nature communications

Article

https://doi.org/10.1038/s41467-026-68718-4

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

Received: 22 June 2025

Accepted: 15 January 2026

Published online: 05 February 2026

Check for updates

Kane A. Foster[1], Elise Rees[1], Louise Ainley[1,2], Annabel Laidler[1], Eileen M. Boyle[1,2], Lydia Lee[1,2], Gwennan Ward[1], Daria Galas-Filipowicz[1], Evelyn Fitzsimons[1], Anna Mikolajczak[1], Emma J. Lyon[1], Dylan Jankovic[1], Jasmine Rahman[1], Mahima Turakhia[1], Dipal Mehta[1,2], Conor Garrod-Ketchley[1], Imran Uddin[3,4], Gordon Beattie[3,4], Yvette Hoade[1], Catherine Zhu[1,2], James L. Reading[5,6], Ieuan Walker[7,8], Michael Chapman[7,8], Karthik Ramasamy[9], Javier Herrero[10], Benny Chain[11,12] ✉, Sergio A. Quezada[6,13] ✉ & Kwee L. Yong[1,2] ✉

Multiple myeloma (MM) is associated with skewed T cell activation and function which is present in asymptomatic myeloma precursor conditions, but underlying mechanisms of progression remain undefined. Here, we assemble a large single-cell RNA sequencing dataset of the bone marrow and blood from patients with MM, precursor conditions, and non-cancer controls. We demonstrate that, unlike solid cancers, MM is not characterized by T cell exhaustion, but by antigen-driven terminal memory differentiation. This is influenced by tumour-intrinsic features including tumour burden and expression of antigen-presentation genes. Expanded TCR clones accumulating in MM are not enriched with viral specificities but accumulate in effector states in highly-infiltrated marrows. Additionally, we identify a role for T cell dynamics in patients treated with autologous stem cell transplantation and demonstrate T cell features predict progression from precursor to symptomatic MM. Together, these results suggest that anti-tumour immunity drives a distinctive form of cancer-associated T cell differentiation in MM.

T cells are polyfunctional immune cells and fundamental players in anti-tumour immunity[1]. In solid cancers, evidence suggests that early in carcinogenesis tumour growth can be curtailed by tumour-reactive T cells. However, persistent activation drives these cells away from functional memory states towards a hypo-responsive state of terminal differentiation termed exhaustion[1,2], characterised by the expression of immune checkpoint molecules like programmed cell death protein 1 (PD-1), contributing to cancer progression in solid cancers. This

[1]Research Department of Haematology, University College London Cancer Institute, London, UK. [2]University College London Hospitals NHS Foundation Trust, London, UK. [3]CRUK City of London Centre Single Cell Genomics Facility, UCL Cancer Institute, University College London, London, UK. [4]Genomics Translational Technology Platform, UCL Cancer Institute, University College London, London, UK. [5]Pre-Cancer Immunology Laboratory, UCL Cancer Institute, London, UK. [6]Cancer Research UK Lung Cancer Centre of Excellence, UCL Cancer Institute, London, UK. [7]MRC Toxicology Unit, Cambridge, UK. [8]Cambridge University Hospitals NHS Trust, Cambridge, UK. [9]Oxford University Hospitals, NHS Foundation Trust, Oxford, UK. [10]Bill Lyons Informatics Centre, University College London Cancer Institute, London, UK. [11]Division of Infection and Immunity, University College London, London, UK. [12]Department of Computer Sciences, University College London, London, UK. [13]Immune Regulation and Tumour Immunotherapy Group, Cancer Immunology Unit, Research Department of Haematology, University College London Cancer Institute, London, UK. ✉e-mail: b.chain@ucl.ac.uk; s.quezada@ucl.ac.uk; kwee.yong@ucl.ac.uk

complex interaction is believed to shape tumours from the early precursor stages to relapsed and refractory disease[1]. Understanding these insights has refined the treatment of solid tumours through the development of immunotherapies targeting exhausted T cells[3].

Multiple myeloma (MM) is a haematological malignancy of bone marrow (BM) plasma cells (PC) that is largely incurable[4], but is invariably preceded by precursor states of increased risk of progression: monoclonal gammopathy of undetermined significance (MGUS) and smouldering multiple myeloma (SMM)[5]. While not every MGUS or SMM patient will progress, virtually every MM patient has transitioned through these stages[6]. Thus, there is a pressing clinical need to understand the determinants of disease progression, so as to accurately identify individuals who may benefit from monitoring and intervention. Current risk factors rely largely on measures of tumour bulk[5], namely the levels of plasma cell infiltration in the bone marrow (BM PC %) and serum concentrations of paraprotein (malignant clone-derived immunoglobulin) and beta2-microglobulin (B2m). However, the role of the BM immune microenvironment, particularly T cells, in progression remains poorly understood. Available reports of altered immune cell function and phenotype suggest that, like solid tumours, the progression from precursor state to frank MM is accompanied by loss of immune control[7]. Understanding how MM drives alterations in T cell state and function is complicated by the influence of patient advanced age and marrow homoeostatic T cell differentiation[8,9]. This, together with high inter-individual immune heterogeneity, confounds attempts to identify associations between T cells and tumour biology or progression.

To solve this, we combined newly-generated single-cell RNA sequencing (scRNA-seq) and T cell receptor (TCR) sequencing (scTCR-seq) data with 12 published studies (see "Methods"), compiling an atlas of over a million single cells from 255 individuals to interrogate T cell dynamics while controlling for natural and tumour-associated sources of inter-individual variation. We show for the first time that the T cell landscape in MM displays features of antigen-driven terminal memory differentiation and highlight the features of tumour biology that drive this. We identify a role for T cell dynamics in the response to autologous stem cell transplantation (ASCT) and observe that alterations in T cells predict progression from precursor SMM to active MM. Our findings provide a framework for understanding tumour-immune interactions in disease evolution and therapy.

Results

Effective integration of scRNA-seq datasets allows a detailed classification of immune cell populations

To study immune differentiation through myeloma disease evolution we generated a large scRNA-seq map of BM ($n = 240$ samples, 77.6%) and peripheral blood (PB; $n = 68$, 22.4%) cells from untreated MGUS ($n = 23$, 9%), SMM ($n = 60$, 25%) and MM ($n = 67$, 25%) patients alongside non-cancer controls ($n = 105$, 41%; Fig. 1, Supplementary Fig. 1A, B and Supplementary Data 1). Together, we collated 1,079,979 cells from 255 individuals with RNA and clinical data, 109 with additional scTCR-seq data, and 1 with cellular indexing of transcriptomes and epitopes (CITE-seq) data. Patients were on average older than controls (controls median 55 range 21–87, MGUS median 62 range: 41–81, SMM median 62 range: 29–81, MM median 62 range: 38–77; Supplementary Fig. 1C). As expected, BM PC % and serum paraprotein levels were higher in MM compared to SMM (PC% $p = 0.001$, paraprotein $p < 0.05$, Wilcoxon test; Supplementary Fig. 1D).

Following quality control and correcting for batch effects (see "Method"; Supplementary Fig. 2A, B), cells were clustered to nine major cell types and phenotyped using RNA expression, protein expression via CITE-seq, and de novo label prediction tools (Fig. 2A, B and Supplementary Fig. 2C–E). T cells (defined as *CD3D*$^+$*CD3E*$^+$*CD3G*$^+$*CD8A*$^+$ or *CD4*$^+$ RNA and CD3$^+$ protein) comprised roughly half (50.2%) of the cells in the dataset, with another quarter occupied by similar proportions of myeloid cells (*FCN1*$^+$*FCER1G*$^+$CD14$^+$; 15%) and haematogenic progenitors (*CD34*$^+$*MPO*$^+$*TYMS*$^+$; 11%). The remainder of the dataset was largely comprised of equivalent numbers of NK cells (*KLRD1*$^+$*FCG3RA*$^+$CD56$^+$; 8%), B cells (*CD79A*$^+$CD19$^+$; 7.5%), and plasma cells (*MZB1*$^+$*SDC1*$^+$; 6.6%).

The use of different sorting strategies across studies precluded in-depth analysis of overall immune composition. However, plasma cells and progenitors were enriched in BM-derived samples relative to PB ($p < 0.001$ and $p < 0.001$, respectively, Wilcoxon test; Fig. 2B), suggesting a relative lack of haemodilution in BM aspirates. As expected, plasma cells were most highly enriched in the BM of patients relative to controls ($p < 0.001$, Wilcoxon test). However, the global distribution of cell types was otherwise similar in diseased and control marrows, suggesting progression to myeloma may be associated with more granular alterations to immune composition.

In-depth T cell phenotyping reveals diverse early and late differentiated CD4$^+$ and CD8$^+$ T cell populations in the marrow and blood

To further probe immune perturbations in disease evolution, T cells were isolated, re-integrated and re-clustered to 19 discrete phenotypes and transcriptional states based on expression of RNA and protein markers (Fig. 2C, D, Supplementary Fig. 3A–C and Supplementary Data 2 and 3)[2,10].

Naïve and central memory T cells (defined by *SELL* and *CCR7*, Tn and Tcm) composed the majority of CD4$^+$ cells (Tn 49% of CD4$^+$ T cells,

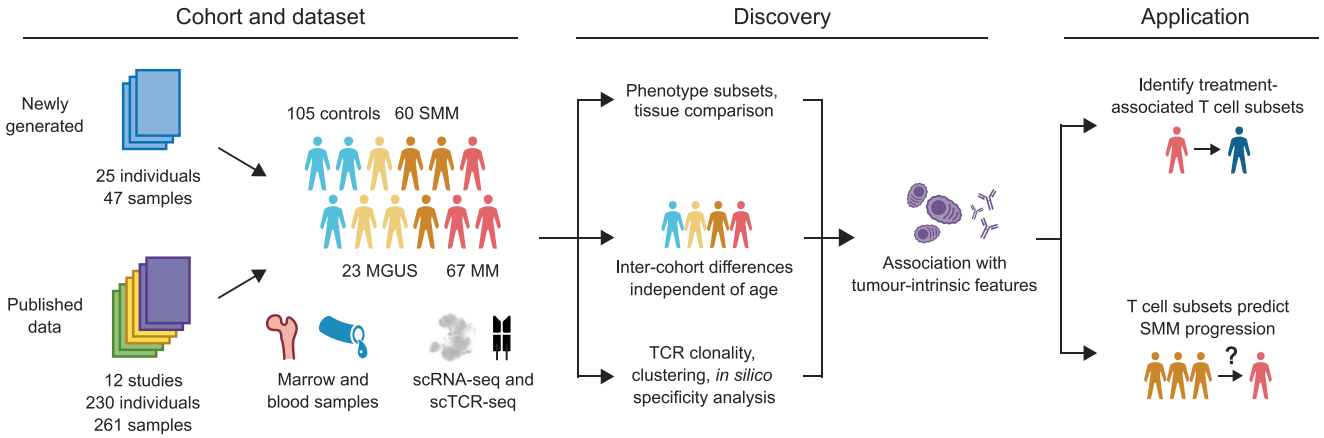

Fig. 1 | Overview of study. A dataset of 308 blood and bone marrow samples from a cohort of 255 patients and controls was derived from newly generated and published single-cell RNA sequencing (scRNA-seq) and single-cell T cell receptor (TCR) sequencing (scTCR-seq) datasets. Phenotyping, differential abundance, and TCR analysis were performed across tissues and clinical groups, compared with tumour-intrinsic features, and explored in the treatment and precursor disease settings.

https://doi.org/10.1038/s41467-026-68718-4

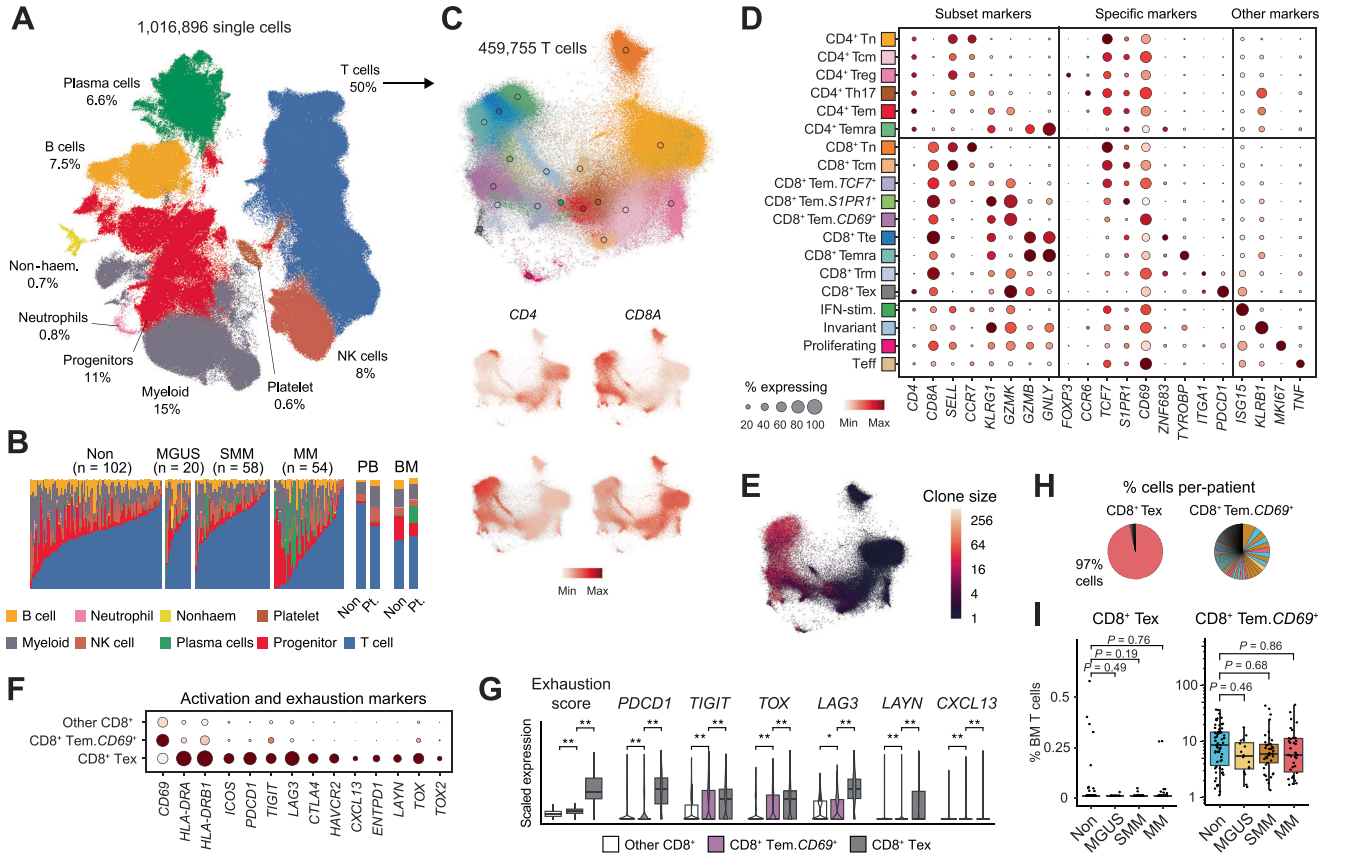

Fig. 2 | Integrated scRNAseq analysis of myeloma patients, precursor conditions, and controls reveals a diverse T cell landscape not defined by T cell exhaustion. A Visualisation of cell type clusters by uniform manifold approximation and projection (UMAP). **B** Bar chart comparing cell type composition in non-cancer controls (Non, $n = 102$), MGUS ($n = 20$), SMM ($n = 58$) and MM ($n = 54$) patients (left) and in the peripheral blood (PB) and bone marrow (BM) of all controls and patient groups (Pt.; right). **C** Visualisation of T cell clusters by minimum-distortion embedding (MDE). The expression of *CD4*, *CD8A*, and naïve and cytotoxic genes signatures is inset below. **D** Dot plot comparing the expression of marker genes in T cell clusters. **E** MDE plot showing the clonal expansion of T cell receptor (TCR) clones. **F** Dot plot comparing the expression of T cell exhausted-associated genes in indicated CD8$^+$ T cell subsets. Legend as in (**D**). **G** Box and violin plots comparing T cell exhaustion signature and exhaustion-associated gene expression in CD8$^+$ Tex, CD8$^+$ Tem.*CD69*$^+$ and other CD8$^+$ clusters. For the three groups a random sample of $n = 1000$ cells from Zheng et al. are shown. $^{**}p = 0.001$, $^{*}p < 0.05$, NS = $p > 0.05$. **H** Pie charts comparing the proportion of CD8$^+$ Tex and CD8$^+$ Tem.*CD69*$^+$ clusters from each individual. Each slice represents a unique individual. **I** Box plots comparing the abundance of CD8$^+$ Tex and CD8$^+$ Tem.*CD69*$^+$ in the BM in non-cancer controls ($n = 71$), MGUS ($n = 16$), SMM ($n = 48$) and MM ($n = 41$) patients. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. *p* values derived by one-way ANOVA followed by Tukey's test.

Tcm: 21%), while the remainder were composed of diverse regulatory (*FOXP3*$^+$ Treg, 8.4%), T helper 17 (*CCR6*$^+$ Th17 7.2%), effector memory (*KLRG1*$^+$*GZMK*$^+$ Tem 6.1%) and Tem re-expressing CD45RA (*GZMB*$^+$*GNLY*$^+$CD45RA$^+$ Temra, 8.7%) states. CD8$^+$ cells were classified by differentiation stage including early differentiated Tn (20% of CD8$^+$ T cells) and Tcm (4.6%), intermediate Tem (37%), and *GZMB* and *GNLY*-expressing terminally differentiated terminal effector memory (Tte, 18%) and Temra (15%) states. CD8$^+$ Tem were further sub-divided by the expression of early differentiation markers like TCF1 (*TCF7*; Tem.*TCF7*$^+$) or *S1PR1* (Tem.*S1PR1*$^+$) alongside activation markers like *CD69* (Tem.*CD69*$^+$). The two terminally differentiated CD8$^+$ subsets were distinguished based on the expression of specific markers like Hobit (*ZNF683*; T terminal effector, Tte) or NK cells genes such as NKG2C (*KLRC*) and CD45RA protein (Supplementary Fig. 3C; Temra) and were enriched in senescence-associated genes (*SenMayo*; CD8$^+$ Tte: $p = 0.03$, NES = 1.6; CD8$^+$ Temra: $p = 0.03$, NES = 1.5, GSEA; Supplementary Data 4). Additionally, rare clusters of tissue resident (Trm, 5.1%) and exhausted (Tex, 0.7%) CD8$^+$ T cells were identified by the expression of *ITGA1* and PD-1 (*PDCD1*), respectively. Additional T cell subsets were identified by the expression of specific markers, including IFN-stimulated cells by *ISG15*, invariant subsets by CD161 (*KLRB1*), proliferating T cells by *MKI67*, and effector T cells (Teff) by effector molecules like TNF-α (*TNF*) and *IFNG* alongside the enrichment of a Teff gene signature (Effector versus memory[11]; $p = 0.005$, NES = 3.5, GSEA). We also explored heterogeneity among CD4$^+$ Treg and invariant T cells by performing additional sub-clusterings of these subsets (see "Methods"; Supplementary Fig. 4A).

This functional annotation was consistent with patterns of TCR expansion (Fig. 2E and Supplementary Fig. 3D), with clonally expanded clusters expressing late differentiation markers. The majority (79.8%) of invariant MAIT TCRs were found in cells of invariant cluster, of which 97% were MAITs (Supplementary Fig. 4A). Our phenotypes showed high concordance with published and predicted cluster labels (Supplementary Fig. 3E), arguing for a faithful representation of T cell phenotypes in our integrated dataset.

The marrow of myeloma patients is not enriched in exhausted T cells

Exhausted CD8$^+$ Tex were identified via high expression of *PDCD1* and *TIGIT*, alongside other RNA markers of exhaustion like *CXCL13* and *LAYN*[10] (Fig. 2F and Supplementary Fig. 3B). Importantly, CD8$^+$ Tex were distinguished from *PDCD1*-expressing CD8$^+$ Tem.*CD69*$^+$. While CD8$^+$ Tem.*CD69*$^+$ cells did express higher levels of exhaustion markers like *PDCD1* and *TOX* than non-exhausted bone marrow T cells

($p < 0.001$ and $p < 0.001$, one-way ANOVA; *PDCD1:* adjusted $p < 0.001$, marker gene testing; Fig. 2G and Supplementary Fig. 3F), it was at lower levels than CD8$^+$ Tex ($p < 0.001$ and $p < 0.001$, Wilcoxon test) and further lacked other markers of exhaustion like *LAYN* and *CXCL13* (Fig. 2G and Supplementary Fig. 3B). A dedicated sub-clustering of CD8$^+$ Tem.*CD69*$^+$ did not identify a subset of these cells with features of CD8$^+$ Tex (Supplementary Fig. 3G). Therefore, despite expressing some exhaustion-associated markers at low levels, CD8$^+$ Tem.*CD69*$^+$ are not phenotypically exhausted T cells.

We extended this phenotypic description by characterising the function of CD8$^+$ Tem.*CD69*$^+$. CD69$^+$ and CD69$^-$ fractions of memory (CCR7$^-$) CD8$^+$ T cells were isolated from 3 MM patient marrows and activated cells in vitro (see "Methods"). After stimulation, CD69$^+$ cells possessed no significant differences in 4-1BB and effector cytokine expression (Supplementary Fig. 4C), suggesting these cells are not functionally impaired relative to other CD8$^+$ Tem.

Interestingly, the CD8$^+$ Tex cluster was almost entirely composed of cells from a single myeloma patient (1181 of 1222 cells, 97%; Fig. 2H) who contributed the majority of exhaustion marker-expressing cells (Supplementary Fig. 3H), suggesting CD8$^+$ Tex were a donor-specific phenomenon. Furthermore, when excluding this patient, CD8$^+$ Tex were not significantly enriched in the BM of MGUS, SMM or MM patients relative to controls ($p = 0.76$, one-way ANOVA; Fig. 2I). In contrast, CD8$^+$ Tem.*CD69*$^+$ cells were seen in a much higher number of patients (Fig. 2H), although, like CD8$^+$ Tex, these cells were not enriched in patients relative to controls ($p = 0.86$, one-way ANOVA; Fig. 2I).

We also explored the contribution of bone marrow tissue on T cell composition. Similar T cell phenotypes were observed in the BM and PB (Supplementary Fig. 4D), but the proportion of T cell clusters differed (Supplementary Fig. 4E). Notably, CD8$^+$ Tem.*CD69*$^+$ were enriched in the BM of both patients and controls ($p < 0.001$ and $p < 0.001$, Supplementary Fig. 4F), suggesting intrinsic marrow biology regulates the abundance of this subset.

These data lead us to suggest that exhausted T cells are rarely seen in myeloma, with the abundant and functional CD8$^+$ Tem.*CD69*$^+$ being phenotypically distinct from exhausted cells. Neither subset was associated with disease progression, although CD8$^+$ Tem.*CD69*$^+$ were enriched in the marrow in health and disease.

T cells in diseased marrow of both pre-malignant SMM and overt MM display skewed differentiation resembling exaggerated age-related changes

We next asked how the relative abundance of T cell subsets in the BM was altered across disease stages. BM T cell composition was strikingly similar in patients and controls (Fig. 2A and Supplementary Fig. 5A, B). Therefore, we compared the T cell composition of controls with each myeloma disease stage in turn, statistically controlling for age. The most prominent difference in BM T cell composition between health and disease was the loss of naïve, CD4$^+$ Th17 and Invariant cells and an enrichment of CD4$^+$ and CD8$^+$ Temra, CD8$^+$ Tte, and IFN-stimulated cells (adjusted $p < 0.1$ for all, linear regression; Fig. 3B and Supplementary Fig. 5C). When removing the one patient who contributed the majority of CD8$^+$ Tex cells (97% cells; Fig. 2H), this cluster was not enriched in MM relative to controls independent of age ($p = 0.54$, linear regression). As non-cancer controls included hip replacement and deceased donors, we repeated our analysis with only healthy donors and obtained the same results (Supplementary Fig. 5D).

Low-risk MGUS possessed a T cell composition with the fewest differences from non-diseased control marrows (Supplementary Fig. 5C). Conversely, T cell composition was similar between the higher-risk but pre-cancerous SMM and symptomatic MM (Supplementary Fig. 5E). While the normalised abundance of CD8$^+$ Tex was lower in SMM than MM independent of age ($p < 0.001$, linear regression), in terms of unnormalized counts this only represented 8 MM patients with a median of 1 CD8$^+$ Tex cells each, suggesting a lack of sufficient cell numbers to properly perform this comparison. Next, we collated available risk data for patients in our dataset, and compared T cell composition between international staging system (ISS) and SMM Mayo risk groups[5,12] (Supplementary Fig. 5F). While this analysis found no differences between risk groups, with the exception for an increased abundance of CD4$^+$ Tem in ISS groups II and III, these comparisons were limited by risk annotations and may not accurately reflect differences between these groups. However, taken together these results suggest that smouldering and overt myeloma are associated with similar T cell alterations independent of patient age.

To quantify global changes to T cell composition, we ran principal component analysis (PCA) on patients' T cell cluster abundance data. The first principal component explaining the highest fraction of variance in T cell composition (PC1, 21.9%; Supplementary Fig. 5H) described a compositional shift from naïve and early subsets towards terminal memory cytotoxic clusters (Fig. 3C, D). Therefore, we termed PC1 "T cell skewing". T cell skewing was greatest (meaning an enrichment of terminal memory clusters) in SMM and MM relative to controls independent of age ($p < 0.004$ and $p < 0.001$, respectively, linear regression; Fig. 3E), demonstrating this metric captured the major alterations to T cells in myeloma. Conversely, T cell skewing was associated with age independent of patient group ($p = 0.013$, linear regression, $R = 0.29$, Pearson correlation; Fig. 3E, F and Supplementary Fig. 5I). Therefore, T cell skewing is a feature of both age and disease, meaning myeloma-associated T cell differentiation directly resembles T cell alterations seen during ageing, or, patients possess an exaggerated form of T cell ageing. Using linear regression, we calculated the degree of exaggerated T cell ageing and excess T cell years (see "Methods") and found both trended to rise with advanced disease (MGUS versus MM $p = 0.13$ and $p < 0.07$ respectively, one-way ANOVA; Fig. 3G and Supplementary Fig. 5J) suggesting this effect scales linearly with disease severity.

PC1 values and the abundance of terminal memory subsets in the BM showed a strong correlation with PB (Fig. 3H), indicating a similarity to systemic T cell alterations seen with ageing[8].

Taken together, these data suggest SMM and MM are associated with a pattern of T cell differentiation skewing closely resembling alterations observed during T cell ageing.

Features of antigen-specific responses underpin myeloma-associated T cell differentiation

Next, we analysed features of the TCR repertoire. Repertoire clonality was associated with T cell skewing in patients independent of age ($p < 0.001$, linear regression, $R = 0.68$; Fig. 4A). We observed similar results when restricting analysis to CD8-expressing memory clones ($p < 0.01$, linear regression, $R = 0.46$; Fig. 4A). Both the clonality and the abundance of expanded CD8$^+$ memory clones trended for enrichment in MM relative to controls (Fig. 4B). T cell skewing and CD8$^+$ memory diversity did not correlate in controls ($p = 0.48$, linear regression; Supplementary Fig. 6A), suggesting clonal expansion may be a unique feature of T cell differentiation in myeloma.

The accumulation of TCRs possessing similar CDR3 sequences indicates responses against common antigens. We grouped all expanded TCRs in the dataset (11,545 clones) into 279 clusters (composed of 1014 clones, 8.8% of input; see "Methods"; Fig. 4C and Supplementary Fig. 6B, C). To prioritise grouping TCRs recognising antigens in the same human leucocyte antigen (HLA) background, TCR clustering was restricted within individuals. An increasingly large fraction of the TCR repertoire was occupied by clustered clones in MM (median 8.7% range 3–35) relative to SMM (median 3.1% range 0–11.3; $p = 0.0009$) and this fraction correlated with T cell skewing ($R = 0.4$, $p = 0.02$, Pearson correlation; Fig. 4D). T cell skewing was specifically associated with the clustering of *GZMB*-expressing CD8$^+$ memory cells ($R = 0.47$, $p = < 0.01$; Fig. 4E), suggesting conserved antigen-specific

https://doi.org/10.1038/s41467-026-68718-4

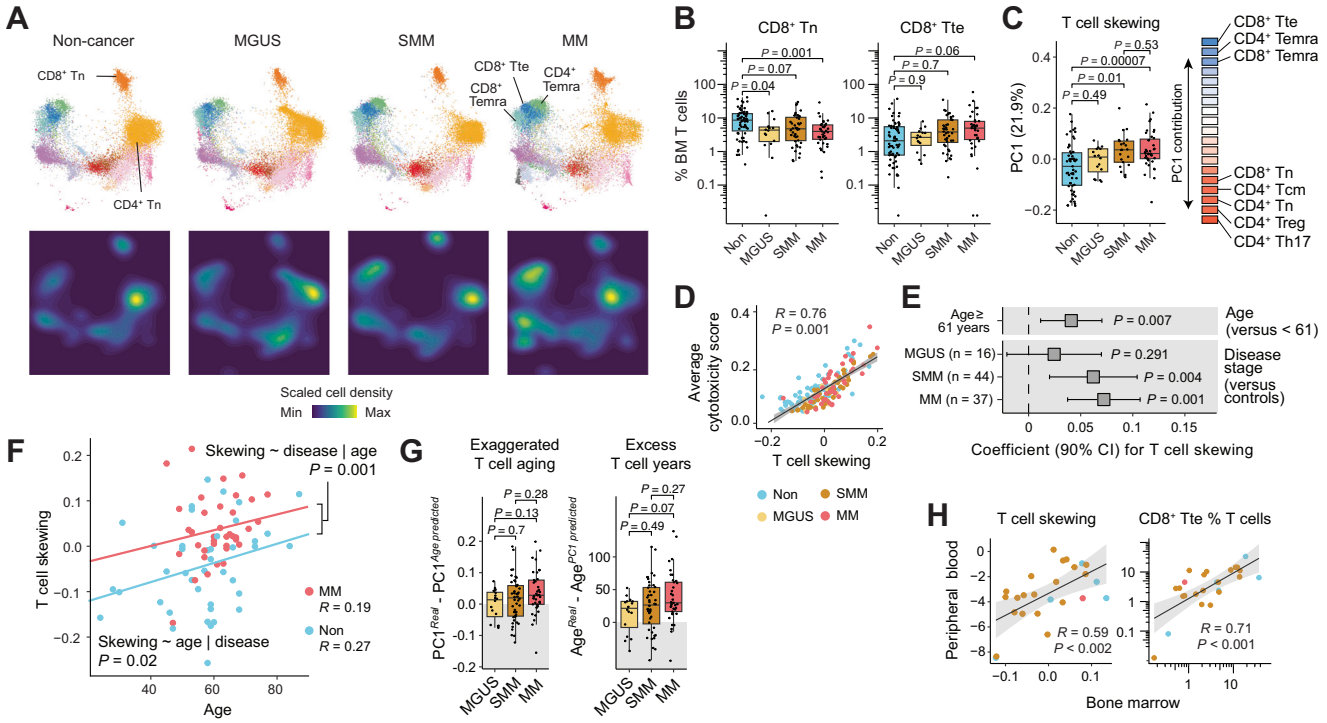

Fig. 3 | Step-wise alterations to bone marrow T cell composition occur through myeloma disease evolution. A Visualisation of bone marrow (BM) T cells clusters (top) and per-cell density (smoothed 2D histogram of cell abundance, bottom) by minimum-distortion embedding (MDE) in non-cancer controls, MGUS, SMM and MM patients. For each group a sample of 20,000 cells is shown. **B** Box plots comparing the normalised abundance of BM CD8⁺ Tn (left) and CD8⁺ Tte (right) in non-cancer controls (Non; $n = 71$), MGUS ($n = 16$), SMM ($n = 48$) and MM ($n = 41$) patients. **C** Left, Box plot comparing the degree of the BM T cell skewing (the first principal component (PC1), 21.9% variance) in Non ($n = 67$), MGUS ($n = 16$), SMM ($n = 44$) and MM ($n = 39$) patients. Right, representation of the clusters with the highest (blue) and lowest (red) contribution (loading) to PC1. **D** Dot plot showing the significant positive correlation between T cell skewing and cytotoxic gene signature expression in BM T cells. **E** Forest plot showing the significant independent associations between T cell skewing and both disease stage and age. *p* values derived from linear regression (see "Methods"). **F** Dot plot showing the absolute difference between T cell skewing in controls and MM patients across the range of ages. Inset *p* values indicate the significance of the association between T cell skewing and age independent of disease (*age* | *disease*) and disease independent of age (*disease* | *age*). *p* values derived from linear regression. *R* values calculated by Pearson correlation. **G** Box plot showing exaggerated T cell ageing and excess T cell years (see "Methods") in MGUS ($n = 16$), SMM ($n = 44$) and MM ($n = 37$) patients. Residuals values of precisely zero or less than zero (indicating T cell skewing or T cell years expected or less for a patient's age) is indicated with shaded region. **H** Dot plots showing the significant positive correlation between T cell skewing (left) and the abundance of CD8⁺ Tte (right) in the peripheral blood (PB) and BM in Non ($n = 4$), SMM ($n = 19$) and MM ($n = 1$) patients. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. For **B**, **C**, **G**, *p* values calculated by one-way ANOVA followed by Tukey's test. For **D**, **H**, *R* and *p* values were derived by Pearson correlation.

responses drive T cell differentiation in myeloma, specifically among *GZMB*-expressing subsets.

We next explored the antigen specificity of these T cells. TCR specificity databases are mostly composed of viral antigens, allowing us to ask if viral antigen specificities were involved in T cell differentiation and clonality in myeloma (see "Methods"). In 19 SMM and MM patients BM samples we found a median of 7 (range 2–88) putative HLA-matched viral-reactive TCRs (Supplementary Fig. 6D, E). Putative viral-reactive clones occupied a range of T cell clusters, including naïve and effector memory phenotypes. Comparing gene expression between clones with and without putative viral specificity annotations, we found clones predicted to not be viral-specific expressed genes characteristic of terminal memory such as *GZMB*, perforin (*PRF1*) and Hobit (*ZNF683*; Fig. 4F and Supplementary Data 4). We summarised the expression of these genes into a non-viral specificity gene signature (Supplementary Data 4). Non-viral signature expression mapped to *GZMB*⁺ terminal memory clusters (Supplementary Fig. 6F), correlated with T cell skewing, clonality and repertoire clustering (Fig. 4G), and was enriched in MM relative to precursor conditions (Fig. 4H). Taken together, these data show MM-associated T cell differentiation occurs alongside alterations to the TCR repertoire resembling antigen-directed T cell immunity which are distinct from viral-specific TCR clones.

Tumour-intrinsic features drive two clonally-related patterns of T cell differentiation

We next asked if features of tumour biology may drive T cell skewing. We examined serum paraprotein and beta2-microglobulin (B2m) concentrations, plasma cell marrow infiltration, and tumour cell transcriptional state (see "Methods", Supplementary Fig. 7). We included all patients irrespective of disease stage.

We did not find an association between T cell skewing and tumour transcriptional state, BM PC %, or B2m levels (Supplementary Fig. 8A, B). However, we observed a positive correlation between high serum paraprotein and enhanced T cell skewing ($R = 0.43$, $p < 0.03$, Pearson correlation; Fig. 5A), alongside the abundance of CD8⁺ Temra ($R = 0.5$, $p = 0.01$) and high TCR clonality ($R = 0.53$, $p < 0.01$; Fig. 5A and Supplementary Fig. 8C). This association was independent of age ($p < 0.05$, linear regression) and exaggerated T cell ageing similarly correlated with paraprotein (Supplementary Fig. 8C). We recapitulated this finding in two additional cohorts. Firstly, 10 SMM and 30 untreated MM patients were profiled by cytometry by time-of-flight (CyTOF; see "Methods"; Supplementary Data 5 and Supplementary Fig. 9A) and the abundance of CD8⁺ Temra (CD28⁻CD45RA⁺) correlated with paraprotein ($R = 0.38$, $p = 0.053$, Pearson; Fig. 5B). Secondly, a large group of 175 precursor patients ($n = 33$ MGUS, $n = 142$ SMM) were profiled at recruitment by flow cytometry (Supplementary Fig. 9B). Using a

https://doi.org/10.1038/s41467-026-68718-4

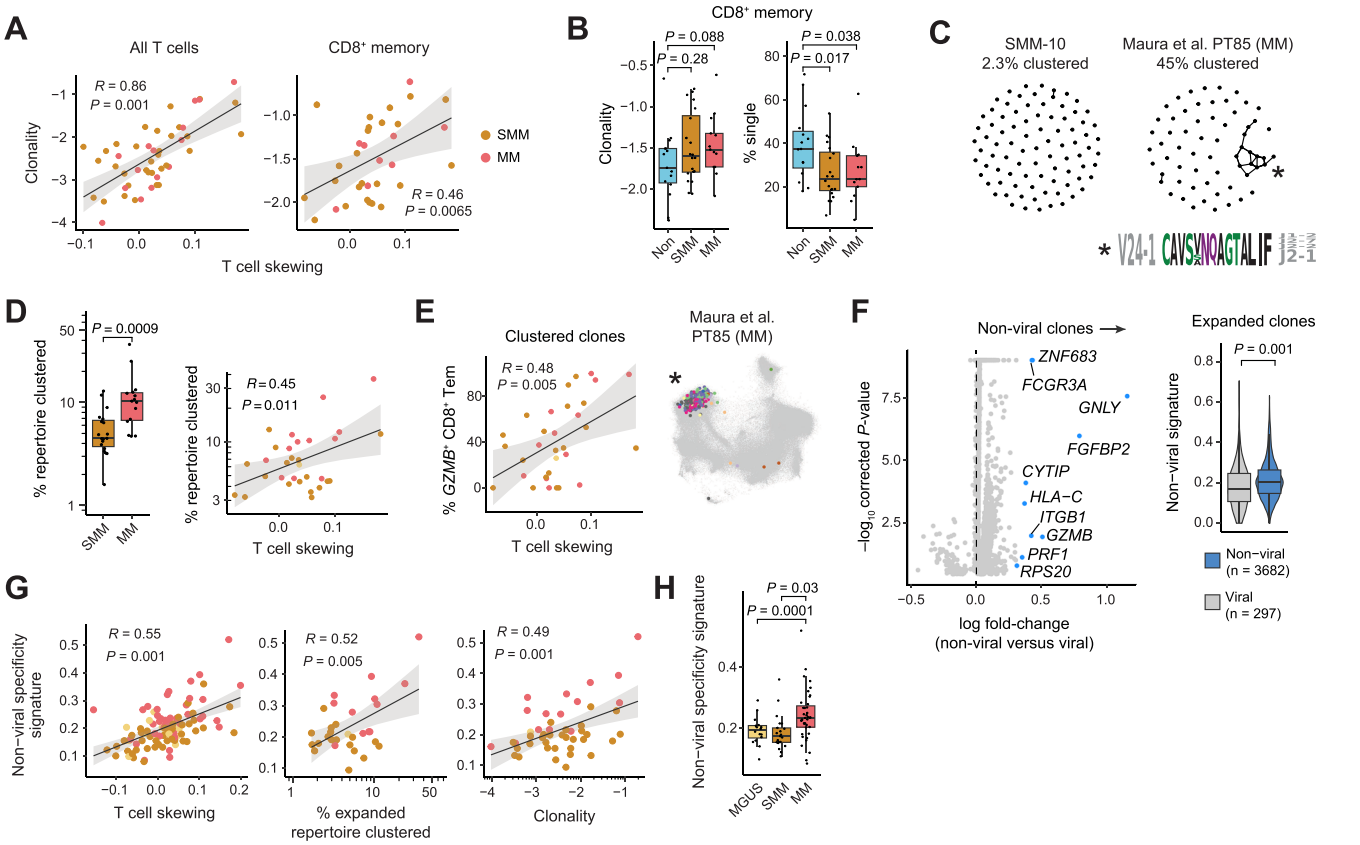

Fig. 4 | Features of antigen-experienced T cell receptor repertoires underpin myeloma-associated T cell differentiation. A Dot plots showing the significant positive correlation between T cell skewing and TCR clonality (log_{10} 1/Simpson's diversity) of all T cells clones (left, SMM $n = 32$, MM $n = 15$) and CD8+ memory clones (right, SMM $n = 22$, MM $n = 9$) in the bone marrow (BM). **B** Box plots comparing the clonality (left) and abundance of non-expanded clones (right) of $CD8^+$ memory clones in the BM of non-cancer controls (Non, $n = 15$), SMM ($n = 19$) and MM ($n = 12$) patients. **C** Network plots showing CDR3 clusters from two representative patients. Each node represents a TCR clone and each connected edge a co-clustering. The beta chain CDR3 sequence logo for the cluster indicated by an asterisk is shown below. **D** Left, Box plot comparing the percentage of clustered expanded TCRs in SMM ($n = 19$) and MM ($n = 12$). Right, dot plot showing the significant positive correlation between T cell skewing and the percentage of clustered expanded TCRs in MGUS ($n = 1$), SMM ($n = 18$) and MM ($n = 12$) patients. **E** Left, Dot plot showing the significant positive correlation of T cell skewing and the percentage of *GZMB*$^+$ $CD8^+$ Tem ($CD8^+$ Tte and $CD8^+$ Temra) cells among clustered TCR clones. Right, minimum-distortion embedding (MDE) showing cells from CDR3 cluster annotated with asterisk in (**C**) with colour representing different clones. **F** Left, differential expression results between T cells possessing TCR clones annotated as viral-reactive (see "Methods") versus all other clones. Labelled genes constitute the non-viral signature. Corrected p values calculated with unpaired T-test and Holm–Bonferroni correction. Right, box and violin plot comparing the average expression of the non-viral signature in expanded viral and non-viral clones. **G** Dot plots showing the significant positive correlation between non-viral specificity signature expression and T cell skewing (left), the percentage of clustered TCRs among expanded TCRs (centre), and TCR clonality (right) in the BM of MGUS, SMM and MM patients. **H** Box plot comparing non-viral specificity signature expression in the BM of MGUS ($n = 16$), SMM ($n = 45$) and MM ($n = 40$) patients. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. For **A**, **D** (right), **E** (left), **G**, R and p values were calculated by Pearson correlation. For **B**, **D** (left), **F** (right), p values derived by two-sided Wilcoxon test. For **H**, p values derived by one-way ANOVA followed by Tukey's test.

minimal set of markers, we approximated T cell skewing via the abundance of $IL7R^-$ (late differentiated) $CD8^+$ T cells (Supplementary Fig. 8D). In this cohort, the abundance of $IL7R^-CD8^+$ T cells significantly correlated with paraprotein independent of age ($R = 0.24$, $p < 0.001$ Pearson correlation; p independent of age $p < 0.001$ linear regression, Fig. 5C). Together, this connects T cell skewing observed in patients with serum paraprotein levels, indicating a relationship with tumour bulk.

We speculated that tumour-intrinsic features were associated with individual T cell clusters. Analysis of T cell cluster abundance with BM PC % and transcriptional pathway activity in tumour cells revealed effector T cells (Teff) were enriched in highly-infiltrated marrows populated by tumour cells expressing stress-associated genes (BM PC %: $R = 0.55$, tumour stress: $R = 0.6$; Fig. 5D and Supplementary Fig. 8D). Several other clusters were significantly correlated with tumour stress, including positive and negative associations with $CD8^+$ Trm and $CD4^+$ Tem, respectively (Supplementary Fig. 8E). The expression of a smaller set of T cell effector genes (*CD69*, *TNF*, *IFNG*) possessed the same correlations (BM PC%: $R = 0.43$, tumour stress: $R = 0.79$, Pearson correlation; Supplementary Fig. 8F), suggesting a T cell effector programme specifically was associated with BM PC % and tumour stress.

The co-occurrence of tumour-associated memory and effector T cells suggested a differentiation process between these two states. We therefore asked if effector T cells were clonally related to terminal memory subsets. Clonally expanded $CD8^+$ effector T cells were significantly enriched among $CD8^+$ Temra ($p < 0.001$, chi-squared test; Fig. 5E). Additionally, clonally expanded Teff cells enriched in marrows containing stressed tumours expressed markers characteristic of $CD8^+$ Temra (Fig. 5F). Therefore, terminally differentiated memory clones accumulating in MM are continuous with effector-like T cells in highly-infiltrated marrows, suggesting these states are linked by a T cell differentiation pathway.

Finally, we asked if tumour genomic features were associated with T cell skewing. $t(4;14)$ (positive $n = 6$, negative $n = 17$) was not

https://doi.org/10.1038/s41467-026-68718-4

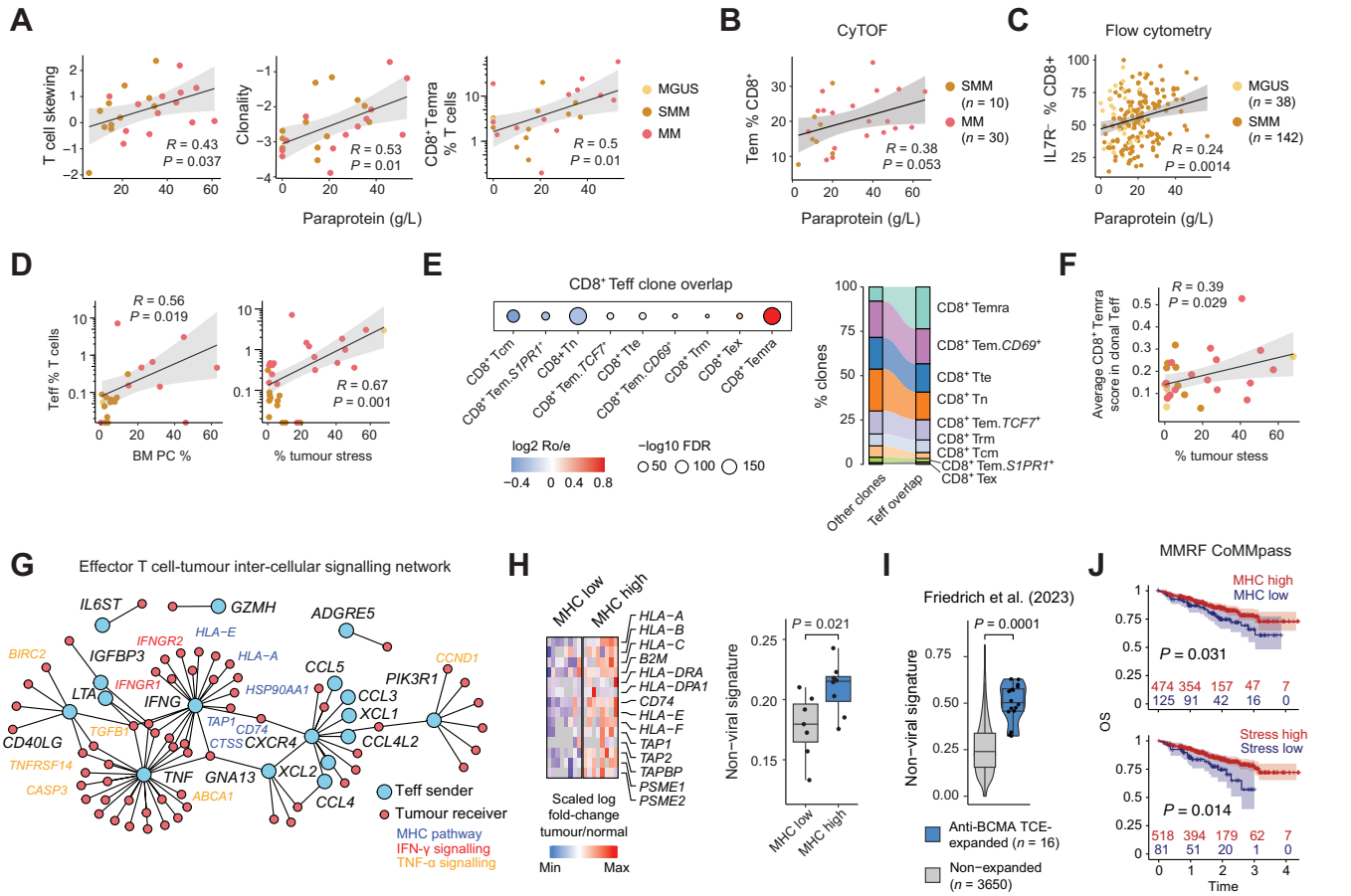

Fig. 5 | Tumour-intrinsic features are associated with T cell differentiation. **A** Dot plot showing the significant positive correlation between serum paraprotein (g/L) and T cell skewing, TCR clonality (log_{10} 1/Simpson's diversity), and the abundance of CD8$^+$ Temra as % T cells in MGUS ($n = 1$), SMM ($n = 11$) and MM ($n = 13$) patients. Dot plot showing the significant positive correlation between serum paraprotein and the abundance of effector memory T cells (Tem) as % of CD8$^+$ T cells by Cytometry by time-of-flight (CyTOF; **B** SMM $n = 10$, MM $n = 30$) and IL7R$^-$ cells among CD8$^+$ T cells by flow cytometry (**C** MGUS $n = 32$, SMM $n = 142$). **D** Dot plot showing the significant positive correlation between the abundance of effector T cells (Teff) and bone marrow (BM) plasma cell (PC) % (left; MGUS $n = 1$, SMM $n = 12$, MM $n = 9$) and the abundance of tumour cells expressing stress-associated genes (right; MGUS $n = 2$, SMM $n = 12$, MM $n = 18$). **E** Left, dot plot representing TCR overlap between CD8$^+$ Teff clones and other CD8$^+$ T cell clusters. Ratio of observed to expected (Ro/e) and p values derived by chi-squared test. Right, bar chart showing the CD8$^+$ T cell cluster composition of clones shared with Teff. **F** Dot plot showing the significant positive correlation between CD8$^+$ Temra signature expression in Teff cells and the % of tumour cells expressing the stress-associated pathway. **G** Network plot showing the ligand-receptor network of Teff and tumour cells. Shown are signalling molecules significantly enriched in effector T cells whose cognate signalling partners were up-regulated by tumour cells. Tumour genes associated with MHC, IFN-γ or TNF-α pathways are labelled. **H** Left, heatmap showing the expression of MHC and antigen-presentation genes in tumour cells relative to non-cancer plasma cells in 16 patients. Patients were grouped into "MHC high" ($n = 8$) and "MHC low" ($n = 7$) based on the co-enrichment of multiple MHC genes. Right, box plot comparing the expression of the non-viral specificity signature in patients with high or low expression of MHC genes. **I** Box and violin plot comparing the expression of the non-viral specificity signature in TCR clones which did and did not significantly expand following T cell engager (TCE) therapy. Scoring was performed in pre-treatment samples from $n = 3$ patients. **J** Kaplan–Meier curve showing the impact of high (red) and low (blue) expression of the MHC and stress pathways on overall survival in 598 newly diagnosed untreated multiple myeloma patients enrolled in the ComMMpass trial. p value calculated with log-rank test. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. For **A**–**D**, **F**, R and p values were calculated by Pearson correlation. For **H**, **I**, p values derived by two-sided Wilcoxon test.

significantly associated with T cell skewing ($p = 0.45$, Wilcoxon test) but a 1q gain (positive $n = 10$, negative $n = 19$) trended to enrich for enhanced T cell skewing (Supplementary Fig. 7G; $p = 0.057$, Wilcoxon test), potentially connecting this adverse prognostic factor to T cell differentiation.

Myeloma-associated T cell differentiation possesses features of anti-tumour immunity

To identify molecular drivers of tumour-immune crosstalk, we analysed the expression of intercellular signalling molecules in tumour-associated T cell subsets (see "Methods"). This analysis identified an intercellular signalling network dominated by T cell effector molecules IFN-γ and TNF-α and antigen-presentation pathways (Fig. 5G). Patients with high T cell effector cytokine expression possessed the highest corresponding pathway activity in tumour cells (IFN-γ: $R = 0.3$, $p < 0.05$; TNF-α: $R = 0.76$, $p < 0.001$, Pearson correlation; Supplementary Fig. 8G), mechanistically linking these processes. Notably, MHC and antigen processing and presentation genes were significantly enriched (GSEA, adjusted $p < 0.1$) in tumour cells for 6 of 16 (37.5%) patients tested (Fig. 5H). Interestingly, non-viral specificity signature expression was highest in patients whose tumour cells significantly upregulated the MHC pathways ($p = 0.021$, Wilcoxon test; Fig. 5H), suggesting that MM-associated T cell differentiation contains a component of tumour reactivity that was connected to tumour MHC class I expression and potentially antigen presentation.

https://doi.org/10.1038/s41467-026-68718-4

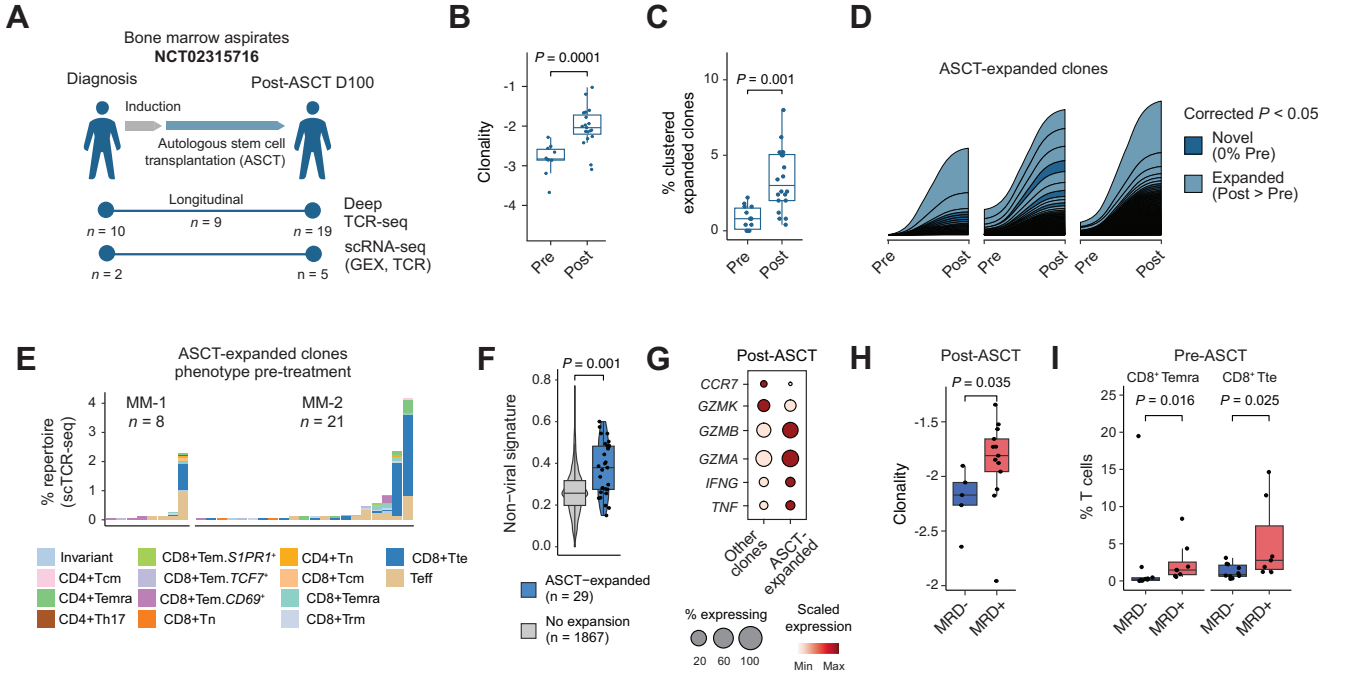

Fig. 6 | T cell skewing is associated with autologous stem cell transplantation. **A** Schematic depicting sampling and assessment of an independent cohort of multiple myeloma patients treated with autologous stem cell transplant (ASCT) as frontline therapy. Samples were taken at diagnosis and at day 100 (D100) post-ASCT. **B** Box plots comparing beta chain TCR clonality ($\log_{10}$ 1/Simpson's diversity) in pre- ($n = 10$) and post-treatment ($n = 19$) samples of patients receiving ASCT. **C** Box plots comparing the abundance of clustered expanded alpha chains in pre- ($n = 10$) and post-treatment ($n = 14$) samples of patients treated with ASCT. **D** Alluvial plots showing the abundance of TCR clones which significantly increased in abundance post-treatment in patients receiving ASCT. Each bar represents a single TCR beta chain and is coloured by whether it was absent pre-treatment (Novel, not present pre-treatment) or was more frequent post-treatment (Expanded, present in low abundance pre-treatment). Corrected p values calculated with Poisson test and Holm–Bonferroni correction. **E** Bar charts showing the pre-treatment abundance and phenotype of TCR clones which expanded significantly post-ASCT. **F** Box and violin plot comparing the pre-treatment expression of the non-viral signature in TCR clones which did and did not expand significantly post-ASCT. **G** Dot plot showing the post-treatment expression of selected marker genes in TCR clones which significantly expanded post-ASCT relative to other clones. **H** Box plot comparing post-ASCT alpha chain TCR clonality in patients who achieved minimal residual disease (MRD) positivity (MRD+; $n = 13$) and negativity (MRD−; $n = 5$) at D100 post-ASCT. **I** Box plot comparing CD8$^+$ T cell CyTOF cluster (see "Methods") abundance in pre-treatment samples from patients who achieved MRD+ ($n = 7$) and MRD− ($n = 10$) at D100 post-ASCT. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. p values derived by two-sided Wilcoxon test.

To explore tumour-reactivity, we utilised a recently-reported gene signature of tumour-reactive BM T cells in MM[13] and found significant positive correlation with our non-viral gene signature ($p < 0.001$, $R = 0.55$ Pearson correlation; Supplementary Fig. 8H). We also reasoned that if MM-associated T cell differentiation was linked to tumour reactivity, this might be revealed in the response to immunotherapy. We acquired additional scRNA-seq data from 3 patients who responded to T cell engager (TCE) therapy[14] and identified TCRs which significantly expanded post-treatment (see "Methods"; Supplementary Fig. 8I). Interestingly, T cells bearing these TCRs highly expressed the non-viral specificity signature pre-treatment (Fig. 5I). Together, this suggests this phenotype contains tumour-reactive TCR clones with the capability to respond to immunotherapy.

Finally, we asked if T cell-associated pathway activity in tumour cells was related to outcome. In an independent cohort of 599 MM patients, we found patients whose tumours highly expressed either MHC or stress pathway genes had superior outcome (MHC $p = 0.031$ and stress $p = 0.014$, log-rank test; Fig. 5J), suggesting that T cell differentiation dynamics associated with these tumour-intrinsic processes may influence clinical outcomes.

Autologous stem cell transplantation enhances T cell skewing through the expansion of pre-existing clones contributing to poor outcome

We next asked how the T cell dynamics we observed in untreated patients are altered by anti-tumour therapy. Autologous stem cell transplantation (ASCT) is standard of care in fitter patients with newly diagnosed MM. Work in patients and mouse models has suggested that the benefit of ASCT relies on the augmentation of anti-tumour T cell immunity[15,16]. To explore this hypothesis, we analysed bone marrow samples from newly diagnosed transplant-eligible patients who received ASCT with deep TCR-seq, scRNA-seq and scTCR-seq (Fig. 6A).

Repertoire clonality rose significantly following ASCT ($p < 0.001$, paired Wilcoxon; Fig. 6B). Similarly, the frequency of CDR3 clusters among expanded TCRs rose post-ASCT (Fig. 6C). We modelled the abundance of TCRs pre- and post-treatment with a Poisson framework (see "Methods") to identify TCRs which significantly increased in abundance post-treatment (Fig. 6D). Among these clones, 29% (range 10.7–48) were present at low abundance pre-treatment. Using pre-treatment samples found in the larger scRNA-seq dataset (Fig. 2), we identified the baseline phenotype of clones which expanded post-ASCT and found they predominantly occupied terminally-differentiated and effector CD8$^+$ T cell clusters (Fig. 6E) and highly expressed the non-viral signature ($p < 0.001$, unpaired Wilcoxon; Fig. 6F). In contrast, clones we annotated as viral-reactive (see "Methods") did not significantly expand post-ASCT (Corrected $p > 0.05$, Poisson model). We independently performed scRNA-seq on four samples from patients taken post-ASCT and found that ASCT-expanded clones post-treatment expressed terminal differentiation and Teff genes such as *GZMB* and *TNF* (Fig. 6G) and expressed significantly higher levels of the non-viral signature ($p < 0.001$, mixed-effect model with patient as random effect). Together, this shows

ASCT is associated with alterations resembling antigen-driven CD8⁺ T cell differentiation.

We next asked if the abundance of clonally expanded terminally differentiated CD8⁺ T cells was reflective of response to therapy. ASCT patients were followed for a median of 3.9 years (range 1–5.8) after which 12 of 19 (63%) patients relapsed. The presence of detectable residual tumour in the marrow post-ASCT (minimal residual disease positive, MRD-positive) is associated with earlier relapse[17]. Notably, we found that post-ASCT marrow T cells from MRD-positive patients demonstrated significantly higher TCR clonality ($p = 0.035$, Fig. 6H). As ASCT-expanded clones were terminally-differentiated pre-treatment (Fig. 6E), we asked if T cell skewing prior to ASCT was associated with poor outcome. We identified MM patients in our CyTOF cohort who received ASCT and clustered CD8⁺ T cells into phenotypes aligned with our scRNA-seq data (Supplementary Fig. 8J). The abundance of CD8⁺ Tte and CD8⁺ Temra in pre-treatment samples was significantly associated with MRD-positivity ($p = 0.02$ and $p = 0.03$ respectively, unpaired Wilcoxon test; Supplementary Fig. 6I), suggesting T cell skewing at diagnosis predicts poor outcome following ASCT.

T cell skewing is not associated with rapid SMM progression

Finally, we explored whether T cell dynamics could be used to improved SMM risk stratification. To assess an association between T cell features and progression, 199 precursor disease patients were followed for a median of 1.1 years (range 1–5.8) after which 25 (15%) patients progressed to active disease (Fig. 7A). As expected, we found an association between time-to-progression and known progression-associated clinical features (BM PC %, paraprotein, FLC ratio: $p < 0001$ log-rank test for all).

Firstly, we asked whether T cell skewing identified SMM patients at high risk of progression. T cell skewing was not significantly enhanced in MM relative to SMM in either our scRNA-seq ($p = 0.53$; Fig. 3C) or CyTOF ($p = 0.37$; Supplementary Fig. 10A) cohorts. Similarly, in our flow cytometry cohort the abundance of late-differentiated IL7R⁻CD8⁺ T cells was not higher in SMM patients who subsequently progressed relative to non-progressors ($p = 0.25$, one-way ANOVA; Fig. 7B). As such, there was only a weak trend between IL7R⁻CD8⁺ T cells and time-to-progression ($p = 0.13$, log-rank test; Fig. 7C). These results demonstrate that T cell skewing presents similarly in asymptomatic SMM and active MM and, in this cohort, does not predict progression.

Regulatory CD4⁺ T cells loss is a biomarker to predict risk of SMM progression

To explore T cell features which might identify SMM patients at increased risk of progression, we returned to our analysis comparing SMM and MM patients (Supplementary Fig. 5E). We observed a reduction in the abundance of CD4⁺ Treg in MM relative to SMM in the scRNA-seq (uncorrected $p = 0.03$) and CyTOF ($p = 0.05$) cohorts (Fig. 7D). In the larger flow cytometry cohort, we found CD4⁺ Treg abundance was reduced in SMM patients that subsequently progressed relative to both non-progressors and MGUS patients (Fig. 7E, left and Supplementary Figs. 9B and 10B). As no individual CD4⁺ Treg sub-cluster correlated with CD4⁺ Treg loss (Supplementary Fig. 4A; $p > 0.1$ for all, Pearson correlation) we concluded this represented an overall depletion of CD4⁺ Treg numbers. As such, this effect was also seen in the reduction of CD4⁺ Treg among total (non-tumour) cells in progressing SMM patients (Fig. 7E, right). Together, this suggests CD4⁺ Treg loss defines active myeloma relative to precursor disease.

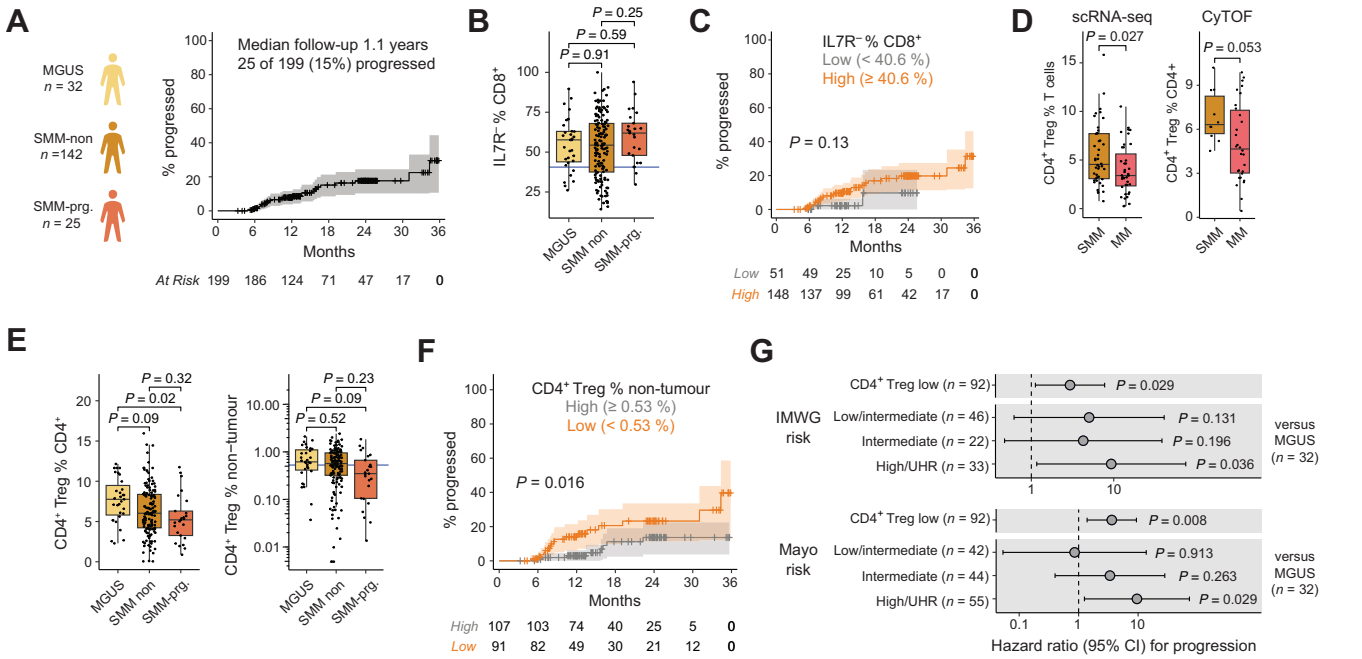

Fig. 7 | CD4⁺ Treg loss but not T cell skewing is associated with progression from SMM to symptomatic disease. A Left, composition of precursor disease flow cytometry cohort. SMM-prg. and SMM-non indicate SMM patients who did and did not progress, respectively. Right, Kaplan–Meier curve showing the frequency of progression in the precursor disease cohort. **B** Box plots comparing the abundance of IL7R⁻ cells among CD8⁺ T cells in MGUS ($n = 32$), SMM-non ($n = 142$) and SMM-prg. ($n = 25$) patients. Vertical line indicates the abundance threshold used in (**C**). **C** Kaplan–Meier curve showing the impact of high (orange) and low (grey) IL7R⁻ % CD8⁺ on the frequency of progression in 199 precursor disease patients. Number of patients at risk inset below. **D** Box plots comparing the abundance of CD4⁺ Treg in SMM and MM patients in scRNA-seq (SMM $n = 44$, MM $n = 39$) and CyTOF (SMM $n = 10$, MM $n = 30$) samples. **E** Box plots comparing the abundance of CD4⁺ Treg among CD4⁺ T cells (left) and non-tumour cells (right) in MGUS ($n = 32$), SMM-non ($n = 142$) and SMM-prg. ($n = 25$) patients. Vertical line indicates the abundance threshold used in (**F**). **F** Kaplan–Meier curve showing the impact of low (orange) and high (grey) CD4⁺ Treg abundance (% non-tumour cells) on the frequency of progression in 199 precursor disease patients. **G** Forest plot showing the significant relationship between risk of progression with the abundance of CD4⁺ Treg among non-tumour cells and IMWG or Mayo SMM risk classification. p values calculated with two-sided Cox proportional hazard model without multiple testing correction are inset. CI confidence interval. Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the IQR. For **B**, **E**, p values derived by one-way ANOVA followed by Tukey's test. For **D**, p values derived by two-sided Wilcoxon test. For **C**, **F**, p value calculated with log-rank test.

We explored the ability of CD4⁺ Treg loss to serve as a marker of SMM progression. We focused on the abundance of CD4⁺ Treg among non-tumour cells as this quantification was independent of existing risk factors (paraprotein $p = 0.64$ and BM PC % $p = 0.58$, Pearson correlation; Supplementary Fig. 10C). MGUS and SMM patients with low CD4⁺ Treg abundance (<0.53% non-tumour cells) were significantly more likely to progress to MM ($p = 0.016$ log-rank test, hazard ratio = 2.7; Fig. 7F). We obtained the same results when removing MGUS patients ($n = 167$ SMM, $p = 0.035$ log-rank test hazard ratio = 2.4). Multivariate analysis showed CD4⁺ Treg abundance predicted progression independently of existing SMM risk models (Mayo $p = 0.04$, IMWG $p = 0.08$; Fig. 7G). Together, this suggests that CD4⁺ Treg loss represents a putative biomarker for SMM risk.

Discussion

In this study, we curated a large cohort of single-cell data across 12 studies and 259 donors to understand alterations to T cells in precursor and symptomatic MM. We enhanced our analysis by analysing the TCR repertoire, the BM and PB, and tumour cells. This allowed us to identify the specific features of tumour biology associated with T cell differentiation in MM, independent of natural heterogeneity attributable to tissue and age. We describe two patterns of myeloma-associated T cell differentiation (Fig. 8): (1) terminal memory T cells with features of antigen-specific differentiation and lacking viral-specificity accumulate in proportion to serum paraprotein (tumour bulk); and (2) effector T cells are enriched in highly-infiltrated marrows populated by stressed tumour cells. As these two MM-associated T cell subsets are clonally related (Fig. 5E, F), we suggest they represent the in situ differentiation of tumour-reactive clones accumulating through disease evolution, which then circulate in the blood. Finally, we demonstrate the relevance of T cell dynamics in newly diagnosed patients receiving ASCT and rapidly-progressing precursor patients, identifying clinical associations with T cell functionality.

Our findings resolve conflicting reports on the presence of exhausted T cells in myeloma and a poor history of checkpoint inhibition in this setting[18–20]. We show exhausted T cells are not pervasively enriched in myeloma and are distinct from the more abundant CD8⁺ Tem.*CD69*⁺ which match descriptions of PD-1-expressing CD8⁺ T cells in healthy donors[21,22]. The presence of these "pseudo-exhausted" cells may have led to misidentification of exhausted cells in myeloma patients[23], especially given their enrichment in the bone marrow (Supplementary Fig. 4F). Our results align closely with a recent report showing CD8⁺ Tex are absent in MM patient marrows[24] but are a feature of extramedullary lesions[25], with the latter report suggesting marrow-intrinsic biology curtails exhaustion. It could be speculated that the donor-specific CD8⁺ Tex (Fig. 2H) represents the accidental or unreported sampling of an extramedullary lesion. While not phenotypically exhausted or functionally impaired (Supplementary Fig. 4C), CD8⁺ Tem.*CD69*⁺ may still be involved in bone marrow pathology, as cells resembling this phenotype possess negative and positive associations with the response to TCE therapy in advanced MM and combination therapy in SMM, respectively[14,26].

We show myeloma is associated with an enrichment of terminally differentiated clonal memory T cells, extending previous reports[27–29] by demonstrating an independence from age and showing concurrent changes to the TCR repertoire and systemic T cells. These alterations are similar to T cell immunosenescence changes seen during ageing[8], implying MM patients have prematurely aged T cell compartments (Fig. 3E–G). Exaggerated T cell ageing in precursor conditions may explain the increased risk of infections in these patients[30], and impede the ability to control tumour growth, facilitating progression. Additionally, this pattern of T cell differentiation could mechanistically represent tumour-targeting T cell responses. Patient-derived T cells show evidence of tumour-reactivity in myeloma[13,31,32], specifically terminal memory CD8⁺ phenotypes[13,33], supporting our in-silico evidence that non-viral reactivity may contribute to myeloma-associated T cell differentiation (Fig. 4F, G). However, inflammation, pervasive in the myeloma marrow[34], can alternatively drive antigen-independent memory T cell differentiation[35].

We have previously reported in newly diagnosed MM patients that induction chemotherapy and ASCT resulted in BM T cell activation and TCR repertoire clonality[36]. Here, we extend those observations by showing ASCT-expanded clones are terminally differentiated CD8⁺ T cells that are detectable in this phenotype pre-treatment, indicating

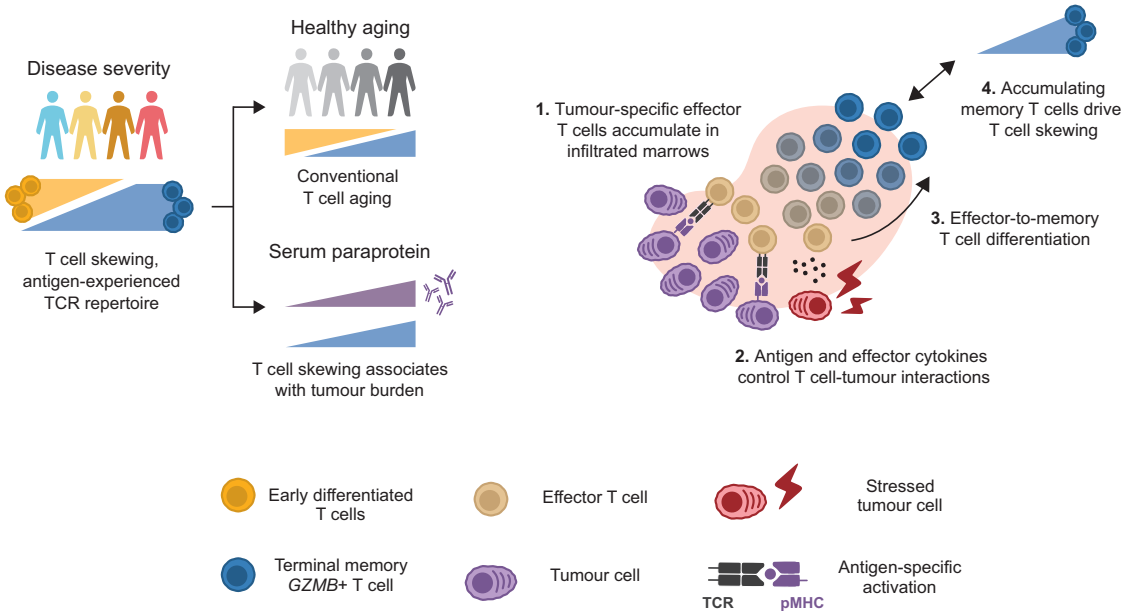

Fig. 8 | Summary of findings and proposed biological model. With increased disease severity, T cells transition from early differentiated subsets to terminal memory GZMB-expressing subsets with features of an antigen-experienced repertoire (collectively, termed T cell skewing). A similar effect occurs throughout healthy ageing, and in patients this effect is associated with serum paraprotein. In highly-infiltrated marrows (high BM PC %), terminal memory subsets (which are also found in the peripheral blood) are clonally related with effector subsets. Here, TCR specificity and effector cytokines (IFN-γ and TNF-α) are involved in T cell-tumour interactions. This process gives rise to the accumulation of terminal memory cells associated with T cell skewing.

that pre-existing antigen-experienced T cells expand following ASCT, consistent with previous reports in patients and mice[15,16]. However, as clonal expansion was most prominent in patients with residual disease post-treatment (Fig. 6H), this response may not be effective in eradicating residual tumour and suggests further therapy may be needed to augment T cell function post-ASCT. Agents that stimulate T cell function, such as immunomodulatory drugs or CD38 monoclonal antibodies, may be required because, while tumour-reactive T cells in SMM and MM are not phenotypically exhausted (Fig. 4), they possess hallmarks of senescent Temra differentiation, including expression of T-bet (*TBX21*) and *ZEB2*[2]. Furthermore, other immunosuppressive immune cells and genomic sub-clonality may also prevent tumour-reactive T cells from mounting effective responses in myeloma.

T cell skewing presented similarly in asymptomatic SMM and overt MM (Fig. 3) and progressing and non-progressing SMM (Fig. 7B, C). This suggests that SMM is an immunologically mature entity in regard to this axis of T cell differentiation and would explain the poor ability of T cell skewing to identify rapidly-progressing SMM patients. Furthermore, T cell skewing was more closely associated with serum paraprotein than disease stage (Fig. 5A–C). Further work is needed to explain this association, but we note malignant immunoglobulin-derived peptides can serve as immunogenic epitopes[37]. While we were unable to comprehensively compare T cell features and tumour genomic classification, a similar pattern of T cell skewing was previously shown to be enhanced in hyperdiploid patients[29].

Our observations indicate that $CD4^+$ Treg loss represents a T cell biomarker for rapidly-progressing SMM patients (Fig. 7D–G). While appearing at odds with earlier reports describing immunosuppressive $CD4^+$ Treg in overt MM[38,39], $CD4^+$ Treg constitute a component of the normal haematopoietic niche[40], hence their depletion may reflect early reshaping of the BM microenvironment towards a state favouring progression.

Our findings provide a conceptual framework for how T cells are altered during myeloma disease evolution and highlight the importance of contextualising immune heterogeneity with tumour biology when exploring immune biomarkers in myeloma.

Methods

Clinical sample acquisition

Bone marrow aspirates from individuals with myeloma or precursor conditions were obtained from patients included in one of four ongoing clinical trials: (1) Defining risk in smouldering myeloma for early detection of multiple myeloma (COSMOS), a multicentre, observational UK study in smouldering myeloma (NCT05047107, COSMOS study UK Research Ethics Committee reference: 270077); (2) Risk-Adapted therapy Directed According to Response (RADAR), a randomised phrase II/III trial in newly diagnosed patients with multiple myeloma eligible for transplant (UK Research Ethics Committee reference[20]:/LO/0238)[41]; (3) Carfilzomib/Cyclophosphamide/Dexamethasone with Maintenance Carfilzomib in Untreated Transplant-eligible Patients with Symptomatic MM to Evaluate the Benefit of Upfront ASCT (CARDAMON), a phase II trial (UK Research Ethics Committee reference: 148600)[42]; (4) Biology of Myeloma, an observational study open to all plasma cell disorder patients treated at University College London Hospitals (Research ethics committee reference: 07/Q0502/17). Bone marrow aspirates from non-cancer controls were collected as a by-product of routine elective orthopaedic surgery (hip or knee replacements) via the UCL/UCLH Biobank for Studying Health and Disease (UK Research Ethics Committee no: 272816). Material was obtained following written informed consent in accordance with the Declaration of Helsinki.

Sample acquisition and processing

Bone marrow aspirates were collected in ethylenediamine-tetraacetic acid (EDTA, Cambridge Bioscience, 60-00030-11) and processed within 24 h of collection. Mononuclear cells (MNCs) were isolated by Ficoll Paque density gradient centrifugation, using SepMate tubes (StemCell Technologies, 85420). Freshly isolated BM MNCs were analysed for tumour infiltration (BM PC %) by multi-parameter flow cytometry. Tumour cell marrow infiltration was determined as the frequency of live BM MNCs cells co-expressing CD38 and CD138 (Supplementary Fig. 1E). For CyTOF experiments, MNCs were cryopreserved in 90% FBS and 10% DMSO prior to use. For flow cytometry experiments, samples were assessed fresh immediately after acquisition.

Multi-parameter flow cytometry antibody staining, data acquisition and analysis

Single cell suspensions from freshly isolated BM mononuclear cells were resuspended in PBS, blocked with 5% mouse and rat serum and stained with the following antibodies: CD138-PE (1:100; MI15, BioLegend, 356504), CD38-PE-Cy7 (1:25; HB7, BioLegend, 356608), CD3-BV785 (1:100; OKT3, BioLegend, 317330), CD56-BV605 (1:100; NCAM16.2, BD Biosciences, 562780), CD4-FITC (1:50, OKT4, BioLegend, 317408), CD8-PB (1:100, RPA-T8 BioLegend, 301033), CD25-BV711 (1:100, M-A251, BioLegend, 356138), CD127-PECy5 (1:100, A019D5 BioLegend, 351324). Fixable Viability Stain-780 (1:250; BD Biosciences, 65-0865-14) was used for dead cell exclusion. Samples were measured by LSRFortessa Cell Analyser (BD Biosciences) and manually gated (Supplementary Fig. 9B) with FlowJo (v10, BD Biosciences).

scRNA-seq and scTCRseq sample and library preparation

For newly-generated "T cell–enriched/depleted" scRNA-seq samples T cells were enriched from freshly isolated BM MNCs by magnetic separation using a Pan T cell Isolation Kit (Miltenyi Biotec, 130-096-535) and CD15 MicroBeads (Miltenyi Biotec, 130-046-601). After sorting, the T cell depleted and enriched compartments were pelleted and resuspended in 0.04% BSA in PBS at 10^6 cells/mL and loaded onto the Chromium Controller (10X Genomics). For newly generated 'CD8-enriched' samples T cells were enriched using the same protocol with the addition of CD4 MicroBeads (Miltenyi Biotec, 120-000-440) and only CD8-enriched samples were loaded. This generated a total of 47 libraries. All samples were processed using the Chromium Next GEM Single Cell 5' Dual Index Kit (10X Genomics, v2) following manufacturer's protocol. T cell and CD8-enriched samples were additionally processed using the VDJ kit (10x Genomics). The libraries were sequenced by Illumina NovoSeq 6000. We used CellRanger v6.0.0 pipeline (10x Genomics) to align gene expression (GEX) and V(D)J (Immune Profiling) experiments using the GRCh38-2020-A and vdj_GRCh38_alts_ensembl-5.0.0 human reference genomes, respectively. Across samples, a median of 6367 cells with a median proportion of 0.76 cells with productive V J spanning TRA and TRB pairs per-sample.

Publicly available data

To assemble our large integrated dataset, scRNA-seq data from 12 published studies was acquired and combined with newly generated data (Supplementary Data 6)[18,26–28,43–50]. Specifically, data shared through the gene expression omnibus (GEO) can be accessed for Maura et al. under the accession GSE161195, Bailur et al. GSE163278, Oetjen et al. GSE120221, Granja et al. GSE139369, Zavidij et al. GSE124310, Kfoury et al. GSE143791, Zheng et al. GSE156728, Botta et al. GSE205393, and Friedrich et al. GSE216571. Data shared via dbGaP for Sklavenitis-Pistofidis et al. can be accessed under accession phs002476.v1.p1. Data shared online can be accessed for Stephenson et al. at covid19cellatlas.org, Conde et al. at tissueimmunecellatlas.org, and Liu et al. at humancellatlas.org/projects/2ad191cd-bd7a-409b-9bd1-e72b5e4cce81.

Filtering, integration, clustering, and dimensionality reduction of scRNA-seq data

scRNA-seq data were analysed using scanpy (1.8.2)[51]. Gene-barcode matrices for all newly generated and re-analysed samples were assigned unique sample-specific barcodes, merged, and subset to high-quality cells for integration (minimum unique genes >200, minimum total counts >500, total percentage mitochondrial chromosome-encoding transcripts <10%, total percentage transcripts encoding haemoglobin genes *HBB*, *HBA1* and *HBA2* < 20%). Cells called as doublets by scrublet (0.2.3)[52] were removed. Samples with <100 high-quality cells were removed before integration.

For integration, we utilised single-cell variational inference (scVI) from the scvi-tools package (0.15.2)[53]. A subset of 7000 highly variable genes across batches were calculated using log ($x+1$) normalised gene expression with the function scanpy.pp.highly_variable_genes (adata, batch_key = "batch") to identify genes with consistently high inter-cellular variation across different batches. Specific gene groups which can vary between cells for technical (mitochondrial, representing cell stress) or irrelevant biological (immunoglobin and TCR genes, representing lymphocyte clonality) reasons were excluded from highly variable genes to prioritise clustering on phenotype-defining genes. The un-normalised expression of these 7000 variable genes was prepared for a scVI model using the function scvi.model.SCVI.setup_anndata () with sample batch as the batch key and sample identifier and 10x chemistry as categorical covariate keys. A scVI model was then initialised with the following non-default parameters: scvi.model.SCVI (n_latent=30, n_layers=2, dropout_rate=0.2, gene_likelihood = "nb"). These parameters (number of variable genes, number of latent dimensions and hidden layers, dropout rate) were selected through a parameter sweep focused on minimising batch influence on integrated latent representation and retaining biological identity (data not shown). Minimisation of batch influence was assessed by linear regression of latent dimensions against batch covariates as implemented by scib (https://github.com/theislab/scib). The retention of biological identity was assessed by analysing the separation of $CD4^+$ and $CD8^+$ T cells (the median log ratio of *CD4*-expressing and *CD8A*-expressing cells closest to zero across clusters). This model was trained for a maximum of 400* (20,000*x) epochs where x was the number of input cells. Integration was first performed on all cells then repeated for just T cell clusters using 5000 highly variable genes but otherwise identical parameters.

The latent representation of the trained scVI model was used to create a k-nearest neighbours graph using scanpy.pp.neighbors (adata, n_neighbors=10) for subsequent graph-based clustering using the Leiden algorithm. The size of the local neighbourhood (n_neighbors=10) and Leiden clustering resolutions were selected for optimum granularity of biological clusters. Analysis of the latent representation was used as input for creation of a uniform manifold approximation and projection (UMAP, scanpy default parameters) or Minimum-Distortion Embedding using pymde (0.1.15)[54]. For visualisation of a large number of cells on either UMAP or MDE, scattermore (1.0) was used to create dot plots.

Differential expression and pathway analysis of scRNA-seq data

Differential expression between specified conditions was performed using scran (1.26.2) function pairwiseTTests () between specified contrasts with batch as the blocking level for each cell to model for batch effects. This restricts differential expression comparisons within individual batches and pools the downstream result, meaning no inter-batch comparisons were performed (which would incur batch effects). Marker genes were combined with supervised analysis of the expression of known RNA and protein markers to phenotype clusters. Genes were identified as significantly differentially expressed with a false discovery rate (FDR, Benjamini and Hochberg-adjusted p value) of <0.1. Pathway analysis of differentially expressed genes was performed using fgsea (1.24.0) with gene set enrichment analysis of gene sets from BIOCARTA, KEGG and REACTOME databases accessed via msigdbr (7.5.1; Supplementary Data 3).

Phenotyping gene expression clusters from scRNA-seq data

Cluster markers genes were calculated using log-normalised expression of all genes in a study-aware fashion using the findMarkers function (scran) specifying test.type = "wilcox" and batch as the blocking level for each cell. Clusters characterised by expression of known stress-associated genes (for example, *JUN*, *FOS*)[55] or by co-expression of marker genes for independent phenotypes (for example, T and B cells) were removed. For T cell cluster phenotyping, clusters lacking expression of *CD3D*, *CD3E* and *CD3G* or both *CD4* and *CD8A* were removed. Cell type prediction tools were run with default parameters: "Azimuth" (https://azimuth.hubmapconsortium.org/) and "Celltypst" (https://www.celltypist.org/). Manually curated T cell naïve and cytotoxicity gene signatures were taken from Chu et al.[56]. Gene sets were applied to cells UCell (2.2.0)[57].

$CD4^+$ Treg sub-clusters were defined by early markers *CCR7* and *TCF7*, suppressive molecules CD25 (*IL2RA*) and *CTLA4*, and activation markers GITR (*TNFRSF18*) and OX40 (*TNFRSF4*; Supplementary Fig. 4A. The invariant cluster was composed of γδ T cells (gdT) (defined by *TRDV2* and *TRDC*) alongside *KLRB1* and *SLC4A10*-expressing MAIT cells (Supplementary Fig. 4B). One MAIT sub-cluster was enriched in Th17 *RORC* and *CCR6* possibly corresponding to MAIT17 cells[58].

Differential abundance analysis of scRNA-seq data

We normalised cell type abundance following a compositional data framework[59]. For each sample, cluster counts were derived and zero values replaced by a Bayesian-multiplicative replacement strategy which preserves the ratios between non-zero clusters, implemented using the zCompositions (1.4.0-1) function cmultRepl ()[60], generating zero-imputed pseudo-counts. The centred log-ratio (CLR) transformation was then used to transform pseudo-counts relative to the geometric mean of all clusters in a given sample, implemented using the compositions package (2.0-6) function clr(). The CLR transformation thus reports cell type abundance relative to the per-sample average seeking to reduce the mutual dependency of proportional data[61].

Normalised cluster abundances were used as input for a combination of intercept-only and additive regression models exploring the relationship between cluster abundance and different conditions (for example, patient group, or patient group and age) as described. Selected comparisons were also performed using a mixed-effect model with an additional random effect term (for sample, study of origin) implemented using lmerTest (3.1-3).

For samples from Stephenson et al.[46] the median age of each age range was used (for example, 52.5 was used for the 50–55 group). For the model in Fig. 3E, age values were binarized to above and below the median (62 years) but results remained significant when modelling age as a continuous variable in both SMM (disease $p < 0.001$, age $p = 0.01$, linear model) and MM (disease $p < 0.001$, age $p = 0.02$, linear model). T cell skewing was greatest (enrichment of terminal memory clusters) in SMM and MM relative to controls independent of age ($p < 0.004$ and $p < 0.001$, respectively, linear regression; Fig. 3E). In the cases of patients with longitudinal sampling[45,48], only the first longitudinal timepoint was analysed unless otherwise specified. For analysis of paraprotein data, only patients with IgG and IgA isotype tumours (which produce heavy chain paraprotein molecules measurable in serum) were analysed.

The abundance of invariant and $CD4^+$ Treg sub-clusters was not significantly different between health and disease (all comparisons $p > 0.1$, linear regression), except for an enrichment of γδ T cells and $CD4^+$ Th17-like MAITs in SMM relative to controls (Supplementary Fig. 5G).

Unsupervised ordination of T cell composition and calculation of exaggerated T cell ageing

Normalised T cell cluster abundance was scaled per-cluster and used as input for PCA using the base R stats function prcomp. For the comparison of T cell skewing with the peripheral blood (Fig. 3H), we first re-scaled normalised peripheral blood T cell cluster abundance to the same centre and range of the scaled BM T cell abundance matrix. Next, we multiplied this scaled PB matrix by the pre-calculated BM PCA feature loadings via feature (cluster)-wise matrix multiplication, yielding the PC1 values for the input PB matrix.

Exaggerated T cell ageing was calculated by first constructing a linear regression model examining the relationship between PC1 and age for non-cancer controls only. The residuals of this model (Supplementary Fig. 5J) were normality distributed as assessed by Shapiro–Wilk test ($p = 0.89$), Kolmogorov–Smirnov test ($p = 0.97$) and visually via a quantile-quantile plot (Supplementary Fig. 5K). Next, the age of cancer patients was used to predict PC1 values for each patient in this model. The difference between predicted and real PC1 values (residuals) for each patient was interpreted as the difference between the T cell skewing expected for each patient's age versus their observed T cell skewing, respectively. These residual-derived values were termed "exaggerated T cell ageing". A patient was considered to have exaggerated T cell ageing if their values were greater than zero (meaning skewing was greater than expected for their age). Similarly, "excess T cell years" were calculated as the residuals between a patient's real age and the age predicted from their PC1 values by a model of PC1 and age in controls.

Label transfer

We validated our phenotypes in an independent scRNA-seq dataset of BM T cells from healthy donors ($n = 3$) alongside precursor disease ($n = 7$) and MM patients ($n = 10$) from Botta et al. (*19*). The scANVI model from the scvi package was used for label transfer, using the Botta et al. dataset as the query and the full T cell dataset as the reference. The query dataset underwent the same pre-processing and gene filtering as the full T cell dataset. This processed query dataset was then used as input for a scANVI model. For the input reference model of this scANVI model, the full T cell dataset integration SCVI model was used. This scANVI model was then trained with the same hyperparameters as this scVI model. Next, the trained scANVI model was used to predict the T cell phenotype of unlabelled query cells based on labelled reference cells following the scArches semi-supervised surgery pipeline[62]. Briefly, a k-nearest neighbours' graph was constructed from the joint query and reference latent space generated by the trained scANVI model. Then, based on the abundance of labelled reference cells near unlabelled reference cells, each reference cell was assigned a weighted prediction for each possible reference label. High confidence predicted labels (uncertainty value > 0.2) were determined to be the T cell phenotype of each query cell and taken forward for downstream analysis. This approach made accurate predictions of cell type identity matching the same expression profile as the larger dataset (Supplementary Fig. 4G).

scTCRseq pre-processing, clonal expansion calculation, T cell subset identification and clustering

TCR paired alpha and beta clones (clonotypes) were defined by CellRanger VDJ (raw_clonotype_id, clonotype_id) by matching shared V and J gene and nucleotide CDR3 sequences for alpha and beta TCR chains. Clonotypes were appended to single cells by matching cell barcodes. For scTCR-seq derived from published data, we utilised published clonotype identifiers. Clonal expansion was calculated as the abundance of cells labelled with each clonotype identified in each sample. $CD8^+$ and $CD4^+$ clones were identified by the presence of >75% of a clone's cells within $CD4^+$ or $CD8^+$ clusters (otherwise, were removed). Repertoire clonality was calculated among each specific subset of cells (such as all T cells or $CD8^+$ memory cells) with a minimum of 100 cells using Simpson's diversity index[63]. Clusters of TCRs with similar sequence features were identified within a single patient's alpha or beta chain repertoire using tcrdist3 (0.2.2)[64] using default parameters. TCR clustering networks were constructed and visualised using igraph (R, 1.4.2). CDR3 sequence logos were created with Logomaker (https://logomaker.readthedocs.io/en/latest/).

For analysis of scTCR-seq from Friedrich et al.[14] (Supplementary Fig. 8I), clones which significantly expanded post-treatment was identified with a Poisson framework, modelling the abundance of each clone pre- and post-treatment as the number of events and the total number of cells in each sample as the rate of Poisson sampling (to account for differences in sample size).

Annotation of HLA-matched viral reactivity-annotated TCR clones

HLA genotypes for 19 patients were derived using arcasHLA[65] ran on Cellranger output bam files (possessorted_genome_bam). All 19 patients were newly sequenced for this study and therefore a combination of T cell-enriched/depleted and CD8-enriched (Supplementary Fig. 1B) samples were available. arcasHLA was ran on every sample for each patient. HLA genotype for class I and class II HLA was almost entirely identical across samples for an individual. In the rare cases of two different samples of the same patient possessed different HLA genotypes, both predicted genotypes were ignored.

Each donor's repertoire was then compared against the annotated TCR reactivity database VDJdb, IEDB and CEDAR[66–68] subset to TCRs with annotated reactivity against an epitope from a single human virus: cytomegalovirus (CMV), Epstein Bar virus (EBV), Influenza A, or severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). TCRs annotated as reactive against more than one human virus were also removed. The viral dataset set was further subset to HLA-matched sequences for each patient's HLA genotype. A query TCR clone was annotated as putatively viral-reactive if at least one alpha or one beta chain CDR3 sequence perfectly matched a CDR3 annotated against the same virus in the database, and this clone's paired chain also perfected matched or possessed a highly similar CDR3 sequence to the same virus in the same HLA background. CDR3 similarity was performed as described previously[69]. Briefly, each TCR chain's CDR3 amino acid sequence was deconstructed into a series of overlapping triplets. Pairwise similarity between two CDR3 was defined as the number of shared triplets normalised to the number of triplets per comparison.

Analysis of tumour cell transcriptional state

To analysed tumour cell transcriptional features, we identified putative malignant clones via clonal immunoglobulin usage. We then scored 67,656 plasma cells from 46 patients with a set of pan-cancer transcriptional pathways (Supplementary Fig. 7E and Supplementary Data 3). We additionally identified individual genes enriched in individual patient's tumour cells relative to normal plasma cells (Supplementary Fig. 7F and Supplementary Data 3) to identify pathways frequently upregulated by tumour cells across multiple patients.

Identification of malignant plasma cell clones in scRNA-seq

Patient plasma cells were isolated from the clustering of all cells (Fig. 2A) and patients with <50 plasma cells were removed. This generated 67,656 plasma cells from 46 patients with a median of 467 plasma cells each (range: 76–13,638). To identify tumour cells among plasma cells, we leveraged the clonal plasma cell origin of myeloma. First, we attached the expression of all available immunoglobin genes to each cell, after removing any gene filtering performed earlier. Then, for each plasma cell, we identified the most highly expressed light variable (kappa or lambda), heavy variable, and heavy constant chains. We next quantified and ranked the abundance of every gene for each chain among an individual's plasma cells (Supplementary Fig. 7A).

Most individuals possessed a single gene for each chain which was expressed by the majority of plasma cells (light variable median 95% range: 31–100%, heavy constant median 96% range: 35–100%, heavy variable median 91% range: 24–100%). The frequency of different light and heavy variable genes among tumour cells matched previously reported frequencies in myeloma[70], including IGHV3-30 in 3 (6.5%) and IGKV1-39 in 2 (4.3%) patients. We inferred that clonal immunoglobulin expression corresponded to clonal plasma cells and labelled any plasma cell expressing the most highly abundant gene for each chain in that donor as a tumour cell. This method yielded 67,048 predicted tumour cells. Predicted tumour cells uniquely co-expressed clonal immunoglobulin genes (Supplementary Fig. 7B) and expressed genes characteristic of their translocation subgroups (Supplementary Fig. 7C)[71], suggesting they did represent malignant cells. These cells composed the majority of plasma cells in all patients but were most abundant in MM (Supplementary Fig. 7D).

Transcriptional pathway analysis of tumour cells in scRNA-seq data

To analysis tumour cell transcriptomes, we scored tumour cells using a set of pan-cancer transcriptional pathways[72] (Supplementary Data 3) using UCell. To compare the expression of individual pathways between patients, we calculated the abundance of cancer cells highly expressing a given pathway as the percentage of cells with expression greater than one standard deviation above the median across all patient's tumour cells.

Identification of novel pathways enriched in malignant cells

To identify novel sets of genes enriched in malignant relative to normal plasma cells, we first isolated each the tumour cells from each patient in turn. Next, we performed differential expression between each patient's tumour cells only and all other plasma cells not classified as malignant. Differential expression was only performed between cells from the same sequencing batch. This yielded a set of malignant-enriched genes for each patient. Pathway analysis was then performed as described. The pathways significantly enriched among malignant-associated genes in four or more patients were identified (Supplementary Fig. 7F).

Cell-cell interaction analysis

Differentially-expressed genes from T and tumour cells were screened for cognate ligand-receptor interactions in the OmniPath database[73] via OmnipathR (3.15.1). From this set of interactions, a cell-cell interaction graph was created with igraph and visualised with ggraph (2.2.1).

Transcriptional pathway expression in CoMMpass

We analysed an association between the expression of the stress and MHC pathways (Supplementary Data 3) with overall survival in bulk RNA sequencing samples from the CoMMpass cohort of newly diagnosed MM patients, with RNA sequencing data processed and normalised as described in Bauer et al.[74]. We calculated the expression of the pathway by taking the expression of each constituent gene, scaling expression between 0 and 1, and taking the average.

Patient outcome analysis

We assessed the predictive power of (1) transcriptional pathway expression for overall survival in patients in the CoMMpass dataset (Fig. 5J); (2) T cell subset abundance and time-to-progression (Fig. 7C, F). Covariate thresholds (such as expression level) for outcome were selected using the maximally selected rank statistic as implemented by maxstat (0.7-25). The association between this threshold and outcome was assessed using univariate or multivariate (as in Fig. 7G) Cox proportional hazards regression models using the "survival" (3.5-5) with default parameters.

Deep TCR sequencing

Before the RNA extractions samples were T cell enriched, by performing Pan T cell isolation (Miltenyi Biotec), following manufacturer protocol. RNA was extracted using RaliaPrep RNA Cell Miniprep System (Promega), following the manufacturer's instructions. RNA integrity was assessed by TapeStation (Agilent Technologies). TCR α-chain and β-chain sequencing was performed by utilising whole RNA extracted from CD138-depleted cells from 19 patients with matched pre- and post-ASCT samples, by using a quantitative experimental and computational TCR sequencing pipeline described[69]. Clonality and clustering of deep TCR-seq samples was calculated identically to scTCR-seq.

$CD69^+$ T cell subset functional assessment

Bone marrow MNCs were obtained from patients enrolled on the RADAR study via density gradient centrifugation with Ficoll-Paque (Cytivia 17144003). Samples were then enriched for T cells using magnetic-activated cell sorting (Pan T cell isolation kit, Miltenyi Biotec, 130-096-535) and stained for flow-activated cell sorting with CD8-eFluor450 (1:100, SK1, Invitrogen 48-0087-42), CD69-BV605 (1:100, FN50, Biolegend 310938), $CCR7^-$BV785 (1:100, G043H7, BioLegend 353229), and Fixable Viability Dye eFluor 780 (1:250, eBioscience, 65-0865-14). An aliquot of unsorted cells was saved for staining (stim−). Cells were sorted into $CD8^+CCR7^-CD69^+$ and $CD8^+CCR7^-CD69^-$ fractions and both were collected. Cells were then resuspended at 1×10^6/ml in RPMI (Gibco 12027599) 10% FBS (Gibco 10500-064) 2 mM L-Glutamine and stimulated with 25 uL/ml ImmunoCult (STEMCELL, 100-0785) or 10 uL/ml TransAct (Miltenyi, 120-111-160) overnight at 37 °C (5% CO_2). Brefeldin A (Biolegend, 420601) was added 4 h prior to harvesting of cells. Cells were then harvested, washed, and extracellular antibodies stained for 30 min in the dark at 4 °C; CD57-FITC (1:100, HNK-1, Biolegend 359603), PD-1-BB700 (1:100, EH12.1, Biolegend 566460), Fixable Viability Dye eFluor 780 (1:250, eBioscience, 65-0865-14), CD3-BV421 (1:100, SK7, Biolegend 344834), CD8-V500 (1:100, RPA-T8, BD Horizon, 560774), CD69-BV605 (1:100, FN50, Biolegend 310938), CCR7-BV785 (1:100, G043H7, Biolegend 353229), and 4-1BB-PE-Dazzle (1:100, 4B4-1, Biolegend 309825). Cells were then washed, permeabilised, and intracellular markers stained (FoxP3/Transcription Factor Staining Buffer Set, eBioscience 00-8333-56) for 30 min in the dark at 4 °C: IFN-G-PE, (1:100, 4S.B3, Invitrogen, 12-7319-41) and TNF-a-APC (1:100, MAb11, Biolegend, 502912). Cells were then washed, and data acquired on NovoCyte Quanteon.

CyTOF antibody staining, data acquisition and analysis

Details on antibodies are listed in Supplementary Data 5. Conjugation of the purified antibodies with metal reporters was performed with the MaxPar X8 and MaxPar MCP9 antibody labelling kits (Fluidigm Sciences) according to the manufacturer's instructions. Frozen bone marrow MNCs or the CD138-negative populations were thawed rapidly at 37 °C and resuspended into pre warmed thawing media of RPMI (Sigma-Aldrich) containing 20% FBS, 2 mM EDTA (pluriSelect) and 5 mg DNase (Sigma-Aldrich). Cell suspensions were washed and filtered to form a single cell suspension. Cells were incubated with 5 mM Cell-ID Cisplatin (Fluidigm Sciences) in serum free RPMI for 3 min at room temperature (rT) to identify dead cells. Cells were then washed and incubated with human Fc block (BioLegend) for 10 min at rT before being barcoded using 6-choose-3 Cadmium CD45 Live Barcoding (Fluidigm Sciences). All samples were stained in the same batch. After live cell barcoding, the combined samples were then stained with metal-conjugated antibodies for surface antigens for 30 min at rT. After staining, cells were washed with MaxPar Cell Staining Buffer (Fluidigm Sciences, 201068) and permeabilised with MaxPar nuclear antigen staining buffer before staining with metal-conjugated antibodies for intracellular antigens. Cells were again washed and fixed using 1.6% paraformaldehyde. Cells were then

incubated with Cell-ID intercalator-Ir (Fluidigm Sciences) to stain all cells in MaxPar Fix and Perm Buffer (Fluidigm Sciences, 201067) and aliquoted and frozen in cryovials. Stained samples were thawed and washed on the day of acquisition.

Cells were acquired on the Helios mass cytometer (Fluidigm Sciences). Data from different days were normalised by using EQ Four Element Calibration Beads (Fluidigm Sciences). Data was debarcoded using the Fluidigm CyTOF software and patient sample fcs files run from different days were concatenated. Before downstream analysis, initial data clean-up was carried out using FlowJo. Live $CD3^+$ cells were exported by manual gating on Event_length, Residual, Offset, DNA (^{191}Ir and ^{193}Ir), live cells (^{195}Ir) and CD3 expression (^{89}Y). Further manual CyTOF gating of T cell phenotypes (Supplementary Fig. 9A) was performed with FlowJo (v10, BD Biosciences).

CyTOF clustering

CyTOF data were analysed using a custom R pipeline using gated $CD8^+$ T cells as input. Batch effects were normalised using the R package CytofIn (0.0.0.9000) using technical replicates (multiple aliquots of the same sample) in each batch. Batch correction was assessed using expression distribution and UMAP plots (not shown). Batch-corrected expression values were arcsine-transformed and clustered using FlowSOM (2.6.0) on a 12×12 node self-organising map. Clusters were assigned to phenotypes based on known $CD8^+$ T cell markers (Supplementary Fig. 8J) including $CD8^+$ Temra ($CD27^-CD28^-GZMB^+Tbet^+CD45RA^+$) and $CD8^+$ Tte ($CD27^-CD28^-GZMB^+CD57^+$).

Statistical analyses

Box plots represent the first and third quartiles around the median with error bars extending 1.5 times the interquartile range (IQR). For comparison of means in box plots, p values were calculated by either two-sided unpaired Wilcoxon test or one-way ANOVA followed by Tukey's test for pairwise comparison evaluation as indicated. Correlation coefficient (R) and p values for correlations were calculated by Pearson correlation. In linear regression slopes, shaded regions represent the 95% confidence interval. For hierarchal clustering on heatmaps, Euclidean distance was used as the default distance measure.

Reporting summary

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

Data availability

The unprocessed sequencing data have been deposited in the Sequence Read Archive under the accession PRJNA1401834. The processed single-cell RNA and TCR data (CellRanger outputs) have been deposited in the Zenodo repository under the accession 13171648. The full integrated single-cell RNA and TCR datasets and cohort information have been deposited in the Zenodo repository under the accession 17418275. All data are included in the Supplementary Information or available from the authors, as are unique reagents used in this Article. The raw numbers for charts and graphs are available in the Source Data file whenever possible. Source data are provided with this paper.

Code availability

The code to reproduce the analysis has been deposited on GitHub under the accession kanefos/myeloma-singlecell.

References

1. de Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell* **41**, 374–403 (2023).
2. Kay Chung, H., McDonald, B. & Kaech, S. M. The architectural design of CD8+ T cell responses in acute and chronic infection: parallel structures with divergent fates. *J. Exp. Med.* **218**, e20201730 (2021).
3. Chow, A., Perica, K., Klebanoff, C. A. & Wolchok, J. D. Clinical implications of T cell exhaustion for cancer immunotherapy. *Nat. Rev. Clin. Oncol.* **19**, 775–790 (2022).
4. van de Donk, N. W. C. J., Pawlyn, C. & Yong, K. L. Multiple myeloma. *Lancet* **397**, 410–427 (2021).
5. Mateos, M. V. et al. International Myeloma Working Group risk stratification model for smoldering multiple myeloma (SMM). *Blood Cancer J.* **10**, 1–11 (2020).
6. Landgren, O. et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood* **113**, 5412–5417 (2009).
7. Cohen, A. D. et al. How to train your T cells: overcoming immune dysfunction in multiple myeloma a C. *Clin. Cancer Res.* **26**, 1541–1554 (2020).
8. Mogilenko, D. A., Shchukina, I. & Artyomov, M. N. Immune ageing at single-cell resolution. *Nat. Rev. Immunol.* **228**, 484–498 (2021).
9. Di Rosa, F. Maintenance of memory T cells in the bone marrow: survival or homeostatic proliferation? *Nat. Rev. Immunol.* **16**, 271–271 (2016).
10. Ren, X. et al. Insights gained from single-cell analysis of immune cells in the tumor microenvironment. *Annu. Rev. Immunol.* **39**, 583–609 (2021).
11. Wherry, E. J. et al. Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* **27**, 670–684 (2007).
12. Greipp, P. R. et al. International staging system for multiple myeloma. *J. Clin. Oncol.* **23**, 3412–3420 (2005).
13. Wagner, T. R. et al. Conserved programs and specificities of T cells targeting hematological malignancies. *bioRxiv* **08**, 668919 (2025).
14. Friedrich, M. J. et al. The pre-existing T cell landscape determines the response to bispecific T cell engagers in multiple myeloma patients. *Cancer Cell* **41**, 711–725 (2023).
15. Vuckovic, S. et al. Bone marrow transplantation generates T cell–dependent control of myeloma in mice. *J. Clin. Invest.* **129**, 106–121 (2019).
16. Chung, D. J. et al. T-cell exhaustion in multiple myeloma relapse after autotransplant: optimal timing of immunotherapy. *Cancer Immunol. Res.* **4**, 61–71 (2016).
17. De Tute, R. M. et al. Minimal residual disease after autologous stem-cell transplant for patients with myeloma: prognostic significance and the impact of lenalidomide maintenance and molecular risk. *J. Clin. Oncol.* **40**, 2889–2900 (2022).
18. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).
19. Kourelis, T. V. et al. Mass cytometry dissects T cell heterogeneity in the immune tumor microenvironment of common dysproteinemias at diagnosis and after first line therapies. *Blood Cancer J.* **9**, 72 (2019).
20. Meermeier, E. W., Bergsagel, P. L. & Chesi, M. Next-generation therapies for multiple myeloma. *Annu. Rev. Cancer Biol.* **8**, 351–371 (2024).
21. Galletti, G. et al. Two subsets of stem-like CD8+ memory T cell progenitors with distinct fate commitments in humans. *Nat. Immunol.* **21**, 1552–1562 (2020).
22. Mogilenko, D. A. et al. Comprehensive profiling of an aging immune system reveals clonal GZMK+ CD8+ T cells as conserved hallmark of inflammaging. *Immunity* **54**, 99–115.e12 (2021).
23. Hudson, W. H. & Wieland, A. Technology meets TILs: deciphering T cell function in the -omics era. *Cancer Cell* **41**, 41–57 (2023).
24. Shasha, C. et al. Hallmarks of T-cell exhaustion and antigen experience are absent in multiple myeloma from diagnosis to maintenance therapy. *Blood* **145**, 3113–3123 (2025).

25. Lutz, R. et al. Bone marrow breakout lesions act as key sites for tumor-immune cell diversification in multiple myeloma. *Sci. Immunol.* **10**, eadp6667 (2025).
26. Sklavenitis-Pistofidis, R. et al. Immune biomarkers of response to immunotherapy in patients with high-risk smoldering myeloma. *Cancer Cell* **40**, 1358–1373.e8 (2022).
27. Zavidij, O. et al. Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nat. Cancer* **1**, 493–506 (2020).
28. Bailur, J. K. et al. Early alterations in stem-like/marrow-resident T cells and innate and myeloid cells in preneoplastic gammopathy. *JCI Insight* **4**, e127807 (2019).
29. Dang, M. et al. Single cell clonotypic and transcriptional evolution of multiple myeloma precursor disease. *Cancer Cell* **41**, 1032–1047.e4 (2023).
30. Ludwig, H., Meckl, A. & Engelhardt, M. Compliance with vaccination recommendations among patients with multiple myeloma: a real world experience. *HemaSphere* https://doi.org/10.1097/HS9.0000000000000597 (2021).
31. Noonan, K. et al. Activated marrow-infiltrating lymphocytes effectively target plasma cells and their clonogenic precursors. *Cancer Res.* **65**, 2026–2034 (2005).
32. Perumal, D. et al. Mutation-derived neoantigen specific T cell responses in multiple myeloma. *Clin. Cancer Res.* **26**, 450 (2020).
33. Vuckovic, S. et al. Inverse relationship between oligoclonal expanded CD69- TTE and CD69+ TTE cells in bone marrow of multiple myeloma patients. *Blood Adv.* **4**, 4593–4604 (2020).
34. de Jong, M. M. E. et al. The multiple myeloma microenvironment is defined by an inflammatory stromal cell landscape. *Nat. Immunol.* **226**, 769–780 (2021).
35. Henson, S. M. & Akbar, A. N. Memory T-cell homeostasis and senescence during aging. *Adv. Exp. Med. Biol.* **684**, 189–197 (2010).
36. Lee, L. et al. Increased immune-regulatory receptor expression on effector T cells as early indicators of relapse following autologous stem cell transplantation for multiple myeloma. *Front. Immunol.* **12**, 618610 (2021).
37. Trojan, A. et al. Immunoglobulin framework-derived peptides function as cytotoxic T-cell epitopes commonly expressed in B-cell malignancies. *Nat. Med.* **66**, 667–672 (2000).
38. Kawano, Y. et al. Blocking IFNAR1 inhibits multiple myeloma-driven Treg expansion and immunosuppression. *J. Clin. Invest.* **128**, 2487–2499 (2018).
39. Alrasheed, N. et al. Marrow-infiltrating regulatory T cells correlate with the presence of dysfunctional CD4þPD-1þ cells and inferior survival in patients with newly diagnosed multiple myeloma. *Clin. Cancer Res.* **26**, 3443–3454 (2020).
40. Méndez-Ferrer, S. et al. Bone marrow niches in haematological malignancies. *Nat. Rev. Cancer* **205**, 285–298 (2020).
41. Royle, K. L. et al. Risk and response adapted therapy following autologous stem cell transplant in patients with newly diagnosed multiple myeloma (RADAR (UK-MRA Myeloma XV Trial): study protocol for a phase II/III randomised controlled trial. *BMJ Open* **12**, e063037 (2022).
42. Yong, K. et al. Upfront autologous haematopoietic stem-cell transplantation versus carfilzomib-cyclophosphamide-dexamethasone consolidation with carfilzomib maintenance in patients with newly diagnosed multiple myeloma in England and Wales (CARDAMON): a randomised, phase 2, non-inferiority trial. *Lancet Haematol.* **10**, e93–e106 (2023).
43. Kfoury, Y. et al. Human prostate cancer bone metastases have an actionable immunosuppressive microenvironment. *Cancer Cell* **39**, 1464–1478.e8 (2021).
44. Botta, C. et al. Large T cell clones expressing immune checkpoints increase during multiple myeloma evolution and predict treatment resistance. *Nat. Commun* **141**, 1–15 (2023).
45. Liu, R. et al. Co-evolution of tumor and immune cells during progression of multiple myeloma. *Nat. Commun.* **121**, 1–18 (2021).
46. Stephenson, E. et al. Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.* **275**, 904–916 (2021).
47. Domínguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **80**, 376 (2022).
48. Oetjen, K. A. et al. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight* **3**, e124928 (2018).
49. Maura, F. et al. Genomic and immune signatures predict clinical outcome in newly diagnosed multiple myeloma treated with immunotherapy regimens. *Nat. Cancer* **412**, 1660–1674 (2023).
50. Granja, J. M. et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **3712**, 1458–1465 (2019).
51. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 1–5 (2018).
52. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
53. Gayoso, A. et al. A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **402**, 163–166 (2022).
54. Agrawal, A., Ali, A. & Boyd, S. Minimum-distortion embedding. *Found. Trends Mach. Learn.* **14**, 211–378 (2021).
55. Araúzo-Bravo, M. J., Izeta, A. & Ascensión, A. M. The need to reassess single-cell RNA sequencing datasets: the importance of biological sample processing. *F1000Res.* **10**, 767 (2021).
56. Chu, Y. et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nat. Med.* **296**, 1550–1562 (2023).
57. Andreatta, M. & Carmona, S. J. UCell: robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
58. Garner, L. C. et al. Single-cell analysis of human MAIT cell transcriptional, functional and clonal diversity. *Nat. Immunol.* **249**, 1565–1578 (2023).
59. Quinn, T. P. et al. A field guide for the compositional analysis of any-omics data. *Gigascience* **8**, 1–14 (2019).
60. Palarea-Albaladejo, J. & Martín-Fernández, J. A. zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemom. Intell. Lab. Syst.* **143**, 85–96 (2015).
61. Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B* **44**, 139–160 (1982).
62. Sikkema, L. et al. An integrated cell atlas of the lung in health and disease. *Nat. Med.* **296**, 1563–1577 (2023).
63. Simpson, E. H. Measurement of diversity [16]. *Nature* **163**, 688 (1949).
64. Dash, P. et al. Quantifiable predictive features define epitope specific T cell receptor repertoires. *Nature* **547**, 89 (2017).
65. Orenbuch, R. et al. arcasHLA: high-resolution HLA typing from RNAseq. *Bioinformatics* **36**, 33–40 (2020).
66. 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).
67. Fleri, W. et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front. Immunol.* **8**, 250318 (2017).
68. Koşaloğlu-Yalçın, Z. et al. The Cancer Epitope Database and Analysis Resource (CEDAR). *Nucleic Acids Res.* **51**, D845–D852 (2023).

69. Joshi, K. et al. Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat. Med.* **2510**, 1549–1559 (2019).
70. González, D. et al. Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. *Blood* **110**, 3112–3121 (2007).
71. Kaiser, M. F. et al. A TC classification-based predictor for multiple myeloma using multiplexed real-time quantitative PCR. *Leukemia* **278**, 1754–1757 (2013).
72. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **6187965**, 598–606 (2023).
73. Türei, D., Korcsmáros, T. & Saez-Rodriguez, J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* **13**, 966–967 (2016).
74. Bauer, M. A. et al. Differential RNA splicing as a potentially important driver mechanism in multiple myeloma. *Haematologica* **106**, 736–745 (2021).

Acknowledgements

This research was funded by Cancer Research UK Early Detection and Diagnosis Programme award No. C9203/A28770). E.M.B is supported by the Wellcome Trust, A.L is supported by Myeloma UK. This work was funded in part by the University College London/UCL Hospitals Biomedical Research Centre. We thank members of the CARDAMON, COSMOS and RADAR teams, the UCL Cancer Trials Centre, the Leeds Clinical Trials Research Unit, and all patients who participated in these studies. We thank the Multiple Myeloma Research Foundation (MMRF), the Perelman Family Foundation, and the International Myeloma Society.

Author contributions

Conceptualisation: B.C., S.A.Q., and K.L.Y. Methodology: E.R., L.A., A.L., E.M.B., L.L., G.W., D.G.F., E.F., A.M., E.J.L., D.J., J.R., M.T., D.M., C.G.K., I.U., G.B., Y.H., C.Z., I.W., and M.C. Investigation: K.A.F. Funding acquisition: L.L., K.R., B.C., S.A.Q., and K.L.Y. Project administration: B.C., S.A.Q., and K.L.Y. Supervision: E.M.B., J.L.R., J.H., B.C., S.A.Q., and K.L.Y. Writing—original draft: K.A.F. Writing—review and editing: K.A.F., E.R., L.A., A.L., E.M.B., L.L., B.C., S.A.Q., and K.L.Y.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-68718-4.

Correspondence and requests for materials should be addressed to Benny Chain, Sergio A. Quezada or Kwee L. Yong.

Peer review information *Nature Communications* thanks P. Leif Bergsagel, Arun Wiita, who co-reviewed with Bonell Patiño-Escobar, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.