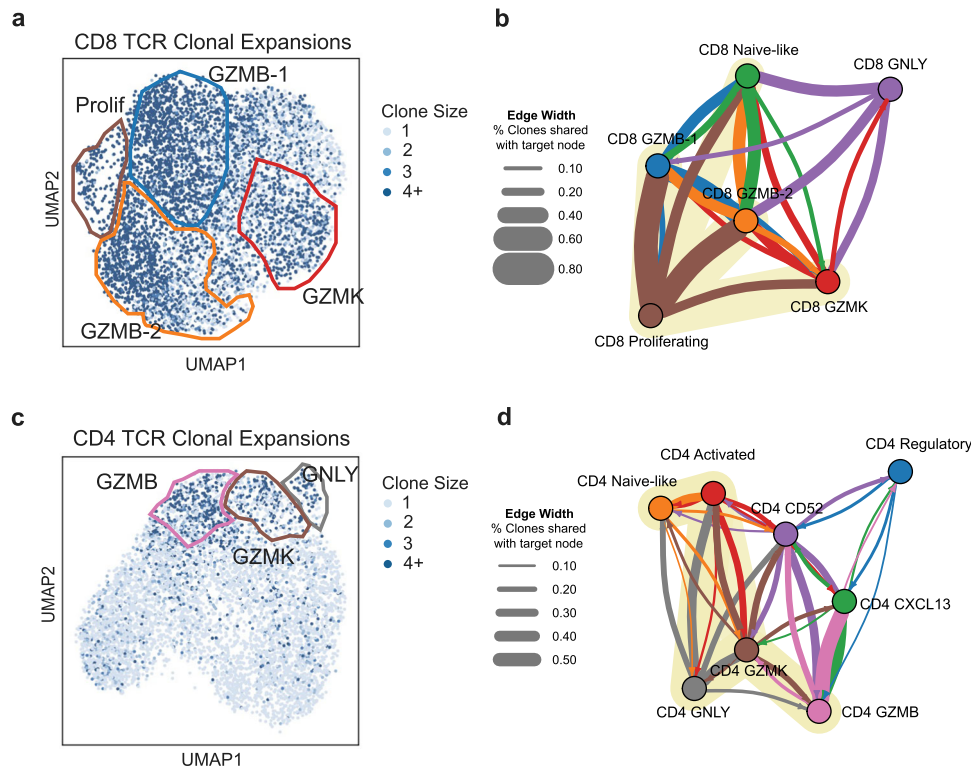


Fig. 3 Cytotoxic CD4⁺ and CD8⁺ T cells are clonally expanded in NSCLC. **a** UMAP projection showing the clone size associated with the TCR for each cell in CD8⁺ and **c** CD4⁺ T cells. **b, d** TCR clonal association between phenotypic states visualized using a graph structure. Each node represents a Leiden cluster and the width of each directed edge represents the percent of clones from the starting node shared with the target node. The color of each edge matches that of the starting node. Edges with weights in the lowest tertile for CD4⁺ and lowest quartile for CD8⁺ T cells are not shown for clarity. Source data are provided as a Source Data file.

CD4_{GZMK} clones and 35% of CD4_{GNLY} clones were shared with CD4_{activated} cells (Fig. 3d). In contrast, only 3% of CD4_{GZMB} cell clones were shared with the CD4_{activated} population (Fig. 3d). This suggests CD4_{GZMK} and CD4_{GNLY} cells are likely earlier-stage cytotoxic cells that are differentiated from recently activated CD4⁺ cells. However, the CD4_{GZMB} population shared over 54% of its clones with the CD4_{follicular} population (Fig. 3d). T follicular helper cells have been hypothesized to promote anti-tumor CD8⁺ T cell responses in NSCLC⁵⁶. The shared clonal lineage suggests that cytotoxic CD4⁺ T cells and those that may provision CD8⁺ T cells help stem from a common precursor. Taken together, our results demonstrate the significant clonal expansion of CD8⁺ and CD4⁺ cytotoxic T cells, suggesting that these populations both have encountered tumor antigens.

NSCLC tumor cells express HLA class II. To determine whether NSCLC tumor cells have the ability to directly present antigen to cytotoxic CD4⁺ T cells, we evaluated HLA expression in the CD45⁻ cell fraction using scHLAcount⁵⁰. We sequenced 2152 CD45⁻ cells and, utilizing Leiden clustering and lineage markers, identified distinct populations of the tumor, endothelial, and fibroblast cells (Supplementary Fig. 6). Since HLA-I is expressed by the vast majority of human cells, the majority of cells contained detectable amounts of HLA-I RNA (Fig. 4a). Expression of HLA-II is typically limited to antigen-presenting cells and endothelial cells. However, HLA-II has been shown to be expressed on lung epithelial cells^{57,58} as well as NSCLC tumor cells^{28–30}. We found that a subset of tumor cells expressed HLA-II and that its expression was significantly correlated with that of its chaperone, CD74 (invariant chain) ($R = 0.627$, $p < 0.0001$, Pearson correlation) (Fig. 4a, b). Notably, HLA-II expression in

tumor cells was not uniform across the genes examined, with significantly higher *HLA-DRB1* than *HLA-DQA1*, *HLA-DQB1*, and *HLA-DPB1* ($p < 0.0001$, Kruskal–Wallis) (Fig. 4c).

We further investigated the presence of cytotoxic CD4⁺ T cells and HLA-II-expressing tumor cells in treatment-naive NSCLC samples using immunofluorescence (Fig. 4d, Supplementary Fig. 7). We tested and optimized a panel of tumor epithelial- and lymphocyte-specific markers. Pan-cytokeratin staining was used to identify epithelial populations within tumor samples, while CD4 and CD8 staining was used to identify tumor-infiltrating T cells. In addition, we used HLA-DR to evaluate HLA-II expression and GZMB to identify cytotoxic lymphocyte populations. We identified regions within tumor samples in which T cells co-expressed CD4 and GZMB, and tumor cells co-expressed pan-CK and HLA-DR. In some instances, close proximity was observed between these populations. Additionally, H&E staining of a neighboring slide revealed clusters of malignant cells adjacent to lymphocyte clusters. These findings indicate the possibility of direct physical interaction between cytotoxic CD4⁺ T cells and HLA-II-expressing NSCLC tumor cells.

Integration of tumor extrinsic and intrinsic features associated with immunotherapy response. To determine the clinical relevance of CD8⁺ and CD4⁺ T cell cytotoxicity, we examined the relationship between the cytotoxic populations identified in the single-cell analysis and real-world ICB response in NSCLC. To begin, we generated gene signatures for each cluster by identifying the top 25 differentially expressed genes within cytotoxic clusters relative to all other clusters in the corresponding CD8⁺ and CD4⁺ compartments. We used the RNA-sequencing data from real-world patient tumor samples and their matched clinical data

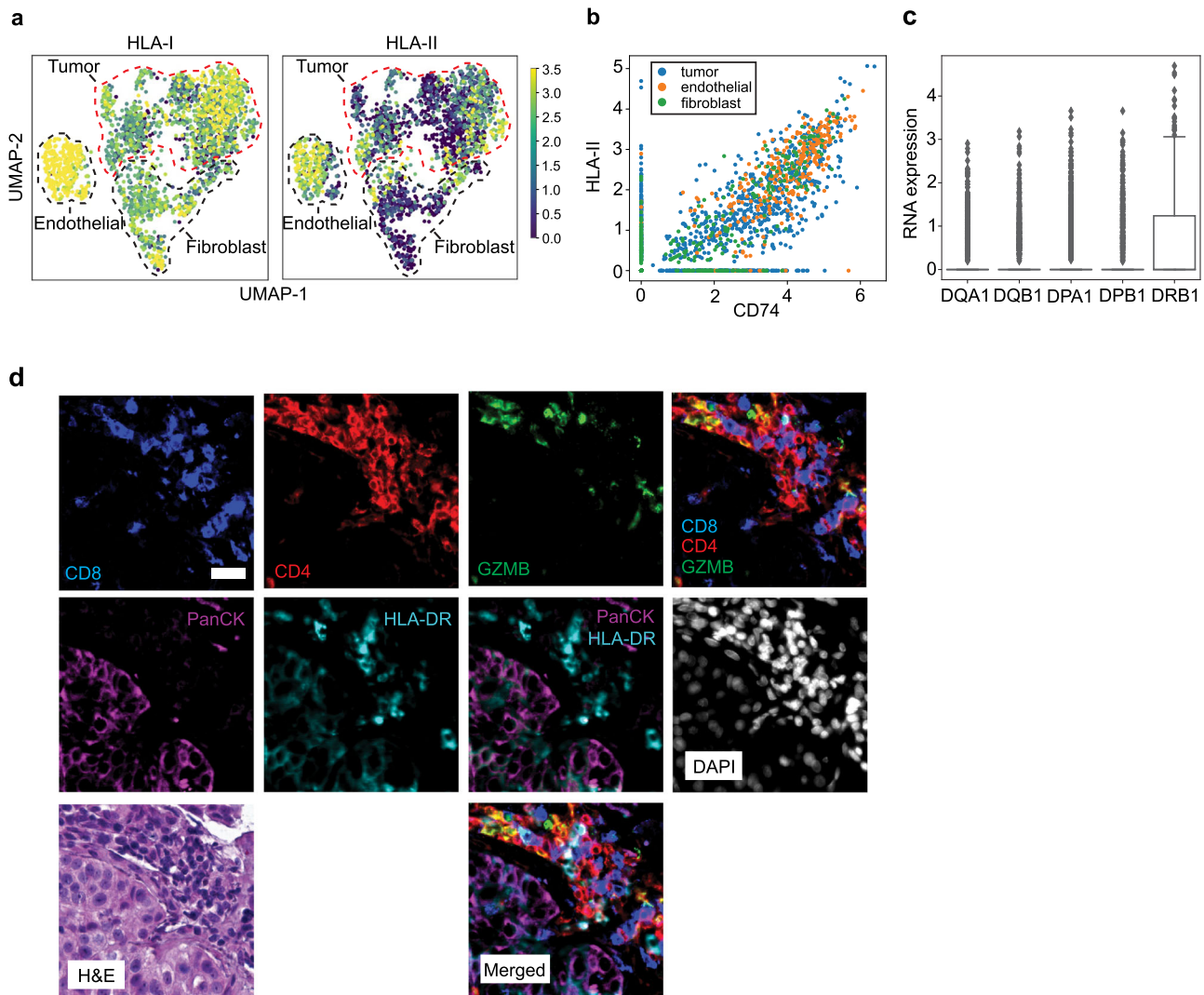


Fig. 4 A subpopulation of tumor cells express HLA class II in NSCLC. UMAP projections showing the expression of **a** HLA-I and HLA-II in the CD45⁺ fraction. HLA-I and II expression was measured by summing the expression of all the individual genes assessed in the locus. **b** Comparison of HLA-II and CD74 (invariant chain) expression ($R = 0.751$, unadjusted $p = 0$, Pearson correlation test). **c** Boxplots show the log-transformed expression of the individual HLA-II genes assessed ($p = 1.01 \times 10^{-178}$, Kruskal-Wallis, $n = 1377$ cells). The box represents the interquartile range, with the center line at the median. The whiskers extend up to 1.5 times the interquartile range (IQR). **d** Representative multiplex immunofluorescent staining of CD8 (blue), CD4 (red), GZMB (green), PanCK (magenta), HLA-DR (cyan), and DAPI (gray) in an NSCLC tumor ($n = 2$). Overlay without DAPI is shown for CD8, CD4, and GZMB, and for PanCK and HLA-DR, along with the corresponding H&E. Scale bar, 25 μm . Source data are provided as a Source Data file.

to assess the expression of specific gene signatures in patients and their association with ICB response. Cox proportional hazards (PH) analysis revealed that cytotoxic CD8⁺ and CD4⁺ signatures had the strongest associations with TTP (Supplementary Fig. 8a) (CD8_{GZMB-1}: HR = 0.77, $p = 0.033$, CD4_{GZMB}: HR = 0.80, $p = 0.044$, CoxPH). Thus, gene signatures from both cytotoxic CD8⁺ and CD4⁺ T cell populations were associated with real-world ICB response in NSCLC.

We next examined whether general CD8⁺ and CD4⁺ T cell infiltration were associated with real-world ICB response. We used estimates of CD8⁺ and CD4⁺ T cell infiltration derived from the xCell⁵⁹ algorithm and found that TTP was not significantly associated with the presence of total CD8⁺ and CD4⁺ T cells (CD8⁺ T cells: HR = 0.86, $p = 0.29$, CD4⁺ T cells: HR = 1.11, $p = 0.36$, CoxPH) (Supplementary Fig. 8b). This demonstrated that the presence of cytotoxic T cells, rather than total T cells, is associated with better prognosis in patients with NSCLC treated with ICB.

Following these findings, we investigated whether a pan-T cell cytotoxicity signature could be used to predict response to ICB in NSCLC patients. We developed a canonical cytotoxic gene signature consisting of genes that were highly expressed by CD4⁺ or CD8⁺ cytotoxic T cells in our single-cell data (Supplementary Fig. 9a, b). The gene list includes 25 genes with the greatest log-fold change in expression in either CD4⁺ or CD8⁺ cytotoxic T cells compared to non-cytotoxic T cells. We calculated a cytotoxic score by taking the arithmetic mean of the selected genes' log-transformed RNA expression values. Our cytotoxic score was found to be positively correlated with the gene signatures derived from the cytotoxic CD4⁺ and CD8⁺ Leiden clusters (Supplementary Fig. 9c). However, the cytotoxic score was not correlated to xCell estimates of total CD8⁺ or CD4⁺ T cell infiltration (Supplementary Fig. 9d, e).

Notably, the cytotoxic score was significantly associated with TTP in the Tempus ICB-treated NSCLC cohort (HR = 0.56,

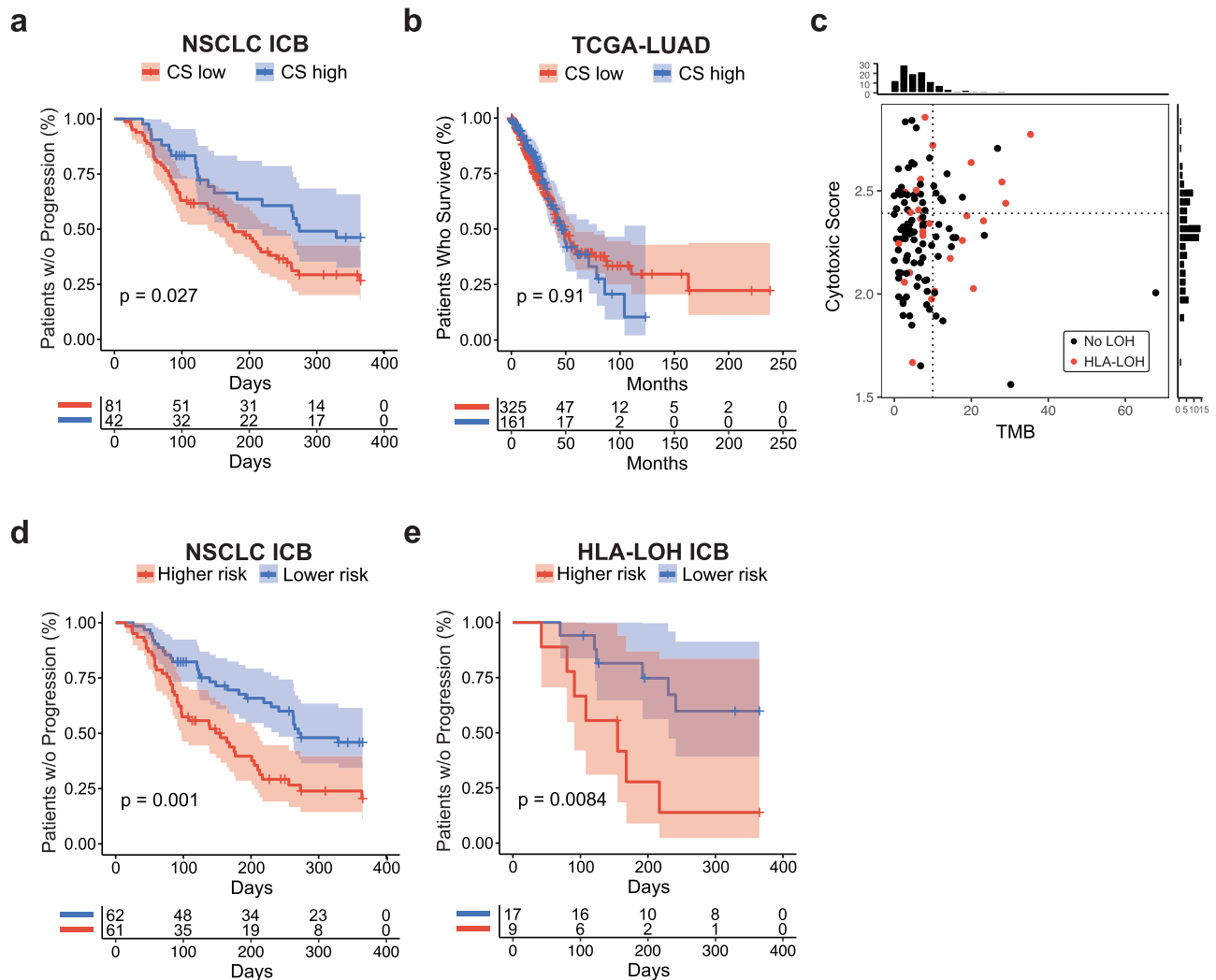


Fig. 5 Cytotoxic gene signature is associated with ICB response in NSCLC. **a** Kaplan-Meier plots showing time to progression stratified by cytotoxic score (CS) status in the Tempus NSCLC ICB cohort ($n = 123$, HR = 0.56, $p = 0.029$, log rank), and **b** the TCGA-LUAD NSCLC cohort ($n = 486$, HR = 1.00, $p = 1.0$, log rank). **c** Correlation of TMB and CS in the NSCLC ICB cohort ($n = 123$, $R = -0.0098$, unadjusted $p = 0.91$, Pearson correlation test). **d** Kaplan-Meier plot showing time to progression on ICB therapy, stratified by multimodal score (MM) status in the Tempus NSCLC ICB cohort ($n = 123$, HR = 1.94, $p = 0.005$, log rank) and **e** the HLA-I-deficient subgroup ($n = 26$, HR = 3.59, $p = 0.017$, log rank). Source data are provided as a Source Data file.

$p = 0.027$, log-rank) (Fig. 5a), but not in the TCGA LUAD⁶⁰ cohort of patients primarily treated with platinum therapy (HR = 1.00, $p = 1.0$, log-rank) (Fig. 5b). This indicated that cytotoxic gene expression is an important tumor extrinsic characteristic associated with immunotherapy response.

We next considered various tumor intrinsic characteristics that may act as a complement to T cell cytotoxicity. We reasoned that a higher mutation burden would increase potential tumor antigen availability to cytotoxic CD4⁺ and CD8⁺ cells. Notably, the cytotoxic score was not correlated with TMB in the ICB-treated NSCLC cohort ($R = -0.0028$, $p = 0.98$, Pearson correlation) (Fig. 5c). We combined TMB and the cytotoxic score to create a simple complementary multimodal model (MM) capable of capturing both the intrinsic availability of potential tumor antigens and the infiltration of T cells with cytotoxic phenotypes. First, TMB was binarized using a 10 mut/Mb cutoff (Fig. 1c), a value previously used in clinical trials^{5,8}. Then, we trained CoxPH models to predict TTP using the cytotoxic score and binarized TMB status on 100 random shuffles of the cohort, with 75% of patients in each shuffle used for training and 25% for evaluation.

The model was evaluated using its out-of-fold performance. Each patient was assigned an MM score based on their average out-of-fold score, and the scores were binarized using the cohort's median score as the cutoff to create two risk categories. In the Tempus ICB-treated NSCLC cohort, the MM score was significantly associated with TTP (HR = 2.17, $p = 0.001$, log-rank) (Fig. 5d).

Finally, we sought to specifically evaluate our MM performance on HLA-I-disrupted NSCLC patients. In the ICB-treated HLA-LOH subcohort, increasing MM score was significantly associated with shorter TTP (HR = 4.11, $p = 0.0084$, log-rank) (Fig. 5e). These findings demonstrate that features capturing T cell cytotoxic phenotype and TMB are associated with TTP in real-world ICB-treated NSCLC patients, including both HLA-I-intact and -disrupted patients.

Discussion

ICB therapy is now used as a standard of care for the majority of metastatic NSCLC patients. In this study, we evaluated the

association of tumor extrinsic features and intrinsic features with ICB response in a real-world cohort of NSCLC patients. We developed a gene signature for cytotoxicity that captures the transcriptional activity of both CD4⁺ and CD8⁺ cytotoxic T cells. Previous studies in other cancer types have demonstrated that cytotoxic gene signatures can be used as ICB biomarkers^{41,61–63}, but the predictive ability of these signatures and their association with the CD4⁺ compartment have not been thoroughly explored in NSCLC. We demonstrate that patients with higher expression of our cytotoxic score had significantly longer TTP on ICB than those with lower expression. This metric was independent of TMB and had more predictive power than PD-L1 IHC status. These findings demonstrate that integrating a canonical cytotoxic gene signature with TMB may be an effective biomarker for identifying NSCLC patients who are more likely to respond to ICB and can remain robust even in populations with high levels of HLA-LOH or other HLA-I antigen presentation deficiencies.

In this study, we also characterized the tumor-infiltrating T cell compartment in 10 NSCLC tumors. We identified a CD4⁺ T cell population that expresses a canonical cytotoxic program similar to classical cytotoxic CD8⁺ T cells. Cytotoxic CD4⁺ T cells in cancer have primarily been characterized in animal models, though recent studies have identified them in patients across multiple cancer types⁵¹. However, the presence, functions, and clinical implications of these cells in NSCLC have remained largely unknown. In this work, we show that CD4⁺ T cells with a cytotoxic phenotype are a notable component of the tumor-infiltrating immune population in NSCLC. Additionally, we demonstrated these cells are clonally expanded, suggesting they are specific for a tumor antigen. These cells also express *IFNG*, which may induce HLA-II expression as observed in some NSCLC cells here. Given the evidence for cytotoxic CD4⁺ T cells in multiple solid tumors⁵¹ as well as now NSCLC, we propose that cytotoxic CD4⁺ T cells are a fundamental component of the tumor immune microenvironment. Recent work by Cohen et al.⁵⁴ has elucidated the priming of CD4⁺ T cells that express cytotoxic genes (*Tht-1*) by antigen-presenting dendritic cells and implicated them in enhancing ICB-induced antitumor immune responses. At the same time, work by Oh et al.²⁷ have demonstrated cytotoxic CD4⁺ T cells can directly kill autologous tumors in an HLA-II-dependent manner. While anti-tumor activity in cancers lacking both HLA-II and HLA-I expression is thought to be coordinated by NK and CD4⁺ T cells^{64,65}, we propose that the distinct expression of HLA-II on NSCLC cells presents a viable pathway for cytotoxic CD4⁺ T cells to drive ICB response in HLA-I-disrupted patients.

More research to further characterize cytotoxic CD4⁺ T cells will also facilitate the development of the next generation of ICB drugs. For example, therapeutically increasing HLA-II expression in myeloid or tumor cells may provide a novel way to rationally combine drugs with ICB treatment. The inhibition of CDK4/6 and MEK has been linked to increased HLA-II expression in tumor cells^{66,67}. At the same time, cytotoxic CD4⁺ T cells upregulate *PDCD1*, *CTLA4*, and *TIGIT*, and are likely to be responsive to current ICB therapies. Notably, anti-CTLA-4 therapy, when combined with a decrease in regulatory T cells, has been shown to enhance cytotoxic CD4⁺ T cell anti-tumor activity⁶⁸. Thus, such rational combinations may successfully enhance tumor antigen presentation and the efficacy of ICB in certain cancers.

We used real-world data to characterize the relationship between tumor extrinsic and intrinsic features and ICB response in NSCLC patients. While real-world data can more accurately reflect the complexities and diversity of patient care in the real world than clinical trials, it does have some drawbacks. Standard clinical trial endpoints, such as RECIST scores and overall

survival, may be unavailable in some patients. Nonetheless, we were able to effectively investigate the impact of HLA-LOH on ICB efficacy and determine whether integrating tumor extrinsic and intrinsic features more accurately model real-world clinical responses³⁸. Further clinical research will be required to validate our integrative model as a predictive or a broadly prognostic NSCLC biomarker for ICB. As a result, we believe that clinical trials and real-world data complement one another and that incorporating real-world data into clinical research can lead to more robust biomarker discovery and guide clinical trial designs.

Finally, the development of pre-treatment biomarkers continues to be an important goal for precision medicine. Such pre-treatment biomarkers can help inform clinical judgment and lead to better outcomes by identifying progression events sooner, limiting usage of ineffective and costly ICB regimens, and improving patient quality of life by potentially transitioning to the next line of therapy before asymptomatic progression becomes symptomatic progression. We note that biomarkers considering HLA-LOH and HLA diversity alone have produced contradictory results in predicting ICB outcomes^{9,14,15,69}. Such HLA-I deficiencies are thought to limit tumor antigen presentation to CD8⁺ T cells but do not account for the potential role of cytotoxic CD4⁺ T cells. We propose that measuring the availability of potential tumor antigens and the infiltration of CD4⁺ and CD8⁺ T cells with cytotoxic phenotypes may aid in resolving the contradictory findings of these studies and provide a more complete picture of the tumor immune microenvironment. Overall, this study highlights the potential and utility of integrating tumor extrinsic and intrinsic features for predicting immunotherapy response in NSCLC.

Methods

Ethics statement. The use of de-identified molecular and clinical data in this study complies with all relevant ethical regulations. De-identified data were obtained from the Tempus Database. All data were de-identified in accordance with the Health Insurance Portability and Accountability Act (HIPAA) using Safe Harbor guidelines. The study protocol was submitted to the Advorra Institutional Review Board (IRB), which determined the research was exempt from IRB oversight.

NSCLC ICB cohort and clinical endpoints. De-identified patient records were selected from the Tempus Database. For inclusion in this study, patients were required to (1) have a diagnosis of metastatic NSCLC with non-squamous histology, (2) have received an ICB regimen, (3) have a documented progression event after treatment initiation, or have at least 90 days of follow up from treatment initiation, (4) completed next-generation DNA and RNA sequencing on an ICB-naive biopsy, and (5) have no actionable *EGFR* or *ALK* alterations. The primary clinical endpoint was real-world time to progression (TTP), defined as the time from the initiation of the ICB regimen to the first progression event, censored on the last known clinical encounter³⁸. Patients who ended treatment due to an adverse event, non-compliance, or another non-progression-related reason were censored at the time of treatment stop.

Clinical data abstraction. Clinical features for this study were derived from unstructured physician progress notes. The physician notes were abstracted using a standardized enriched curation process. A data dictionary and template were developed and reviewed with a panel of oncologists. The data dictionary included every field, the associated value sets, the definition of the fields, and scenarios to clarify for abstraction. Each patient case was curated by two abstractors, blinded to each other's curation. Discordances were reviewed by a third abstractor and escalated to a lung oncologist if adjudication was necessary. Two oncologists reviewed a random selection of patients from the final curated dataset to ensure validity.

DNA and RNA sequencing. Formalin-fixed, paraffin-embedded (FFPE) patient samples were profiled using Tempus xT targeted panel (596 genes), Tempus xO targeted panel (1700 genes), or Tempus xE whole-exome DNA sequencing, as well as Tempus xT whole-transcriptome RNA sequencing. Each sample underwent expert pathologist assessment for tumor cellularity and other quality measures. All tumor samples had at least 20% tumor content. Total nucleic acid was extracted from the FFPE tissue sections and digested by proteinase K. RNA was purified from the total nucleic acid by DNase-I digestion. A matched normal sample of blood or saliva was also obtained when possible. Germline DNA was extracted from either

650 μ l of saliva or 200 μ l of blood. DNA and RNA sequencing were performed as previously described^{34–36}.

PD-L1 immunohistochemistry. PD-L1 status for the NSCLC ICB cohort was preferentially determined by clinical Tempus testing with the 22C3 anti-PD-L1 antibody (Agilent). Slides were scored by a pathologist using the tumor proportion score (TPS), which is the percentage of tumor cells with complete or partial membrane staining. Samples with a TPS <1% are considered PD-L1 negative, 1–49% are considered PD-L1 low, and \geq 50% are considered PD-L1 high. For patients without Tempus 22C3 staining, PD-L1 IHC status was assessed from the abstracted clinical notes. For patients with recorded TPS scores, the same thresholds as described above were used. If no TPS score was available, but a PD-L1 IHC status was recorded, that status was used. PD-L1 IHC was performed on samples collected prior to ICB treatment start in all cases with known sample collection dates. Seven patients with abstracted PD-L1 IHC results had an unknown sample collection date. If there were multiple PD-L1 IHC tests recorded, the result from the test closest in time to the start of ICB treatment was used.

TCGA LUAD cohort. FASTQ files from RNA sequencing data for the TCGA LUAD cohort were downloaded from the Genomic Data Commons⁷⁰ and processed through the Tempus RNA pipeline as described below. The clinical data for the cohort was obtained from cBioportal⁷¹.

Tumor mutational burden. TMB was calculated by dividing the number of non-synonymous mutations by the megabase size of the panel, as previously described³⁶. All non-silent somatic coding mutations, including missense, indel, and stop-loss variants with coverage >100 \times and an allelic fraction >5% for targeted gene panels and coverage >30 \times and an allelic fraction >10% for the whole exome, were counted as non-synonymous mutations.

HLA loss of heterozygosity. HLA-LOH status was determined as previously described^{72,73}. Briefly, we assessed HLA-LOH status for all patients with a matched normal sample. Four-digit class, HLA-I typing was performed on the matched normal samples using Optitype (version 1.3.4)⁷⁴ and a custom HLA reference file for each patient was generated. All reads mapped to the HLA locus, as well as unmapped reads were extracted from the tumor and normal BAM files and remapped to the patient's HLA reference. After accounting for potential germline variants present in the sample's HLA, the alignments were updated and allele-specific coverage was determined. Changes in coverage between alleles, in the context of the expected tumor purity and copy states in the flanking genome, were assessed to determine if any reduction in allele coverage was consistent with a clonal loss of a specific HLA allele.

Transcriptomic analysis. Transcript-level quantification to GRCh37 was performed using Kallisto (version 0.44). Transcript counts were then corrected for GC content and length using quantile normalization and adjusted for sequencing depth via a size factor method. Normalized counts in protein-coding transcripts covered by the exome panel were then summed to obtain gene-level counts. Subsequent expression analyses were performed on log₁₀-transformed counts.

Single-cell multi-omics sequencing. Samples for single-cell multi-omic sequencing were previously frozen dissociated tumor cells (DTCs) (Discovery Life Sciences, Huntsville, AL). DTCs were thawed and washed with FACS buffer (PBS, 0.04% BSA). Up to 1 million cells were suspended in 45 μ l cell staining buffer (BioLegend Cat. No. 420201). The cells were blocked with Human 5 μ l TruStain FcX (Fc Receptor Blocking Solution, Cat. No. 422301) for 10 min on ice. An antibody pool was created by combining equal volumes of each TotalSeq-C antibody, including the anti-Human-CD45 antibody (Cat# 368545, BioLegend), anti-Human-CD3 antibody (Cat# 300479, BioLegend), anti-Human-CD4 antibody (Cat# 300567, BioLegend), anti-Human-CD8 antibody (Cat# 344753, BioLegend), anti-Human-CD20 antibody (Cat# 302363, BioLegend), and FITC-conjugated anti-Human-CD45 antibody (Cat# 304006, BioLegend), into a pool. The final concentration of each antibody in the pool was 1 μ g/ μ l as per the manufacturer's recommendations. 1 μ l (=1 μ g) of mixed TotalSeq-C + FITC-anti-CD45 antibody pool was then added to the 50 μ l volume and the mixture was incubated for 30 min on ice. Cells were washed by spinning at 400 \times g for 3 min, removing the upper liquid portion avoiding the pellet, and finally resuspension of the cell pellet in 500 μ l cell staining buffer (first 2 washes). The third wash was with FACS buffer (PBS + 0.04% BSA) including DAPI (10 min RT incubation). Finally, the cells were resuspended in FACS buffer without DAPI prior to cell sorting.

Samples were sorted using the SH800S cell sorter (Sony Biotechnology). Live cells were gated on DAPI⁻ cells sorted as CD45⁺ and CD45⁻ populations and collected in RPMI. The sorted CD45⁺ and CD45⁻ cells were pelleted and resuspended to recover a target of 3000 cells after 10 \times droplet formation. Cellular suspensions were barcoded using a Chromium Single Cell Controller instrument (10x Genomics) and 10 \times Genomics Chromium Single Cell A Chip Kit (P/N 120236, 10 \times Genomics) to generate single-cell Gel Beads-in-Emulsion (GEMs) for reverse transcription. Single-cell RNA-Seq libraries were prepared using the

Chromium Single Cell 5' Library and Gel Bead Kit (P/N 1000020, 10 \times Genomics) as per the manufacturer's instructions. For each sample, four libraries were generated: CD45⁻ 5' gene expression library, CD45⁺ 5' gene expression library, CD45⁺ TCR library, and CD45⁺ cell surface protein library.

Single-cell multi-omic analysis. Raw sequencing files were processed through the Cell Ranger pipeline (version 3.1.0) and then analyzed using Scanpy (version 1.6) and Scirpy (version 0.4). Cells with detectable gene expression in less than 200 genes, >6% mitochondrial genes for immune cells, >20% mitochondrial genes for tumor cells, or more than 2500 genes were removed from downstream analyses, as were any genes expressed in <3 cells. Scrublet⁷⁵ was used for doublet detection and removal. Gene expression values were normalized to 10,000 counts per cell and log-transformed. Protein expression values were normalized using the centered log-ratio normalization.

Data from the CD45⁺ fraction was then filtered on the CD4⁺ and CD8⁺ T cell populations, based on protein expression of CD45, CD3, CD4, CD8, and CD20 and RNA expression of CD68. Data from the CD45⁻ fraction was filtered to remove a minor population of contaminating immune cells. Genes from the T cell receptor and HLA loci were removed from the gene expression data. The gene expression data was then batch corrected using BBKNN (version 1.5.1)⁷⁶, filtered on highly variable genes, and scaled to unit variance. Leiden clustering was performed using Scanpy with a resolution of 0.7 for the CD8 cells and 0.8 for the CD4 cells. Cell cycle status was assessed using the score_genes_cell_cycle function in Scanpy with the gene list from Tirosh et al.⁶⁹ PAGA graphs were generated with the paga function in Scanpy with 6 neighbors and the first 20 principal components. Differential expression analysis was performed using the rank_genes_groups function in Scanpy with the Wilcoxon Rank-Sum test.

To assess HLA expression, we first performed HLA typing using ArcasHLA (version 0.2.5)⁷⁷ and then quantified HLA expression using sHLAcount (version 0.1.0)⁷⁸. Gene-level raw counts were then normalized by library size and log-transformed.

For the T cell receptor analysis, cells without a paired TRA and TRB, or with multiple TRA or TRB chains were removed. A TCR clonotype was defined as a group of cells with identical TRA and TRB CDR3 sequences. An expanded clone was considered as any clone consisting of more than one cell. Shannon entropy was calculated using the alpha_diversity function in scirpy and Hutcheson's *t*-test was performed using the ecolTest R package (version 0.0.1).

Immunofluorescence staining. Multiplex immunofluorescence staining was performed on FFPE sections as previously described⁷⁹. In brief, slides were deparaffinized and re-hydrated, followed by antigen retrieval. Slides were first stained with a cocktail of Tagged primary antibodies against CD8a (Clone: EPR10640(2), UltraTag: UT015, Cell Idx), CD4 (Clone: EPR6855, UltraTag: UT014, Cell Idx), Granzyme B (Clone: EPR20129-217, UltraTag: UT021, Cell Idx), panCK (Clone: AE1/AE3, UltraTag: UT016, Cell Idx), and HLA-DR (Clone: EPR3692, UltraTag: UT019, Cell Idx), and diluted with antibody diluent (PBS/1% BSA/0.2% Tween 20/15 mM Sodium Azide) for 1 h (UltraPlex detection system, Cell Idx). Slides were then washed with wash buffer and a cocktail of anti-Tag detection antibodies (UltraPlex detection system, Cell Idx), anti-UT015 (clone: CXC015, Flour: CL490, Cell Idx), anti-UT014 (clone: CXC015, Flour: CL550, Cell Idx), anti-UT021 (clone: CXC021, Flour: CL650, Cell Idx), anti-UT016 (clone: CXC016, Flour: CL480XL [megastoke dye], Cell Idx), and anti-UT019 (clone: CXC019, Flour: CL750, Cell Idx), were diluted with antibody diluent (PBS/1% BSA/0.2% Tween 20/15 mM Sodium Azide) and added to the slide and incubated for 1 h. As a negative control, slides were incubated with the secondary anti-Tag detection cocktail alone. Slides were then mounted using Fluoroshield with DAPI (Immunobiosciences) and coverslips applied prior to scanning at 20 \times using the Leica Versa scanner. Analysis was performed on the Aperio ImageScope, (v12.4.2.5010) using the Leica Quantitative Algorithm (v1).

Gene signatures. Gene lists for the Leiden cluster gene signatures were generated by taking the top 25 ranked differentially expressed genes for each cluster using the rank_genes_group Scanpy function. The gene score was then calculated by taking the arithmetic mean of the log-transformed RNA expression of the gene list. Scores were mean-centered and scaled prior to use in survival analyses.

The cytotoxic score was calculated by taking the arithmetic mean of 25 genes that are highly expressed in either CD4⁺ or CD8⁺ cytotoxic T cells. The genes are *NGG7*, *CXCL13*, *GZMH*, *HAVCR2*, *CCL5*, *GZMK*, *CCL4*, *GZMA*, *CCL3*, *CST7*, *CCLAL2*, *ACP5*, *TNFRSF9*, *TIGIT*, *GZMB*, *PDCD1*, *PRF1*, *LYST*, *SIRPG*, *LAG3*, *CARD16*, *TUBA4A*, *PTMS*, *CD74*, *KLRD1*.

Immune infiltration estimates. Estimates of immune cell infiltration were generated using the R package xCell⁵⁹ (version 1.1.0).

Multimodal model training. The MM score was generated via a Cox proportional hazards (CoxPH) model using cytotoxic score and the binarized TMB status to predict TTP on 100 random shuffles of the cohort. The train_test_split function from sklearn was used to randomly split 75% of patients in each shuffle into the training set and 25% of patients into the evaluation set. The

CoxPHSurvivalAnalysis function from the scikit-survival (version 0.17.2)⁸⁰ package was used to train the CoxPH models with the alpha parameter set to 1 and Breslow tie handling. Each patient was scored based on the arithmetic mean of their out-of-fold scores. The scores were then binarized using the median score of the cohort as the threshold to create two risk categories.

Survival analysis. Kaplan–Meier plots were generated using the survminer R package (version 0.4.8). Forest plots were generated using the survivalAnalysis R package (version 0.1.3). The log-rank test was used to compare survival curves and hazard ratios were calculated using a Cox proportional hazards model.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Patient demographics data are available in Supplementary Data 1 and 2. The raw sequencing data are not publicly available due to data privacy regulations and commercial restrictions on the use of such data. As such, the genomic data analyzed here, including the de-identified clinical data, DNA variant data, RNA expression data and single-cell multiomic profiling data, is available and may be obtained in accordance with Tempus's data sharing policy (<https://vivli.org/ourmember/tempus/>) as part of an external data access request (Accession ID: T21.02) linked here: <https://doi.org/10.25934/PR00007504>. The approximate time for processing a data access request is one month and inquiries about the process can be directed to publications@tempus.com. The TCGA data used in this study are available on the NIH Genomic Data Commons (<https://portal.gdc.cancer.gov/projects/TCGA-LUAD>) and the cBioPortal (https://www.cbioportal.org/study/summary?id=luad_tcga_pan_can_atlas_2018). The remaining data are available within the Article, Source Data, and Supplementary Information and Data files. Source data are provided with this paper.

Code availability

Requests for code used to generate the figures from Source Data are subject to a non-commercial data software agreement and review by Tempus and can be directed to publications@tempus.com. All other code used in this study is previously published and specified in the “Methods” section.

Received: 19 October 2021; Accepted: 1 July 2022;

Published online: 13 July 2022

References

- Hanna, N. H. et al. Therapy for stage IV non–small-cell lung cancer without driver alterations: ASCO and OH (CCO) joint guideline update. *J. Clin. Oncol.* **38**, 1608–1632 (2020).
- Brahmer, J. R. et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of non-small cell lung cancer (NSCLC). *J. Immunother. Cancer* **6**, 1–15 (2018).
- Gandhi, L. et al. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. *N. Engl. J. Med.* **378**, 2078–2092 (2018).
- Borghaei, H. et al. Five-year outcomes from the randomized, phase III trials checkmate 017 and 057: nivolumab versus docetaxel in previously treated non–small-cell lung cancer. *J. Clin. Oncol.* **39**, 723–733 (2021).
- Hellmann, M. D. et al. Nivolumab plus ipilimumab in advanced non–small-cell lung cancer. *N. Engl. J. Med.* **381**, 2020–2031 (2019).
- Herbst, R. et al. Association between tissue TMB (tTMB) and clinical outcomes with pembrolizumab monotherapy (pembro) in PD-L1-positive advanced NSCLC in the KEYNOTE-010 and-042 trials. *Ann. Oncol.* **30**, v916–v917 (2019).
- Paz-Ares, L. et al. Pembrolizumab (pembro) plus platinum-based chemotherapy (chemo) for metastatic NSCLC: tissue TMB (tTMB) and outcomes in KEYNOTE-021, 189, and 407. *Ann. Oncol.* **30**, v917–v918 (2019).
- Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* **362**, eaar3593 (2018).
- Litchfield, K. et al. Meta-analysis of tumor-and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614. e514 (2021).
- Sade-Feldman, M. et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat. Commun.* **8**, 1–11 (2017).
- Gettinger, S. et al. Impaired HLA class I antigen processing and presentation as a mechanism of acquired resistance to immune checkpoint inhibitors in lung cancer. *Cancer Discov.* **7**, 1420–1435 (2017).
- Zaretsky, J. M. et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N. Engl. J. Med.* **375**, 819–829 (2016).
- McGranahan, N. et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* **171**, 1259–1271. e1211 (2017).
- Chowell, D. et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).
- Montesion, M. et al. Somatic HLA class I loss is a widespread mechanism of immune evasion which refines the use of tumor mutational burden as a biomarker of checkpoint inhibitor response. *Cancer Discov.* **11**, 282–292 (2021).
- Shim, J. et al. HLA-corrected tumor mutation burden and homologous recombination deficiency for the prediction of response to PD-(L) 1 blockade in advanced non-small-cell lung cancer patients. *Ann. Oncol.* **31**, 902–911 (2020).
- Rodig, S. J. et al. MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. *Sci. Transl. Med.* **10**, eaar3342 (2018).
- Anagnostou, V. et al. Multimodal genomic features predict outcome of immune checkpoint blockade in non-small-cell lung cancer. *Nat. Cancer* **1**, 99–111 (2020).
- Middha, S. et al. Majority of B2M-mutant and-deficient colorectal carcinomas achieve clinical benefit from immune checkpoint inhibitor therapy and are microsatellite instability-high. *JCO Precis. Oncol.* **3**, 1–14 (2019).
- Borst, J., Ahrends, T., Bąbala, N., Melief, C. J. & Kastenmüller, W. CD4+ T cell help in cancer immunology and immunotherapy. *Nat. Rev. Immunol.* **18**, 635–647 (2018).
- Ferris, S. T. et al. cDC1 prime and are licensed by CD4+ T cells to induce anti-tumour immunity. *Nature* **584**, 624–629 (2020).
- Xie, Y. et al. Naive tumor-specific CD4+ T cells differentiated in vivo eradicate established melanoma. *J. Exp. Med.* **207**, 651–667 (2010).
- Quezada, S. A. et al. Tumor-reactive CD4+ T cells develop cytotoxic activity and eradicate large established melanoma after transfer into lymphopenic hosts. *J. Exp. Med.* **207**, 637–650 (2010).
- Cachot, A. et al. Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer. *Sci. Adv.* **7**, eabe3348 (2021).
- Lowery, F. J. et al. Molecular signatures of antitumor neoantigen-reactive T cells from metastatic human cancers. *Science* **375**, 877–884 (2022).
- Oliveira, G. et al. Landscape of helper and regulatory antitumour CD4+ T cells in melanoma. *Nature* **605**, 532–538 (2022).
- Oh, D. Y. et al. Intratumoral CD4+ T cells mediate anti-tumor cytotoxicity in human bladder cancer. *Cell* **181**, 1612–1625. e1613 (2020).
- Datar, I. J. et al. Spatial analysis and clinical significance of HLA class-I and class-II subunit expression in non–small cell lung cancer. *Clin. Cancer Res.* **27**, 2837–2847 (2021).
- He, Y. et al. MHC class II expression in lung cancer. *Lung Cancer* **112**, 75–80 (2017).
- Redondo, M. et al. Differential expression of MHC class II genes in lung tumour cell lines. *Eur. J. Immunogenet.* **25**, 385–391 (1998).
- Johnson, A. M. et al. Cancer cell–intrinsic expression of MHC Class II regulates the immune microenvironment and response to anti-PD-1 therapy in lung adenocarcinoma. *J. Immunol.* **204**, 2295–2307 (2020).
- Johnson, D. B. et al. Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat. Commun.* **7**, 1–10 (2016).
- Alspach, E. et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
- Beaubier, N. et al. Clinical validation of the Tempus xO assay. *Oncotarget* **9**, 25826 (2018).
- Beaubier, N. et al. Clinical validation of the tempus xT next-generation targeted oncology sequencing assay. *Oncotarget* **10**, 2384 (2019).
- Beaubier, N. et al. Integrated genomic profiling expands clinical options for patients with cancer. *Nat. Biotechnol.* **37**, 1351–1360 (2019).
- Lau, D., Bobe, A. M. & Khan, A. A. RNA sequencing of the tumor microenvironment in precision cancer immunotherapy. *Trends cancer* **5**, 149–156 (2019).
- Griffith, S. D. et al. Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced non-small-cell lung cancer data set. *JCO Clin. Cancer Inform.* **3**, 1–13 (2019).
- Neu, K. E., Tang, Q., Wilson, P. C. & Khan, A. A. Single-cell genomics: approaches and utility in immunology. *Trends Immunol.* **38**, 140–149 (2017).
- Ma, A., Xin, G. & Ma, Q. The use of single-cell multi-omics in immunology. *Nat. Commun.* **13**, 1–5 (2022).
- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
- Bratke, K., Kuepper, M., Bade, B., Virchow, J. C. Jr. & Luttmann, W. Differential expression of human granzymes A, B, and K in natural killer cells and during CD8+ T cell differentiation in peripheral blood. *Eur. J. Immunol.* **35**, 2608–2616 (2005).

43. Takeuchi, A. et al. CRTAM confers late-stage activation of CD8+ T cells to regulate retention within lymph node. *J. Immunol.* **183**, 4220–4228 (2009).
44. Boles, K. S., Barchet, W., Diacovo, T., Cella, M. & Colonna, M. The tumor suppressor TSLCI/NECL-2 triggers NK-cell and CD8+ T-cell responses through the cell-surface receptor CRTAM. *Blood* **106**, 779–786 (2005).
45. Duhon, T. et al. Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat. Commun.* **9**, 1–13 (2018).
46. Franczkiewicz, K. et al. Intratumoral induction of CD103 triggers tumor-specific CTL function and CCR5-dependent T-cell retention. *Cancer Res.* **69**, 6249–6255 (2009).
47. Thommen, D. S. et al. A transcriptionally and functionally distinct PD-1+ CD8+ T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat. Med.* **24**, 994–1004 (2018).
48. Liu, F. et al. Identification of FABP5 as an immunometabolic marker in human hepatocellular carcinoma. *J. Immunother. Cancer* **8**, e000501 (2020).
49. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 1–9 (2019).
50. Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
51. Oh, D. Y. & Fong, L. Cytotoxic CD4+ T cells in cancer: expanding the immune effector toolbox. *Immunity* **54**, 2701–2711 (2021).
52. Takeuchi, A. et al. CRTAM determines the CD4+ cytotoxic T lymphocyte lineage. *J. Exp. Med.* **213**, 123–138 (2016).
53. Kitano, S. et al. Enhancement of tumor-reactive cytotoxic CD4+ T-cell responses after ipilimumab treatment in four advanced melanoma patients. *Cancer Immunol. Res.* **1**, 235–244 (2013).
54. Cohen, M. et al. The interaction of CD4+ helper T cells with dendritic cells shapes the tumor microenvironment and immune checkpoint blockade response. *Nat. Cancer* **3**, 303–317 (2022).
55. Kamma, H., Yazawa, T., Ogata, T., Horiguchi, H. & Iijima, T. Expression of MHC class II antigens in human lung cancer cells. *Virchows Arch. B* **60**, 407–412 (1991).
56. Cui, C. et al. Neoantigen-driven B cell and CD4 T follicular helper cell collaboration promotes anti-tumor CD8 T cell responses. *Cell* **184**, 6101–6118. e6113 (2021).
57. Kalb, T. H., Chuang, M. T., Marom, Z. & Mayer, L. Evidence for accessory cell function by class II MHC antigen-expressing airway epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **4**, 320–329 (1991).
58. Glanville, A. R. et al. The distribution of MHC class I and II antigens on bronchial epithelium. *Am. Rev. Respir. Dis.* **139**, 330–334 (1989).
59. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 1–14 (2017).
60. Collisson, E. et al. Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. *Nature* **511**, 543–550 (2014).
61. Ayers, M. et al. IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
62. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).
63. Prat, A. et al. Immune-related gene expression profiling after PD-1 blockade in non-small cell lung carcinoma, head and neck squamous cell carcinoma, and melanoma. *Cancer Res.* **77**, 3540–3550 (2017).
64. Badrinath, S. et al. A vaccine targeting resistant tumours by dual T cell plus NK cell attack. *Nature* **606**, 992–998 (2022).
65. Doorduijn, E. M. et al. CD4+ T cell and NK cell interplay key to regression of MHC class II low tumors upon TLR7/8 agonist therapy. *Cancer Immunol. Res.* **5**, 642–653 (2017).
66. Schaer, D. A. et al. The CDK4/6 inhibitor abemaciclib induces a T cell inflamed tumor microenvironment and enhances the efficacy of PD-L1 checkpoint blockade. *Cell Rep.* **22**, 2978–2994 (2018).
67. Neuwelt, A. J. et al. Cancer cell-intrinsic expression of MHC II in lung cancer cell lines is actively restricted by MEK/ERK signaling and epigenetic mechanisms. *J. Immunother. Cancer* **8**, e000441 (2020).
68. Śledzińska, A. et al. Regulatory T cells restrain interleukin-2 and Blimp-1-dependent acquisition of cytotoxic function by CD4+ T cells. *Immunity* **52**, 151–166. e156 (2020).
69. Chhibber, A. et al. Germline HLA landscape does not predict efficacy of pembrolizumab monotherapy across solid tumor types. *Immunity* **55**, 56–64. e4 (2022).
70. Grossman, R. L. et al. Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).
71. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, p11–p11 (2013).
72. Larsen, B. M. et al. A pan-cancer organoid platform for precision medicine. *Cell Rep.* **36**, 109429 (2021).
73. Perera, J., Mapes, B., Lau, D., Salahudeen, A. & Khan, A. Detection of human leukocyte antigen class I loss of heterozygosity in solid tumor types by next-generation DNA sequencing. *Lung* **26**, 50 (2019).
74. Szolek, A. et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
75. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291. e289 (2019).
76. Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
77. Orenbuch, R. et al. arcasHLA: high-resolution HLA typing from RNAseq. *Bioinformatics* **36**, 33–40 (2020).
78. Darby, C. A., Stubbington, M. J., Marks, P. J., Martínez Barrio, Á. & Fiddes, I. T. scHLAcount: allele-specific HLA expression from single-cell gene expression data. *Bioinformatics* **36**, 3905–3906 (2020).
79. Durante, M. A. et al. Single-cell analysis reveals new evolutionary complexity in uveal melanoma. *Nat. Commun.* **11**, 1–10 (2020).
80. Pölsterl, S. scikit-survival: a Library for Time-to-Event Analysis Built on Top of scikit-learn. *J. Mach. Learn. Res.* **21**, 1–6 (2020).

Acknowledgements

The authors thank all patients and their families and caregivers as well as physicians and clinical care teams that have enabled this study. Figure 2a was created with BioRender.com. The authors thank Matt Levin, Helen Snyder and Zohreh AlkavanAghdam from Cell IDx for assistance with mIF; Pete Savage, Justin Guinney, Alice Walsh, Tim Taxter, Gautam Kumar, Jason Perera, Ariane Lozac’hmeur, Mario Rosasco, Ken Carson, Kathryn Sands, Matthew Kase and Calvin Chao with helpful discussions.

Author contributions

D.L., A.A.S., and A.A.K. conceived and designed the overall study. S.K. and T.A.R. performed the single-cell sample processing and sequencing. D.L. and A.A.K. designed and performed the single-cell computational analyses. D.L., A.A.K., M.M.S., P.J., and Y.G. designed and performed the NSCLC cohort analyses. A.A.S., A.B., D.L., and A.A.K. designed and performed the imaging study. D.L. and A.A.K. wrote the manuscript and all authors reviewed and approved the final manuscript.

Competing interests

D.L., S.K., M.M.S., P.J., Y.G., A.B., T.A.R., A.A.S., and A.A.K. are either current or previous employees at Tempus Labs, Inc. and may be listed as co-inventors on various pending patent applications related to the multimodal model presented in this study, including “Determination of cytotoxic gene signature and associated systems and methods for response prediction and treatment” (US 2022-0154284).

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-31769-4>.

Correspondence and requests for materials should be addressed to Denise Lau or Aly A. Khan.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022