Article

Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers

https://doi.org/10.1038/s41586-021-03752-4

Received: 8 January 2021

Accepted: 18 June 2021

Published online: 21 July 2021

Open access

Check for updates

Justina X. Caushi[1,2,3,15], Jiajia Zhang[1,2,3,15], Zhicheng Ji[4,11], Ajay Vaghasia[3], Boyang Zhang[4], Emily Han-Chung Hsiue[3,5,6], Brian J. Mog[3,5,6], Wenpin Hou[4], Sune Justesen[7], Richard Blosser[1,3], Ada Tam[1,3], Valsamo Anagnostou[1,3], Tricia R. Cottrell[1,2,3,12], Haidan Guo[1,2,3], Hok Yee Chan[1,2,3], Dipika Singh[1,2,3], Sampriti Thapa[1,2,3], Arbor G. Dykema[1,2,3], Poromendro Burman[1,2,3], Begum Choudhury[1,2,3], Luis Aparicio[1,2,3], Laurene S. Cheung[1,2,3], Mara Lanis[1,3], Zineb Belcaid[1,3], Margueritta El Asmar[1,3], Peter B. Illei[3], Rulin Wang[3], Jennifer Meyers[3], Kornel Schuebel[3], Anuj Gupta[3], Alyza Skaist[3], Sarah Wheelan[3], Jarushka Naidoo[1,3,13], Kristen A. Marrone[1,3], Malcolm Brock[3], Jinny Ha[3], Errol L. Bush[3], Bernard J. Park[8], Matthew Bott[8], David R. Jones[8], Joshua E. Reuss[3,14], Victor E. Velculescu[1,3], Jamie E. Chaft[8], Kenneth W. Kinzler[3,5,6], Shibin Zhou[3,5,6], Bert Vogelstein[3,5,6], Janis M. Taube[1,2,3], Matthew D. Hellmann[8], Julie R. Brahmer[1,3], Taha Merghoub[8,9,10], Patrick M. Forde[1,3], Srinivasan Yegnasubramanian[1,2,3]✉, Hongkai Ji[4]✉, Drew M. Pardoll[1,2,3]✉ & Kellie N. Smith[1,2,3]✉

PD-1 blockade unleashes CD8 T cells[1], including those specific for mutation-associated neoantigens (MANA), but factors in the tumour microenvironment can inhibit these T cell responses. Single-cell transcriptomics have revealed global T cell dysfunction programs in tumour-infiltrating lymphocytes (TIL). However, the majority of TIL do not recognize tumour antigens[2], and little is known about transcriptional programs of MANA-specific TIL. Here, we identify MANA-specific T cell clones using the MANA functional expansion of specific T cells assay[3] in neoadjuvant anti-PD-1-treated non-small cell lung cancers (NSCLC). We use their T cell receptors as a 'barcode' to track and analyse their transcriptional programs in the tumour microenvironment using coupled single-cell RNA sequencing and T cell receptor sequencing. We find both MANA- and virus-specific clones in TIL, regardless of response, and MANA-, influenza- and Epstein–Barr virus-specific TIL each have unique transcriptional programs. Despite exposure to cognate antigen, MANA-specific TIL express an incompletely activated cytolytic program. MANA-specific CD8 T cells have hallmark transcriptional programs of tissue-resident memory (TRM) cells, but low levels of interleukin-7 receptor (IL-7R) and are functionally less responsive to interleukin-7 (IL-7) compared with influenza-specific TRM cells. Compared with those from responding tumours, MANA-specific clones from non-responding tumours express T cell receptors with markedly lower ligand-dependent signalling, are largely confined to HOBIT^{high} TRM subsets, and coordinately upregulate checkpoints, killer inhibitory receptors and inhibitors of T cell activation. These findings provide important insights for overcoming resistance to PD-1 blockade.

The efficacy of PD-1- and PD-L1-blocking agents is predicated upon CD8 T cell-mediated anti-tumour immunity[1]. Early studies focused on tumour-associated antigens, whereas recent work has shifted attention to T cell recognition of mutation-associated neoantigens (MANA), owing to the large numbers of somatic mutations acquired by many cancers during their development[4]. The association of improved anti-PD-1 and anti-PD-L1 clinical responses with high tumour mutational burden[5] strongly suggests that MANA

[1]Bloomberg-Kimmel Institute for Cancer Immunotherapy at Johns Hopkins, Baltimore, MD, USA. [2]The Mark Center for Advanced Genomics and Imaging at Johns Hopkins, Baltimore, MD, USA. [3]Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA. [4]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [5]Ludwig Center and Howard Hughes Medical Institute at Johns Hopkins, Baltimore, MD, USA. [6]Lustgarten Pancreatic Cancer Research Laboratory, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA. [7]Immunitrack, Copenhagen, Denmark. [8]Department of Medicine, Memorial Sloan Kettering Cancer Center and Weill Cornell Medicine, New York, NY, USA. [9]The Swim Across America and Ludwig Collaborative Laboratory, Immunology Program, Parker Institute for Cancer Immunotherapy at Memorial Sloan Kettering Cancer Center, New York, NY, USA. [10]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [11]Present address: Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, USA. [12]Present address: Ontario Institute for Cancer Research, Queens University, Kingston, Ontario, Canada. [13]Present address: Beaumont Hospital Dublin, RCSI University of Medicine and Health Science, Dublin, Ireland. [14]Present address: Georgetown Lombardi Comprehensive Cancer Center at Georgetown University, Washington, DC, USA. [15]These authors contributed equally: Justina X. Caushi, Jiajia Zhang. ✉e-mail: syegnasu@jhmi.edu; hji@jhu.edu; dpardol1@jhmi.edu; kellie@jhmi.edu

are important targets of anti-tumour immunity induced by PD-1 blockade.

Despite the success of immune checkpoint blockade (ICB) in improving clinical outcomes, most cancers still do not respond[6]. Improving response rates to ICB will require an understanding of the functional state of tumour-specific T cells, particularly in the tumour microenvironment. However, a fundamental limitation in the current understanding of the T cell functional programs that underpin the response to ICB has been the absence of transcriptional profiling of true MANA-specific TIL. A related problem is the paucity of information regarding the differences between MANA-specific TIL in ICB-responsive versus resistant tumours. Indeed, MANA-specific T cells represent a small fraction of total TIL[2,7], particularly in lung cancer, in which they have been shown to selectively upregulate CD39. This highlights the challenges confronting characterization of the cells responsible for the activity of T cell-targeting immunotherapies.

Global gene expression of NSCLC TIL

For this study, we used peripheral blood and tissue biospecimens obtained from the first-in-human clinical trial of neoadjuvant anti-PD-1 (nivolumab) in resectable non-small cell lung cancer[8] (NSCLC; ClinicalTrials.gov identifier: NCT02259621; Fig. 1a, top) to study the transcriptional programs of MANA-specific TIL. Nine out of 20 patients with NSCLC (45%) treated in this trial had a major pathologic response (MPR) at the time of resection, defined as no more than 10% viable tumour at the time of surgery; previous studies have established an association between MPR and improved overall survival[9–12]. A schematic of the study design and experimental approach is shown in Fig. 1a, bottom. Combined single-cell RNA sequencing (scRNA-seq) and T cell receptor sequencing (TCR-seq) was performed on TIL ($n = 15$), paired adjacent normal lung ($n = 12$), tumour-draining lymph nodes (TDLN, $n = 3$) and a distant metastasis (Extended Data Fig. 1a, Supplementary Tables 1–3). In total, 560,916 T cells passed quality control (Fig. 1b, Supplementary Table 3) and were carried forward for analyses.

Uniform manifold approximation and projection (UMAP) analysis of cells from all samples on the basis of filtered and normalized transcript counts defined 15 T cell clusters (Fig. 1b, c, Extended Data Fig. 1b–e, Supplementary Data 1.1). Expression of subset-defining markers and T cell checkpoints was visualized in red scale on the UMAP (Fig. 1d). The two clusters designated as TRM had the highest expression of the canonical TRM genes, *ZNF683* (also known as *HOBIT*) and *ITGAE* (also known as *CD103*), and the highest expression of a TRM gene set[13] (Extended Data Fig. 1f, Supplementary Data 2.1). Principal component analysis (PCA) of samples based on concatenated cell-cluster-level pseudobulk profiles distinguished adjacent normal-lung T cells from TIL (Fig. 1e), but did not distinguish MPR from non-MPR TIL (Fig. 1f). We did not observe notable differentially expressed gene programs between MPR and non-MPR TIL (Supplementary Data 3), indicating that gene expression profiling of total TIL has limited sensitivity in distinguishing the pathologic response to PD-1 blockade.

Expression programs of MANA-specific TIL

We next performed the MANA functional expansion of specific T cells assay (MANAFEST)[3] on 9 of the 16 individuals on whom scRNA-seq–TCR-seq was conducted. This assay detects in vivo antigen-experienced T cell responses and identifies the clonal identity of the T cell receptor (TCR) corresponding to these cells. Of these nine, four were classed as MPR and five were non-MPR (results from one individual have been previously described[8]). Putative MANA (Supplementary Tables 4–6), peptide pools representing influenza matrix and nucleoproteins, and a pool of major histocompatibility complex (MHC) class I-restricted cytomegalovirus (CMV), Epstein–Barr virus (EBV) and influenza virus epitopes (CEF) were queried for $CD8^+$ T cell reactivity in parallel (Supplementary Tables 6, 7). From 7 (3 with MPR and 4 without MPR) of the 9 individuals, 72 total unique MANA-specific TCRs, 33 unique CEF-specific TCRs, and 52 unique influenza-specific TCRs were identified (Extended Data Fig. 2, Supplementary Tables 8, 9, Supplementary Data 4, 5). Out of 33 CEF-specific TCRs, 6 matched known public EBV-specific TCRs and 3 matched known public influenza-specific TCRs[14]. No CMV-reactive TCRs were mapped from our CEF-specific TCRs. Notably, 4 of the 41 MANA-specific TCRVβ complementarity-determining region 3 (CDR3) clonotypes identified in a patient without MPR (patient ID MD01-004) (Extended Data Fig. 2) were specific for a MANA (MD01-004-MANA12) derived from a p53 R248L hotspot mutation, and were found at appreciable frequency in the pre- and post-treatment tumour (Extended Data Fig. 3), despite the tumour not attaining MPR. Most MANA-specific clones were detected at very low frequency (median: 0.001%) in the peripheral blood across all available time points (Fig. 2a, Extended Data Fig. 3). Overall, pathologic response was not associated with the prevalence, frequency or intratumoral representation of MANA-specific T cells (Extended Data Figs. 2, 3 Supplementary Table 9). In fact, more MANA-specific TIL were observed among non-MPR TIL than among MPR TIL. No consistent pattern was observed for the frequency of viral-specific T cells in the tissue or peripheral blood (Extended Data Figs. 2, 3).

Ten MANA-specific clonotypes, for which the TCRα could be confidently identified from the single-cell analysis, were selected for validation of MANA recognition via TCR cloning and introduction into a Jurkat–NFAT luciferase reporter system[15]. Seventy per cent of tested clonotypes (representing 95.2% of total cells bearing TCRs identified by MANAFEST) were validated as MANA-specific (Extended Data Fig. 4a–c). Peptide–human leukocyte antigen (HLA) binding assays demonstrated that two MANA peptides—MD01-005-MANA7 and MD01-004-MANA12—displayed comparably high MHC class I affinity (measured dissociation constants (K_d) = 5.1 nM and 17.5 nM, respectively) and stability (Extended Data Fig. 4d, e).

We next evaluated the transcriptional programming of MANA- and viral-specific $CD8^+$ T cells. Refined clustering of all $CD8^+$ T cells ($n = 235{,}851$) identified 15 unique clusters (Fig. 2b, Extended Data Fig. 5a, Supplementary Data 1.2). Clusters were named on the basis of previously defined T cell states from single-cell transcriptomic studies[16]. Six clusters had gene expression programs consistent with TRM T cells, characterized by high expression of *HOBIT*, *LINC02446*, *CD103* and a previously published TRM gene set (Extended Data Fig. 5b). Selective genes and linkage to the global CD3 T cell clusters shown in Fig. 1 were visualized (Extended Data Figs. 5c, d). The six TRM subsets were heterogenous in their expression of an exhaustion gene set described previously in NSCLC[17] (Extended Data Fig. 5e, Supplementary Data 2.2). None of the most frequent tumour-infiltrating clonotypes were restricted to a single cluster (Extended Data Fig. 5f). Among all tested individuals, a total of 28 MANA-specific CD8 clonotypes (1,350 total cells from 3 patients with MPR and 3 patients without MPR) as identified by MANAFEST were detected in the single-cell data, of which 20 clonotypes (890 cells) were in the tumour (Fig. 2c, Supplementary Table 8). Of the viral-specific T cell clonotypes, 23 influenza-specific (866 cells) and 2 EBV-specific (281 cells) clones were found in the CD8 single-cell analysis.

Overlay of these clonotypes onto the $CD8^+$ T cell UMAP demonstrated a marked distinction between the clonotypes with different antigen specificities (Fig. 2c, Extended Data Fig. 6a–c). EBV-reactive T cells primarily resided in effector T (T_{eff}) cell clusters, whereas influenza- and MANA-specific T cells largely occupied distinct TRM clusters. Notably, because influenza is a respiratory virus, influenza-specific T cells may be considered the archetypal lung-resident memory T cells[18]. None of the patients in our study were symptomatic for influenza in the six weeks preceding surgery. It is therefore not surprising that influenza-specific CD8 cells were TRM rather than T_{eff} cells. By contrast, EBV-specific T cells exclusively occupied T_{eff} clusters, consistent with periodic acute

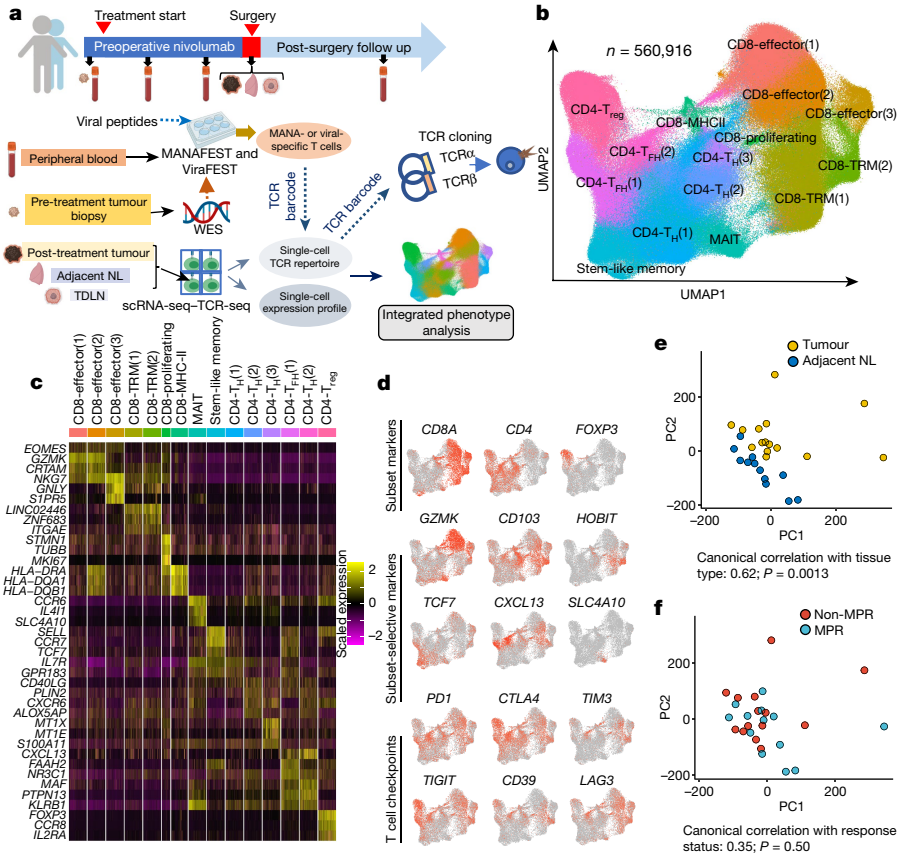


Fig. 1 | Profiling single T cells in NSCLC treated with neoadjuvant PD-1 blockade. Twenty patients with resectable NSCLC were treated with two doses of PD-1 blockade before surgical resection. **a**, An overall schematic of the clinical trial, biospecimen collection (top) and study design (bottom). scRNA-seq–TCR-seq was performed on T cells isolated from resected tumour ($n = 15$), adjacent normal lung (NL; $n = 12$), TDLN ($n = 3$), and a resected brain metastasis ($n = 1$) from patients with NSCLC treated with two doses of neoadjuvant anti-PD-1 (bottom). The MANAFEST and ViraFEST assays were used to identify MANA- and viral (EBV and influenza)-specific TCRs, respectively. WES, whole-exome sequencing. **b**, UMAP projection of the expression profiles of the 560,916 T cells that passed quality control. Immune cell subsets, defined by 15 unique clusters, are annotated and marked by colour code. **c**, Relative expression of the top-3 most differentially expressed genes. Five-thousand cells (or all cells in the cluster if the cluster size was fewer than 5,000 cells) were randomly sampled from each cluster for visualization. MAIT, mucosal-associated invariant T cells; T_{FH}, T follicular helper cells; T_{reg}, regulatory T cells. **d**, Expression of T cell subset-defining genes, T cell subset-selective genes and major T cell checkpoint genes. *CD39* is also known as *ENTPD1*. **e**, PCA of cell-cluster-level pseudobulk gene expression for individual samples for tumour (yellow, $n = 15$) and adjacent normal lung (dark blue, $n = 12$). One-sided permutation test. **f**, PCA of cell-cluster-level pseudobulk gene expression for non-MPR (red, $n = 9$) and MPR (light blue, $n = 6$) tumours. One-sided permutation test.

stimulation upon latent EBV reactivation. Whereas influenza-specific cells were the most abundant in normal lung, MANA-specific CD8 cells were more numerous in the tumour (Extended Data Fig. 6d, e).

There was considerable shared expression of selective cytotoxic T lymphocyte (CTL) activation genes between MANA- and EBV-specific T cells, in particular genes encoding T cell activation and CTL activity, such as *HLA-DRA*, *GZMH*, *IFNG* and *NKG7* (Fig. 2d, Supplementary Data 1.3). However, genes encoding certain canonical cytolytic molecules, such as *GZMK*, were expressed at low levels in MANA-specific TIL. Most notably, *EOMES*, which encodes a transcription factor that is critical for CTL activity[19], was expressed in EBV-specific CD8 cells but was minimally expressed in most MANA-specific cells. Multiple checkpoints were significantly upregulated in MANA-specific TIL compared with EBV-specific TIL. Notably, MANA-specific cells expressed higher levels of *PRDM1*, which encodes BLIMP-1 and has been reported to participate in coordinated transcriptional activation of multiple checkpoint genes, including *PD-1* (also known as *PDCD1*), *LAG3*, *TIGIT* and *HAVCR2*[20]. *TOX*, which encodes a chromatin modifier important for exhaustion programs of chronic virus-specific and tumour-specific T cells in mouse models[21,22], was only marginally increased in MANA-specific cells, whereas its homologue, *TOX2*, which has also been reported to drive T cell exhaustion[23], showed much higher upregulation in MANA-specific versus EBV-specific CD8 TIL. *HOBIT*, which is selectively upregulated in TRM T cells[24], was also upregulated in MANA-specific TIL, even relative to influenza-specific TRM (Fig. 2e). Indeed, MANA-specific T cells demonstrated the highest immune checkpoint and exhaustion signatures[17] (Extended Data Fig. 6f). These findings demonstrate that MANA-specific CD8 T cells in the tumour have an unconventional hybrid transcriptional program characterized by incomplete activation of effector programs and significant upregulation of checkpoint molecules such as PD-1, CTLA-4, TIM3, TIGIT and CD39. Genes encoding each of these checkpoint molecules were more highly expressed among MANA-specific CD8 cells than either influenza- or EBV-specific CD8 cells, with *CD39* being the most highly differentially expressed (Fig. 2d, e), congruent with previous flow cytometry findings on MANA-specific lung cancer TIL[2].

Influenza-specific TRM were distinguished from MANA-specific TRM by low levels of both activation and effector CTL programs and had lower expression of multiple checkpoint molecules, but had the highest levels of genes associated with T memory stem cells, such as TCF7 and IL-7R (Fig. 2e, f). Indeed, IL-7R expression was 4.6-fold higher on influenza-specific TIL relative to MANA-specific TIL. In TIL

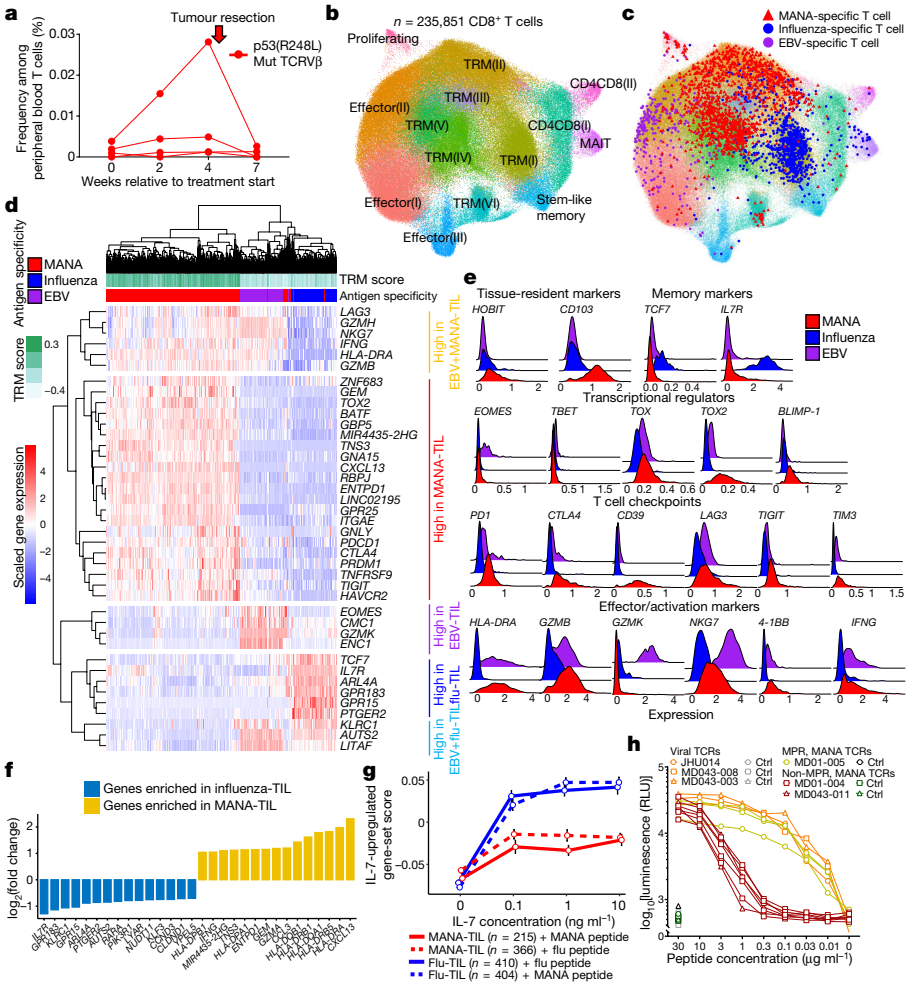


Fig. 2 | Characterization of antigen-specific T cells in NSCLC treated with neoadjuvant PD-1 blockade. The MANAFEST assay was performed on four patients with MPR and five patients without MPR. Results are shown in Extended Data Fig. 2 and Supplementary Data 5. **a**, Four TCRs recognizing p53(R248L)-derived MD01-004-MANA12 were identified in patient without MPR MD01-004. Their frequency was tracked in serial peripheral blood. Mut, mutant. **b**, Refined clustering was performed on 235,851 CD8$^+$ T cells from tumour (n = 15), adjacent normal lung (n = 12), TDLN (n = 3) and one resected brain metastasis (MD043-011). Fourteen unique clusters were visualized and were using T cell gene programs described in previous studies[16]. Cluster-defining genes are shown in Extended Data Fig. 5a. **c**, MANA-specific (red), influenza-specific (blue) and EBV-specific (purple) clonotypes were visualized on the CD8 UMAP. **d**, Antigen-specific gene programs in the TIL were visualized as a heat map. Comparisons were performed at the individual cell level using a two-sided Wilcoxon rank-sum test with P-value adjustment using Bonferroni correction. **e**, Expression levels of key markers are shown. *TBET* is also known as *TBX21*; *4-1BB* is also known as *TNFRSF9*. **f**, Transcriptional programs of influenza-specific and MANA-specific TIL were compared. The top-15 significantly upregulated genes in influenza -specific T cells (blue) and in MANA-specific T cells (yellow) are shown. **g**, TIL from MD01-004 were cultured with MD01-004-MANA-12 or influenza peptide and titrating concentrations of IL-7, followed by scRNA-seq–TCR-seq. In total, 814 influenza-specific TIL (410 co-cultured with influenza peptide and 404 co-cultured with MANA peptide) and 581 MANA-specific TIL (366 co-cultured with influenza peptide and 215 co-cultured with MANA peptide) were detected from a single experiment and were analysed. Composite expression of an IL-7 gene set by influenza-specific and MANA-specific TIL (as determined by their TCRVβ CDR3) was analysed. Dose–response curve of the IL-7-upregulated gene set-score is shown (mean ± s.e.m.). **h**, TCRs corresponding to seven MANA-specific clonotypes from two patients without MPR (red lines), three MANA-specific clonotypes from a patient with MPR (yellow lines), two influenza-specific TCRs, and one EBV-specific TCR (orange lines) were tested for ligand-dependent TCR-signalling capacity. Ctrl, control; RLU, relative luminescence units.

obtained from patient without MPR MD01-004, culture with titrating concentrations of IL-7 in vitro induced much higher levels of IL-7R-regulated genes (Supplementary Data 2.3) in influenza-specific TIL than in MANA-specific TIL (Fig. 2g, Extended Data Fig. 7). Nonetheless, supraphysiological levels of IL-7 induced appreciable upregulation of IL-7R-induced genes in MANA-specific TIL. Given the distinct transcriptional programs of the identified MANA-specific CD8 cells, we hypothesized that other CD8 T cells in the same TRM cluster showing differential expression relative to influenza-specific T cells (Fig. 2g) may also recognize MANA that were not detected by the MANAFEST assay. We cloned seven TCRs corresponding to CD8$^+$ T cells with highly differential gene expression relative to influenza-specific T cells. We screened each TCR with a library of candidate MANA (Supplementary Table 6) and confirmed MANA recognition in three of these TCRs, one TCR each from patients MD01-004, MD01-005 and MD043-011 (Extended Data Fig. 8a–d).

To next investigate the ligand-dependent TCR signalling capacity of antigen-specific T cells, we performed a dose–response curve with cognate peptides matched to the ten total Jurkat-validated MANA-specific TCRα–TCRβ pairs (Supplementary Table 10). Peptide dose–response curves of MPR-derived TCRs were comparable to those of EBV- and influenza-specific TCRs, suggesting that these TCRs were capable of strong ligand-dependent signalling (sometimes referred to as functional avidity). However, the peptide dose–response curves of TCRs

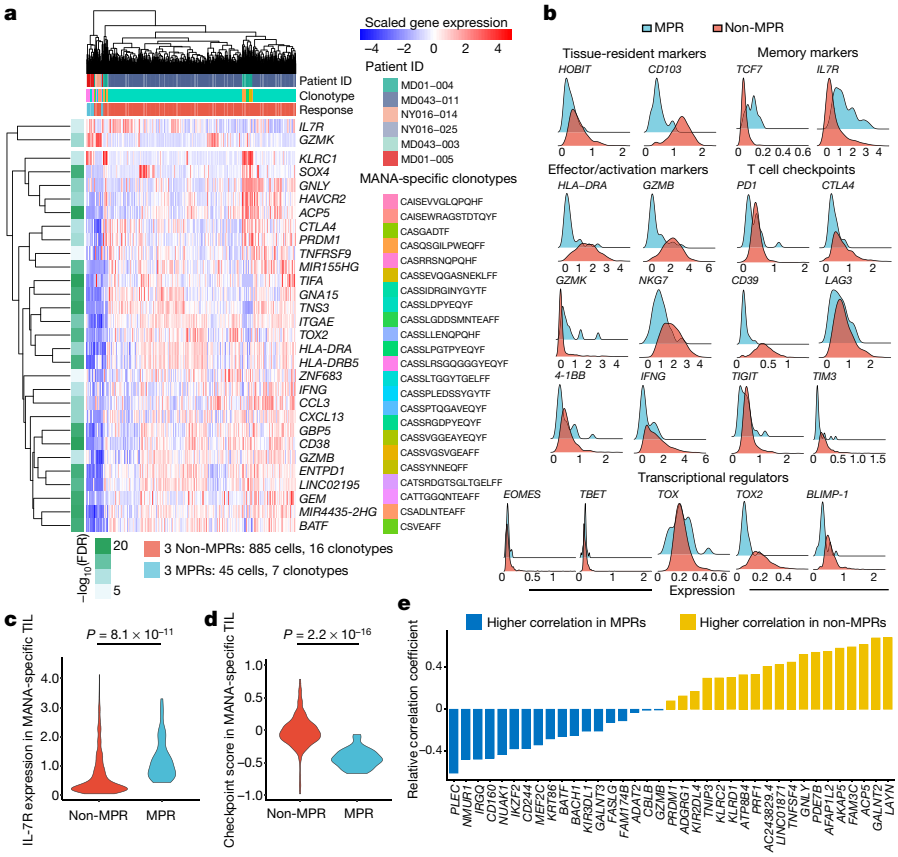


Fig. 3 | Differential gene expression programs of MANA-specific CD8 T cells in MPR versus non-MPR tumours. Seven unique MANA-specific clonotypes, representing 45 total transcriptomes, were identified in MPR TIL: 39 from MD01-005, 2 from MD043-003 and 4 from NY016-025. In non-MPR TIL, 16 unique clonotypes, representing 885 total transcriptomes, were identified: 782 from MD043-011, 62 from MD01-004 and 22 from NY016-014 (Supplementary Table 8). Differential gene expression analysis was performed on the MANA-specific T cells detected in MPR ($n = 3$) and non-MPR ($n = 3$) tumours. **a**, The top differential genes and selective immune markers of tumour-infiltrating MANA-specific T cells from MPR and non-MPR tumours. Comparisons were performed at the individual cell level using two-sided Wilcoxon rank sum test. *P*-value adjustment was performed using Bonferroni correction. Side bar shows the adjusted *P* value (green scale) and response status (red, TIL from MPR; light blue, TIL from non-MPR). **b**, Histograms show the expression of key genes among MANA-specific T cells from MPR (light blue) and non-MPR (red) tumours. **c**, A violin plot shows IL-7R expression by each MANA-specific CD8 T cell in MPR (red) and non-MPR (light blue) tumours. Comparisons were performed at the individual cell level using two-sided Wilcoxon rank-sum test. **d**, A T cell immune checkpoint score was calculated for each MANA-specific CD8 T cell detected in MPR (red) and non-MPR (light blue) tumours. This checkpoint score was compared between MPR and non-MPR using two-sided Wilcoxon rank-sum test. **e**, The relative correlation coefficient (MPR MANA-specific TIL versus non-MPR MANA-specific TIL) with the immune checkpoint score is shown for genes more highly correlated in non-MPR (yellow) and MPR (blue) TIL.

derived from patients without MPR were markedly lower (approximately 2 $\log_{10}$ leftward shift in peptide dose–response curve) (Fig. 2h, Extended Data Fig. 8e). Together, our data show that despite similar measured MANA–HLA binding affinities (Extended Data Fig. 4c, d), TCR from expandable MANA-specific clones from the patient with MPR had significantly higher functional avidity than MANA-specific clones from patients without MPR.

MANA-specific TIL programs correlate with MPR

To explore determinants of ICB sensitivity, we examined differences in gene expression patterns between MPR and non-MPR MANA-specific TIL. The neoadjuvant clinical trial format enabled us to make this distinction through pathological analysis of surgically resected tissue. In total, we compared 45 MPR TIL transcriptomes (39 from MD01-005, 2 from MD043-003 and 4 from NY016-025) with 885 non-MPR TIL transcriptomes (782 from MD043-011, 62 from MD01-004 and 22 from NY016-014; Extended Data Fig. 9, Supplementary Table 8). We observed highly significant differences between pathologic MPR and non-MPR tumours (Fig. 3a, Supplementary Data 1.4). Significantly higher levels of genes associated with T cell dysfunction such as *TOX2*, *CTLA4*, *HAVCR2* and *ENTPD1* were observed for non-MPR MANA-specific T cells, whereas MPR MANA-specific T cells had higher expression of genes associated with memory (*IL7R* and *TCF7*) and effector function (*GZMK*) (Fig. 3a–c). Both the checkpoint score and exhaustion score were higher in MANA-specific TIL from patients without MPR (Fig. 3d, Extended Data Fig. 10a, b). Of note, *CXCL13* is one of the genes most highly correlated with checkpoint-associated genes in non-MPR MANA-specific TIL, and was also found to be highly expressed in MANA-specific cells relative to virus-specific cells among CD8 TIL (Fig. 2d–f).

A number of genes encoding T cell inhibitory molecules were more highly correlated with a composite immune checkpoint score of MANA-specific TIL from patients without MPR than those from patients with MPR (Fig. 3e, Extended Data Fig. 10c). In two patients without MPR (MD01-004 and MD043-011) and one patient with MPR (MD01-005), we also detected MANA-specific cells upon single-cell

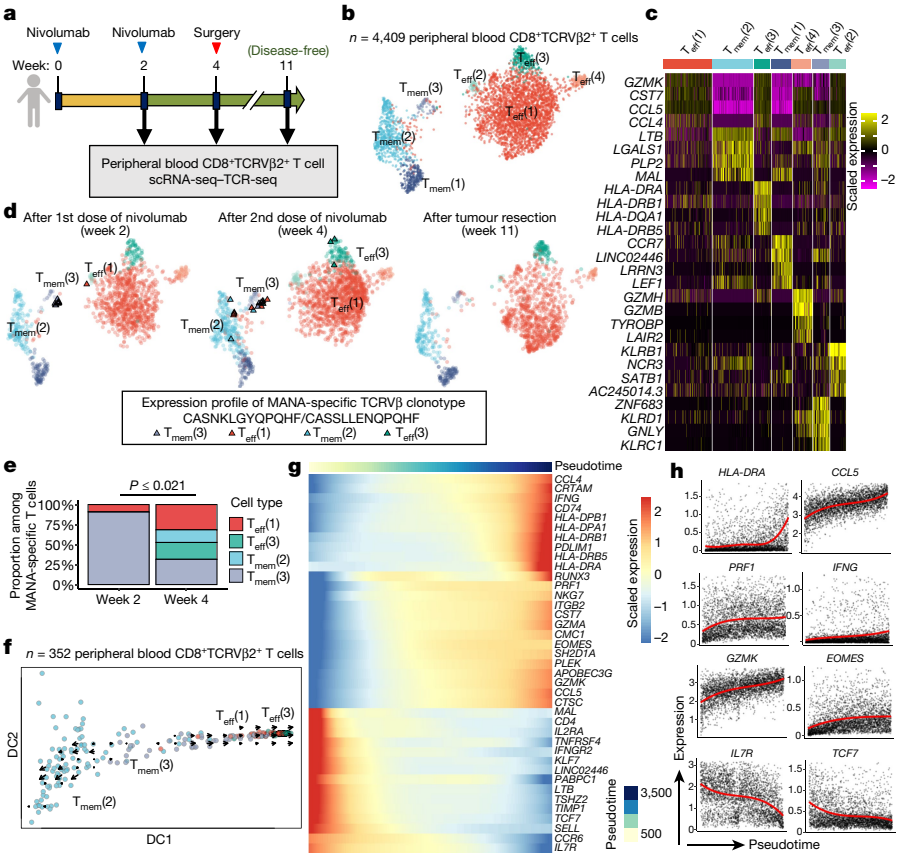


Fig. 4 | Neoadjuvant PD-1 blockade promotes systemic transcriptional reprogramming in MANA-specific T cells from a patient with complete pathologic response. a, Longitudinal peripheral blood mononuclear cells were collected from complete pathologic responder MD01-005 (0% residual tumour) during treatment and in post-surgery follow up. Peripheral blood $CD8^+$ T cells were sorted using FACS on the basis of expression of TCRVβ2, which corresponds to the MANA-specific CDR3 CASNKLGYQPQHF, as identified by the MANAFEST assay (Extended Data Fig. 2a). scRNA-seq–TCR-seq was performed on the sorted population from each time point. **b**, UMAP projection of expression profiles of 4,409 peripheral blood $CD8^+TCRV\beta2^+$ T cells. **c**, Heat map of the top-5 differential genes, ranked by average fold change, for each T cell cluster. **d**, UMAP projection of MANA-specific T cells, identified via the CASNKLGYQPQHF or CASSLLENQPQHF TCRVβ CDR3, is shown for each time point. Clusters were coloured using the same colour scheme as in **b**. MANA-specific T cells are highlighted as triangles. **e**, The proportions of cells in each T cell cluster among all MANA-specific cells identified at week 2 and week 4 were compared (two-sided Fisher's exact test and a two-sided test accounting for background cell proportion, both smaller than 0.021; Methods). **f**, Diffusion plot with RNA velocity for clusters in which MANA-specific T cells were detected. Cells were randomly downsampled to 100 cells (or all cells in the cluster if cluster size was smaller than 100 cells) for each cluster for visualization. **g**, Heat map of the top differential genes along the pseudotime trajectory from $T_{mem}(3)$ to $T_{eff}(3)$. **h**, Pseudotemporal expression of genes that significantly change along the pseudotime from $T_{mem}(3)$ to $T_{eff}(3)$. Red curves represent the mean temporal function estimates of the three samples from this individual (Methods). Cells with gene expression levels above the top one percentile were removed as outliers.

profiling of CD8 T cells from TDLN (Extended Data Fig. 10d, e). Tracking the MANA-specific CD8 clonotypes from the primary tumour, we detected those clones among TIL from a brain metastasis resected from patient MD043-011 24 months after primary tumour resection (Extended Data Fig. 10f). Relative to the primary tumour, even-higher levels of three checkpoints—LAG3, TIGIT and HAVCR2—were expressed on MANA-specific TIL in the metastasis (Extended Data Fig. 10g, Supplementary Data 1.5).

Going back to overall TIL transcriptomic patterns, we hypothesized that MANA-specific T cells and/or a MANA-specific T cell-like signature might correlate with response to ICB, even though total TIL single-cell transcriptomic patterns did not (Fig. 1e). Among CD8 TIL from six MPR tumours and nine non-MPR tumours, the greatest correlation with pathologic response status was observed by combining four TIL clusters most highly enriched in MANA-specific cells, whereas the expression profile of total CD8 TIL did not distinguish MPR from non-MPR (Extended Data Fig. 11). These data suggest that additional T cells with this profile may contribute to the anti-tumour response.

Systemic reprogramming of MANA-specific T cells

We next performed scRNA-seq–TCR-seq of serial peripheral blood T cells from patient with MPR MD01-005 after enriching for expression of the TCR-Vβ genes corresponding to this patient's MANA-specific TCRs using fluorescence-activated cell sorting (FACS) (Fig. 4a–c, Extended Data Fig. 12a). Nine out of ten MANA-specific clones mapped to a TRM-like cluster ($T_{mem}(3)$; T_{mem}, memory T cell), with some transcriptional features of TRM, such as expression of *HOBIT*) two weeks after the initiation of anti-PD-1 treatment (Fig. 4d). By four weeks (time of tumour resection), a significant diversification of phenotype was observed ($P \leq 0.021$; Methods). Half of the MANA-specific cells were in T_{eff} clusters (Fig. 4e). By 11 weeks (7 weeks after tumour resection), the MANA-specific cells were below the limit of detection in the blood, consistent with known TRM patterns in the peripheral blood[25]. Using RNA velocity, we observed a clear bidirectional flow of TRM-like memory MANA-specific T cells in the $T_{mem}(3)$ cluster towards either an activated effector ($T_{eff}(3)$) or a $T_{mem}(2)$ transcriptional profile (Fig. 4f). Genes associated with T_{eff} cell function and activation, T cell homing and migration, and tissue retention were

upregulated along the pseudotime from $T_{mem}(3)$ to $T_{eff}(3)$, whereas there was a decrease in genes associated with resting memory T cells (Fig. 4g, h). Gene Ontology (GO) analysis revealed significant enrichment of an IFNγ-mediated signalling pathway along the differentiation trajectory (Extended Data Fig. 12b–f). Although all these tissue compartments were only available for one MPR, these findings are consistent with our hypothesis that, upon activation, functional effector MANA-specific T cells enter the blood and traffic into tissues, including normal lung, in search of micro-metastatic tumour[26], and are compatible with a previous study showing that TRM cell plasticity can influence systemic memory T cell responses[27].

Discussion

Here we describe the transcriptional programming of MANA-specific TIL after ICB in lung cancer, and further, differential gene programs between patients whose tumours show MPR versus those that do not. Using the MANAFEST platform, MANA-specific CD8 T cells in peripheral blood were detected in the majority of patients who were treated with anti-PD-1; these were also found among TIL in roughly a third of these individuals. Detection of these T cells was independent of tumour response, suggesting that factors in the tumour microenvironment affecting T cell function probably contribute to anti-tumour responsiveness. Indeed, the most frequent MANA-specific clonotype, representing 782 TIL, was observed in a patient with no MPR. This tumour had dual *KRAS* and *STK11* oncogenic mutations, which are known to be highly associated with non-response to PD-1 blockade[28]. Consistent with an earlier study[2], *CD39* expression was a key difference between MANA-specific and viral-specific T cells. Among MANA-specific CD8 TIL, roughly 90% were TRM with high expression of *HOBIT* that also displayed a partial but incompletely activated T_{eff} program, along with upregulation of several targetable checkpoints in non-MPR tumours. MANA-specific T cells also express far less IL-7R relative to influenza TRM, translating functionally into poor IL-7 responsiveness. These features may all contribute to their limited tumour-specific responsiveness in contrast to anti-viral responses. Future studies are warranted to assess the diminished functional capacity of MANA-specific T cells that was suggested by the transcriptomic profiles observed in our study.

One hypothesis for the lack of ICB response in some patients is that tumour-specific T cells exhibit low activity owing to poor avidity or affinity of their TCR for its cognate peptide MHC. Our finding comparing the ligand-induced TCR signalling of three MANA-specific TCRs from MPR TIL with seven from patients without MPR supports this notion, although additional studies of this type are necessary to definitively test the hypothesis. An overall limitation of these studies is the modest number of MANA-specific cells among TIL that we were able to detect, representing three responders and three non-responders. Indeed, identification of MANA-specific cells is experimentally challenging, and only a few studies have successfully identified these cells in NSCLC[2,3,8,29,30], yet none of these profiled the transcriptome of MANA-specific T cells at single-cell resolution. Among the 930 MANA-specific transcriptomes that we identified in TIL, there was high consistency among cells from each response group in highly differential expression of key genes known to regulate T cell function. These findings inform on potential ICB combination therapies to overcome anti-PD-1 resistance that occurs even in the presence of potent MANA-specific T cells. For example, our data demonstrated reduced activation of transcriptional programs downstream of IL-7 ligation in MANA-specific TIL relative to influenza-specific TIL, but the MANA-specific TIL retain their ability to respond to supraphysiological levels of IL-7. Because IL-7 signalling is a requisite for maintenance of T cell homeostasis and long-lived memory, it is conceivable that targeting the IL-7 pathway could enhance ICB response. Our findings thus provide a platform for follow-up studies to more rigorously test the generalizability of our conclusions in the setting of resectable and metastatic NSCLC.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03752-4.

1. Tumeh, P. C. et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
2. Simoni, Y. et al. Bystander CD8$^+$ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).
3. Danilova, L. et al. The mutation-associated neoantigen functional expansion of specific T Cells (MANAFEST) assay: a sensitive platform for monitoring antitumor immunity. *Cancer Immunol. Res.* **6**, 888–899 (2018).
4. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
5. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* **377**, 2500–2501 (2017).
6. Yarchoan, M., Johnson, B. A., III, Lutz, E. R., Laheru, D. A. & Jaffee, E. M. Targeting neoantigens to augment antitumour immunity. *Nat. Rev. Cancer* **17**, 209–222 (2017).
7. Scheper, W. et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat. Med.* **25**, 89–94 (2019).
8. Forde, P. M. et al. Neoadjuvant PD-1 blockade in resectable lung cancer. *N. Engl. J. Med.* **378**, 1976–1986 (2018).
9. Blank, C. U. et al. Neoadjuvant versus adjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma. *Nat. Med.* **24**, 1655–1661 (2018).
10. Huang, A. C. et al. A single dose of neoadjuvant PD-1 blockade predicts clinical outcomes in resectable melanoma. *Nat. Med.* **25**, 454–461 (2019).
11. Kaunitz, G. J. et al. Melanoma subtypes demonstrate distinct PD-L1 expression profiles. *Lab. Invest.* **97**, 1063–1071 (2017).
12. Topalian, S. L. et al. Neoadjuvant nivolumab for patients with resectable Merkel cell carcinoma in the CheckMate 358 Trial. *J. Clin. Oncol.* **38**, 2476–2487 (2020).
13. Clarke, J. et al. Single-cell transcriptomic analysis of tissue-resident memory T cells in human lung cancer. *J. Exp. Med.* **216**, 2128–2149 (2019).
14. Bagaev, D. V. et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res.* **48**, D1057–D1062 (2020).
15. Dykema, A. G. et al. Functional characterization of CD4$^+$ T cell receptors crossreactive for SARS-CoV-2 and endemic coronaviruses. *J. Clin. Invest.* **131**, 146922 (2021).
16. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8$^+$ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).
17. 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).
18. Pizzolla, A. et al. Influenza-specific lung-resident memory T cells are proliferative and polyfunctional and maintain diverse TCR profiles. *J. Clin. Invest.* **128**, 721–733 (2018).
19. Pearce, E. L. et al. Control of effector CD8$^+$ T cell function by the transcription factor eomesodermin. *Science* **302**, 1041–1043 (2003).
20. Chihara, N. et al. Induction and transcriptional regulation of the co-inhibitory gene module in T cells. *Nature* **558**, 454–459 (2018).
21. Scott, A. C. et al. TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).
22. Yao, C. et al. Single-cell RNA-seq reveals TOX as a key regulator of CD8$^+$ T cell persistence in chronic infection. *Nat. Immunol.* **20**, 890–901 (2019).
23. Seo, H. et al. TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8$^+$ T cell exhaustion. *Proc. Natl Acad. Sci. USA* **116**, 12410–12415 (2019).
24. Mackay, L. K. et al. Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459–463 (2016).
25. Kok, L. et al. A committed tissue-resident memory T cell precursor within the circulating CD8$^+$ effector T cell pool. *J. Exp. Med.* **217**, e20191711 (2020).
26. Topalian, S. L., Taube, J. M. & Pardoll, D. M. Neoadjuvant checkpoint blockade for cancer immunotherapy. *Science* **367**, eaax0182 (2020).
27. Behr, F. M. et al. Tissue-resident memory CD8$^+$ T cells shape local and systemic secondary T cell responses. *Nat. Immunol.* **21**, 1070–1081 (2020).
28. Skoulidis, F. et al. *STK11/LKB1* mutations and PD-1 inhibitor resistance in *KRAS*-mutant lung adenocarcinoma. *Cancer Discov.* **8**, 822–835 (2018).
29. Smith, K. N. et al. Persistent mutant oncogene specific T cells in two patients benefitting from anti-PD-1. *J. Immunother. Cancer* **7**, 40 (2019).
30. Anagnostou, V. et al. Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer. *Cancer Discov.* **7**, 264–276 (2017).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Patients and biospecimens

This study was approved by the Institutional Review Boards (IRB) at Johns Hopkins University (JHU) and Memorial Sloan Kettering Cancer Center and was conducted in accordance with the Declaration of Helsinki and the International Conference on Harmonization Good Clinical Practice guidelines. The patients described in this study provided written informed consent. All biospecimens were obtained from patients with stage I-IIIA NSCLC who were enrolled to a phase II clinical trial evaluating the safety and feasibility of administering two doses of anti-PD-1 (nivolumab) before surgical resection. Pathological response assessments of primary tumours were reported previously[8,31]. Tumours with no more than 10% residual viable tumour cells were considered to have a MPR.

scRNA-seq–TCR-seq

Cryobanked T cells were thawed and washed twice with pre-warmed RPMI with 20% FBS and gentamicin. Cells were resuspended in PBS and stained with a viability marker (LIVE/DEAD Fixable Near-IR; ThermoFisher) for 15 min at room temperature in the dark. Cells were the incubated with Fc block for 15 min on ice and stained with antibody against CD3 (BV421, clone SK7) for 30 min on ice. After staining, highly viable $CD3^+$ T cells were sorted into 0.04% BSA in PBS using a BD FACSAria II Cell Sorter. Sorted cells were manually counted using a hemocytometer and prepared at the desired cell concentration (1,000 cells per µl), when possible. The Single Cell 5′ V(D)J and 5′ DGE kits (10X Genomics) were used to capture immune repertoire information and gene expression from the same cell in an emulsion-based protocol at the single-cell level. Cells and barcoded gel beads were partitioned into nanolitre-scale droplets using the 10X Genomics Chromium platform to partition up to 10,000 cells per sample followed by RNA capture and cell-barcoded cDNA synthesis using the manufacturer's standard protocols. Libraries were generated and sequenced on an Illumina NovaSeq instrument using 2 × 150-bp paired end sequencing. 5′ VDJ libraries were sequenced to a depth of ~5,000 reads per cell, for a total of 5 million to 25 million reads. The 5′ DGE libraries were sequenced to a target depth of ~50,000 reads per cell.

Whole-exome sequencing, mutation calling and neoantigen prediction

Genomic data for most individuals in our study were reported previously[8], and whole-exome sequencing, variant calling and neoantigen predictions for individuals MD043-003 and NY016-025 were performed prospectively for the present study. Whole-exome sequencing was performed on pre-treatment tumours unless otherwise noted (Supplementary Table 4) and matched normal samples. DNA was extracted from tumours and matched peripheral blood using the Qiagen DNA kit (Qiagen). Fragmented genomic DNA from tumour and normal samples was used for Illumina TruSeq library construction (Illumina) and exonic regions were captured in solution using the Agilent SureSelect v.4 kit (Agilent,) according to the manufacturers' instructions as previously described[32]. Paired-end sequencing, resulting in 100 bases from each end of the fragments for the exome libraries was performed using Illumina HiSeq 2000/2500 instrumentation (Illumina). The depth of total and distinct coverage is shown in Supplementary Table 4. Somatic mutations, consisting of point mutations, insertions, and deletions across the whole exome were identified using the VariantDx custom software for identifying mutations in matched tumour and normal samples as previously described[32,33]. Somatic mutations, consisting of nonsynonymous single base substitutions, insertions and deletions, were evaluated for putative MHC class I neoantigens using the ImmunoSelect-R pipeline (Personal Genome Diagnostics) as previously described[30]. Somatic sequence alterations are listed in Supplementary Table 5.

Identification of neoantigen-specific TCRVβ CDR3 clonotypes

We used the MANAFEST assay[3] to evaluate T cell responsiveness to MANA and viral antigens. In brief, pools of MHC class I-restricted CMV, EBV and influenza peptide epitopes (CEFX, jpt Peptide Technologies), pools representing the matrix protein and nucleoprotein from H1N1 and H3N2 (jpt Peptide Technologies), and putative neoantigenic peptides defined by the ImmunoSelect-R pipeline (jpt Peptide Technologies; Supplementary Table 6) were each used to stimulate 250,000 T cells in vitro for 10 days as previously described[3]. The time point of peripheral blood collection used for each MANAFEST assay is described in Supplementary Tables 2, 7. In brief, on day 0, T cells were isolated from PBMC by negative selection (EasySep; STEMCELL Technologies). The T cell-negative fraction was co-cultured with an equal number of selected T cells in culture medium (IMDM/5% human serum with 50 µg ml^{-1} gentamicin) with 1 µg ml^{-1} relevant neoantigenic peptide, 1 µg ml^{-1} of an MHC class I-restricted CMV, EBV, and influenza peptide epitope pool (CEFX, jpt Peptide Technologies), 1 µg ml^{-1} of pools representing the matrix protein and nucleoprotein from H1N1 and H3N2 (jpt Peptide Technologies), or no peptide (to use as a reference for non-specific or background clonotypic expansion). On day 3, half the medium was replaced with fresh medium containing cytokines for a final concentration of 50 IU ml^{-1} IL-2 (Chiron), 25 ng ml^{-1} IL-7 (Miltenyi) and 25 ng ml^{-1} IL-15 (PeproTech). On day 7, half the medium was replaced with fresh culture medium containing cytokines for a final concentration of 100 IU ml^{-1} IL-2 and 25 ng ml^{-1} IL-7 and IL-15. On day 10, cells were harvested, washed twice with PBS, and the $CD8^+$ fraction was isolated using a $CD8^+$ negative enrichment kit (EasySep; STEMCELL Technologies). DNA was extracted from each CD8-enriched culture condition using the Qiamp micro-DNA kit according to the manufacturer's instructions. TCR sequencing was performed on each individual peptide-stimulated T cell culture using survey-level sequencing (max depth ~60,000 reads) by Adaptive Biotechnologies using their established platform[34] or by the Sidney Kimmel Comprehensive Cancer Center FEST and TCR Immunogenomics Core (FTIC) facility using the Oncomine TCR Beta short-read assay (Illumina) and sequenced on an Illumina iSeq 100 using unique dual indexes, for a maximum of ~40,000 reads per sample.

Data pre-processing was performed to eliminate non-productive TCR sequences and to align and trim the nucleotide sequences to obtain only the CDR3 region. Sequences not beginning with C or ending with F or W and having less than seven amino acids in the CDR3 were eliminated. TCR sequencing samples with less than 1,000 productive reads were excluded from downstream analysis. MD043-011-MANA_22 was the only such sample in the present study (see Supplementary Table 7). Resultant processed data files were uploaded to our publicly available MANAFEST analysis web app (http://www.stat-apps.onc.jhmi.edu/FEST) to bioinformatically identify antigen-specific T cell clonotypes.

Bioinformatic analysis of productive clones was performed to identify antigen-specific T cell clonotypes meeting the following criteria: (1) significant expansion (Fisher's exact test with Benjamini–Hochberg correction for false discovery rate (FDR), $P < 0.05$) compared to T cells cultured without peptide, (2) significant expansion compared to every other peptide-stimulated culture (FDR <0.05) except for conditions stimulated with similar neoantigens derived from the same mutation, (3) an odds ratio >5 compared to the no peptide control, and (4) present in at least 10% of the cultured wells to ensure adequate distribution among culture wells. A lower read threshold of 300 was used for assays sequenced by the FTIC and a lower threshold of 30 was used for samples sequenced by Adaptive Biotechnologies. In MANAFEST assays testing less than 10 peptides or peptide pools, cultures were

performed in triplicate and reactive clonotypes were defined as being significantly expanded relative to T cells cultured without peptide (FDR <0.05) in two out of three triplicates, and not significantly expanded in any other well tested. When available, TCRseq was also performed on DNA extracted from tumour, normal lung, and lymph node tissue obtained before treatment and at the time of surgical resection, as well as serial peripheral blood samples. The assays performed on each biospecimen are outlined in Supplementary Table 2.

Peptide affinity and stability measurements

Peptide affinity for cognate HLA molecules was assessed using a luminescent oxygen channeling immunoassay (LOCI; AlphaScreen, Perkin Elmer) as previously described[35]. This is a proximity-based system using a donor and acceptor bead, each conjugated with an epitope tag. When the donor bead is excited with light at 650 nm and can activate an acceptor bead, resulting in a signal at 520–620 nm, which can be quantified per second as a surrogate of affinity. A higher number of counts per second indicates higher affinity of the peptide:HLA pair. The stability of peptide loaded complexes was measured by refolding MHC with peptide and subsequently challenging complexes with a titration of urea. The denaturation of MHC was monitored by ELISA as described previously[36].

TCR reconstruction and cloning

Ten MANAFEST+ TCR sequences for which the TCRα chain could be enumerated (>3 cells in single-cell data with the same TCRα–TCRβ pair) were selected for cloning. In addition, seven clones (from three individuals: MD01-004, MD01-005 and MD043-011) that have high composite signature (using the AddModuleScore function) consisting of differential gene programs of MANA-specific T cell relative to influenza-specific T cells in the TRM were selected for cloning. Relevant TCRs were analysed with the IMGT/V-Quest database (http://www.imgt.org). The database allows us to identify the TRAV and TRBV families with the highest likelihood to contain the identified segments which match the sequencing data. To generate the TCRs, the identified TCRA V-J region sequences were fused to the human TRA constant chain, and the TCRB V-D-J regions to the human TRB constant chain. The full-length TCRA and TCRB chains were then synthesized as individual gene blocks (IDT) and cloned into the pCI mammalian expression vector, containing a CMV promoter, and transformed into competent *Escherichia coli* cells according to the manufacturer's instructions (NEBuilder HiFi DNA Assembly, NEB). Post transformation and plasmid miniprep, the plasmids were sent for Sanger sequencing to ensure no mutations were introduced (Genewiz).

T cell transfection, transient TCR expression and MANA-recognition assays

To generate a Jurkat reporter cell in which we could transfer our TCRs of interest, the endogenous TCR α- and β-chains were knocked out of a specific Jurkat line that contains a luciferase reporter driven by an NFAT response element (Promega) using the Alt-R CRISPR system (Integrated DNA Technologies, IDT). Two sequential rounds of CRISPR knockout were performed using crDNA targeting the TCRα constant region (AGAGTCTCTCAGCTGGTACA) and the TCRβ constant region (AGAAGGTGGCCGAGACCCTC). Limiting dilution was then used to acquire single cell clones and clones with both TCRα and TCRβ knocked out, as confirmed by Sanger sequencing and restoration of CD3 expression only by the co-transfection of TCRα or TCRβ chains, were chose. CD8α and CD8β chains were then transduced into the $TCR\alpha^{-}TCR\beta^{-}$ Jurkat reporter cells using the MSCV retroviral expression system (Clontech). Jurkat reporter cells were then co-electroporated with the pCI vector encoding the TCRB and TCRA gene blocks, respectively, using ECM830 Square wave electroporation system (BTX) at 275 V for 10 ms in OptiMem media in a 4-mm cuvette. Post electroporation, cells were rested overnight by incubating in in RPMI 10% FBS at 37 °C, 5% CO_2. TCR expression was confirmed by flow cytometric staining for CD3 on a BD FACSCelesta and 50,000 $CD3^{+}$ T cells were plated in each well of a 96-well plate. Reactivity of the TCR-transduced Jurkat cells was assessed by co-culturing with 1×10^5 autologous EBV-transformed B cells, loaded with titrating concentrations of MANA peptides, viral peptide pools or negative controls. After overnight incubation, activation of the NFAT reporter gene was measured by the Bio-Glo Luciferase Assay per manufacturer's instructions (Promega). Jurkat cells were routinely tested for mycoplasma contamination. No cell line authentication was performed.

COS-7 transfection with HLA allele and p53 plasmids

gBlocks (IDT) encoding HLA A*68:01, p53(R248L) and wild-type p53 were cloned into pcDNA3.4 vector (Thermo Fisher Scientific, A14697). COS-7 cells were transfected with plasmids at 70–80% confluency using Lipofectamine 3000 (Thermo Fisher Scientific, L3000015) and incubated at 37 °C overnight in T75 flasks. A total of 30 μg plasmid (1:1 ratio of HLA plasmid per target protein plasmid in co-transfections) was used. Post transfection, COS-7 cells were plated with TCRαβ-transfected JurkaT cells containing NFAT reporter gene at a 1:1 ratio. After overnight incubation, activation of the NFAT reporter gene was measured by the Bio-Glo Luciferase Assay per manufacturer's instructions (Promega).

Single-cell data pre-processing and quality control

Cell Ranger v3.1.0 was used to demultiplex the FASTQ reads, align them to the GRCh38 human transcriptome, and extract their cell and unique molecular identifier (UMI) barcodes. The output of this pipeline is a digital gene expression (DGE) matrix for each sample, which records the number of UMIs for each gene that are associated with each cell barcode. The quality of cells was then assessed based on (1) the number of genes detected per cell and (2) the proportion of mitochondrial gene/ribosomal gene counts. Low-quality cells were filtered if the number of detected genes was below 250 or above 3× the median absolute deviation away from the median gene number of all cells. Cells were filtered out if the proportion of mitochondrial gene counts was higher than 10% or the proportion of ribosomal genes was less than 10%. For single-cell VDJ sequencing, only cells with full-length sequences were retained. Dissociation/stress associated genes[37,38], mitochondrial genes (annotated with the prefix 'MT-'), high abundance lincRNA genes, genes linked with poorly supported transcriptional models (annotated with the prefix 'RP-')[39] and TCR (TR) genes (TRA/TRB/TRD/TRG, to avoid clonotype bias) were removed from further analysis. In addition, genes that were expressed in less than five cells were excluded.

Single-cell data integration and clustering

Seurat[40] (3.1.5) was used to normalize the raw count data, identify highly variable features, scale features, and integrate samples. PCA was performed based on the 3,000 most variable features identified using the vst method implemented in Seurat. Gene features associated with type I Interferon (IFN) response, immunoglobulin genes and specific mitochondrial related genes were excluded from clustering to avoid cell subsets driven by the above genes[39]. Dimension reduction was done using the RunUMAP function. Cell markers were identified by using a two-sided Wilcoxon rank sum test. Genes with adjusted P <0.05 were retained. Clusters were labelled based on the expression of the top differential gene in each cluster as well as canonical immune cell markers. Global clustering on all CD3 T cells and refined clustering on CD8 T cells were performed using same procedure. To select for $CD8^{+}$ T cells, SAVER[41] was used to impute dropouts by borrowing information across similar genes and cells. A density curve was fitted to the $\log_2$-transformed SAVER imputed CD8A expression values (using the 'density' function in R) of all cells from all samples. A cut-off is determined as the trough of the bimodal density curve (that is, the first location where the first derivative is zero and the second derivative is positive). All cells with $\log_2$-transformed SAVER imputed

CD8A expression larger than the cut-off are defined as $CD8^+$ T cells. TRB amino acid sequences were used as a biological barcode to match MANA, EBV or influenza A-specific T cell clonotypes identified from the FEST assay with single-cell VDJ profile and were projected onto $CD8^+$ T cell refined UMAP.

Single-cell subset pseudobulk gene expression analysis

PCA was performed on a standardized pseudobulk gene expression profile, where each feature was standardized to have a mean of zero and unit variance. In global CD3 and CD8 TIL PCA, for each cell cluster we first aggregated read counts across cells within the cluster to produce a pseudobulk expression profile for each sample and normalized these pseudobulk expression profiles across samples by library size. Combat function in the sva R package[42,43] was applied to address potential batch effects on the normalized pseudobulk profile. Highly variable genes (HVGs) were selected for each cell cluster by fitting a locally weighted scatterplot smoothing (LOESS) regression of standard deviation against the mean for each gene and identifying genes with positive residuals. For each sample, all cell clusters were then concatenated by retaining each cluster's HVGs to construct a concatenated gene expression vector consisting of all highly variable features identified from different cell clusters. Each element in this vector represents the pseudobulk expression of a HVG in a cell cluster. Samples were embedded into the PCA space based on these concatenated gene expression vectors. Canonical correlation[44,45] between the first two PCs (that is, PC1 and PC2) and a covariate of interest (that is, tissue type or response status) was calculated. Permutation test was used to assess the significance by randomly permuting the sample labels 10,000 times. In the MANA-specific PCA (Extended Data Fig. 11), MANA-enriched cell clusters, defined by clusters with MANA-specific T cell frequency at least two fold higher than randomly expected, were aggregated as one combined cell cluster. Then, a similar procedure by first identifying HVGs, computing the first 2 PCs and then calculating the canonical correlation was repeated for the combined MANA-enriched cell cluster and each of the other CD8 clusters.

Differential analysis comparing MPR and non-MPR by total CD8 or CD4 TIL and by cell cluster

The gene expression read counts were adjusted by library size. SAVER[41] was used to impute the dropouts, and further log2-transformed the imputed values after adding a pseudocount of 1. A linear mixed-effect model[46] was constructed to identify genes that are significantly differential between MPR and non-MPR among total CD8/CD4 TIL and by each cell cluster, respectively. The B-H procedure[47] was used to adjust the *P* values for multiple testing, and the statistical significance is determined using a cut-off of FDR <0.05.

Differential-expression tests and antigen-specific T cell marker genes

Differential-expression tests for antigen-specific T cells were performed using FindAllMarkers functions in Seurat with Wilcoxon rank-sum test on SAVER imputed expression values. Genes with >0.25 $\log_2$-fold changes, at least 25% expressed in tested groups, and Bonferroni-corrected p values <0.05 were regarded as significantly differentially expressed genes (DEGs). Antigen-specific (MANA versus influenza versus EBV) T cell marker genes were identified by applying the differential expression tests for upregulated genes between cells of one antigen specificity to all other antigen-specific T cells in the dataset. MANA-specific T cell genes associated with response to ICB were identified by applying the differential expression tests comparing MANA-specific T cells from MPR versus those from non-MPR. Top ranked DEGs (by log-fold changes) with a $\log_2$-fold changes >0.8 and DEGs relating to T cell function were extracted for further visualization in a heat map using pheatmap package. SAVER-imputed expression values of selective marker genes (transcriptional regulators, memory markers, tissue-resident markers, T cell checkpoints, effector and activation markers) were plotted using the RidgePlot function in Seurat.

In vitro short-term TIL stimulation with IL-7

Cryopreserved TIL from patient MD01-004 were thawed, counted, and stained with the LIVE/DEAD Fixable Aqua (ThermoFisher) viability marker and antibodies specific for CD3 (PE, clone SK1) and CD8 (BV786, clone RPA-T8). Thirty-thousand $CD8^+$ T cells per condition were sorted on a BD FACSAria II Cell Sorter into a 96-well plate. Autologous T cell-depleted PBMC were added as antigen presenting cells (APC) at 1:1 ratio. The cells were stimulated with either influenza A or MD01-004-MANA 12 peptide and titrating concentrations of recombinant human IL-7 (Miltenyi) for 12 h in a round-bottomed 96-well plate.

Gene expression analysis of IL-7-stimulated MANA- and influenza-specific TIL

Following 12 h of antigen and IL-7 stimulation, cells were spun down, counted and re-suspended in 1% BSA at desired concentration. scRNA-seq and VDJ libraries were prepared using 10X Chromium single cell platform using 5′ DGE library preparation reagents and kits according to manufacturer's protocols (10X Genomics) and as described above. MANA- or influenza-specific T cell clonotypes from the single-cell dataset were identified by using the TRB amino acid sequences as a biological barcode. SAVER imputed gene expression was scaled and centred using the ScaleData function in Seurat. A composite score for the IL-7-upregulated gene set[48] (Supplementary Data 2.3) expression was computed using the AddModuleScore function and subsequently visualized using ridgeplot. Mean ± standard error was used to show the dose–response curve of the IL-7-upregulated gene-set score by antigen-specific T cells and peptide-stimulation groups.

Immune checkpoint and exhaustion score generation and highly correlated genes

To characterize dysfunctional CD8 MANA TIL, six best-characterized (and clinically targeted) checkpoints: CTLA4, PDCD1, LAG3, HAVCR2, TIGIT and ENTPD1, were used to compute the T cell checkpoint score, and a published gene list from exhausted T cells was used to compute the T cell exhaustion score, using AddModuleScore function in Seurat. Applying T cell checkpoint score as an anchor, genes that were maximally correlated to the score were identified using linear correlation in MANA-specific TIL from MPR and non-MPR, respectively. Top-30 genes (from HVG selected using FindVariableGenes function in Seurat and excluded the 6 genes included in immune checkpoint score generation) with the highest correlation coefficients were plotted as a bar plot. The difference of correlation coefficients of the above genes was additionally computed between MPR and non-MPR and visualized using waterfall plot.

Evaluation of peripheral MANA-specific T cell transcriptome changes during treatment

Peripheral blood T cells from patient MD01-005 were sorted based on expression of CD8 and TCRVβ2, followed by scRNA-seq–TCR-seq and clustering on conventional $CD8^+$ T cells (MAIT cells excluded). To evaluate whether there was a statistically significant change in the cell types of MANA cells between week 2 (W2) and week 4 (W4) samples in Fig. 4d, e, we first conducted a Fisher's exact test, which yields a $P = 0.021$, indicating a statistically significant phenotype change in MANA-specific cells (Fig. 4e). We also conducted a more sophisticated test that adjust for potential background differences in cell type abundance between W2 and W4 samples. In this test, we let $m_{c,t}$ denote the probability that a MANA-specific T cell collected at time point t (W2 or W4) comes from cell type c, and let $p_{c,t}$ denote the proportion of all cells in time point t that come from cell type c. We evaluated the ratio $R_{c,t} = m_{c,t}/p_{c,t}$, which characterizes the relative abundance of MANA-specific T cells in each cell type. We compared the null model

where this ratio does not change over time (H_0: $R_{c,W2} = R_{c,W4}$ for all cell type c) versus the alternative model where W2 and W4 T cells have different ratios (H_1: $R_{c,W2} \neq R_{c,W4}$). To do this, we computed the test statistic $S = \sum_c (R_{c,W2} - R_{c,W4})^2$ using the observed data and compared it to its null distribution obtained using Monte Carlo simulations. To construct the null distribution for S, we pooled cells from W2 and W4 together and treated them as one sample to estimate the common ratio $R_{c,W2} = R_{c,W4} = R_c$ shared by W2 and W4, and then derived the probability that a MANA-specific T cell collected at time point t comes from cell type c under the null model H_0, which is proportional to $p_{c,t}R_c$ (that is, the product of the sample-specific background cell type proportion $p_{c,t}$ and the common MANA-abundance ratio R_c shared between samples). The MANA-specific T cells at time point t were then redistributed to different cell types randomly based on a multinomial distribution with this expected MANA-specific T cell type proportion (that is, the expected probability that a MANA-specific T cell at time point t comes from cell type c under H_0 is $p_{c,t}R_c/(\sum_{c'} p_{c',t}R_{c'})$), while keeping the total number of MANA-specific T cells at each time point the same as the observed MANA-specific T cell number at that time point. The test statistic S was then computed using this simulated sample. We repeated this simulation 10,000 times to derive the null distribution of S. Comparing the observed S to its null distribution yields a $P < 10^{-4}$.

RNA velocity-based differentiation-trajectory tracing

The RNA velocity analysis was performed by first recounting the spliced reads and unspliced reads based on aligned bam files of scRNA-seq data using the velocyto Python package. The calculation of RNA velocity values for each gene in each cell and embedding RNA velocity vector to low-dimension space were done using the SeuratWapper workflow for estimating RNA velocity using Seurat (https://github.com/satijalab/seurat-wrappers/blob/master/docs/velocity.md). The first two diffusion components from Diffusion map were used to construct the coordinates along with velocity. TSCAN (v.1.7.0) was used to reconstruct the cellular pseudotime on diffusion maps space for the PBMC T cells from three time points (samples) of one patient (MD01-005). Based on velocity analysis, the $T_{mem}(3)$ cluster was specified as the starting cluster for the pseudotemporal trajectory which has branches. For each branch, $\log_2$-transformed and library size-normalized SAVER-imputed gene expression values were used for analysing gene expression dynamics along the pseudotime. 10,325 genes with normalized expression ≥ 0.01 in at least 1% of cells were retained. For each gene g, the gene expression along pseudotime t in each sample s was described as a function $f_{gs}(t)$ which was obtained by fitting B-spline regression to the gene's normalized expression values in single cells. The red curves in Fig. 4h are the mean of the function $f_{gs}(t)$ of the three samples. In order to test whether the gene expression shows a significant change along pseudotime, we compared the above model with a null model in which $f_{gs}(t)$ is assumed to be a constant over time. The likelihood ratio statistic between the two models was computed. To determine the P value, the null distribution of the likelihood ratio statistic was constructed by permuting the pseudotime of cells in each sample, refitting the models and recomputing the likelihood ratio statistic. The P value was calculated as the number of permutations out of a total of 1,000 permutations that produce a likelihood ratio statistic larger than the observed one. The P values from all genes were converted to FDR by Benjamini–Hochberg procedure to adjust for multiple testing. Genes with FDR < 0.05 were considered as dynamic genes with statistical significance. k-Means clustering was applied to group genes with similar dynamic expression patterns into clusters. topGO (v.2.42.0) was used to identify the enriched Gene Ontology terms by comparing the genes in each cluster to all 10,325 genes as background.

Reporting summary

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

Data availability

Bulk TCRVβ sequencing data generated by Adaptive Biotechnologies are available in the Adaptive Biotechnologies ImmuneACCESS repository under DOI 10.21417/JC2021N, at https://clients.adaptivebiotech.com/pub/caushi-2021-n. Bulk TCRVβ raw and processed sequencing data generated by the Sidney Kimmel Comprehensive Cancer Center FTIC are available in the Gene Expression Omnibus with accession number GSE173351. Raw scRNA-seq–TCR-seq data reported in this paper are available in the European Genome-phenome Archive under controlled access with accession number EGAS00001005343. Owing to the personal, sensitive and inherently identifying nature of raw genomic data, access to raw RNA-seq–TCR-seq data is controlled and full instructions to apply for data access can be found at https://ega-archive.org/access/data-access. Approvals will be granted immediately upon confirmation that all requirements are met. Processed and de-identified single-cell data are available in the Gene Expression Omnibus with accession number GSE176022.

Code availability

Scripts to reproduce the analyses used in this study are available at https://github.com/BKI-immuno/neoantigen-specific-T-cells-NSCLC.

31. Cottrell, T. R. et al. Pathologic features of response to neoadjuvant anti-PD-1 in resected non-small-cell lung carcinoma: a proposal for quantitative immune-related pathologic response criteria (irPRC). *Ann. Oncol.* **29**, 1853–1860 (2018).
32. 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).
33. Jones, S. et al. Personalized genomic analyses for cancer mutation discovery and interpretation. *Sci. Transl. Med.* **7**, 283ra53 (2015).
34. Carlson, C. S. et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat. Commun.* **4**, 2680 (2013).
35. Harndahl, M. et al. Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays. *J. Biomol. Screen.* **14**, 173–180 (2009).
36. Sylvester-Hvid, C. et al. Establishment of a quantitative ELISA capable of determining peptide - MHC class I interaction. *Tissue Antigens* **59**, 251–258 (2002).
37. O'Flanagan, C. H. et al. Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol.* **20**, 210 (2019).
38. van den Brink, S. C. et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
39. Li, H. et al. Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell* **181**, 747 (2020).
40. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 747–1902 (2019).
41. Huang, M. et al. SAVER: gene expression recovery for single-cell RNA sequencing. *Nat. Methods* **15**, 539–542 (2018).
42. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
43. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
44. Hotelling, H. *Relations Between Two Sets of Variates* Vol. 28, 321–377 (Biometrika, 1936).
45. Härdle, W. K. & Simar, L. *Applied Multivariate Statistical Analysis* 1st edn (Springer-Verlag, 2003).
46. McCulloch, C. E., Searle, S. R. & Neuhaus, J. M. *Generalized, Linear and Mixed Models* 2nd edn (Wiley, 2008).
47. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
48. Belarif, L. et al. IL-7 receptor blockade blunts antigen-specific memory T cell responses and chronic inflammation in primates. *Nat. Commun.* **9**, 4483 (2018). 10.1038/s41467–018–06804-y

Acknowledgements We thank the FEST and TCR Immunogenomics Core (FTIC) and the Experimental and Computational Genomics Core (ECGC) at the Sidney Kimmel Comprehensive Cancer Center; P. F. Robbins and R. Yossef for assistance with developing TCR cloning methods; G. Pereira and A. Curry from the Johns Hopkins Upper Aerodigestive Biorepository; the Melanoma DMT Tissue Repository Team and Memorial Sloan Kettering; I. Beadles and C. Barkley for clinical support; and K. Maly, L. Hartman, R. Carlson and our respective administrative teams. Graphical images were created by BioRender.com. This work was funded by grants from the Lung Cancer Foundation of America (K.N.S.), The Mark Foundation for Cancer Research (D.M.P., J.M.T., V.E.V. and V.A.), The SU2C/Mark Foundation Lung Cancer Dream Team Convergence Award (D.M.P.), the SU2C INTIME and SU2C-LUNGevity-American Lung Association Lung Cancer Interception Dream Team (J.R.B. and V.E.V.), Bristol-Myers Squibb, the SU2C DCS International Translational Cancer Research Dream Team (SU2C-AACR-DT1415, V.E.V.), The IASLC Foundation (K.N.S.), Swim Across America (K.N.S. and V.A.), The LUNGevity Foundation (K.N.S. and V.A.), The Commonwealth Foundation (K.N.S., S.Z., D.M.P. and S.Y.), The Banks Family Foundation (M. Brock), Subsidies for Current

Expenditures to Private Institutions of Higher Education from the PMAC, through a sub-award from Juntendo University (M. Brock), the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (V.E.V.), The Virginia and D.K. Ludwig Fund for Cancer Research (B.V., K.W.K. and S.Z.), the Ludwig Center for Cancer Immunotherapy at Memorial Sloan Kettering (T.M.), the Cancer Research Institute (T.M.), the Parker Institute for Cancer Immunotherapy (T.M.), The Lustgarten Foundation for Pancreatic Cancer Research (B.V., K.W.K. and S.Z.), The Conquer Cancer Foundation of ASCO (K.A.M. and J.E.R.), Bloomberg Philanthropies, the Maryland Cigarette Restitution Fund (S.Y.), the V Foundation (V.A. and V.E.V.), the Allegheny Health Network–Johns Hopkins Research Fund (V.A. and V.E.V), the Damon Runyon Cancer Research Foundation (CI-98-18, M.D.H.), US NIH grants R37CA251447 (K.N.S.), R01HG010889 (H.J.), R01HG009518 (H.J.), R01CA121113 (V.A. and V.E.V.), R01CA217169 (J.E.C.), R01CA240472 (J.E.C.), CA62924 (B.V. and K.W.K.), T32 CA193145 (T.R.C.), T32 CA009110 (J.X.C.) and T32 GM136577 (B.V. and K.W.K.), and NIH Cancer Center Support Grants (P30 CA008748 and P30 CA006973). Stand Up To Cancer is a programme of the Entertainment Industry Foundation administered by the American Association for Cancer Research.

Author contributions A.V. and Z.J. contributed equally to this work. J.X.C., J.Z., S.Y., H.J., D.M.P. and K.N.S. conceived the project and wrote the manuscript. J.X.C., A.V., R.B., A.T., H.G., H.Y.C., D.S., M.E.A., S.T., A.G.D., B.C., L.S.C., R.W., J.M., K.S., A.G. and A.S. designed and performed functional T cell and single-cell experiments. T.R.C., J.M.T., P.B.I., E.L.B., J.H., M. Brock, J.E.R., P.M.F., J.R.B., K.A.M., J.N., B.J.P., D.R.J., M.D.H., J.E.C. and M. Bott. provided pathologic assessments, response classifications and clinical annotation. E.H.-C.H., B.J.M., K.W.K., S.Z. and B.V. designed and optimized the $CD8^+$ Jurkat cell line. V.A., V.E.V., M.L. and Z.B. performed whole-exome sequencing, mutation calls, neoantigen prediction and interpretation of genetic data. S.J. performed peptide–MHC binding and stability assays. J.Z., Z.J., B.Z., P.B., S.W., W.H., L.A., S.Y. and H.J. performed computational, bioinformatic and biostatistical analyses. T.M., D.M.P. and K.N.S. led analysis and interpretation of immunological experiments. All authors provided input on the manuscript. S.Y., H.J., D.M.P. and K.N.S. supervised all aspects of this work.

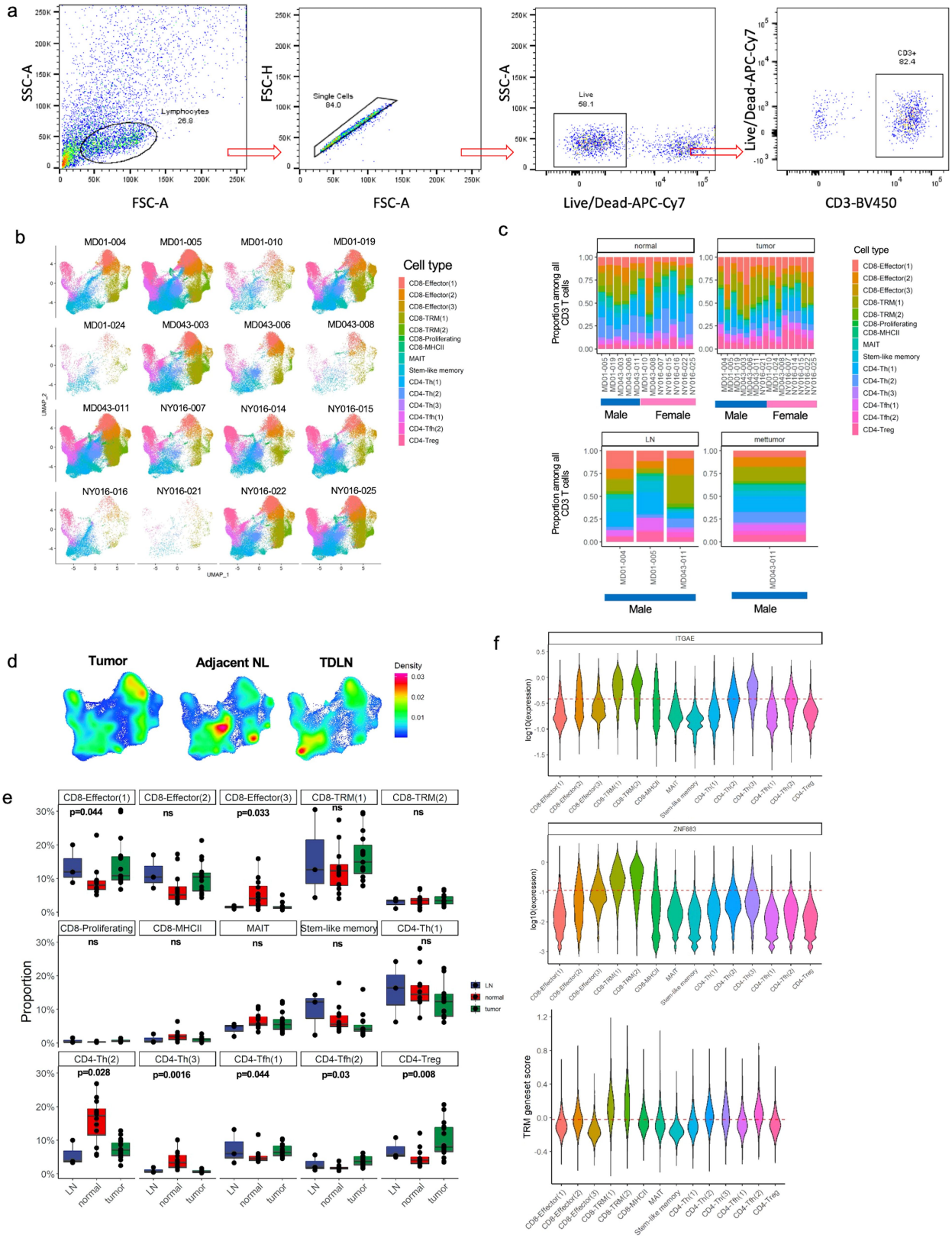
Competing interests V.A. receives research funding from Bristol-Myers Squibb and Astra Zeneca. J.M.T. receives research funding from Bristol-Myers Squibb and serves a consulting/advisory role for Bristol-Myers Squibb, Merck, and Astra Zeneca. P.B.I. receives research funding from Bristol-Myers Squibb and Erbe Elektromedizin GmbH and serves a consulting/advisory role for AstraZeneca and Veran Medical Technologies. J.N. receives research funding from AstraZeneca, Bristol-Myers Squibb, and Merck, and serves a consulting/advisory role for AstraZeneca, Daiichi Sankyo, Bristol-Myers Squibb, Merck, and Roche/Genentech. K.A.M. is a consultant for Amgen and Astra Zeneca. D.R.J. is a consultant for More Health and AstraZeneca and a Steering Committee Member for Merck. B.J.P. is a consultant for AstraZeneca and Regeneron and has received honoraria from Intuitive Surgical. J.E.C. is a consultant for AstraZeneca, Genentech, Merck, Flame Bioscience, and Novartis. V.E.V. is a founder of Delfi Diagnostics and Personal Genome Diagnostics, serves on the Board of Directors and as a consultant for both organizations, and owns Delfi Diagnostics and Personal Genome Diagnostics stock, which are subject to certain restrictions under university policy. Additionally, Johns Hopkins University owns equity in Delfi Diagnostics and Personal Genome Diagnostics. V.E.V. is an advisor to Bristol-Myers Squibb, Genentech, Merck, and Takeda Pharmaceuticals. Within the last five years, V.E.V. has been an advisor to Daiichi Sankyo, Janssen Diagnostics, and Ignyta. M.D.H. receives research support from Bristol-Myers Squibb; has been a compensated consultant for Merck, Bristol-Myers Squibb, AstraZeneca, Genentech/Roche, Nektar, Syndax, Mirati, Shattuck Labs, Immunai, Blueprint Medicines, Achilles, Arcus, and Natera; received travel support/honoraria from AstraZeneca, Eli Lilly, and Bristol-Myers Squibb; has options from Shattuck Labs, Immunai, and Arcus; and has a patent filed by his institution related to the use of tumor mutation burden to predict response to immunotherapy (PCT/US2015/062208), which has received licensing fees from PGDx. The Johns Hopkins University is in the process of filing patent applications related to technologies described in this paper on which E.H.-C.H., B.V., K.W.K. and S.Z. are listed as inventors. B.V. and K.W.K. are founders of Thrive Earlier Detection. K.W.K. is a consultant to and was on the Board of Directors of Thrive Earlier Detection. B.V., K.W.K. and S.Z. own equity in Exact Sciences. B.V., K.W.K. and S.Z. are founders of, hold or may hold equity in, and serve or may serve as consultants to manaT Bio, and hold or may hold equity in manaT Holdings, LLC. B.V., K.W.K., and S.Z. are founders of, hold equity in, and serve as consultants to Personal Genome Diagnostics. S.Z. has a research agreement with BioMed Valley Discoveries. K.W.K. and B.V. are consultants to Sysmex, Eisai, and CAGE Pharma and hold equity in CAGE Pharma. B.V. is a consultant to and holds equity in Catalio. K.W.K., B.V. and S.Z. are consultants to and hold equity in NeoPhore. The companies named above, as well as other companies, have licensed previously described technologies related to the work from this lab at Johns Hopkins University. Licenses to these technologies are or will be associated with equity or royalty payments to the inventors as well as to Johns Hopkins University. T.M. is a cofounder and holds equity in IMVAQ Therapeutics, is a consultant for Immunos Therapeutics, ImmunoGenesis, and Pfizer, receives research funding from Bristol-Myers Squibb, Surface Oncology, Kyn Therapeutics, Infinity Pharmaceuticals, Inc., Peregrine Pharmaceuticals, Inc., Adaptive Biotechnologies, Leap Therapeutics, Inc., and Aprea, and holds patents on applications related to work on oncolytic viral therapy, alpha virus-based vaccines, neoantigen modeling, CD40, GITR, OX40, PD-1, and CTLA-4. J.R.B. serves an advisory/consulting role for Amgen, AstraZeneca, Bristol-Myers Squibb, Genentech/Roche, Eli Lilly, GlaxoSmithKline, Merck, Sanofi, and Regeneron, receives research funding from AstraZeneca, Bristol-Myers Squibb, Genentech/Roche, Merck, RAPT Therapeutics, Inc., and Revolution Medicines, and is on the Data and Safety Monitoring Board of GlaxoSmithKline, Janssen, and Sanofi. P.M.F. receives research support from AstraZeneca, Bristol-Myers Squibb, Novartis, and Kyowa, and has been a consultant for AstraZeneca, Amgen, Bristol-Myers Squibb, Daichii Sankyo, and Janssen and serves on a data safety and monitoring board for Polaris. S.Y. receives research funding from Bristol-Myers Squibb/Celgene, Janssen, and Cepheid, has served as a consultant for Cepheid, and owns founder's equity in Astra Therapeutics and Digital Harmonic. K.N.S., D.M.P., B.V. and K.W.K. have filed for patent protection on the MANAFEST technology described herein (serial no. 16/341,862). K.N.S., D.M.P., J.E.C., B.V., E.H.-C.H. D.M.P., K.W.K. and S.Z. have filed for patent protection on the p53 R248L mutation-specific TCR described herein (serial no. 63/168,878). D.M.P. is a consultant for Compugen, Shattuck Labs, WindMIL, Tempest, Immunai, Bristol-Myers Squibb, Amgen, Janssen, Astellas, Rockspring Capital, Immunomic, Dracen and owns founder's equity in manaT Holdings, LLC, WindMIL, Trex, Jounce, Anara, Tizona, Tieza, RAPT and receives research funding from Compugen, Bristol-Myers Squibb, and Anara. K.N.S. has received travel support/honoraria from Illumina, Inc., receives research funding from Bristol-Myers Squibb, Anara, and Astra Zeneca, and owns founder's equity in manaT Holdings, LLC. The terms of all these arrangements are being managed by the respective institutions in accordance with their conflict-of-interest policies.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03752-4.

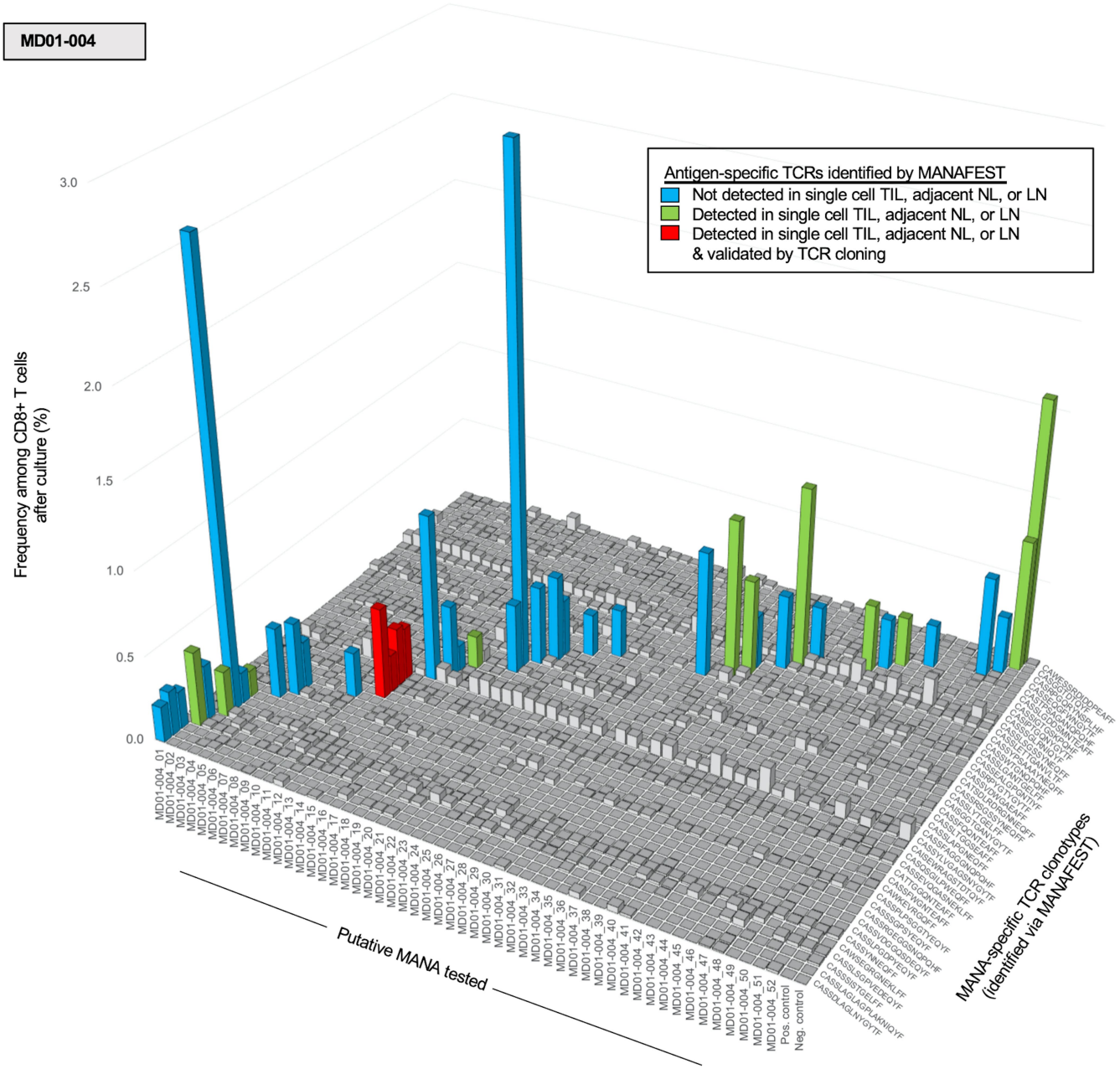
Correspondence and requests for materials should be addressed to S.Y., H.J., D.M.P. or K.N.S.

Peer review information *Nature* thanks Benny Chain, Aude Chapuis, Evan Newell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

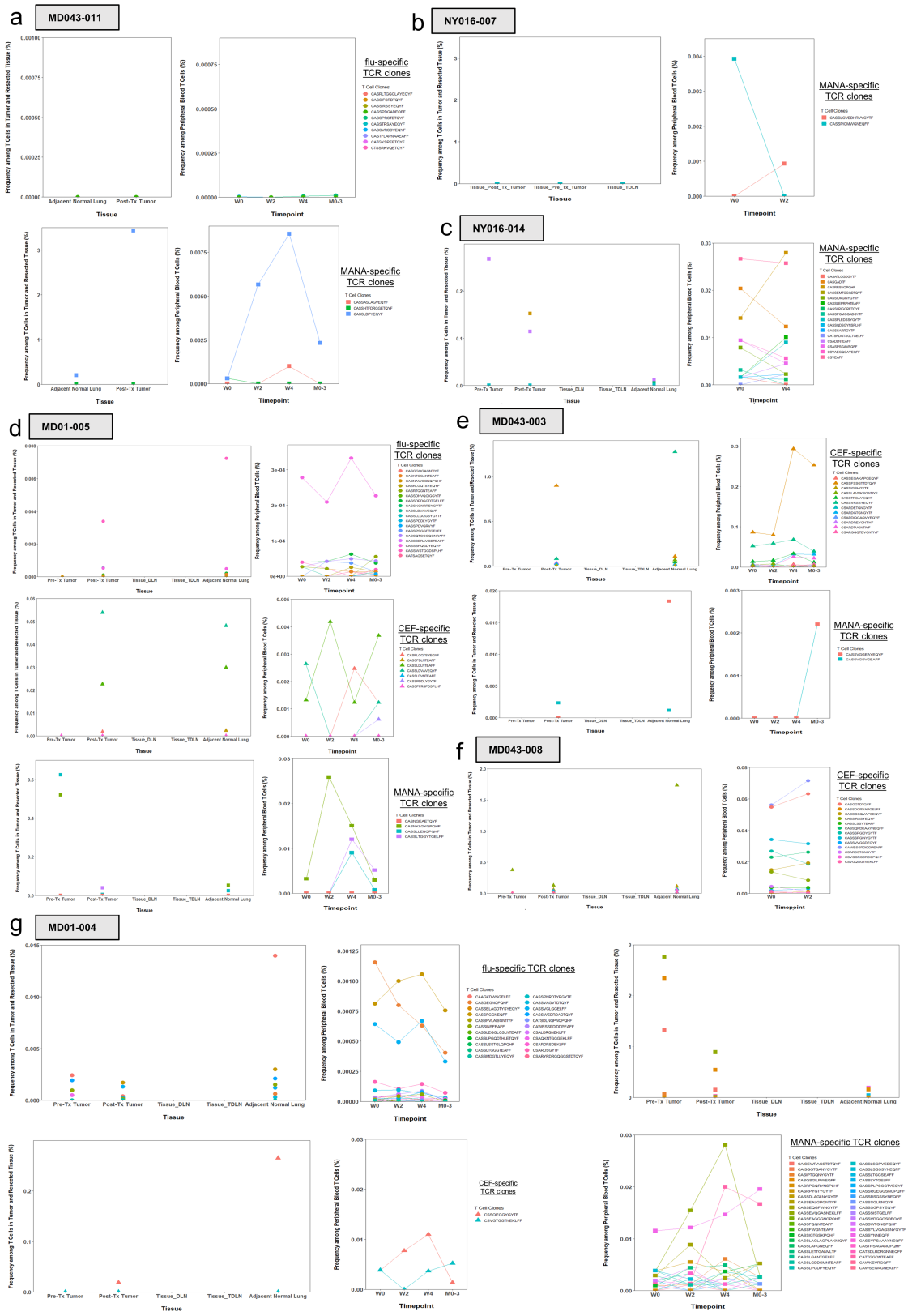


Extended Data Fig. 1 | See next page for caption.

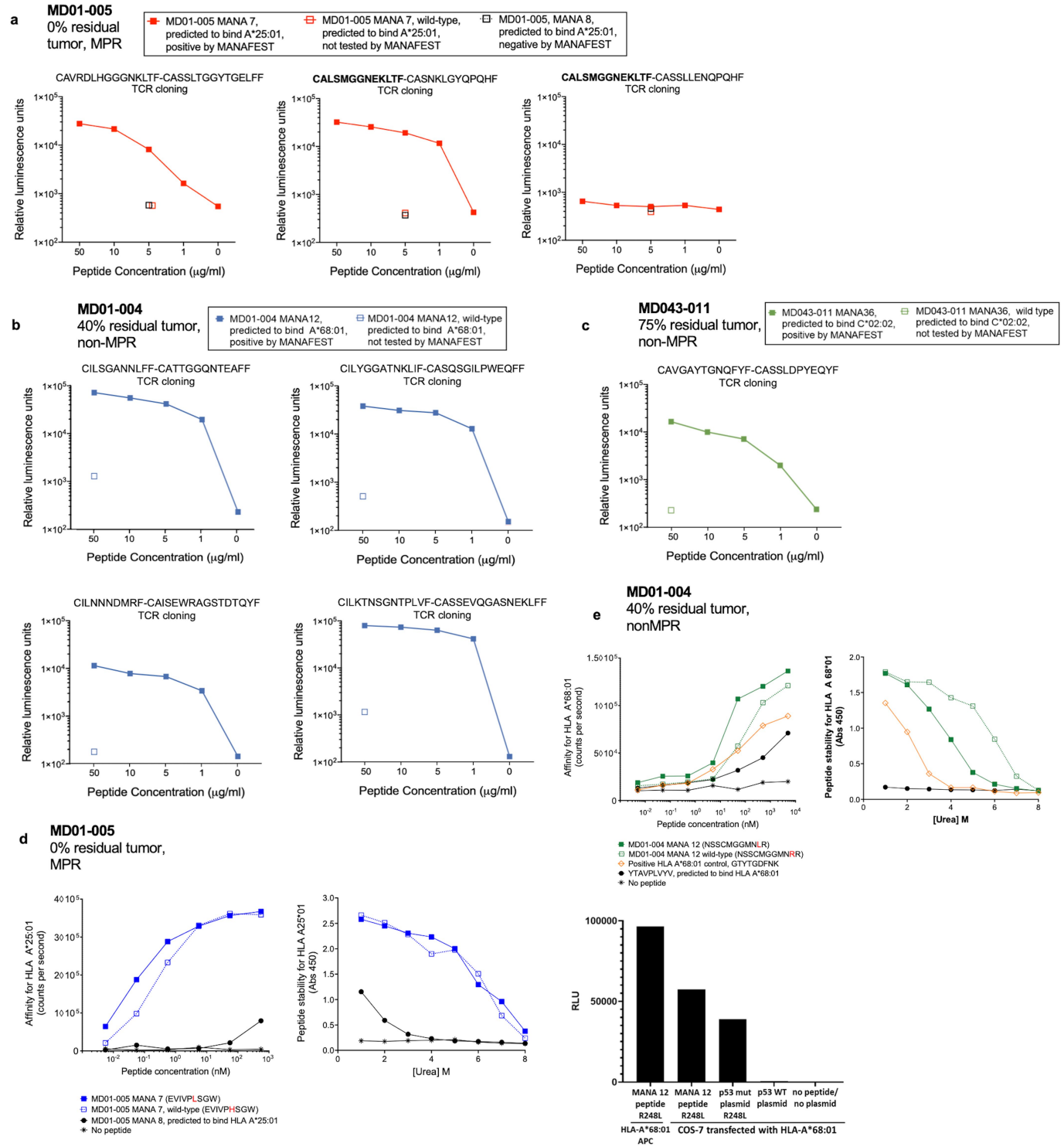
Extended Data Fig. 1 | Defining CD3$^+$ T cell subsets in patients with non-small cell cancer treated with anti-PD-1. a, FACS gating strategy for sorting CD3+ T cells. The gating strategy is shown for sorting live CD3$^+$ T cells from tumour, normal lung, lymph node, or metastasis, when available, on a BD FACSAria. **b**, Patient and tissue compartment variability across clusters on UMAP. scRNA-seq–TCR-seq was performed on available resected biospecimens (tumour, adjacent NL, TDLN, and a brain metastasis) from 16 patients treated with neoadjuvant PD-1 blockade. CD3$^+$ T cells stratified by patient are visualized using UMAP. Each cluster is annotated and marked by colour code. **c**, Barplots show the proportion of each T cell cluster in the TDLN, brain metastasis, tumour, and adjacent NL of each patient. Each cluster as shown on the UMAP is denoted by colour code. No clusters were driven by a particular patient based. **d**, A density plot of all CD3$^+$ T cells on the UMAP, stratified by tissue compartment, is shown. Cells were obtained from 15 tumours, 12 adjacent NL specimens, and 3 TDLN. Because a metastasis was sequenced in only one patient, this specimen is not included in this analysis. **e**, The proportion (%) of total CD3$^+$ T cells made up by each T cell cluster was compared between tumour (n = 15 biologically independent samples), adjacent NL (n = 12 biologically independent samples), and TDLN (n = 3 biologically independent samples). P values were obtained using Kruskal–Wallis Test and were adjusted for multiple comparisons using Benjamini-Hochberg method. Each dot represents a patient and all data points are shown. Individual data points are superimposed over a Box and Whiskers plot summarizing the data. The middle bar shows the median, with the lower and upper hinges corresponding to the 25th and 75th percentiles, respectively (interquartile range, IQR). The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. **f**, Tissue-resident defining genes and core TRM gene set signature on different T cell cluster. The top and middle violin plots show the expression of TRM-defining genes (ITGAE, ZNF683) by each cell in each cluster. The dashed line indicates the mean expression of the respective gene among all CD3$^+$ T cells. Expression values were $\log_{10}$ transformed for visualization. The bottom violin plot shows the TRM gene-set score for each cluster. This gene-set is comprised of TRM-associated genes as published previously (Supplementary Data 2.1). The dashed line shows the mean TRM gene-set score among all T cells. Because the proliferating cluster is driven by proliferation-associated genes and is comprised of mixed cell types, this cluster was not shown in the violin plots.



Extended Data Fig. 2 | MANA-specific TCRs detected in patient without MPR MD01-004 using MANAFEST and ViraFEST assays. Antigen-specific responses identified using the MANAFEST assay are shown for patient without MPR MD01-004. MANAFEST assays for all other patients are shown in Supplementary Data 5. Each antigen-specific clonotypic expansion is colour coded to indicate if the clone was not detected in the single-cell data (blue), detected in the single-cell data but not tested via TCR cloning (green), or detected in the single-cell data and validated with TCR cloning (red). Data are shown as the percent of MANAFEST+ clonotypes among CD8+ T cells after 10 day culture.

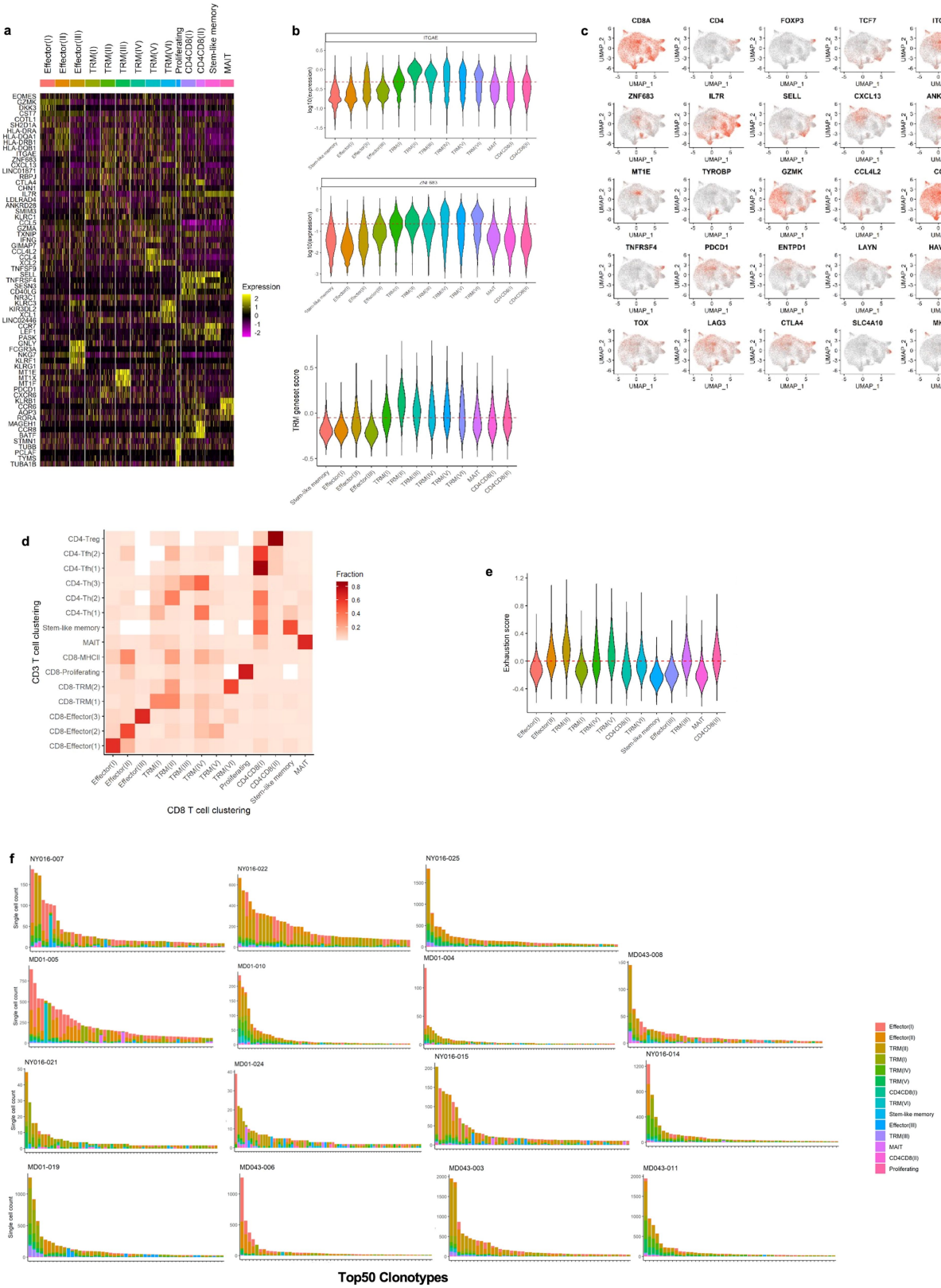


Extended Data Fig. 3 | Peripheral dynamics and cross-compartment representation of antigen-specific T cells. Bulk TCRseq was performed on pre- and post-treatment tissue (left panels) and peripheral blood (right panels) for each patient in whom antigen-specific TCRs were identified by ViraFEST/MANAFEST (as shown in Extended Data Fig. 2 and Supplementary Data 5). Data are shown as the frequency of each influenza-, CEF-, and MANA-specific TCR clonotype among all TCRs detected by bulk TCR sequencing of the indicated tissue or peripheral blood time point. Antigen-specific clonotypes were not detected by bulk TCRseq of any available tissue/peripheral blood time point in patient NY016-025. TDLN, tumour draining lymph node; DLN, draining lymph node.



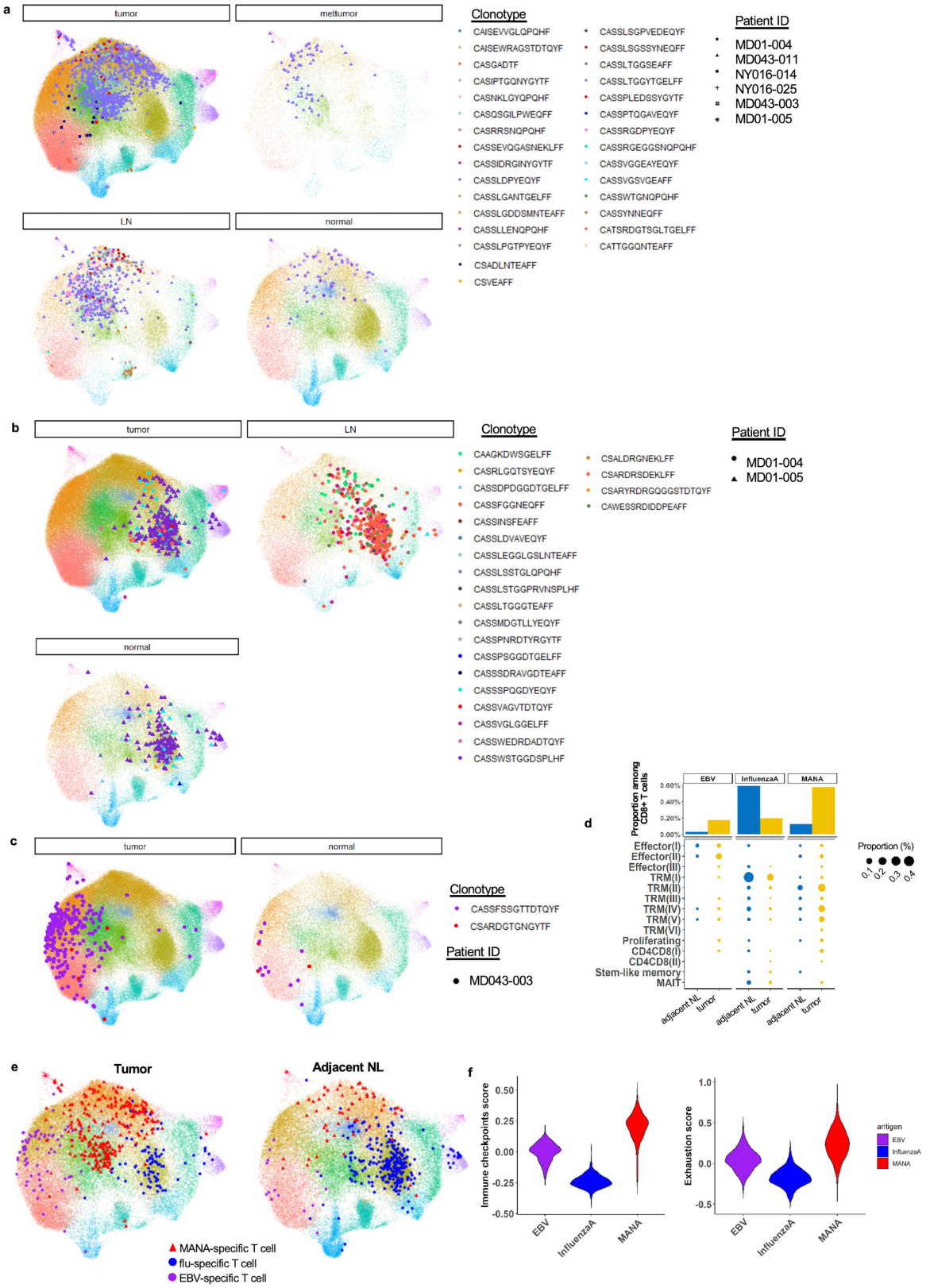
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | TCR cloning validation of MANA-specific TCRs and MANA binding kinetics. Ten TCRs identified via the MANAFEST assay were selected for TCR cloning and transfer into our NFAT/luciferase Jurkat reporter system. Seven of these TCRs recognized the cognate MANA. **a**, In MD01-005, three TCR Vβ clonotypes recognizing the ARVCF H497L-derived EVIVPLSGW MANA were identified by MANAFEST. Single-cell analysis determined that the Vβ CDR3 s CASNKLGYQPQHF and CASSLLENQPQHF were consistently detected in the same cell and paired with the same Vα CDR3, CALSMGGNEKLTF, likely the result of incomplete allelic exclusive at the beta locus. To validated that these TCRs recognized MD01-005-MANA7, and to determine which Vβ CDR3 was responsible for recognition in the case of incomplete allelic exclusion, all three TCRs were cloned into the Jurkat NFAT luciferase reporter system and tested against autologous LCL loaded with titrating concentrations of MD01-005-MANA7. Data are shown as relative luminescence units (RLU) for MD01-005-MANA7 (solid red square), the cognate wild-type peptide (open red square), or MD01-005-MANA8, which was predicted to bind A*25:01, for each individual TCR. **b**, In patient without MPR MD01-004, four TCRs recognizing the p53 R248L-derived NSSCMGGMNLR MANA (MD01-004-MANA12) were identified by MANAFEST and were detected in the single-cell data. Each Vβ chain paired exclusively with a single Vα chain. These four TCRs were cloned into the Jurkat NFAT/luciferase reporter system and tested against autologous LCL loaded with titrating concentrations of MD01-004-MANA12. Data are shown as relatively luminescence units (RLU) in response to MD01-004-MANA12 (solid blue square) or the cognate wild-type peptide (open blue square). **c**, In patient without MPR MD043-011, a TCR recognizing the CARM1 R208W-derived FAAQAGAWKIY MANA (MD043-011-MANA36) was a candidate for positivity by MANAFEST and was detected in the single-cell data. This Vβ chain paired exclusively with a single Vα chain. This TCR was cloned into the Jurkat NFAT/luciferase reporter system and tested against autologous LCL loaded with titrating concentrations of MD043-011-MANA36. Data are shown as relatively luminescence units (RLU) in response to MD01-004-MANA12 (solid green square) or the cognate wild-type peptide (open green square). **d**, The affinity of MD01-005-MANA7 for HLA A*25:01 was assessed using a luminescent oxygen channeling immunoassay (LOCI, left). This is a proximity-based system using a "donor" and "acceptor" bead, each conjugated with an epitope tag. When the donor bead is excited with light at 650nm and can activate an acceptor bead, resulting in a signal at 520-620nm, which can be quantified per second as a surrogate of affinity. A higher number of counts per second indicates higher affinity of the peptide:HLA pair. Data are shown as the number of counts per second for titrating concentrations of MD01-005-MANA7 (solid blue square), the cognate wild-type (open blue square), MD01-005-MANA8, which is predicted to bind HLA A*68:01 (black circle), or no peptide (star). Stability of these same peptides in the HLA A*68:01 complex was also evaluated using a urea-based assay, whereby the stability of the peptide:HLA complex is measured at increasing concentrations of urea (right). Data are shown as the absorbance at 450nm. Data points represent the mean +/− s.d. of two independent experiments. **e**, Binding (top left) and stability (top right) assays were conducted as in (b) for the p53 R248L-derived MD01-004-MANA12 (solid green square), the cognate wild-type peptide (open green square), a positive control peptide for HLA A*68:01 (orange diamond), the YTAVPLVYV peptide which is predicted to bind A*68:01 (black circle), or no peptide (black star). Data points represent the mean +/− s.d. of two independent experiments. To determine if MD01-004-MANA12 is endogenously processed and presented by HLA A*68:01, COS-7 cells were transfected with HLA-A*68:01 plasmid and p53 R248L mutant plasmid or p53 wild type plasmid. HLA- and p53-transfected COS-7 cells, autologous APC loaded with MD01-004-MANA12, and HLA-A*68:01-transfected COS-7 were co-cultured with CD8+ Jurkat reporter cells expressing the MD01-004-MANA12-reactive TCR, Vβ: CATTGGQNTEAFF, Vα: CILSGANNLFF. Data are shown as relative luminescence units (RLU) for each condition (bottom).



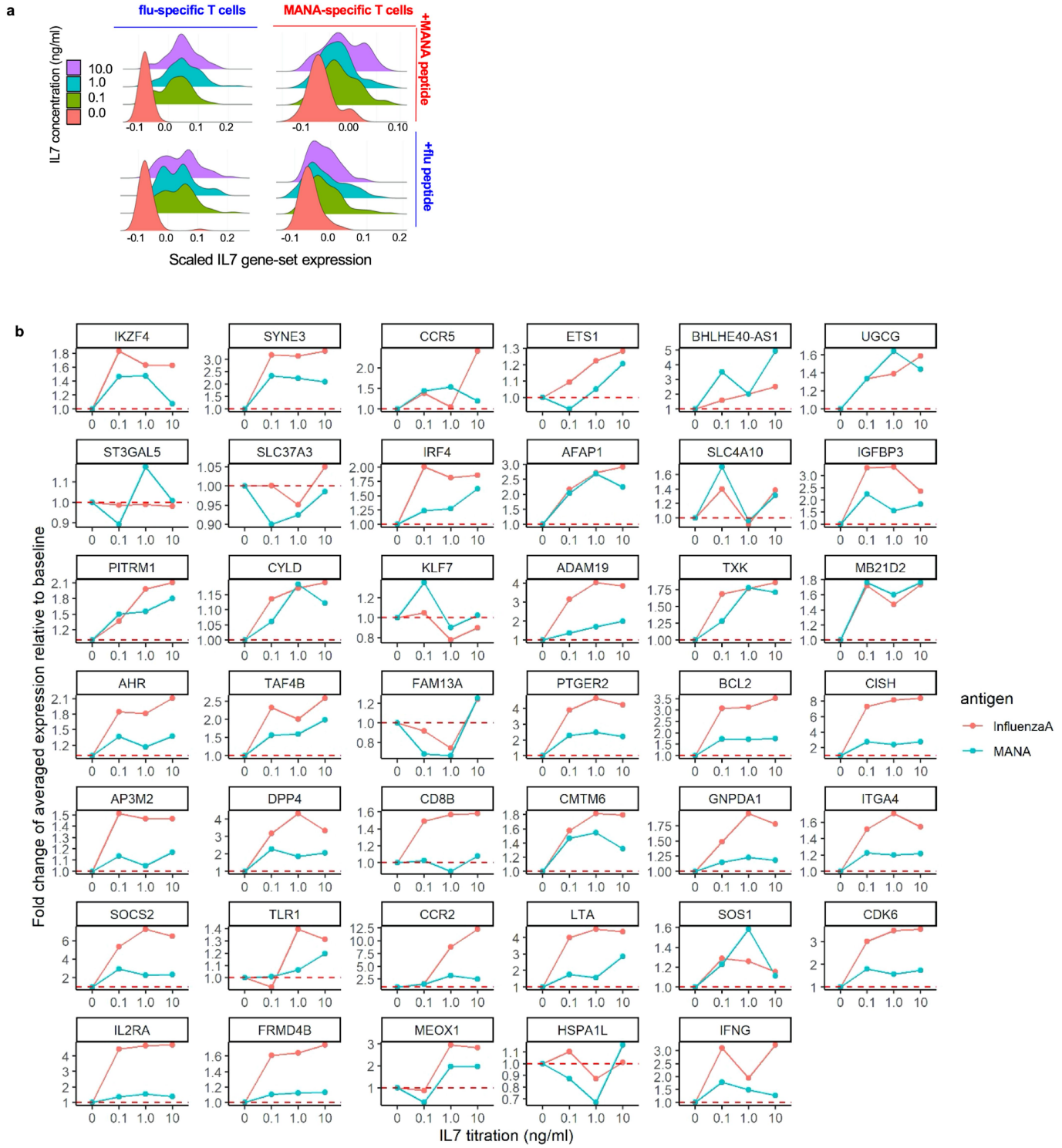
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Refined clustering on CD8 T cells. a, A heat map shows the top differential genes, ranked by average fold change, for each refined CD8 T cell cluster. 5,000 cells (or all cells in the cluster if cluster size <5000 cells) were randomly sampled from each cluster for visualization (n = 16 patients). **b**, Violin plots show the log10 expression of the TRM-defining genes, ITGAE (top) and ZNF683 (HOBIT, middle), and a TRM gene-set score (bottom) for each CD8 T cell cluster. The dashed line indicates the mean expression of the respective gene or gene-set score among all CD8 T cells. Because the proliferating cluster is driven by proliferation-associated genes and represents mixed cell types, this cluster was not shown in the plot. **c**, 2D UMAP red-scale projection of canonical T cell subset marker genes, cell subset selective genes, and immune checkpoints on CD8 T cell subsets. **d**, A heat map shows the proportion of each refined CD8 T cell cluster (Fig. 2b) that is found within each global UMAP T cell cluster (Fig. 1b). This enables visualization of the "parent" cluster for the refined CD8 T cell clusters. **e**, A violin plot shows the exhaustion gene-set score, comprised of a published exhaustion gene list (Supplementary Data 2.2), for each refined CD8 T cell cluster. The dashed line shows the mean exhaustion gene-set score among all CD8 T cells. Because the proliferating cluster is driven by proliferation-associated genes and represents mixed cell types, this cluster was not shown in the plot. **f**, $CD8^+$ T cell clonotypic cluster composition. The top 50 $CD8^+$ TCR clonotypes in the tumour are shown for each patient, and the proportion of each clonotype that was found within each cluster is designated by the colour code.

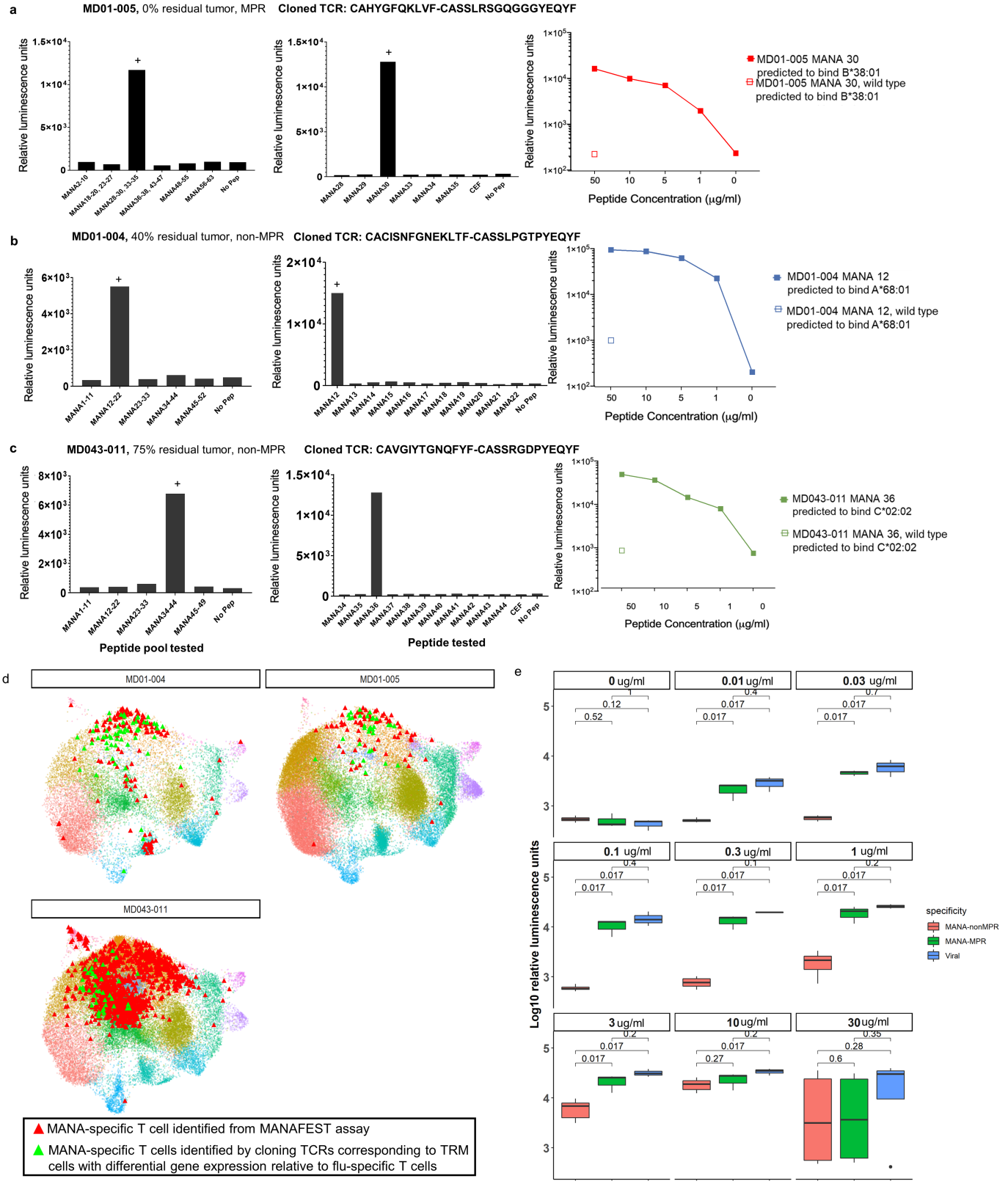


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Distinct phenotype of antigen-specific T cells. **a**, Distribution of MANA-specific T cells on UMAP. Individual MANA-specific clonotypes are shown on the UMAP, stratified by tissue compartment and patient ID. Each colour represents a unique MANA-specific clonotype, and each symbol represents a patient. **b**, Distribution of EBV-specific T cells on UMAP. Individual EBV-specific clonotypes are shown on the UMAP, stratified by tissue compartment. Each colour represents a unique EBV-specific clonotype and each symbol represents a patient. **c**, Distribution of influenza-specific T cells on UMAP. Individual influenza-specific clonotypes are shown on the UMAP, stratified by tissue compartment and patient ID. Each colour represents a unique influenza-specific clonotype, and each symbol represents a patient. The CD8 T cell clusters are annotated according to the designation in Fig. 2b. **d**, The barplot (upper) shows the proportion of antigen-specific T cells among total CD8 T cells by tissue compartment (blue bar, adjacent NL; yellow bar, tumour). The dotplot (bottom) shows the proportion of antigen-specific T cells stratified by subset, with the size of the dot representing the proportion among total CD8 T cells (blue dot, adjacent NL; yellow dot, tumour). **e**, TIL and adjacent NL CD8 T cells were downsampled to equal numbers of cells on UMAP before visualization of antigen-specific clonotypes in tumour (left) and adjacent normal lung (right). **f**, The immune checkpoint score and exhaustion score of antigen-specific T cells. A violin plot shows a composite immune checkpoint score (left) and exhaustion score (right) for EBV(purple)-, influenza (blue)-, and MANA (red)-specific T cells.

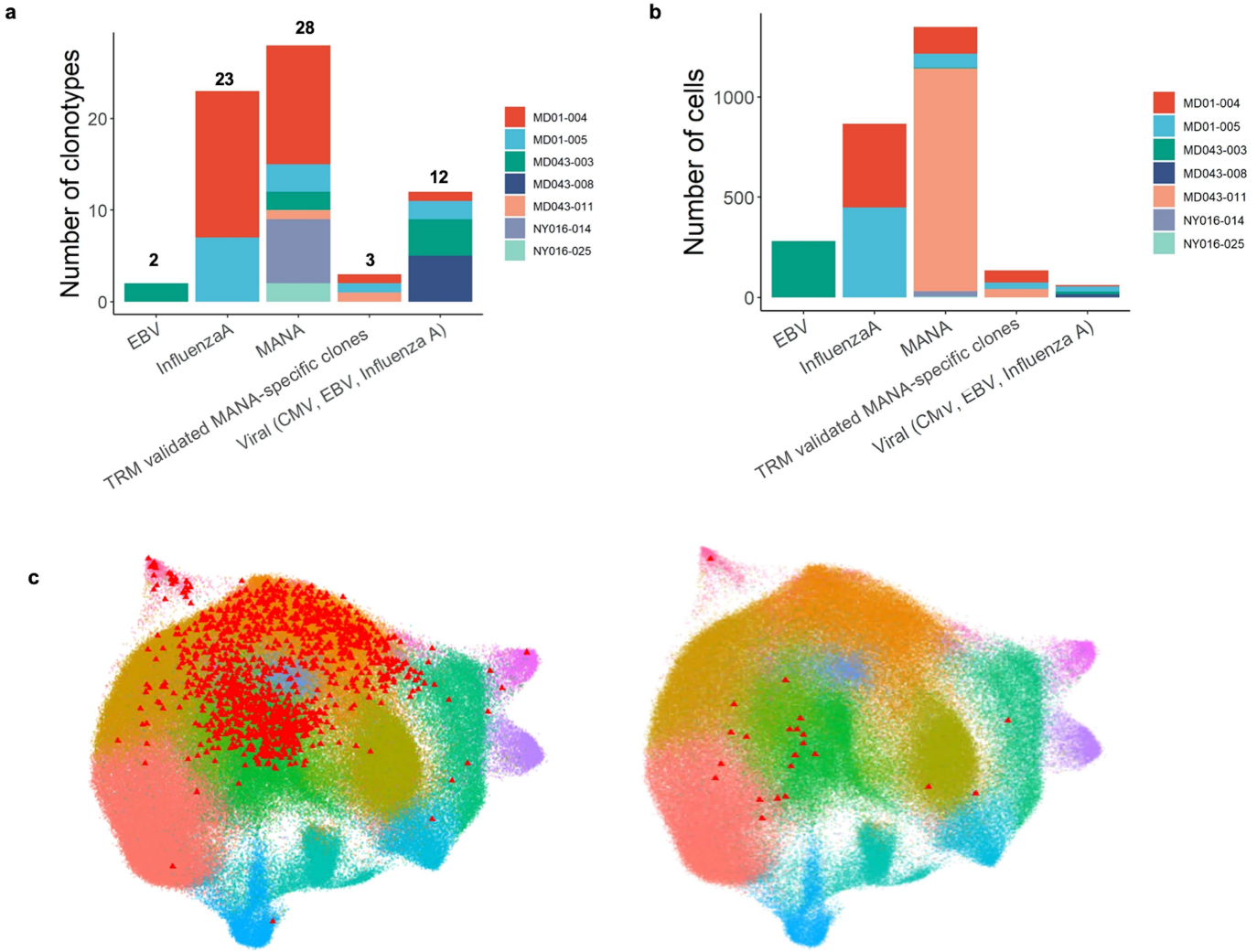


Extended Data Fig. 7 | IL-7-induced gene signature between MANA-specific and influenza-specific TIL. TIL from patient MD01-004 were cultured with MD01-004-MANA-12 or influenza A peptide and titrating concentrations of recombinant human IL-7, followed by coupled scRNA-seq–TCR-seq. A total of 814 influenza-specific (410 co-cultured with influenza peptide, 404 co-cultured with MANA peptide) and 581 MANA-specific TIL (366 co-cultured with influenza peptide, 215 co-cultured with MANA peptide) were detected in the single-cell data from a single experiment and were analysed. **a**, Composite expression of an IL-7 gene set by influenza-specific and MANA-specific TIL (as determined by their TCR Vβ CDR3) stimulated with cognate or non-cognate antigen is shown. **b**, Dose–response curve showing the fold change of averaged expression of IL-7-induced genes (Supplementary Data 2.3) that significantly changed from baseline (no IL-7 vs 0.1 ng/ml) in influenza-specific (red) or MANA-specific (blue) T cells. Comparisons were performed using two-sided Wilcoxon rank sum test and adjusted for multiple comparisons using BH method.

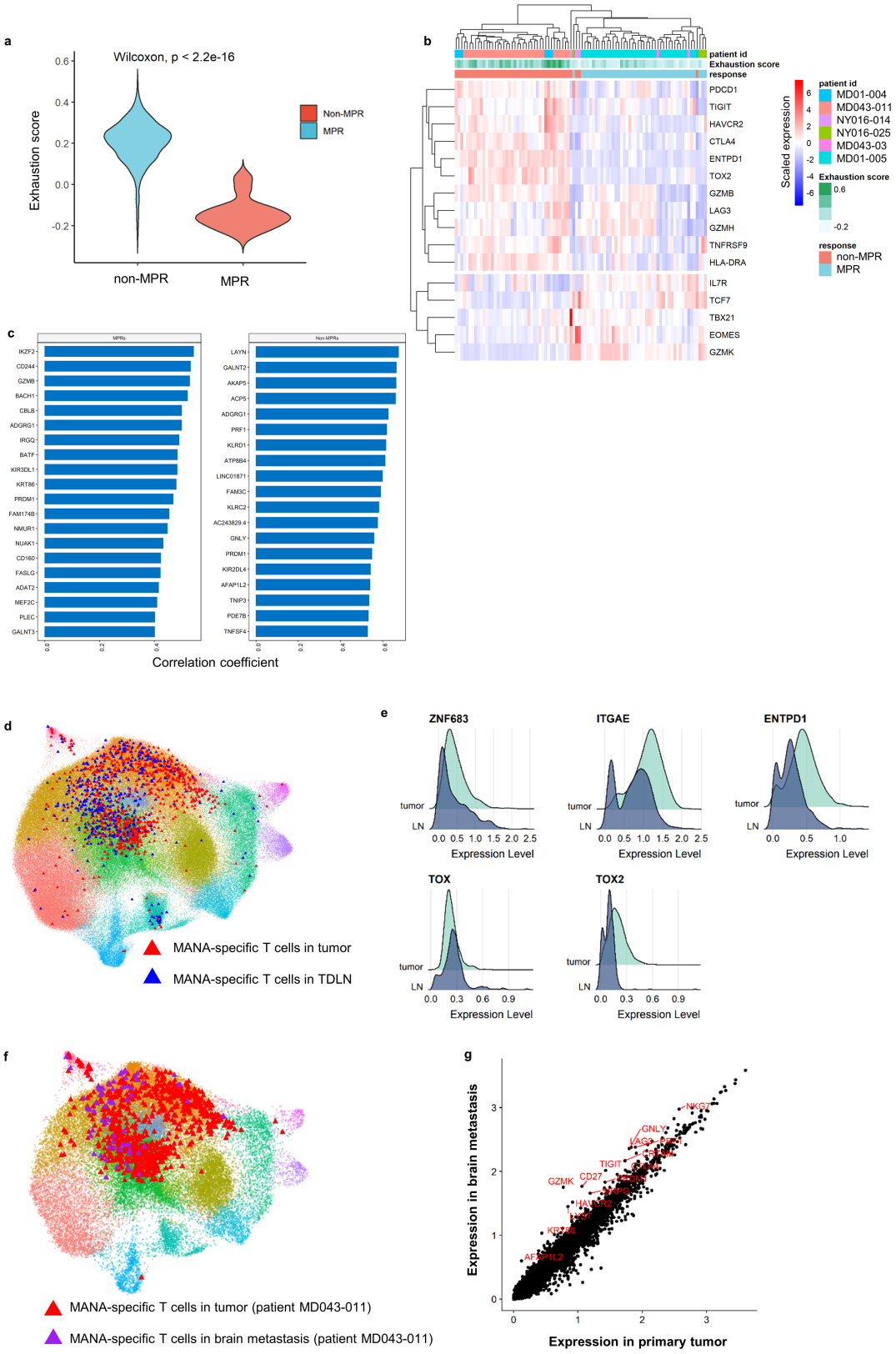


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Cloning and dose response of antigen-specific T cells. a–**c**, Cloning and screening of TCRs corresponding to CD8 T cells with highly differential gene expression relative to influenza-specific T cells. Seven TCRs were selected from the refined CD8 sc data based on highly differential gene expression relative to influenza-specific T cells. These TCRs were cloned into the Jurkat/NFAT luciferase reporter system and first screened against autologous LCL pre-loaded with pools of putative MANA peptides (10μg/ml) based on the respective patient's WES and MANA predictions. Three TCRs recognized a MANA peptide pool, one each from patients MD01-005 (**a**), MD01-004 (**b**), and MD043-011 (**c**). The reactive MANA was then mapped from the reactive peptide pool by stimulating the TCR-transfected Jurkat cell with autologous LCL pre-loaded with 10μg/ml of each individual MANA within the reactive pool (centre). Dose–response curves were then generated for each MANA-specific TCR (right). Data are shown as relative luminescence units. A (+) sign indicates the positive response. **d**, Functional characterization of MANAFEST-identified and screening-identified TCRs. 2D projection of clones identified from the MANAFEST assay (red) and clones identified via cloning of TCRs corresponding to T cells with differential gene expression relative to influenza-specific T cells (green) is shown for patients MD01-004, MD01-005, and MD043-011. CD8 T cell clusters are marked with the same colour code as Fig. 2b. **e**, Viral-specific TCRs and MANA-specific TCRs from one patient with MPR and two patients without MPR were cloned into the Jurkat reporter system and tested against titrating concentrations of relevant peptide. The average $\log_{10}$ relative luminescence of viral-specific TCRs (blue, 3 clonotypes from 3 different patients), MANA-specific MPR TCRs (green, 3 clonotypes from 1 patient with MPR), and MANA-specific non-MPR TCRs (red, 7 clonotypes from 2 patients without MPR) was compared at each peptide titration. Data are shown as a Box and Whiskers plot. The middle bar shows the median, with the lower and upper hinges corresponding to the 25^{th} and 75^{th} percentiles, respectively (interquartile range, IQR). The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Comparisons of relative luminescence units for viral-specific vs MANA-specific T cell clonotypes at different titrations were performed using two-sided Wilcoxon rank sum test. ns: $P > 0.05$; *, $0.01 < P < 0.05$.

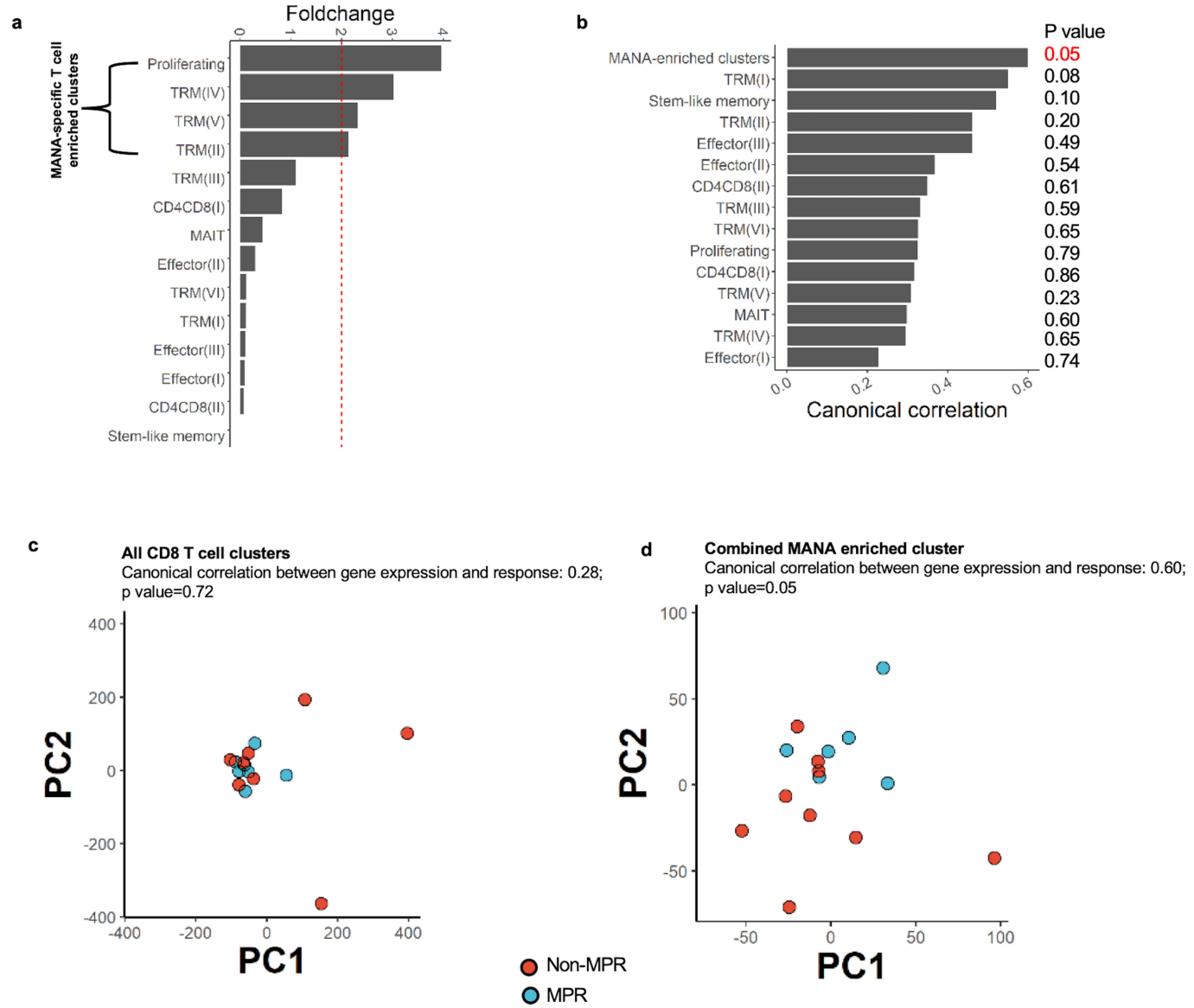


Extended Data Fig. 9 | Patient representation of antigen-specific clonotypes. a, **b**, Barplots summarize the total number of unique tumour-infiltrating clonotypes (**a**) and cells (**b**), stratified by antigen specificity and method of detection (MANAFEST or based on the TRM gene signature and cloning/peptide screen). Different colours represent the patient identity. **c**, Visualization of clonotypes included in the MANA-specific analysis. The individual UMAP projections of clonotypes that were validated (left) and were not validated (right) by TCR cloning are shown. Of the cells that corresponded to a MANAFEST-identified, MANA-specific clonotype that was detected in the single-cell data, >94% were validated by the jurkat/luciferase TCR cloning system.

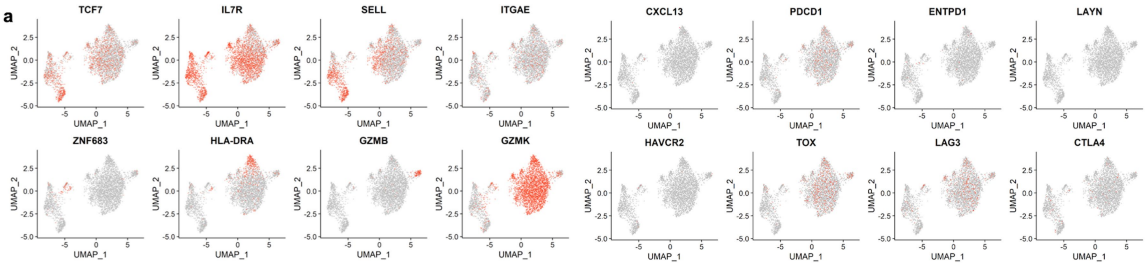


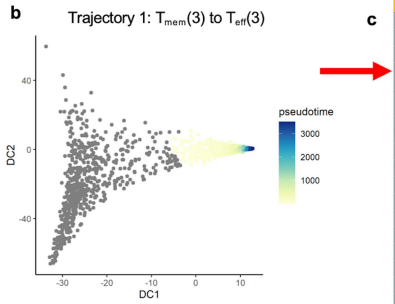
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Signatures of MANA-specific T cells according to response and tissue compartment. a, Exhaustion score and co-expression of immune checkpoints/effector/memory function gene on MANA-specific TIL. Violin plot shows the exhaustion gene-set score (Supplementary Data 2.2) of MANA-specific TIL of non-MPR (red, $n = 3$) and MPR (light blue, $n = 3$) tumours. Comparisons were performed at the individual cell level using two-sided Wilcoxon rank sum test without multiple comparison adjustment. **b**, Heat map shows co-expression of immune checkpoints and effector/memory genes on MANA-specific TIL. Each column represent a cell. The exhaustion score, response status, and patient IDs are designated by the relevant colour bar. For visualization, MANA-specific T cells were downsampled to the same number of cells from MPR ($n = 3$) and non-MPR ($n = 3$). **c**, Top ranked genes correlated with the immune checkpoint score in MANA-specific TIL. Barplots show the correlation coefficients of the top ranked genes highly correlated with the immune checkpoint score in MPR (left) and non-MPR (right) MANA-specific TIL. **d**, MANA-specific T cells found in the tumour (red triangles) and TDLN (blue triangles) of patients MD01-004, MD01-005, and MD043-011 were projected on the refined CD8 UMAP. **e**, Expression of selective genes is shown for MANA-specific T cells in the tumour and TDLN ($n = 3$). **f**, MANA-specific T cells found in the tumour (red triangle) and brain metastasis (purple triangle) are shown on the UMAP for patient MD043-011. **g**, The scatterplot shows the average expression of genes comparing all refined CD8 T cells from the primary tumour and metastatic brain resection in patient MD043-011. The top differential genes enriched in the brain metastasis are labelled in red. Comparisons were performed at the individual cell level using two-sided Wilcoxon rank sum test. *P*-value adjustment was performed using bonferroni correction. A complete list of differential genes comparing primary tumour at resection vs. the distant brain metastasis is shown in Supplementary Data 1.5. CD8 T cell clusters are marked by the same colour code as Fig. 2b.



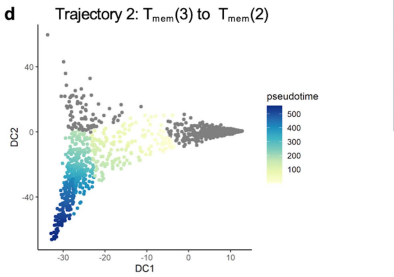
Extended Data Fig. 11 | Canonical correlations of CD8 T cell clusters with pathologic response. The canonical correlation between pathologic response status and CD8 T cell clusters vs. a MANA-specific T cell-enriched cluster was evaluated. **a**, Selection of MANA-specific T cell enriched clusters (Proliferating, TRM(IV), TRM (V) and TRM (II)) based on >2 fold change (red dotted line) of MANA-specific T cell frequency relative to random expectation. The above 4 clusters were combined as a 'MANA-combined' cluster. **b**, Combined MANA-specific T cell enriched clusters showed the highest canonical correlation with pathologic response. **c**, PCA of pseudobulk gene expression from all CD8 T cell clusters for individual tumour samples ($n = 15$, 6 MPRs and 9 non-MPRs), coloured by response status (MPR as blue blue dots, non-MPR as red dots). **d**, PCA of pseudobulk gene expression from combined MANA enriched T cell cluster for individual tumour samples ($n = 15$, 6 MPRs and 9 non-MPRs), coloured by response status (MPR as light blue dots, non-MPR as red dots). *P* values were obtained using a one-sided permutation test, without correction for multiple comparisons.





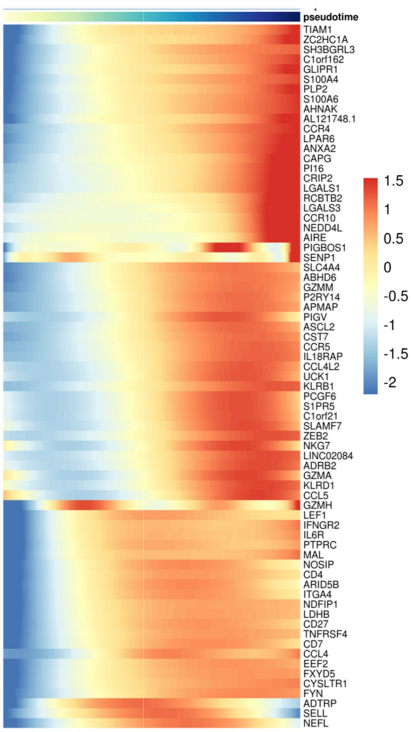
c

	GO.ID	Term	Annotated	Significant	Expected	classicFisher	FDR	FC
1	GO:0060333	interferon-gamma-mediated signaling pathway	68	9	0.22	3.6E-13	9.3093E-10	41.1666666666667
2	GO:0071346	cellular response to interferon-gamma	104	10	0.33	3.9E-13	9.3093E-10	29.9074074074074
3	GO:0034341	response to interferon-gamma	114	10	0.37	9.9E-13	1.57542E-09	27.2839506172839
4	GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	68	8	0.22	2.4E-11	2.70526666666667E-08	36.5925925925926
5	GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	70	8	0.23	3E-11	2.70526666666667E-08	35.5470899470899
6	GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	71	8	0.23	3.4E-11	2.70526666666667E-08	35.0464267083985
7	GO:0019221	cytokine-mediated signaling pathway	433	13	1.39	1.6E-10	1.0912E-07	9.33829441450689
8	GO:0006955	immune response	1145	18	3.68	3.2E-10	1.9096E-07	4.88966521106259
9	GO:0002684	positive regulation of immune system process	606	14	1.95	7.5E-10	3.97833333333333E-07	7.18567412296785
10	GO:0006952	defense response	791	15	2.54	2.1E-09	1.00254E-06	5.89830032307909
11	GO:0002429	immune response-activating cell surface receptor signaling pathway	253	10	0.81	2.8E-09	1.2152E-06	12.2939540330845
12	GO:0050851	antigen receptor-mediated signaling pathway	185	9	0.59	3.3E-09	1.31285E-06	15.1315315315315
13	GO:0002768	immune response-regulating cell surface receptor signaling pathway	272	10	0.87	5.6E-09	2.05649230769231E-06	11.4351851851852
14	GO:0002478	antigen processing and presentation of exogenous peptide antigen	135	8	0.43	6.2E-09	2.1142E-06	18.4318244170096
15	GO:0019884	antigen processing and presentation of exogenous antigen	137	8	0.44	7E-09	2.22786666666667E-06	18.1627466882941
16	GO:0071345	cellular response to cytokine stimulus	597	13	1.92	8.3E-09	2.4765125E-06	6.77300080650164
17	GO:0048002	antigen processing and presentation of peptide antigen	145	8	0.47	1.1E-08	2.91744444444444E-06	17.1606641123883
18	GO:0002376	immune system process	1624	19	5.22	1.1E-08	2.91744444444444E-06	3.63898011311804
19	GO:0045087	innate immune response	508	12	1.63	1.6E-08	4.02021052631579E-06	7.34733158355206
20	GO:0034097	response to cytokine	639	13	2.05	1.9E-08	4.5353E-06	6.32782704457196
21	GO:0019882	antigen processing and presentation	166	8	0.53	3.2E-08	7.27466666666667E-06	14.9897367246765
22	GO:0002757	immune response-activating signal transduction	348	10	1.12	5.9E-08	1.2803E-05	8.93784589186888
23	GO:0002682	regulation of immune system process	861	14	2.77	7.2E-08	1.49446956521739E-05	5.0575127973502
24	GO:0050776	regulation of immune response	588	12	1.89	8.2E-08	1.63111666666667E-05	6.34769463340892
25	GO:0050778	positive regulation of immune response	470	11	1.51	8.7E-08	1.661352E-05	7.2795902285264
26	GO:0002764	immune response-regulating signaling pathway	366	10	1.18	9.5E-08	1.74434615384615E-05	8.49827970046549
28	GO:0002253	activation of immune response	375	10	1.21	1.2E-07	2.12177777777778E-05	8.29432098765432
29	GO:0050852	T cell receptor signaling pathway	152	7	0.49	3.6E-07	6.138E-05	14.3240740740741
31	GO:0002696	positive regulation of leukocyte activation	187	7	0.6	1.5E-06	0.000246931034482759	11.6430976430976



e

	GO.ID	Term	Annotated	Significant	Expected	classicFisher	FDR
1	GO:0071345	cellular response to cytokine stimulus	590	11	1.62	1.2E-07	0.0002374
3	GO:0006955	immune response	1126	14	3.09	1.5E-07	0.0002374
2	GO:0007166	cell surface receptor signaling pathway	1335	15	3.67	1.4E-07	0.0002374
4	GO:0034097	response to cytokine	632	11	1.74	2.4E-07	0.00028488
6	GO:0007159	leukocyte cell-cell adhesion	190	7	0.52	5E-07	0.000395666666666667
5	GO:0002682	regulation of immune system process	845	12	2.32	4.9E-07	0.000395666666666667
7	GO:0022407	regulation of cell-cell adhesion	208	7	0.57	9.2E-07	0.000501177777777778
9	GO:0098609	cell-cell adhesion	309	8	0.85	9.5E-07	0.000501177777777778
8	GO:0040011	locomotion	722	11	1.98	9.3E-07	0.000501177777777778
10	GO:0002376	immune system process	1603	15	4.4	1.6E-06	0.00075968
11	GO:1903037	regulation of leukocyte cell-cell adhesion	169	6	0.46	4.7E-06	0.00175310769230769
12	GO:0007155	cell adhesion	520	9	1.43	4.8E-06	0.00175310769230769
13	GO:0022610	biological adhesion	520	9	1.43	4.8E-06	0.00175310769230769
15	GO:0042110	T cell activation	281	7	0.77	6.9E-06	0.002047575
14	GO:0046649	lymphocyte activation	402	8	1.1	6.8E-06	0.002047575
16	GO:0071310	cellular response to organic substance	1297	13	3.56	6.9E-06	0.002047575
17	GO:0045321	leukocyte activation	716	10	1.97	8E-06	0.00223435294117647
19	GO:0050900	leukocyte migration	192	6	0.53	9.8E-06	0.002374
20	GO:0019221	cytokine-mediated signaling pathway	426	8	1.17	1E-05	0.002374
18	GO:0010033	response to organic substance	1574	14	4.32	9.6E-06	0.002374
21	GO:0016477	cell migration	589	9	1.62	1.3E-05	0.00293923809523381
22	GO:0002684	positive regulation of immune system process	595	9	1.63	1.4E-05	0.00302145454545455
23	GO:0001775	cell activation	779	10	2.14	1.7E-05	0.00350939130434783
25	GO:0097191	extrinsic apoptotic signaling pathway	125	5	0.34	1.9E-05	0.00360848
24	GO:0030155	regulation of cell adhesion	327	7	0.9	1.9E-05	0.00360848
26	GO:0009605	response to external stimulus	990	11	2.72	2.1E-05	0.00383492307692308
32	GO:0002250	adaptive immune response	228	6	0.63	2.6E-05	0.00385775
30	GO:0048870	cell motility	640	9	1.76	2.6E-05	0.00385775
31	GO:0051674	localization of cell	640	9	1.76	2.6E-05	0.00385775
28	GO:0006928	movement of cell or subcellular component	812	10	2.23	2.5E-05	0.00385775



Extended Data Fig. 12 | See next page for caption.

Extended Data Fig. 12 | Phenotypic characteristics of FACS-sorted peripheral blood CD8+/Vβ2+ T cells from MPR MD01-005. a, Selective gene expression of 2D UMAP red-scale projection is shown of canonical T cell subset marker genes, cell subset selective genes, and immune checkpoints on CD8 T cell subsets sorted from longitudinal peripheral blood of one patient (MD01-005) with complete pathologic response. **b-d**, Pseudotime reconstruction and pseudo-temporal dynamic gene identification in peripheral blood CD8 T cells from a complete pathologic responder. Longitudinal PBMC were collected from complete pathologic responder MD01-005 (0% residual tumour) during treatment and in post-surgery follow up. Peripheral blood $CD8^+$ T cells were FACS sorted based on expression of TCR Vβ2, which corresponds to the MANA-specific CDR3 CASNKLGYQPQHF as identified previously via the MANAFEST assay (Extended Data Fig. 2a). scRNA-seq–TCR-seq was performed on the sorted population from each time point. **b**, Constructing the pseudotime axis on the diffusion map from $T_{mem}(3)$ to $T_{eff}(3)$ as trajectory 1. **c**, GO analysis for genes that significantly change along trajectory 1, ranked by FDR. **d**, Constructing the pseudotime axis on the diffusion map from $T_{mem}(3)$ to $T_{mem}(2)$ as trajectory 2. **e**, GO analysis for genes that significantly change along trajectory 2, ranked by FDR. **f**, Heat map showing genes that significantly change along trajectory 2 (FDR < 0.05).

nature research

Corresponding author(s): Kellie N. Smith, Drew M. Pardoll, Hongkai Ji, Srinivasan Yegnasubramanian

Last updated by author(s): 2021/06/17

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	Alignment and preprocessing was performed using cellranger 3.1.0 (10X Genomics). Data were processed using Seurat 3.1.5 in R 4.0.3. Gene expression heatmap was generated using heatmap3 1.1.7 package. Plots were generated using PRISM 8.2.0 and ggplot2 3.3.2 package in R. Bioinformatic identification of antigen specific T cells in bulk TCR sequencing was performed in http://www.stat-apps.onc.jhmi.edu/FEST/. Scripts to reproduce the analyses used in this study are available at: https://github.com/BKI-immuno/neoantigen-specific-T-cells-NSCLC.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Bulk TCR Vbeta sequencing data generated by Adaptive Biotechnologies is available in the Adaptive Biotechnologies ImmuneACCESS repository under DOI 10.21417/JC2021N, at clients.adaptivebiotech.com/pub/caushi-2021-n. Bulk TCR Vbeta raw and processed sequencing data generated by the SKCCC FTIC are available in GEO with accession number GSE173351. Raw scRNAseq/TCRseq data reported in this paper are available in the European Genome-phenome archive

under controlled access with accession number EGAS00001005343. Due to the personal, sensitive and inherently identifying nature of raw genomic data, access to rawRNAseq/TCRseq data are controlled and full instructions to apply for data access can be found at https://ega-archive.org/access/data-access. Approvals will be granted immediately upon confirmation that all requirements are met. Processed and de-identified single cell data are available in GEO with accession number GSE176022. Somatic mutations, predicted neoantigens, and the identity of all antigen-specific TCR Vbeta clonotypes are shown in the Supplementary Tables.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	21 patients with non-small cell lung cancer were recruited to a prospective phase 2 clinical trial that investigated the safety and feasibility of administering two doses of anti-PD-1 (nivolumab) prior to surgical resection. Twenty patients had surgical resection of their primary tumor. Major pathologic response (MPR) was defined as ≤ 10% residual tumor as assessed by two independent pathologists. All participants signed the informed consent. In total, we performed combined single cell RNA/TCR sequencing of 16 patients (7 MPR, 9 non-MPR) and 9 patients were tested for MANA reactivity (4 MPR, 5 non-MPR).
Data exclusions	Single cell sequencing was performed on all tumor and normal lung specimens with sufficient viably banked T cells. MANAFEST/ViraFest was performed on all patients with adequate blood samples and available WES. Single cell TCRseq/RNAseq was performed on peripheral blood T cells sorted for positive expression of the TCR Vb2 gene. Unconventional/MAIT CD8+ T cells were excluded from this analysis to enable better visualization of canonical CD8+ T cell subsets.
Replication	Epidemiological replication/validation: In this study, we included all NSCLC patients enrolled at two academic cancer centers in the US, who were willing to participate in this study and signed the informed consent. For replication purposes an additional validation cohort would have been desirable. However, due to the limited number of patients in a phase 2 study per study design and increased interests to test combinational regimens in the neoadjuvant setting, we were not able to perform an independent replication trial. Technical replication: We performed single-cell RNA sequencing experiments with high numbers of cells per patient and, in general, high quality sequencing libraries. Sample replicates were performed for 4 patients and showed high consistency in RNA expression profile, and were subsequently merged as one single sample for downstream analysis.
Randomization	*Randomization was not relevant for our study*
Blinding	*Blinding was not relevant for our study*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*

Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

LIVE/DEAD™ Fixable Near-IR; ThermoFisher
CD3-BV605, clone SK7; BD Bioscience
LIVE/DEAD™ Fixable Aqua; ThermoFisher
CD3-PE, clone HIT3a; BD Bioscience
CD8-BV786, clone RPA-T8; BD Bioscience
Anti-TCR Vb2-PE, clone REA654; Miltenyi
Anti-TCR Vb5-FITC, clone MEM-262, Biolegend

Validation

All new antibodies were titrated to determine the optimal working concentration. Specificity was validated using control primary cells or cell lines. Isotype controls were used to gate on cells staining with the antibody of interest.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Jurkat cell line - Promega

Authentication

TCRa/b knockout of the jurkat cell line was confirmed by Sanger sequencing and restoration of CD3 expression only by the co-transfection of TCRα or TCRβ chains.

Mycoplasma contamination

The jurkat cell line used in this study tested negative for mycoplasma contamination

Commonly misidentified lines
(See ICLAC register)

We did not use any commonly misidentified cell lines

Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics

No patient was excluded from the original clinical trial on the basis of sex, ethnic background, or socio-economic status. Only patients greater than age 18 were eligible for the study. Patients ranged in age from 55 to 84. Health status was good to excellent and the breakdown for race was 5.7% Hispanic, 5.7% African American, 8.6% Other, and 80% Caucasian. Special classes of subjects such as pregnant women, children, prisoners, or other institutionalized individuals were not included in

	the original trial and are therefore not included in our present study.
Recruitment	Participants were not prospectively recruited for this study. The original clinical trial was performed on 21 patients with resectable NSCLC. Twenty of these patients underwent successful surgical resection of their primary tumor. Patients included in the present study were selected solely based on the availability of biospecimens for analysis.
Ethics oversight	This study was approved by the Institutional Review Boards of Johns Hopkins and MSKCC

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT02259621
Study protocol	The results from this clinical trial and associated study protocol have been published previously (Forde PM, et al, N Engl J Med, 2018). The study protocol can be accessed at: https://www.nejm.org/doi/suppl/10.1056/NEJMoa1716078/suppl_file/nejmoa1716078_protocol.pdf
Data collection	All biospecimens were obtained between 2015 and 2018 from patients enrolled to a phase II study of neoadjuvant PD-1 blockade in resectable lung cancer at Johns Hopkins Sidney Kimmel Comprehensive Cancer Center and Memorial Sloan Kettering Cancer Center. Analyses on these biospecimens were continuously performed between 2017 and 2021.
Outcomes	Primary lung tumor and lymph-node surgical specimens were staged according to the criteria of the American Joint Committee on Cancer (seventh edition) for evaluating tumor size, affected lymph nodes, and metastases.7 Primary tumors were assessed for the percentage of residual viable tumor that was identified on routine hematoxylin and eosin staining, and tumors with no more than 10% viable tumor cells were considered to have had a major pathological response. These methods and clinical outcomes have been reported previously (Forde PM, et al, N Engl J Med, 2018).

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were resuspended in PBS and stained with a viability marker (LIVE/DEAD™ Fixable Near-IR; ThermoFisher) for 15mins at RT in the dark. Cells were the incubated with FC block for 15 mins on ice and stained with antibody against CD3 (BV605, clone SK7) for 30mins on ice. After staining, highly viable CD3+ T cells were sorted into 0.04% BSA in PBS using a BD FACSAria II Cell Sorter. Sorted cells were manually counted using a hemocytometer and prepared at the desired cell concentration (1000 cells/ul), when possible. CD3+ cells were then sorted and immediately used in single cell experiments.
Instrument	FACS Aria II
Software	FACS Diva
Cell population abundance	Cells that were FACS sorted were immediately used in single cell TCRseq/RNAseq experiments. Only T cells were sequenced in this platform, and the cells analyzed were validated using gene expression programs, so there is no risk of contaminating cell types/phenotypes.
Gating strategy	Gating was based on staining with negative isotype control antibodies

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis	*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*