

# **Tumour-intrinsic features shape T cell differentiation through precursor to symptomatic multiple myeloma**

Kane A. Foster, Benny Chain, Sergio A. Quezada, Kwee L. Yong et al.

## **Supplementary Information (SI)**

### **1. Supplemental Figures**

**Supplementary Fig. 1** Acquisition of single-cell RNA sequencing cohort.

**Supplementary Fig. 2** Pre-processing, integration, and immunophenotyping of scRNA-seq dataset.

**Supplementary Fig. 3** Pre-processing, integration, and immunophenotyping of T cells.

**Supplementary Fig. 4** Sub-cluster, functional, inter-tissue, and re-projection analyses.

**Supplementary Fig. 5** Quantification of changes in T cell abundance in non-cancer controls, MGUS, SMM and MM patients.

**Supplementary Fig. 6** Analysis of T cell receptor repertoire structure and specificity.

**Supplementary Fig. 7** Recurrent tumour cell transcriptional dynamics in tumour cells.

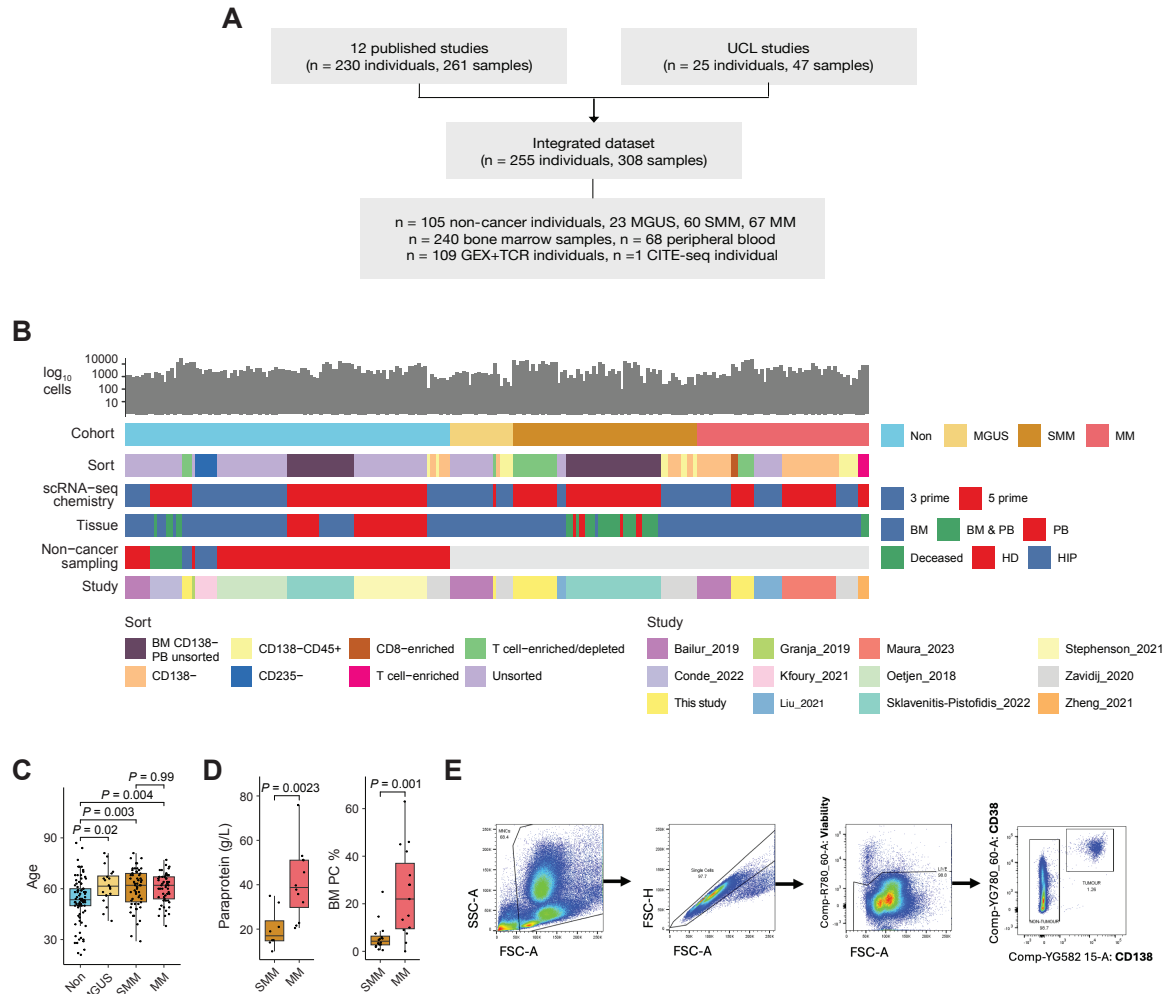
**Supplementary Fig. 8** Tumour-intrinsic drivers of T cell differentiation.

**Supplementary Fig. 9** T cell gating strategies.

**Supplementary Fig. 10** Extended SMM progression analysis.

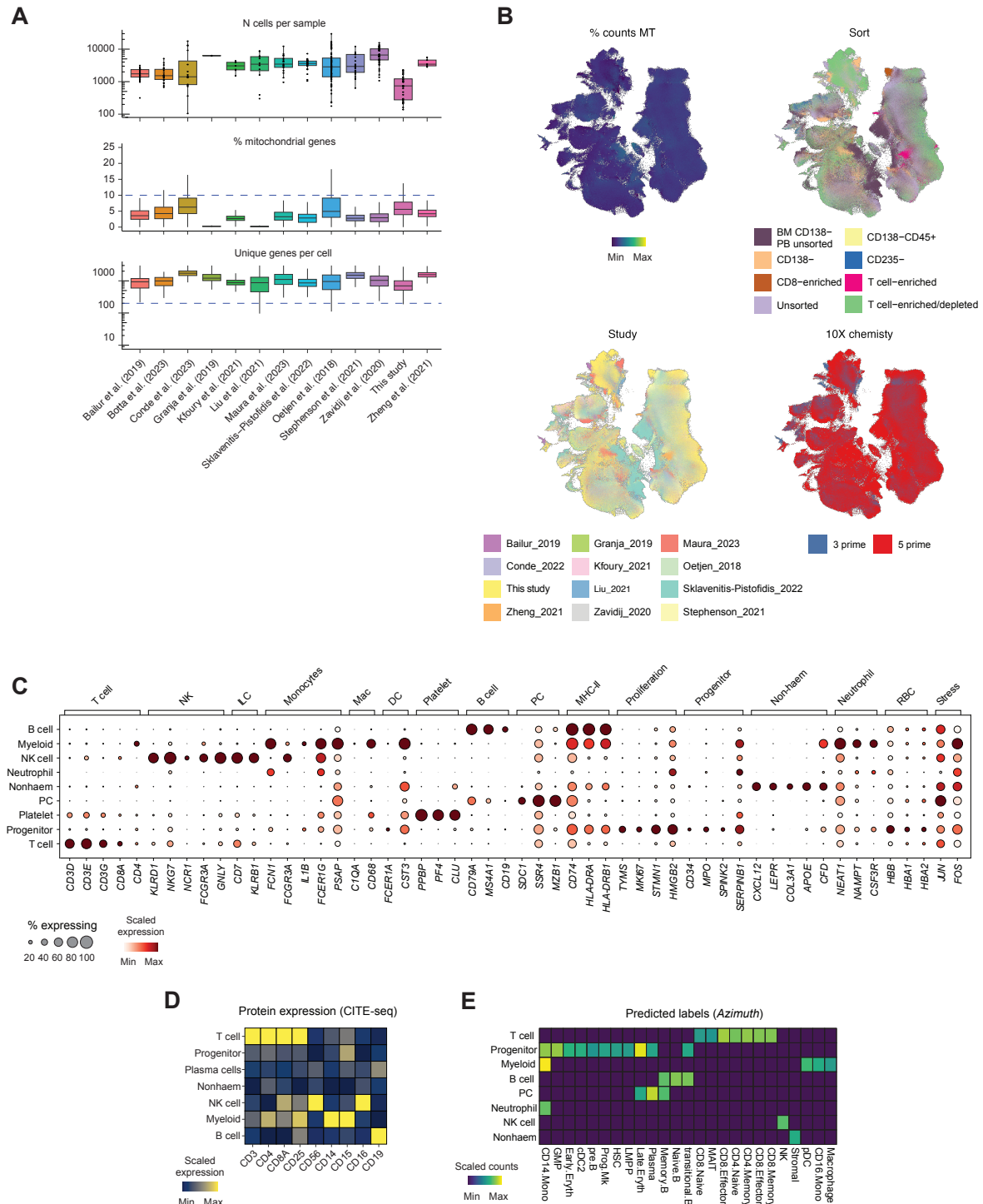
## Supplemental figures

### Supplementary Fig. 1. Acquisition of single-cell RNA sequencing cohort.



**(A)** Sample collection for scRNA-seq cohort. **(B)** Heatmap representing sample metadata for the individuals included in the study. BM, bone marrow; PB, peripheral blood; HD, healthy donor; HIP, hip replacement surgery donor; Deceased, sampled posthumously. **(C)** Box plots comparing the age of non-cancer controls (Non), MGUS, SMM and MM patients. **(D)** Box plots comparing the serum paraprotein (g/L, left) and BM PC % (right) between SMM and MM patients confirming that both are higher in the MM group. **(E)** Gating strategy used to calculate BM PC % in primary bone marrow samples. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. For (C)  $P$ -values derived by one-way ANOVA followed by Tukey's test. For (D)  $P$ -values derived by two-sided Wilcoxon test.

## Supplementary Fig. 2 Pre-processing, integration, and immunophenotyping of scRNA-seq dataset.

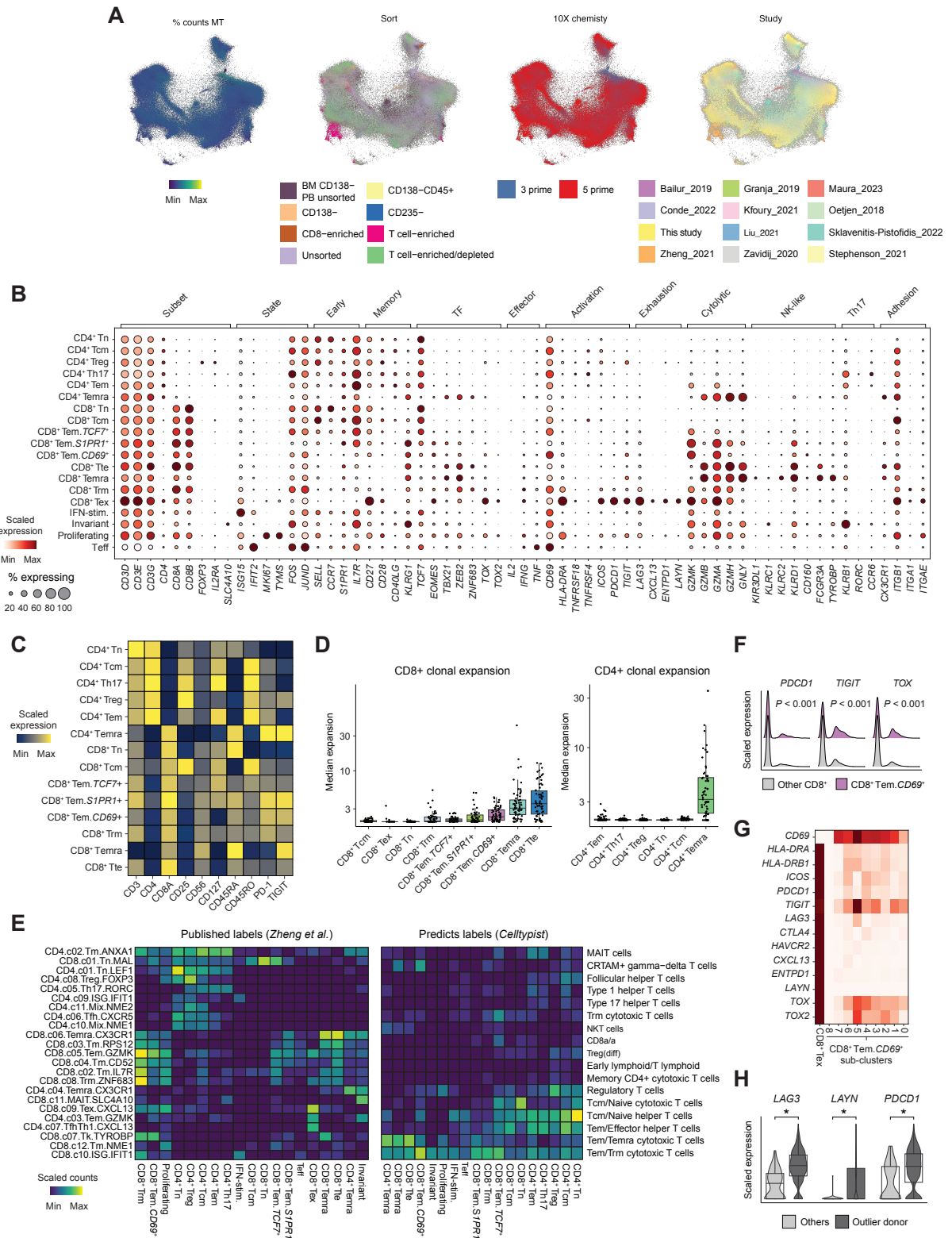


(A) Box plots comparing the number of cells, fraction of mitochondrial genome-derived genes, and number of unique genes per-cell across different studies. (B) Visualisation of all cells by uniform manifold approximation and projection (UMAP). Individual cells are coloured by the proportion of mitochondrial genes (% counts MT) and sample sorting strategy, study of origin, and 10X chemistry. (C) Dot plot of gene expression in all cell type clusters. The mean expression of each gene is scored from low (white) to high (red), with the percentage of cells

expressing each gene represented by the size of each point. Nonhaem, non-haematopoietic cell; PC, plasma cell; ILC, innate lymphoid cell; DC, dendritic cell; RBC, red blood cell. **(D)** Heatmap showing scaled average protein expression derived by cellular indexing of transcriptomes and epitopes sequencing (CITE-seq) in cell type clusters derived from Granja et al. **(E)** Heatmap showing the total number of each predicted label (Azimuth, see Methods) for each cell type cluster. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR.



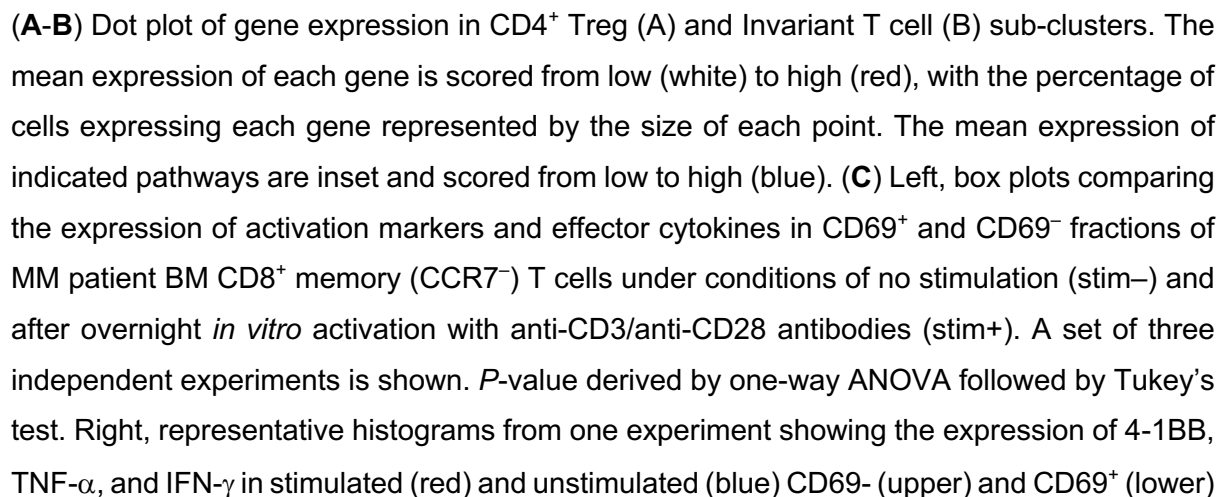
## Supplementary Fig. 3 Pre-processing, integration, and immunophenotyping of T cells.



(A) Visualisation of T cells by minimum-distortion embedding (MDE). Individual cells are coloured by the proportion of mitochondrial genes (% counts MT) and sample sorting strategy, study of origin, and 10X chemistry. (B) Dot plot of gene expression in T cell clusters. The mean expression of each gene is scored from low (white) to high (red), with the percentage of cells

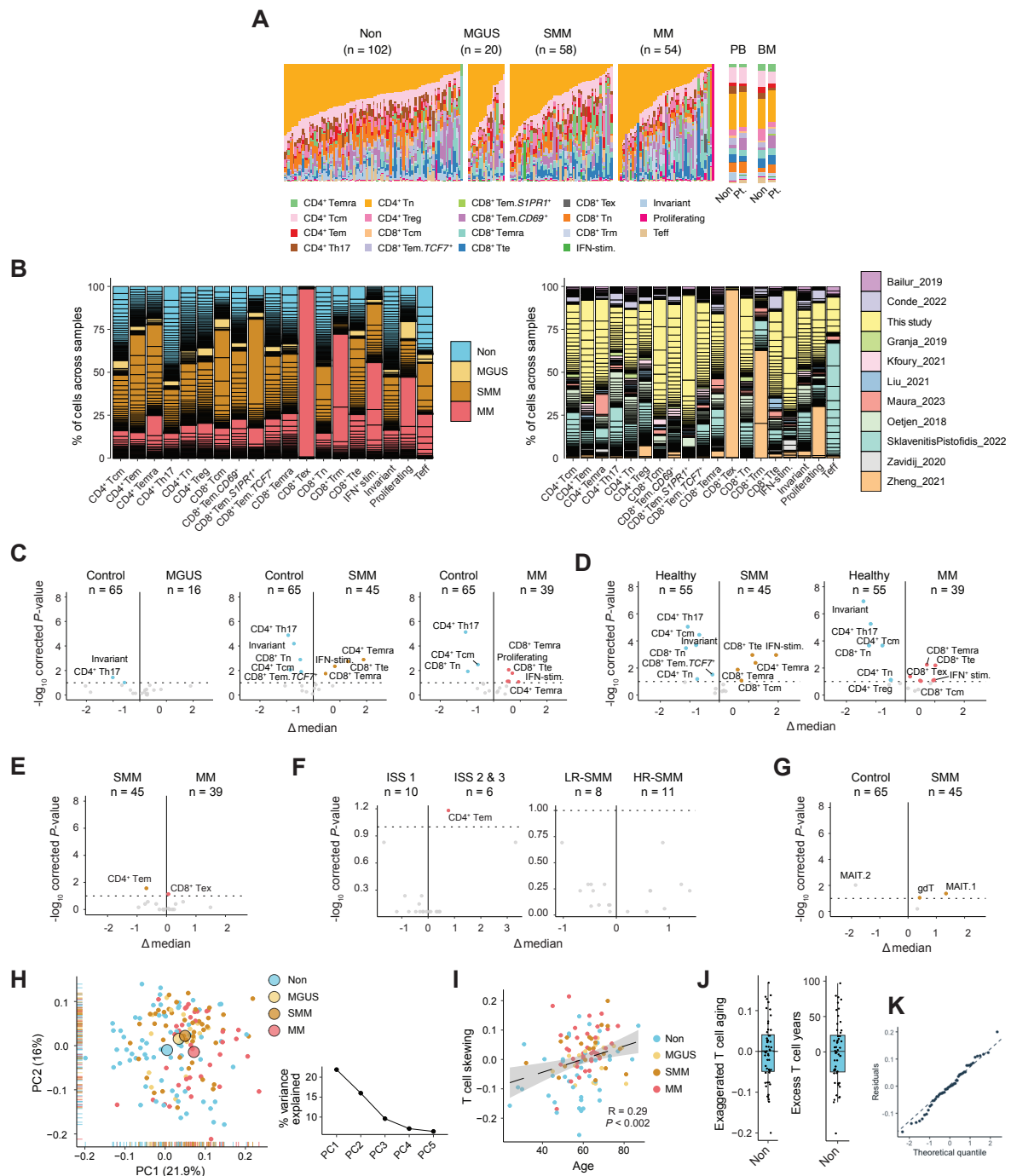
expressing each gene represented by the size of each point. TF, transcription factor; Th17, T-helper 17 (C) Heatmap showing scaled average protein expression derived by cellular indexing of transcriptomes and epitopes sequencing (CITE-seq) in T cell clusters derived from Granja et al. (D) Box plot showing the median TCR clonal expansion per-patient for each CD8<sup>+</sup> (left) and CD4<sup>+</sup> (right) cluster. (E) Heatmap showing the total number of each published label (Zheng et al., left) and predicted label (Celltypist, right) for each T cell cluster. (F) Ridge plot comparing exhaustion-associated gene expression CD8<sup>+</sup> Tem.CD69<sup>+</sup> and other CD8<sup>+</sup> clusters. (G) Heatmap showing scaled average RNA expression of exhaustion-associated markers in CD8<sup>+</sup> Tem.CD69<sup>+</sup> sub-clusters and CD8<sup>+</sup> Tex. (H) Box and violin plots comparing the expression of indicated T cell exhaustion-associated genes in CD8<sup>+</sup> Tex cells derived from the single outlier donor and all other donors. P-values calculated by two-sided Wilcoxon test, \* =  $P < 0.001$ .

A



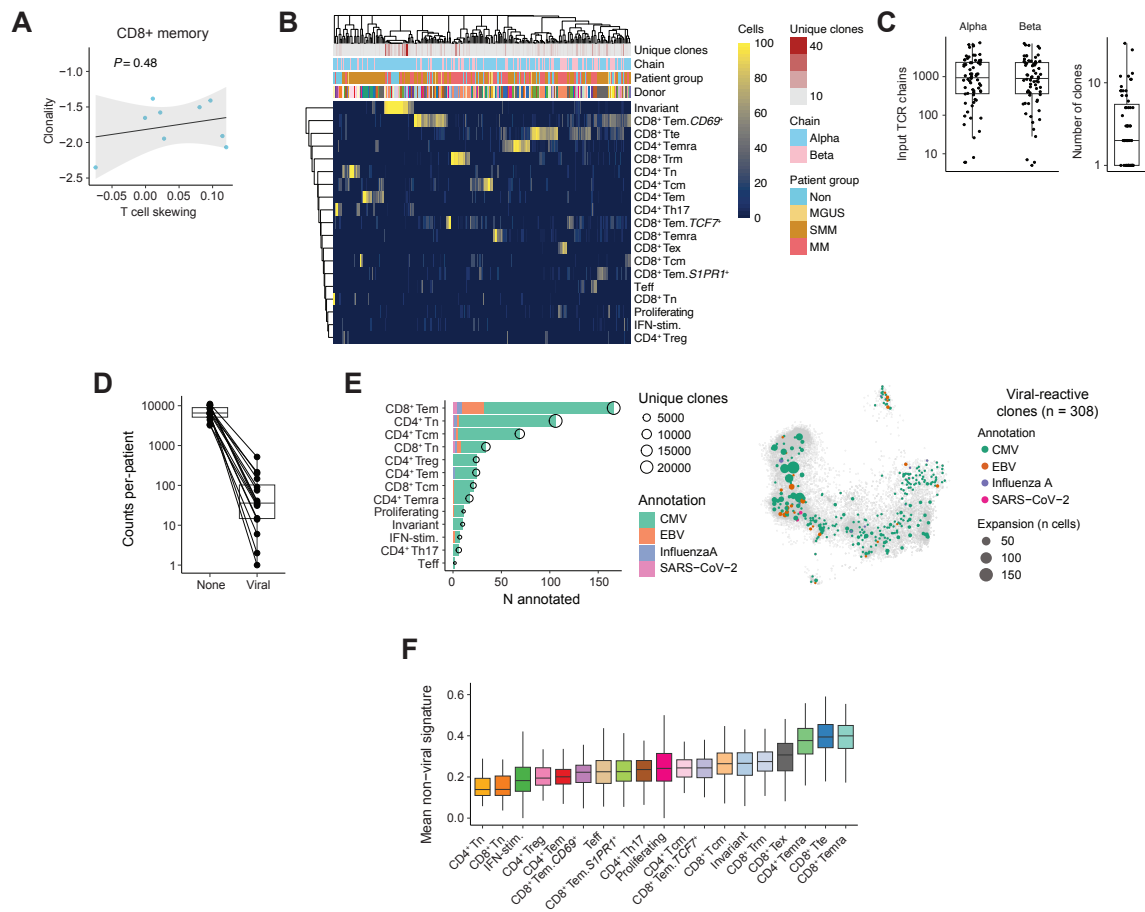
cells. **(D)** Visualisation of T cells by minimum-distortion embedding (MDE) in across tissues (rows) and clinical group (columns) with colour representing different T cell clusters. For each group a sample of 10,000 cells is shown. **(E)** Dot plot showing corrected *P*-values and difference in the median normalised abundance ( $\Delta$  median) of T cell clusters and tissue (bone marrow, BM; peripheral blood, PB) for non-cancer controls (left) and SMM and MM patients (right). To account for the fact a subset of samples were taken from donors with both PB and BM samples, corrected *P*-values were calculated with a linear mixed-effects regression model with donor as a random effect and Holm–Bonferroni correction. **(F)** Box plot showing the abundance of CD8<sup>+</sup> Tem.CD69<sup>+</sup> in PB and BM samples from all controls and patients (left, *P*-value calculated by unpaired two-sided Wilcoxon test) and controls and patients with paired PB and BM samples (right, *P*-value calculated by paired two-sided Wilcoxon test). **(G)** Dot plot of gene expression of T cell clusters in validation dataset (Botta et al.). Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR.

**(A)** Bar chart comparing T cell cluster composition in non-cancer controls (Non), MGUS, SMM and MM patients (left) and in the peripheral blood (PB) and bone marrow (BM) of all controls and patients (Pt.; right). **(B)** Bar chart showing the abundance of each T cell cluster. Individual bar segments represent individual patients and are coloured by cohort (left) or study of origin (right). **(C-F)** Analysis of the abundance of T cell cluster abundance between non-cancer controls and specified patient group (C), between healthy donors sampled for research and SMM or MM patients (D), SMM and MM patients (E), and ISS or SMM (Mayo) risk groups (F).



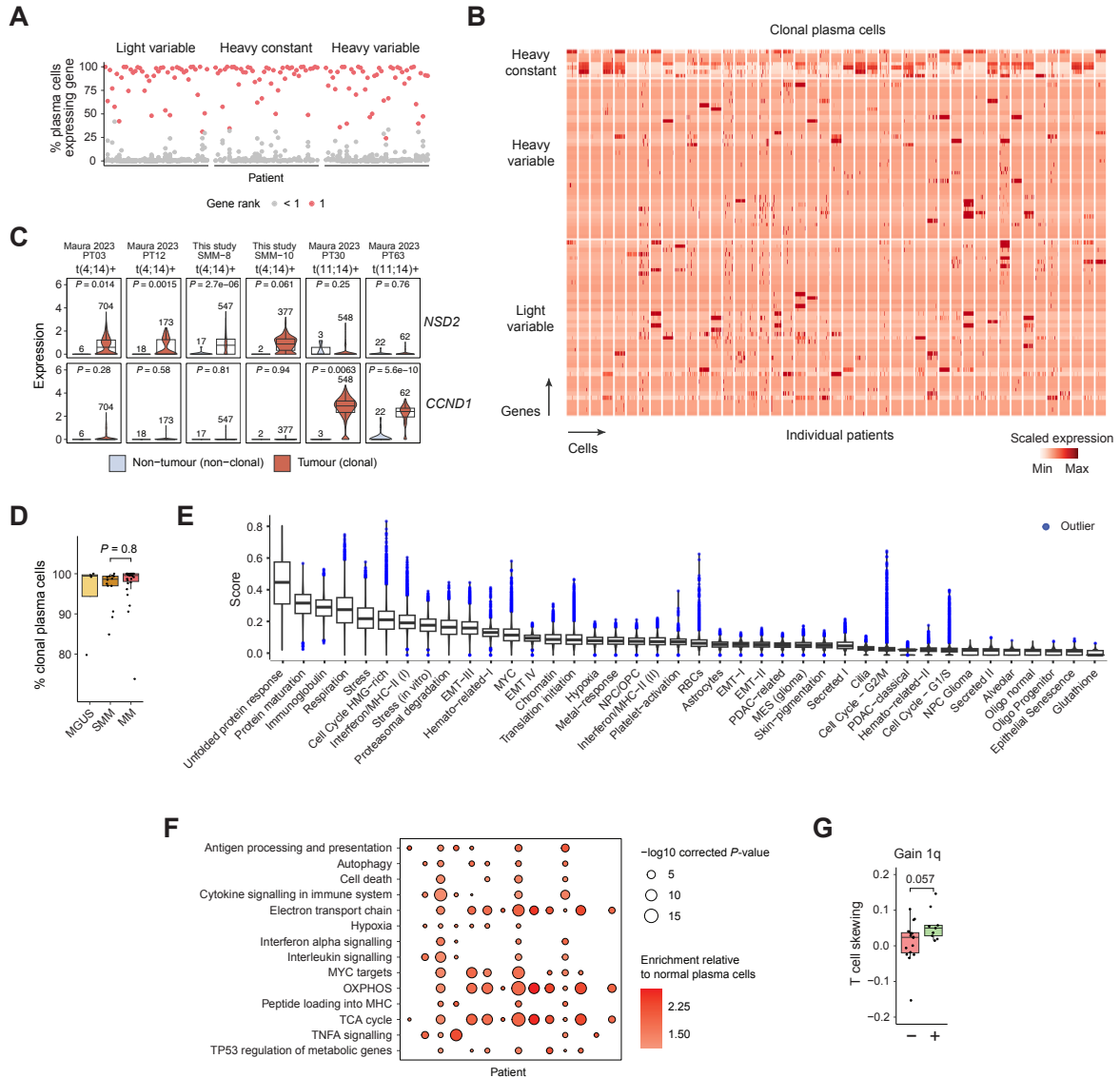
The number of individuals per group is inset. Only individuals with age information are included in this comparison. Corrected *P*-values and the difference in the median normalised abundance ( $\Delta$  median, as a percentage of T cells) for each cluster is represented as volcano plots. To account for patient age, corrected *P*-values were calculated by a linear regression with an intercept term for age and Holm–Bonferroni correction. Horizontal dashed line represents the corrected *P*-value threshold of 0.1. Clusters significantly enriched in either condition are labelled. **(G)** Volcano plot as in (B-E) showing the changes in Invariant sub-cluster abundance between controls and SMM patients. **(H)** Left, dot plot showing the first two principal components (PCs) calculated using normalised T cell abundance in all individuals. Larger points represent the average PC1 and PC2 value for each group. Right, scree plot showing the percentage variance explained by the first five principal components. **(I)** Dot plot showing the correlation between T cell skewing and age (years) in non-cancer controls and MGUS, SMM and MM patients. **(J)** Box plot showing exaggerated T cell ageing and excess T cell years (see Methods) in non-cancer control patients. **(K)** Quantile-quantile plot used to assess the normality of residuals in the linear model assessing the effect of age on T cell skewing. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. *R* and *P*-values were calculated by Pearson correlation.

## Supplementary Fig. 6 Analysis of T cell receptor repertoire structure and specificity.



(A) Dot plots showing the lack of correlation between T cell skewing and clonality ( $\log_{10} 1/\text{Simpson's diversity}$ ) in CD8<sup>+</sup> memory clones from non-cancer controls ( $n=9$ ). (B) Heatmap showing the T cell cluster composition (rows) of 279 individual TCR clusters (columns). The number of unique clones, TCR chain, patient group, and donor of origin for each individual TCR cluster is shown above the heatmap. (C) Box plots showing the number of input alpha and beta chains (left) and number of clusters (right) per donor. (D) Box plot showing the number of clones with viral annotation or no annotation in patient TCR repertoires. Each pair of points connected by a line represents an individual patient. (E) Left, bar chart showing the fraction of annotated TCR specificities against different viruses across different T cell clusters. Right, multi-dimensional embedding plot showing the specificity and expansion of clones predicted to be viral-reactive (see Methods). (F) Box plots showing the median expression of the non-viral specificity gene signature among T cell clusters. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR.

## Supplementary Fig. 7 Recurrent tumour cell transcriptional dynamics in tumour cells.

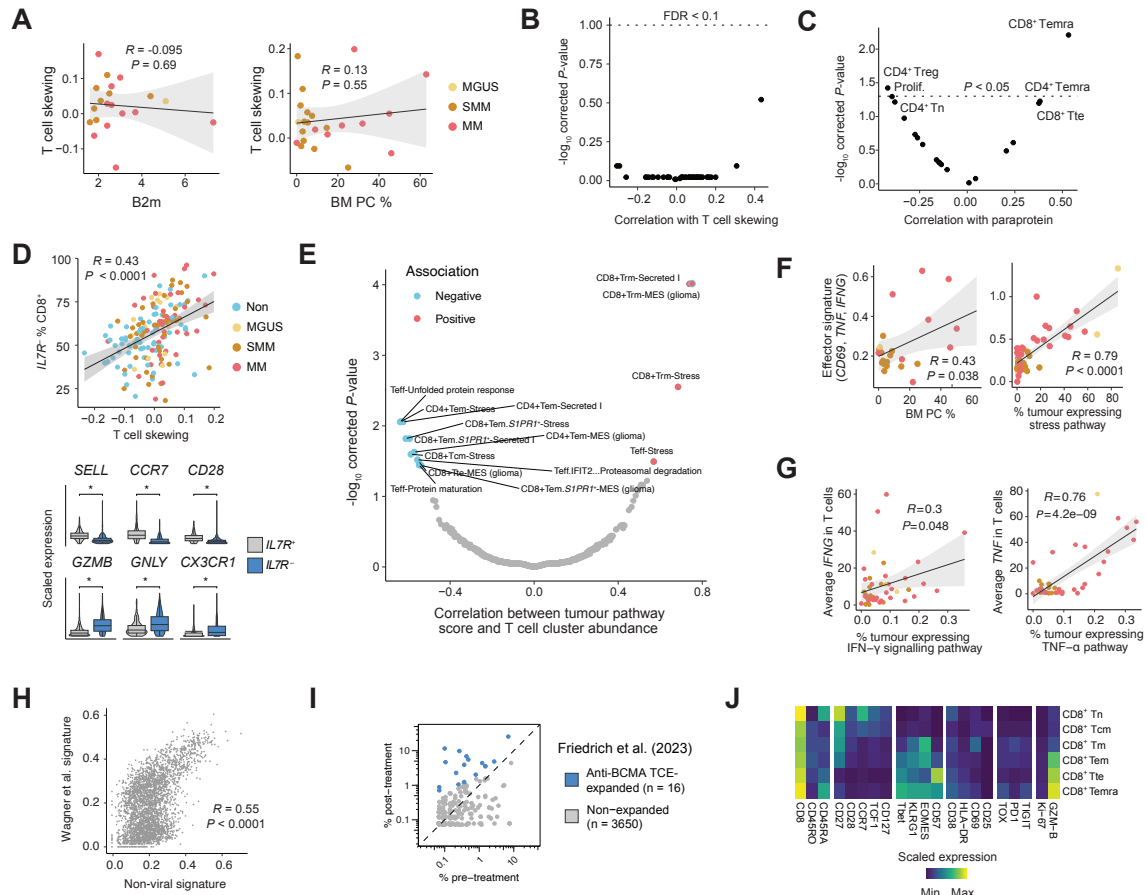


(A) Dot plot showing, for each patient ( $n=46$ ), the % of plasma cells expressing each light variable, heavy constant, and heavy variable chain genes. The chain with the highest expression (rank 1) in each patient is shown in red. (B) Heatmap showing expression of immunoglobulin genes in plasma cells predicted to be tumour cells based on clonal expression of immunoglobulin genes. Each column represents a cell and each row a gene. Columns and rows are split by individual donor and immunoglobulin chain, respectively. A random subset of 50 plasma cells called as tumour cells per donor are shown. (C) Box and violin plots showing the expression of *NSD2* (upper) and *CCND1* (lower) in plasma cells classified as tumour (clonal immunoglobulin usage, red) or non-tumour (non-clonal immunoglobulin usage, grey) cells in six patients with available genomic translocation information. The number of tumour or non-tumour cells is inset above each box plot. The translocation information (t(4;14) or t(11;14)) is inset below each patient identifier. *NSD2* is characteristic of t(4;14) positive tumours and *CCND1* of t(11;14) positive tumours. (D) Box plot showing the proportion of



plasma cells with clonal immunoglobulin usage in MGUS ( $n=4$ ), SMM ( $n=15$ ) and MM ( $n=27$ ) patients. **(E)** Box plots showing the expression of indicated pan-cancer transcriptional pathways in 67,048 individual tumour cells from all patients. Cells with expression values classified as outliers are shown in blue. **(F)** Heatmap showing the enrichment of *de novo* transcriptional pathways on cancer cells relative to non-cancer plasma cells across patients selected from batches with sufficient non-tumour plasma cells (see Methods). Corrected *P*-values were calculated by gene set enrichment analysis and Holm–Bonferroni correction. **(G)** Box plot showing T cell skewing in patients with confirmed 1q gain positivity (+) and negativity (–). Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. For (C), and (G) *P*-values derived by two-sided Wilcoxon test. For (C), *P*-values derived by one-way ANOVA followed by Tukey’s test.

## Supplementary Fig. 8 Tumour-intrinsic drivers of T cell differentiation.

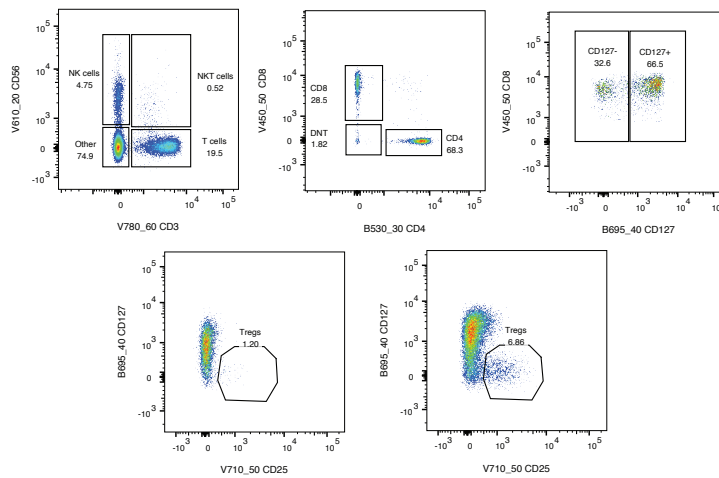


(A) Dot plots showing the absence of significant correlation between T cell skewing and serum B2m (left) and BM PC % (right). (B) Dot plot showing corrected  $P$ -values and correlation coefficients between patient PC1 values and individual pan-cancer transcriptional pathway expression in tumour cells. No significant correlations were present after  $P$ -value correction. (C) Dot plot showing corrected  $P$ -values and correlation coefficients between patient Paraprotein values and the abundance of T cell clusters. The dashed line indicates the  $P$ -value threshold of 0.05. Clusters which significantly (corrected  $P < 0.05$ ) correlated with Paraprotein are labelled. (D) Dot plot showing the correlation between T cell skewing and the abundance of  $IL7R^-$  CD8<sup>+</sup> T cells in scRNA-seq samples. The expression of indicated genes in  $IL7R^+$  and  $IL7R^-$  cells is inset below with \* indicating  $P < 0.001$ .  $P$ -values derived by two-sided Wilcoxon test. (E) Dot plot showing corrected  $P$ -values and correlation coefficients between T cell cluster abundance and pan-cancer transcriptional pathway expression in tumour cells. T cell clusters and tumour pathway which were significantly associated are labelled. (F) Dot plots showing the correlation between the average expression of an effector T cell gene signature ( $CD69$ ,  $IFNG$ ,  $TNF$ ) among all T cells with BM PC % (left) and the abundance of cancer cells highly expressing the stress pathway (right). (G) Dot plots showing the correlation between indicated effector molecules ( $IFNG$  and  $TNF$ ) in T cells and the abundance of cancer cells highly expressing downstream signalling pathways for each

effector molecule. **(H)** Dot plot showing the correlation between the expression of the non-viral gene signature and the Tumor-reactive Features in T cells gene signature from Wagner et al. (2025). **(I)** Dot plot showing the abundance of TCR clones in pre- and post-treatment samples from 3 patients treated with T cell engager (TCE) therapy (Friedrich et al., 2023). Clones with significantly expanded post-treatment are shown in blue. **(J)** Heatmap showing the scaled expression of T cell marker proteins in CD8<sup>+</sup> T cell clusters identified by Cytometry by Time-of-Flight (CyTOF; see Methods). Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. For (A), (D, upper), (F) and (G) *R* and *P*-values were calculated by Pearson correlation. For (B), (C) and (E) corrected *P*-values were calculated by Pearson correlation and Holm–Bonferroni correction.

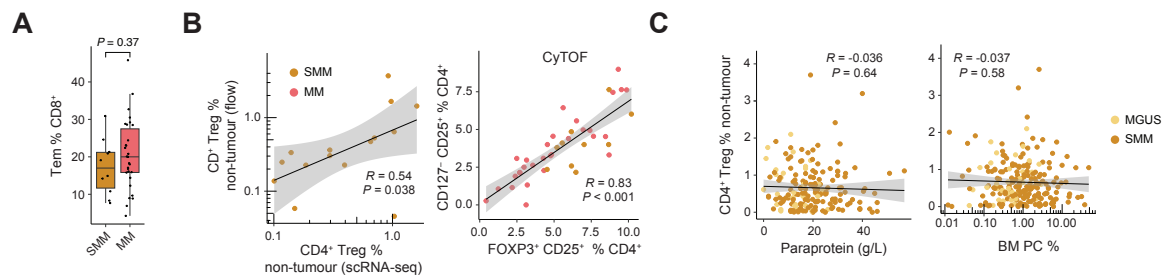
**(A-C)** Gating strategy for samples asses by Cytometry by time-of-flight (CyTOF; A), disease evolution flow cytometry cohort (B), autologous stem cell transplantation flow cytometry cohort (C).

**B**



Flow cytometry plots showing the isolation of NK cells from a mixed population. The process starts with a mixed population (FSC-A vs FSC-H) and proceeds through several gates to isolate NK cells (CD56+). The final plot shows NK cells (CD56+).

# Supplementary Fig. 10 Extended SMM progression analysis.



(A) Box plot comparing the abundance of Tem among CD8<sup>+</sup> T cells in SMM ( $n=9$ ) and MM ( $n=28$ ) patients assayed by Cytometry by time-of-flight (CyTOF).  $P$ -value derived by two-sided Wilcoxon test. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. (B) Dot plots showing the positive correlation between CD4<sup>+</sup> Treg abundance in matched scRNA-seq and flow cytometry samples (left) and different gating strategies in CyTOF samples (right). (C) Dot plots showing the absence of correlation between serum paraprotein and BM PC % with CD4<sup>+</sup> Treg as a % of non-tumour cells. For (B) and (C)  $R$  and  $P$ -values were calculated by Pearson correlation.