

## Supplementary Methods

### Patient selection

Patients treated with chemotherapy were selected from the IMvigor 211 trial (NCT02302807), which was a multicenter, open label, randomized phase 3 study of mUC patients treated with atezolizumab versus chemotherapy combination of taxane or vinflunine <sup>1</sup>. Both trials were sponsored by Genentech, Inc., a member of the Roche group. Of 30 patients treated with atezolizumab (validation cohort), 16 patient samples were processed for scRNA-seq coupled with scTCR-seq and CITE-seq. The best confirmed overall responses were assessed by complete response (CR), partial response (PR), and progressive disease (PD) according to RECIST v1.1. The baseline characteristics of the study population are separated by cohorts (discovery and validation cohorts) and shown in **table S1**. Data are presented with mean (+/- SD), number (%), or median (interquartile range). NA refers to data not available.

### Neoantigen prediction

Briefly, whole exome sequencing data was generated from tumor and PBMCs of 20 mUC patients from the discovery cohort using the Agilent SureSelect v5 (51 MB) kit and a HiSeq 2500 (Illumina®) sequencer. HLA typing was performed using Polysolver (v1.0) <sup>2</sup> with the whole exome data from PBMCs. Somatic variants were called using Strelka (v1.0.14) <sup>3</sup> and Lofreq2 (v2.1.2) <sup>4</sup> and annotated for effects on transcripts using the Ensembl Variant Effect Predictor (v74) on RefSeq-based gene models. RNA-seq data was generated from tumor samples, and expressed mutations were identified from the exome data by tallying base substitution counts from the RNA-seq alignments using the tallyVariants function from the VariantTools R package (v1.20.0) <sup>5</sup>. Mutations with 2 or more supporting RNA reads were retained, while the rest were discarded. For

each retained somatic mutation resulting in an amino acid change, HLA-peptide binding affinities were predicted across all pairs of HLA alleles and 8-11mer peptides containing the mutation using NetMHCcons-1.1<sup>6</sup>. The HLA-peptide pair with the lowest binding affinity percentile rank score (i.e. the highest binding likelihood) across all pairs derived from a given mutation was chosen as representative. Neoantigens were then defined as mutations with a percentile rank score of  $\leq 2$ , while the remaining putative non-binders were discarded.

### Peptides

A total of 656 patient-specific neoantigen sequences restricted to HLA-A\*02:01, HLA-A\*01:01, HLA-A\*11:01, HLA-A\*03:01 and HLA-B\*07:02 were used for antigen-specific T cell screening in PBMC samples from the discovery cohort. (**table S3**). Individual patient samples were probed for 25 to 98 antigen specificities across up to three different HLAs (**table S2**). The total number of mutations, predicted and assayed neoantigens in each patient is included in **table S2**. The shared neoantigens among patients are also included in **table S2**. All peptides were ordered from Genscript (China) or Mimotopes (Australia) with a purity above 85% by HPLC purification and mass spectrometry.

### Phenotypic panel set-up

Purified antibodies lacking carrier proteins (100  $\mu\text{g}$ /antibody) were conjugated to DN3 MAXPAR chelating polymers loaded with heavy metal isotopes according to the manufacturer's recommendations (Fluidigm). Specific antibody staining panels consisting of lineage markers, descriptive markers, labels for live/dead discrimination (cisplatin) and DNA were set up for all samples from discovery cohort (**table S4**) and validation cohort (**table S5**), respectively. All metal conjugated antibodies were titrated and tested by assessing the relative marker expression intensity on relevant immune cell subsets in PBMCs from healthy individuals.

### **Sample staining and acquisition**

PBMCs were thawed at 37°C, immediately transferred into complete RPMI medium 10% hiFCS (fetal calf serum), 1% penicillin/streptomycin/glutamine, 10mM HEPES, 55µM 2-mercaptoethanol (2-ME) supplemented with 50 U/ml Benzonase (Sigma) and washed twice before staining. Samples undergoing tetramer staining (discovery cohort) were further sorted for live lymphocytes using an ARIA II flow sorting device. Sorted cells from each patient sample were split and transferred into two wells of a 96-well plate and each sample was stained with 100µl of tetramer cocktail (with a separate staining approach for each tetramer configuration) for 1h at RT. For antibody staining, all samples were incubated with primary metal-labelled surface antibody mixtures for 30 min on ice and incubated with 200 µM cisplatin during the last 5 min on ice for the discrimination of live and dead cells. Cells were washed and fixed in 2% paraformaldehyde in PBS overnight at 4 °C. For intracellular staining, cells were incubated in 1x permeabilization buffer (Biolegend) for 5 min on ice and incubated with metal-labelled intracellular antibodies for 30 min on ice. For samples undergoing an intranuclear staining step, cells were fixed using the Foxp3 transcription factor staining buffer set (eBioscience) after surface antibody and cisplatin staining. To minimize cell loss, samples with total cell counts below  $1 \times 10^6$  cells were further buffered using  $1 \times 10^6$  cells from healthy donor PBMCs (STEMCELL). Cells were then incubated with a mixture of all panel specific intracellular and intranuclear antibodies for 30 min on ice. Cells were washed and longitudinal samples, including samples stained with two tetramer configurations from the same patient, were barcoded with a unique combination of two distinct palladium barcodes for 30 min on ice. Cells were stained with 250 nM iridium intercalator (DNA staining) in 2% paraformaldehyde/PBS. Cells were washed, pooled together and adjusted to 0.5 million cells

per ml in H<sub>2</sub>O together with 1% equilibration beads (EQ Four element calibration beads, Fluidigm) for acquisition on a HELIOS mass cytometer (CyTOF, Fluidigm).

### **Multiplexed tetramer staining**

To screen for neoantigen-specific CD8 T cells, we set up a three-metal combinatorial tetramer staining approach, as described <sup>7, 8</sup>. Briefly, three out of ten different heavy metal-labeled streptavidins were randomly combined by using an automated pipetting device (TECAN) resulting in a total of 120 unique possible barcodes for single peptide candidates. For tetramerization, these combinations were incubated with single UV-exchanged peptide–MHC complexes <sup>9</sup> at a final molar ratio of 1:4 (total streptavidin:peptide–MHC). All samples were run in technical replicates by using a second configuration staining with a completely different barcoding scheme. For each staining configuration the tetramerized pMHC complexes were combined, concentrated in a cytometry buffer (PBS, 2% fetal calf serum, 2 mM EDTA, 0.05% sodium azide) and filtered before staining the cells.

### **CyTOF data and statistical analyses**

Mass cytometry signals for each parameter were normalized based on EQ beads and zero values were randomized using a custom R Script that uniformly distributed values between minus-one and zero. Each sample was manually de-barcoded followed by sequential gating on live (CD45<sup>+</sup> DNA<sup>+</sup> cisplatin<sup>-</sup>) CD8 T cells after gating out monocytes (CD14<sup>+</sup>) and B cells (CD14<sup>-</sup> CD19<sup>+</sup>), NK cells (CD14<sup>-</sup> CD19<sup>-</sup> CD3<sup>-</sup> CD56<sup>+</sup>), gd-T cells (CD14<sup>-</sup> CD19<sup>-</sup> CD3<sup>+</sup> gdTCR<sup>+</sup>) and CD4<sup>+</sup> T cells (CD14<sup>-</sup> CD19<sup>-</sup> CD3<sup>+</sup> gdTCR<sup>-</sup> CD4<sup>+</sup> CD8<sup>-</sup>) using FlowJo (Tree Star Inc) software. Gating for all phenotypic markers used in the present study was performed on total live CD45<sup>+</sup> immune cells for each patient separately. Cut-off values for the positive expression of each marker were defined based on the expression intensity on relevant immune cell subsets from the same

individual. Antigen-specific triple tetramer positive cells were identified by an automated peptide-MHC gating method <sup>8</sup> followed by an unbiased objective scoring system based on (i) detection cut-off thresholds, (ii) frequency correspondence between 2 tetramer staining configurations, (iii) background noise and (iv) phenotypic homogeneity as described earlier <sup>10</sup>. Phenotypic profiles and sample distributions were shown using principal component analysis (PCA), heatmaps, bar plots and uniform manifold approximation and projection (UMAP) for high-dimensionality reduction. Data analysis was performed using CYTOGRAPHER®, ImmunoScape's cloud based analytical software, custom R-scripts, Flowjo and GraphPad Prism. Comparative analysis of frequencies and marker expression levels of immune cell subsets between time points and patients were done using Wilcoxon (matched-pairs) signed rank test and paired t-tests where applicable, extended to Kruskal-Wallis tests by ranks for more than 2 levels in a grouping variable. P-values were adjusted for multiple testing using the Benjamini-Hochberg method to control the false discovery rate. Significance was set at a threshold of \*\* $p < 0.01$  and \* $p < 0.05$ . Correlations were calculated using the Pearson correlation coefficient.

### **RNA-seq 10X Genomics library construction and sequencing**

Frozen PBMCs were thawed, washed 2x in RPMI 2% FCS, treated with the ACK lysis buffer (Lonza) to remove red blood cells (RBCs), and briefly incubated with DAPI. 300,000 cells were then sorted on a DAPI negative gate, stained for 30 minutes at room temperature with a custom panel of 138 Total-Seq-C antibodies (Biolegend) <sup>11</sup>, and washed 3x using the HT1000 laminar wash system (Curiox). Cells were then counted using the Cellca MX High-throughput Automated Cell Counter (Nexcelom), pooled from 5 samples, and loaded on the 10x Chromium Next GEM Chip G Kit using a superloading strategy. TCR CDR3 sequences were enriched using human

V(D)J T cell enrichment. Libraries were prepared according to manufacturer's protocol (10x Genomics) and sequenced on a NovaSeq 6000 System using the S4 2x 150 kit (Illumina).

### **Preprocessing of scRNA-seq, CITE-seq, and TCR-seq data**

The scRNA-seq reads were aligned to the human transcriptome (GRCh38) and unique molecular identifier (UMI) counts were quantified to generate a gene-barcode matrix using the Cell Ranger pipeline (10X Genomics, version cellranger-5.0.1). CITE-seq (cellular indexing of transcriptomics and epitopes by sequencing) antibody expression matrices were generated using the Cell Ranger pipeline (10X Genomics, version cellranger-5.0.1). TCR reads were aligned to the GRCh38 reference genome and consensus TCR annotation was performed using the Cell Ranger vdj pipeline (10X Genomics, version cellranger-5.0.1). To assign cells to their respective samples of origin, cells were demultiplexed with a modified *HTOdemux* function from the *Seurat* package, whereby the negative cluster was defined by minimal non-zero expression. TCR clonality was calculated using the Gini coefficient index based on the TCR CDR3 nucleotide sequences derived from single cell analysis. The Gini coefficient measures the inequality of clonotype distribution within a repertoire and ranges from 0 to 1, with 0 representing complete equal clonal distribution and 1 complete expanded clonal distribution.

### **Differential gene expression analysis and cluster marker genes**

Differential gene expression analysis was performed using the Wilcoxon rank-sum test implemented in the *Seurat*. *FindMarkers* function was used to define the differentially expressed genes (DEGs) between cells from responders and those from non-responders. Marker genes for each cluster were identified by comparing cells in one cluster to cells in all other clusters using *FindAllMarkers* function. For DEGs and cluster marker genes, we required  $\text{ave\_logFC} > \log(1.25)$ , at least 10% of testing cells expressing the genes, and adjusted p values  $< 0.05$ .

### **Unsupervised clustering of CD57<sup>+</sup> CD8 T cells**

To investigate the cellular heterogeneity of CD57<sup>+</sup> CD8 T cells, we extracted all cells classified as CD8 T cells in the all-immune-cell analysis. To define CD57 positivity, a gating strategy was applied to CITE-seq data, which requires a normalized CD57 protein expression > 1.0. Based on this cutoff, 8,406 CD8 T cells were classified as CD57<sup>+</sup>, out of a total of 26,112 CD8 T cells profiled in this study. The expression matrix of CD57<sup>+</sup> CD8 T cells was imported into the Seurat (version 3.2.2) package<sup>12</sup> and normalized using the NormalizeData function with 'LogNormalize' method and 'scaled.factor' equal to 10,000. The data was then scaled and the effects of UMI counts and percent mitochondrial gene contents were regressed using the ScaleData function. Principal component analysis was performed on the scaled data cut to the 2,000 most variable genes identified by the FindVariableFeatures function with the 'vst' method. Sample batch effect correction was performed on the top 20 principal components using the RunHarmony function from the Harmony (version 1.0) package<sup>13</sup>. The cells were embedded in a shared nearest neighbor (SNN) graph based on the first 20 'harmony' dimensions and clustered using the FindClusters function with a resolution of 0.2. UMAP was generated using the RunUMAP function with the same 'harmony' dimensions used in clustering analysis.

### **Cluster analysis of all immune cells**

The processed expression matrix was subjected to several preprocessing steps. Genes that were expressed in less than 10 cells were removed. As a quality control step, cells were filtered based on the number of detected genes, number of detected UMIs, house-keeping gene expression and percentage of mitochondrial gene expression. Cells that expressed less than 10 house-keeping genes were removed. For UMIs, detected genes and mitochondrial gene expression, cutoffs were defined as the more conservative value between a hard predefined cutoff (UMI: lower 1000 - upper

20,000; Genes: lower 200 - upper 5,000; mitochondrial gene expression: 10%) and a dataset specific cutoff computed using interquartile ranges. RBC and platelet contaminants were removed via automated filtering algorithms. Expression counts of genes were divided by the total UMIs per cell and multiplied by a scaling factor of 10,000, and the log transformed with a pseudo count of 1 ( $\log(\text{TPM}/100 + 1)$ ). Surface proteins were normalized using the centered-log ratio (CLR) method. Variable genes were identified, and a dimensionality-reduced representation of the cells was created based on those variable genes. Batch effects were mitigated using the Harmony (version 1.0) package<sup>13</sup>. Shared nearest neighbors were computed and cells were then clustered using graph community clustering methods. Cells were annotated using a cell type classifier taking into account RNA, surface proteins and TCR sequences, and further validated and refined by using immunai's curated in-house signatures. Multi-omic data was further utilized to remove low quality cells and previously undetected doublets (e.g. cells which express both CD8 and CD4 protein tags, cells which express both a high B cell signature and have a detected TCR).

#### **TCR analysis from scRNA-seq data**

Single cell TCR clonotypes were grouped using exact CDR3 nucleotide sequence matching with both alpha and beta chains. For cells with more than one alpha and/or beta chains, sequence matching with all chains were required to be grouped in the same clonotype. TCR repertoire diversity was measured using the Gini index, which ranges from 0 (uniform clonal distribution) to 1 (uneven clonal distribution). To be included in the TCR diversity/clonality analysis, the sample needed to contain at least 10 cells with sequenced TCR. Percentage of T cells with clonal expansion  $\geq 2$  or  $\geq 10$  were calculated based on the frequency of clonotypes present at least twice or 10 times in a sample, respectively. For one of the patients (**Fig. 7D**), the overlap of tumor TCRB repertoire (the TCRB repertoire profiling of the tumor was conducted by the ImmunoSEQ assay;



Adaptive Biotechnologies) and corresponding TCRB repertoire in blood (from scTCR-seq data) was evaluated by comparing the CDR3 aa sequences.

#### **T-cell receptor (TCR) repertoire reconstruction from bulk RNA-seq data (TRUST4)**

To find shared TCR clonotypes between peripheral CD8 T cells and TILs, we extracted the TCR sequences from the tumor bulk RNA-seq data. Specifically, the TCRA repertoire of tumor samples was reconstructed by the TRUST4 algorithm using the bulk RNA-seq data (from baseline tumor tissue)<sup>14</sup>. The alignment of RNA-seq reads in BAM format alongside the genomic sequence and coordinate of V, J, and C genes, generated from the human IMGT database, were inputted into the TRUST4 and computations were run on a local HPC system.

#### **PD-L1 immunohistochemistry**

The pre-screening biopsies were collected from archived FFPE tissue. Patients were required to have tissue sent to the central laboratory before study entry. Samples were processed at the time of screening. FFPE tumor tissue was stained prospectively for PD-L1 by immunohistochemistry using a proprietary diagnostic anti-human PD-L1 monoclonal antibody (SP142) as described earlier<sup>15</sup>. Samples were scored for PD-L1 expression on tumor-infiltrating immune cells (IC), which included macrophages, dendritic cells and lymphocytes or on tumor cells (TC). PD-L1 scoring bins are IC0/TC0 <1%; IC1/TC1 ≥1% but <5%; IC2 >5% but <10%; TC2 ≥5% but <50% and IC3 ≥10% whereas TC3 ≥50%. PD-L1 scores in patients with multiple specimens from different time points or samples were based on the highest score. This assay was validated for investigational use in clinical trials.

#### **Tumor mutation burden (TMB) calculation**

TMB was determined using the FDA approved FoundationOne CDx™ (F1CDx) platform<sup>16</sup>. TMB was defined as the number of somatic, coding, base substitution, and indel mutations per megabase

of the genome examined. All base substitutions and indels in the coding region of targeted genes, including synonymous alterations, are initially counted before filtering as described below. Synonymous mutations are counted in order to reduce sampling noise. Non-coding alterations were not counted. Alterations listed as known somatic alterations in COSMIC and truncations in tumor suppressor genes were not counted, since our assay genes are biased toward genes with functional mutations in cancer. Alterations predicted to be germline by the somatic-germline-zygosity algorithm were not counted. Alterations that were recurrently predicted to be germline in our cohort of clinical specimens were not counted. Known germline alterations in dbSNP were not counted. Germline alterations occurring with two or more counts in the ExAC database were not counted. To calculate the TMB per megabase, the total number of mutations counted was divided by the size of the coding region of the targeted territory.

## Supplementary Figures

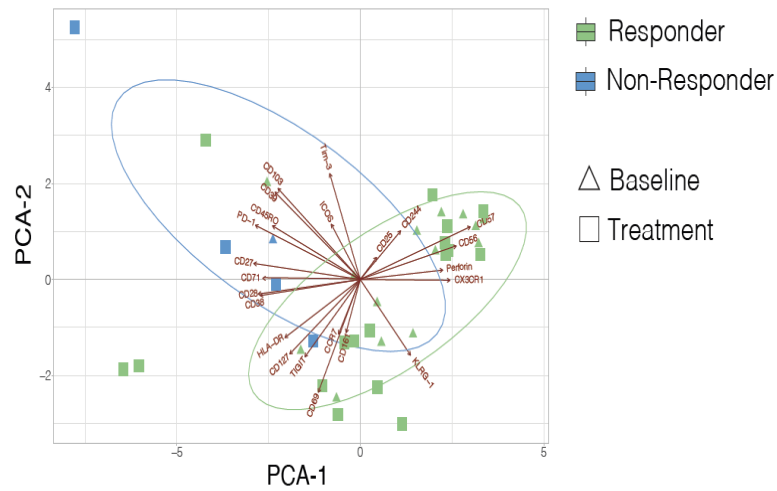


Fig. S1. Phenotypes of neoantigen-specific CD8 T cells in atezolizumab-treated responder and non-responder patients. PCA of all neoantigen-specific CD8 T cell populations identified in the discovery cohort. PCA is based on phenotypic profiling (percent of antigen-specific CD8 T cells positive for all phenotypic markers assessed).

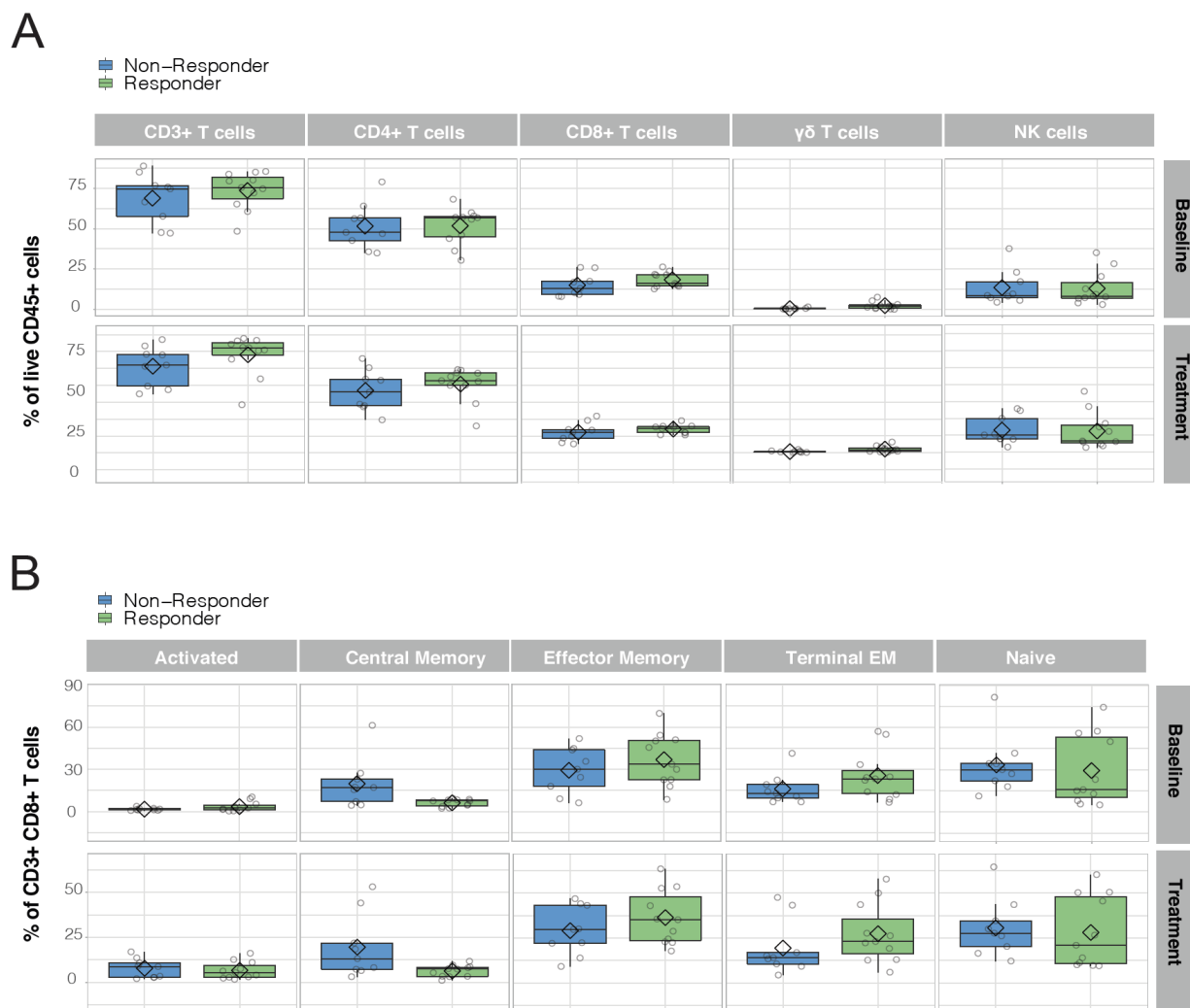


Fig. S2. Immune cell subset frequencies in atezolizumab-treated responder and non-responder patients. **(A)** Frequencies of CD3<sup>+</sup> T cells, CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells,  $\gamma\delta$ -T cells and NK cells at baseline (top) and on atezolizumab (bottom) treatment (discovery cohort). **(B)** Frequencies of major CD8 T cell subsets (naïve: CD45RO<sup>-</sup>, CCR7<sup>+</sup>; central memory (CM): CD45RO<sup>+</sup>, CCR7<sup>+</sup>; effector memory (EM): CD45RO<sup>+</sup>, CCR7<sup>-</sup>; terminal EM: CD45RO<sup>-</sup>, CCR7<sup>-</sup>; and activated cells: CD38<sup>+</sup>, HLA-DR<sup>+</sup> at baseline (top) and on atezolizumab (bottom) treatment (discovery cohort).

Wilcoxon rank sum test. P values were adjusted for multiple testing using the Benjamini-Hochberg method to control the false discovery rate.

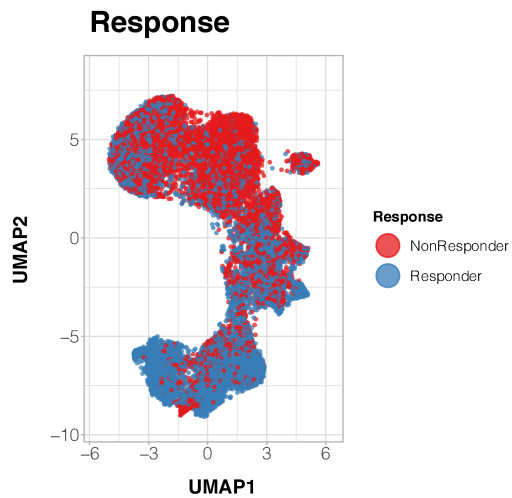
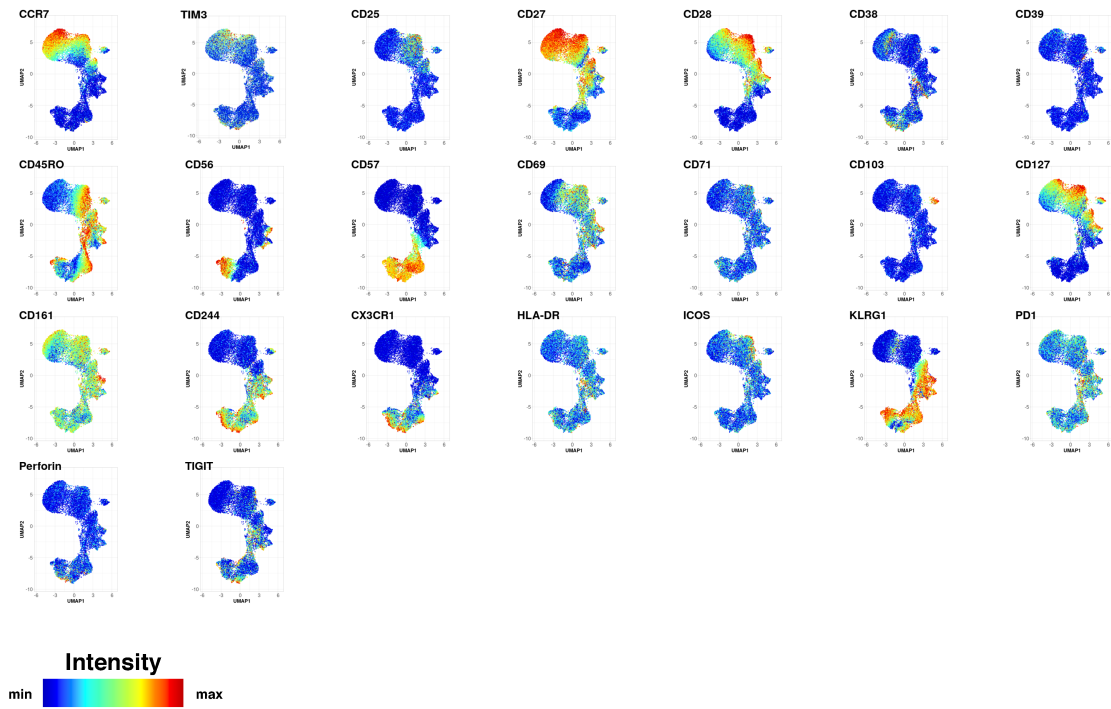
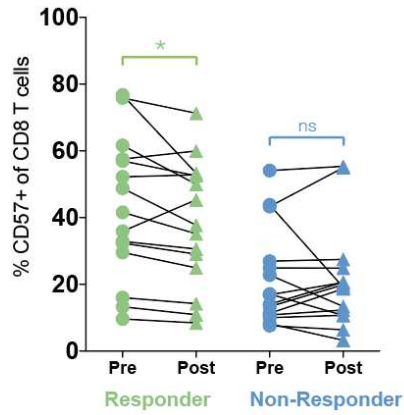


Fig. S3. High-dimensional immune profiles of CD8 T cells from atezolizumab-treated responder and non-responder patients. UMAP plots display relative expression intensities of all phenotypic

markers assessed (top). Shown are combined UMAP plots from CD8 T cells from all responder and non-responder patients at baseline (discovery cohort), total cell number 20,000 per plot. Bottom plot shows visualization of responder and non-responder CD8 T cells on the two-dimensional UMAP plot.

**A**



**B**

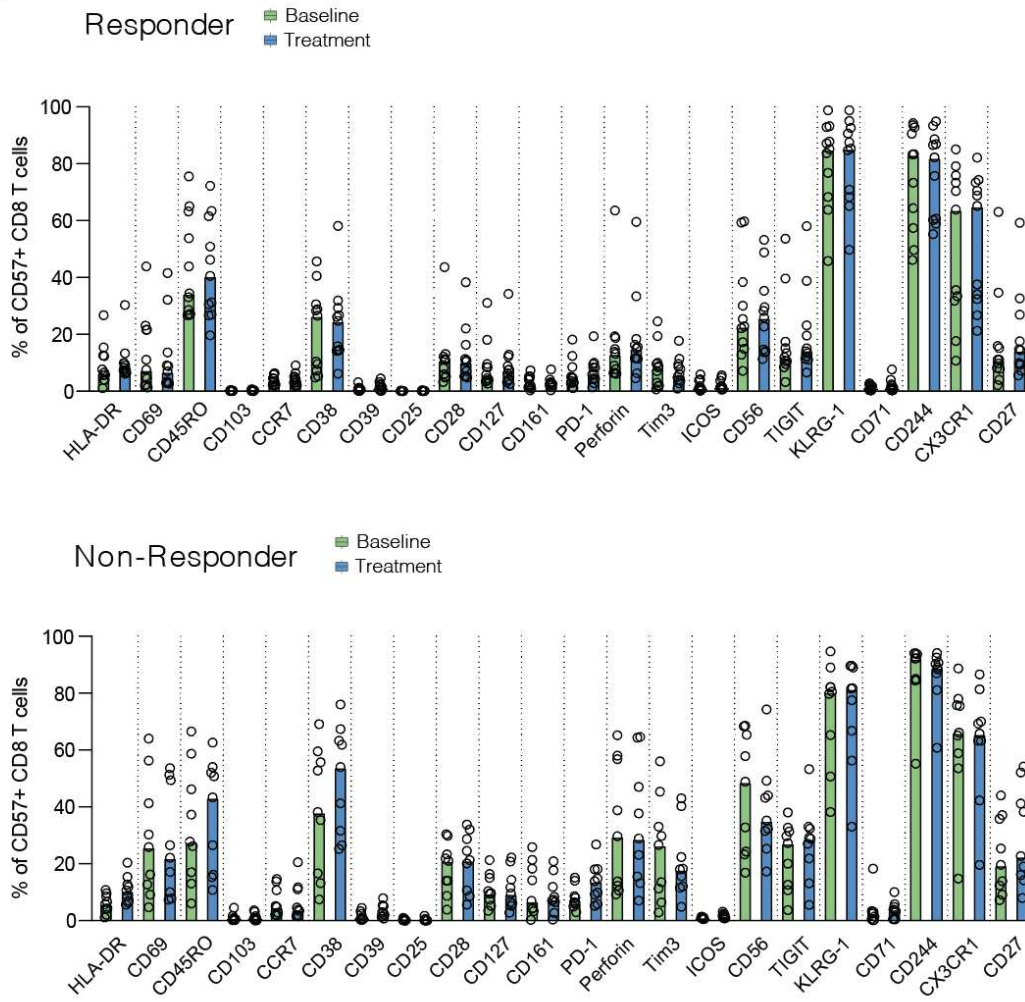




Fig. S4. Frequencies and phenotypes of CD57<sup>+</sup> CD8 T cells from atezolizumab treated responder and non-responder patients at baseline and on-treatment (3 weeks post atezolizumab treatment).

(A) CD57<sup>+</sup> CD8 T cells are enriched in atezolizumab-treated responder patients. CD57 expression on CD8 T cells from responders and non-responders at baseline (pre) and on-treatment (post). Validation cohort.  $p^* < 0.05$ . Ns; non-significant. Wilcoxon matched-pairs signed rank test. (B) Phenotypic marker expression on CD57<sup>+</sup> CD8 T cells from responder and non-responder patients is similar at baseline and on-treatment. Shown are the frequencies of CD57<sup>+</sup> CD8 T cells positive for 22 phenotypic marker molecules in responders (n=11, top) and non-responders (n=9, bottom) assessed at baseline and on-treatment (median values). Wilcoxon matched-pairs signed rank test (paired t-test where applicable). P values were adjusted for multiple testing using the Benjamini-Hochberg method to control the false discovery rate.

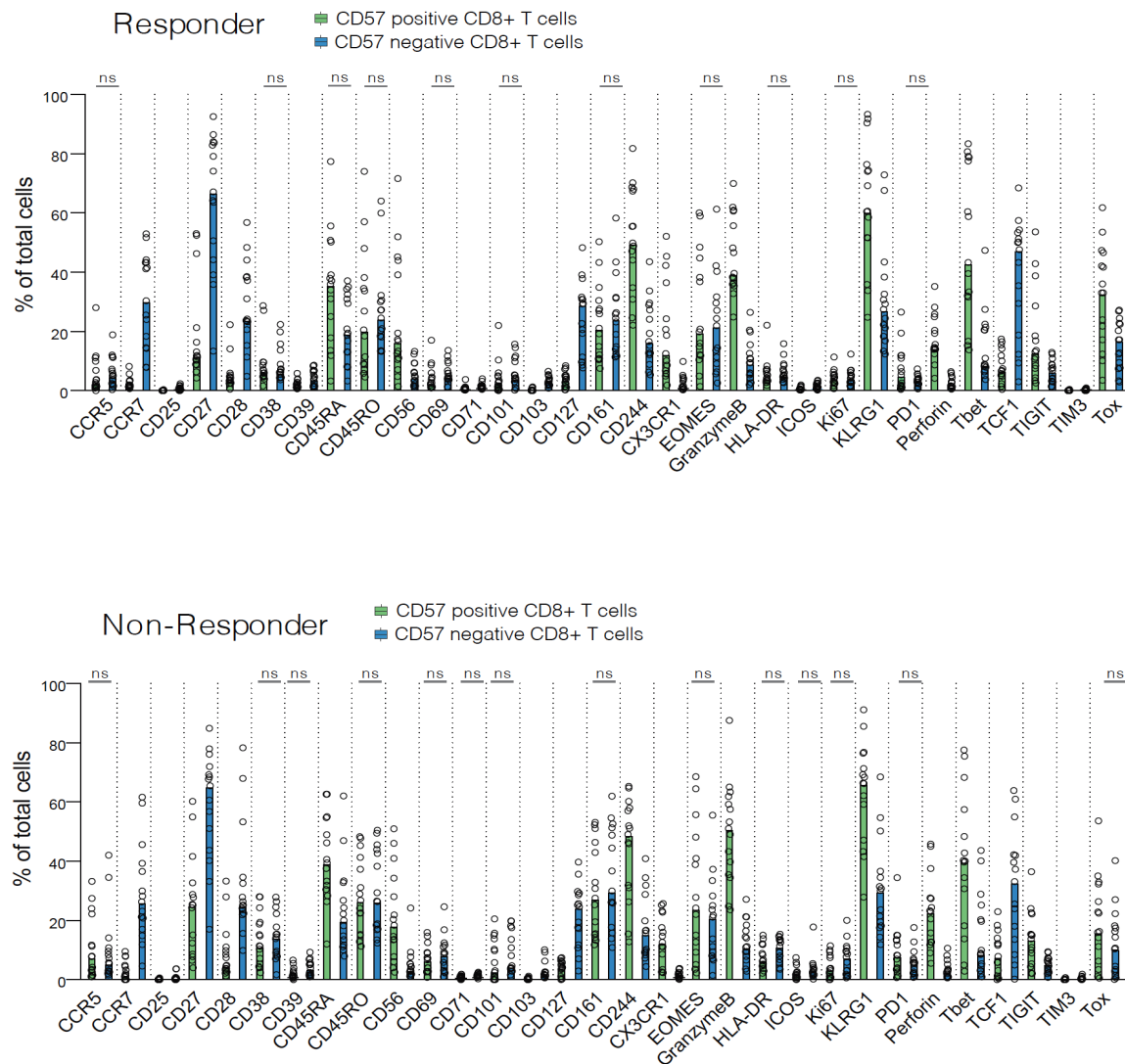
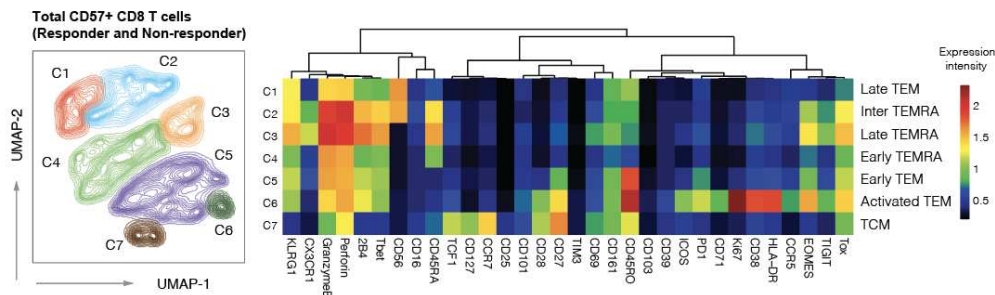


Fig. S5. CD57<sup>+</sup> CD8 T cells in responder and non-responder patients are skewed towards a higher differentiation compared to CD57<sup>-</sup> CD8 T cells. Frequencies of CD57<sup>+</sup> and CD57<sup>-</sup> CD8 T cells positive for 31 phenotypic marker molecules in responders (n=15, top) and non-responders (n=15, bottom) assessed at baseline. Data shown are median values. Ns; non-significant. Unless ns is indicated, each marker is significantly different between CD57<sup>+</sup> and CD57<sup>-</sup> CD8 T cells.

Wilcoxon rank sum test. P values were adjusted for multiple testing using the Benjamini-Hochberg method to control the false discovery rate.

A



B

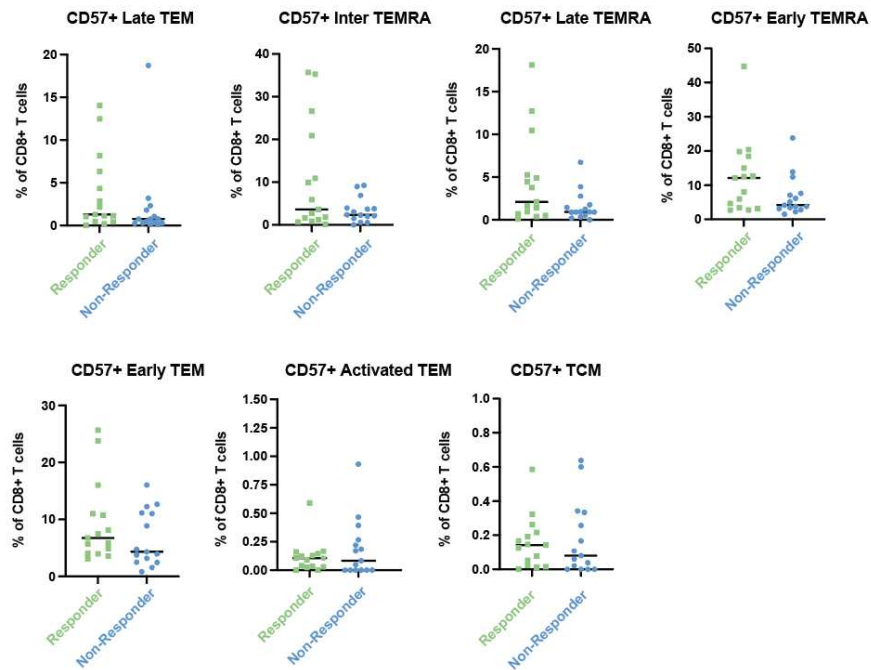


Fig. S6. High-dimensional profiles of CD57<sup>+</sup> CD8 T cells in baseline blood skew toward early memory effector phenotypes in atezolizumab-treated responding patients at baseline. **(A)** UMAP representation and delineated heatmap of combined phenotype clusters based on PhenoGraph automated clustering on CD57<sup>+</sup> CD8 T cells from responder and non-responder patients. In total, seven clusters could be delineated based on the differential median expression levels of each phenotypic marker assessed. **(B)** Frequencies of cells found for all patients across the clusters defined in (A) in relation to their total CD8 T cell counts. Responder patients show trends toward higher frequencies of early effector memory and early terminally effector memory T cells among the CD57<sup>+</sup> CD8 T cell population. T<sub>EM</sub>; effector memory T cells, T<sub>EMRA</sub>; terminally effector memory T cells, T<sub>CM</sub>; central memory T cells. Wilcoxon rank-sum test.

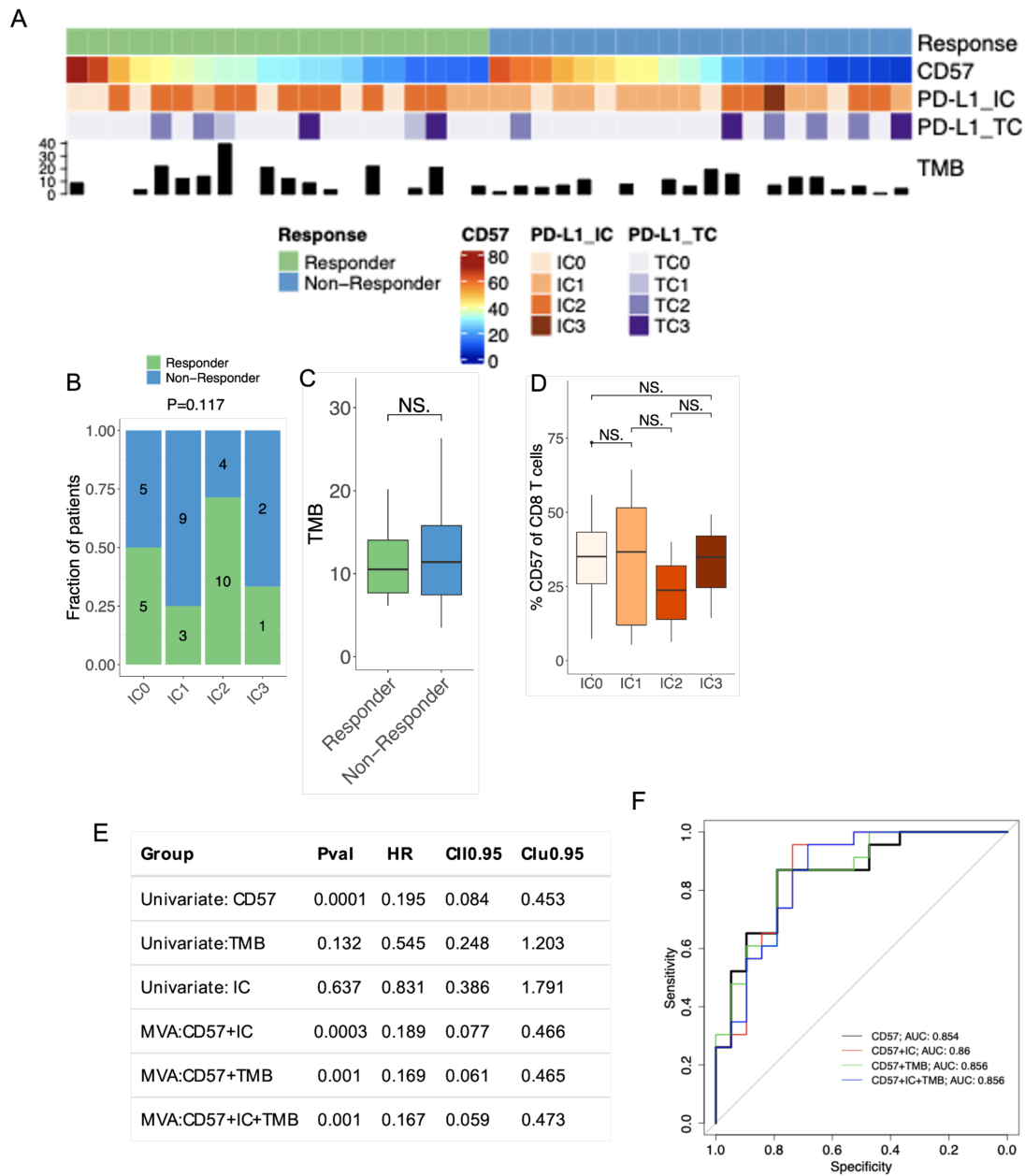


Fig. S7. Comparison of CD57<sup>+</sup> CD8 T cells, PD-L1 immune cell (IC) and tumor cell (TC) scores, tumor mutation burden (TMB) in chemotherapy-treated responding and non-responding patients, and survival analysis from atezolizumab-treated patients. (A) Heatmap showing frequency of CD57<sup>+</sup> peripheral blood CD8 T cells, PD-L1 IC and TC scores as well as TMB score in responders

and non-responders in the chemotherapy arm. Patients with PD-L1 IC and TC together with TMB evaluable population are pooled and shown here in responders (n=20) and non-responders (n=20). **(B)** Response rate by PD-L1 IC score. Bar chart represents the proportion of responders to non-responders by PD-L1 IC score status. P value is calculated by Chi-square test. **(C)** TMB score in responder and non-responder patients. Bar chart represents the median TMB score in each response group. P value between responders and non-responders is calculated by Wilcoxon rank-sum test. Ns; non-significant. **(D)** Frequency of CD57<sup>+</sup> peripheral blood CD8 T cells in patients separated based on PD-L1 IC score status. P values between IC groups are calculated by Wilcoxon rank-sum test. Ns; non-significant. **(E)** Tables showing univariate and multivariate logistic regression analyses of CD57 with different biomarkers (PD-L1 IC and/or TMB) added as covariates in atezolizumab-treated patients. HR and CI are calculated using stratified Cox proportional hazard regression models, and p values are calculated using a log-rank test. **(F)** Receiver operating characteristic (ROC) curve showing CD57 levels and other biomarkers (PD-L1 IC and/or TMB) added as covariates in predicting response in atezolizumab monotherapy. The diagonal line represents random expectation. Area under the curve (AUC) values are displayed in the figure legend.

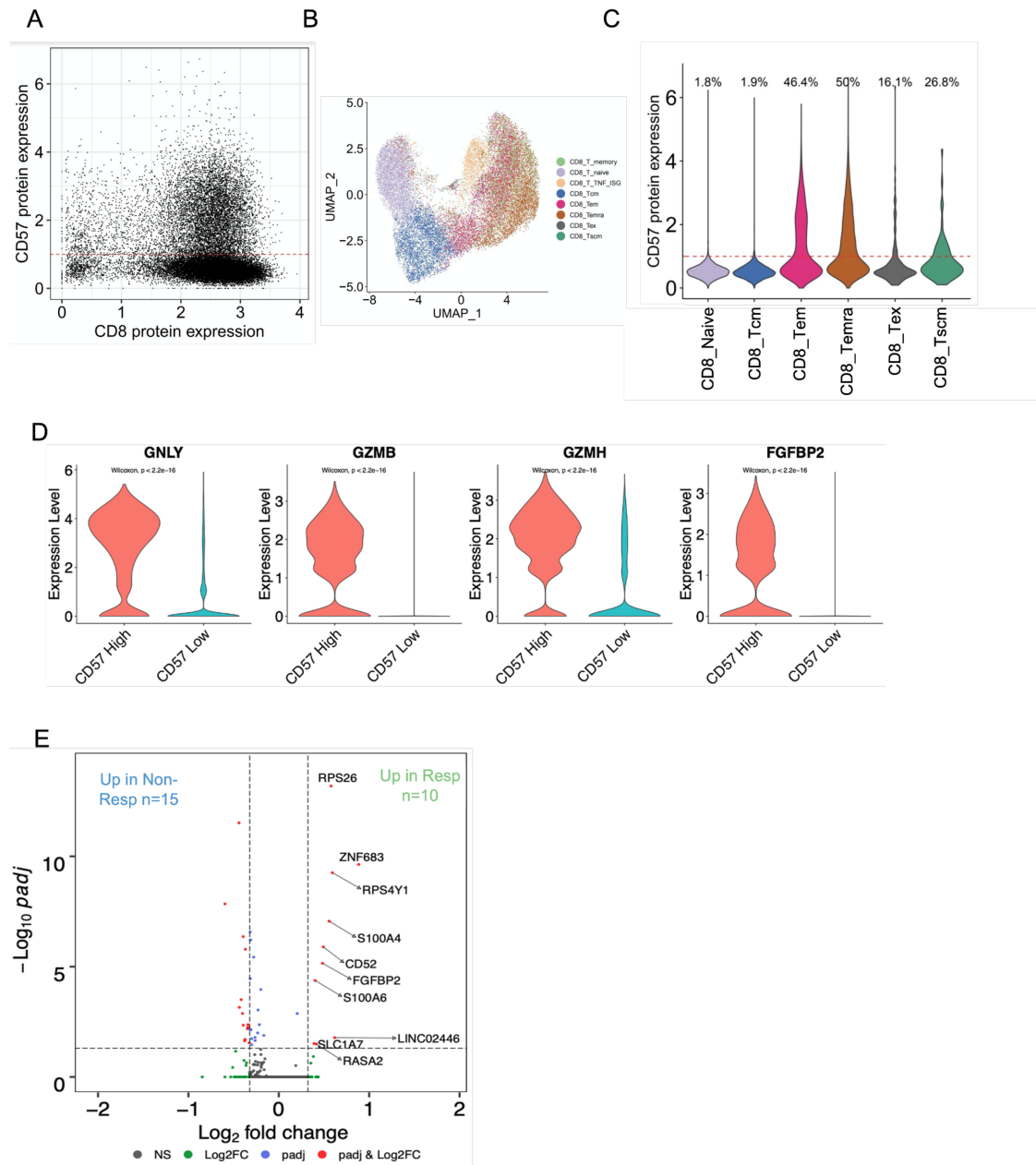


Fig. S8 CD57 expression is different in T cell subsets determined using CITE-seq. (A) Scatter plot showing the gating strategy for CD57 expressing CD8 T cells using CITE-seq analysis. Normalized CD8 and CD57 protein expressions are shown on the x-axis and y-axis, respectively. The red line represents the expression cutoff at 1.0. (B) UMAP projection of CD8 T cells

(n=26,112). All CD8 T cells profiled with scRNA-seq and CITE-seq are shown. **(C)** Violin plots showing CD57 expression in various CD8 T cell subsets as shown in (B). Percentage of CD57<sup>+</sup> CD8 T cells for each cell subset is shown at the top. **(D)** Expression of selected genes associated with CD8 T effector functions in CD57<sup>+</sup> vs CD57<sup>-</sup> CD8 T cells. **(E)** Volcano plot showing differentially expressed genes (DEGs) in the C1 cluster of CD57<sup>+</sup> CD8 T cells from responders and non-responders at baseline. DEGs are nominated by requiring at least 1.25 times fold change and an adjusted p value of < 0.05 with gene expression detected in at least 10% of cells in either one of the two comparison groups. Upregulated DEGs in responders are labeled in the volcano plot and red dots represent DEGs.



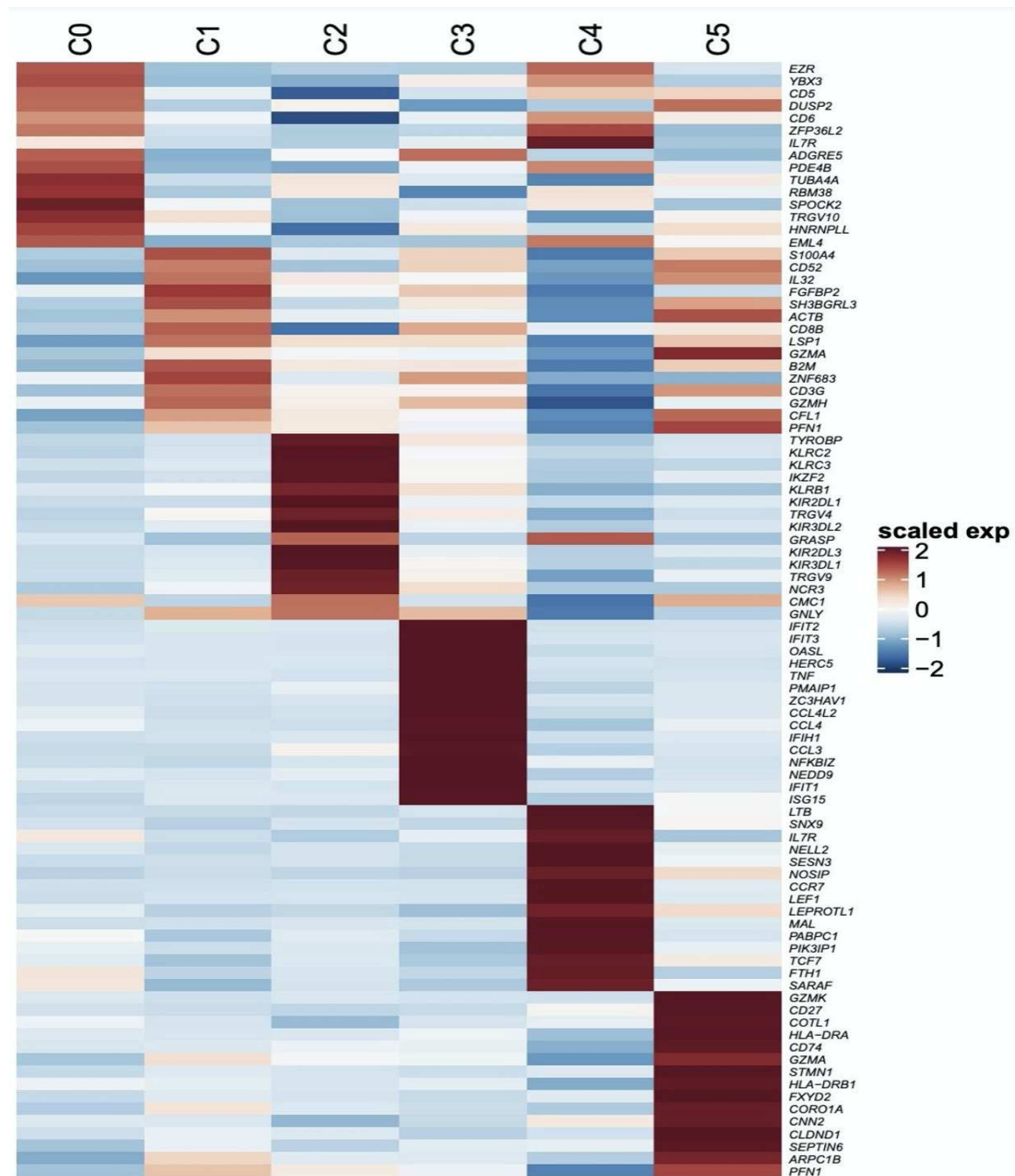


Fig. S9. A, Heatmap showing the scaled expression of top 15 cell markers ranked by fold change for each cell cluster identified among CD57<sup>+</sup> CD8 T cells in Fig 6D.

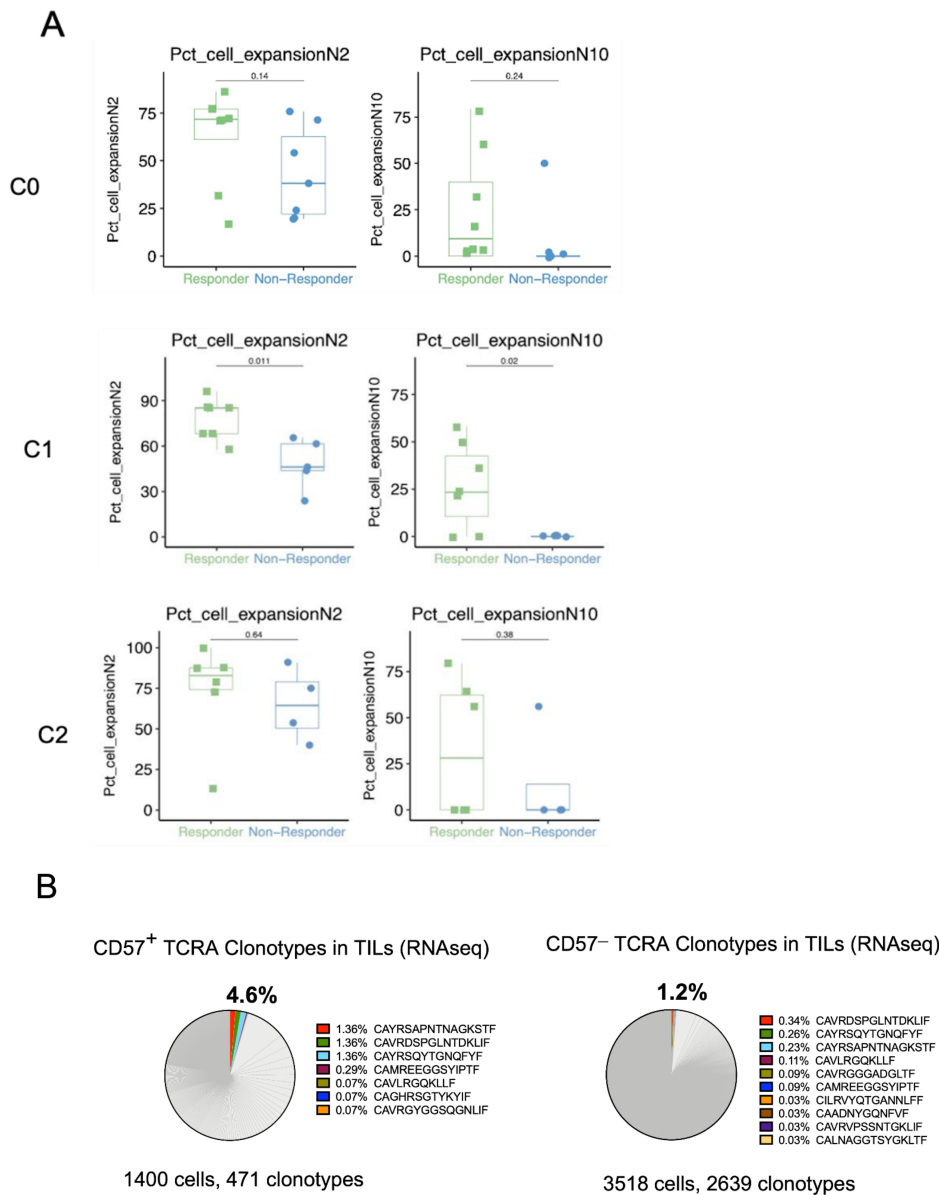


Fig. S10. Clonal expansion and overlap with TILs among CD57<sup>+</sup> CD8 T cells responders compared to non-responders. (A) Box plots showing the proportion of clonal cells in responders and non-responders at baseline in clusters C0 (top), C1 (middle) and C2 (bottom) identified among CD57<sup>+</sup> CD8 T cells (referring to Fig. 7C). The clonotypes are categorized as clonal based on the cell numbers ( $n \geq 2$ , left;  $n \geq 10$ , right). Percentage of cells belonging to clone size of  $\geq 2$  or  $\geq 10$  are

shown. Two-tailed unpaired Student's t-test. **(B)** Frequency distribution of peripheral CD57<sup>+</sup> (left) and CD57<sup>-</sup> (right) CD8 T cells based on individual TCRA clonotype sequences. TIL TCRA were reconstructed from bulk RNA-seq. Each slice size represents the percentage of cells with individual TCR clonotypes. TCR sequences that overlap between blood and tumor are highlighted in color.

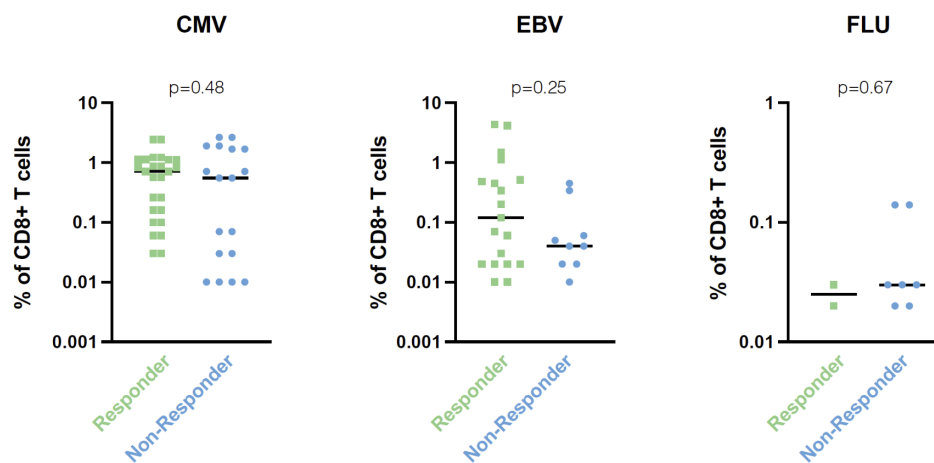


Fig. S11. Frequencies of common virus-specific T cells in responding and non-responding patients. Antigens in the screening approach included epitopes derived from Cytomegalovirus (CMV), Epstein Barr Virus (EBV) or Influenza (FLU). Data shown are median values of all antigen specificities detected. Wilcoxon rank-sum test.

## References

1. Powles T, Duran I, van der Heijden MS, et al. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*. 2018;391(10122):748-757. doi: 10.1016/S0140-6736(17)33297-X.
2. Shukla SA, Rooney MS, Rajasagi M, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol*. 2015;33(11):1152-8. doi: 10.1038/nbt.3344.
3. Saunders CT, Wong WS, Swamy S, et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-7. doi: 10.1093/bioinformatics/bts271.
4. Wilm A, Aw PP, Bertrand D, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. 2012;40(22):11189-201. doi: 10.1093/nar/gks918.
5. Lawrence M, Gentleman R. VariantTools: an extensible framework for developing and testing variant callers. *Bioinformatics*. 2017;33(20):3311-3313. doi: 10.1093/bioinformatics/btx450.
6. Karosiene E, Lundegaard C, Lund O, et al. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics*. 2012;64(3):177-86. doi: 10.1007/s00251-011-0579-8.
7. Fehlings M, Chakarov S, Simoni Y, et al. Multiplex peptide-MHC tetramer staining using mass cytometry for deep analysis of the influenza-specific T-cell response in mice. *J Immunol Methods*. 2018;453:30-36. doi: 10.1016/j.jim.2017.09.010.

8. Newell EW, Sigal N, Nair N, et al. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat Biotechnol.* 2013;31(7):623-9. doi: 10.1038/nbt.2593.
9. Rodenko B, Toebe M, Hadrup SR, et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat Protoc.* 2006;1(3):1120-32. doi: 10.1038/nprot.2006.121.
10. Fehlings M, Jhunjhunwala S, Kowanetz M, et al. Late-differentiated effector neoantigen-specific CD8<sup>+</sup> T cells are enriched in peripheral blood of non-small cell lung carcinoma patients responding to atezolizumab treatment. *J Immunother Cancer.* 2019;7(1):249. doi: 10.1186/s40425-019-0695-9.
11. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods.* 2017;14(9):865-868. doi: 10.1038/nmeth.4380.
12. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell.* 2019;177(7):1888-1902 e21. doi: 10.1016/j.cell.2019.05.031.
13. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* 2019;16(12):1289-1296. doi: 10.1038/s41592-019-0619-0.
14. Song L, Cohen D, Ouyang Z, et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat Methods.* 2021;18(6):627-630. doi: 10.1038/s41592-021-01142-2.
15. Vennapusa B, Baker B, Kowanetz M, et al. Development of a PD-L1 Complementary Diagnostic Immunohistochemistry Assay (SP142) for Atezolizumab. *Appl Immunohistochem Mol Morphol.* 2019;27(2):92-100. doi: 10.1097/PAI.0000000000000594.
16. Hellmann MD, Ciuleanu TE, Pluzanski A, et al. Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. *N Engl J Med.* 2018;378(22):2093-2104. doi: 10.1056/NEJMoa1801946.

